

1 **Title: The upper respiratory tract microbiome of indigenous Orang Asli in north-**
2 **eastern Peninsular Malaysia**

3

4 **Authors:** Cleary D. W. *^{1,2}, Morris D. E.¹, Anderson R. A.¹, Jones J.¹, Alattraqchi A. G.³,
5 Rahman N. I. A.³, Ismail S., Razali M. S.³, Mohd Amin R.³, Abd Aziz A.³, Esa N. K.³,
6 Amiruddin S.³, Chew C. H.⁴, Amat Simin M. H.⁵, Abdullah R.⁵, Yeo C. C.³ and Clarke S.
7 C.^{1,2,6,7,8}

8

9 **Affiliations:**

10 ¹. Faculty of Medicine and Institute for Life Sciences, University of Southampton,
11 Southampton, UK

12 ². Southampton NIHR Biomedical Research Centre, University Hospital Southampton NHS
13 Trust, Southampton, UK

14 ³. Faculty of Medicine, Universiti Sultan Zainal Abidin, Medical Campus, 20400 Kuala
15 Terengganu, Terengganu, Malaysia

16 ⁴. Faculty of Health Sciences, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300
17 Kuala Nerus, Terengganu, Malaysia

18 ⁵. Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin, Gong Badak Campus,
19 21300 Kuala Nerus, Terengganu, Malaysia.

20 ⁶. Global Health Research Institute, University of Southampton, Southampton, UK.

21 ⁷. School of Postgraduate Studies, International Medical University, Kuala Lumpur,
22 Malaysia.

23 ⁸. Centre for Translational Research, IMU Institute for Research, Development and
24 Innovation (IRDI), Kuala Lumpur, Malaysia.

25

26 *Corresponding author

27 **Abstract (350 words max):**

28

29 Background: Microbiome research has focused on populations that are predominantly of
30 European descent, and from narrow demographics that do not capture the socio-economic
31 and lifestyle differences which impact human health. This limits our understanding of
32 human-host microbiota interactions in their broadest sense. Here we examined the airway
33 microbiology of the Orang Asli, the indigenous peoples of Malaysia. In addition to exploring
34 the carriage and antimicrobial resistance of important respiratory pathobionts, we also present
35 the first investigation of the nasal microbiomes of these indigenous peoples, in addition to
36 their oral microbiomes.

37

38 Results: A total of 130 participants were recruited to the study from Kampung Sungai
39 Pergam and Kampung Berua, both sites in the north-eastern state of Terengganu in
40 Peninsular Malaysia. High levels of *Staphylococcus aureus* carriage were observed,
41 particularly in the 18-65 age group (n=17/36; 47.2% 95% CI: 30.9-63.5). The highest carriage
42 of pneumococci was in the <5 and 5 to 17 year olds, with 57.1% (4/7) and 49.2% (30/61)
43 respectively. Sixteen pneumococcal serotypes were identified, the most common being the
44 non-vaccine type 23A (14.6%) and the vaccine type 6B (9.8%). The nasal microbiome was
45 significantly more diverse in those aged 5-17 years compared to 50+ years ($p = 0.023$). In
46 addition, samples clustered by age (PERMANOVA analysis of the Bray-Curtis distance, $p =$
47 0.001). Hierarchical clustering of Bray-Curtis dissimilarity scores revealed six microbiome
48 types. The largest cluster (n=28; 35.4%) had a marked abundance of *Corynebacterium*.
49 Others comprised *Corynebacterium* with *Dolosigranulum*, two clusters were definable by the
50 presence of *Moraxella*, one with and the other without *Haemophilus*, a small grouping of
51 *Delftia/Ochrobactum* profiles and one with *Streptococcus*. No *Staphylococcus* profiles were

52 observed. In the oral microbiomes *Streptococcus*, *Neisseria* and *Haemophilus* were dominant.
53 Lower levels of *Prevotella*, *Rothia*, *Porphyromonas*, *Veillonella* and *Aggregatibacter* were
54 also among the eight most observed genera.

55

56 **Conclusions:** We present the first study of Orang Asli airway microbiomes and pathobiont
57 microbiology. Key findings include the prevalence of pneumococcal serotypes that would be
58 covered by pneumococcal conjugate vaccines if introduced into a Malaysian national
59 immunisation schedule, and the high level of *S. aureus* carriage. The dominance of
60 *Corynebacterium* in the airway microbiomes is particularly intriguing given its' consideration
61 as a potentially protective commensal with respect to acute infection and respiratory health.

62

63 **Keywords:** 16S rRNA; microbiome; upper respiratory tract; antimicrobial resistance;
64 Malaysia; *Streptococcus pneumoniae*; *Haemophilus influenzae*; *Staphylococcus aureus*;
65 *Moraxella catarrhalis*; Orang Asli; indigenous people

66 **Background:**

67 Much microbiome research has, to-date, been focused on populations that are predominantly
68 of European descent, and from demographics that do not capture the socio-economic and
69 lifestyle challenges which impact human health [1]. The study of non-Western,
70 unindustrialised populations is therefore an important extension of microbiome research so
71 that we may understand human-host microbiota interactions in their broadest sense. Notable
72 endeavours here, which have principally focussed on gut microbiomes, include studies of the
73 Hadza hunter-gatherers of Tanzania [2], Amerindians of South America [3, 4], agriculturalist
74 communities in Burkina Faso [5] Malawi and Venezuela [6] and the Cheyenne and Arapaho
75 Tribes of North America [7]. Fewer studies have gone beyond gut microbiomes to sample
76 additional body sites, such as those of the airways. The analysis of salivary microbiota of the
77 Batwa Pygmies [8], and the Yanomami of the Venezuelan Amazon [3] have shown the
78 benefits of including these body sites where, respectively, previously undiscovered genera
79 have been identified and novel combinations of taxa described. However, there remains a
80 paucity of data regarding microbiomes of other anatomical sites of the upper respiratory tract
81 (URT). With respiratory infectious diseases continuing to be a significant component of
82 global morbidity and mortality [9], the interest in the microbiome of the upper airways stems
83 from its' importance in an individual's susceptibility to respiratory infection, in part through
84 the presence of resilient taxa which prevent colonisation and/or outgrowth of specific
85 pathobionts [10]. Understanding the variability in airway microbiomes is a key first step to
86 leveraging these interactions for health. To date no study has undertaken a comprehensive
87 analysis of the URT microbiome of the indigenous populations of Peninsular Malaysia, the
88 Orang Asli.

89

90 Although the name Orang Asli was only introduced around 1960 by the British, as an ethnic
91 group they are believed to be the first settlers of Peninsular Malaysia, moving from Northern
92 Thailand, Burma and Cambodia 3-8000 years ago. Orang Asli is in fact a catch-all term for
93 three tribal groups, the Senoi, Proto-Malays and Negrito, each of which in turn can be further
94 divided into six ethnic groups [11]. They number ~179 000, 0.6% of the population of
95 Malaysia but are disproportionately impacted by economic marginalisation and
96 discrimination [12]. As a consequence, nearly 80% of the Orang Asli population are beneath
97 the poverty line compared to 1.4% of the national population, and this hardship is reflected in
98 a 20-year lower average life expectancy of just 53 years [13]. Several studies have
99 highlighted the susceptibility of these populations to specific diseases [14, 15]. More recently
100 a study of 73 adults from the semi-urbanized Temiar, an Orang Asli tribe of Kampong Pos
101 Piah, in Perak, examined salivary microbiomes in the context of obesity – a growing concern
102 and evidence of the epidemiological shifts in disease burden as these communities leave their
103 more traditional existences [16].

104

105 Here we examined the microbiology of the upper airway of two Orang Asli communities
106 located in Terengganu state, North-east of Peninsular Malaysia. The nasopharyngeal and
107 nasal (anterior nares) carriage of important pathobionts was determined by culture, and the
108 resistance of these isolates to clinically relevant antibiotics tested. We also present the first
109 investigation of the nasal microbiomes of these indigenous peoples, in addition to their oral
110 microbiomes.

111 **Results:**

112 A total of 130 participants were recruited to the study, with 68 from Kampung Sungai
113 Pergam (Site 1) and 62 from Kampung Berua (Site 2) (Figure 1). The age and gender
114 distribution are shown in Figure 1. Of those for whom age was accurately recorded, 49%
115 were female and 45% male. There were notable challenges in the recruitment of children
116 under five years old (accounting for only 11.5% of the total population sampled). All those
117 with missing age data were adults.

118

119 The carriage prevalence of the four most commonly isolated pathobionts in the anterior nares
120 and nasopharynx are shown in Figure 2. Here carriage was defined as the isolation of a
121 pathobiont from either of the two nasal swabs and for visual clarity all adults were combined
122 into one age group (18-65). The most commonly isolated pathobionts were *S. aureus* and *S.*
123 *pneumoniae* both with a total of 39 individuals colonised. The highest carriage for *S. aureus*
124 was in the 18-65 age group at 47.2% (n=17/36; 95%CI: 30.9-63.5), followed by 30%
125 (n=18/60, 95%CI: 18.4-41.6) in the 5 to 17 year olds. In contrast, the highest carriage of
126 pneumococci was in the <5 and 5 to 17 year olds, with 57.1% (4/7) and 49.2% (30/61)
127 respectively. Looking at the latter cases in more detail shows that the average age of a
128 pneumococcal carrier in this age group was ~9 years old (range: 6-13). The next most carried
129 bacterium was *H. influenzae*, isolated from seventeen individuals, the majority of which
130 (88.2%) were in the 5 to 17 years old age group. In contrast, *M. catarrhalis* was rarely
131 isolated (n=6), as was *N. meningitidis* (2), *P. aeruginosa* (n=3), *K. pneumoniae* (n=6) and β-
132 haemolytic *Streptococci* (n=7). Carriage and co-carriage are summarised in Table 1. Carriage
133 of only *S. aureus* or *S. pneumoniae* accounted for 31.5% of the profiles observed. A limited
134 number of co-carriage incidents were found which included seven and five cases of
135 pneumococci isolation with *H. influenzae* or *S. aureus* respectively.

136

137 Serotyping of pneumococcal isolates revealed 16 different serotypes (Figure 3). To avoid
138 repeated sampling bias, if isolates recovered from nose and nasopharyngeal swabs in a single
139 individual were the same serotype this was only counted once. The most common serotypes
140 were the non-VT 23A (14.6%) and the VT 6B (9.8%). Overall, five of the 16 pneumococcal
141 serotypes would be covered by Prevenar 7® (Pfizer, PCV7) (4, 6B, 14, 19F and 23F) with
142 two additional serotypes covered by Prevenar 13® (Pfizer, PCV13) (3 and 6A). Only four of
143 the serotypes (4, 14, 19F and 23F) are included in Synflorix® (GSK, PCV10). Six isolates
144 were not serotypeable. All but one of the vaccine-type pneumococci (a single 23F) were
145 isolated from the site at Kampung Berua (Site 2).

