

1 **Development of RNAseq methodologies to profile the *in vivo* transcriptome of *Bordetella***
2 ***pertussis* during murine lung infection**

3

4 Short title: *In vivo* RNAseq analysis of *B. pertussis*

5

6 Ting Wong^a, Jesse Hall^a, Dylan Boehm^a, Mariette Barbier^a, and F. Heath Damron^a

7

8

9 ^a Department of Microbiology, Immunology, and Cell Biology, West Virginia University,
10 Morgantown, WV, USA

11

12 *Corresponding author

13 Email: fdamron@hsc.wvu.edu

14

15

16 **Abstract**

17 *Bordetella pertussis* is an obligate human respiratory pathogen that causes the disease whooping
18 cough. A whole cell vaccine (DTP) was developed in the 1940s and was subsequently replaced
19 in the 1990s with a protein-based subunit acellular vaccine (DTaP; tdap). Today, we are
20 observing a resurgence of whooping cough due to evolution of the pathogen and waning vaccine
21 immunity. The use of vaccines decreased the need for basic research on this pathogen. As a
22 result, numerous questions on the basic pathogenesis of *B. pertussis* remain to be answered.
23 Microarrays and more recently, RNA sequencing (RNAseq), have allowed the field to describe
24 the *in vitro* gene expression profiles of the pathogen growing in both virulent and avirulent phases;
25 however, no published studies have described an *in vivo* transcriptome of the pathogen. To
26 address this need, we have designed and evaluated workflows to characterize the *in vivo*
27 transcriptome of *B. pertussis* during infection of the murine lung. During our initial studies, we
28 observed that only 0.014% of the ~100 million 2x50bp illumina reads corresponded to the
29 pathogen, which is insufficient for analysis. Therefore, we developed a simple protocol to filter the
30 bacteria out of the tissue homogenates and separate bacterial cells from the host tissue. RNA is
31 then prepared, quantified, and the *B. pertussis* to host RNA ratio is determined. Here, we present
32 the protocol and discuss the uses and next directions for which this RNAseq workflow can be
33 applied. With this strategy we plan to fully characterize the *B. pertussis* transcriptome when the
34 pathogen is infecting the murine lung in order to identify expressed genes that encode potential
35 new vaccine antigens that will facilitate the development of the next generation of pertussis
36 vaccines.

37

38 Introduction

39 *Bordetella pertussis* is a bacterial respiratory pathogen and the causative agent of whopping
40 cough or pertussis. While whopping cough is a vaccine-preventable disease, the number of cases
41 have risen during the past decade (1). Re-emergence of pertussis is a multi-factorial problem that
42 highlights the need to re-design and improve current vaccination strategies. Historically, whole
43 cell vaccines (WCV), made of formalin inactivated bacteria were first introduced in the 1940s.
44 Following implementation of WCV, efforts by multiple research teams led to the identification of
45 *B. pertussis* major virulence factors, the adenylate cyclase toxin (ACT) and pertussis toxin (PT)
46 (2). The global regulatory system of virulence genes (*Bordetella* virulence genes system or Bvg)
47 was discovered (3, 4) and important surface adhesins such as the filamentous hemagglutinin
48 (FHA), fimbriae, and pertactin (PRN) were identified. This knowledge led to the development of
49 acellular vaccines which are composed of PT, FHA, fimbriae, and PRN (5). Since the
50 implementation of acellular vaccination against pertussis, research efforts on the pathogenesis
51 and basic bacteriology of *Bordetella pertussis* have decreased. As a result, there are still very
52 important gaps in the knowledge of how the bacterium causes infection. These gaps need to be
53 filled to rationally design the next generation of pertussis vaccines.

54 Recent advances in sequencing technology have opened new doors to study bacterial
55 pathogenesis using a global transcriptomics approach. In a previous study, we demonstrated the
56 feasibility of sequencing bacterial RNA in complex samples (6). We were able to isolate both
57 bacterial and host RNA during an acute murine pneumonia, and perform dual-seq to determine
58 both transcriptional profiles during infection. In the present study, we hypothesized that this
59 methodology could be applied for the study of other bacterial pathogens such as *Bordetella*
60 *pertussis*. Here, we describe the first transcriptome of *B. pertussis* during lung infection and
61 identify some of the technical pitfalls of this approach for tissues in which the bacterial burden is
62 low. To circumvent these technical difficulties, we propose a novel methodology to facilitate

63 bacterial RNA recovery from these samples. We expect that this method will be broadly applicable
64 for the study of the transcriptome of bacterial pathogens during infection.

65

66 **Materials and Methods**

67 **Bacterial strains and growth conditions.**

68 *B. pertussis* strain UT25 (UT25Sm1) (7) was cultured on Bordet Gengou (BG) agar (8) (remel)
69 supplemented with 15% defibrinated sheep blood (Hemostat Laboratories) for 48 h at 36°C. *B.*
70 *pertussis* was then transferred from BG plates to three flasks of 12 ml of modified Stainer-Scholte
71 liquid medium (SSM) (9). SSM cultures were not supplemented with cyclodextrin (Heptakis(2,6-
72 di-O-methyl)- β -cyclodextrin). SSM cultures were grown for ~22 h at 36°C with shaking at 180
73 rpm until the OD₆₀₀ reached 0.5 on a 1 cm path width spectrophotometer (Beckman Coulter DU
74 530). The cultures were then diluted to provide a challenge dose of 2×10^7 CFU in 20 μ l. For
75 growth of *Pseudomonas aeruginosa* strain PAO1, *Pseudomonas* Isolation Agar (Difco) was used.
76 *Bordetella bronchiseptica* RB50 and *Escherichia coli* TOP10 were cultured on Lysogeny Agar (10
77 g NaCl, 5 g Yeast Extract, 10 g tryptone). *P. aeruginosa*, *B. bronchiseptica*, and *E. coli* were
78 cultured at 36°C for 18 hr.

79

80 **Murine *B. pertussis* challenge**

81 Outbred CD1 mice were obtained from Charles River and NSG mice (NOD.Cg-Prkdcscid
82 Il2rgtm1Wjl/SzJ; Jackson Labs stock number 005557) raised in house by the WVU Transgenic
83 Animal Core Facility. Mice were anesthetized by intraperitoneal injection of ketamine and xyalzine
84 in saline. Two 10 μ l doses of the *B. pertussis* strain were pipetted directly into each nostril of the
85 mouse. Five to seven mice were infected with strains UT25, and at 1 and 3 days post challenge,
86 mice were euthanized for determination of bacterial burden in the nasal wash, trachea, and lungs.
87 To determine the number of *B. pertussis* in the nares, 1 ml of PBS was flushed up through the

88 nares and collected. Trachea and lungs were extracted and homogenized. Serial dilutions in
89 PBS were plated on BG containing streptomycin (100 µg/ml) to ensure that only UT25 *B. pertussis*
90 were cultured. All murine infection experiments were performed according to protocols approved
91 by the West Virginia University Institutional Animal Care and Use Committee (IACUC) protocol
92 number 14-1211), conforming to AAALAC International accreditation guidelines.

