



**HAL**  
open science

## First insight into oral microbiome diversity in Papua New Guineans reveals a specific regional signature

Nicole Pedro, Nicolas Brucato, Bruno Cavadas, Valentine Lisant, Rui Camacho, Christopher Kinipi, Matthew Leavesley, Luisa Pereira, Francois-Xavier Ricaut

### ► To cite this version:

Nicole Pedro, Nicolas Brucato, Bruno Cavadas, Valentine Lisant, Rui Camacho, et al.. First insight into oral microbiome diversity in Papua New Guineans reveals a specific regional signature. *Molecular Ecology*, In press, 10.1111/mec.16702 . hal-03808087

**HAL Id: hal-03808087**

**<https://hal.science/hal-03808087v1>**

Submitted on 10 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pedro Nicole (Orcid ID: 0000-0001-5667-1549)

## First insight into oral microbiome diversity in Papua New Guineans reveals a specific regional signature

**Running title:** Oral microbiome diversity in Papua New Guinea

Nicole Pedro<sup>1,2,3</sup>, Nicolas Brucato<sup>4</sup>, Bruno Cavadas<sup>1,2</sup>, Valentine Lisant<sup>4</sup>, Rui Camacho<sup>5,6</sup>, Christopher Kinipi<sup>7</sup>, Matthew Leavesley<sup>8,9,10</sup>, Luisa Pereira<sup>1,2\*</sup>, François-Xavier Ricaut<sup>4\*</sup>

1 – Instituto de Investigação e Inovação em Saúde, Universidade do Porto (i3S), 4200-135 Porto, Portugal

2 – Instituto de Patologia e Imunologia Molecular, Universidade do Porto (Ipatimup), 4200-465 Porto, Portugal

3 – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto (ICBAS), 4050-313 Porto, Portugal

4 – Laboratoire Évolution and Diversité Biologique (EDB UMR5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, 118 route de Narbonne, Bat 4R1, 31062 Toulouse, France

5 – FEUP – Faculdade de Engenharia, Universidade do Porto, 4200-465 Porto, Portugal.

6 – INESC TEC-Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Universidade do Porto, 4200-465 Porto, Portugal.

7 – Medical Clinic, University of Papua New Guinea, Waigani Drive, Port Moresby, PO111, Papua New Guinea

8 – Strand of Anthropology, Sociology and Archaeology, School of Humanities and Social Sciences, University of Papua New Guinea, National Capital District, Papua New Guinea

9 – College of Arts, Society and Education, James Cook University, Cairns, Queensland, Australia

10 – ARC Centre of Excellence for Australian Biodiversity and Heritage, University of Wollongong, Wollongong, New South Wales, Australia

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1111/mec.16702](https://doi.org/10.1111/mec.16702)

This article is protected by copyright. All rights reserved.

\* These authors should be considered joint senior authors.

**Corresponding author:** François-Xavier Ricaut - Laboratoire Évolution and Diversité Biologique (EDB UMR5174), Université de Toulouse Midi-Pyrénées, CNRS, IRD, UPS, 118 route de Narbonne, Bat 4R1, 31062 Toulouse, France; email: [francois-xavier.ricaut@univ-tlse3.fr](mailto:francois-xavier.ricaut@univ-tlse3.fr).

Accepted Article

## Abstract

The oral microbiota is a highly complex and diversified part of the human microbiome. Being located at the interface between the human body and the exterior environment, this microbiota can deepen our understanding of the environmental impacts on the global status of human health. This research topic has been well addressed in Westernized populations, but these populations only represent a fraction of human diversity. Papua New Guinea hosts very diverse environments and one of the most unique human biological diversities worldwide. In this work we performed the first known characterization of the oral microbiome in 85 Papua New Guinean individuals living in different environments, using a qualitative and quantitative approach. We found a significant geographical structure of the Papua New Guineans oral microbiome, especially in the groups most isolated from urban spaces. In comparison to other global populations, two bacterial genera related to iron absorption were significantly more abundant in Papua New Guineans and Aboriginal Australians, which suggests a shared oral microbiome signature. Further studies will be needed to confirm and explore this possible regional-specific oral microbiome profile.

**Keywords:** Oral microbiome; Papua New Guinea; Environment; Diet; Sahul.

## Introduction

In recent years the field of host-associated microbes has greatly increased due to a desire to understand how the microbiome impacts a broad array of hosts, such as animals, plants and algae (Parfrey, Moreau, & Russell, 2018). Microbes have been shown to be involved in host immunity, digestion and nutrient scavenging activities, facilitate metabolism and cellular growth, regulate behavior, and release microbial toxins, among others (Miller, Westlake, Cross, Leigh, & Bordenstein, 2021). However, the first step in understanding host-microbiota dynamics is to identify the composition of the microbiome, and for that unique microbiomes across diverse host populations must be studied (Parfrey et al., 2018).

The human oral microbiota is one of the most complex microbial communities in the human body, colonized by around 700 different types of microorganisms (Lamont, Koo, & Hajishengallis, 2018). Awareness of the role of oral microbes in health and disease has been increasing, not only in its obvious relation to oral diseases, such as periodontal disease (Costalonga & Herzberg, 2014) and oral cancer (Wang & Ganly, 2014), but also in how it relates to systemic (Graves, Correa, & Silva, 2019), cardiovascular (Bryan, Tribble, & Angelov, 2017) and gastrointestinal diseases (Flemer et al., 2018).

The oral microbiome is formed mainly by bacteria, but it also contains lower amounts of fungi, viruses, archaea, and protozoa. The top bacterial phyla detected in the oral microbiome are Firmicutes, Fusobacteria, Proteobacteria, Actinobacteria, Bacteroidetes and Spirochaetes (Deo & Deshmukh, 2019), but abundance of these taxa differs between populations. Several studies on oral microbiome have been performed for Westernized populations (Burcham et al., 2020; Nasidze, Li, Quinque, Tang, & Stoneking, 2009), which have been associated with a specific Western diet, low in fiber and fruit and high in refined sugars and fat (Cordain et al., 2005), but also in more remote populations, such as Native South Americans (Contreras et al., 2010), native Alaskans (J. Li et al., 2014), Philippines hunter-gatherers (Lassalle et al., 2018) and Batwa Pygmies (Nasidze et al., 2011).

Accepted Article

These studies revealed a high diversity in oral microbiome within and between the different populations. Different environments, lifestyles, eating habits, or genetic background are possible reasons for these variations (J. Li et al., 2014; Mousa, Chehadeh, & Husband, 2022).

Papua New Guinea (PNG) is one of the most culturally diverse and least urbanized countries in the world, with over 800 spoken languages and thousands of ethnic groups (Pawley, 2005). It also hosts a very high human genetic diversity (Brucato et al., 2021), including the highest proportion of archaic introgressions from Denisovans (~4%) in the world (Jacobs et al., 2019; Reich et al., 2011). The majority of Papua New Guineans live in a semi-traditional way with an agriculture-based diet, including sago, sweet potato, taro, cassava, yam, rice, maize, breadfruit, various nuts, sugarcane, pig, chicken and fruits, like coconut, banana and pineapple (Schapper, 2017). The PNG populations have a history of low levels of protein consumption, which has been improving in the last years with an increase in the importation of poultry and fish (Schmidt & Fang, 2021). However, with the importation of animal-sourced foods there has also been an increased importation of ultra-processed foods, and meals containing higher saturated fat (Schmidt & Fang, 2021). PNG has poor general health due to low socio-economic status and high burden of infectious diseases, such as pneumonia, malaria, tuberculosis and meningitis (Riley, 2009). While the gut microbiota of different regions within PNG showed higher bacterial diversity and different abundance profiles in comparison with Westernized populations (Martinez et al., 2015), a comprehensive analysis of the oral microbiome for diverse ethnic groups and regions in PNG has not been performed.

In this work we combined the results from two different methodologies (Figure 1) to characterize the microbial community from DNA extracted from PNG saliva samples: (1) qualitative results from a microbial detection array, which contains specific probes to the different microbial species, and (2) quantitative from whole genome sequencing (WGS) reads unmapped to the human genome, which were aligned to a microbial database. The array methodology is cost-effective, highly sensitive and

accurate (Thissen et al., 2014), and easy to analyze (Gardner, Jaing, McLoughlin, & Slezak, 2010; McLoughlin, 2011), being the ideal technique to validate an off-target microbiome inference from the human WGS. In fact, the recovery of off-target reads obtained from human WGS provides researchers with excellent secondary opportunities to study data beyond the intended target regions (Cavadas et al., 2020; Gomes et al., 2019; Samuels et al., 2013). This technique permits users to detect a larger number of microorganisms (Ranjan, Rani, Metwally, McGee, & Perkins, 2016) and to obtain the abundance of the identified microbial communities. However, it is more expensive than the array methodology and involves much more complex bioinformatic analysis.

The quantitative oral microbiome data allowed us to assess differences between diverse PNG populations in light of diet and/or environmental factors, and against other worldwide populations. Our working hypothesis was that a specific oral microbiome signature might be found in these populations due to adaptation to their specific biodiversity, cultural lifestyle or genetic background, since their 50kya-establishment in the ancient Sahul region (New Guinea/Australia).

## Material and Methods

### Samples

Ninety-two unrelated individuals from different regions of PNG contributed saliva samples during field seasons between 2016 and 2019. Four different sampling sites were selected in order to access (1) population from different environments (coastal, highlands, lowlands), (2) different lifestyles (communities with a traditional subsistence or urban lifestyles and with variable access to health centers, grocery shops, medicine and Westernized food), and (3) being geographically distinct from each other. This sample set permits capture of some of the diversity of the PNG populations by defining five main groups: Port Moresby, includes individuals sampled in the east coastal lowlands (n=20, Central, Oro and Milne Bay provinces); Simbu province, Highlanders sampled in the Highlands (n=20, Simbu province); East Sepik province, individuals sampled in the north inland lowlands (n=20); Western province, individuals sampled in the west coastal lowlands (n=20); and East highlanders living in Port Moresby, Highlanders sampled in the east coastal lowlands (n=12), (Figure 2A, Table S1). The saliva samples were collected with the Oragene DNA OG-500 collection kit (DNA Genotek Inc., Ottawa, Canada), and the nucleic acids were extracted following the manufacturer's instructions.