146

147 The extent of antimicrobial resistance for the key pathobionts is shown in Figure 4. Notable
148 levels of resistance were observed for *S. aureus* where 73.3% (n=44) of isolates were
149 resistant to penicillin, 33.0% (n=20) to tetracycline and 18.3% (n=11) to ciprofloxacin.
150 However, all isolates were sensitive to ceftazidime, and all but one to chloramphenicol. Three of
151 the *H. influenzae* isolates (15%) were resistant to penicillin, all of which were negative for β -
152 lactamase activity. Of the *S. pneumoniae*, 17.9% (n=12) were resistant to tetracycline. No
153 resistance was seen for *M. catarrhalis* against any antibiotic tested. *A. baumannii* were
154 sensitive to meropenem and ciprofloxacin (n=14). Similarly, all *K. pneumoniae* were
155 sensitive to all antibiotics tested including ceftazidime and meropenem.

156

157 The microbiome of the nasal (anterior nares) and oral (whole mouth) was examined using
158 16S rRNA sequencing of the V4 hypervariable region. Here, the median number of
159 sequences per sample, including negative controls, was 11,431 with a maximum of
160 1,072,721. Eighteen nasal and two oral samples were excluded based on low numbers of

161 ASVs (<1000). Alpha diversity (Figure 5A and 5B) was similar between all ages apart from
162 nasal samples from those aged 5-17 and 50-65 where a significantly greater diversity was
163 observed in the younger group ($p = 0.0023$); although notably fewer samples were available
164 from the older participants. No clustering was observed using DPCoA for oral samples
165 (Figure 5C – right column). In contrast, groups are clearly visible for nasal samples (Figure
166 5C – left column), with stratification based on age. PERMANOVA analysis of the Bray-
167 Curtis distance showed that this was a significant difference ($p = 0.001$). Examining the most
168 abundant phyla (Figure 6) showed nasal samples were dominated, as expected, by Firmicutes,
169 Proteobacteria and Actinobacteria. For oral samples Bacteroidetes was a much more
170 prevalent phyla in addition to Fusobacteria, although those same three phyla which
171 dominated the nasal samples were also present at high levels here as well. Hierarchical
172 clustering of Bray-Curtis dissimilarity scores revealed six clusters that could be characterised
173 by the dominance of only one or two Genera (Figure 7). The largest cluster ($n=28$; 35.4%)
174 contained individuals with a marked abundance of *Corynebacterium*, accounting for an
175 average ASV relative abundance of 73.2% (range: 59.5 – 87.2%). The second largest cluster
176 ($n=16$; 20.3%) also had a substantial proportion of *Corynebacterium*, in this case with
177 *Dolosigranulum* as significant co-carried genus. Here, on average, *Corynebacterium*
178 accounted for 48.88% and *Dolosigranulum* 31.6% of the relative abundance; combined these
179 Genera accounted for between 71.5 and 89.4% of the ASV relative abundance of this profile.
180 Two groups were then definable by the presence of *Moraxella*, one with and the other
181 without *Haemophilus*. In the *Moraxella*-dominated profile, between 48.2 and 75.2% of the
182 ASV relative abundance could be attributed to this Genus. When combined with
183 *Haemophilus* this proportion dropped slightly from an average of 58.1% to 41.5% and here
184 *Haemophilus* accounted for 13.2 to 47.9% of the profile. A small grouping of
185 *Delftia/Ochrobactum* profiles were also found. The final cluster consisted of only two

186 profiles, where in one *Streptococcus* accounted for 63.8% and the second with
187 *Streptococcus/Haemophilus* at 25.9 and 52.6% respectively. Interestingly, no *Staphylococcus*
188 profiles were observed and none that were characterised by *Haemophilus* alone. No
189 correlation between profile (in terms of genus composition) and culture outcomes for *S.*
190 *aureus* or *S. pneumoniae* were seen (Figure 7). To determine if the abundance of either genus
191 was related to culture positivity, the relative abundance of each was compared between
192 samples from which either *S. aureus* or *S. pneumoniae* was isolated (Supplementary Figure
193 1). The relative abundance of Staphylococcal ASVs was not significantly different between
194 groups, however those from whom *S. pneumoniae* were cultured did have a significantly
195 higher relative abundance of Streptococci ASVs ($p = 0.016$); although this may be due to the
196 presence of two profiles characterised by a significant dominance of *Streptococcus*. To
197 explore the differences in microbiomes between the younger and older age groups,
198 differentially abundant ASVs were identified using DESeq2 (Figure 8). The 5-17 year age
199 group had ASVs classified as *Haemophilus* (including *H. influenzae*), *Moraxella* and
200 *Streptococcus*. In contrast, *Propionibacterium*, *Peptoniphilus* and *Corynebacterium* ASVs
201 were significantly increased in adults.
202
203 *Streptococcus*, *Neisseria* and *Haemophilus* dominated the oral microbiota (Figure 9). Lower
204 levels of *Prevotella*, *Rothia*, *Porphyromonas*, *Veillonella* and *Aggregatibacter* were also
205 among the eight most observed genera. Streptococcal-dominated profiles were characterised
206 by 43.7% ($\pm 15.7\%$) relative abundance of ASVs belonging to this genus. Those with a
207 *Streptococcus/Haemophilus* profile still had 30.7% ($\pm 7.6\%$) Streptococci ASV relative
208 abundance but with 15.0% ($\pm 10\%$) *Haemophilus*. The final profile, where clear dominance of
209 one particular genus was observed, were those where *Neisseria* accounted for 30.6% (± 12.5)
210 of the relative abundance of all ASVs.

211 **Discussion:**

212 As a site that can be colonised by commensals, as well as opportunistic pathogens such as the
213 pneumococcus [17], *S. aureus* [18] and *H. influenzae* [19], understanding the microbiology of
214 the upper respiratory tract is key in tackling respiratory disease. The Anna Karenina principle
215 of microbiomes, that “all healthy microbiomes are similar; each dysbiotic microbiome is
216 dysbiotic in its own way” is a useful maxim for examining microbiomes in the context of
217 disease [20]. However, from previous studies of communities that live unique, traditional
218 lifestyles a clear challenge to the paradigm of health mirroring health has been shown [3, 8].
219 Here we examined the respiratory microbiomes of Orang Asli communities living traditional
220 lifestyles in Terengganu, a rural state in the North-east of Peninsular Malaysia. These
221 communities are known to face specific health challenges to that of the broader population of
222 Malaysia, a country where respiratory disease is still a significant burden [21]. Our study
223 showed a high prevalence of *S. aureus* carriage but interestingly not the presence of
224 Staphylococci-dominated nasal microbiomes. Instead, *Corynebacterium*, with
225 *Dolosigranulum* and *Moraxella* dominated as commensal flora. We also showed the potential
226 impact that PCV introduction in Malaysia may have on circulating pneumococcal serotypes.
227
228 Although pneumococcal carriage can vary globally between 10 and 90% [22-24], the highest
229 carriage prevalence is usually seen in young children, particularly those <5 years [24].
230 Although only a small number of this age group was sampled here (n=7), over half were
231 colonised suggesting high levels of carriage. Given that in the 5-17 year age group the mean
232 age was young at ~9 years old the high carriage (50%; 95%CI, 37.3 and 62.7%) seen in this
233 age group supports this observation, being higher than has been seen elsewhere in older
234 children [25] and similar to the 60.9% observed in Aboriginal children over 5 years old – a
235 demographic known to experience high rates of invasive pneumococcal disease (IPD) [26].

236 As expected, the carriage in adults was low at 8.3% (95%CI, 0% to 17.4%), which is similar
237 to the NP carriage prevalence of 11.1% (95%CI, 9.8 and 12.6%) that has been seen in the
238 USA [27], but higher than recent surveys in the UK which found a 2.8% (95%CI: 1.2–5.5)
239 prevalence in this age group [25]. As for serotypes, whilst there is growing epidemiological
240 data for the region [28], there is limited data for Malaysia. A study of 245 cases of paediatric
241 IPD between 2014 and 2017 showed that serotypes 14, 6B, 19A, 6A and 19F were the most
242 common, accounting for ~75% of disease [29]. This is concordant with this study where
243 serotypes 14 and 6B were also the most common in carriage with 6A and 19F being
244 identified. Of the non-VT serotypes isolated only 34, 35F and 18A did not feature in those
245 identified in the IPD study. The lack of 19A is intriguing but we suspect a consequence of the
246 relatively low sampling. Both PCV10 and PCV13 are available in Malaysia however only in
247 private practice and therefore the presence of VT serotypes is not unsurprising. These data do
248 suggest introduction of PCV into the national immunisation programme as planned may
249 prove efficacious.

250

251 Only one of the children less than 5 years old was culture positive for *H. influenzae* but given
252 the low numbers in this age group this must be interpreted with caution. Whilst certainly
253 higher than in the UK where only 6.5% *H. influenzae* carriage was seen in older children
254 [30], the 25% (95%CI, 14.0 and 40.0%) found here is similar to that observed in other
255 developing countries such as Kenya [31], but lower than the 52.8% seen in Aboriginal
256 children aged between 5 and 15 years old [26]. Whether this increased carriage translates into
257 disease risk is difficult to determine; there is no data on the burden of otitis media among
258 Orang Asli, although the country wide incidence in children <12 year olds is low at 2.3% for
259 the Asia-Pacific region [32].

260

261 The prevalence of *M. catarrhalis* at 6.7% (95%CI, 0.35 and 13.0%) was very similar to that
262 of our previous UK study where 5.7% of the same age group were colonised [30], a level also
263 observed in other western countries [33]. This is a stark contrast to the 67% in Aboriginal
264 children of a similar age [26], however this low level of carriage (6%) has been seen in
265 Warao Amerindians in Venezuela [34] – arguably a similar demographic to the one being
266 presented here.