93

94 **Isolation of RNA, library construction, and Illumina sequencing**

95 For the RNA sequencing experiment described in Fig. 2 and 3 of this study, RNA was isolated
96 using RNeasy Mini Kit (Qiagen) as specified by the instructions of the manufacturer. The resulting
97 RNA was treated with RNase-free DNase (Qiagen). To remove the DNase, the samples were
98 then cleaned up on another RNeasy Mini column. The resulting RNA was quantified on a Qubit
99 3.0 fluorometer (ThermoFisher). Next, the RNA integrity was assessed by running the samples
100 on an Agilent BioAnalyzer RNA Pico chip. Due to the fact that lysis of lung tissue causes RNA
101 degradation and that the RNA has both bacterial and murine ribosomal RNA, we observed low
102 RIN numbers of 2-4. However, there was sufficient RNA from 200 to 1000 nucleotides to prepare
103 libraries. Overall, all of the samples were observed to have similar RIN numbers. For comparison
104 with *in vitro* grown *B. pertussis*, we used samples from SSM *in vitro* grown UT25 extracted and
105 analyzed by RNAseq using an identical procedure previously published by our laboratory (10).
106 These *In vitro* grown *B. pertussis* UT25 RNA samples had RIN numbers of 9-10. For library
107 preparation, illumina Scriptseq complete gold (epidemiology; for mouse/bacteria ribosomal
108 depletion) was used. Resulting libraries passed standard illumina quality control PCR and were
109 sequenced on an illumina Hiseq 1500 by the Marshall University Genomics Core. Three
110 biological samples were pooled into 3 technical samples (9 total mice).

111

112 **RNAseq and bioinformatics analyses**

113 The reads were aligned to the *B. pertussis* Tohama I genome (11) using CLC Genomics
114 Workbench version 9.5. Reads per kilobase per million (RPKM) and fold change for each gene
115 were calculated. Approximately 180,000-200,000 reads mapped per 100 million 2x50bp samples.
116 Upon inspection of the reads, we observed cross mapping reads that are likely due to the
117 presence of residual murine RNA in the sample. To avoid non-specific mapping and remove
118 these reads, we increased mapping stringency to 100% identity and extracted only reads that
119 mapped as pairs and counted them as fragments. These manipulations removed ~90% of the
120 total reads. Due to the low number of total reads obtained from the lungs of infected mice, we
121 sampled *in vitro* grown RNAseq reads from 10 million 2x50bp down to 15,000 total reads. Fold
122 change was calculated by comparison of the *in vitro* UT25 *B. pertussis* reads to the *in vivo* NSG
123 mouse UT25 reads. The raw data reads will soon be submitted to the Sequence Read Archive
124 (SRA). If requested, we would also be happy to provide the raw reads directly.

125

126 **Filtration of *B. pertussis* from culture and infected murine lungs.**

127 To determine the percent recovery of bacteria from *in vitro* cultures, *B. pertussis*, *B.*
128 *bronchiseptica*, *E. coli*, and *P. aeruginosa* were grown in the culture conditions described above.
129 Various amounts from OD₆₀₀ 0.3 to OD₆₀₀ 0.003 were filtered through 5 µm syringe filters
130 (Sartorius Minisart part ref 17594 Cellulose Acetate). The input culture and outflow were plated
131 on agar plates (BG or LA) to determine the amount of input and flow through bacteria. To isolate
132 bacteria from the infected mouse lungs, mice were dissected and the lungs were extracted and
133 placed into 1 ml of PBS. The lungs were then homogenized with glass Dounce tissue grinders
134 (Sigma-Aldrich). The homogenate was then strained through a 70 µm nylon cell strainer (VWR).
135 The filtrate was then passed through a 5 µm syringe filter. The suspension was next pelleted by
136 16,100 x g for 4 min in 1.5 ml Eppendorf tubes. The supernatant was then discarded, RNAprotect
137 bacterial reagent (Qiagen) was added to the pellet, and samples were stored at -80°C.

138

139 **RNA isolation and reverse transcriptase PCR (qRT-PCR).**

140 For qRT-PCR analysis, cells stored at -80°C in RNAlater (Qiagen) were lysed with lysozyme,
141 and RNA was isolated using the RNeasy method (12). Briefly, cell pellet were resuspended in
142 RNA extraction solution (18mM EDTA, 0.025% SDS, 1% 2-mercaptoethanol, 95% formamide) by
143 vortexing vigorously. Samples were incubate at 95°C for 7 min and pelleted by centrifugation at
144 16,000g for 5 min at room temperature. The supernatant containing RNA was then pipetted into
145 a fresh tube without disturbing the clear gelatinous pellet, and DNA and RNA concentrations were
146 measured using a Qubit 3.0 fluorometer. After extraction, DNA was digested using an off-column
147 digestion with RNase-free DNase and re-isolated with another RNeasy column. RNA
148 concentration and quality was assessed on a Molecular Devices i3 Spectramax Spectra drop
149 plate and on a Qubit 3.0 fluorometer. To ensure RNA was DNA-free, 25 ng of RNA was checked
150 by PCR amplification and was only used for cDNA if no amplicon was observed and a C_T of >32.
151 cDNA was synthesized using M-MLV reverse transcriptase (Promega) per the manufacturer's
152 instructions using 200 ng of RNA and gene specific reverse primers for targets. Twenty five
153 microliter qPCR mixtures were setup with Excella SYBR Green PCR master mix (Worldwide
154 Medical Products), per manufacturer's instructions using 1 μ l of cDNA. A minimum of three
155 technical replicate reactions were ran per gene target per sample on a Step One Plus qPCR
156 thermocycler (Applied Biosystems). Primers were designed on Primer3 (Primer-Blast; NCBI) and
157 checked for specificity by PCR. Melt curve analysis as well as subsequent agarose gel
158 electrophoresis were performed on all reactions. Gene expression was normalized to the *rpoB*
159 reference using the $2^{-\Delta\Delta C_T}$ method (13). For statistical analysis, the ΔC_T for the three biological
160 replicate experiments was calculated and a Student's *t*-test was performed using Microsoft Excel
161 2013. Standard error of mean was calculated based on the variability of the ΔC_T of three biological
162 replicates. For the *rpoB* to *Gapdh* absolute quantitation PCR, amplicons of *rpoB* and *Gapdh* were
163 generated, purified and quantified. The number of copies per ng of DNA was determined and
164 standard curves were generated for *rpoB* (5×10^8 copies to 500) and *Gapdh* (4.4×10^8 to 440

