

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

Authors: David W. Cleary* ^{1,2}, Denise E. Morris ¹, Rebecca A. Anderson ¹, Jessica Jones¹, Ahmed Ghazi Alattraqchi ³, Nor Iza A. Rahman ³, Salwani Ismail ³, Mohamad Sayuti Razali³, Rahmah Mohd Amin³, Aniza Abd Aziz³, Nor Kamarruzaman Esa³, Salman Amiruddin³, Ching Hoong Chew ⁴, Mohamad Hafis Amat Simin⁵, Ramle Abdullah⁵, Chew Chieng Yeo ³ and Stuart C. Clarke ^{1,2,6,7,8}

Affiliations:

¹: Faculty of Medicine and Institute for Life Sciences, University of Southampton, Southampton, UK

²: NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Trust, Southampton, UK

³: Faculty of Medicine, Universiti Sultan Zainal Abidin, Medical Campus, 20400 Kuala Terengganu, Terengganu, Malaysia

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

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

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

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

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

26

27 *Corresponding author

28 **School of Clinical and Experimental Sciences & Clinical Informatics Research Unit**

29 **Faculty of Medicine, University of Southampton**

30 **Room LC71A, Mailpoint 814, Level C**

31 **Sir Henry Wellcome Laboratories South Block**

32 **University Hospital Southampton Foundation NHS Trust**

33 **Southampton, SO16 6YD**

34 **+44 (0)2381 208896**

35 **d.w.cleary@soton.ac.uk**

Abstract:

Much microbiome research has focused on populations that are predominantly of European descent, and from narrow demographics that do not capture the socio-economic and lifestyle differences which impact human health. Here we examined the airway microbiomes of the Orang Asli, the indigenous peoples of Malaysia. A total of 130 participants were recruited from two sites in the north-eastern state of Terengganu in Peninsular Malaysia. Using 16S rRNA sequencing, the nasal microbiome was significantly more diverse in those aged 5-17 years compared to 50+ years ($p = 0.023$) and clustered by age (PERMANOVA analysis of the Bray-Curtis distance, $p = 0.001$). Hierarchical clustering of Bray-Curtis dissimilarity scores revealed six microbiome types. The largest cluster ($n=28$; 35.4%) had a marked abundance of *Corynebacterium*. In the oral microbiomes *Streptococcus*, *Neisseria* and *Haemophilus* were dominant. Using conventional microbiology, high levels of *Staphylococcus aureus* carriage were observed, particularly in the 18-65 age group ($n=17/36$; 47.2% 95%CI: 30.9-63.5). The highest carriage of pneumococci was in the <5 and 5 to 17 year olds, with 57.1% (4/7) and 49.2% (30/61) respectively. Sixteen pneumococcal serotypes were identified, the most common being the non-vaccine type 23A (14.6%) and the vaccine type 6B (9.8%). The prevalence of pneumococcal serotypes covered by pneumococcal conjugate vaccines support introduction into a Malaysian national immunisation schedule. In addition, the dominance of *Corynebacterium* in the airway microbiomes is intriguing given role as a potentially protective commensal with respect to acute infection and respiratory health.

59 **Keywords:** 16S rRNA; microbiome; upper respiratory tract; antimicrobial resistance;
60 Malaysia; *Streptococcus pneumoniae*; *Haemophilus influenzae*; *Staphylococcus aureus*;
61 *Moraxella catarrhalis*; Orang Asli

Introduction:

Much microbiome research has, to-date, been focused on populations that are predominantly of European descent, and from demographics that do not capture the socio-economic and lifestyle challenges which impact human health ¹. The study of non-Western, unindustrialised populations is therefore an important extension of microbiome research so that we may understand human-host microbiota interactions in their broadest sense. Notable endeavours here, which have principally focussed on gut microbiomes, include studies of the Hadza hunter-gatherers of Tanzania ², Amerindians of South America ^{3,4}, agriculturalist communities in Burkina Faso ⁵ Malawi and Venezuela ⁶ and the Cheyenne and Arapaho Tribes of North America ⁷. Fewer studies have gone beyond gut microbiomes to sample additional body sites, such as those of the airways. The analysis of salivary microbiota of the Batwa Pygmies ⁸, the Yanomami of the Venezuelan Amazon ³ and Guahibo of Venezuela ⁹ have shown the benefits of including these body sites where, respectively, previously undiscovered genera have been identified and novel combinations of taxa described. However, there remains a paucity of data regarding microbiomes of other anatomical sites of the upper respiratory tract (URT). This lack of data requires addressing given the evidence that geographically isolated and/or indigenous populations endure distinct burdens of respiratory infections. For example children of Warao communities in Venezuela have been shown to be at increased risk of pneumococcal disease¹⁰, and it is well established that Aboriginal infants are highly susceptible to otitis media caused by Non-typeable *Haemophilus influenzae* (NTHi)¹¹. With respiratory infectious diseases continuing to be a significant component of global morbidity and mortality ¹², the interest in the microbiome of the upper airways stems from its' importance in an individual's susceptibility to respiratory infection, in part through the presence of resilient taxa which prevent colonisation and/or outgrowth of specific pathobionts ¹³. Understanding the variability in airway microbiomes is

a key first step to leveraging these interactions for health. Data from the Gambia during the introduction of the seven valent pneumococcal conjugate vaccine (PCV7) highlighted changes in nasopharyngeal microbiome following implementation; a finding only possible due to the efforts to understand the microbial ecology as PCV7 was being rolled out¹⁴. To date no study has undertaken a comprehensive analysis of the URT microbiome of the indigenous populations of Peninsular Malaysia, the Orang Asli.

Although the name Orang Asli was only introduced around 1960 by the British, as an ethnic group they are believed to be the first settlers of Peninsular Malaysia, moving from Northern Thailand, Burma and Cambodia 3-8000 years ago. Orang Asli is in fact a catch-all term for three tribal groups, the Senoi, Proto-Malays and Negrito, each of which in turn can be further divided into six ethnic groups ¹⁵. They number ~179 000, 0.6% of the population of Malaysia but are disproportionately impacted by economic marginalisation and discrimination ¹⁶. As a consequence, nearly 80% of the Orang Asli population are beneath the poverty line compared to 1.4% of the national population, and this hardship is reflected in a 20-year lower average life expectancy of just 53 years ¹⁷. Several studies have highlighted the susceptibility of these populations to specific diseases ^{18,19}. More recently a study of 73 adults from the semi-urbanized Temiar, an Orang Asli tribe of Kampong Pos Piah, in Perak, examined salivary microbiomes in the context of obesity – a growing concern and evidence of the epidemiological shifts in disease burden as these communities leave their more traditional existences ²⁰. However, there is no published data on the rates and characteristics of respiratory tract infections in the Orang Asli.

Here we examined the microbiology of the upper airway of two Orang Asli communities located in Terengganu state, North-east of Peninsular Malaysia. The nasopharyngeal and

112 nasal (anterior nares) carriage of important pathobionts was determined by culture, and the
113 resistance of these isolates to clinically relevant antibiotics tested. We also present the first
114 investigation of the nasal microbiomes of these indigenous peoples, in addition to their oral
115 microbiomes.

Results:

Participant Demographics

A total of 130 participants were recruited to the study, with 68 from Kampung Sungai Pergam (Site 1) and 62 from Kampung Berua (Site 2) (Figure 1). The age and gender distribution are shown in Figure 1. Of those for whom age was accurately recorded, 49% were female and 45% male. Recruitment of children under five years old was low accounting for only 11.5% of the total population sampled, for which there was no obvious reason. All those with missing age data were adults. A comparison of participants from the two sites is shown in Table 1. A significant difference was seen for Site 1 which had a greater proportion of adults (18 to <85 year olds; $p < 0.05$) compared to Kampung Berua. In addition, self-reported respiratory illness was higher in both the previous month and three months prior to the site visits ($p < 0.001$). No difference in self-reported smoking or vaccination status was observed.

Carriage of Bacterial Pathobionts

The carriage prevalence of the four most commonly isolated pathobionts in the anterior nares and nasopharynx are shown in Figure 2. Here carriage was defined as the isolation of a pathobiont from either the nasal swab used for culture or the nasopharyngeal swab. The most commonly isolated pathobionts were *S. aureus* and *S. pneumoniae* both with a total of 39 individuals colonised. The highest carriage for *S. aureus* was in the 18-65 age group at 47.2% ($n=17/36$; 95%CI: 30.9-63.5), followed by 30% ($n=18/60$, 95%CI: 18.4-41.6) in the 5 to 17 year olds. In contrast, the highest carriage of pneumococci was in the <5 and 5 to 17 year olds, with 57.1% (4/7) and 49.2% (30/61) respectively. Looking at the latter cases in more detail shows that the average age of a pneumococcal carrier in this age group was ~9 years old (range: 6-13). Examining differences in carriage of *S. pneumoniae* between female and

male participants showed no significant differences (Supplementary Figure 1). However, for *S. aureus* there was an increased carriage prevalence in females (57.7% (n=15/36, 95% CI:38.7-76.7) in the 18 to <65 years old age group, giving a significant ($p = 0.0257$) OR of 28.30, albeit it with large 95% CI: 1.50-534.25. The next most carried bacterium was *H. influenzae*, isolated from seventeen individuals, the majority of which (88.2%) were in the 5 to 17 years old age group. In contrast, *M. catarrhalis* was rarely isolated (n=6), as was *N. meningitidis* (2), *P. aeruginosa* (n=3), *K. pneumoniae* (n=6) and α -haemolytic *Streptococci* (n=7). Carriage and co-carriage are summarised in Supplementary Table 1. Carriage of only *S. aureus* or *S. pneumoniae* accounted for 31.5% of the profiles observed. A limited number of co-carriage incidents were found which included seven and five cases of pneumococci isolation with *H. influenzae* or *S. aureus* respectively.

Pneumococcal Serotypes

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

Antibiotic Resistance

High levels of resistance were observed for *S. aureus* where 73.3% (n=44) of isolates were resistant to penicillin, 33.0% (n=20) to tetracycline and 18.3% (n=11) to ciprofloxacin (Supplementary Figure 2). However, all isolates were sensitive to ceftazidime, and all but one to chloramphenicol. Three of the *H. influenzae* isolates (15%) were resistant to penicillin, all of which were negative for β -lactamase activity. Of the *S. pneumoniae*, 17.9% (n=12) were resistant to tetracycline. No resistance was seen for *M. catarrhalis* against any antibiotic tested. *A. baumannii* were sensitive to meropenem and ciprofloxacin (n=14). Similarly, all *K. pneumoniae* were sensitive to all antibiotics tested including ceftazidime and meropenem.

Upper Respiratory Tract Microbiomes

The microbiome of the nasal (anterior nares) and oral (whole mouth) was examined using 16S rRNA sequencing of the V4 hypervariable region. Here, the mean number of sequences per sample, excluding negative controls, was 64,694 with a maximum of 1,072,721. Negative control samples had a median of 1077 sequences, ranging from 692 to 1700. The frequency and classification of ASVs identified at Phylum and Genus level can be found in Supplementary Figure 3. Sixteen nasal and two oral samples were excluded based on low numbers of ASVs (<1000). Eleven of these samples had ASVs only classified as Bacteria. The remaining five (comprising two oral and three nasal samples) also had a sequencing depth of <1000 reads. Alpha diversity (Figure 4A and 4B) was similar between all ages apart from nasal samples from those aged 5-17 and 50-65 where a significantly greater diversity was observed in the younger group ($p = 0.0023$), although notably fewer samples were available from the older participants. Nasal samples are shown to cluster by age using DPCoA (Figure 4C - left column) and PERMANOVA analysis of the Bray-Curtis distance showed that this was a significant difference ($p = 0.001$). In contrast, again using age as the grouping factor, this was not seen for oral samples (Figure 4C – left column). Examining the

most abundant phyla (Supplementary Figure 4) showed nasal samples were dominated, as expected, by Firmicutes, Proteobacteria and Actinobacteria. For oral samples Bacteroidetes was a much more prevalent phyla in addition to Fusobacteria, although those same three phyla which dominated the nasal samples were also present at high levels here as well. Hierarchical clustering of Bray-Curtis dissimilarity scores from nasal samples revealed six clusters that could be characterised by the dominance of only one or two Genera (Figure 5). The largest cluster (n=28; 35.4%) contained individuals with a marked abundance of *Corynebacterium*, accounting for an average ASV relative abundance of 73.2% (range: 59.5 – 87.2%). The second largest cluster (n=16; 20.3%) also had a substantial proportion of *Corynebacterium*, in this case with *Dolosigranulum* as significant co-carried genus. Here, on average, *Corynebacterium* accounted for 48.9% and *Dolosigranulum* 31.6% of the relative abundance; combined these Genera accounted for between 71.5 and 89.4% of the ASV relative abundance. Two clusters were then definable by the presence of *Moraxella*, one with and the other without *Haemophilus*. In the *Moraxella*-dominated cluster, between 48.2 and 75.2% of the ASV relative abundance could be attributed to this Genus. When combined with *Haemophilus* this proportion dropped slightly from an average of 58.1% to 41.5% and here *Haemophilus* accounted for 13.2 to 47.9%. A small grouping of samples with *Delftia/Ochrobactrum* as the dominant genera were also found. The final cluster consisted of two distinct profiles, where in one *Streptococcus* accounted for 63.8% and the second with *Streptococcus/Haemophilus* at 25.9 and 52.6% respectively.

Co-occurrence of ASVs are shown in Figure 6. Evidence of co-occurrence between ASVs of *Haemophilus* with *Moraxella* and *Streptococcus* can be seen, in particular the latter. There appears both negative and positive associations between *Corynebacterium* and *Haemophilus*. *Dolosigranulum* is linked to ASVs of both *Moraxella* and *Corynebacterium*. The

Ochrobactrum-Delftia ASVs appear within a collection of nodes with other environmental taxa – further suggesting profiles dominated by these taxa (Figure 5) are the result of contamination.

Double PCoA ordination and Jaccard distance-based network analyses were used to examine nasal microbiomes in the context of self-reported symptoms of respiratory illness (Supplementary Figure 5). It was not possible to delineate samples based on symptomology given that there were significant differences between location. We further examined log₂-fold changes in Genera abundances using DESeq2 but found no significant taxa associated with individuals that reported URT illness within the previous one month or three months.

No correlation between microbiome composition, in terms of dominant genus, and culture outcomes for *S. aureus* or *S. pneumoniae*, as the two most frequently cultured pathobionts, were seen (Figure 5). To determine if the abundance of either genus was related to culture positivity, the relative abundance of each was compared between samples from which either *S. aureus* or *S. pneumoniae* was isolated (Supplementary Figure 6). The relative abundance of Staphylococcal ASVs was not significantly different between groups, however those from whom *S. pneumoniae* were cultured did have a significantly higher relative abundance of Streptococci ASVs ($p = 0.016$); although this may be due to the presence of two profiles characterised by a significant dominance of *Streptococcus*. To explore the differences in microbiomes between the younger and older age groups, differentially abundant ASVs were identified using DESeq2 (Supplementary Figure 7). The 5-17-year age group had ASVs classified as *Haemophilus* (including *H. influenzae*), *Moraxella* and *Streptococcus*. In contrast, *Propionibacterium*, *Peptoniphilus* and *Corynebacterium* ASVs were significantly increased in adults.

241
242 *Streptococcus*, *Neisseria* and *Haemophilus* dominated the oral microbiota (Supplementary
243 Figure 8). Lower levels of *Prevotella*, *Rothia*, *Porphyromonas*, *Veillonella* and
244 *Aggregatibacter* were also among the eight most observed genera. Streptococcal-dominated
245 profiles were characterised by 43.7% ($\pm 15.7\%$) relative abundance of ASVs belonging to this
246 genus. Those with a *Streptococcus*/*Haemophilus* profile still had 30.7% ($\pm 7.6\%$) Streptococci
247 ASV relative abundance but with 15.0% ($\pm 10\%$) *Haemophilus*. The final profile, where clear
248 dominance of one particular genus was observed, were those where *Neisseria* accounted for
249 30.6% (± 12.5) of the relative abundance of all ASVs.

Discussion:

As a site that can be colonised by commensals, as well as opportunistic pathogens such as the pneumococcus²¹, *S. aureus*²² and *H. influenzae*²³, understanding the microbiology of the upper respiratory tract is key in tackling respiratory disease. The Anna Karenina principle of microbiomes, that “all healthy microbiomes are similar; each dysbiotic microbiome is dysbiotic in its own way” is a useful maxim for examining microbiomes in the context of disease²⁴. However, from previous studies of communities that live unique, traditional lifestyles a clear challenge to the paradigm of health mirroring health has been shown^{3,8}. Here we examined the respiratory microbiomes of Orang Asli communities living traditional lifestyles in Terengganu, a rural state in the North-east of Peninsular Malaysia. These communities are known to face specific health challenges to that of the broader population of Malaysia, a country where respiratory disease is still a significant burden²⁵. Our study showed a high prevalence of *S. aureus* carriage but interestingly not the presence of Staphylococci-dominated nasal microbiomes. Instead, *Corynebacterium*, with *Dolosigranulum* and *Moraxella* dominated as commensal flora. We also showed the potential impact that PCV introduction in Malaysia may have on circulating pneumococcal serotypes.