267

268 *S. aureus* carriage is known to vary markedly, by age but also by geographic location [35].
269 Although the lowest prevalence observed here was in <5 year olds (28.6%; 95%CI: 0-62.0),
270 which is surprising as one would expect the highest carriage would be seen in this age group,
271 it could be either a feature of the low recruitment or that the average age of this group was
272 2.5 and carriage is known decline between 1 and 10 [18]. As for the carriage in adults, a
273 previous study in Malaysia of 384 adults (students with an average age of 25) found a
274 carriage prevalence of 23.4% [36] which is much lower than the 47.2% in our population of
275 Orang Asli. A similarly high level of adult carriage (41.7% and 57.8% at two separate
276 sampling timepoints) was observed in Wayampi Amerindians who, in terms of remote living,
277 are similar to the Orang Asli [37]. Our own research has previously shown adult carriage at
278 24.4% in a UK population [30] and thus the differences we observed are unlikely to be
279 methodological. More likely it is a consequence of greater co-habitation (the average
280 household size of our communities being 5.2 people) and reduced access to hygiene, both of
281 which have been highlighted as reasons why *S. aureus* carriage has declined in other parts of
282 the world [18].

283

284 Of the ESKAPE pathogens isolated and tested only *S. aureus* exhibited any substantial level
285 of resistance; importantly all isolates were sensitive to methicillin. Whilst there appears to be

286 limited resistance it is important to note that this represents the only data on the AMR in
287 these communities and is therefore merely a starting point for future studies. This is
288 particularly important given that AMR burden in indigenous populations has been seen to
289 increase elsewhere [38].
290
291 URT microbiome research has shown that the composition and/or dominance of particular
292 bacterial taxa can be indicative of nasal community state types (CSTs). Previously, in a study
293 of adults in the USA, seven CSTs were identified that were characterised by *S. aureus*
294 (CST1), or other *Staphylococci* (CST3), a mix of Proteobacteria including *Escherichia* and
295 *Enterobacteriaceae* sp. (CST2), *Corynebacterium* either with *Propionibacterium acnes*
296 (CST4) or without (CST5), *Moraxella* (CST6) and *Dolosigranulum* (CST7) [39]. It is here
297 that we see the most startling contrasts to our Orang Asli population. Although we also
298 identified seven clusters based on what is/are the dominant taxa we did not see anything
299 resembling CST 1 or 3 (*Staphylococci*), and no CST2. The lack of a *Staphylococcus*
300 dominated profile is an enigma given the high carriage rate we observed. We hypothesise
301 that the high prevalence of profiles similar to CST5, *Corynebacterium* dominated, along with
302 another that is defined by high abundance of both *Corynebacterium* and *Dolosigranulum*, a
303 composition that has been found in the nasopharynx [40], explains this absence and is based
304 on previous work showing that *Corynebacterium* reduces *S. aureus* carriage [41]. A similar
305 interaction may also explain the absence of *Streptococci* [42] – although we recognise in both
306 cases that our hypothesis requires further testing due to the identification of both *S. aureus*
307 and pneumococcal carriage among our participants. A *Moraxella* dominated profile similar to
308 CST6, is also present in the Orang Asli, a profile which has also been observed in healthy
309 adults in a European setting [43], although we also see a *Moraxella-Haemophilus* profile
310 which hasn't been described previously. The *Streptococcus* profiles are both from older

311 children (6 and 10 years old) and this not uncommon as a profile among this age group [44].
312 The significant proportion of individuals who harbour profiles that are being considered to be
313 potentially protective either in terms of acute respiratory infections [45] or chronic disease
314 such as asthma [46-48], is intriguing. Lastly, whilst *Delftia* is not uncommon as an oral
315 bacterium [49], given the common finding of *Delftia* and *Ochrobactum* as common
316 contaminants in microbiome research [50] we interpret these four samples with extreme
317 caution. The presence of these profiles is in spite of our efforts to control for such
318 contaminants through the collection and sequencing of control swabs. Epidemiologically, all
319 four samples come from adult females, but two each from Kampung Berua and Kampung
320 Sungai Pergam. The samples are below the 25% quantile for ASV number (n=3831). It is
321 possible that these reflect true profiles, but it is impossible to rule out process error during
322 sample handling, extraction and/or sequencing.

323

324 The microbiota recovered from the oral cavity was substantially more diverse than the nasal
325 samples, as expected [49]. The top three genera were unremarkable in terms of what is
326 considered to be most commonly observed, mirroring exactly that found in a study of 447
327 datasets deposited as part of large-scale microbiome projects [51]. A previous study of
328 Temiar, Orang Asli in Perak found a significant difference in oral microbiota compositions
329 between males and females [16], however a PERMANOVA analysis of the data presented
330 here showed no such distinction ($p = 0.41$).

331

332 Whilst this piece of research was highly novel as a first glimpse into the respiratory
333 microbiology of Orang Asli communities, it could nevertheless have been strengthened in a
334 number of ways. Firstly, it was only possible to visit two sites on a single occasion, and to
335 conduct a pragmatic survey of the communities. As a consequence, the study lacks follow-up

336 and longitudinal data. Whilst we were very successful at recruiting older children and adults,
337 future studies will have to identify how to recruit greater numbers of younger children in
338 particular. In terms of pathobiont carriage, serotyping of more than one pneumococcal isolate
339 from each individual would have enabled a determination of the prevalence of multiple
340 serotype carriage. Serotyping of *H. influenzae* would also have been informative, although
341 future work on exploring the genomics of all isolates will address this in more detail. This
342 additional insight could also be expanded to include all the pathogens/pathobionts isolated.
343 As with all microbiome research, longitudinal sampling taking into account an individual's
344 temporal variability, the seasonality and medical history would be desirable.

345 **Conclusions:** Here we present the first study of Orang Asli airway microbiomes and
346 pathobiont microbiology, including the antimicrobial resistance profiles of ESKAPE
347 pathogens. Important findings include the prevalence of pneumococcal serotypes that would
348 be covered by pneumococcal conjugate vaccines, supporting their introduction as part of a
349 national immunisation programme. The high prevalence of *S. aureus* carriage is noteworthy
350 and warrants further study. In terms of the airway microbiomes the dominance of
351 *Corynebacterium*, additionally with *Dolosigranulum* and *Moraxella* in nasal profiles is
352 particularly intriguing and future work should explore these commensals in the context of
353 burden and susceptibility to both acute and chronic respiratory conditions in these
354 communities.

355 **Methods:**

356 Study sites and participants: Two Orang Asli villages were visited in August 2017 –
357 Kampung Sungai Pergam in Kemaman district and Kampung Berua in Hulu Terengganu
358 district. Both sites are located in the state of Terengganu which lies in the north-east of
359 Peninsular Malaysia. Participant recruitment was with consent, across all ages with no
360 exclusion criteria. Questionnaires were used to capture participant metadata which included
361 gender, age, number of dwelling co-occupants and occupation in addition to health-related
362 questions including current or recent (within the last month and last three months) respiratory
363 symptoms, antibiotic use and vaccination status. Questionnaires were translated into Malay
364 by A.S.M.H., A.R., M.A.R., and C.C.Y., and conducted by research assistants from the
365 Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin, who were trained on
366 how to administer the questionnaires for the purpose of this study.

367

368 Swab collection: A total of four swabs were taken from each participant. A nasal swab
369 (anterior nares) and a nasopharyngeal swab, using rayon tipped transport swabs containing
370 Amies media with charcoal (Medical Wire and Equipment, Corsham, UK), were used for
371 bacterial pathobiont isolation. A further nasal swab, taken from the un-swabbed nostril, and
372 an oral (whole mouth) swab, supplied by uBiome Inc. (San Francisco, USA), were taken for
373 microbiome analysis.

374

375 Bacteriology: Isolation of the common respiratory pathobionts *S. pneumoniae*, *H. influenzae*,
376 *M. catarrhalis*, *N. meningitidis*, *S. aureus*, *K. pneumoniae* and *P. aeruginosa* was done by
377 culture. Swabs were plated onto CBA (Columbia blood agar with horse blood), CHOC
378 (Columbia blood agar with chocolated horse blood), CNA (Columbia Blood Agar with
379 Colisitin and Naladixic Acid), BACH (Columbia Agar with Chocolated Horse Blood and

380 Bacitracin), GC (Lysed GC Selective Agar) and, for *P. aeruginosa*, CFC (*Pseudomonas* CFC
381 Selective agar) (all Oxoid, UK). Primary identification of all bacterial species was by colonial
382 morphology. Swabs were then vortexed in skim milk, tryptone, glucose and glycerine
383 (STGG), and 10 μ L of the suspension was then
384
385 Serotyping of *S. pneumoniae*: Pneumococcal isolates were serotyped by slide agglutination
386 reactions using a Neufeld *S. pneumoniae* antisera kit following manufacturer's guidelines
387 (Statens Serum Institute, Copenhagen, Denmark).
388
389 Antimicrobial Resistance Testing: All bacteria were phenotypically tested for antibiotic
390 resistance using antibiotic discs and/or minimum inhibitory concentration (MIC) strips, in
391 accordance with EUCAST. Firstly, 10 μ L (a suspension of cells in liquid STGG) of each
392 isolate was plated onto CBA (Oxoid, UK) or CHOC agar (Oxoid, UK). *M. catarrhalis*, *S.*
393 *pneumoniae*, *S. aureus* and *K. pneumoniae* isolates were plated on CBA, whilst *H. influenzae*
394 and *N. meningitidis* isolates were plated onto CHOC agar. Plates were incubated for 24 hours
395 at 37°C in 5% CO₂. Pure colonies were added to 1ml of saline to get an inoculum of 0.5
396 McFarland. For *M. catarrhalis*, *S. pneumoniae*, *H. influenzae* and *N. meningitidis*, a sterile
397 swab was used to spread this inoculum over Mueller-Hinton agar + 5% defibrinated horse
398 blood and 20 mg⁻¹ β -NAD plates (MHF, Oxoid, UK). For *S. aureus* and *K. pneumoniae* a
399 sterile swab was used to spread the inoculum over Mueller-Hinton agar plates (MH, Oxoid,
400 UK). Antibiotic discs (Oxoid, UK) (four per plate) or MIC strips (E-tests; Oxoid, UK) (one
401 per plate) were added and plates were incubated at 37°C in 5% CO₂ for 18 hours (\pm 2 hours).
402 *M. catarrhalis* were tested with amoxicillin-clavulanic acid (2-1 μ g), cefotaxime (5 μ g),
403 ceftriaxone (30 μ g), erythromycin (15 μ g), tetracycline (30 μ g), chloramphenicol (30 μ g),
404 ciprofloxacin (5 μ g) and meropenem (10 μ g) antibiotic discs. *S. pneumoniae* were tested with

405 oxacillin (1µg), erythromycin (15µg), tetracycline (30µg) and chloramphenicol (30µg)
406 antibiotic discs. *H. influenzae* were tested with benzylpenicillin (1µg), tetracycline (30µg),
407 chloramphenicol (30µg) and ciprofloxacin (5µg) antibiotic discs as well as erythromycin
408 (15µg) MIC strips. *S. aureus* were tested with benzylpenicillin (1µg), erythromycin (15µg),
409 cefoxitin (30µg), tetracycline (30µg), chloramphenicol (30µg) and ciprofloxacin (5µg)
410 antibiotic discs. *K. pneumoniae* were tested with amoxicillin - clavulanic acid (20-10µg),
411 cefotaxime (5µg), ciprofloxacin (5µg), meropenem (10µg) and ceftazidime (10µg) antibiotic
412 discs. *N. meningitidis* were tested with amoxicillin, benzylpenicillin (1µg), cefotaxime (5µg),
413 ceftriaxone (30µg), chloramphenicol (30µg), ciprofloxacin (5µg) and meropenem (10µg)
414 MIC strips.