165 copies) with three technical replicates. Microsoft Excel was used to plot standard curves based
166 on the CTs of the known standards described above and number of copies in each sample was
167 calculated using a linear regression with R^2 values of 0.999 for both *rpoB* and *Gapdh*. The ratio
168 of *Gapdh* to *rpoB* RNA was used to estimate the relative amount of host to pathogen RNA in each
169 sample. The following primers sequences used in this study were described by Bibova *et al.* (14):
170 *cyaF*(CGAGGCGGTCAAGGTGAT), *cyaR*(GCGGAAGTTGGACAGATGC),
171 *ptxAF*(CCAGAACGGATTCACGGC), *ptxAR* (CTGCTGCTGGTGGAGACGA), *bvgAF*
172 (AGGTCATCAATGCCGCCA), *bvgAR* (GCAGGACGGTCAGTTCGC), *fhaBF*
173 (CAAGGGCGGCAAGGTGA), *fhaBR* (ACAGGATGGCGAACAGGCT), *rpoBF*
174 (GCTGGGACCCGAGGAAAT), *rpoBR* (CGCCAATGTAGACGATGCC). BP2497 were designed
175 in a previous study: BP2497F (TCGGATCGCACCAATTACTTC) and BP2497R
176 (CCTTGGCGATCAGCGAGTT) (10). The following primers were used for *Gapdh*: *gapdhf*
177 (CATGGCCTTCCGTGTTCT) and *gapdhr* (GCGGCACGTCAGATCCA).

178

179 **Results and Discussion**

180 **Establishment of a murine model to perform *in vivo* transcriptomic analysis of *B. pertussis*.**

181 The main objective of this study was to perform RNA sequencing on *B. pertussis* during lung
182 infection. Our previous studies with the respiratory pathogen *P. aeruginosa* demonstrated the
183 feasibility of sequencing both bacterial and host pathogen during infection. However, they also
184 highlighted the need to use high bacterial loads (10^8 to 10^9 CFU/organ) to obtain sufficient RNA
185 for purification and analysis. In addition, our previous attempts to perform RNAseq from CD1
186 mouse trachea yielded very few reads of bacterial RNA, demonstrating the challenge of these
187 types of workflows. Therefore, in this work, we first optimized the murine model of pertussis to
188 maximize the recovery of bacteria during infection. To this end, we performed intranasal infections
189 of outbred CD1 mice (traditional model used for vaccine development) and in immunodeficient
190 NSG mice. We hypothesized that immunodeficient mice would be unable to clear bacterial

191 infection and have higher bacterial burdens in the airways which would increase the recovery rate
192 of bacterial RNA for sequencing. Mice were infected with 10^7 CFU/mouse (CD1) or 4×10^6
193 CFU/mouse (NSG) and euthanized 24 or 72 hr post infection. Bacterial burden was determined
194 in the nasal passages by performing a nasal wash, and in the trachea and lung by generating
195 tissue homogenates. As expected, significantly higher bacterial burden was observed in the nasal
196 wash, trachea, and lung of NSG mice compared to CD1 mice (Fig. 1). The bacterial burden in the
197 nasal wash and lung remained constant between day 1 and 3 in CD1 mice, and even significantly
198 decreased in the trachea (Fig. 1). However, bacterial burden significantly increased over time in
199 NSG infected mice in both the upper and the lower respiratory airways (Fig 1). Overall, the lung
200 samples of NSG mice at the later time point had the highest bacterial burden. Therefore, for
201 subsequent experiment, the lungs from NSG mice infected with *B. pertussis* strain UT25 were
202 harvested two days post infection and used for RNA preparation for NGS.

203
204 **RNAseq analysis of NSG lungs infected with *B. pertussis*.** To perform dual RNAseq (analyze
205 both host and pathogen), RNA was isolated from the infected NSG lungs per the standard RNeasy
206 column based kit. Illumina Scriptseq libraries were prepared and sequenced on a Hiseq 1500.
207 100 million 2x50bp reads were devoted to each sample which contained 3 pooled *B. pertussis*
208 infected NSG lungs. Approximately 200,000 reads mapped to the *B. pertussis* Tohama I reference
209 genome for each sample. Upon inspection of the reads, it was apparent that the mouse RNA of
210 the sample was cross mapping to the bacterial genome. This non-specific cross-mapping was
211 due to sequence similarities between the bacterial and the murine genomes. In order to focus on
212 only true *B. pertussis* reads, we extracted only the reads that mapped with 100% identity in pairs.
213 This removed the false positive mappings as observed in Fig. 2. However, this procedure also
214 significantly decreased the number of *B. pertussis* reads. Therefore, we sampled the *in vitro* UT25
215 reads down to 15,000 to allow for a more accurate comparison to the NSG mouse *B. pertussis*
216 reads. Overall we observed that only 0.014% of the total reads corresponded to the transcriptome

217 of *B. pertussis*, which is insufficient for high quality transcriptome analysis. The murine reads
218 greatly outnumbered the pathogen reads (Table 1). We were able to detect up to 828 genes with
219 at least 1 true *B. pertussis* read. When we compared the *in vitro* and *in vivo* *B. pertussis* reads
220 (Table 2), we observed promising data suggesting that the changes in gene expression observed
221 in these samples are relevant for the pathogenesis of *B. pertussis*. We performed String analysis
222 to better understand the associations between the genes differentially regulated between the two
223 datasets (Fig. 3). In this analysis, we observed the alcaligin siderophore biosynthesis genes in
224 the infected NSG lungs (Table 2 and Fig. 3A). Alcaligin is a siderophore that sequesters iron for
225 the pathogen and is an important factor required for infection. We also detected *fhaB* in the
226 infected lung tissue. FhaB is an adhesin that allows *B. pertussis* to adhere to the airway during
227 infection. In previous experiments we observed that *fhaB* is one of the most abundant transcripts
228 in the transcriptome (10). We also detected both pertussis and adenylate cyclase toxin transcripts
229 in the infected lung tissue (Table 2). Overall in this limited analysis, *B. pertussis* gene expression
230 was higher *in vitro*. It is likely that *B. pertussis* grows faster *in vitro* and is transcriptionally more
231 active (Fig. 3B). From this analysis, we concluded that method optimization to enrich the RNA of
232 *B. pertussis* was required in order to characterize the *in vivo* transcriptome of the bacterium in the
233 mouse.

234

235 **Method optimization for the extraction of bacterial RNA using differential filtration.** Despite
236 murine model optimization to maximize bacterial RNA recovery the total number of reads from
237 bacterial RNA remained low. To increase bacterial sequence coverage without rising sequencing
238 cost by increasing the total number of sequencing lanes and total reads, we developed a novel
239 methodology to separate bacterial from eukaryotic RNA. This method is based on the difference
240 in size between bacterial and eukaryotic cells present in lung homogenates. We first performed
241 method optimization using *in vitro* grown bacterial samples of *E. coli*, *P. aeruginosa*, *B.*
242 *bronchiseptica*, and *B. pertussis*. *In vitro* grown cultures of various bacterial pathogens were

243 filtered through a 5 μm filter, and plated before and after filtration to estimate percentage recovery
244 of the bacteria. Overall, 30 to 60% of the bacteria present in the suspension were recovered after
245 filtration (Figure 4). Recovery rate was different between the species, with *B. bronchiseptica*
246 showing the highest recovery rate and *P. aeruginosa* the lowest. These differences are potentially
247 due to the variation in size of these bacteria, *P. aeruginosa* cells measuring up to 3 μm in length,
248 while *Bordetellae* cells measure less than 0.7 μm in length. From these results, we concluded that
249 5 μm filters could be used to recover a significant number of bacteria from suspensions.