Although pneumococcal carriage can vary globally between 10 and 90%²⁶⁻²⁸, the highest carriage prevalence is usually seen in young children, particularly those <5 years²⁸. Here, over half of the <5 year olds were colonised which is suggestive of high levels of carriage in this age group, although we note a very limited sample size (n=7). Examining the next age group, 5-17, we observed a higher carriage than seen elsewhere in older children²⁹. However, given the relatively young age of this cohort at ~9 years old, the higher carriage observed (50%; 95%CI, 37.3 and 62.7%) perhaps supports our proposition of high paediatric carriage. Of interest is that this level of carriage is similar to the 60.9% observed in Aboriginal

children over 5 years old – a demographic known to experience high rates of invasive pneumococcal disease (IPD) ³⁰. As expected, the carriage in adults was low at 8.3% (95%CI, 0% to 17.4%). For context this is similar to the NP carriage prevalence of 11.1% (95%CI, 9.8 and 12.6%) seen in the USA ³¹, slightly higher than recent surveys in the UK which found a 2.8% (95%CI: 1.2–5.5) prevalence in this age group ²⁹ but significantly lower than that seen in Gambian adults where pneumococcal carriage exceeds 50% in all age groups³². As for serotypes, whilst there is growing epidemiological data for the region ³³, there is limited data for Malaysia. A study of 245 cases of paediatric IPD between 2014 and 2017 showed that serotypes 14, 6B, 19A, 6A and 19F were the most common, accounting for ~75% of disease ³⁴. This is concordant with this study where serotypes 14 and 6B were also the most common in carriage with 6A and 19F being identified. Of the non-VT serotypes isolated only 34, 35F and 18A did not feature in those identified in the IPD study. The lack of 19A is intriguing but we suspect a consequence of the relatively low sampling. Both PCV10 and PCV13 are available in Malaysia however only in private practice and therefore the presence of VT serotypes is not surprising. Given the presence of vaccine types here that have all but been eliminated from carriage in countries where public health initiatives have introduced these or similar vaccine these data do suggest introduction of PCV into the national immunisation programme as planned may prove efficacious.

Only one of the children less than 5 years old was culture positive for *H. influenzae* but given the low numbers in this age group this must be interpreted with caution. Whilst certainly higher than in the UK where only 6.5% *H. influenzae* carriage was seen in older children ³⁵, the 25% (95%CI, 14.0 and 40.0%) found here is similar to that observed in other developing countries such as Kenya ³⁶, but lower than the 52.8% seen in Aboriginal children aged between 5 and 15 years old ³⁰. Whether this increased carriage translates into disease risk is

difficult to determine; there is no data on the burden of otitis media among Orang Asli, although the country wide incidence in children <12 year olds is low at 2.3% for the Asia-Pacific region ³⁷. It is likely that there is not a burden of *H. influenzae* serotype B given the high coverage (98% in 2019) of the Hib vaccine in Malaysia³⁸, although accurate health records for these communities renders this an assumption.

The prevalence of *M. catarrhalis* at 6.7% (95%CI, 0.35 and 13.0%) was very similar to that of our previous UK study where 5.7% of the same age group were colonised ³⁵, a level also observed in other western countries ³⁹. This is a stark contrast to the 67% in Aboriginal children of a similar age ³⁰, however this low level of carriage (6%) has been seen in Warao Amerindians in Venezuela ⁴⁰ – arguably a similar demographic to the one being presented here.

S. aureus carriage is known to vary markedly, by age but also by geographic location ⁴¹. Although the lowest prevalence observed here was in <5 year olds (28.6%; 95%CI: 0-62.0), which is surprising as one would expect the highest carriage would be seen in this age group, it could be either a feature of the low recruitment or that the average age of this group was 2.5 and carriage is known decline between 1 and 10 ²². As for the carriage in adults, a previous study in Malaysia of 384 adults (students with an average age of 25) found a carriage prevalence of 23.4% ⁴² which is much lower than the 47.2% in our population of Orang Asli. A similarly high level of adult carriage (41.7% and 57.8% at two separate sampling timepoints) was observed in Wayampi Amerindians who, in terms of remote living, are similar to the Orang Asli ⁴³. Our own research has previously shown adult carriage at 24.4% in a UK population ³⁵ and thus the differences we observed are unlikely to be methodological. More likely it is a consequence of greater co-habitation (the average

household size of our communities being ~5 people) and poorer sanitation infrastructure, both of which have been highlighted as reasons why *S. aureus* carriage has declined in other parts of the world ²². It also is interesting that in our study females had significantly greater odds of carrying *S. aureus* which is in contrast to the typical burden in males⁴⁴.

Of the ESKAPE pathogens isolated and tested only *S. aureus* exhibited any substantial level of resistance although the prevalence of resistance to Penicillin was in keeping with the most recent National Antibiotic Resistance Surveillance Report from 2018 where 80.6% of 34 564 isolates were also resistant ⁴⁵; importantly all isolates from this study were sensitive to methicillin. Whilst there appears to be limited resistance it is important to note that this represents the only data on the antimicrobial resistance (AMR) in these communities and is therefore merely a starting point for future studies. This is particularly important given that AMR burden in indigenous populations has been seen to increase elsewhere ⁴⁶.

URT microbiome research has shown that the composition and/or dominance of particular bacterial taxa can be indicative of nasal community state types (CSTs). Previously, in a study of adults in the USA, seven CSTs were identified that were characterised by *S. aureus* (CST1), or other *Staphylococci* (CST3), a mix of Proteobacteria including *Escherichia* and *Enterobacteriaceae* sp. (CST2), *Corynebacterium* either with *Propionibacterium acnes* (CST4) or without (CST5), *Moraxella* (CST6) and *Dolosigranulum* (CST7) ⁴⁷. It is here that we see the most startling contrasts to our Orang Asli population. Although we also identified seven clusters based on what is/are the dominant taxa we did not see anything resembling CST 1 or 3 (*Staphylococci*), and no CST2. Although we observed a high carriage prevalence of *S. aureus*, it is likely that this discordance with CSTs is, in part, an artefact of the ease with which this bacterium is cultured from nasal samples. Nevertheless, the lack of a dominant

350 molecular signature is intriguing. Here we hypothesise that the high prevalence of profiles
351 similar to CST5, *Corynebacterium* dominated, along with another that is defined by high
352 abundance of both *Corynebacterium* and *Dolosigranulum*, a composition that has been found
353 in the nasopharynx⁴⁸, explains this absence and is based on previous work showing that
354 *Corynebacterium* reduces *S. aureus* carriage⁴⁹. A similar interaction may also explain the
355 absence of *Streptococci*⁵⁰ – although we recognise in both cases that our hypothesis requires
356 further testing due to the identification of both *S. aureus* and pneumococcal carriage among
357 our participants. A *Moraxella* dominated profile similar to CST6, is also present in the Orang
358 Asli, a profile which has also been observed in healthy adults in a European setting⁵¹, as well
359 as in both Kenyan⁵² and Fijian children⁵³ (where it is more common in iTaukei, the
360 indigenous Fijians). We also saw a *Moraxella-Haemophilus* profile which hasn't been
361 described previously. The *Streptococcus* profiles are both from older children (6 and 10 years
362 old) and this not uncommon as a profile among this age group^{52,54}. Given the low numbers it
363 is impossible to determine whether we would have found more *Streptococcus/Haemophilus*
364 profiles as been described elsewhere⁵². The significant proportion of individuals who harbour
365 profiles that are being considered to be potentially protective either in terms of acute
366 respiratory infections⁵⁵ or chronic disease such as asthma⁵⁶⁻⁵⁸, is intriguing. Similar profiles
367 were observed in Fijian children with either *Dolosigranulum* or
368 *Haemophilus/Corynebacterium* dominant profiles⁵³ although there are important contrasts,
369 remembering that these were observed in older children and adults, we observed
370 *Corynebacterium* and *Corynebacterium/Dolosigranulum* profiles in the Orang Asli. Lastly,
371 whilst *Delftia* is not uncommon as an oral bacterium⁵⁹, given the common finding of *Delftia*
372 and *Ochrobactrum* as common contaminants in microbiome research⁶⁰ we interpret these
373 four samples with extreme caution. The presence of these profiles is in spite of our efforts to
374 control for such contaminants through the collection and sequencing of control swabs.

Epidemiologically, all four samples come from adult females, but two each from Kampung Berua and Kampung Sungai Pergam. The samples are below the 25% quantile for ASV number ($n=3831$). It is possible that these reflect true profiles, but it is impossible to rule out process error during sample handling, extraction and/or sequencing.

The microbiota recovered from the oral cavity was substantially more diverse than the nasal samples, as expected⁵⁹. The top three genera were unremarkable in terms of what is considered to be most commonly observed, mirroring exactly that found in a study of 447 datasets deposited as part of large-scale microbiome projects⁶¹. A previous study of Temiar, Orang Asli in Perak, found a significant difference in oral microbiota compositions between males and females²⁰, however a PERMANOVA analysis of the data presented here showed no such distinction ($p = 0.41$).

Whilst this piece of research was highly informative as a first glimpse into the respiratory microbiology of Orang Asli communities, it could nevertheless have been strengthened in a number of ways. Firstly, it was only possible to visit two sites on a single occasion, and to conduct a pragmatic survey of the communities. As a consequence, the study lacks follow-up and longitudinal data. Whilst we were very successful at recruiting older children and adults, future studies will have to identify how to recruit greater numbers of younger children in particular. In addition, those >65 years old were poorly represented as well. Whilst we have included self-reported symptoms of respiratory infection, we have consciously avoided over-interpretation of these data. Health records are non-existent for these communities, and there are suggestions that these communities suffer from a lack of adequate provision and/or are reluctant to seek out healthcare⁶², and thus it is impossible to determine their validity. It is intriguing that participants from Site 1 reported significantly greater incidences of respiratory

400 infection, both in the preceding 1 and 3 months, and suggests community-specific health
401 burdens. That this does not correlate to the greater incidence of pneumococcal serotypes
402 associated with disease, and is not associated with any notable difference in microbiome
403 composition in individuals from this particular site opens this as a question for future study.
404 In terms of pathobiont carriage, serotyping of more than one pneumococcal isolate from each
405 individual would have enabled a determination of the prevalence of multiple serotype
406 carriage. Serotyping of *H. influenzae* would also have been informative, although future work
407 on exploring the genomics of all isolates will address this in more detail. This additional
408 insight could also be expanded to include all the pathogens/pathobionts isolated. Moreover,
409 here we have relied on culture to determine carriage which can be less sensitive than
410 molecular methods as we have and others have shown previously ^{35,63}. As with all
411 microbiome research, longitudinal sampling taking into account an individual's temporal
412 variability, the seasonality and medical history would be desirable. Finally, the lack of data
413 on bacterial load leaves open the question, particularly with respect to nasal profiles, of
414 whether this may be a factor in the differences we observed in CSTs.

415

416 Important findings from this study include the prevalence of pneumococcal serotypes that
417 would be covered by pneumococcal conjugate vaccines, supporting their introduction as part
418 of a national immunisation programme. The high prevalence of *S. aureus* carriage is
419 noteworthy and warrants further study. In terms of the airway microbiomes the dominance of
420 *Corynebacterium*, additionally with *Dolosigranulum* and *Moraxella* in nasal profiles is
421 particularly intriguing and future work should explore these commensals in the context of
422 burden and susceptibility to both acute and chronic respiratory conditions in these
423 communities.

Methods:

Ethics: Ethical approval for this study was provided by Universiti Sultan Zainal Abidin (UniSZA) Ethics Committee: approval no. UniSZA/C/1/UHREC/628-1(85) dated 27th June 2016, the Department of Orang Asli Affairs and Development (JAKOA): approval no. JAKOA/PP.30.052Jld11 (42), and by the University of Southampton Faculty of Medicine Ethics Committee (Submission ID: 20831). Informed consent was taken for all participants, with parents/guardians providing consent for those <17 years old. Participation in the study involved reading (or being read) and understanding the Malay-translated participant information sheet, and the completion of a consent form and questionnaire. These documents were translated into Malay by A.S.M.H and A.R from the Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin who are familiar with the Orang Asli and have on-going social sciences-based research with them, and by M.A.R and C.C.Y from the Faculty of Medicine, Universiti Sultan Zainal Abidin, who are proficient in Malay and who ensured that the appropriate research and clinical context have been preserved in the translated documents. The information sheet includes guidelines about the confidentiality and anonymity of the data obtained. This process was facilitated by native-speaking research assistants from the Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin, who underwent prior training on how to administer the questionnaires for the purpose of this study.

Study sites and participants: Two Orang Asli villages were visited in August 2017 – Kampung Sungai Pergam in Kemaman district on the 8th August, and Kampung Berua in Hulu Terengganu district on the 9th August. Both sites are located in the state of Terengganu which lies in the north-east of Peninsular Malaysia. Maps were created using the R packages

tmap⁶⁴ and ggmap⁶⁵. Participant recruitment was with consent, across all ages with no exclusion criteria. Questionnaires were used to capture participant metadata which included gender, age, number of dwelling co-occupants and occupation in addition to health-related questions including current or recent (within the last month and last three months) respiratory symptoms, antibiotic use and vaccination status.

Swab collection: A total of four swabs were taken from each participant on the day of visit to the study site. A nasal swab (anterior nares) and a nasopharyngeal swab, using rayon tipped transport swabs containing Amies media with charcoal (Medical Wire and Equipment, Corsham, UK), were used for bacterial pathobiont isolation. A further nasal swab, taken from the un-swabbed nostril, and an oral (whole mouth) swab, supplied by uBiome Inc. (San Francisco, USA), were taken for microbiome analysis. For each study site visited four swabs (two each of nasal and oral) were opened as per the swabbing protocol but then placed, without use, into the uBiome Inc. transport containers for processing alongside used swabs as negative controls.

Bacteriology: Isolation of the common respiratory pathobionts *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, *N. meningitidis*, *S. aureus*, *K. pneumoniae* and *P. aeruginosa* was done by culture. Swabs were plated onto CBA (Columbia blood agar with horse blood), CHOC (Columbia blood agar with chocolate horse blood), CNA (Columbia Blood Agar with Colistin and Naladixic Acid), BACH (Columbia Agar with Chocolate Horse Blood and Bacitracin), GC (Lysed GC Selective Agar) and, for *P. aeruginosa*, CFC (*Pseudomonas* CFC Selective agar) (all Oxoid, UK). Swabs were then vortexed in skim milk, tryptone, glucose and glycerol (STGG)⁶⁶, and 10µL of the suspension was then plated onto the appropriate media. Plates were then incubated for 24-48 hours at 37°C in 5% CO₂. Primary identification

of all bacterial species was by colonial morphology. Carriage for each pathobiont was then determined for age groups known to exhibit differences in prevalence: young children (aged <5 years), older children (>5 years to 17) and adults (18 to 65 years old).

Serotyping of *S. pneumoniae*: Pneumococcal isolates were serotyped by latex agglutination on a slide using a Neufeld *S. pneumoniae* antisera kit following manufacturer's guidelines (Statens Serum Institute, Copenhagen, Denmark). Isolates were cultured from frozen STGG vials onto Columbia Blood agar (CBA) and incubated overnight at 37°C in 5% CO₂. A single colony was then purity plated onto a second CBA plate and again incubated overnight. Growth from this plate was inoculated into 3ml of Todd-Hewitt Broth (THB) and incubated overnight at 37°C in 5% CO₂. Following o/n culture, cells were pelleted and resuspended in 100 µl THB. On a clean slide, 3 µl of the suspension was mixed with 3 µl of the pooled sera and examined for signs of agglutination. The chessboard for the identification of pneumococcal groups/types indicated if the isolate was a type (e.g. serotype 8) in which case no further serotyping was necessary, or a group (e.g. serogroup 15) in which case factor sera were used to further identify the serotype.