415

416 16S rRNA Sequencing: The V4 region was amplified using primers 515F (5'-
417 GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') to
418 generate an amplicon of 460 bp [52] at uBiome Inc. (San Francisco, USA). Samples were
419 individually barcoded and sequenced on a NextSeq 500 (Illumina, San Diego, USA) to
420 generate 2 × 150 bp paired-end reads.

421

422 Microbiome Analysis: Initial data handling was done using QIIME 2 2018.8 [53]. Only
423 forward reads were processed owing to low quality scores of reverse reads. Raw sequence
424 data were denoised with DADA2 [54]. Amplicon sequence variants (ASVs) were aligned
425 with mafft [55] (via q2 alignment) and a phylogeny constructed using fasttree2 [56].

426 Taxonomic classification was done using the q2 feature classifier [57] classify sklearn
427 naïve Bayes taxonomy classifier using the Greengenes 13_8 (99%) reference sequences [58].

428 The feature table, taxonomy, phylogenetic tree and sample metadata were then combined into
429 a Phyloseq object using qiime2R [59] via qza_to_phyloseq. All further analysis was done in

430 R v3.6.0 [60] in RStudio and figures were produced using the package ggplot2 [61]. Phyloseq
431 v1.29.0 [62] was used following a workflow previously described [63]. Potential
432 contaminants were identified by prevalence in the blank swab controls and removed using the
433 R package ‘decontam’ [64]. Samples with <1000 ASVs were excluded and taxa present in
434 less than 5% of samples were removed using the prune_taxa() phyloseq function. Alpha
435 diversity was calculated using the estimate_richness(). The R function stat_compare_means()
436 was used to compare age groups using the nonparametric Wilcoxon test. Beta diversity was
437 determined using Double Principle Co-ordinates Analysis and plot_ordination(). Hierarchical
438 clustering was done using Bray-Curtis Dissimilarity calculated using vegdist() from the R
439 package vegan [65]. Differentially abundant ASVs were identified using the DESeq2 [66]
440 package using an adjusted p-value cut-off of 0.05 and a Log₂ fold change of 1.5.

441 **Abbreviations:**

442 AMR – antimicrobial resistance; ASV – amplicon sequencing variant; PCV – pneumococcal
443 conjugate vaccine; VT – vaccine type pneumococci; NVT- non-vaccine type pneumococci;
444 ESKAPE - *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*,
445 *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.

446

447 **Ethics Approval and Consent to Participate:**

448 Ethical approval for this study was provided by Universiti Sultan Zainal Abidin (UniSZA)
449 Ethics Committee: approval no. UniSZA/C/1/UHREC/628-1(85) dated 27th June 2016, the
450 Department of Orang Asli Affairs and Development (JAKOA): approval no.
451 JAKOA/PP.30.052Jld11 (42), and by the University of Southampton Faculty of Medicine
452 Ethics Committee (Submission ID: 20831). Informed consent was taken for all participants,
453 with parents/guardians providing consent for those <17 years old. Participation in the study
454 involved reading (or being read) and understanding the translated participant information
455 sheet, and the completion of a consent form and questionnaire. This process was facilitated
456 by native speakers.

457

458 **Consent for Publication:**

459 Not applicable.

460

461 **Availability of Data and Materials:**

462 Sequence files and associated metadata have been deposited in the European Nucleotide
463 Archive (ENA) in project PRJEB38610 under experiment accessions ERX4147052 to
464 ERX4147288.

465

466 **Competing Interests:**

467 SCC acts as principal investigator on studies conducted on behalf of University Hospital
468 Southampton NHS Foundation Trust/University of Southampton that are sponsored by
469 vaccine manufacturers but receives no personal payments from them. SCC has participated in
470 advisory boards for vaccine manufacturers but receives no personal payments for this work.
471 SCC has received financial assistance from vaccine manufacturers to attend conferences.
472 DWC was a post-doctoral researcher on projects funded by Pfizer and GSK between April
473 2014 and October 2017. All grants and honoraria are paid into accounts within the respective
474 NHS Trusts or Universities, or to independent charities. All other authors have no conflicts of
475 interest.

476

477 **Funding:**

478 The laboratory microbiology work was funded by a grant from the Higher Education Funding
479 Council for England (HEFCE) Newton Fund Official Development Assistance (ODA) fund.
480 The microbiome sequencing was funded through a uBiome Inc. Microbiome Grants Initiative
481 award.

482

483 **Authors Contributions:**

484 DWC and SCC conceived the project with YCC. DE, RAA and HM were the UoS contingent
485 who undertook the sampling and helped with study planning and design. AGA, NIAR, SI,
486 MSR, RMA, AAA, NKE, SA, CHC MHAS and RA helped with study design, coordinated
487 sampling visits to Orang Asli settlements, liaised with those communities and assisted with
488 sampling and questionnaires. AGA did the microbiological culture and antibiotic testing with
489 RAA and DE. DWC carried out the data analysis and wrote the paper. All authors provided
490 critical feedback and helped shape the manuscript.

491

492 **Acknowledgments:**

493 We would like to thank the Department of Orang Asli Affairs and Development (JAKOA),
494 Additionally, the authors would especially like to thank the communities of Kampung Sungai
495 Pergam and Kampung Berua for their support and participation in this study.

496

497 **Figure Legends:**

498 **Figure 1: Location of Orang asli sites with age and gender demographics.** The location of
499 Orang Asli settlements were in the north-east of Peninsula Malaysia (A) to the south of the
500 Terengganu state's capital city, Kuala Terengganu (B). An even gender split was achieved
501 (C) but with disproportionate recruitment in older children >5 years of age and adults less
502 than <65 (D).

503

504 **Figure 2: Carriage Prevalence of *S. pneumoniae*, *H. influenzae*, *S. aureus* and *M.***
505 ***catarrhalis*.** Point estimates for percentage carriage prevalence of four key upper respiratory
506 tract pathobionts are shown across three age groups <5, 5 to 17 and 18 to 65. Those aged 65+
507 are not shown due to limited numbers. All 'Unknowns' were of adult age. Carriage was
508 estimated overall i.e. a positive swab culture from either nasal or nasopharyngeal swab (dark
509 blue), and also separately for each swab type (nasal – black; NP – grey). Error bars show
510 95% CI intervals.

511

512 **Figure 3: Prevalence of PCV7, PCV13 and non-vaccine pneumococcal serotypes.** Bar
513 plot with 95% CI showing prevalence of serotypes that are included in PCV7 (black),
514 additional types covered by PCV13 (grey) or non-vaccine serotypes (white).

515

516 **Figure 4: Antibiotic resistance of *S. pneumoniae* (Spn), *S. aureus* (Sa) and *H. influenzae***
517 **(Hi).** Columns represent the proportion of isolates that were resistant to tetracycline,
518 chloramphenicol, erythromycin, ciprofloxacin and penicillin.

519

520 **Figure 5: Alpha and Beta Diversity of Nasal (left column) and Oral (right column)**
521 **Samples.** Observed richness and Simpsons 1-D (a measure of diversity) are shown in rows A
522 and B respectively. Only a significant difference (**) was observed for nasal samples for
523 those aged 5-17 when compared against 50-65 ($p = 0.0023$). Beta diversity is shown using
524 Double Principle Coordinates Analysis (DPCoA) (row C). Here, clusters are observed for
525 nasal samples (left) with grouping based on age-group. No clusters were observed for oral
526 samples (right).

527

528 **Figure 6: Log Transformed Abundances of Phylum-level ASV Classifications in Nasal**
529 **(top) and Oral (bottom) Samples.** Bars are coloured by site of location. Firmicutes,
530 Actinobacteria and Proteobacteria were the most abundant phyla in Nasal samples with
531 Bacteroidetes, Firmicutes and Proteobacteria the most common in Oral samples. A clear
532 increase in Firmicutes in nasal samples from Site 2 (Kampung Berua) can be seen.

533

534 **Figure 7: Hierarchical Clustering of Nasal Samples with Bray-Curtis Dissimilarity**
535 **using Genus-level Relative Abundances.** The dendrogram (top) shows clustering of
536 samples with the below bar chart showing the relative abundance of the eight most
537 commonly observed Genera. Six clear clusters were observed including a *Corynebacterium*
538 dominant profile (n=28; 35.4%), a *Corynebacterium/Dolosigranulum* profile (n=16; 20.3%),
539 a *Moraxella* profile (n=10; 12.7%) a *Moraxella/Haemophilus* profile (n=10; 12.7%), a
540 *Delftia/Orchobactum* profile (n=4; 5.1%) and a final group of two samples characterised by

541 high *Streptococcus*, in one case with *Haemophilus*. Culture results for *S. aureus* (dark blue)
542 and *S. pneumoniae* (pink) are shown as coloured circles at the tips of the dendrogram. There
543 is no clear distribution of individuals who were culture positive for either bacterium.

544

545 **Figure 8: Differentially Abundant ASVs Between Older Children (5-17 years) and**
546 **Adults (18-49).** Volcano plot showing the ASVs that were differentially abundant in
547 comparisons between the two age groups. Here a p-value cut-off of 0.05 and a Log₂ fold
548 change of 1.5 was applied. Those ASVs that pass both these thresholds are labelled where a
549 genus and/or species taxonomic assignment was available.

550

551 **Figure 9: Hierarchical Clustering of Oral Samples with Bray-Curtis Dissimilarity using**
552 **Genus-level Relative Abundances.** The dendrogram (top) shows clustering of samples with
553 the below bar chart showing the relative abundance of the six most commonly observed
554 Genera. Three genera are seen to be responsible for the majority of ASVs – *Streptococcus*
555 (brown), *Haemophilus* (orange) and *Neisseria* (purple). Profiles characterised by an
556 overabundance of *Streptococcus* or *Neisseria* are seen, as too are those containing both
557 *Streptococcus* and *Haemophilus*.