250

251 **Method optimization for the extraction of bacterial RNA from infected tissue.** In the next
252 phase of our protocol development, we applied filtration to homogenized tissue to separate
253 bacterial cells from the rest of the homogenate. Infected tissues were first gently homogenized
254 using a mechanical Dounce glass homogenizer to separate the cells in the lung while causing
255 minimal cell lysis compared to enzymatic digestion. Homogenates were then filtered through a 70
256 μm filter to remove large clumps and debris. The resulting filtrate was pushed through a 5 μm
257 filter to remove eukaryotic cells. This second filtrate, containing bacteria and residual eukaryotic
258 DNA released from cell lysis during homogenization, was then centrifuged to pellet bacteria. The
259 bacterial pellet was finally re-suspended in RNA protect and processed for extraction (Fig. 5).
260 Bacterial loads in nasal wash, trachea and lung were quantified in the tissue homogenate and in
261 the final bacterial pellet after filtration (Fig. 6). From these data, we estimated the range of viable
262 bacteria needed for RNAseq analysis. We estimate that RNA from 10^4 *B. pertussis* CFU is the
263 minimum that can be used for *in vivo* seq (assuming 100% of the sample is pathogen). This
264 estimation is based on our filtration method and RNA purification using RNAsnap, and assumes
265 that the RNA obtained is DNA-free. For samples in which these yields cannot be achieved, low
266 input kits such as the Nugen Ovation RNAseq V2 library preparation kit can be used to perform
267 the analysis with picogram amounts of RNA.

268

269 **qPCR estimation of *rpoB* to *Gapdh* transcript levels.** To sequence the transcriptome of *B.*
270 *pertussis* during infection, it is important to have a high bacteria to host RNA ratio in the sample.
271 To estimate this ratio, we performed absolute qPCR to count the numbers of transcripts in cDNA
272 samples from the *B. pertussis* infected NSG mice (Fig. 7). We added this quality control metric
273 to select samples with high amounts of bacterial RNA that have the best probability of successful
274 RNAseq analysis. Standard curves of *rpoB* and *Gapdh* templates were generated and used to
275 calculate *rpoB* and *Gapdh* transcript copy numbers. We observed that not all samples had the
276 same pathogen to host copy ratios (Fig. 7). However, the data suggested that in most samples
277 there was a 1:4 ratio of *rpoB* to *Gapdh*. We also filtered a *B. pertussis* infected CD1 lung and
278 observed a similar *rpoB* to *Gapdh* ratio. This promising data suggests that we can obtain sufficient
279 RNA, even from mice that are not highly susceptible.

280

281 **qRT-PCR analysis of *in vitro* and *in vivo* grown *B. pertussis*.** To confirm that this method
282 could allow detection of bacterial transcript during infection, we performed qRT-PCR analysis on
283 *B. pertussis* RNA samples proceeding from SSM *in vitro* cultures and RNA from *B. pertussis*
284 infected NSG lung samples (Fig. 6 and 8). We selected the *cya* and *ptx* genes that encode the
285 major toxins as well as *fhaB*, *bvgA*, and *BP2497*. We hypothesized that the expression of the
286 genes encoding these toxins would be more highly expressed in the mouse compared to *in vitro*.
287 We observed increased *cya* and *ptx* gene expression in all of our NSG mice (Fig. 8). Expression
288 of *fhaB*, the major adhesin, was decreased *in vivo* compared to *in vitro*. The genes *cya* and *ptx*
289 are classified as “class II Bvg system genes” (3) whereas *fhaB* is a “class I gene”. Class I genes
290 are usually expressed at different levels than class II genes, which could explain why *cya*, *ptx* and
291 *fhaB* are differentially regulated. In addition, it is possible that the bacteria tightly adhered to
292 epithelial cells during infection are not collected using this purification method, which would
293 introduce bias and explain differences in gene expression involved in bacterial adhesion. This
294 caveat will be further investigated as the project moves forward. At this phase, we are most

295 interested in successfully performing the feasibility of performing *in vivo* RNAseq on *B. pertussis*
296 in the murine lung.

297

298

299 **Conclusions**

300 In this report, we have detailed our progress on developing RNAseq workflows to characterize
301 the transcriptome of *B. pertussis* infecting the murine lung. Bacterial pathogens are small yet can
302 cause substantial harm to the host. *B. pertussis* releases toxins which block or suppress both the
303 innate and adaptive immune responses. These toxins are essential to infection but little is known
304 about the roles of the majority of the genes encoded in the genome of *B. pertussis*. We
305 hypothesize that the *B. pertussis* transcriptome *in vivo* will be significantly different than the *in*
306 *vitro* transcriptome in media such as on BG or SSM. It is possible genes that are highly expressed
307 during infection could be exploited as vaccine antigens. By answering the simple question of
308 “what does the pathogen express when it infects?”, we believe that significant advancement can
309 be made which will potentially impact vaccines and treatments. Our initial attempts showed that
310 performing dual-seq on a pathogen that is both growing slowly and is present in at low levels in
311 tissue samples is challenging. These technical difficulties led us to re-formulate our ideas and
312 come up with an innovate strategy. By combining a novel filtration method to separate the bacteria
313 from host tissue and exhaustive qRT-PCR quality controls, we have obtained RNA with a
314 pathogen to host ration high enough to perform RNAseq in the near future. Next generation
315 sequencing applications such as RNAseq have greatly expanded over the past few years but we
316 believe there is still a need to further develop RNA preparation workflows to truly take advantage
317 of these technologies. There are many other bacterial pathogens for this type of methodology
318 could potentially be applied. We envision that this strategy could also be used in other models
319 systems such as the baboon model of pertussis. We hypothesize that depending on the lower
320 limits of detection, it may also be possible to isolate pathogens directly from human patients and

321 determine the “human specific” transcriptomes of the pathogen. Our current goal is to fully
322 characterize the transcriptome of the human pathogen *B. pertussis* and we will also seek to apply
323 this strategy to other pathogens and extend our knowledge of how pathogens infect.

324

325

326

327 **Funding Information**

328 This work was supported by funding from National Institutes of Health HHSN272201200005C-
329 416476 and laboratory startup funds from West Virginia University to F.H.D. The WVU and
330 Marshall University CORE facilities were funded by the WV InBRE grant: GM103434.

331

332 **Acknowledgements**

333 *B. pertussis* strains UT25 was kindly provided by Dr. Sandra Armstrong (University of Minnesota).
334 We would like to thank the WVU Transgenic Animal Core Facility for providing the NSG mice for
335 this study.