Antimicrobial Resistance Testing: All bacteria were phenotypically tested for antibiotic resistance using antibiotic discs and/or minimum inhibitory concentration (MIC) strips, in accordance with EUCAST. Firstly, 10µL (a suspension of cells in liquid STGG) of each isolate was plated onto CBA (Oxoid, UK) or CHOC agar (Oxoid, UK). *M. catarrhalis*, *S. pneumoniae*, *S. aureus* and *K. pneumoniae* isolates were plated on CBA, whilst *H. influenzae* and *N. meningitidis* isolates were plated onto CHOC agar. Plates were incubated for 24 hours at 37°C in 5% CO₂. Pure colonies were added to 1ml of saline to get an inoculum of 0.5 McFarland. For *M. catarrhalis*, *S. pneumoniae*, *H. influenzae* and *N. meningitidis*, a sterile

swab was used to spread this inoculum over Mueller-Hinton agar + 5% defibrinated horse blood and 20 mg⁻¹ β-NAD plates (MHF, Oxoid, UK). For *S. aureus* and *K. pneumoniae* a sterile swab was used to spread the inoculum over Mueller-Hinton agar plates (MH, Oxoid, UK). Antibiotic discs (Oxoid, UK) (four per plate) or MIC strips (E-tests; Oxoid, UK) (one per plate) were added and plates were incubated at 37°C in 5% CO₂ for 18 hours (±2 hours). *M. catarrhalis* were tested with amoxicillin-clavulanic acid (2-1µg), cefotaxime (5µg), ceftriaxone (30µg), erythromycin (15µg), tetracycline (30µg), chloramphenicol (30µg), ciprofloxacin (5µg) and meropenem (10µg) antibiotic discs. *S. pneumoniae* were tested with oxacillin (1µg), erythromycin (15µg), tetracycline (30µg) and chloramphenicol (30µg) antibiotic discs. *H. influenzae* were tested with benzylpenicillin (1µg), tetracycline (30µg), chloramphenicol (30µg) and ciprofloxacin (5µg) antibiotic discs as well as erythromycin (15µg) MIC strips. *S. aureus* were tested with benzylpenicillin (1µg), erythromycin (15µg), cefoxitin (30µg), tetracycline (30µg), chloramphenicol (30µg) and ciprofloxacin (5µg) antibiotic discs. *K. pneumoniae* were tested with amoxicillin - clavulanic acid (20-10µg), cefotaxime (5µg), ciprofloxacin (5µg), meropenem (10µg) and ceftazidime (10µg) antibiotic discs. *N. meningitidis* were tested with amoxicillin, benzylpenicillin (1µg), cefotaxime (5µg), ceftriaxone (30µg), chloramphenicol (30µg), ciprofloxacin (5µg) and meropenem (10µg) MIC strips.

16S rRNA Sequencing: The V4 region was amplified using primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') to generate an amplicon of 460 bp⁶⁷ at uBiome Inc. (San Francisco, USA). Samples were individually barcoded and sequenced on a NextSeq 500 (Illumina, San Diego, USA) to generate 2 × 150 bp paired-end reads. Reads were supplied having undergone primer and

low-quality end trimming. Maximum read lengths for forward and reverse were capped at 125bp and 124bp respectively.

Microbiome Analysis: Initial data handling was done using QIIME 2 2018.8 ⁶⁸. Only forward reads were processed owing to additional low-quality scores (<Q30) of reverse reads and that merging of read pairs resulted in significant loss of data. Raw sequence data were denoised with DADA2 ⁶⁹. Amplicon sequence variants (ASVs) were aligned with mafft ⁷⁰ (via q2-alignment) and a phylogeny constructed using fasttree2 ⁷¹. Taxonomic classification was done using the q2-feature-classifier ⁷² classify-sklearn naïve Bayes taxonomy classifier using the Greengenes 13_8 (99%) reference sequences ⁷³. The feature table, taxonomy, phylogenetic tree and sample metadata were then combined into a Phyloseq object using qiime2R ⁷⁴ via qza_to_phyloseq. All further analysis was done in R v3.6.0 ⁷⁵ in RStudio and figures were produced using the package ggplot2 ⁷⁶. Phyloseq v1.29.0 ⁷⁷ was used following a published workflow ⁷⁸. Potential contaminants were identified by prevalence in the blank swab controls and removed using the R package ‘decontam’ ⁷⁹. Ten ASVs were removed as a result. Only one was identified beyond “Bacteria” and belonged to the Genus *Meiothermus*. In addition, ASVs with Phylum classifications of ‘NA’ or ‘uncharacterised’ and any Phyla with a total of fewer than five ASVs whose provenance suggested low level environmental contamination were removed using the subset_taxa() function. Finally ASVs with Kingdom classifications other than ‘Bacteria’, Phylum classifications of ‘Cyanobacteria’, Family classifications of ‘mitochondria’ and Class classifications of ‘Chloroplast’ were also removed. Samples with <1000 ASVs were excluded and taxa present in less than 5% of samples were removed using the prune_taxa() phyloseq function. Alpha diversity was calculated using the estimate_richness(). The R function stat_compare_means() was used to compare age groups using the nonparametric Wilcoxon test. Beta diversity was determined

548 using Double Principle Co-ordinates Analysis and `plot_ordination()`. Hierarchical clustering
549 was done using Bray-Curtis Dissimilarity calculated using `vegdist()` from the R package `vegan`
550 ⁸⁰. Differentially abundant ASVs were identified using the `DESeq2` ⁸¹ package using an
551 adjusted p-value cut-off of 0.05 and a Log₂ fold change of 1.5. Networks were generated
552 using the R implementation of SPIEC-EASI (`SpiecEasi`) ⁸² using the `meinshausen-`
553 `buhlmann's` neighbourhood selection method with results visualised using `ggnet()`.

554

555 **Data Availability:**

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

559

560 **Code Availability:**

561 Code is available at:

562 https://github.com/davidwcleary/orang_asli_16S/blob/master/OrangAsli.Rmd

563

564 **Acknowledgments:**

565 The laboratory microbiology work was funded by a grant from the Higher Education Funding
566 Council for England (HEFCE) Newton Fund Official Development Assistance (ODA) fund.
567 The microbiome sequencing was funded through a uBiome Inc. Microbiome Grants Initiative
568 award. We would like to thank the Department of Orang Asli Affairs and Development
569 (JAKOA), Additionally, the authors would especially like to thank the communities of
570 Kampung Sungai Pergam and Kampung Berua for their support and participation in this
571 study.

572

Competing Interests:

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

Authors Contributions:

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

592 **References:**

- 593 1 Rogers, G. B., Ward, J., Brown, A. & Wesselingh, S. L. Inclusivity and equity in
594 human microbiome research. *The Lancet* **393**, 728-729,
595 doi:[https://doi.org/10.1016/S0140-6736\(18\)33204-5](https://doi.org/10.1016/S0140-6736(18)33204-5) (2019).
- 596 2 Schnorr, S. L. *et al.* Gut microbiome of the Hadza hunter-gatherers. *Nature*
597 *Communications* **5**, 3654, doi:10.1038/ncomms4654 (2014).
- 598 3 Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Science Advances*
599 **1** (2015).
- 600 4 Obregon-Tito, A. J. *et al.* Subsistence strategies in traditional societies distinguish gut
601 microbiomes. *Nature Communications* **6**, 6505, doi:10.1038/ncomms7505 (2015).
- 602 5 Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative
603 study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* **107**,
604 doi:10.1073/pnas.1005963107 (2010).
- 605 6 Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography.
606 *Nature* **486** (2012).
- 607 7 Sankaranarayanan, K. *et al.* Gut Microbiome Diversity among Cheyenne and Arapaho
608 Individuals from Western Oklahoma. *Current Biology* **25**, 3161-3169,
609 doi:<https://doi.org/10.1016/j.cub.2015.10.060> (2015).
- 610 8 Nasidze, I. *et al.* High Diversity of the Saliva Microbiome in Batwa Pygmies. *PloS*
611 *one* **6**, e23352, doi:10.1371/journal.pone.0023352 (2011).
- 612 9 Contreras, M. *et al.* The bacterial microbiota in the oral mucosa of rural Amerindians.
613 *Microbiology* **156**, 3282-3287, doi:<https://doi.org/10.1099/mic.0.043174-0> (2010).
- 614 10 Rivera-Olivero, I. *et al.* Pneumococcal carriage among indigenous Warao children in
615 Venezuela: serotypes, susceptibility patterns, and molecular epidemiology. *Clinical*
616 *infectious diseases : an official publication of the Infectious Diseases Society of*
617 *America* **45**, 1427-1434, doi:10.1086/522984 (2007).
- 618 11 Sun, W. *et al.* Association between early bacterial carriage and otitis media in
619 Aboriginal and non-Aboriginal children in a semi-arid area of Western Australia: a
620 cohort study. *BMC Infectious Diseases* **12**, 366, doi:10.1186/1471-2334-12-366
621 (2012).
- 622 12 Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age
623 groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease
624 Study 2010. *The Lancet* **380**, 2095-2128, doi:[https://doi.org/10.1016/S0140-](https://doi.org/10.1016/S0140-6736(12)61728-0)
625 [6736\(12\)61728-0](https://doi.org/10.1016/S0140-6736(12)61728-0) (2012).
- 626 13 Man, W. H., de Steenhuijsen Pijters, W. A. A. & Bogaert, D. The microbiota of the
627 respiratory tract: gatekeeper to respiratory health. *Nat Rev Micro* **15**, 259-270,
628 doi:10.1038/nrmicro.2017.14 (2017).
- 629 14 Kwambana-Adams, B. *et al.* Rapid replacement by non-vaccine pneumococcal
630 serotypes may mitigate the impact of the pneumococcal conjugate vaccine on
631 nasopharyngeal bacterial ecology. *Scientific Reports* **7**, 8127, doi:10.1038/s41598-
632 017-08717-0 (2017).
- 633 15 SyedHussain, T., Krishnasamy, DS., Hassan, AAG. Distribution and demography of
634 the Orang Asli in Malaysia. *International Journal of Humanities and Social Science*
635 *Invention* **6**, 40-45 (2017).
- 636 16 Masron T., M. F., Ismail N. Orang Asli in Peninsular Malaysia: Population, Spatial
637 Distribution and Socio-Economic Condition. *J Ritsumeikan Soc Sci Humanit* **6**, 75-
638 115 (2013).

- 639 17 Mohd Asri, M. N. Advancing the Orang Asli through Malaysia Clusters of Excellence
640 Policy. *Journal of International and Comparative Education* **1**,
641 doi:10.14425/00.45.76 (2012).
- 642 18 Al-Mekhlafi, H. M. *et al.* Prevalence and risk factors of *Strongyloides stercoralis*
643 infection among Orang Asli schoolchildren: new insights into the epidemiology,
644 transmission and diagnosis of strongyloidiasis in Malaysia. *Parasitology* **146**, 1602-
645 1614, doi:10.1017/S0031182019000945 (2019).
- 646 19 Al-Mekhlafi, H. M. *et al.* Protein-energy malnutrition and soil-transmitted
647 helminthiasis among Orang Asli children in Selangor, Malaysia. *Asia Pacific journal*
648 *of clinical nutrition* **14**, 188-194 (2005).
- 649 20 Yeo, L.-F., Aghakhanian, F. F., Tan, J. S. Y., Gan, H. M. & Phipps, M. E. Health and
650 saliva microbiomes of a semi-urbanized indigenous tribe in Peninsular Malaysia.
651 *FL1000Research* **8**, 175-175, doi:10.12688/fl1000research.17706.3 (2019).
- 652 21 Bogaert, D., de Groot, R. & Hermans, P. W. M. *Streptococcus pneumoniae*
653 colonisation: the key to pneumococcal disease. *The Lancet Infectious Diseases* **4**, 144-
654 154, doi:[http://dx.doi.org/10.1016/S1473-3099\(04\)00938-7](http://dx.doi.org/10.1016/S1473-3099(04)00938-7) (2004).
- 655 22 Wertheim, H. F. L. *et al.* The role of nasal carriage in *Staphylococcus aureus*
656 infections. *The Lancet Infectious Diseases* **5**, 751-762,
657 doi:[http://dx.doi.org/10.1016/S1473-3099\(05\)70295-4](http://dx.doi.org/10.1016/S1473-3099(05)70295-4) (2005).
- 658 23 Slack, M. P. E. A review of the role of *Haemophilus influenzae* in community-
659 acquired pneumonia. *Pneumonia* **6**, doi:10.15172/pneu.2015.6/520 (2015).
- 660 24 Zaneveld, J. R., McMinds, R. & Vega Thurber, R. Stress and stability: applying the
661 Anna Karenina principle to animal microbiomes. *Nature Microbiology* **2**, 17121,
662 doi:10.1038/nmicrobiol.2017.121 (2017).
- 663 25 *Malaysia DoS* (ed Department of Statistics Malaysia) (2018).
- 664 26 Hill, P. C. *et al.* Nasopharyngeal Carriage of *Streptococcus pneumoniae* in Gambian
665 Infants: A Longitudinal Study. *Clinical Infectious Diseases* **46**, 807-814,
666 doi:10.1086/528688 (2008).
- 667 27 Regev-Yochay, G. *et al.* Nasopharyngeal Carriage of *Streptococcus pneumoniae* by
668 Adults and Children in Community and Family Settings. *Clinical Infectious Diseases*
669 **38**, 632-639, doi:10.1086/381547 (2004).
- 670 28 Bogaert, D. *et al.* Colonisation by *Streptococcus pneumoniae* and *Staphylococcus*
671 *aureus* in healthy children. *Lancet* **363**, doi:10.1016/s0140-6736(04)16357-5 (2004).
- 672 29 Southern, J. *et al.* Pneumococcal carriage in children and their household contacts six
673 years after introduction of the 13-valent pneumococcal conjugate vaccine in England.
674 *PloS one* **13**, e0195799, doi:10.1371/journal.pone.0195799 (2018).
- 675 30 Mackenzie, G. A., Leach, A. J., Carapetis, J. R., Fisher, J. & Morris, P. S.
676 Epidemiology of nasopharyngeal carriage of respiratory bacterial pathogens in
677 children and adults: cross-sectional surveys in a population with high rates of
678 pneumococcal disease. *BMC Infectious Diseases* **10**, 304, doi:10.1186/1471-2334-10-
679 304 (2010).
- 680 31 Watt, J. P. *et al.* Nasopharyngeal versus Oropharyngeal Sampling for Detection of
681 Pneumococcal Carriage in Adults. *Journal of Clinical Microbiology* **42**, 4974-4976
682 (2004).
- 683 32 Hill, P. C. *et al.* Nasopharyngeal Carriage of *Streptococcus pneumoniae* in Gambian
684 Villagers. *Clinical Infectious Diseases* **43**, 673-679, doi:10.1086/506941 (2006).
- 685 33 Jauneikaite, E., Jefferies, J. M., Hibberd, M. L. & Clarke, S. C. Prevalence of
686 *Streptococcus pneumoniae* serotypes causing invasive and non-invasive disease in
687 South East Asia: A review. *Vaccine* **30**, 3503-3514,
688 doi:<https://doi.org/10.1016/j.vaccine.2012.03.066> (2012).

- 34 Arushothy, R. *et al.* Pneumococcal serotype distribution and antibiotic susceptibility in Malaysia: A four-year study (2014–2017) on invasive paediatric isolates. *International Journal of Infectious Diseases* **80**, 129-133, doi:<https://doi.org/10.1016/j.ijid.2018.12.009> (2019).
- 35 Coughtrie, A. L. *et al.* Ecology and diversity in upper respiratory tract microbial population structures from a cross-sectional community swabbing study. *Journal of Medical Microbiology* **67**, 1096-1108, doi:<https://doi.org/10.1099/jmm.0.000773> (2018).
- 36 Abdullahi, O., Nyiro, J., Lewa, P., Slack, M. & Scott, J. A. G. The descriptive epidemiology of *Streptococcus pneumoniae* and *Haemophilus influenzae* nasopharyngeal carriage in children and adults in Kilifi district, Kenya. *The Pediatric infectious disease journal* **27**, 59-64, doi:10.1097/INF.0b013e31814da70c (2008).
- 37 Mahadevan, M. *et al.* A review of the burden of disease due to otitis media in the Asia-Pacific. *International Journal of Pediatric Otorhinolaryngology* **76**, 623-635, doi:<https://doi.org/10.1016/j.ijporl.2012.02.031> (2012).
- 38 WHO. WHO UNICEF estimates time series for Malaysia (MYS), <https://apps.who.int/immunization_monitoring/globalsummary/estimates?c=MYS> (2020).
- 39 Kovács, E. *et al.* Co-carriage of *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Moraxella catarrhalis* among three different age categories of children in Hungary. *PloS one* **15**, e0229021, doi:10.1371/journal.pone.0229021 (2020).
- 40 Verhagen, L. M. *et al.* Nasopharyngeal carriage of respiratory pathogens in Warao Amerindians: significant relationship with stunting. *Tropical Medicine & International Health* **22**, 407-414, doi:10.1111/tmi.12835 (2017).
- 41 Sollid, J. U. E., Furberg, A. S., Hanssen, A. M. & Johannessen, M. *Staphylococcus aureus*: Determinants of human carriage. *Infection, Genetics and Evolution* **21**, 531-541, doi:<http://dx.doi.org/10.1016/j.meegid.2013.03.020> (2014).
- 42 Choi, C. S. *et al.* Nasal carriage of *Staphylococcus aureus* among healthy adults. *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi* **39**, 458-464 (2006).
- 43 Ruimy, R. *et al.* Are host genetics the predominant determinant of persistent nasal *Staphylococcus aureus* carriage in humans? *The Journal of Infectious Diseases* **202**, 924-934, doi:10.1086/655901 (2010).
- 44 Humphreys, H., Fitzpatrick, F. & Harvey, B. J. Gender Differences in Rates of Carriage and Bloodstream Infection Caused by Methicillin-Resistant *Staphylococcus aureus*: Are They Real, Do They Matter and Why? *Clinical Infectious Diseases* **61**, 1708-1714, doi:10.1093/cid/civ576 (2015).
- 45 Hashim, R. National Antibiotic Resistance Surveillance Report 2018. (Antibiotic Resistance Surveillance Reference Laboratory, Institute for Medical Research 2018).
- 46 Tong, S. Y. C., Varrone, L., Chatfield, M. D., Beaman, M. & Giffard, P. M. Progressive increase in community-associated methicillin-resistant *Staphylococcus aureus* in Indigenous populations in northern Australia from 1993 to 2012. *Epidemiology and Infection* **143**, 1519-1523, doi:10.1017/S0950268814002611 (2015).
- 47 Liu, C. M. *et al.* *Staphylococcus aureus* and the ecology of the nasal microbiome. *Science Advances* **1**, doi:10.1126/sciadv.1400216 (2015).
- 48 Cremers, A. J. *et al.* The adult nasopharyngeal microbiome as a determinant of pneumococcal acquisition. *Microbiome* **2**, 1-10, doi:10.1186/2049-2618-2-44 (2014).