558

559 **Supplementary Figure 1: Comparison of the relative abundance of Streptococcal and**
560 **Staphylococcus ASVs between those that were culture positive or negative for**
561 ***Streptococcus pneumoniae* or *Staphylococcus aureus*. P values were calculated using the**
562 **nonparametric Wilcoxon test.**

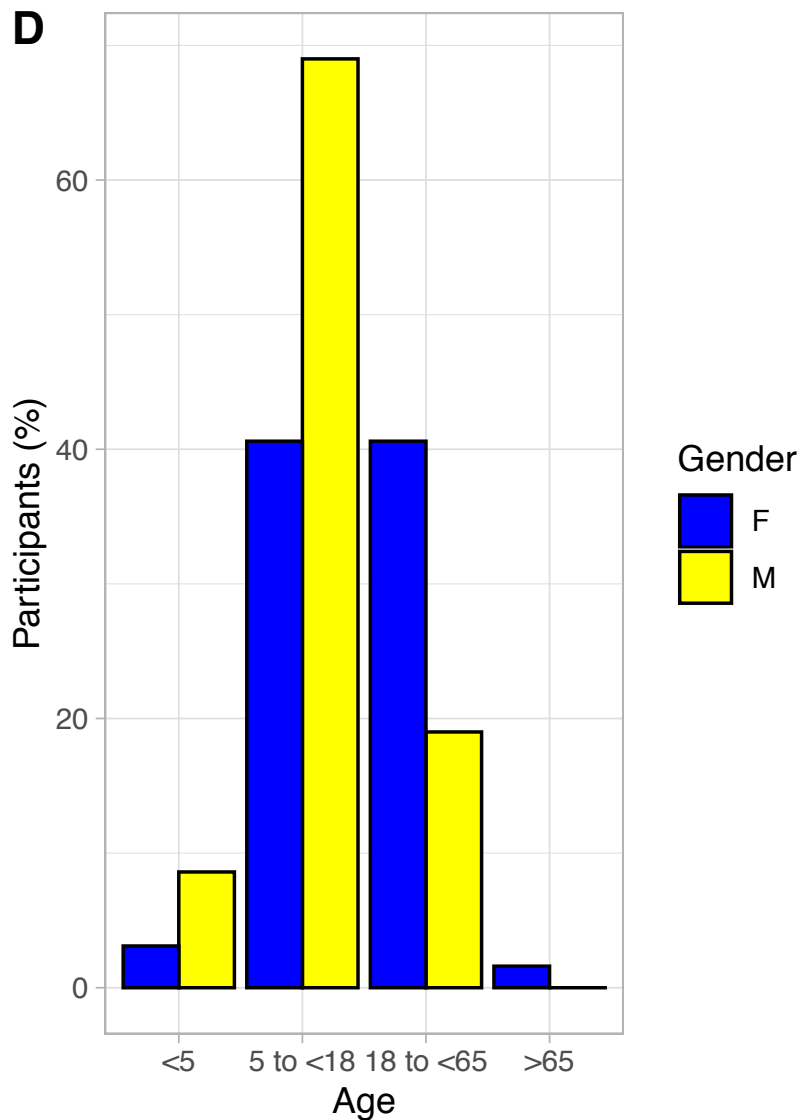
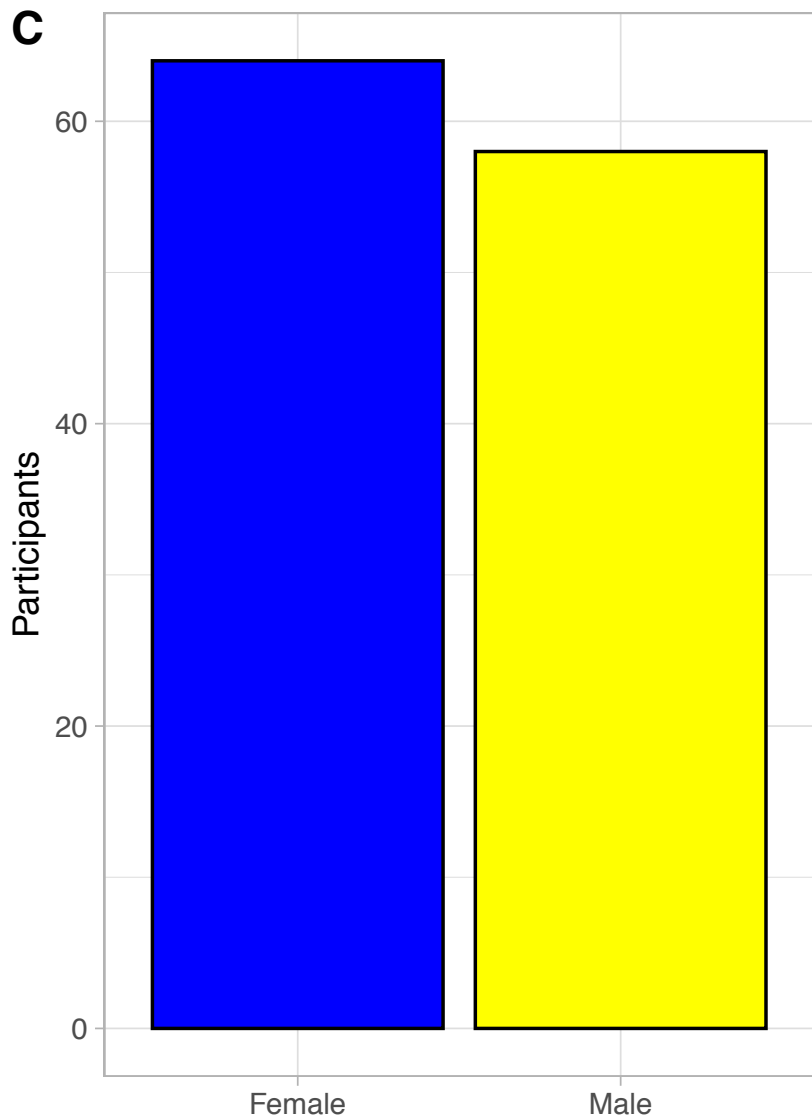
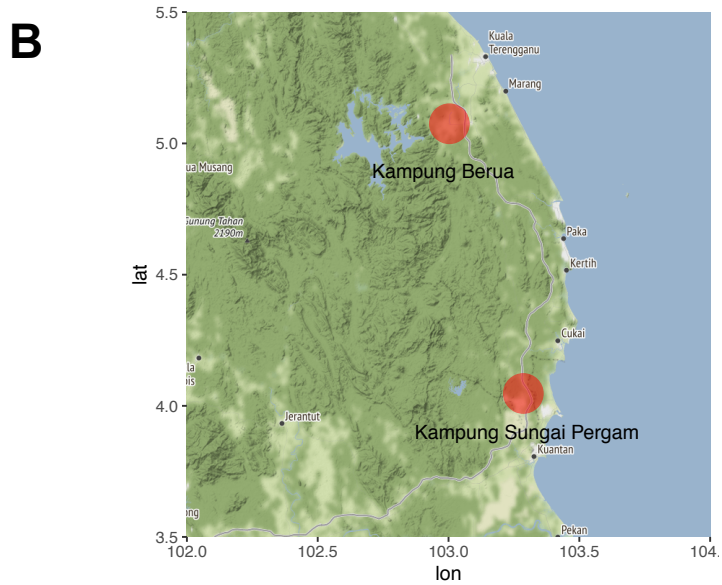
563 **References:**