336

337 **Author contributions**

338 T.W. and J.H. performed murine experiments, prepared RNA, validated RNA for RNAseq,
339 performed qRT-PCR, analyzed data, and composed the manuscript. D.T.B performed murine
340 experiments analyzed data, and composed the manuscript. M.B. designed / performed murine
341 experiments, analyzed data, and composed the manuscript. F.H.D designed / performed
342 experiments (murine infection model, RNAseq, and etc), analyzed data, and composed the
343 manuscript.

344

345

346

347 **References**

- 348 1. **Allen A.** 2013. Public health. The pertussis paradox. *Science* **341**:454-455.
- 349 2. **Hewlett EL, Burns DL, Cotter PA, Harvill ET, Merkel TJ, Quinn CP, Stibitz ES.** 2014.
- 350 Pertussis pathogenesis--what we know and what we don't know. *J Infect Dis* **209**:982-
- 351 985.
- 352 3. **Melvin JA, Scheller EV, Miller JF, Cotter PA.** 2014. *Bordetella pertussis* pathogenesis:
- 353 current and future challenges. *Nat Rev Microbiol* **12**:274-288.
- 354 4. **Smith AM, Guzman CA, Walker MJ.** 2001. The virulence factors of *Bordetella pertussis*:
- 355 a matter of control. *FEMS Microbiol Rev* **25**:309-333.
- 356 5. **Sato Y, Sato H.** 1999. Development of acellular pertussis vaccines. *Biologicals* **27**:61-69.
- 357 6. **Damron FH, Oglesby-Sherrouse AG, Wilks A, Barbier M.** 2016. Dual-seq
- 358 transcriptomics reveals the battle for iron during *Pseudomonas aeruginosa* acute murine
- 359 pneumonia. *Scientific Reports* **6**.
- 360 7. **Brickman TJ, Armstrong SK.** 1996. The ornithine decarboxylase gene *odc* is required
- 361 for alcaligin siderophore biosynthesis in *Bordetella* spp.: putrescine is a precursor of
- 362 alcaligin. *J Bacteriol* **178**:54-60.
- 363 8. **Bordet J, O G.** 1906. Le microbe de la coqueluche. *Ann Inst Pasteur (Paris)* **20**:731-741.
- 364 9. **Stainer DW, Scholte MJ.** 1970. A simple chemically defined medium for the production
- 365 of phase I *Bordetella pertussis*. *J Gen Microbiol* **63**:211-220.
- 366 10. **Barbier M, Boehm DT, Sen-Kilic E, Bonnin C, Pinheiro T, Hoffman C, Gray M, Hewlett**
- 367 **E, Damron FH.** 2017. Modulation of Pertussis and Adenylate Cyclase Toxins by Sigma
- 368 Factor RpoE in *Bordetella pertussis*. *Infect Immun* **85**.
- 369 11. **Parkhill J, Sebahia M, Preston A, Murphy LD, Thomson N, Harris DE, Holden MT,**
- 370 **Churcher CM, Bentley SD, Mungall KL, Cerdeno-Tarraga AM, Temple L, James K,**
- 371 **Harris B, Quail MA, Achtman M, Atkin R, Baker S, Basham D, Bason N, Cherevach**
- 372 **I, Chillingworth T, Collins M, Cronin A, Davis P, Doggett J, Feltwell T, Goble A,**
- 373 **Hamlin N, Hauser H, Holroyd S, Jagels K, Leather S, Moule S, Norberczak H, O'Neil**
- 374 **S, Ormond D, Price C, Rabinowitsch E, Rutter S, Sanders M, Saunders D, Seeger**
- 375 **K, Sharp S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Unwin L, et**
- 376 **al.** 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*,
- 377 *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* **35**:32-40.
- 378 12. **Stead MB, Agrawal A, Bowden KE, Nasir R, Mohanty BK, Meagher RB, Kushner SR.**
- 379 2012. RNAsnap (TM): a rapid, quantitative and inexpensive, method for isolating total RNA
- 380 from bacteria. *Nucleic Acids Research* **40**.
- 381 13. **Livak KJ, Schmittgen TD.** 2001. Analysis of relative gene expression data using real-
- 382 time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**:402-408.
- 383 14. **Bibova I, Skopova K, Masin J, Cerny O, Hot D, Sebo P, Vecerek B.** 2013. The RNA
- 384 chaperone Hfq is required for virulence of *Bordetella pertussis*. *Infect Immun* **81**:4081-
- 385 4090.

386

387

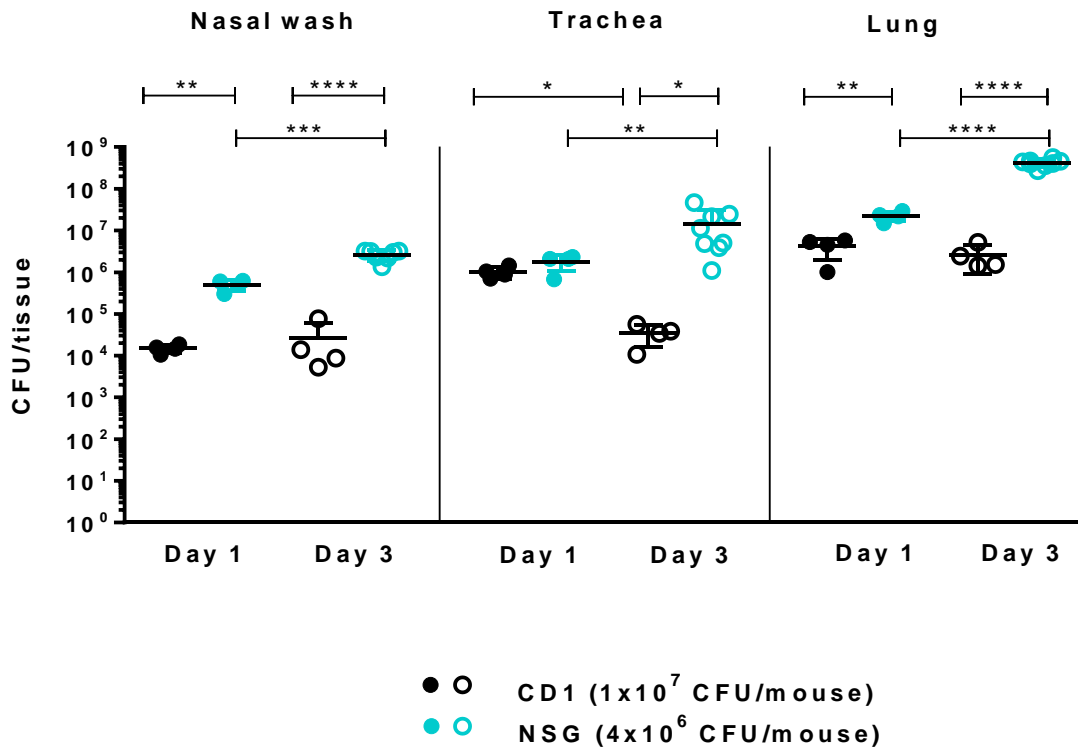
388

389

390

391 **Figures and Legends**

392



393

394 **Figure 1. *B. pertussis* burden in outbred CD1 mice compared to highly sensitive NSG mice.**

395 NSG mice were infected with 4x10⁶ CFU and CD1 mice were infected with 2x10⁷ CFU of the *B.*

396 *pertussis* strains UT25. At 1 and 3 days post infection the bacterial burdens in the nasal wash,

397 trachea, and lungs were determined. Four to eight mice were used for each time-point. Groups

398 were compared independently using a *t*-test with Tukey correction (* *p*<0.05; ** *p*<0.01; ***

399 *p*<0.001; **** *p*<0.0001).