738 49 Yan, M. *et al.* Nasal microenvironments and interspecific interactions influence nasal
739 microbiota complexity and *S. aureus* carriage. *Cell host & microbe* **14**, 631-640,
740 doi:10.1016/j.chom.2013.11.005 (2013).

741 50 Bomar, L., Brugger, S. D., Yost, B. H., Davies, S. S. & Lemon, K. P.
742 *Corynebacterium accolens* Releases Antipneumococcal Free Fatty Acids from Human
743 Nostril and Skin Surface Triacylglycerols. *mBio* **7**, doi:10.1128/mBio.01725-15
744 (2016).

745 51 De Boeck, I. *et al.* Comparing the Healthy Nose and Nasopharynx Microbiota
746 Reveals Continuity As Well As Niche-Specificity. *Frontiers in microbiology* **8**, 2372-
747 2372, doi:10.3389/fmicb.2017.02372 (2017).

748 52 Feazel, L. M. *et al.* Effects of Vaccination with 10-Valent Pneumococcal Non-
749 Typeable Haemophilus influenza Protein D Conjugate Vaccine (PHiD-CV) on the
750 Nasopharyngeal Microbiome of Kenyan Toddlers. *PloS one* **10**, e0128064-e0128064,
751 doi:10.1371/journal.pone.0128064 (2015).

752 53 Boelsen, L. K. *et al.* The association between pneumococcal vaccination, ethnicity,
753 and the nasopharyngeal microbiota of children in Fiji. *Microbiome* **7**, 106-106,
754 doi:10.1186/s40168-019-0716-4 (2019).

755 54 Biesbroek, G. *et al.* Early respiratory microbiota composition determines bacterial
756 succession patterns and respiratory health in children. *Am J Respir Crit Care Med*
757 **190**, doi:10.1164/rccm.201407-1240OC (2014).

758 55 Laufer, A. S. *et al.* Microbial Communities of the Upper Respiratory Tract and Otitis
759 Media in Children. *mBio* **2**, doi:10.1128/mBio.00245-10 (2011).

760 56 Teo Shu, M. *et al.* The infant nasopharyngeal microbiome impacts severity of lower
761 respiratory infection and risk of asthma development. *Cell Host Microbe* **17**,
762 doi:10.1016/j.chom.2015.03.008 (2015).

763 57 Durack, J., Boushey, H. A. & Lynch, S. V. Airway Microbiota and the Implications of
764 Dysbiosis in Asthma. *Current Allergy and Asthma Reports* **16**, 52,
765 doi:10.1007/s11882-016-0631-8 (2016).

766 58 Ege, M. J. *et al.* Exposure to environmental microorganisms and childhood asthma.
767 *New Engl J Med* **364**, doi:10.1056/NEJMoa1007302 (2011).

768 59 Verma, D., Garg, P. K. & Dubey, A. K. Insights into the human oral microbiome.
769 *Archives of Microbiology* **200**, 525-540, doi:10.1007/s00203-018-1505-3 (2018).

770 60 Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact
771 sequence-based microbiome analyses. *BMC Biol* **12**, doi:10.1186/s12915-014-0087-z
772 (2014).

773 61 Wang, J., Li, D., Wang, J. & Zhang, Z. Human oral microbiome characterization and
774 its association with environmental microbiome revealed by the Earth Microbiome
775 Project. *bioRxiv*, 732123, doi:10.1101/732123 (2019).

776 62 Wong, Y. S., Allotey, P. & Reidpath, D. D. Why we run when the doctor comes:
777 Orang Asli responses to health systems in transition in Malaysia. *Critical Public*
778 *Health* **29**, 192-204, doi:10.1080/09581596.2018.1438588 (2019).

779 63 Olwagin, C. P., Adrian, P. V. & Madhi, S. A. Comparison of traditional culture and
780 molecular qPCR for detection of simultaneous carriage of multiple pneumococcal
781 serotypes in African children. *Scientific Reports* **7**, 4628, doi:10.1038/s41598-017-
782 04915-y (2017).

783 64 Tennekkes, M. tmap: Thematic Maps in R. *Journal of Statistical Software* **84**,
784 doi:10.18637/jss.v084.i06 (2018).

785 65 Kahle, D. W., H. in *The R Journal* Vol. 5 144-161.

786 66 O'Brien, K. L. *et al.* Evaluation of a medium (STGG) for transport and optimal
787 recovery of *Streptococcus pneumoniae* from nasopharyngeal secretions collected

- during field studies. *Journal of clinical microbiology* **39**, 1021-1024, doi:10.1128/JCM.39.3.1021-1024.2001 (2001).
- 67 Almonacid, D. E. *et al.* 16S rRNA gene sequencing and healthy reference ranges for 28 clinically relevant microbial taxa from the human gut microbiome. *PloS one* **12**, e0176555, doi:10.1371/journal.pone.0176555 (2017).
- 68 Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* **37**, 852-857, doi:10.1038/s41587-019-0209-9 (2019).
- 69 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, doi:10.1038/nmeth.3869 (2016).
- 70 Katoh, K., Misawa, K., Kuma, K.-i. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059-3066, doi:10.1093/nar/gkf436 (2002).
- 71 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one* **5**, doi:10.1371/journal.pone.0009490 (2010).
- 72 Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90, doi:10.1186/s40168-018-0470-z (2018).
- 73 McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* **6**, 610-618, doi:10.1038/ismej.2011.139 (2012).
- 74 Bisanz, J. E. *qiime2R: Importing QIIME2 artifacts and associated data into R sessions*, <<https://github.com/jbisanz/qiime2R>> (2018).
- 75 Team, R. C. R: *a language and environment for statistical computing*. (R Foundation for Statistical Computing, 2014).
- 76 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
- 77 McMurdie, P. J. H., S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. . *PloS one* **8**, doi:10.1371/journal.pone.0061217 (2013).
- 78 Callahan, B., Sankaran, K., Fukuyama, J., McMurdie, P. & Holmes, S. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses [version 2; peer review: 3 approved]. *F1000Research* **5**, doi:10.12688/f1000research.8986.2 (2016).
- 79 Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226, doi:10.1186/s40168-018-0605-2 (2018).
- 80 Oksanen, J., Blanchet, F., Kindt, R., Legendre, P. & O'Hara, R. (2016).
- 81 Love, M. I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. . *Genome Biology* **15**, doi:10.1186/s13059-014-0550-8 (2014).
- 82 Kurtz, Z. D. *et al.* Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS Computational Biology* **11**, e1004226, doi:10.1371/journal.pcbi.1004226 (2015).

Figure Legends:

Figure 1: Location of Orang Asli sites with age and gender demographics. The location of Orang Asli settlements were in the north-east of Peninsula Malaysia (a.) to the south of the Terengganu state's capital city, Kuala Terengganu (b.). An even gender split was achieved (c.) but with disproportionate recruitment in older children >5 years of age and adults less than <65 (d.).

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

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

Figure 4: Alpha and Beta Diversity of Nasal (left column) and Oral (right column) Samples. Observed richness and Simpsons 1-D (a measure of diversity) are shown in rows a. and b. respectively. Only a significant difference (**) was observed for nasal samples for those aged 5-17 when compared against 50-65 ($p = 0.0023$). Beta diversity is shown using Double Principle Coordinates Analysis (DPCoA) (row c.). Here, clusters are observed for

nasal samples (left) with grouping based on age-group. No clusters were observed for oral samples (right).

Figure 5: Hierarchical Clustering of Nasal Samples with Bray-Curtis Dissimilarity

using Genus-level Relative Abundances. The dendrogram (top) shows clustering of samples with the below bar chart showing the relative abundance of the eight most commonly observed Genera. Six clear clusters were observed including a *Corynebacterium* dominant profile (n=28; 35.4%), a *Corynebacterium/Dolosigranulum* profile (n=16; 20.3%), a *Moraxella* profile (n=10; 12.7%) a *Moraxella/Haemophilus* profile (n=10; 12.7%), a *Delftia/Orchobactum* profile (n=4; 5.1%) and a final group of two samples characterised by high *Streptococcus*, in one case with *Haemophilus*. Culture results for *S. aureus* (dark blue) and *S. pneumoniae* (pink) are shown as coloured circles at the tips of the dendrogram. There is no clear distribution of individuals who were culture positive for either bacterium.

Figure 6: Co-occurrence network of the 100 most abundant ASVs. Node size represents the Log2 abundance of each ASV and are labelled by Genus. Edge colours denote the direction of association with positive (green) and negative (red). Edge size shows strength of co-occurrence with a weighting value >2.