### Ethics approval

This study was approved by the Medical Research Advisory Committee of PNG (National Department of Health) under research ethics clearance MRAC 16.21 and by the French Ethics Committees (Committees of Protection of Persons IE-2015-837 (1)). Permission to conduct research in PNG was granted by the National Research Institute of PNG (permit 99902292358), with full support from the School of Humanities and Social Sciences, University of PNG.

All samples were collected from healthy unrelated adult donors, all of whom provided written informed consent. In each sampling location, a full presentation of the project was made, followed



by discussion with each donor to ensure that they fully understood the project. Participants were surveyed for language affiliation(s), current residence, date and place of birth, and a short genealogy up to three generations to establish regional ancestry.

#### Microbiome inference from Axiom Microbiome Array

From the DNA extracted, we performed microbiome diversity characterization with the Axiom™ Microbiome Array from Thermo Fisher Scientific. First, cDNA synthesis was performed using the SuperScript™ VILO™ cDNA Synthesis Kit from Invitrogen™ (Waltham, MA, USA). Afterwards, microbiome diversity was characterized with the Axiom™ Microbiome Array (MDA) from Thermo Fisher Scientific (Waltham, MA, USA), which can rapidly identify the microbial community at the species or strain level, giving a profile of which microorganisms are present in the analyzed samples. The Microbiome array harbors a total of 1.38 million probes that detect over 12,513 microbial species including archaea, bacteria, fungi, protozoa, and viruses, and the probes were designed against both conserved regions (for the detection of higher taxonomic ranks) and unique regions (species/strain specific) of the microbial genomes with a sensitivity of 100-1000 genome copies and achieving 100% accuracy in species identification. The 96-well array plate version of the Microbiome array was used to run all the samples on the GeneTitan instrument (Thermo Fisher Scientific, Waltham, MA, USA). A volume of 20µL of DNA from all samples was used as input into the MDA, including a no-template control and a positive control for quality control. The standard manufacturer's protocol was followed for all steps of the workflow. Quality control (QC) was performed using the QC option in the Axiom™ Microbial Detection Analysis Software (MiDAS), which analyzes specific probes included in the Microbiome array to target regions of the human genome that are non-polymorphic (sites that do not vary in sequence from one individual to the next). This QC analysis generates a metric termed Dish QC (DQC) and applies a threshold of DQC above 0.95 for the samples to be considered as having passed QC. Six samples were removed from the analysis after

QC: PNG237 and PNG238 (East Sepik province), PNG128, PNG514 and PNG525 (Western province), and PNG205 (East highlanders living in Port Moresby). The remaining samples (n=86) were then further analyzed with MiDAS, which analyses all targets considered positive and reaches a consensus for the microbial community present in the samples. Positive targets are the probes with signal intensity above the 99<sup>th</sup> percentile of the random control probe intensities and with more than 20% of target-specific probes. These probe sequences are available from Thermo Fisher Scientific (Waltham, MA, USA) as part of the MiDAS software. To note, this technology is unable to provide a view of the varying prevalence of the microorganisms present in analyzed samples.

#### Microbiome inference from human whole genome sequences

For 68 individuals, also characterized by Microbiome array, we used reads extracted from WGS from (1) previously published studies (Brucato et al., 2021), and (2) from newly unpublished whole genome sequences data (Table S1). The sequencing procedure has been described previously (Brucato et al., 2021). Briefly, sequencing libraries were prepared using the TruSeq DNA PCR-Free HT kit. About 150-bp paired-end sequencing was performed on the Illumina HiSeq X5 sequencer. We aligned the fastq files against the human reference GRCh38 (GCA 000001405.15) using minimap2-2.23 (H. Li, 2018) and the unmapped reads were retrieved. For taxonomic assignment of the unmapped sequencing reads we resorted to KRAKEN 2 tool (Wood, Lu, & Langmead, 2019) using the Standard plus protozoa & fungi database (PlusPF; includes archaea, bacteria, viral, plasmid, human, UniVec\_Core, protozoa and fungi; updated on January 27<sup>th</sup>, 2021). We then used Bracken (Lu, Breitwieser, Thielen, & Salzberg, 2017) (Bayesian Reestimation of Abundance after Classification with KRAKEN) to estimate abundance at the genus level, using the taxonomic assignments made by Kraken. Since these PNG populations are understudied, we are unable to consider possible unidentified microbiota species that may be specific to them. We therefore performed all analysis on genus level to reduce the bias caused by using the Microbiome array and the WGS databases. We

applied Bracken's default threshold of 10 reads as the minimum number of reads required for a classification at the specified rank. To remove low background noise on the WGS results due to minor contaminants, we applied a threshold for which all genera below a given percentage would be removed. This threshold was determined by calculating the best correlation of alpha-diversity values obtained from the Chao1 index between each sample using the qualitative and quantitative results, applying several thresholds, from 1% to 0.0001% (1%, 0.5%, 0.1%, 0.05%, 0.01%, 0.005%, 0.001%, 0.0005%, 0.0001%). Applying a threshold of 0.01% to the WGS results, the Chao1 values obtained from both methodologies presented the best correlation from all the tested thresholds.

### Microbial composition

We created phylogenetic trees and calculated the respective abundance of the genera found in the five groups from the qualitative results (Microbiome array). The abundance in this case refers to the percentage of samples in the group in which the specific taxon was detected. We used RStudio software (RStudio Team, 2020) to perform pairwise chi-squared tests and correction with Holm's method to see which genera were significantly different. Venn diagrams with the percentage of Bacteria and Virus were created to help to visualize the similarities and differences between the five groups resorting to the webtool Bioinformatics & Evolutionary Genomics (<https://bioinformatics.psb.ugent.be/webtools/Venn/>). Discrimination between the five different groups using quantitative data was performed with the Linear discriminant analysis Effect Size (LEfSe) algorithm (Segata et al., 2011) with the calculation of a linear discriminant analysis (LDA) score, available at the Huttenhower lab Galaxy server (<http://huttenhower.sph.harvard.edu/galaxy/>) (Afgan et al., 2018). Statistical differences were found under a Kruskal-Wallis test which identified the features with significant variation across the five groups. Then, pairwise Wilcoxon tests were performed between each pair of groups, using the "one-against-all" option for multi-class analysis, which only requires that one of the pairwise Wilcoxon tests to be significant in order to keep the

feature in the list of significantly discriminating features (Segata et al., 2011). Only taxa with a p-value < 0.05 and with a LDA value greater than 3 were considered enriched.

#### Alpha-diversity

The alpha-diversity of each sample was calculated by estimating the genus and species richness, evenness, and diversity using the Phyloseq package v1.23.1 (McMurdie & Holmes, 2013). Alpha-diversity measures were calculated differently for qualitative and quantitative results. For the qualitative results (from the Microbiome array data) we calculated Chao1 (Hughes, Hellmann, Ricketts, & Bohannan, 2001), which measures the richness of the samples and is more suitable for presence/absence data. We created a ggplot for each alpha-diversity metric and performed pairwise comparisons using t tests and correction with Holm's method using RStudio (RStudio Team, 2020). After alpha-diversity analysis, one sample was considered an outlier, PNG54 (Port Moresby), as its Chao1 index value falls out of the mean +/- 3SD for all samples (Figure S1), so it was removed from further analyses.

#### Beta Diversity and unrooted neighbor-joining tree

The phylogenetic information for the genera present in our samples was retrieved (on October 20<sup>th</sup>, 2021) from the Time Tree website (<http://www.timetree.org/>). This information was used to calculate Jaccard (Koleff, Gaston, & Lennon, 2003), unweighted and weighted UniFrac distances (Lozupone, Lladser, Knights, Stombaugh, & Knight, 2011), beta diversity measures that analyze the similarities and differences among groups. The Jaccard and the unweighted UniFrac distances were applied to the qualitative data (Microbiome array) as they do not consider abundances in the calculations and the weighted UniFrac distances were used on the quantitative data (WGS results). The unrooted neighbor-joining tree was used to represent distances between the five groups. A permutation multivariate analysis (1000 permutations) of variance (PERMANOVA; (Anderson, 2017)) of the beta diversity distances (to reflect the phylogenetic diversity between genera) was also

performed and plotted in a non-metric multidimensional scale (nMDS). The null hypothesis tested by PERMANOVA was that the centroids of each group are equivalent, under the assumption of exchangeability of the samples among groups.

#### Comparative analysis of oral microbiome data from other populations

The qualitative results (Microbiome array) were used to perform the analysis of the PNG oral microbiome (Alpha-diversity) as we have more available samples. To have a global perspective of the oral microbiome of PNG populations in the global oral microbiome we compared the abundance of the core microbiome obtained from the quantitative results (WGS) in this study with other published studies for Philippines hunter-gatherers and farmers, Batwa Pygmies, Aboriginal Australians and Westernized individuals (Handsley-Davis et al., 2021; Lassalle et al., 2018; Nasidze et al., 2011). For the hunter-gatherers and traditional farmer groups from the Philippines and their curated data of the Western controls from the Human Microbiome Project (Lassalle et al., 2018), we downloaded their published microbial metagenomic read data sets from ENA ([www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena); BioSample Accession nos. ERS1202862—ERS1202885) and processed their dataset with our workflow. These files were merged with our dataset to perform beta-diversity analysis and to calculate the abundances of the core microbiome. For the Australian Aboriginals and Torres Strait Islanders (Handsley-Davis et al., 2021) and the Batwa Pygmies populations (Nasidze et al., 2011), we were not able to obtain WGS raw data, as they only sequenced a region of the microbial 16S rRNA gene (V4 for the Australian Aboriginals and Torres Strait Islanders and V1 and V2 for the Batwa pygmies). For these, we retrieved the values of abundance for the core microbiome that were described and available in their publications. Previous research has shown that microbiome results might differ quite substantially depending on the underlying method used (Durazzi et al., 2021; Hillmann et al., 2018), but a high correlation has been found between these different methods regarding the most abundant species in the samples. As such, we performed comparative analysis of only the most

abundant taxon at genus level to limit inter-study variation due to the comparison between results from different methodologies (WGS and 16S).