400

401



402

403 **Figure 2. Visualization of the *bvgRSA* and *fhaB* loci mapped reads.** Reads were mapped to
404 the *B. pertussis* genome (1st pass mapping) and cross-mapping murine RNA was observed. The
405 stringency of identify was increased (2nd pass mapping) and only paired reads were considered.
406 In the blue boxes, examples of cross mapping reads are shown.

407

408

409

410

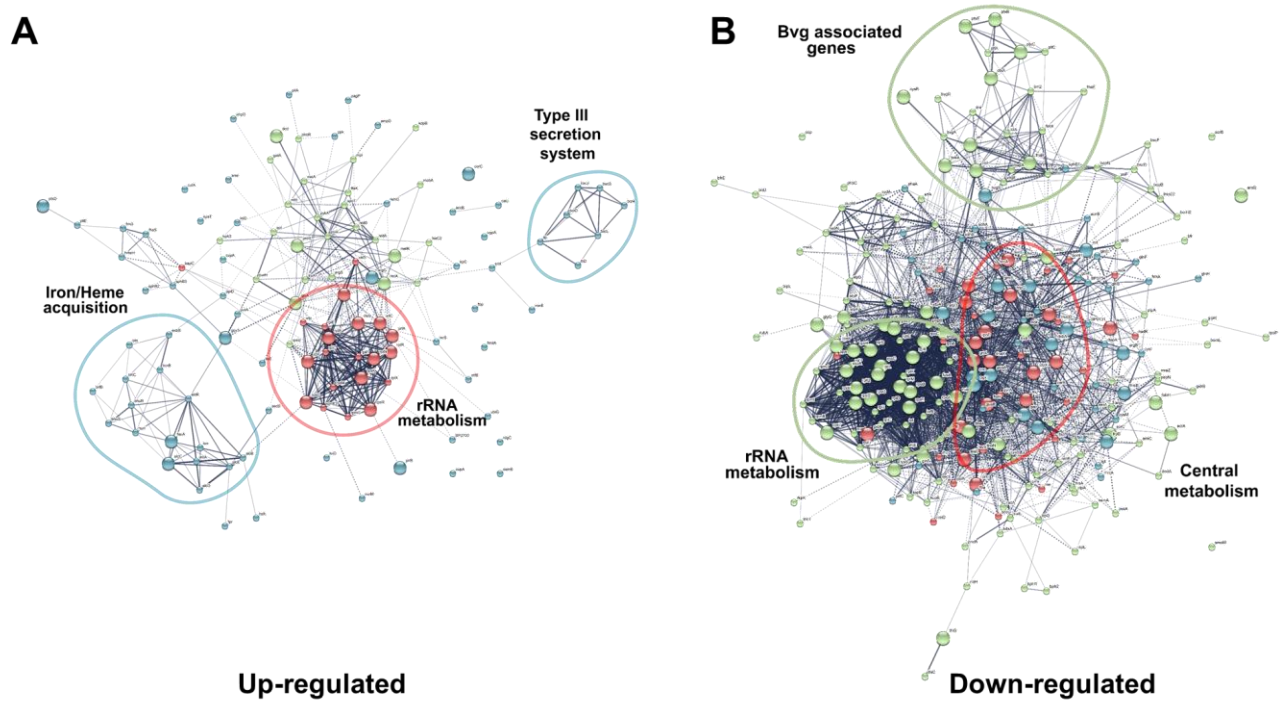
411

412

413

414

415



416

417

418

419

420

421

422

423

424

425

426

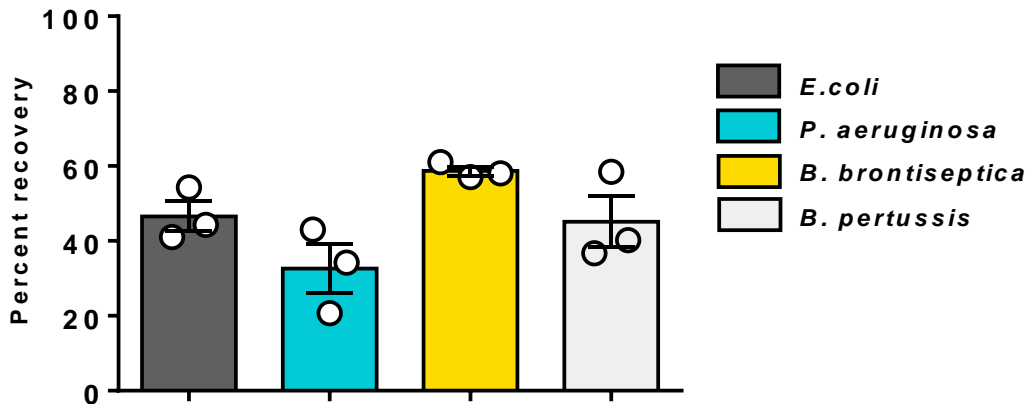
427

428

429

430

Figure 3. String analysis of differentially regulated genes during infection. Up- (A) and down-regulated (B) annotated *B. pertussis* genes in the lung compared to *in vitro* were analyzed and represented using the String database (REF) and kmeans clustering.



431

432 **Figure 4. Recovery of bacteria after syringe filtration through a 5 µm filter.** Percent recovery

433 of the bacteria present in a suspension prepared from *in vitro* grown cultures after filtration with a

434 5 µm filter. Each bar shows the mean percentage recovery of bacteria from suspensions at

435 densities ranging from OD₆₀₀ 0.3 to OD₆₀₀ 0.003 to illustrate that the recovery rate does not depend

436 on bacterial suspension density but on bacterial size.