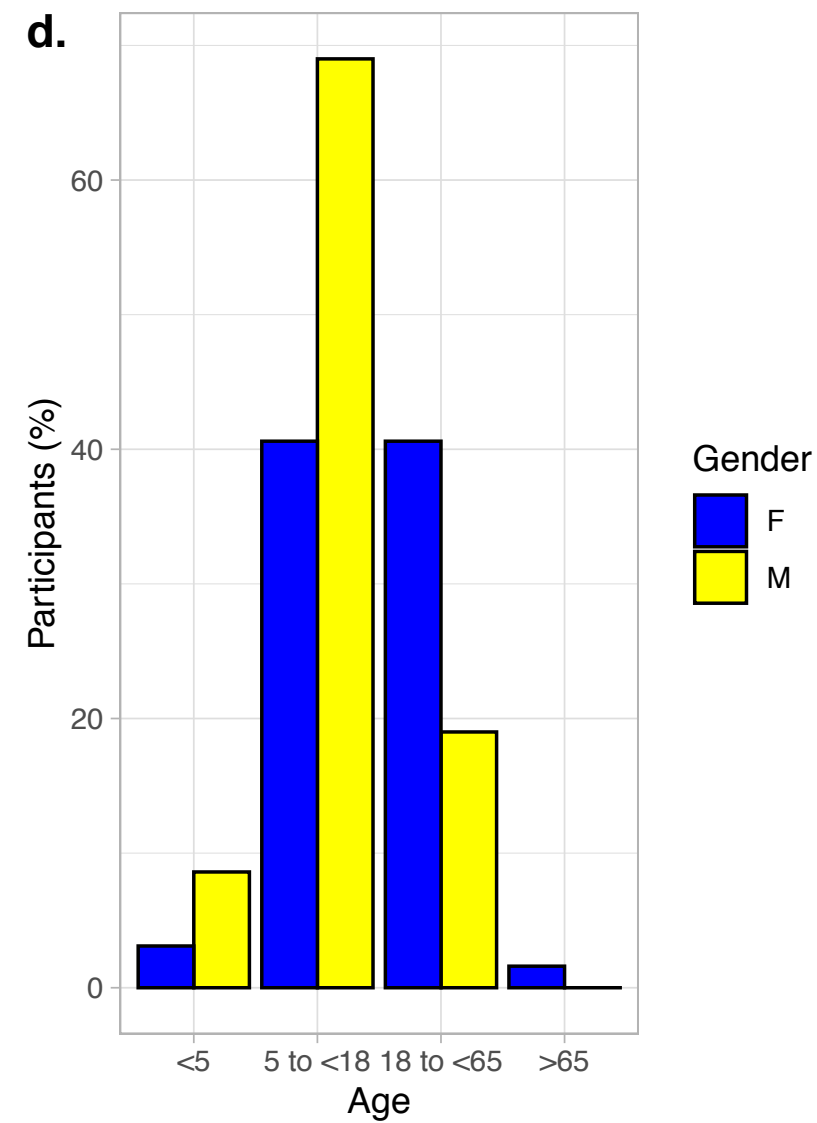
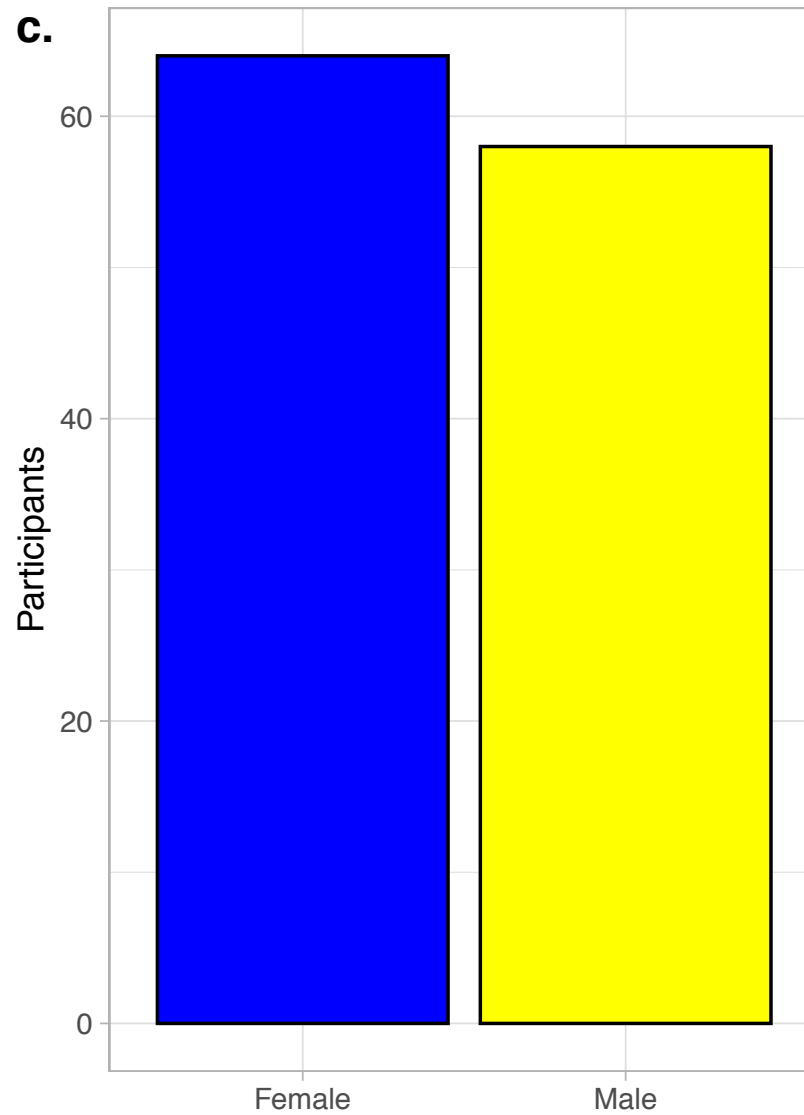
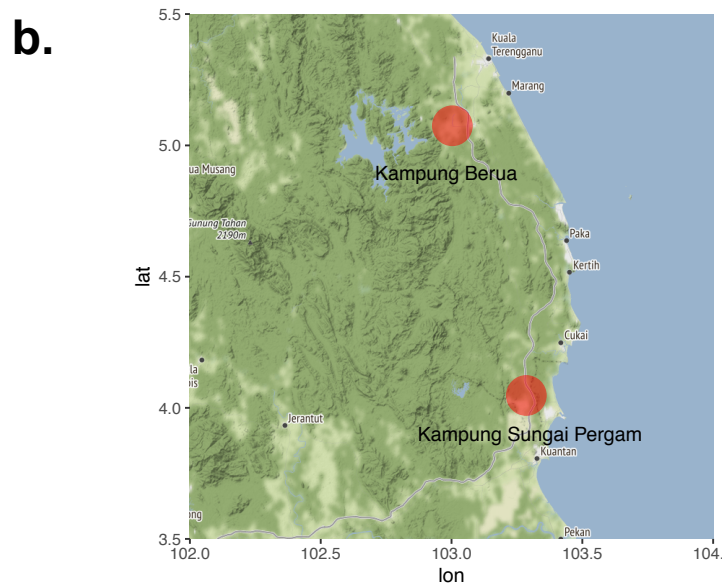
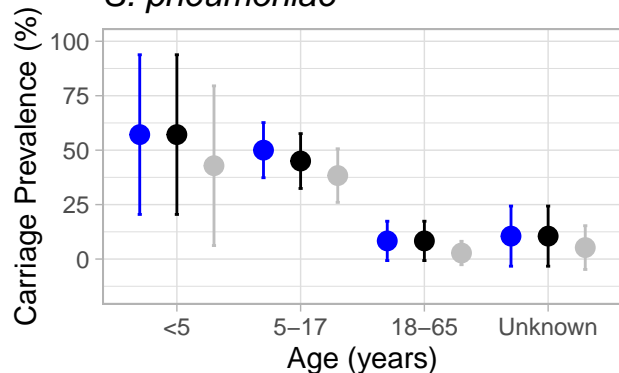
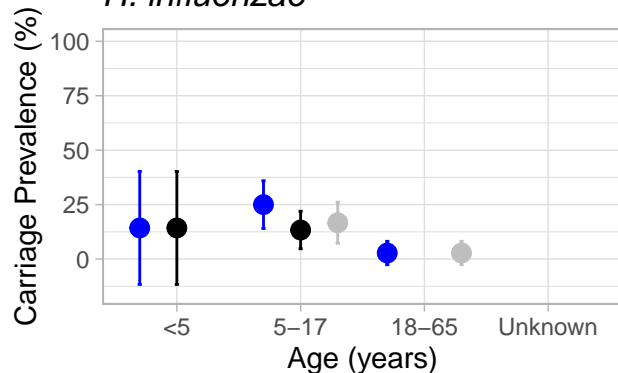
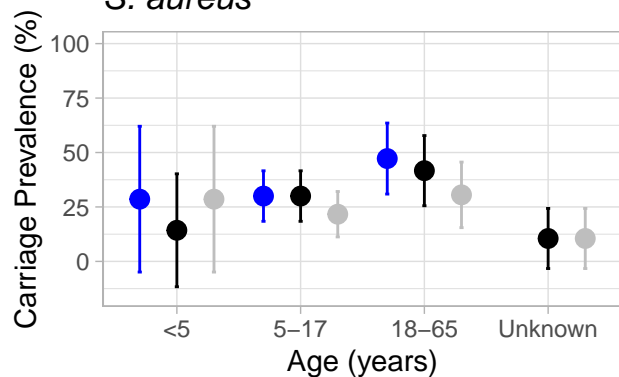
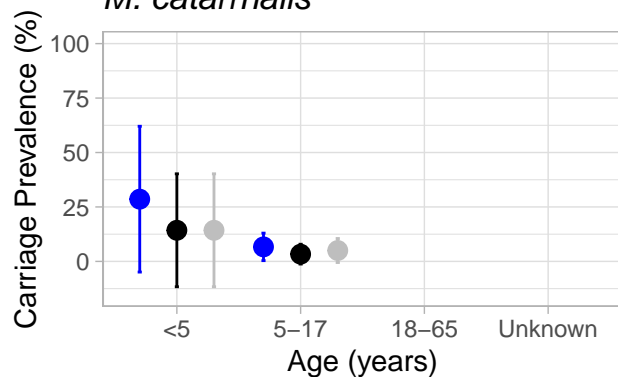
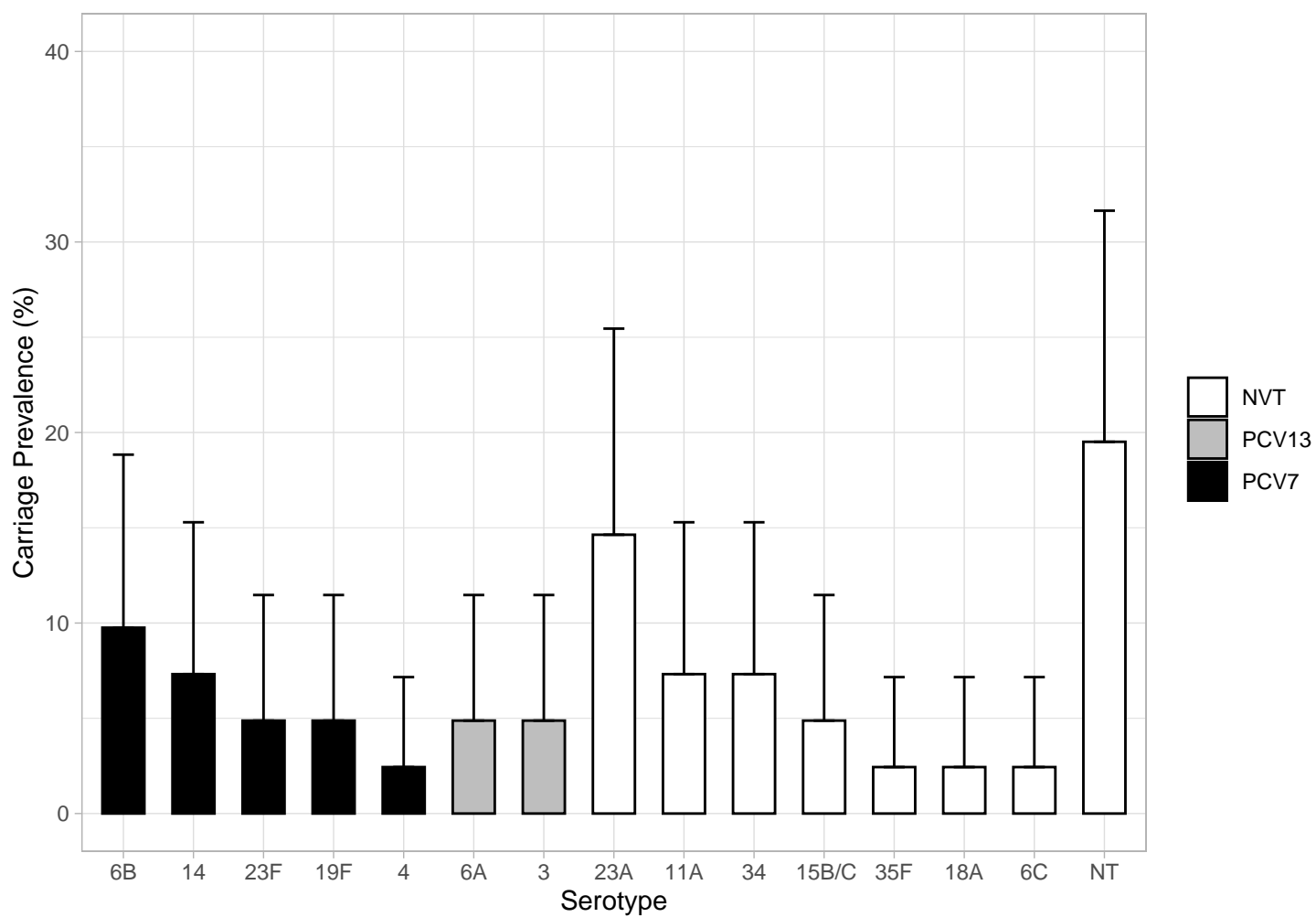


Table 1: Comparison of populations from Kampung Sungai Pergam (Site 1) and Kampung Berua (Site 2). † respiratory symptoms are self-reported incidences of cough, sore throat, flu-like symptoms, ear infection or fever; *p* values are z-test of proportions; ns = non-significant.

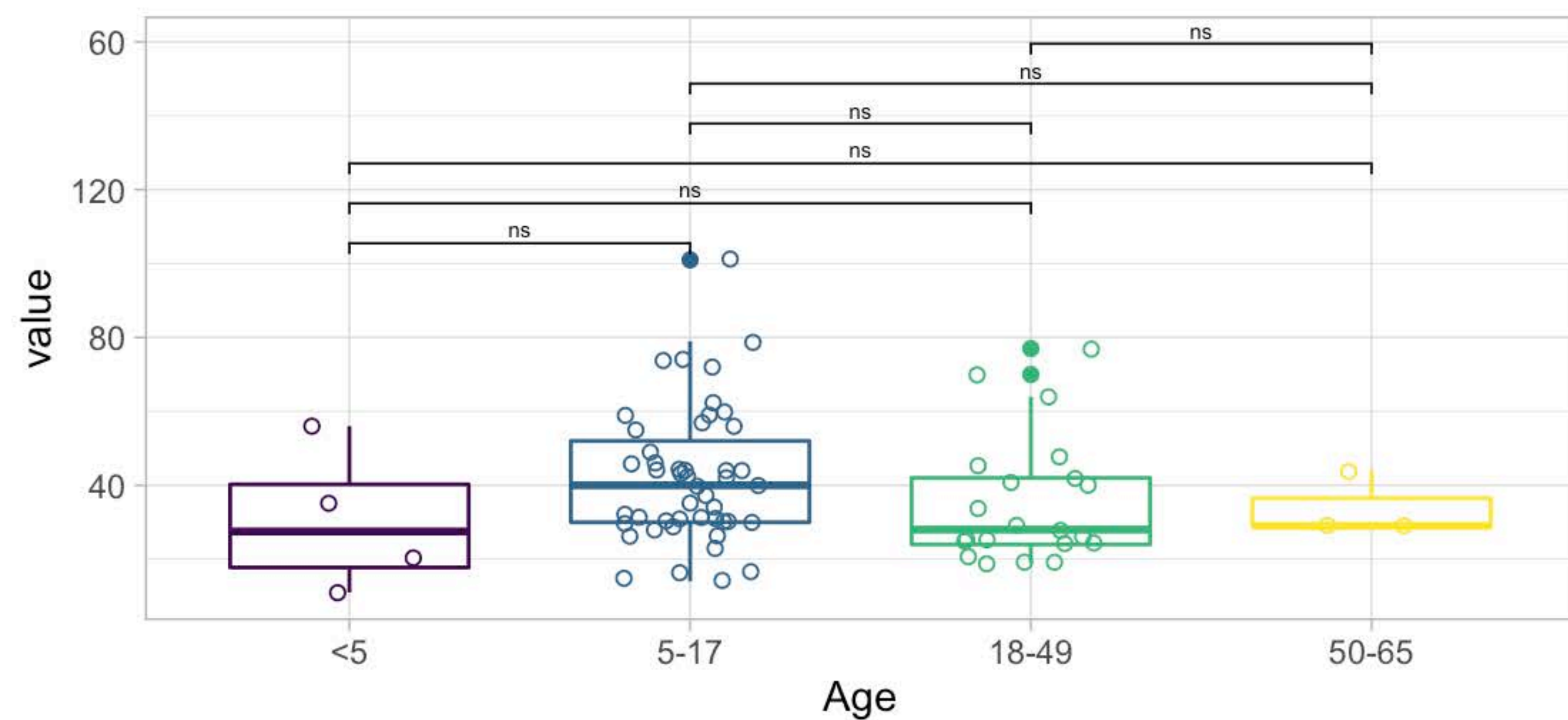
		Kampung Sungai Pergam (Site 1)	Kampung Berua (Site 2)	<i>p</i> value
Total (n)		68	62	
Gender	Female	34	30	ns
	Male	32	26	ns
	NA	2	6	ns
Household Size	(Average, n)	4.5	5.2	NA
Age (years)	<5	1	6	ns
	5 to <18	27	39	ns
	18 to <65	25	12	<0.05
	65+	1	0	ns
	NA	14	5	<0.05
Vaccinations	Yes	61	50	ns
	No	5	4	ns
	NA	2	8	ns
Smoking	Yes	14	7	ns
	No	53	51	ns
	NA	1	4	ns
Respiratory Symptoms†	Within last 30 days	40	16	<0.001
	Within last 90 days	36	6	<0.001