## Results

A total of 85 samples from the Microbiome array (qualitative results) were analyzed in this study divided into five groups, based on the sampling locations and individual origins from: Port Moresby (n=19); Simbu province (n=20); East Sepik province (n=17); Western province (n=18); and East highlanders living in Port Moresby (n=11) (Figure 2A). Our cohort includes 71 males (84%) and 14 females (16%), with an average age at time of collection of 38.6 years (range from 20-67 years; Supplementary table S1). After QC, 81 microbial genera were detected in the cohort. We controlled for substructure due to age or gender, and found no correlation (for age, p-value = 0.36; Figure S2A; for gender, p-value = 0.4; Figure S2B), confirming previous results (Nasidze et al., 2009).

A subset of the cohort was also characterized using available WGS to obtain quantitative results (n=65): Port Moresby (n=13); Simbu province (n=18); East Sepik (n=9); Western province (n=18); and East highlanders living in Port Moresby (n=7) (Supplementary Table S1). Reads from WGS data not mapping to the human genome were analyzed with KRAKEN 2 (Wood et al., 2019), resulting in an average of 18 652 535 reads matching microbial genomes per individual (ranging from 1 483 445 - 44 976 955 reads). As expected, the WGS data identified a much higher number of genera (1757 genera; Table S2) than with the Microbiome array data. In fact, the BRAKEN method used in the WGS inference limits the number of reads to 10, while the Microbiome array has a sensitivity of 100-1000 genome copies. Hence, when comparing the concordance in inferred genera between the two methods, these values are high for the most abundant genera (Figure 2B; Table S3): an average of 92% match filtering to the genera with abundance above 1% (28 out of the 65 samples achieving 100% match); 84% when filtering to the genera with abundance above 0.5% (11 out of 65 samples reaching a 100% match); 58% when lowering the threshold to 0.1%; and 30% for threshold 0.01%. Even so, the overall diversity (alpha-diversity, Chao1) per sample is highly correlated between the

two methods when using all genera above 0.01% of presence ( $R= 0.64$ ;  $p\text{-value} = 1.3 \times 10^{-08}$ ; Figure 2C). Further analyses based on WGS were done for this threshold.

#### Core oral microbiome in PNG

Our qualitative analysis (Microbiome array) of the PNG oral microbiome detected 75 bacterial genera and 9 viral genera (Table S4). Although the array includes probes for RNA viruses and we performed RNA to cDNA reaction, we were not able to identify any RNA viruses in the PNG samples. This is most probably due to the storage of samples at room temperature, leading to the instability of RNA molecules. Around 43% of the detected bacteria genera are common to all the groups (Figure 3A). Of these, there are 12 genera that were present in more than 50% of the samples in all groups, namely *Veilonella*, *Streptococcus*, *Prevotella*, *Fusobacterium*, *Rothia*, *Haemophilus*, *Neisseria*, *Porphyromonas*, *Treponema*, *Leptotrichia*, *Actinomyces* and *Selenomonas* (Figure S3).

On exploration the quantitative results (WGS), the 12 genera mentioned above account for 80% of the oral microbiome of the PNG population (Table S5). *Prevotella* (~17%), *Neisseria* (~13%) and *Streptococcus* (~13%) are the three most abundant taxa. They represent more than 40% of the PNG oral microbiome, followed by *Haemophilus* (~9%), *Rothia* (~8%), *Veilonella* (~7%), *Fusobacterium* (~4%), *Porphyromonas* (~2%), *Actinomyces* (~2%), *Leptotrichia* (~2%), *Treponema* (~2%) and *Selenomonas* (~1%). These 12 taxa represent the core oral microbiome of PNG populations. The remaining 20% is composed by lower abundance taxa which differentiate PNG groups.

#### Oral Microbiome differences between PNG populations

We detected significant differences between PNG groups at the global level of oral microbiome, comparing their alpha-diversity (Chao1), based on qualitative data (Microbiome array; Figure 3B). The East Sepik group, the most remote (as in difficult to reach by transportation, and with less access to health care services, electricity, education, processed food, etc.) community of our sampling, has a significantly lower alpha-diversity value than all other PNG groups ( $p < 0.05$ ), except the Port Moresby



group (Figure 3B). The four other PNG groups show relative similar alpha-diversity ( $p > 0.05$ , Figure 3B). These results mirror the results of beta diversities (Jaccard and unweighted UniFrac; Figure S4), although the PERMANOVA test was not significant for the East Sepik differentiation ( $p > 0.05$ ). The East Sepik has one of the highest percentages of unique genera (8%; Figure 3A) and only shares 54% of genera with the four other groups, while the values are higher for the other groups: 57% for Port Moresby; 67% for Simbu Province; 70% for Western Province; and 62% for East highlanders living in Port Moresby. This regional particularity is partly explained by the significantly higher frequency (in number of individuals of the group) of specific taxa such as *Vibrio*, *Capnocytophaga*, *Oribacterium*, *Peptostreptococcus*, *Megasphaera*, *Granulicatella*, *Gemella*, *Mycoplasma* and *Atopobium* when compared to each of the other PNG groups ( $p < 0.05$ , Figure 3C).

Quantitatively, the East Sepik oral microbiome is significantly enriched (medium abundance in the group) in several taxa such as *Haemophilus*, *Tannerella*, *Anabaena*, *Pseudonocardia*, *Aromatoleum*, *Actinobacillus*, *Histophilus*, *Dickeya*, *Aquifaea* and *Dietzia* ( $LDA > 3$ , Figure 4).

The Western province group, another relatively remote coastal community, shows one of the highest alpha-diversities and the highest number of specific taxa in PNG (8%, Figures 3A and 3B). Although qualitatively, no genera were present in more samples in the Western province group than in the rest of PNG, a high number of taxa are quantitatively enriched ( $LDA > 3$ , Figure 4) such as *Klebsiella*, *Fusobacterium*, *Starkeya*, *Corynebacterium*, *Anaerococcus*, *Mogibacterium*, *Liberibacter*, *Enterocloster*, *Parvimonas*, *Mediterraneibacter*, *Aeromonas*, *Anareocolumna* and *Pseudomonas*.

Groups from Port Moresby, Simbu province and East highlanders living in Port Moresby show relatively similar alpha-diversity in the oral microbiome (Figure 3B). Qualitative data show that the Simbu province group has 6% of unique genera while the Port Moresby group has no unique genera (Figure 3A). No statistical difference in the presence of specific taxa was observed despite living in very different geographical regions (e.g., capital city on the southeast coast and east highlands respectively, Figure 3B). Correspondingly, the group of East highlanders living in Port Moresby shows

no significant difference in the presence of taxa and only 3% of unique taxa (Figure 3A). However, each of these three groups can be characterized by the enrichment of some taxa like *Cronobacter* and *Flaviflexus* in Port Moresby, *Streptococcus* and *Gemella* in Simbu, and *Schaalia* in the East highlanders living in Port Moresby (LDA>3, Figure 4). The group of East highlanders living in Port Moresby shares 59% of their genera with the Simbu group but only 49% with the Port Moresby group (Figure 3A). The Alpha-diversity values in the East highlanders living in Port Moresby are more similar to the one from the Simbu group than to the Port Moresby group (Figure 3B). This lack of significant differences might be due to the incorporation of both microbiotas (initial oral microbiota from the highlands and the acquired microbiota from the coast) in the group of East Highlanders living in Port Moresby.

#### Oral microbiome of PNG and other populations

To have a broader perspective of the oral microbiome of PNG population samples, we merged the WGS results from our PNG populations with the available WGS samples from Philippines hunter-gatherers, Philippines traditional farmers and Western controls (Lassalle et al., 2018) and processed their dataset with our workflow. We generated an unrooted neighbor-joining for the eight populations (Figure 5A) which showed that the PNG populations are closer to one another and separated from the Philippines and the Western controls.

We also compared the abundance of the PNG core microbiome (Figure 5B; Table S5) obtained in this WGS-based dataset (Human Microbiome Project, 2012; Lassalle et al., 2018) with other 16S rRNA-based published studies for Australian Aboriginals and Torres Strait Islanders (Handsley-Davis et al., 2021) and Batwa Pygmies (Nasidze et al., 2011). Interestingly, the *Rothia* genera abundances in the five PNG groups (average 8.24%) are significantly higher (p-value =0.0013) than in all the other compared groups (Philippines hunter-gatherers, Philippines traditional farmers, Batwa Pygmies and Western Controls; average 3.87%), except for the Australian Aboriginals and Torres Strait Islanders

(8%). The above pattern is also significantly different for the *Neisseria* genera (p-value = 0.0405), with high abundances in the PNG groups (average 15.05%) and Australian Aboriginals and Torres Strait Islanders (14%) than in the remaining groups (9.58%). This enrichment of *Rothia* and *Neisseria* appear to be limited to the New Guinea-Australia region (Figure 5B), which might indicate a specific trait of these genera on the oral microbiome, potentially influenced by environmental conditions found in this geographic region.