437

438

439

440

441

442

443

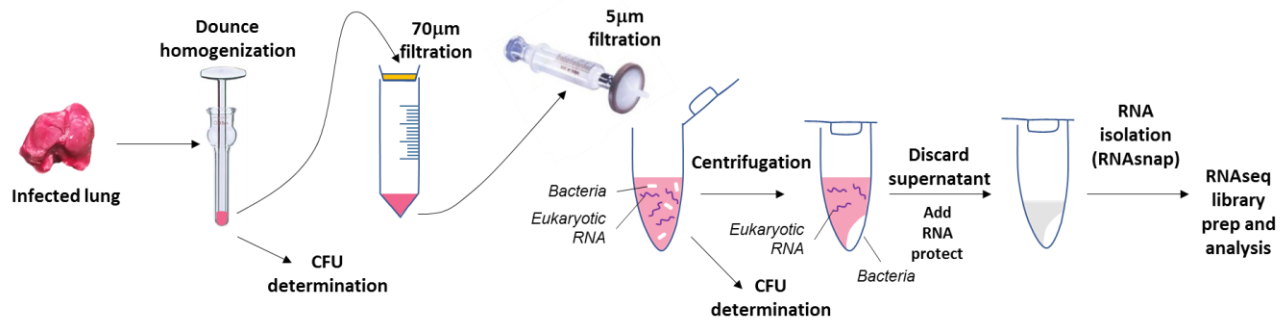
444

445

446

447

448



449

450 **Figure 5. Isolation of *B. pertussis* from infected murine lung for preparation of RNA for *in***

451 ***vivo* RNAseq.** Samples are homogenized using a Dounce glass homogenizer and tissue debris

452 are removed by filtration through a 70 µm filter. Eukaryotic cells are excluded by 5 µm filtration.

453 Bacteria are isolated by centrifugation and contaminating murine RNA (due to lysis of cells) is

454 removed. Bacterial cells are stabilized by RNAprotect until RNAsnap RNA isolation.

455

456

457

458

459

460

461

462

463

464

465

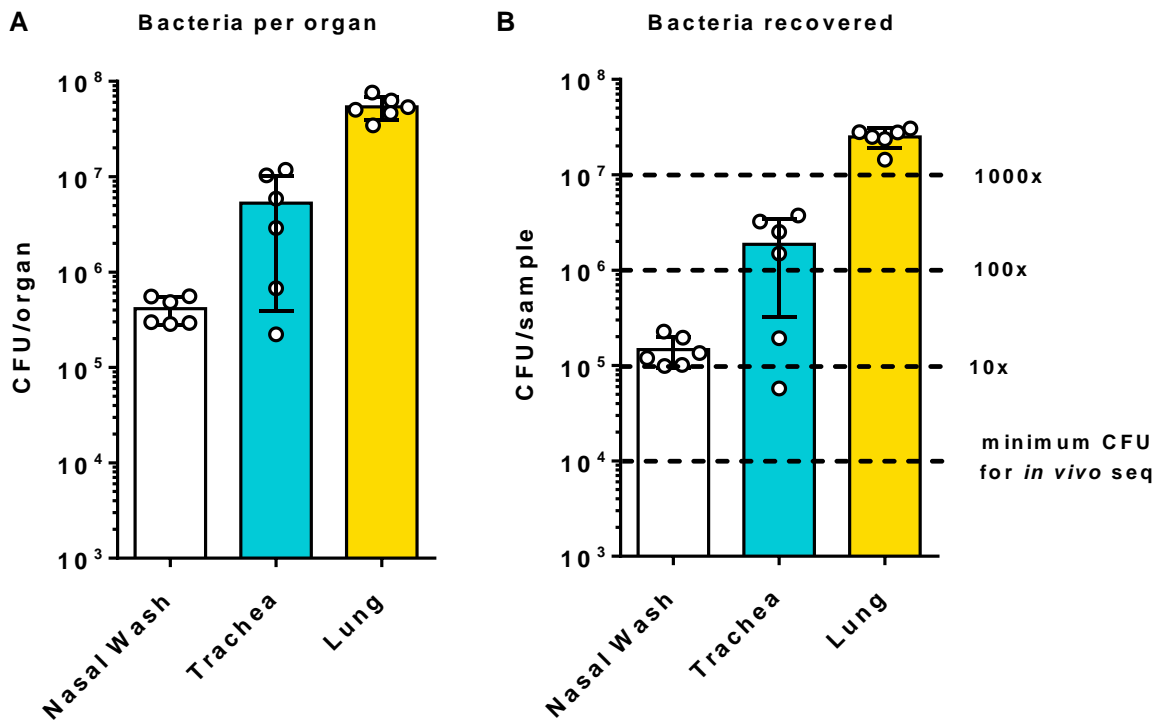
466

467

468

469

470



471

472 **Figure 6. Isolation of *B. pertussis* from infected NSG mouse lungs.** NSG mice were infected
473 with *B. pertussis* strain UT25. At two days post infection, the tissues were collected and the
474 number of input and output (post filtration) were quantified. A. Bacterial load in each tissue after
475 homogenization. B. Number of bacteria recovered from each tissue after implementing the
476 filtration strategy described in Fig. 5. The estimated number of bacteria needed to perform
477 RNAseq assuming 100% bacteria RNA in the sample are shown using dashed lines.

478

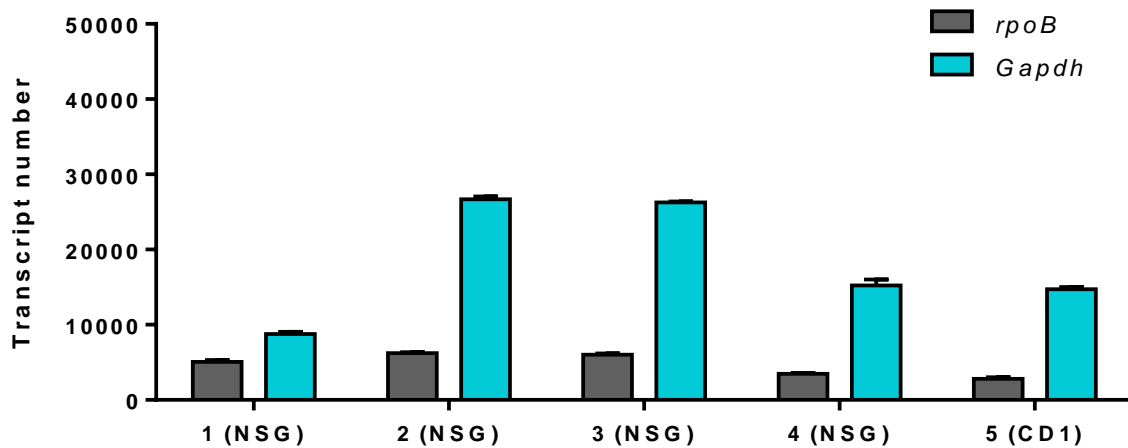
479

480

481

482

483



484

485 **Figure 7. Absolute quantification of *rpoB* and *Gapdh* transcripts in the filtered *B. pertussis***

486 **infected NSG lungs.** Standard curves of known amounts of *rpoB* or *Gapdh* targets were

487 established and used to calculate the number of transcripts of each gene in cDNA from the lung

488 of *B. pertussis* infected NSG mice. Four NSG mice were analyzed along with one outbred CD1

489 mouse. The CFU present in the lung of the NSG mice are shown in Figure 6A. The CD1 mouse

490 had 10^7 viable *B. pertussis* when the lung was harvested. These absolute qPCR methods also