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

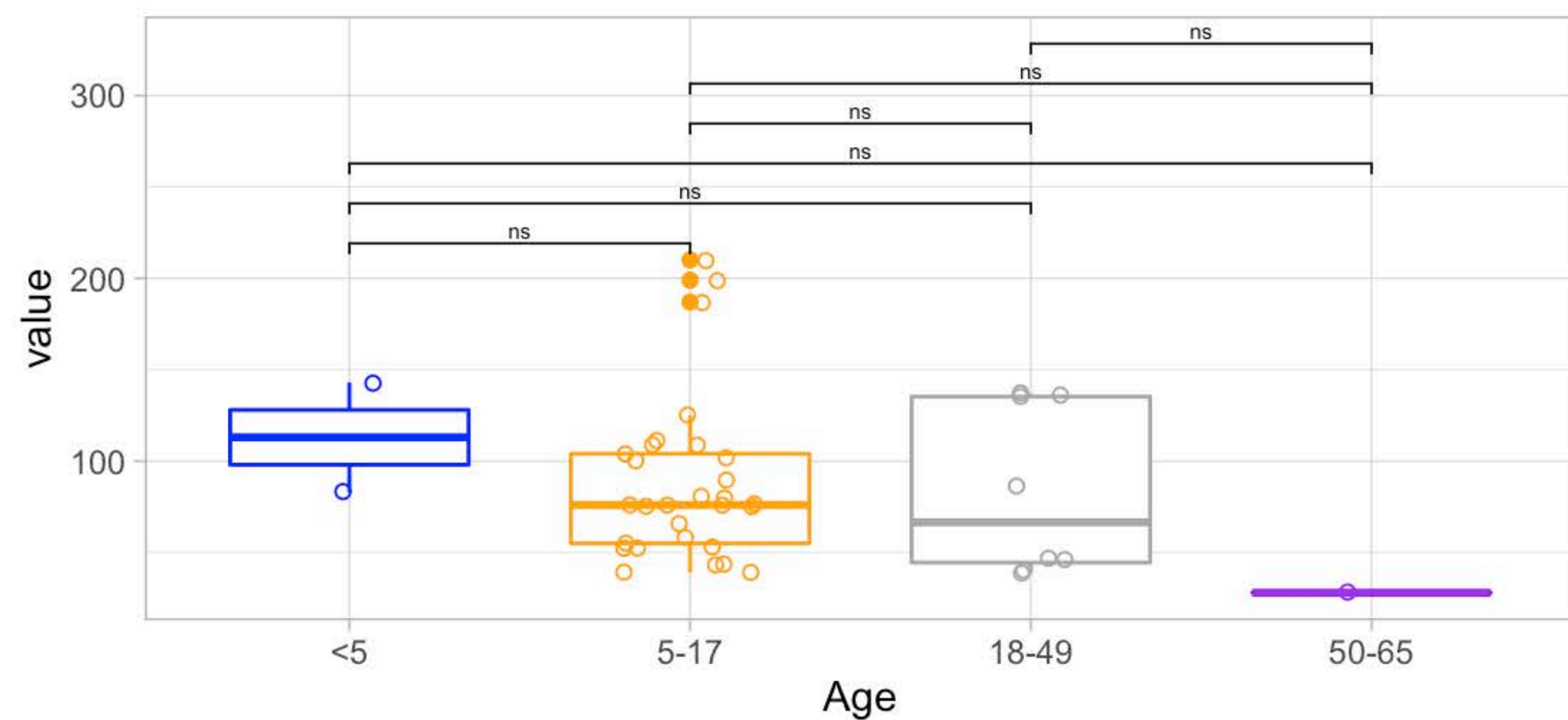
● Combined ● Nose ● NP



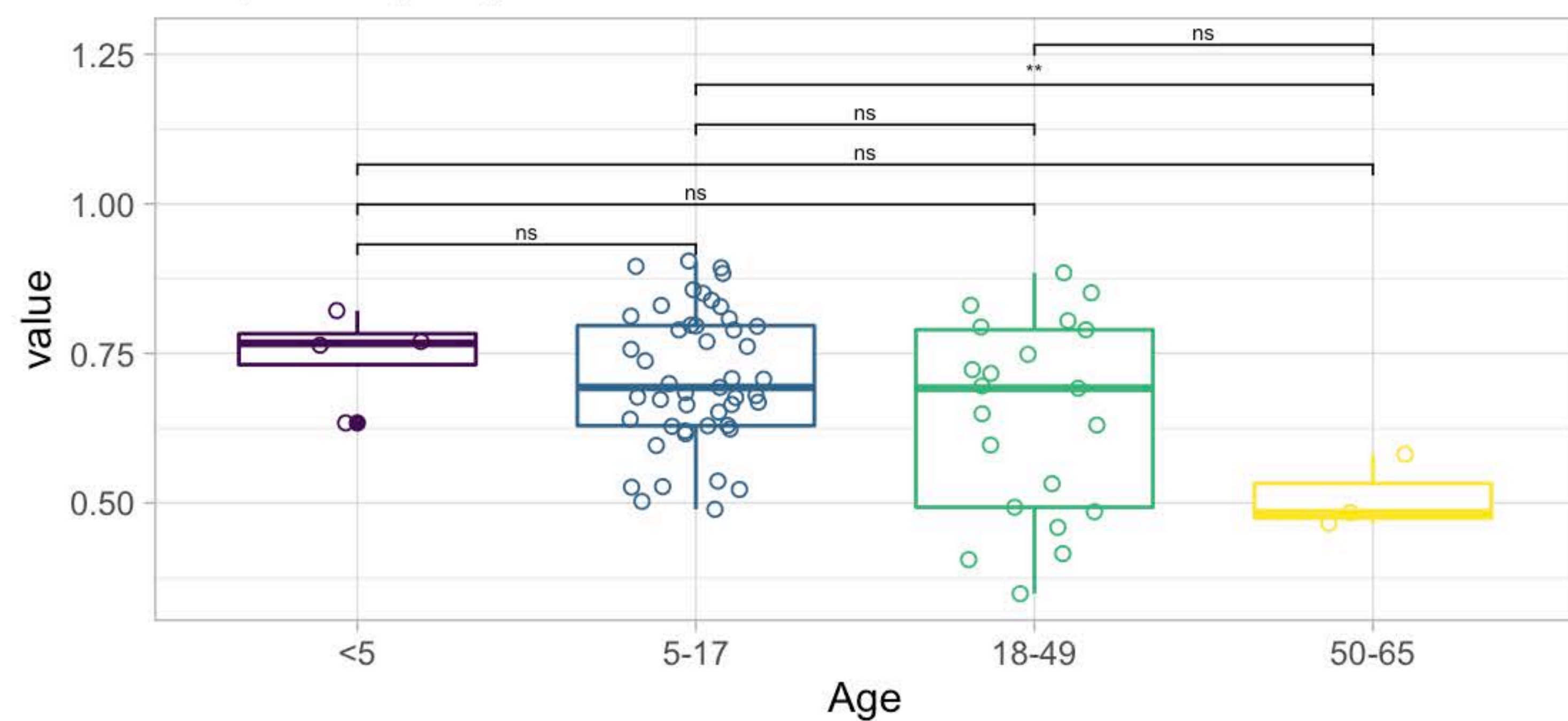
a. Observed



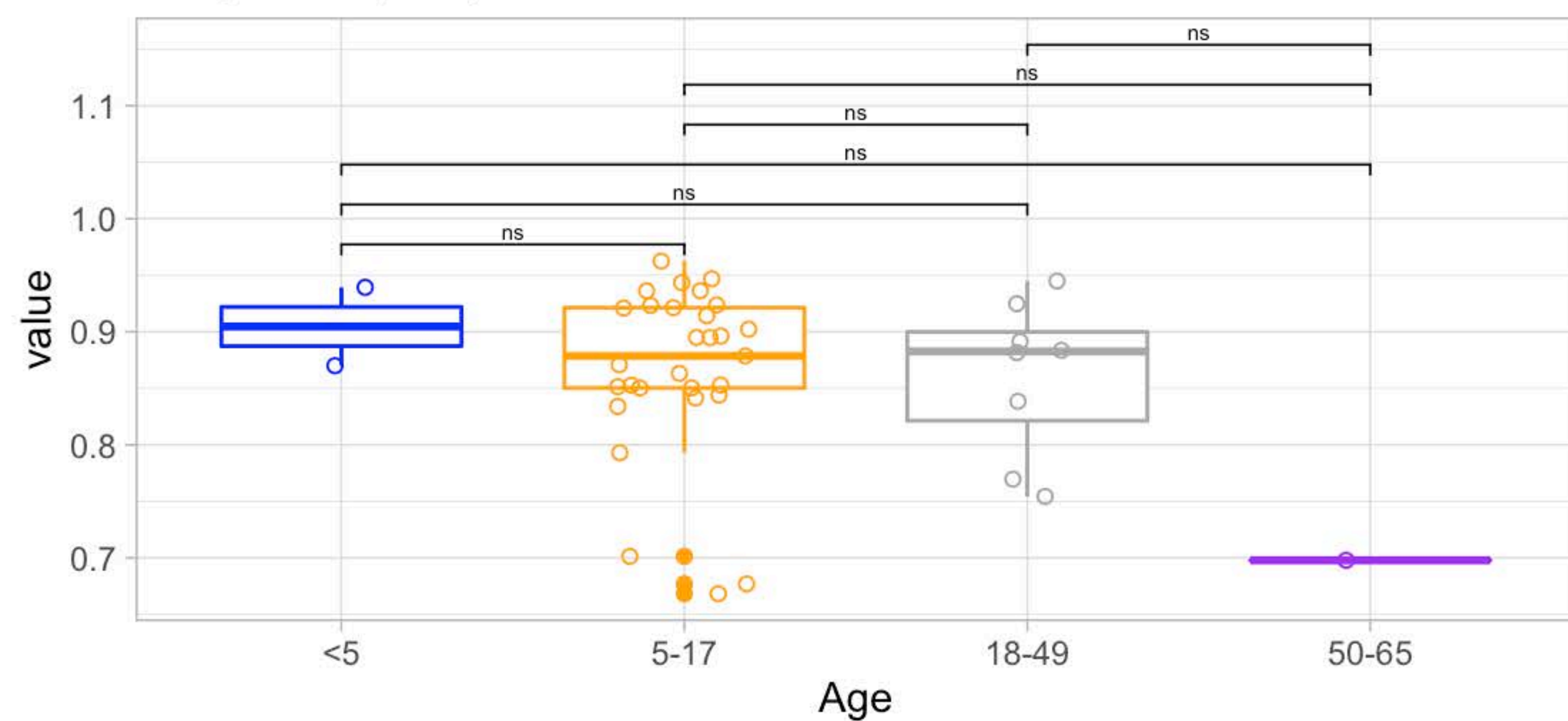
Observed



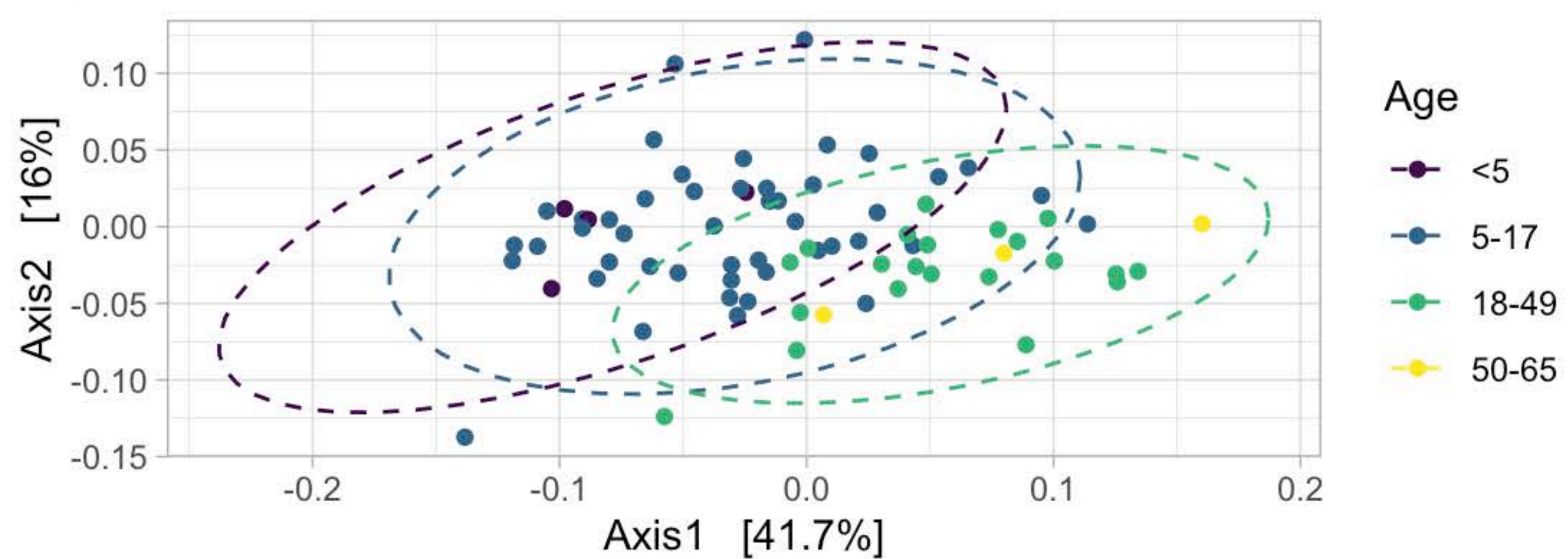
b. Simpsons (1-D)



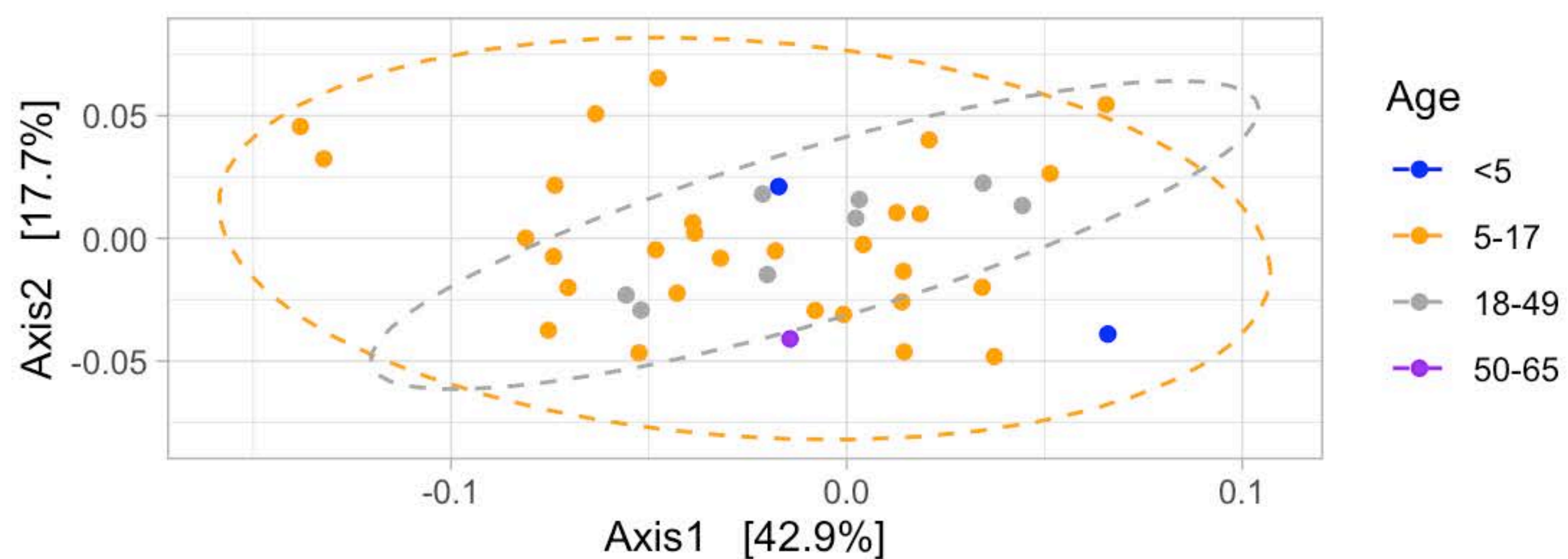
Simpsons (1-D)