*Veilonella* appears in higher abundances in the Western controls (11.86%), the Philippines traditional farmers (11.75%), in the Port Moresby group (9.05%) and the East Highlanders living in Port Moresby group (9.07%). This genus has been associated with groups having a Westernized diet (Clemente et al., 2015; Lassalle et al., 2018). These studies also reported an enrichment of *Haemophilus* in hunter-gatherers, which is not apparent in our results (Table S5), as the Western controls have a higher abundance of *Haemophilus* than the Hunter-gatherer groups. (Lassalle et al., 2018)

## Discussion

The core oral microbiome of PNG agrees with the normal microbiota in the oral cavity obtained in other studies (Aas, Paster, Stokes, Olsen, & Dewhirst, 2005; Deo & Deshmukh, 2019; Dewhirst et al., 2010). Bacterial genera such as *Prevotella*, *Neisseria* and *Streptococcus* represent most of the human oral microbiome diversity. However, as seen within our PNG populations, the abundances of each taxon might vary, potentially due to differential lifestyles, diets or host genetics (J. Li et al., 2014; Mousa et al., 2022). Both quantitative (WGS) and qualitative (Microbiome array) results show significant differences between the East Sepik group and the other PNG groups. This might be explained by the relative geographic isolation and a more traditional form of subsistence, with limited access to Westernized food, contrary to the other four groups (Pawley, 2005; Roscoe, 2005). Similar results were described in the oral microbiomes of isolated groups across the globe, such as the Native American Yanomami population (Clemente et al., 2015), the African Batwa (J. Li et al., 2014), and in the Philippines hunter-gatherer group (Lassalle et al., 2018), when compared to groups with a Westernized diet. Apart from the East Sepik group, all the other PNG groups included in this study have a mixed traditional and Westernized lifestyle (Simbu and Western provinces and Port Moresby), which might explain the similarity of their respective oral microbiome compositions. Our results for the group of East Highlanders living in Port Moresby suggests an incorporation of both microbiotas (initial oral microbiota from the highlands and the acquired microbiota from the capital city), with higher similarity with the initial microbiota. There are no studies of the effect of human migration on oral microbiome, but for gut microbiome (Vangay et al., 2018), a complete change of gut microbiome diversity and function from the native to a new environment was observed. Our results indicate a more complex interaction between the original and the recently acquired oral microbiome, emphasizing the need for further study on the plasticity of the human oral microbiome.

An interesting pattern of genera enrichment was found in the populations from New Guinea and Australia (PNG groups and Australian Aboriginals and Torres Strait Islanders). We detected higher abundances for the *Neisseria* and *Rothia* genera than other populations of the world. New Guinean and Aboriginal Australian populations share a common history, since the early settlement of these territories 50,000 years ago (Allen & O'Connell, 2020; Brucato et al., 2021; Pedro et al., 2020), then connected into the ancient continent Sahul. A secondary gene flow linked to the Austronesian dispersal 3kya led to a gradient of Asian genetic ancestry in coastal groups of the region (Brucato et al., 2021; Mallick et al., 2016). However, our limited dataset (no group from New Guinea with a predominant Austronesian genetic ancestry) and current knowledge on microbiome research, do not support a strong link between a population genetic ancestry (e.g. Austronesian) and a specific oral microbiome signature, despite some encouraging new studies (Weyrich, 2021). It is, nevertheless, remarkable that our study revealed a specific pattern in the oral microbiome shared by New Guineans and Aboriginal Australian groups, that are genetically separated for at least 30 000 years (Brucato et al., 2021; Malaspinas et al., 2016). This specific pattern may have been accentuated by limited gene flow between southeast coastal New Guinean lowlanders and northeast Indigenous Australians until at least 10kya, and during the Austronesian period of the last 3ky (Bellwood, 2007; Brucato et al., 2021; Karmin et al., 2022; Lipson et al., 2014; Pedro et al., 2020). This may suggest a Sahul-specific oral microbiome diversity, caused by a shared environment and/or a shared ancient ancestry. Among the Sahul-specific oral microbiome, the *Rothia* genus has been shown to have positive effects on the host's iron pool, by the production of the iron-scavenging enterobactin (Uranga, Arroyo, Duggan, Gerwick, & Edlund, 2020). Enterobactin is a siderophore protein released by bacteria into the surrounding environment to bind to available iron, which is then used in several biochemical pathways of the bacteria and the host (Raymond, Dertz, & Kim, 2003). Both *Rothia* and *Neisseria* genera appear to be depleted in the saliva of iron-deficient individuals, compared to healthy controls, which establishes their importance in the iron homeostasis (Xi et al., 2019). Iron

Accepted Article

deficiency is an important health matter in PNG and Indigenous Australian communities, associated with increased morbidity and mortality (Leonard, Buttner, Thompson, Makrides, & McDermott, 2019; Manning et al., 2012). These genera might be more abundant in the human oral microbiome of populations from the ancient Sahul region to, in part, compensate a potential iron deficiency. Other worldwide studies of oral microbiome in healthy individuals did not report this *Rothia* and *Neisseria* pattern, except for Nasidze et al. (2009) that show a higher abundance of the *Rothia* genus in the German population, but no interpretation was provided. Additionally, in the gut microbiome of Papua New Guineans individuals, the *Lactobacillus* genus, which is able to sense iron deprivation and communicate this iron need to the host (Das et al., 2020), was found highly enriched when compared to US individuals (Martinez et al., 2015). Iron plays an important role in infection and immunity, for both pathogens and host cells (Nairz & Weiss, 2020). While it is well-known that pathogens and diet are motors for human adaptation, understudied populations still have the potential to shed light on new adaptations (Pedro, Pinto, Cavadas, & Pereira, 2021). In this case, the environmental pressure (e.g. high pathogenic pressure, specific diet and plant use) observed in New Guinea and north Australia (Horwood et al., 2019) might have led to the development of a Sahul-specific oral microbiome.

This unexpected pattern of the PNG oral microbiome clearly needs further exploration. As this data was collected opportunistically, potential environmental and cultural cofactors were not controlled for. In particular, betel nut chewing (Hernandez et al., 2017; Zhong et al., 2021) has shown to be highly associated with variation in the oral microbiome in other populations and is a common habit to all the PNG locations. The impact of this specific cultural behavior would presumably result in an alteration of the global oral microbiome in all PNG populations, which would not greatly affect our results. Nevertheless, the effect of betel nut chewing in oral microbiome is an important question and should be further explored in the PNG populations. The time of sample collection might also influence the results obtained for oral microbiome, as seasonality in sample collection should be

Accepted Article

considered in further studies, and might increase the power in discriminating between the oral microbiomes of the PNG populations (Clemente et al., 2015). Oral health is also one of the factors that influence oral microbiota (Chukkapalli et al., 2015; Costalonga & Herzberg, 2014) and is common to all PNG populations. We have no information available concerning oral hygiene habits of the participants, but the influence of dental hygiene should not be ignored. All these external factors should be integrated in future analyses to evaluate their impact on our results. However, the convergence of our results in PNG with those from another study on Australian Aboriginals and Torres Strait Islanders (Handsley-Davis et al., 2021) suggests that the influence of these potential cofactors is unlikely to be major.

The detection of a possible Sahul-specific oral microbiome diversity may be potentially related to adaptation to their specific biodiversity, cultural lifestyle or genetic background. The results obtained in this study highlight the value of studying the microbiomes of more isolated groups to fully encompass the ecology and biogeography of the host–microbiota interactions. As the microbiota can affect the host’s phenotypic traits, host-microbe relationships can evolve in selective environments, conferring fitness advantages to the host. The influence of the microbiome in host ecology and evolution is more easily studied in non-human mammals, which are subjected to variation in nutritional intake, due to the seasonal availability of foods, changes in their natural habitat and higher disease burden (Amato, 2016; Clayton et al., 2018). In fact, Amato et al. (2015) reported this relationship in black howler monkeys (*Alouatta pigra*). The researchers showed that the gut microbiota adapts its composition based on the higher consumption of available fruit versus a higher consumption of leaves. In most modern human populations, which have access to a broad availability of food and medicine year-round, there is limited selective pressure that could affect the microbiome of the host (Amato, 2016; Colquhoun & Lyon, 2001). However, human populations residing in rural or remote areas (as for PNG populations, especially the East Sepik province) are still subjected to the pressure of seasonal food availability. A potential study including more samples and more relevant

data from these populations will be developed to further study, validate and understand how ecology, ancestry and human lifestyle impacts microbiome structure. Only by studying the different microbiotas in the ecosystem of diverse hosts, can we truly highlight the unique characteristics that defines the symbiotic relationship between microbe and host (Parfrey et al., 2018).



## Acknowledgements

We especially thank all of our study participants. We also thank Mathilde André and Georgi Hudjashov (University of Tartu) for their help in data analysis. We thank Roxanne Tsang, Jason Kariwiga, Teppsy Beni, Kylie Sesuki (University of Papua New Guinea), Alois Kuaso, and Kenneth Miamba (National Museum and Art Gallery, Papua New Guinea), and John Muke (Social Research Institute, Papua New Guinea) for their help during the sampling campaigns. We acknowledge support from the GenoToul bioinformatics facility of Genopole Toulouse Midi-Pyrénées, France; the LabEx TULIP, France; and the Centre National de Recherche en Génomique Humaine (CNRGH), Evry, France. This work was supported by the French National Research Agency (ANR; grant PAPUAEVOL n° ANR-20-CE12-0003-01), the French Ministry of Foreign and European Affairs (French Prehistoric Mission in Papua New Guinea), and the French Embassy in Papua New Guinea. The Portuguese Foundation for Science and Technology (FCT) financed the PhD grant to NP (SFRH/BD/136299/2018). i3S is supported by Fundo Europeu de Desenvolvimento Regional (FEDER) funds through the COMPETE 2020 – Operational Program for Competitiveness and Internationalization (POCI), Portugal 2020, and by Portuguese funds through FCT/Ministério da Ciência, Tecnologia e Inovação in the framework of the project ‘Institute for Research and Innovation in Health Sciences’ (POCI-01-0145-FEDER-007274). The authors also acknowledge the support of the i3S Genomics Scientific Platform, member of the GenomePT project (POCI-01-0145-FEDER-022184), supported by FEDER funds through COMPETE 2020, POCI, Lisboa Portugal Regional Operational Program (Lisboa2020), Algarve Portugal Regional Operational Program (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement and FCT. Lastly, we would like to thank the Reviewers for the time and effort spent in reviewing the manuscript. All their valuable comments and suggestions helped us improve the quality of the manuscript.