491 provides a quality control metric to validate samples before proceeding to RNAseq library

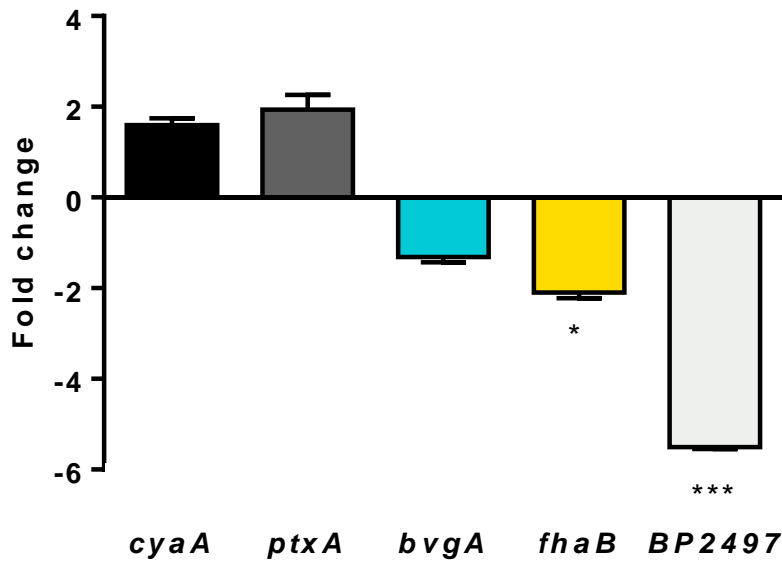
492 preparation.

493

494

495

496



497

498 **Figure 8. qRT-PCR analysis of *B. pertussis* growing *in vitro* compared to *in vivo* in the NSG**

499 **lung.** *B. pertussis* strain UT25 was grown in SSM liquid media for control comparison. The NSG

500 lung samples prepared by the filtration protocol were compared to the SSM control data. Genes

501 encoding the adenylate cyclase toxin (*cya*) and the pertussis toxin (*ptxA*) were increased *in vivo*

502 compared to *in vitro*. The *bvgA* and *fhaA* genes were slightly decreased in expression. *BP2497*

503 encodes a putative protease and it was also decreased in the murine lung. Data from three

504 biological replicates with three technical replicates each. Standard deviations are calculated from

505 variations between biological replicates and compared using a *t*-test (* $p < 0.05$; *** $p < 0.001$).

506

507

508

509

510 **Tables**

511 **Table 1. Summary of RNA-seq gene reads and mapping.**

Sample	Total reads in sample	% of reads mapped to <i>B. pertussis</i> genome	Reads mapped to <i>B. pertussis</i> genome	Number of genes with 10 reads	Number of genes with 1 read	% of the genes in the genome with at least 1 read
NSG mouse lung infected with UT25	109,878,438	0.01365	14,995	69	828	21.47
<i>B. pertussis</i> UT25 growing in SSM	15,000	97.35	14,602	94	1,343	34.48

512

513 **Table 2.**

Gene	gene product	Fold change	SSM reads	SSM RPKM	NSG reads	NSG RPKM
Genes of interest and virulence factors						
<i>alcA</i>	Alcaligin biosynthesis enzyme	∞	0	0	27	1256
<i>alcB</i>	Alcaligin biosynthesis protein	∞	0	0	16	1702
<i>alcC</i>	Alcaligin biosynthesis protein	∞	0	0	12	417
<i>alcE</i>	Putative iron-sulfur protein	∞	0	0	1	80
<i>bcr4</i>	Siderophore [Alcaligin] translocase AlcS	11	2	525	47	5804
<i>bcrD</i>	Type III secretion inner membrane channel Protein (LcrD,HrcV,EscV,SsaV)	-2	8	491	9	260
<i>bfrC</i>	Ferrichrome-iron receptor	∞	0	0	12	353
<i>bfrI</i>	Ferrichrome-iron receptor	∞	0	0	12	367
<i>brkA</i>	Serum resistance protein	-10	23	1039	5	106
<i>bvgA</i>	Virulence factors transcription regulator	-17	8	1739	1	102
<i>bvgS</i>	Hybrid sensory histidine kinase	-20	19	700	2	35
<i>cyaA</i>	Hemolysin-adenylate cyclase precursor	-3	13	348	11	138
<i>fhaB</i>	Filamentous hemagglutinin/adhesin	-7	275	3496	83	497
<i>fhaL</i>	Adhesin	-4	9	98	5	26
<i>fhaS</i>	Adhesin	6	1	18	12	101
<i>fim2</i>	Serotype 2 fimbrial subunit precursor	1	1	216	3	306
<i>fim3</i>	Serotype 3 fimbrial subunit precursor	3	17	3786	96	10063
<i>groEL</i>	Heat shock protein 60 family chaperone GroEL	-3	67	5582	45	1765
<i>ompA</i>	Outer membrane protein A precursor	-28	66	15532	5	554
<i>prn</i>	Pertactin precursor					
<i>ptxA</i>	Pertussis toxin subunit 1 precursor	-3	16	2706	11	875
<i>ptxB</i>	Pertussis toxin subunit 2 precursor	-6	12	2414	4	379
<i>ptxC</i>	Pertussis toxin subunit 3 precursor	-2	10	2002	14	1319

<i>ptxD</i>	Pertussis toxin subunit 4 precursor	1	5	1492	13	1826
<i>ptxE</i>	Pertussis toxin subunit 5 precursor	-11	16	802	3	71
<i>ssrA</i>	Transfer-messenger RNA (tmRNA)	-20	1094	388193	114	19039
<i>pcfA</i>	Tracheal colonization factor precursor	-7	38	2677	11	365
<i>tuf_1</i>	Translation elongation factor Tu	-1	22	2530	33	1786
<i>vag8</i>	Autotransporter	-5	46	2293	19	446
rRNAs						
<i>BPr01</i>	SSU rRNA	82	17	1520	2959	124514
<i>BPr02</i>	LSU rRNA	35	81	3850	6103	136515
<i>BPr05</i>	LSU rRNA	4	77	3659	638	14271
<i>BPr06</i>	SSU rRNA	15	18	1609	573	24112
<i>BPr08</i>	LSU rRNA	3	80	3802	583	13041
<i>BPr09</i>	SSU rRNA	17	17	1520	596	25080
<i>BPs02</i>		-14	285	94289	42	6540
Hypothetical genes						
<i>BP0168</i>	Putative membrane protein	14	1	152	30	2142
<i>BP0840</i>	Putative membrane porin protein precursor	-46	303	35839	14	779
<i>BP2374</i>	Hypothetical gene	∞	0	0	42	13675
<i>BP2410</i>	Hypothetical gene	19	1	102	40	1926
<i>BP3084</i>	Putative metal chaperone, Zn homeostasis	3	10	1227	61	3524
<i>BP3402</i>	Zinc metalloprotease YfgC precursor	∞	0	0	94	3992

514

515

516

517

518