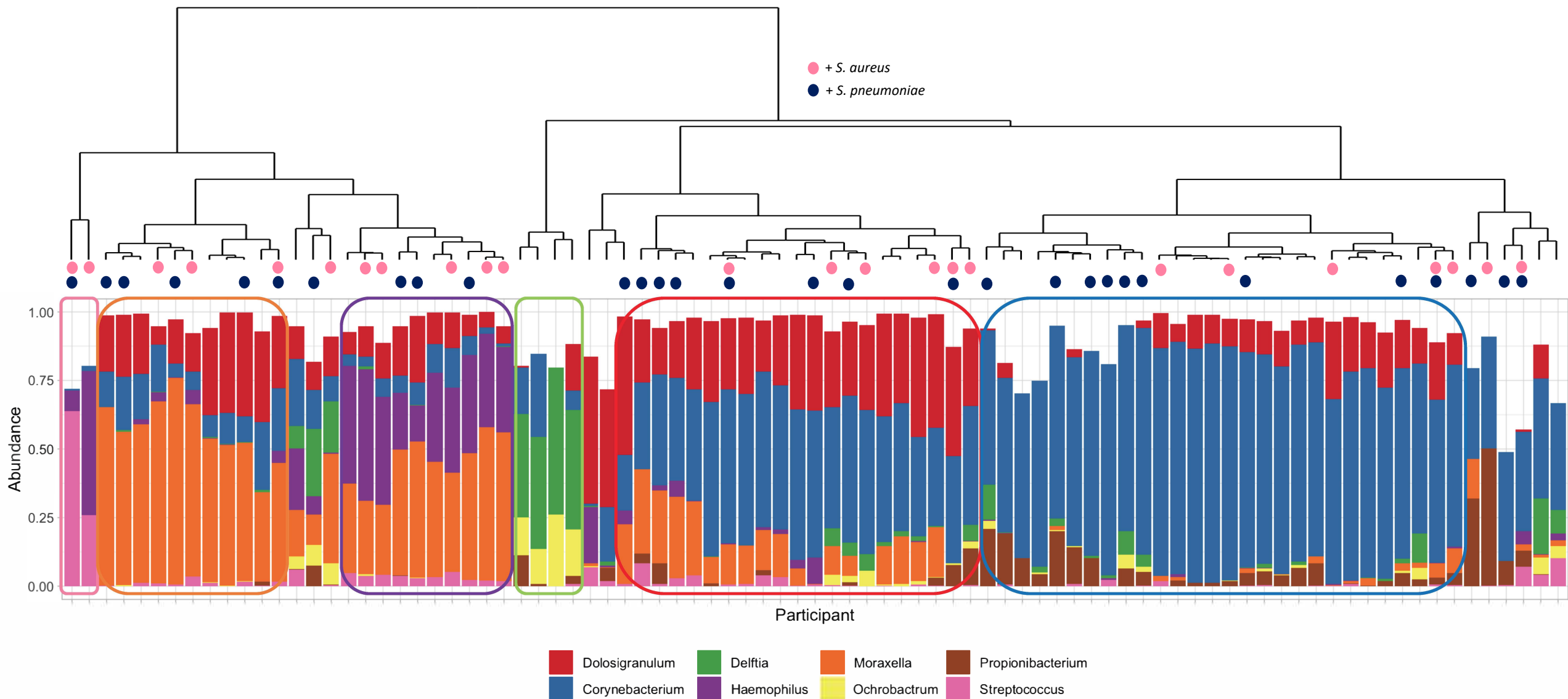


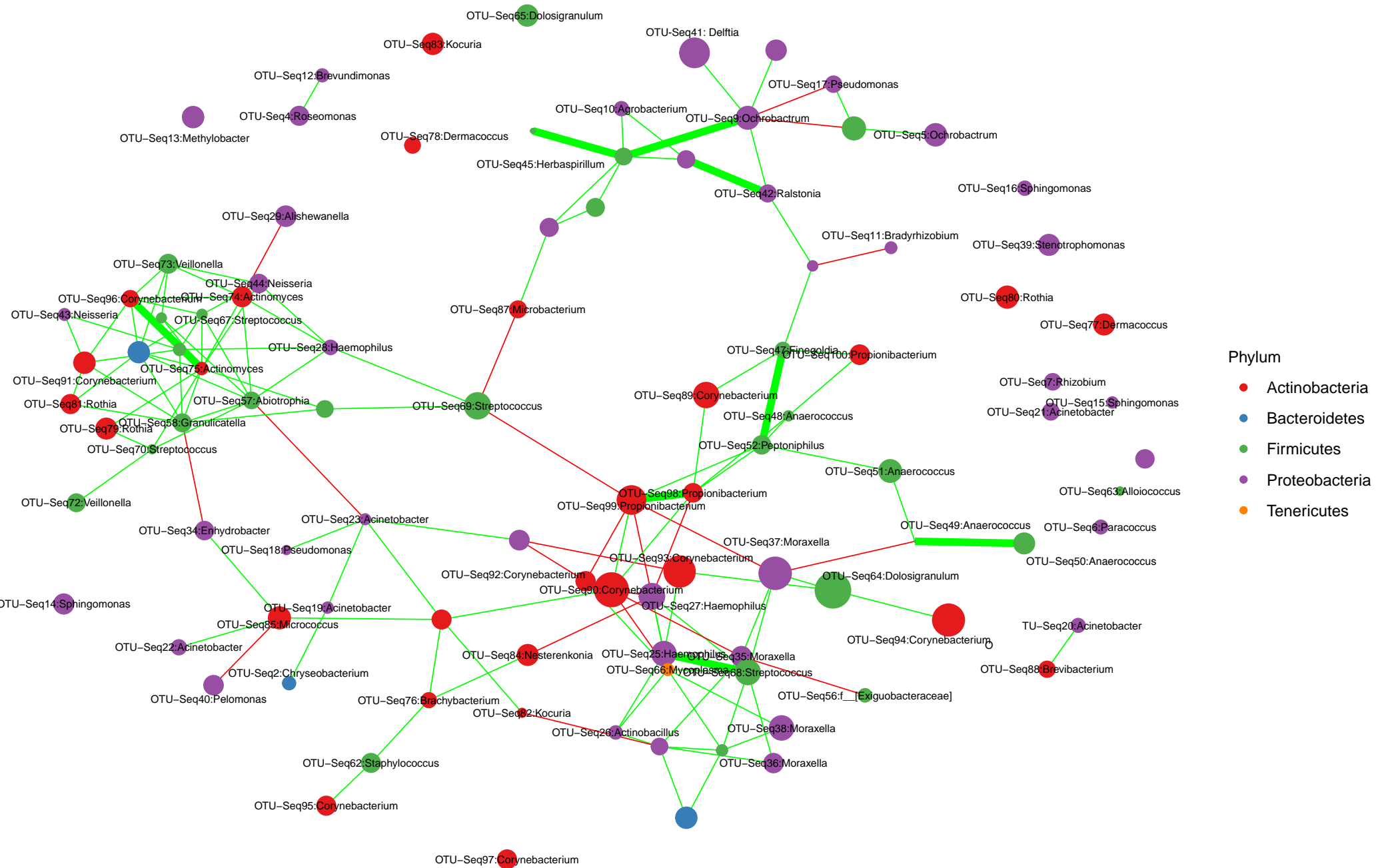
c. Double PCoA

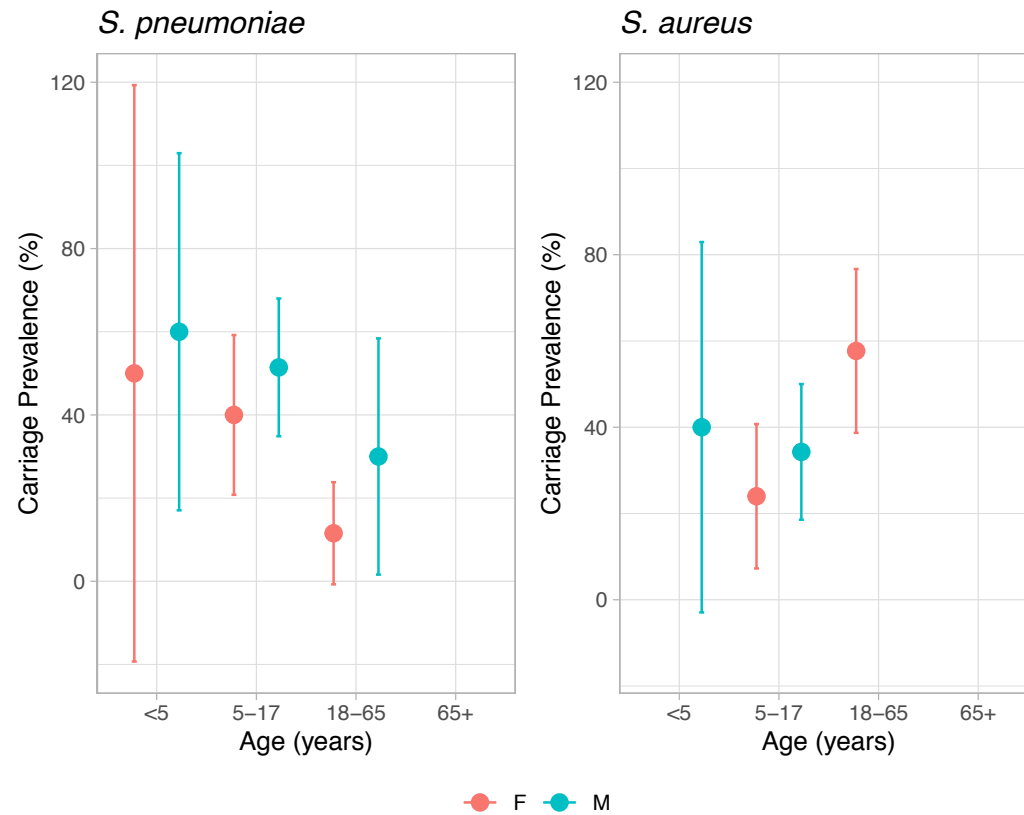


Double PCoA





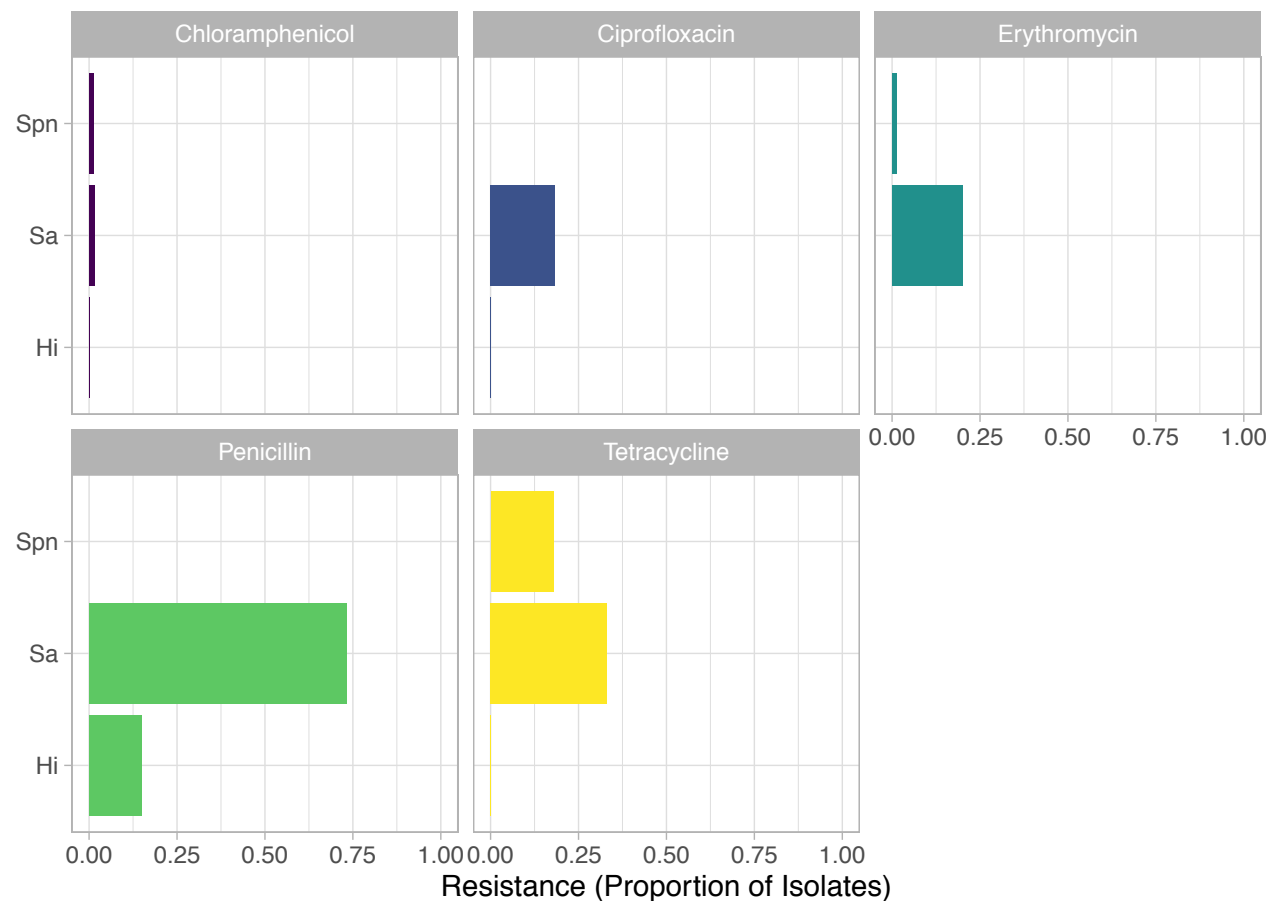




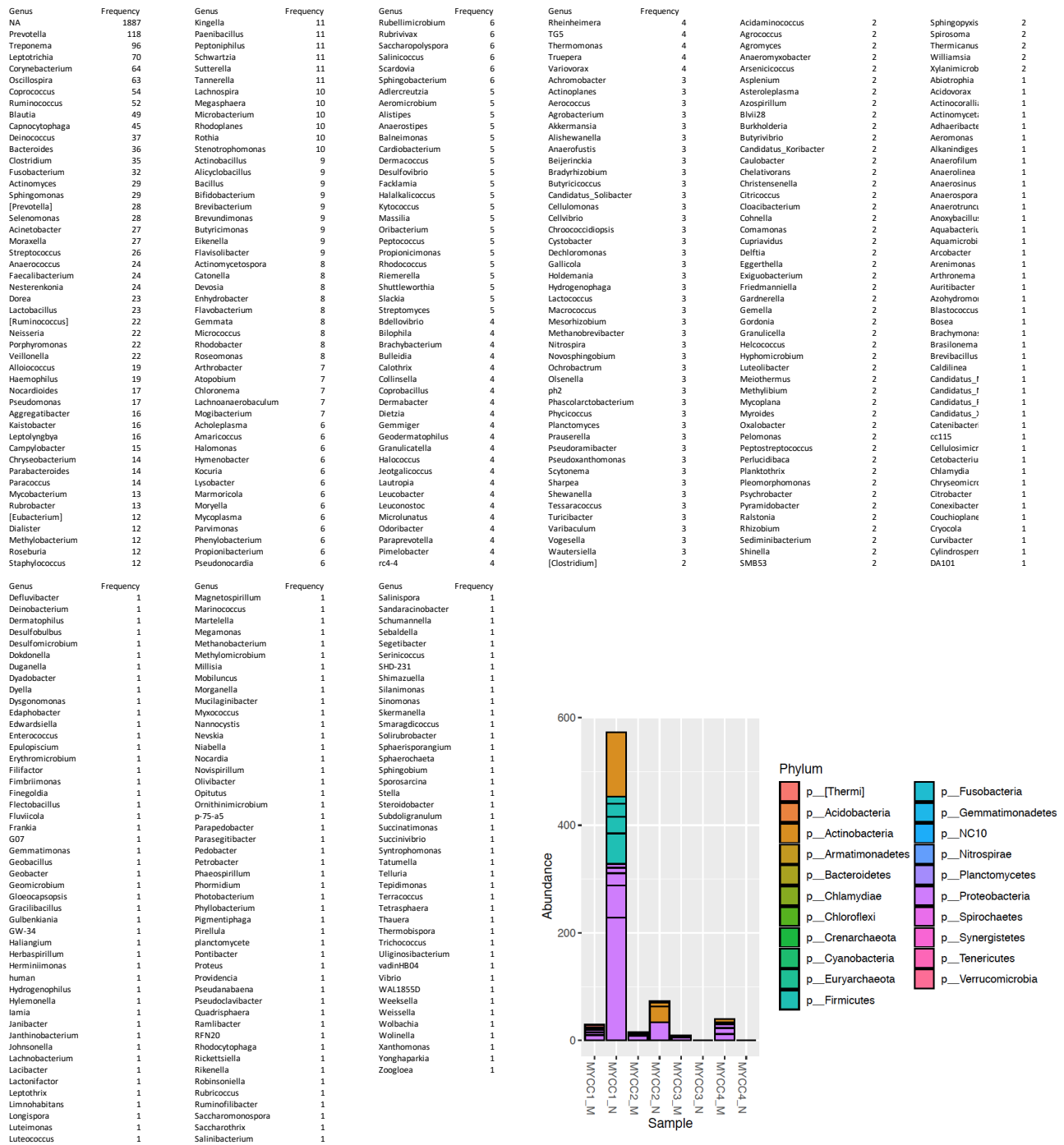
Supplementary Figure 1: Carriage Prevalence of *S. pneumoniae* and *S. aureus* in Female and Male Study Participants. Point estimates for percentage carriage prevalence of *S. pneumoniae* and *S. aureus* are shown across three age groups <5, 5 to 17, 18 to 65 and 65+. Missing data indicates either carriage was not absent, not determined, or gender was not recorded. Carriage was estimated overall i.e. a positive swab culture from either nasal or nasopharyngeal swab. Error bars show 95% CI intervals.

Supplementary 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.

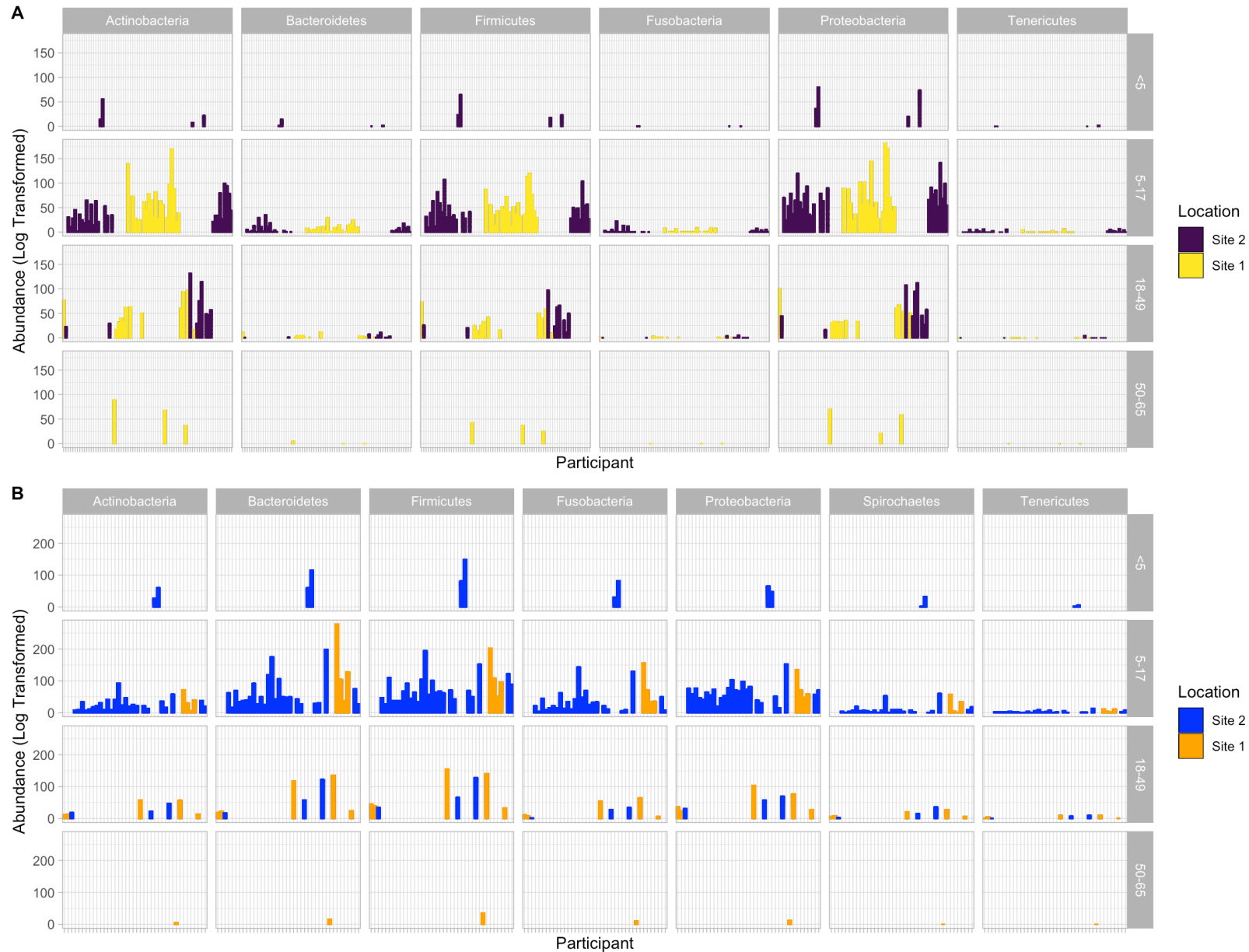
Bacterial Species	N	%
Sa	25	19.23
Spn	16	12.31
Spn-Hi	7	5.38
Spn-Sa	5	3.85
Hi	2	1.54
Mc	1	0.77
Spn-Mc	1	0.77
Hi-Sa	1	0.77
Hi-Mc	1	0.77
Spn-Hi-Sa	1	0.77
Spn-Hi-Mc	1	0.77
Sa-Mc	0	0.00
Hi-Sa-Mc	0	0.00
Sa-Mc-Spn	0	0.00
Sa-Mc-Spn-Hi	0	0.00