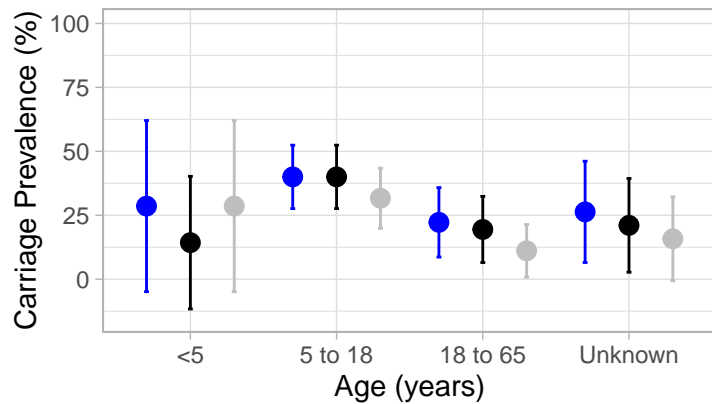
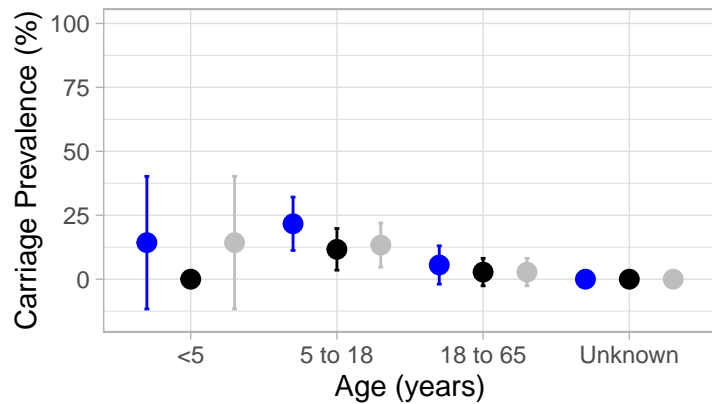
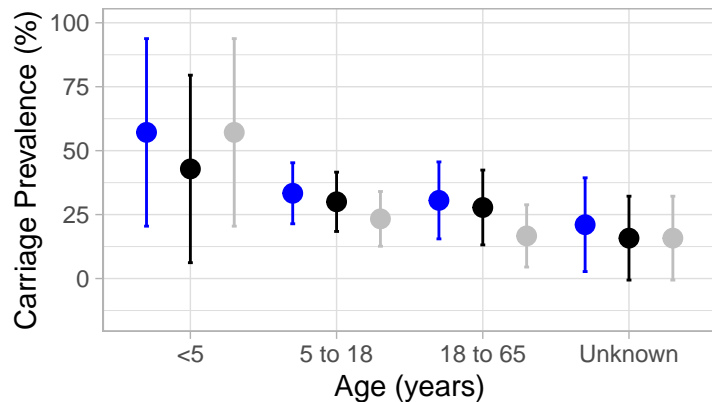
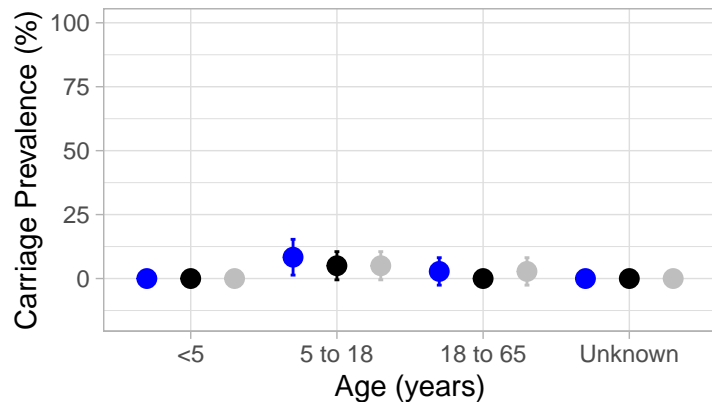
- 564 1. Rogers, G.B., et al., *Inclusivity and equity in human microbiome research*. The
565 Lancet, 2019. **393**(10173): p. 728-729.
- 566 2. Schnorr, S.L., et al., *Gut microbiome of the Hadza hunter-gatherers*. Nature
567 Communications, 2014. **5**(1): p. 3654.
- 568 3. Clemente, J.C., et al., *The microbiome of uncontacted Amerindians*. Science
569 Advances, 2015. **1**.
- 570 4. Obregon-Tito, A.J., et al., *Subsistence strategies in traditional societies distinguish
571 gut microbiomes*. Nature Communications, 2015. **6**(1): p. 6505.
- 572 5. Filippo, C., et al., *Impact of diet in shaping gut microbiota revealed by a comparative
573 study in children from Europe and rural Africa*. Proc Natl Acad Sci U S A, 2010. **107**.
- 574 6. Yatsunenkov, T., et al., *Human gut microbiome viewed across age and geography*.
575 Nature, 2012. **486**.
- 576 7. Sankaranarayanan, K., et al., *Gut Microbiome Diversity among Cheyenne and
577 Arapaho Individuals from Western Oklahoma*. Current Biology, 2015. **25**(24): p.
578 3161-3169.
- 579 8. Nasidze, I., et al., *High Diversity of the Saliva Microbiome in Batwa Pygmies*. PLOS
580 ONE, 2011. **6**(8): p. e23352.
- 581 9. Lozano, R., et al., *Global and regional mortality from 235 causes of death for 20 age
582 groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease
583 Study 2010*. The Lancet, 2012. **380**(9859): p. 2095-2128.
- 584 10. Man, W.H., W.A.A. de Steenhuijsen Piters, and D. Bogaert, *The microbiota of the
585 respiratory tract: gatekeeper to respiratory health*. Nat Rev Micro, 2017. **15**(5): p.
586 259-270.
- 587 11. SyedHussain, T., Krishnasamy, DS., Hassan, AAG., *Distribution and demography of
588 the Orang Asli in Malaysia*. International Journal of Humanities and Social Science
589 Invention, 2017. **6**(1): p. 40-45.
- 590 12. Masron T., M.F., Ismail N., *Orang Asli in Peninsular Malaysia: Population, Spatial
591 Distribution and Socio-Economic Condition*. J Ritsumeikan Soc Sci Humanit, 2013.
592 **6**: p. 75-115.
- 593 13. Mohd Asri, M.N., *Advancing the Orang Asli through Malaysia Clusters of Excellence
594 Policy*. Journal of International and Comparative Education, 2012. **1**(2).
- 595 14. Al-Mekhlafi, H.M., et al., *Prevalence and risk factors of Strongyloides stercoralis
596 infection among Orang Asli schoolchildren: new insights into the epidemiology,
597 transmission and diagnosis of strongyloidiasis in Malaysia*. Parasitology, 2019.
598 **146**(12): p. 1602-1614.
- 599 15. Al-Mekhlafi, H.M., et al., *Protein-energy malnutrition and soil-transmitted
600 helminthiasis among Orang Asli children in Selangor, Malaysia*. Asia Pac J Clin
601 Nutr, 2005. **14**(2): p. 188-94.
- 602 16. Yeo, L.-F., et al., *Health and saliva microbiomes of a semi-urbanized indigenous
603 tribe in Peninsular Malaysia*. F1000Research, 2019. **8**: p. 175-175.
- 604 17. Bogaert, D., R. de Groot, and P.W.M. Hermans, *Streptococcus pneumoniae
605 colonisation: the key to pneumococcal disease*. The Lancet Infectious Diseases, 2004.
606 **4**(3): p. 144-154.
- 607 18. Wertheim, H.F.L., et al., *The role of nasal carriage in Staphylococcus aureus
608 infections*. The Lancet Infectious Diseases, 2005. **5**(12): p. 751-762.
- 609 19. Slack, M.P.E., *A review of the role of Haemophilus influenzae in community-acquired
610 pneumonia*. Pneumonia, 2015. **6**.

- 611 20. Zaneveld, J.R., R. McMinds, and R. Vega Thurber, *Stress and stability: applying the*
612 *Anna Karenina principle to animal microbiomes*. *Nature Microbiology*, 2017. **2**(9): p.
613 17121.
- 614 21. *Statistics on Causes of Death, Malaysia*, D.o.S. Malaysia, Editor. 2018.
- 615 22. Hill, P.C., et al., *Nasopharyngeal Carriage of Streptococcus pneumoniae in Gambian*
616 *Infants: A Longitudinal Study*. *Clinical Infectious Diseases*, 2008. **46**(6): p. 807-814.
- 617 23. Regev-Yochay, G., et al., *Nasopharyngeal Carriage of Streptococcus pneumoniae by*
618 *Adults and Children in Community and Family Settings*. *Clinical Infectious Diseases*,
619 2004. **38**(5): p. 632-639.
- 620 24. Bogaert, D., et al., *Colonisation by Streptococcus pneumoniae and Staphylococcus*
621 *aureus in healthy children*. *Lancet*, 2004. **363**.
- 622 25. Southern, J., et al., *Pneumococcal carriage in children and their household contacts*
623 *six years after introduction of the 13-valent pneumococcal conjugate vaccine in*
624 *England*. *PLOS ONE*, 2018. **13**(5): p. e0195799.
- 625 26. Mackenzie, G.A., et al., *Epidemiology of nasopharyngeal carriage of respiratory*
626 *bacterial pathogens in children and adults: cross-sectional surveys in a population*
627 *with high rates of pneumococcal disease*. *BMC Infectious Diseases*, 2010. **10**(1): p.
628 304.
- 629 27. Watt, J.P., et al., *Nasopharyngeal versus Oropharyngeal Sampling for Detection of*
630 *Pneumococcal Carriage in Adults*. *Journal of Clinical Microbiology*, 2004. **42**: p.
631 4974-4976.
- 632 28. Jauneikaite, E., et al., *Prevalence of Streptococcus pneumoniae serotypes causing*
633 *invasive and non-invasive disease in South East Asia: A review*. *Vaccine*, 2012.
634 **30**(24): p. 3503-3514.
- 635 29. Arushothy, R., et al., *Pneumococcal serotype distribution and antibiotic susceptibility*
636 *in Malaysia: A four-year study (2014–2017) on invasive paediatric isolates*.
637 *International Journal of Infectious Diseases*, 2019. **80**: p. 129-133.
- 638 30. Coughtrie, A.L., et al., *Ecology and diversity in upper respiratory tract microbial*
639 *population structures from a cross-sectional community swabbing study*. *Journal of*
640 *Medical Microbiology*, 2018. **67**(8): p. 1096-1108.
- 641 31. Abdullahi, O., et al., *The descriptive epidemiology of Streptococcus pneumoniae and*
642 *Haemophilus influenzae nasopharyngeal carriage in children and adults in Kilifi*
643 *district, Kenya*. *The Pediatric infectious disease journal*, 2008. **27**(1): p. 59-64.
- 644 32. Mahadevan, M., et al., *A review of the burden of disease due to otitis media in the*
645 *Asia-Pacific*. *International Journal of Pediatric Otorhinolaryngology*, 2012. **76**(5): p.
646 623-635.
- 647 33. Kovács, E., et al., *Co-carriage of Staphylococcus aureus, Streptococcus pneumoniae,*
648 *Haemophilus influenzae and Moraxella catarrhalis among three different age*
649 *categories of children in Hungary*. *PLOS ONE*, 2020. **15**(2): p. e0229021.
- 650 34. Verhagen, L.M., et al., *Nasopharyngeal carriage of respiratory pathogens in Warao*
651 *Amerindians: significant relationship with stunting*. *Tropical Medicine &*
652 *International Health*, 2017. **22**(4): p. 407-414.
- 653 35. Sollid, J.U.E., et al., *Staphylococcus aureus: Determinants of human carriage*.
654 *Infection, Genetics and Evolution*, 2014. **21**: p. 531-541.
- 655 36. Choi, C.S., et al., *Nasal carriage of Staphylococcus aureus among healthy adults*. *J*
656 *Microbiol Immunol Infect*, 2006. **39**(6): p. 458-64.
- 657 37. Ruimy, R., et al., *Are host genetics the predominant determinant of persistent nasal*
658 *Staphylococcus aureus carriage in humans?* *The Journal of Infectious Diseases*,
659 2010. **202**(6): p. 924-934.