## References

- Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I., & Dewhirst, F. E. (2005). Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol*, *43*(11), 5721-5732. doi:10.1128/JCM.43.11.5721-5732.2005
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., . . . Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*, *46*(W1), W537-W544. doi:10.1093/nar/gky379
- Allen, J., & O'Connell, J. F. (2020). A different paradigm for the initial colonisation of Sahul. *Archaeology in Oceania*, *55*(1), 1-14. doi:10.1002/arco.5207
- Amato, K. R. (2016). Incorporating the gut microbiota into models of human and non-human primate ecology and evolution. *Am J Phys Anthropol*, *159*(Suppl 61), S196-215. doi:10.1002/ajpa.22908
- Amato, K. R., Leigh, S. R., Kent, A., Mackie, R. I., Yeoman, C. J., Stumpf, R. M., . . . Garber, P. A. (2015). The gut microbiota appears to compensate for seasonal diet variation in the wild black howler monkey (*Alouatta pigra*). *Microb Ecol*, *69*(2), 434-443. doi:10.1007/s00248-014-0554-7
- Anderson, M. J. (2017). Permutational Multivariate Analysis of Variance (PERMANOVA). In *Wiley StatsRef: Statistics Reference Online* (pp. 1-15).
- Bellwood, P. (2007). *Prehistory of the Indo-Malaysian Archipelago Revised Edition*: ANU Press.
- Brucato, N., Andre, M., Tsang, R., Saag, L., Kariwiga, J., Sesuki, K., . . . Ricaut, F. X. (2021). Papua New Guinean Genomes Reveal the Complex Settlement of North Sahul. *Mol Biol Evol*, *38*(11), 5107-5121. doi:10.1093/molbev/msab238
- Bryan, N. S., Tribble, G., & Angelov, N. (2017). Oral Microbiome and Nitric Oxide: the Missing Link in the Management of Blood Pressure. *Curr Hypertens Rep*, *19*(4), 33. doi:10.1007/s11906-017-0725-2
- Burcham, Z. M., Garneau, N. L., Comstock, S. S., Tucker, R. M., Knight, R., Metcalf, J. L., & Genetics of Taste Lab Citizen, S. (2020). Patterns of Oral Microbiota Diversity in Adults and Children: A Crowdsourced Population Study. *Sci Rep*, *10*(1), 2133. doi:10.1038/s41598-020-59016-0
- Cavadas, B., Camacho, R., Ferreira, J. C., Ferreira, R. M., Figueiredo, C., Brazma, A., . . . Pereira, L. (2020). Gastric Microbiome Diversities in Gastric Cancer Patients from Europe and Asia Mimic the Human Population Structure and Are Partly Driven by Microbiome Quantitative Trait Loci. *Microorganisms*, *8*(8), E1196. doi:10.3390/microorganisms8081196
- Chukkapalli, S. S., Rivera-Kweh, M. F., Velsko, I. M., Chen, H., Zheng, D., Bhattacharyya, I., . . . Kesavalu, L. (2015). Chronic oral infection with major periodontal bacteria *Tannerella forsythia* modulates systemic atherosclerosis risk factors and inflammatory markers. *Pathogens and disease*, *73*(3), ftv009. doi:10.1093/femspd/ftv009
- Clayton, J. B., Gomez, A., Amato, K., Knights, D., Travis, D. A., Blekhman, R., . . . Johnson, T. J. (2018). The gut microbiome of nonhuman primates: Lessons in ecology and evolution. *Am J Primatol*, *80*(6), e22867. doi:10.1002/ajp.22867
- Clemente, J. C., Pehrsson, E. C., Blaser, M. J., Sandhu, K., Gao, Z., Wang, B., . . . Dominguez-Bello, M. G. (2015). The microbiome of uncontacted Amerindians. *Sci Adv*, *1*(3), e1500183. doi:10.1126/sciadv.1500183
- Colquhoun, A., & Lyon, P. (2001). To everything there was a season: deconstructing UK food availability. *Food Service Technology*, *1*(2), 93-102. doi:10.1046/j.1471-5740.2001.00010.x
- Contreras, M., Costello, E. K., Hidalgo, G., Magris, M., Knight, R., & Dominguez-Bello, M. G. (2010). The bacterial microbiota in the oral mucosa of rural Amerindians. *Microbiology (Reading)*, *156*(Pt 11), 3282-3287. doi:10.1099/mic.0.043174-0

- Cordain, L., Eaton, S. B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B. A., . . . Brand-Miller, J. (2005). Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr*, *81*(2), 341-354. doi:10.1093/ajcn.81.2.341
- Costalonga, M., & Herzberg, M. C. (2014). The oral microbiome and the immunobiology of periodontal disease and caries. *Immunol Lett*, *162*(2 Pt A), 22-38. doi:10.1016/j.imlet.2014.08.017
- Das, N. K., Schwartz, A. J., Barthel, G., Inohara, N., Liu, Q., Sankar, A., . . . Shah, Y. M. (2020). Microbial Metabolite Signaling Is Required for Systemic Iron Homeostasis. *Cell Metab*, *31*(1), 115-130 e116. doi:10.1016/j.cmet.2019.10.005
- Deo, P. N., & Deshmukh, R. (2019). Oral microbiome: Unveiling the fundamentals. *Journal of oral and maxillofacial pathology: JOMFP*, *23*(1), 122-128. doi:10.4103/jomfp.JOMFP\_304\_18
- Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C., Yu, W. H., . . . Wade, W. G. (2010). The human oral microbiome. *J Bacteriol*, *192*(19), 5002-5017. doi:10.1128/JB.00542-10
- Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., & De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep*, *11*(1), 3030. doi:10.1038/s41598-021-82726-y
- Flemer, B., Warren, R. D., Barrett, M. P., Cisek, K., Das, A., Jeffery, I. B., . . . O'Toole, P. W. (2018). The oral microbiota in colorectal cancer is distinctive and predictive. *Gut*, *67*(8), 1454-1463. doi:10.1136/gutjnl-2017-314814
- Gardner, S. N., Jaing, C. J., McLoughlin, K. S., & Slezak, T. R. (2010). A microbial detection array (MDA) for viral and bacterial detection. *BMC Genomics*, *11*(1), 668. doi:10.1186/1471-2164-11-668
- Gomes, S., Cavadas, B., Ferreira, J. C., Marques, P. I., Monteiro, C., Sucena, M., . . . Seixas, S. (2019). Profiling of lung microbiota discloses differences in adenocarcinoma and squamous cell carcinoma. *Sci Rep*, *9*(1), 12838. doi:10.1038/s41598-019-49195-w
- Graves, D. T., Correa, J. D., & Silva, T. A. (2019). The Oral Microbiota Is Modified by Systemic Diseases. *J Dent Res*, *98*(2), 148-156. doi:10.1177/0022034518805739
- Handsley-Davis, M., Skelly, E., Johnson, N. W., Kapellas, K., Lalloo, R., Kroon, J., & Weyrich, L. S. (2021). Biocultural Drivers of Salivary Microbiota in Australian Aboriginal and Torres Strait Islander Children. *Front Oral Health*, *2*, 641328. doi:10.3389/froh.2021.641328
- Hernandez, B. Y., Zhu, X., Goodman, M. T., Gatewood, R., Mendiola, P., Quinata, K., & Paulino, Y. C. (2017). Betel nut chewing, oral premalignant lesions, and the oral microbiome. *PLoS One*, *12*(2), e0172196. doi:10.1371/journal.pone.0172196
- Hillmann, B., Al-Ghalith, G. A., Shields-Cutler, R. R., Zhu, Q., Gohl, D. M., Beckman, K. B., . . . Knights, D. (2018). Evaluating the Information Content of Shallow Shotgun Metagenomics. *mSystems*, *3*(6). doi:10.1128/mSystems.00069-18
- Horwood, P. F., Tarantola, A., Goarant, C., Matsui, M., Klement, E., Umezaki, M., . . . Greenhill, A. R. (2019). Health Challenges of the Pacific Region: Insights From History, Geography, Social Determinants, Genetics, and the Microbiome. *Front Immunol*, *10*, 2184. doi:10.3389/fimmu.2019.02184
- Hughes, J. B., Hellmann, J. J., Ricketts, T. H., & Bohannon, B. J. (2001). Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol*, *67*(10), 4399-4406. doi:10.1128/AEM.67.10.4399-4406.2001
- Human Microbiome Project, C. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207-214. doi:10.1038/nature11234
- Jacobs, G. S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C. C., Lawson, D. J., . . . Cox, M. P. (2019). Multiple Deeply Divergent Denisovan Ancestries in Papuans. *Cell*, *177*(4), 1010-1021 e1032. doi:10.1016/j.cell.2019.02.035
- Karmin, M., Flores, R., Saag, L., Hudjashov, G., Brucato, N., Crenna-Darusallam, C., . . . Cox, M. P. (2022). Episodes of Diversification and Isolation in Island Southeast Asian and Near Oceanian Male Lineages. *Mol Biol Evol*, *39*(3). doi:10.1093/molbev/msac045

- Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring beta diversity for presence-absence data. *Journal of Animal Ecology*, *72*(3), 367-382. doi:DOI 10.1046/j.1365-2656.2003.00710.x
- Lamont, R. J., Koo, H., & Hajishengallis, G. (2018). The oral microbiota: dynamic communities and host interactions. *Nat Rev Microbiol*, *16*(12), 745-759. doi:10.1038/s41579-018-0089-x
- Lassalle, F., Spagnoletti, M., Fumagalli, M., Shaw, L., Dyble, M., Walker, C., . . . Balloux, F. (2018). Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet. *Mol Ecol*, *27*(1), 182-195. doi:10.1111/mec.14435
- Leonard, D., Buttner, P., Thompson, F., Makrides, M., & McDermott, R. (2019). Anaemia in early childhood among Aboriginal and Torres Strait Islander children of Far North Queensland: a retrospective cohort study. *Aust N Z J Public Health*, *43*(4), 319-327. doi:10.1111/1753-6405.12911
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-3100. doi:10.1093/bioinformatics/bty191
- Li, J., Quinque, D., Horz, H. P., Li, M., Rzhetskaya, M., Raff, J. A., . . . Stoneking, M. (2014). Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. *BMC Microbiol*, *14*(1), 316. doi:10.1186/s12866-014-0316-1
- Lipson, M., Loh, P. R., Patterson, N., Moorjani, P., Ko, Y. C., Stoneking, M., . . . Reich, D. (2014). Reconstructing Austronesian population history in Island Southeast Asia. *Nat Commun*, *5*, 4689. doi:10.1038/ncomms5689
- Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J*, *5*(2), 169-172. doi:10.1038/ismej.2010.133
- Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*. doi:ARTNe10410.7717/peerj-cs.104
- Malaspinas, A. S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., . . . Willerslev, E. (2016). A genomic history of Aboriginal Australia. *Nature*, *538*(7624), 207-214. doi:10.1038/nature18299
- Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., . . . Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201-206. doi:10.1038/nature18964
- Manning, L., Laman, M., Rosanas-Urgell, A., Michon, P., Aipit, S., Bona, C., . . . Davis, T. M. (2012). Severe anemia in Papua New Guinean children from a malaria-endemic area: a case-control etiologic study. *PLoS Negl Trop Dis*, *6*(12), e1972. doi:10.1371/journal.pntd.0001972
- Martinez, I., Stegen, J. C., Maldonado-Gomez, M. X., Eren, A. M., Siba, P. M., Greenhill, A. R., & Walter, J. (2015). The gut microbiota of rural Papua New Guineans: composition, diversity patterns, and ecological processes. *Cell Rep*, *11*(4), 527-538. doi:10.1016/j.celrep.2015.03.049
- McLoughlin, K. S. (2011). Microarrays for pathogen detection and analysis. *Brief Funct Genomics*, *10*(6), 342-353. doi:10.1093/bfpg/blr027
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, *8*(4), e61217. doi:10.1371/journal.pone.0061217
- Miller, A. K., Westlake, C. S., Cross, K. L., Leigh, B. A., & Bordenstein, S. R. (2021). The microbiome impacts host hybridization and speciation. *PLoS Biol*, *19*(10), e3001417. doi:10.1371/journal.pbio.3001417
- Mousa, W. K., Chehadeh, F., & Husband, S. (2022). Recent Advances in Understanding the Structure and Function of the Human Microbiome. *Front Microbiol*, *13*, 825338. doi:10.3389/fmicb.2022.825338
- Nairz, M., & Weiss, G. (2020). Iron in infection and immunity. *Mol Aspects Med*, *75*, 100864. doi:10.1016/j.mam.2020.100864

- Nasidze, I., Li, J., Quinque, D., Tang, K., & Stoneking, M. (2009). Global diversity in the human salivary microbiome. *Genome Res*, *19*(4), 636-643. doi:10.1101/gr.084616.108
- Nasidze, I., Li, J., Schroeder, R., Creasey, J. L., Li, M., & Stoneking, M. (2011). High diversity of the saliva microbiome in Batwa Pygmies. *PLoS One*, *6*(8), e23352. doi:10.1371/journal.pone.0023352
- Parfrey, L. W., Moreau, C. S., & Russell, J. A. (2018). Introduction: The host-associated microbiome: Pattern, process and function. *Mol Ecol*, *27*(8), 1749-1765. doi:10.1111/mec.14706
- Pawley, A., Attenborough, R., Golson, J et al. (2005). Papuan Pasts: cultural, linguistic and biological histories of the Papuan-speaking peoples. *Pacific Linguistics*, 555-584. doi:10.15144/PL-572
- Pedro, N., Brucato, N., Fernandes, V., Andre, M., Saag, L., Pomat, W., . . . Ricaut, F. X. (2020). Papuan mitochondrial genomes and the settlement of Sahul. *J Hum Genet*, *65*(10), 875-887. doi:10.1038/s10038-020-0781-3
- Pedro, N., Pinto, R. J., Cavadas, B., & Pereira, L. (2021). Sub-Saharan African information potential to unveil adaptations to infectious disease. *Hum Mol Genet*, *30*(R1), R138-R145. doi:10.1093/hmg/ddab001
- Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun*, *469*(4), 967-977. doi:10.1016/j.bbrc.2015.12.083
- Raymond, K. N., Dertz, E. A., & Kim, S. S. (2003). Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci U S A*, *100*(7), 3584-3588. doi:10.1073/pnas.0630018100
- Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., . . . Stoneking, M. (2011). Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am J Hum Genet*, *89*(4), 516-528. doi:10.1016/j.ajhg.2011.09.005
- Riley, I. (2009). *Demography and the Epidemiology of Disease in Papua New Guinea* (Vol. 52): Papua New Guinea Institute of Medical Research.
- Roscoe, P. (2005). Foraging, ethnographic analogy and Papuan pasts: Contemporary models for the Sepik-Ramu past. *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*, 555-584.
- RStudio Team. (2020). RStudio: Integrated Development for R. Boston, MA. Retrieved from <http://www.rstudio.com/>.
- Samuels, D. C., Han, L., Li, J., Quanguo, S., Clark, T. A., Shyr, Y., & Guo, Y. (2013). Finding the lost treasures in exome sequencing data. *Trends Genet*, *29*(10), 593-599. doi:10.1016/j.tig.2013.07.006
- Schapper, A. (2017). Farming and the trans-new Guinea family: a consideration chapter. In M. R. a. A. Saveliev (Ed.), *Language Dispersal Beyond Farming* (pp. 155–181). Amsterdam: John Benjamins Publishing Company.
- Schmidt, E., & Fang, P. (2021). Papua New Guinea agri-food trade and household consumption trends point towards dietary change and increased overweight and obesity prevalence. *Global Health*, *17*(1), 135. doi:10.1186/s12992-021-00787-0
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol*, *12*(6), R60. doi:10.1186/gb-2011-12-6-r60
- Thissen, J. B., McLoughlin, K., Gardner, S., Gu, P., Mabery, S., Slezak, T., & Jaing, C. (2014). Analysis of sensitivity and rapid hybridization of a multiplexed Microbial Detection Microarray. *J Virol Methods*, *201*, 73-78. doi:10.1016/j.jviromet.2014.01.024
- Uranga, C. C., Arroyo, P., Jr., Duggan, B. M., Gerwick, W. H., & Edlund, A. (2020). Commensal Oral *Rothia mucilaginosa* Produces Enterobactin, a Metal-Chelating Siderophore. *mSystems*, *5*(2), e00161-00120. doi:10.1128/mSystems.00161-20

- Vangay, P., Johnson, A. J., Ward, T. L., Al-Ghalith, G. A., Shields-Cutler, R. R., Hillmann, B. M., . . . Knights, D. (2018). US Immigration Westernizes the Human Gut Microbiome. *Cell*, *175*(4), 962-972 e910. doi:10.1016/j.cell.2018.10.029
- Wang, L., & Ganly, I. (2014). The oral microbiome and oral cancer. *Clin Lab Med*, *34*(4), 711-719. doi:10.1016/j.cll.2014.08.004
- Weyrich, L. S. (2021). The evolutionary history of the human oral microbiota and its implications for modern health. *Periodontol 2000*, *85*(1), 90-100. doi:10.1111/prd.12353
- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol*, *20*(1), 257. doi:10.1186/s13059-019-1891-0
- Xi, R., Wang, R., Wang, Y., Xiang, Z., Su, Z., Cao, Z., . . . Li, J. (2019). Comparative analysis of the oral microbiota between iron-deficiency anaemia (IDA) patients and healthy individuals by high-throughput sequencing. *BMC Oral Health*, *19*(1), 255. doi:10.1186/s12903-019-0947-6
- Zhong, X., Lu, Q., Zhang, Q., He, Y., Wei, W., & Wang, Y. (2021). Oral microbiota alteration associated with oral cancer and areca chewing. *Oral Dis*, *27*(2), 226-239. doi:10.1111/odi.13545

### **Data accessibility**

The MiDAS output files obtained from the Array methodology and the KRAKEN output files obtained from the WGS unmapped reads are available as supplementary files. The reads mapping to the microbial community (after removal of human reads) were submitted to ENA ([www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena)), for the study number ERP139829 under the BioSample Accession numbers ERS12516930 – ERS12516994.

### **Benefit-sharing statement**

Benefits Generated: Benefits from this research accrue from the sharing of our data and results on public databases as described above.

### **Author contributions**

N.B., L.P. and F-X.R. designed the study; C.K. and M.L. performed sample collection; N.P., N.B. and V.L. performed laboratory procedures; N.P., B.C and R.C. performed the bioinformatics and computational analyses; N.P., N.B., L.P. and F-X.R. wrote the manuscript; all authors read and reviewed the manuscript.