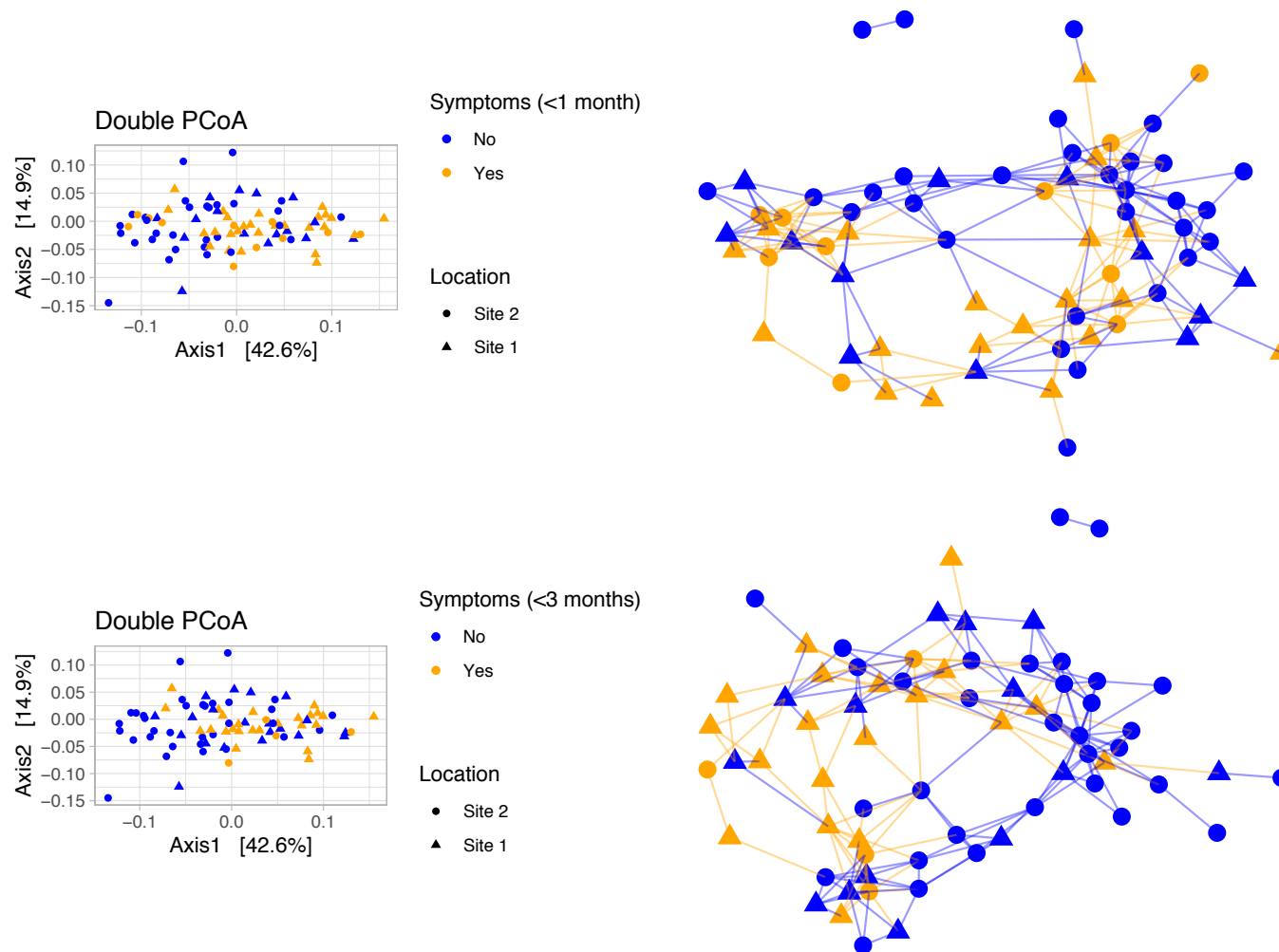
Supplementary Figure 2: Antibiotic resistance of isolates of *S. pneumoniae* (Spn; n=67), *S. aureus* (Sa; n=60) and *H. influenzae* (Hi; n=10). Columns represent the proportion of isolates that were resistant to tetracycline, chloramphenicol, erythromycin, ciprofloxacin and penicillin. Forty-four (73.3%) *S. aureus* isolates were resistant to penicillin, 20 (33.0%) to tetracycline, 12 (20%) to erythromycin and 11 (18.3%) to ciprofloxacin. Of the *H. influenzae* isolates three (15%) were resistant to penicillin. Twelve *S. pneumoniae* (17.9%) were resistant to tetracycline.



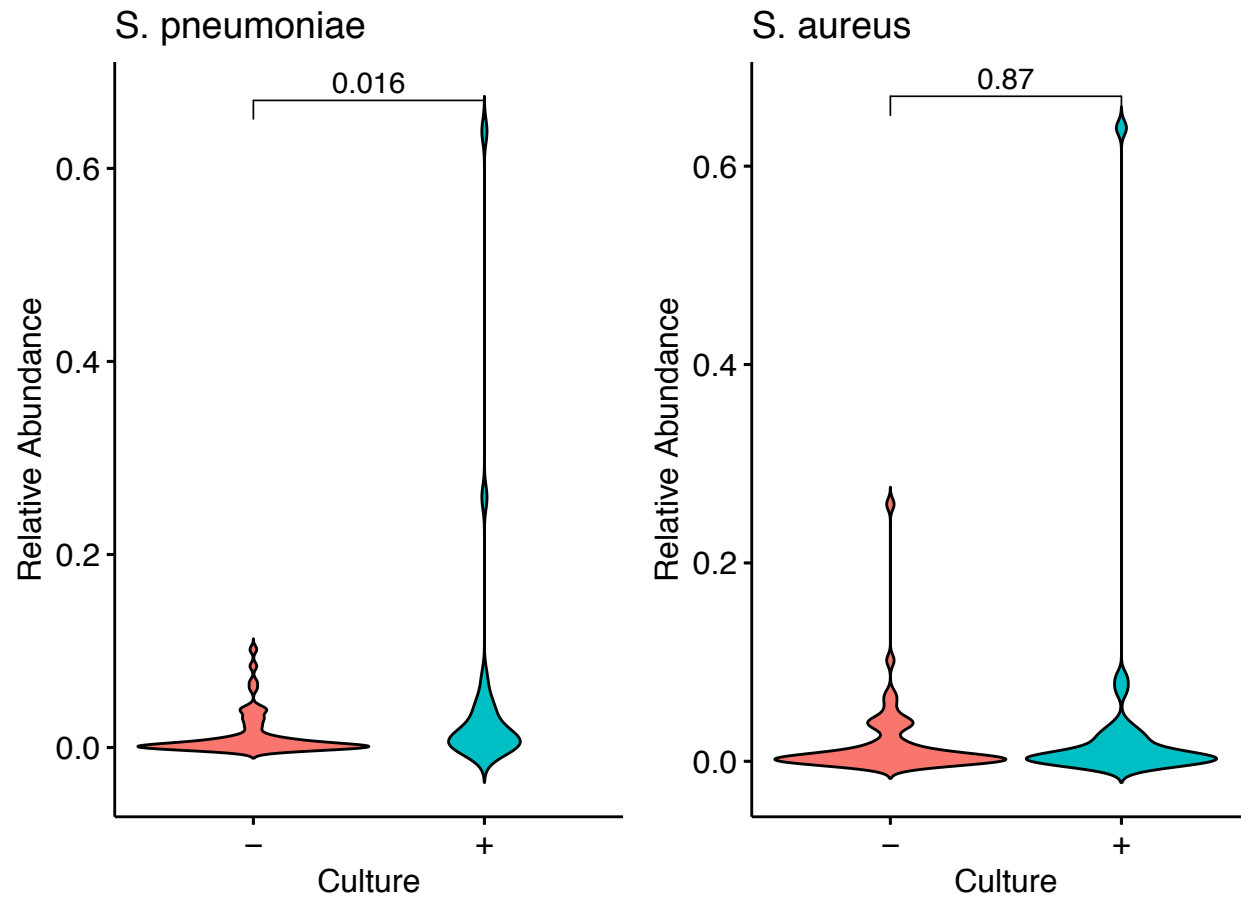
Supplementary Figure 3: Composition of control swab samples at Phylum and Genus level. Frequency indicates the cumulative frequency of ASVs with Genus X from the eight control samples.



Supplementary Figure 4: Log Transformed Abundances of Phylum-level ASV Classifications in Nasal (top) and Oral (bottom) Samples. Bars are coloured by site of location. Firmicutes, Actinobacteria and Proteobacteria were the most abundant phyla in Nasal samples with Bacteroidetes, Firmicutes and Proteobacteria the most common in Oral samples. A clear increase in Firmicutes in nasal samples from Site 2 (Kampung Berua) can be seen.



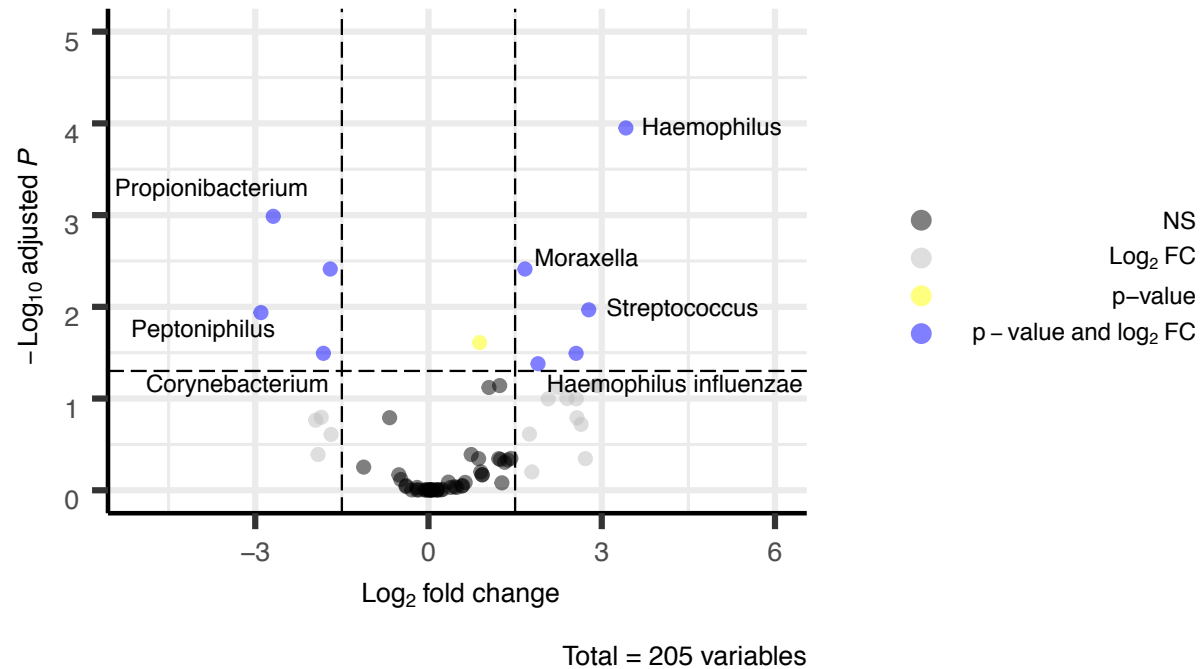
Supplementary Figure 5: Network and Double PCoA Beta-diversity plots comparing individuals that had reported respiratory symptoms within the last month (top) or within the last three months (bottom). Individuals reporting respiratory symptoms are shown in blue. Shape denotes the location with triangles Site 1 - Kampung Sungai Pergam and circles being Site 2 - Kampung Berua. Network construction was based on Jaccard distance with a maximum distance of 0.7.



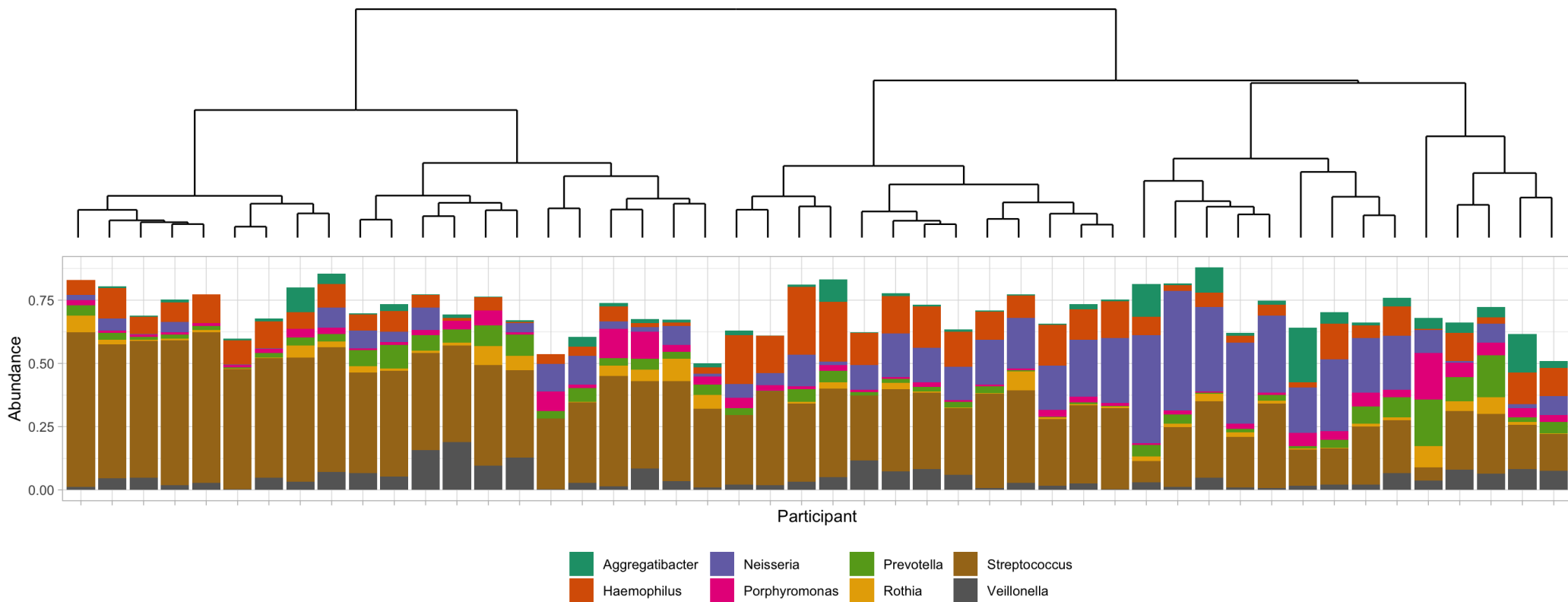
Supplementary Figure 6: Comparison of the relative abundance of Streptococcal and Staphylococcus ASVs between those that were culture positive or negative for *Streptococcus pneumoniae* or *Staphylococcus aureus*. *P* values were calculated using the nonparametric Wilcoxon test.

Age Group Comparison 5-17 vs.18-49

EnhancedVolcano



Supplementary Figure 7: Differentially Abundant ASVs Between Older Children (5-17 years) and Adults (18-49). Volcano plot showing the ASVs that were differentially abundant in comparisons between the two age groups. Here a p-value cut-off of 0.05 and a Log₂ fold change of 1.5 was applied. Those ASVs that pass both these thresholds are labelled where a genus and/or species taxonomic assignment was available. The right hand-side of the plot shows ASVs that had increased abundance in the 5-17 years olds compared to the 18-49. Conversely those to the left are those more abundant in the older age group.



Supplementary Figure 8: Hierarchical Clustering of Oral Samples with Bray-Curtis Dissimilarity using Genus-level Relative Abundances. The dendrogram (top) shows clustering of samples with the below bar chart showing the relative abundance of the six most commonly observed Genera. Three genera are seen to be responsible for the majority of ASVs – *Streptococcus* (brown), *Haemophilus* (orange) and *Neisseria* (purple). Profiles characterised by an overabundance of *Streptococcus* or *Neisseria* are seen, as too are those containing both *Streptococcus* and *Haemophilus*.