- 660 38. Tong, S.Y.C., et al., *Progressive increase in community-associated methicillin-*
661 *resistant Staphylococcus aureus in Indigenous populations in northern Australia from*
662 *1993 to 2012.* *Epidemiology and Infection*, 2015. **143**(7): p. 1519-1523.
- 663 39. Liu, C.M., et al., *Staphylococcus aureus and the ecology of the nasal microbiome.*
664 *Science Advances*, 2015. **1**(5).
- 665 40. Cremers, A.J., et al., *The adult nasopharyngeal microbiome as a determinant of*
666 *pneumococcal acquisition.* *Microbiome*, 2014. **2**(1): p. 1-10.
- 667 41. Yan, M., et al., *Nasal microenvironments and interspecific interactions influence*
668 *nasal microbiota complexity and S. aureus carriage.* *Cell host & microbe*, 2013.
669 **14**(6): p. 631-640.
- 670 42. Bomar, L., et al., *Corynebacterium accolens Releases Antipneumococcal Free Fatty*
671 *Acids from Human Nostril and Skin Surface Triacylglycerols.* *mBio*, 2016. **7**(1).
- 672 43. De Boeck, I., et al., *Comparing the Healthy Nose and Nasopharynx Microbiota*
673 *Reveals Continuity As Well As Niche-Specificity.* *Frontiers in microbiology*, 2017. **8**:
674 p. 2372-2372.
- 675 44. Biesbroek, G., et al., *Early respiratory microbiota composition determines bacterial*
676 *succession patterns and respiratory health in children.* *Am J Respir Crit Care Med*,
677 2014. **190**.
- 678 45. Laufer, A.S., et al., *Microbial Communities of the Upper Respiratory Tract and Otitis*
679 *Media in Children.* *mBio*, 2011. **2**(1).
- 680 46. Teo Shu, M., et al., *The infant nasopharyngeal microbiome impacts severity of lower*
681 *respiratory infection and risk of asthma development.* *Cell Host Microbe*, 2015. **17**.
- 682 47. Durack, J., H.A. Boushey, and S.V. Lynch, *Airway Microbiota and the Implications*
683 *of Dysbiosis in Asthma.* *Current Allergy and Asthma Reports*, 2016. **16**(8): p. 52.
- 684 48. Ege, M.J., et al., *Exposure to environmental microorganisms and childhood asthma.*
685 *New Engl J Med*, 2011. **364**.
- 686 49. Verma, D., P.K. Garg, and A.K. Dubey, *Insights into the human oral microbiome.*
687 *Archives of Microbiology*, 2018. **200**(4): p. 525-540.
- 688 50. Salter, S.J., et al., *Reagent and laboratory contamination can critically impact*
689 *sequence-based microbiome analyses.* *BMC Biol*, 2014. **12**.
- 690 51. Wang, J., et al., *Human oral microbiome characterization and its association with*
691 *environmental microbiome revealed by the Earth Microbiome Project.* *bioRxiv*, 2019:
692 p. 732123.
- 693 52. Almonacid, D.E., et al., *16S rRNA gene sequencing and healthy reference ranges for*
694 *28 clinically relevant microbial taxa from the human gut microbiome.* *PLOS ONE*,
695 2017. **12**(5): p. e0176555.
- 696 53. Bolyen, E., et al., *Reproducible, interactive, scalable and extensible microbiome data*
697 *science using QIIME 2.* *Nature Biotechnology*, 2019. **37**(8): p. 852-857.
- 698 54. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina*
699 *amplicon data.* *Nat Methods*, 2016. **13**.
- 700 55. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment*
701 *based on fast Fourier transform.* *Nucleic acids research*, 2002. **30**(14): p. 3059-3066.
- 702 56. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree 2--approximately maximum-*
703 *likelihood trees for large alignments.* *PLoS One*, 2010. **5**.
- 704 57. Bokulich, N.A., et al., *Optimizing taxonomic classification of marker-gene amplicon*
705 *sequences with QIIME 2's q2-feature-classifier plugin.* *Microbiome*, 2018. **6**(1): p.
706 90.
- 707 58. McDonald, D., et al., *An improved Greengenes taxonomy with explicit ranks for*
708 *ecological and evolutionary analyses of bacteria and archaea.* *The ISME journal*,
709 2012. **6**(3): p. 610-618.

- 710 59. Bisanz, J.E. *qiime2R: Importing QIIME2 artifacts and associated data into R*
711 *sessions*. 2018; Available from: <https://github.com/jbisanz/qiime2R>.
712 60. Team, R.C., *R: a language and environment for statistical computing*. 2014, Vienna,
713 Austria: R Foundation for Statistical Computing.
714 61. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016: Springer-Verlag
715 New York.
716 62. McMurdie, P.J.H., S., *phyloseq: an R package for reproducible interactive analysis*
717 *and graphics of microbiome census data*. . PLoS One, 2013. **8**(4).
718 63. Callahan, B., et al., *Bioconductor Workflow for Microbiome Data Analysis: from raw*
719 *reads to community analyses [version 2; peer review: 3 approved]*. F1000Research,
720 2016. **5**(1492).
721 64. Davis, N.M., et al., *Simple statistical identification and removal of contaminant*
722 *sequences in marker-gene and metagenomics data*. Microbiome, 2018. **6**(1): p. 226.
723 65. Oksanen, J., et al., *Vegan: Community ecology package*. *R Packag.* 2.3-3. 2016.
724 66. Love, M.I., Huber, W., Anders, S., *Moderated estimation of fold change and*
725 *dispersion for RNA-seq data with DESeq2*. . Genome Biology, 2014. **15**(12).
726

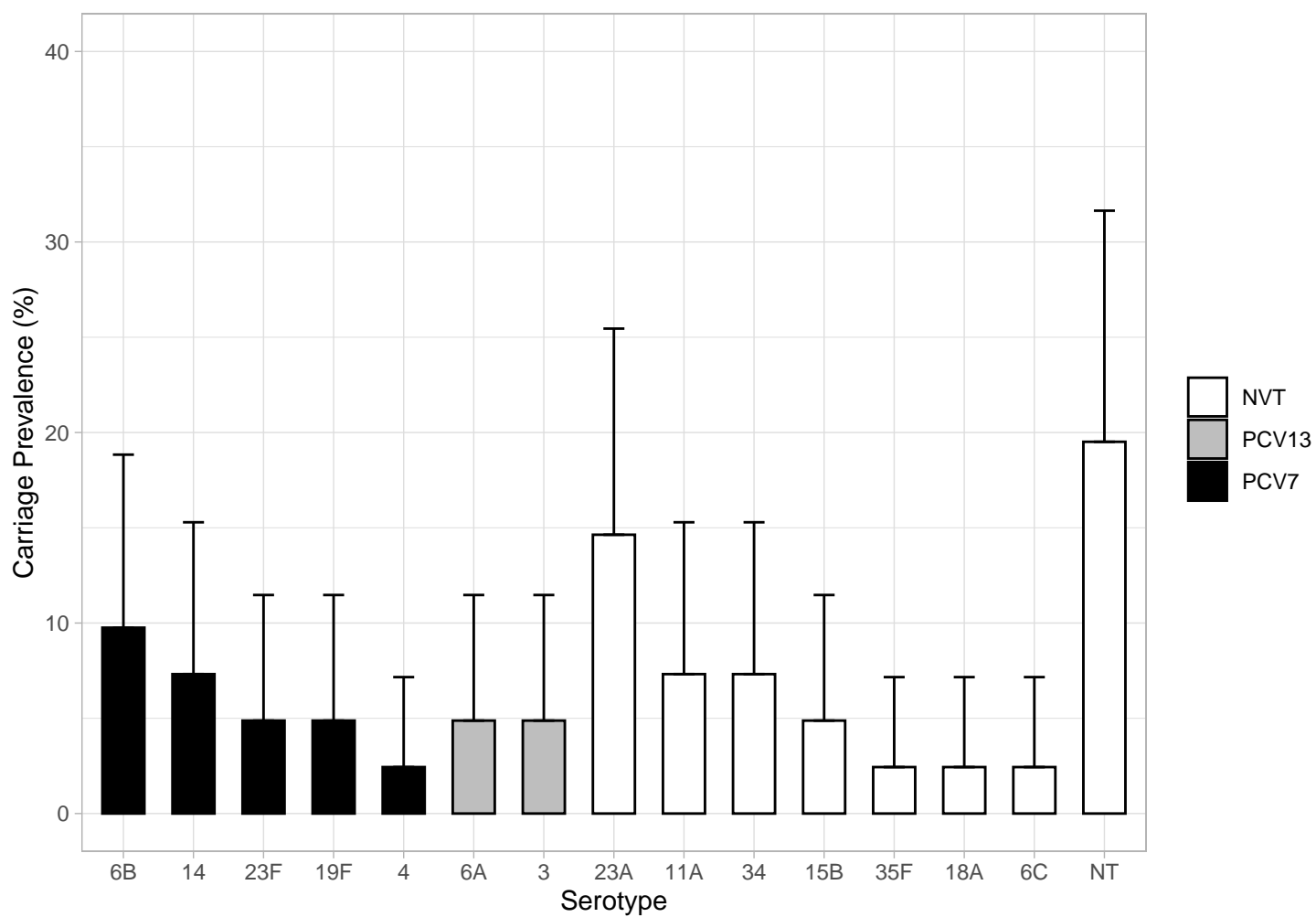


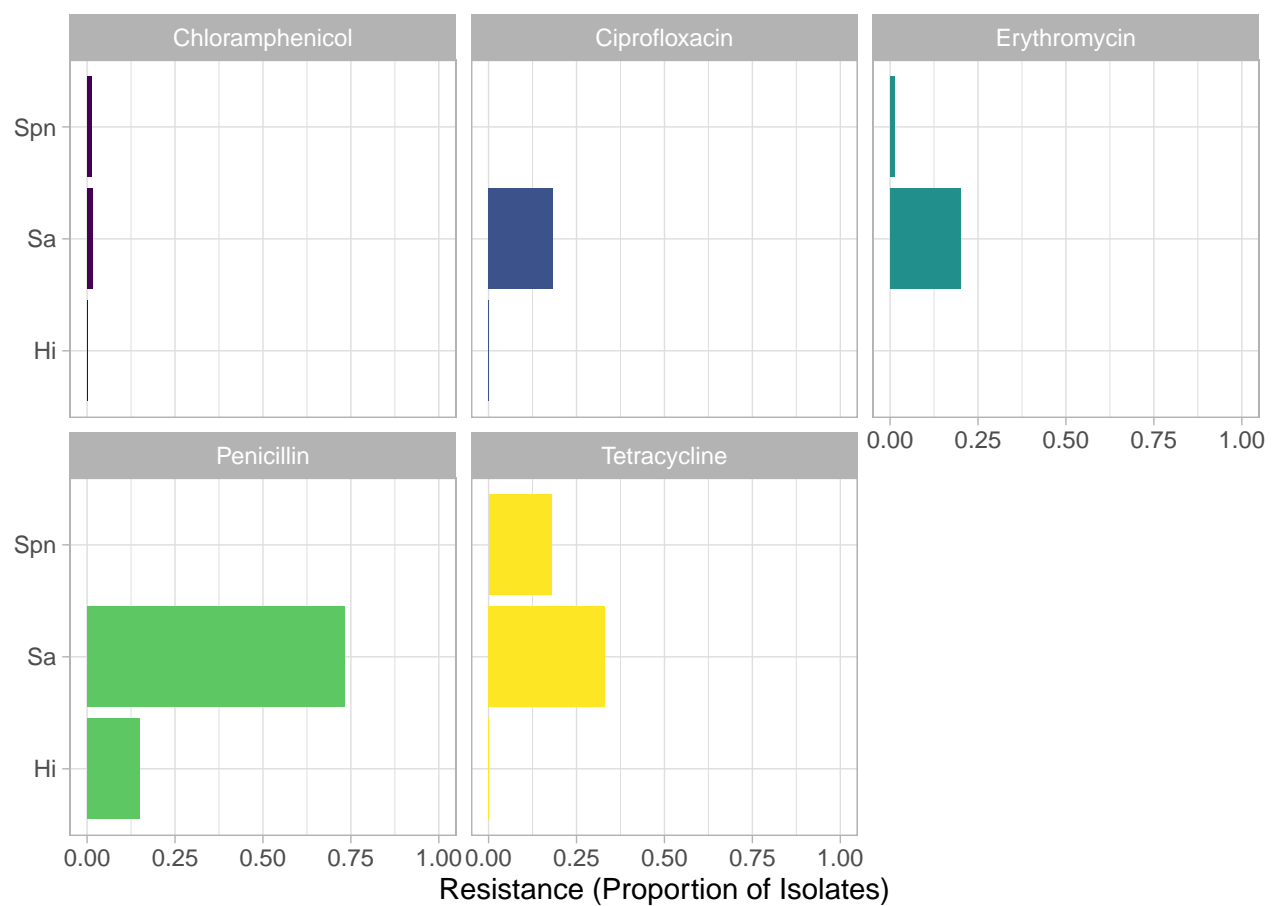
S. pneumoniae*H. influenzae**S. aureus**M. catarrhalis*