### Microbiome from Array

**Sample preparation** DNA extracted from saliva samples (DNA genotek)

**Sample Analysis** Microbiome characterization with the Axiom Microbiome Array (Thermo Fischer Scientific)

↓ .CEL files

**Data Analysis** Axiom Microbial Detection Analysis Software (MiDAS)

**Output** Qualitative results only  
Presence/Absence of species

### Microbiome from WGS

**Sample preparation** DNA extracted from saliva samples (DNA genotek)

**Sample Analysis** WGS using Illumina XTen shotgun sequencing

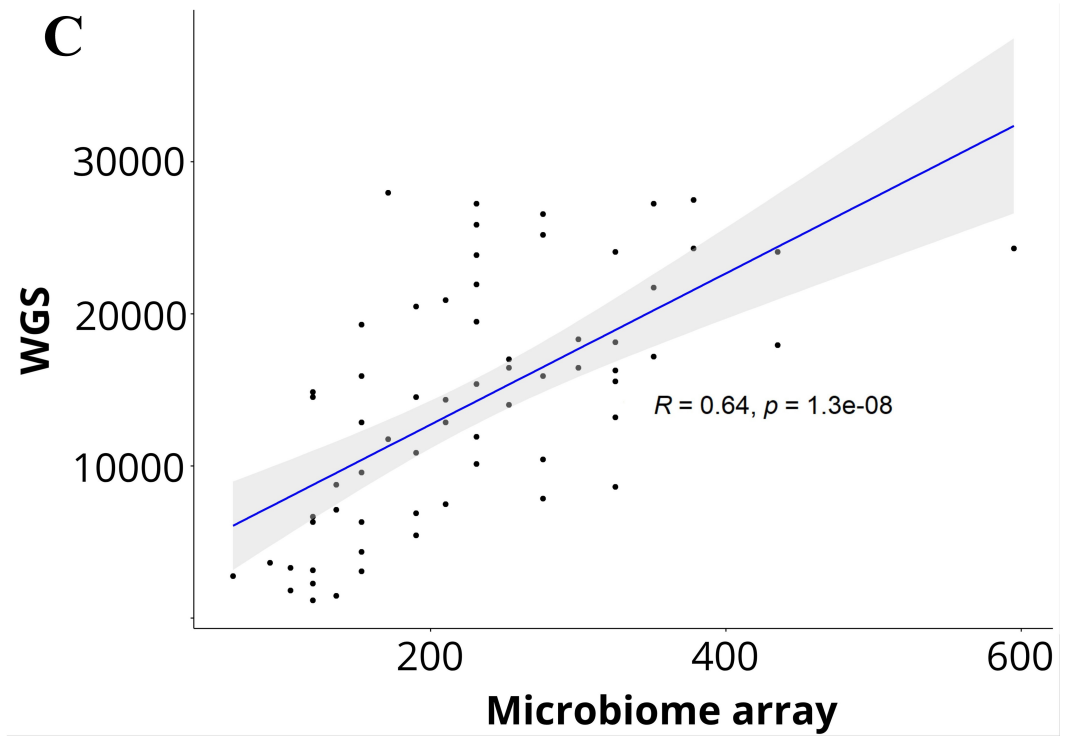
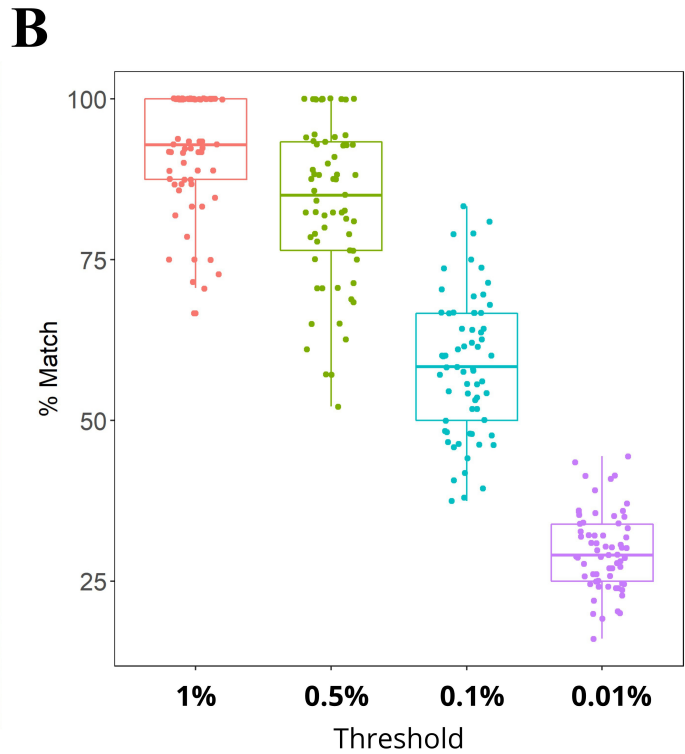
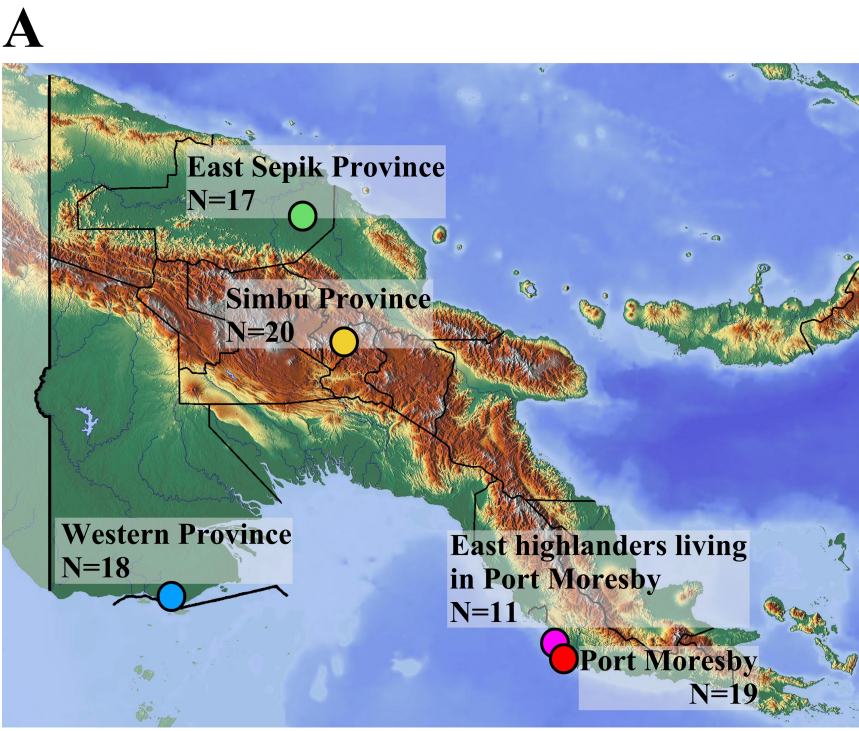
↓ .fastq files

**Data Analysis** Alignment against human reference sequence using minimap2-2.23  
▼  
Recovery of off-target reads  
▼  
Kraken - metagenomic sequence classification  
Bracken - estimating species abundance

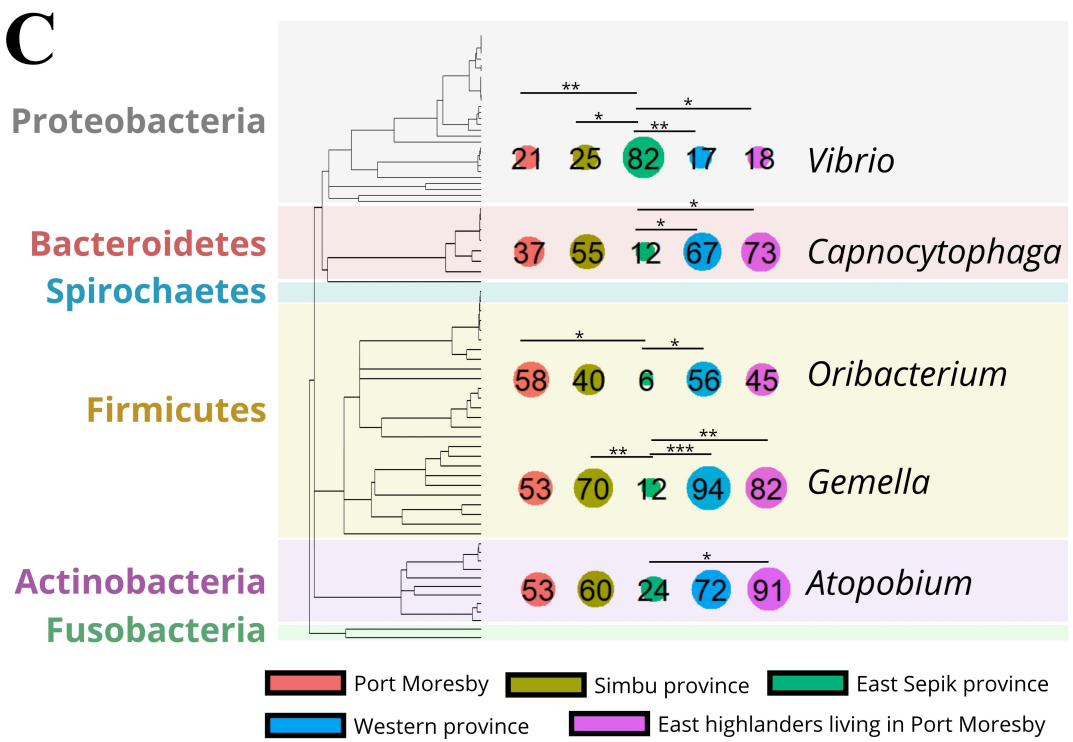
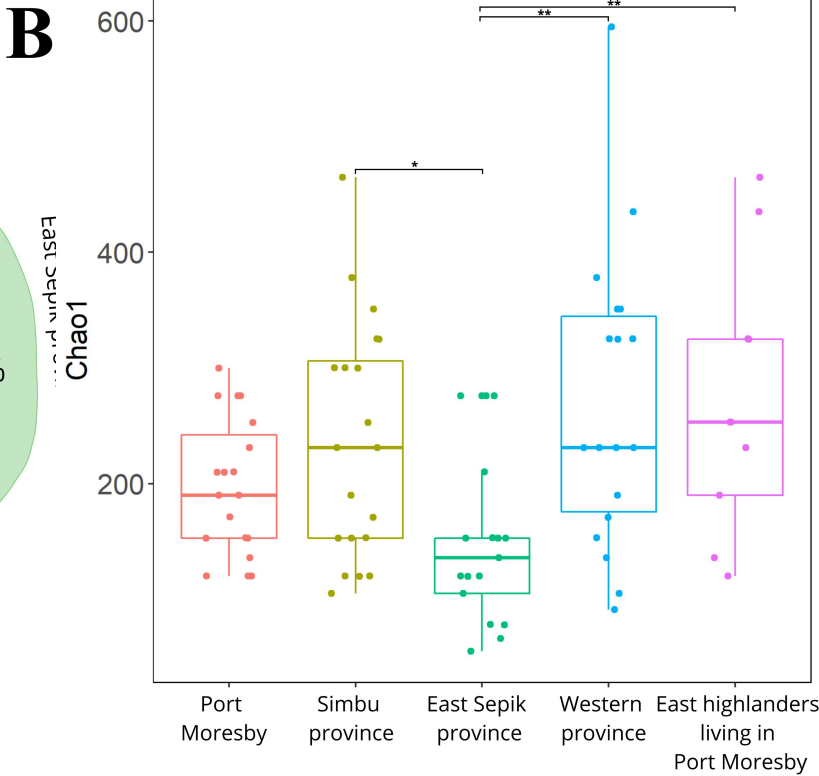
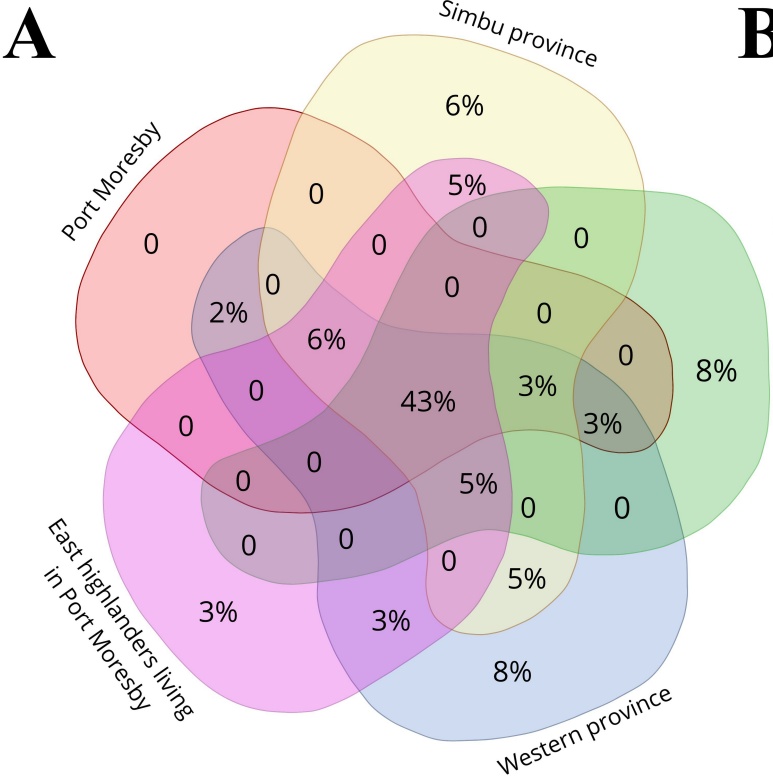
**Output** Qualitative and Quantitative results  
Presence/Absence and Abundance of species

MEC\_16702\_Figure 1.jpg

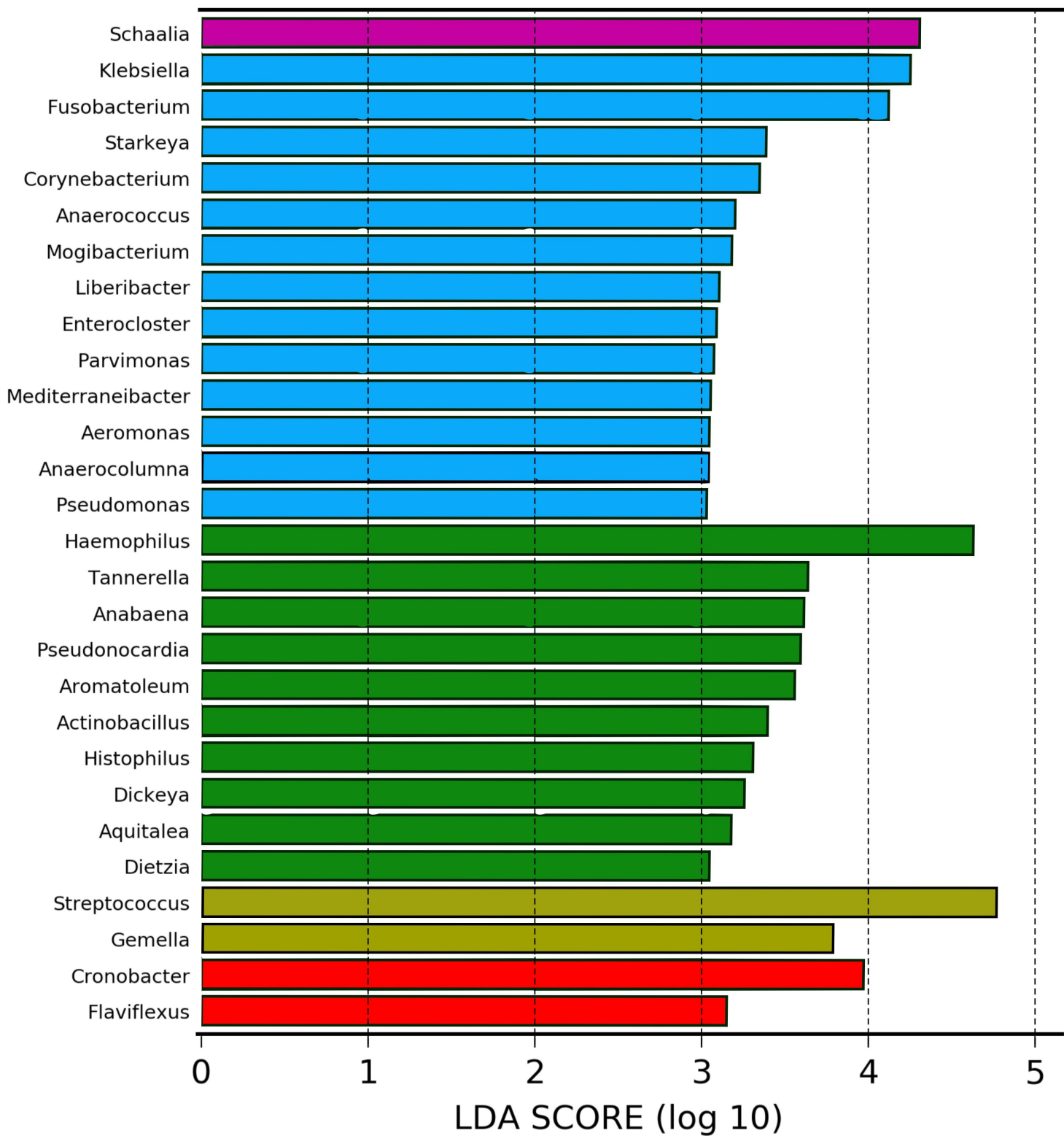
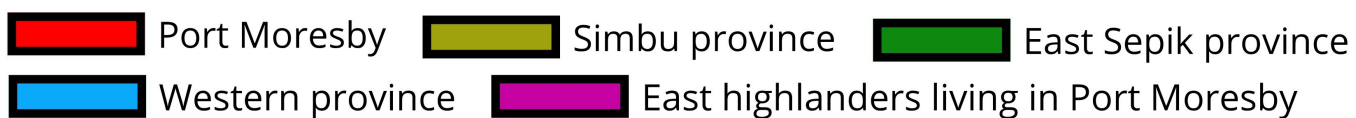




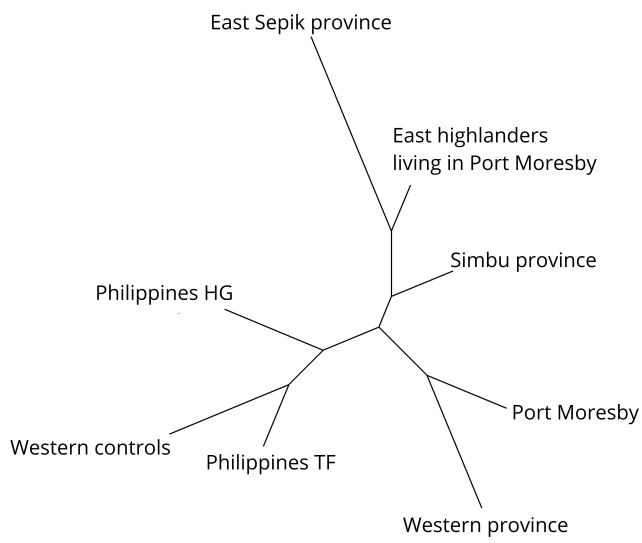
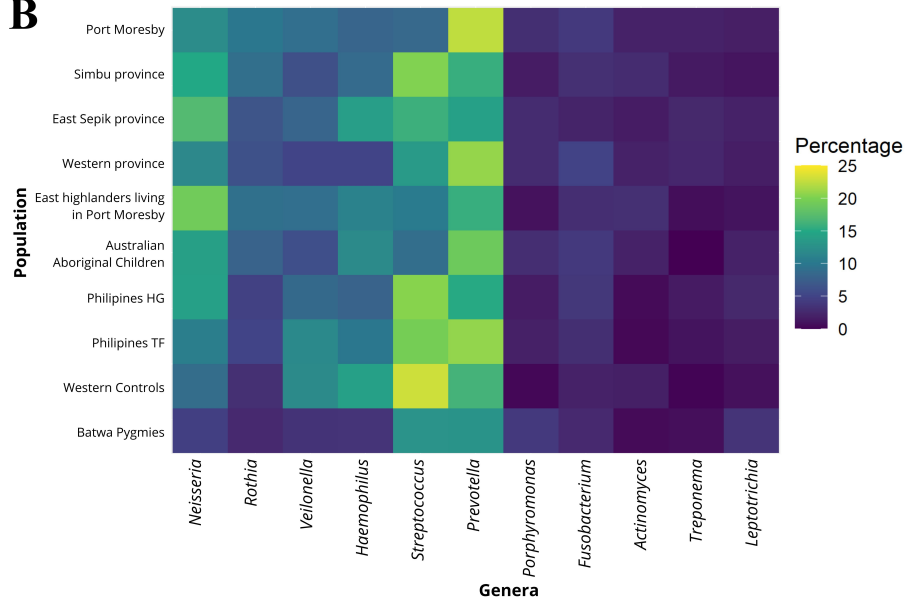
MEC\_16702\_Figure 2.jpg



MEC\_16702\_Figure 3.jpg



MEC\_16702\_Figure 4.jpg

**A****B**

MEC\_16702\_Figure 5.jpg