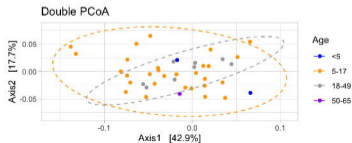
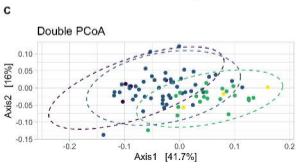
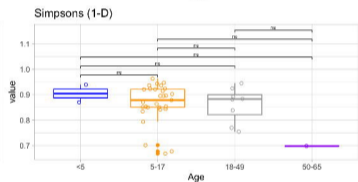
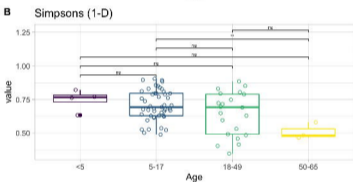
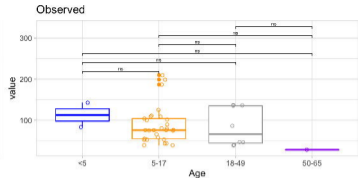
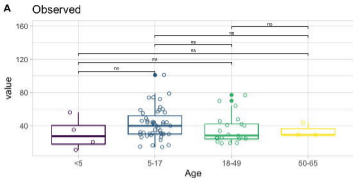
● Combined ● Nose ● NP

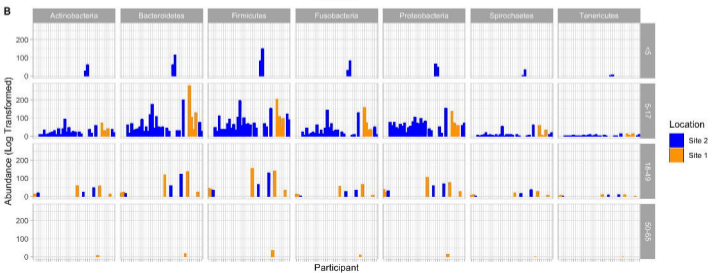
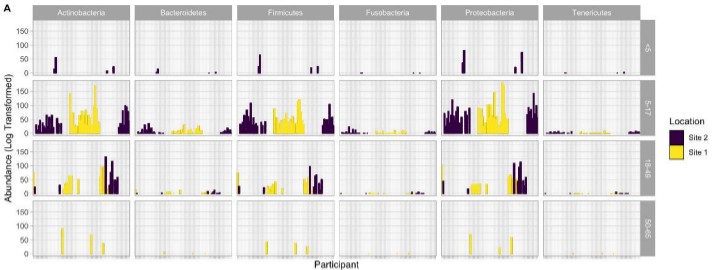
Table 1: Multi- versus single-species carriage of *Streptococcus pneumoniae* (Spn), *Staphylococcus aureus* (Sa), *Haemophilus influenzae* (Hi) and *Moraxella catarrhalis* (Mc). Forty-seven individuals were culture negative. Seven had single-species profiles involving *Klebsiella pneumoniae* (n=3), α -haemolytic *Streptococci* (n=2) or *Pseudomonas aeruginosa* (n=2). A final thirteen had different multi-species profiles to those listed below.

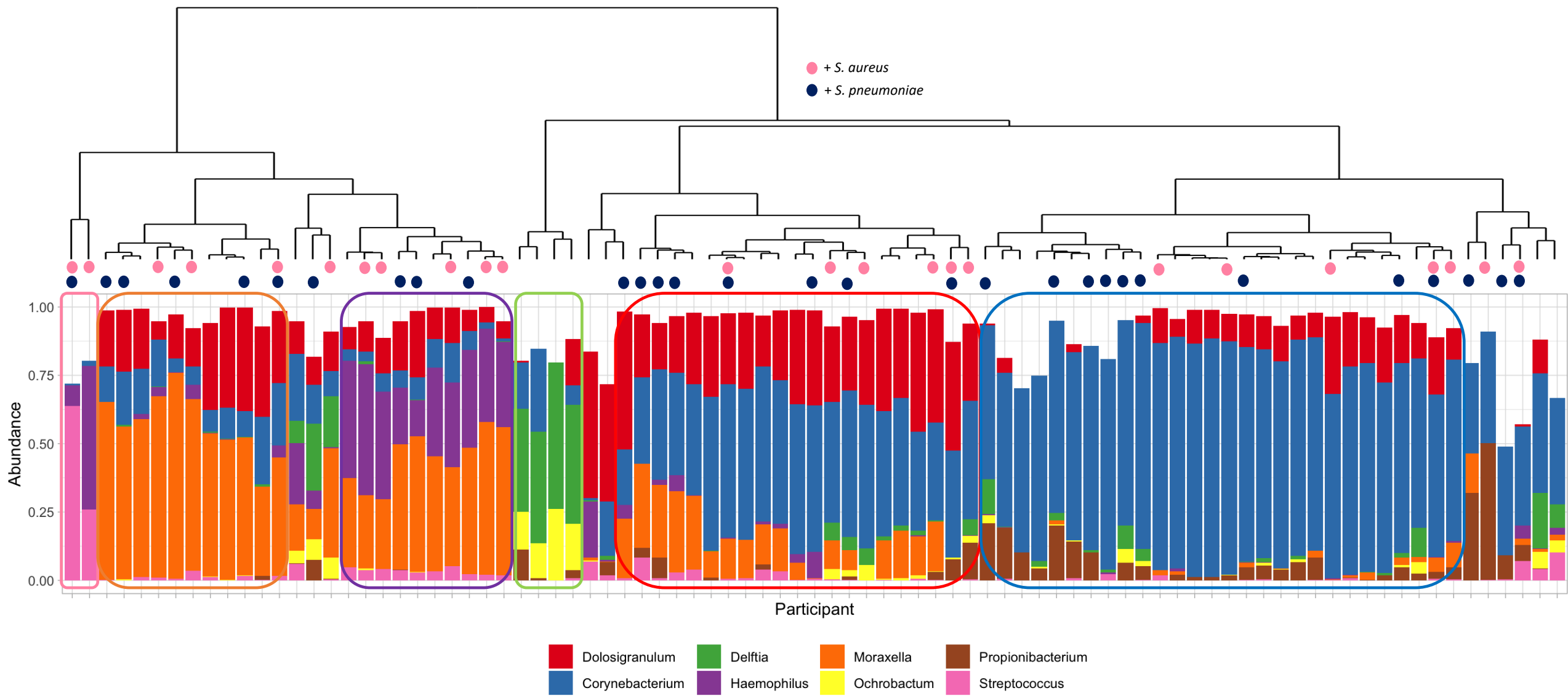
Bacterial Species	N	%
Sa	25	19.23
Spn	16	12.31
Spn-Hi	7	5.38
Spn-Sa	5	3.85
Hi	2	1.54
Mc	1	0.77
Spn-Mc	1	0.77
Hi-Sa	1	0.77
Hi-Mc	1	0.77
Spn-Hi-Sa	1	0.77
Spn-Hi-Mc	1	0.77
Sa-Mc	0	0.00
Hi-Sa-Mc	0	0.00
Sa-Mc-Spn	0	0.00
Sa-Mc-Spn-Hi	0	0.00





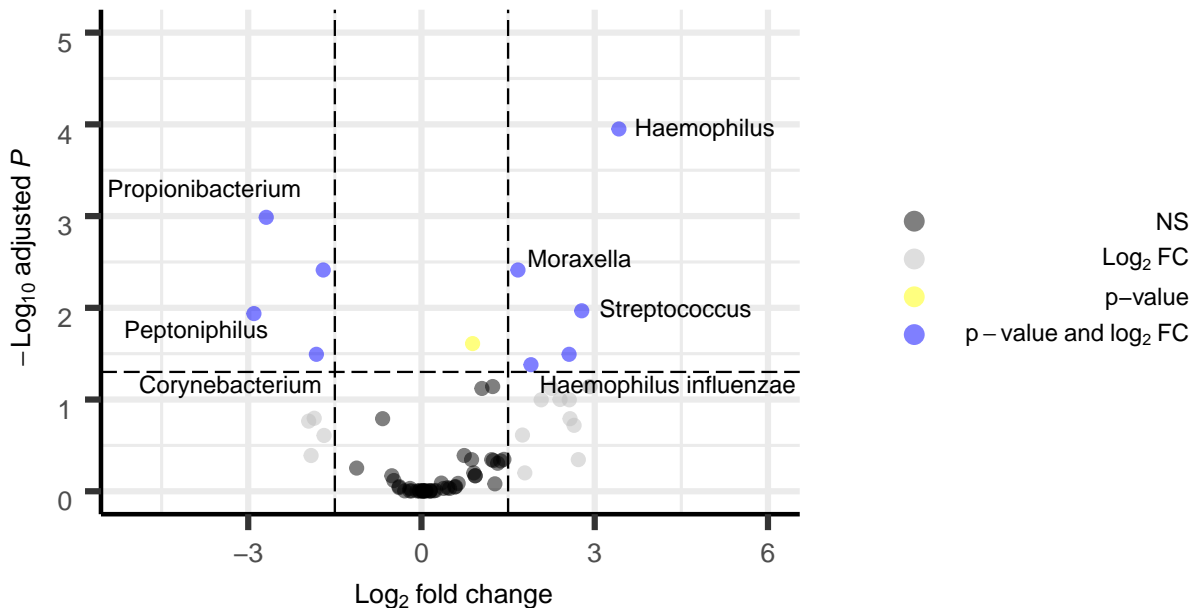






Age Group Comparison 5-17 vs.18-49

EnhancedVolcano



Total = 205 variables

