The making of the oral microbiome in Agta hunter–gatherers

Begoña Dobon[1,2][†] [ID], Federico Musciotto[1,3][†], Alex Mira[4,5], Michael Greenacre[6,7],
Rodolph Schlaepfer[1], Gabriela Aguileta[2], Leonora H. Astete[8], Marilyn Ngales[8], Vito Latora[9,10,11],
Federico Battiston[1,12], Lucio Vinicius[1,13][‡] [ID], Andrea B. Migliano[1,13][*][‡] [ID] and Jaume Bertranpetit[2][*][‡]

[1]Department of Anthropology, University of Zurich, Switzerland, [2]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain, [3]Dipartimento di Fisica e Chimica, Università di Palermo, Italy, [4]Department of Health and Genomics, Center for Advanced Research in Public Health, FISABIO Foundation, Valencia, Spain, [5]CIBER Center for Epidemiology and Public Health, Madrid, Spain, [6]Department of Economics and Business, Universitat Pompeu Fabra and Barcelona Graduate School of Economics, Barcelona, Spain, [7]Faculty of Biosciences, Fisheries and Economics, University of Tromsø, Norway, [8]Lyceum of the Philippines University, Intramuros, Manila, Philippines, [9]School of Mathematical Sciences, Queen Mary University of London, UK, [10]Dipartimento di Fisica ed Astronomia, Università di Catania and INFN, Catania, Italy, [11]Complexity Science Hub Vienna, Vienna, Austria, [12]Department of Network and Data Science, Central European University, Vienna 1100, Austria and [13]Department of Anthropology, University College London, UK
*Corresponding author. E-mail: jaume.bertranpetit@upf.edu, andrea.migliano@uzh.ch

Abstract

Ecological and genetic factors have influenced the composition of the human microbiome during our evolutionary history. We analysed the oral microbiota of the Agta, a hunter–gatherer population where some members have adopted an agricultural diet. We show that age is the strongest factor modulating the microbiome, probably through immunosenescence since we identified an increase in the number of species classified as pathogens with age. We also characterised biological and cultural processes generating sexual dimorphism in the oral microbiome. A small subset of oral bacteria is influenced by the host genome, linking host collagen genes to bacterial biofilm formation. Our data also suggest that shifting from a fish/meat diet to a rice-rich diet transforms their microbiome, mirroring the Neolithic transition. All of these factors have implications in the epidemiology of oral diseases. Thus, the human oral microbiome is multifactorial and shaped by various ecological and social factors that modify the oral environment.

Keywords: hunter–gatherers; oral microbiome; Neolithic transition; pathogen transmission

Social media summary: Our study shows how ecology, genetics and development shape the oral microbiome of Agta hunter–gatherers.

Introduction

The composition and diversity of the human oral microbiota have been influenced by several factors during our evolutionary history (Cornejo Ulloa et al., 2019; Weyrich, 2021). Some are intrinsic biological characteristics of the host, such as age, sex and genetic composition, while others are external factors (Gomez et al., 2017; Willis et al., 2018), including diet, drinking water sources, oral hygiene, lifestyle and social interactions. Such factors modulate the physiological conditions of the oral cavity

---

[†]These authors contributed equally to this work.
[‡]These authors contributed equally to this work.

and affect the composition and diversity of oral microbiota. Many studies performed in farming and urban populations have indicated that while the oral microbiome is highly diverse between individuals, it remains stable over time (Costello et al., 2009). However, less is known about how the composition of the oral microbiome is modulated in hunter–gatherer populations, where the fully mature oral biofilm microbiome can be studied without the confounding effects of tooth brushing or professional dental cleaning (Velsko et al., 2019).

To investigate the multiple ecological and genetic factors shaping the human oral microbiome, we have analysed the oral microbiome of the Agta hunter–gatherers from the Philippines. The Agta were predominantly hunter–gatherers (fishing, hunting and gathering) when data collection took place (Page et al., 2018), and while their main source of animal protein was obtained by riverine and marine spearfishing or hunting, other activities such as inter-tidal foraging, wild food gathering, low-intensity cultivation, wage labour and trade complement their economy (Minter, 2010). There is high variability in the amount of hunting, gathering and sea foraging products traded for rice and other items (such as tobacco) with farming neighbours (Page et al., 2018). As a result, the Agta lifestyle ranges from completely mobile foragers with a protein-rich diet, to settled low-intensity farmers with a rice-rich diet (Page et al., 2016; Dyble et al., 2019).

To detect fine-scale variation in the oral microbiome of Agta hunter–gatherers, we collected saliva samples from 138 Agta aged 5–65 years (see Supplementary Table S1), sequenced the 16S rRNA region of oral bacteria and identified 5430 amplicon sequence variants (ASVs; see Callahan et al., 2017) belonging to 110 genera. Since 16S rRNA sequencing only provides relative abundance data, we refrained from making any inferences based on total yield. To study genetic host factors associated with microbiome composition, we also genotyped all individuals with the Axiom Genome-wide Human Origins array. We combined this information with additional individual data on household composition, age, sex and diet, measured as the proportion of meals including meat/fish (animal protein), and the proportion of meals that consisted exclusively of agricultural products (rice) (Supplementary Figure S1). Based on this rich dataset, we attempted to determine the contributions of age, sex, diet and host genetics in the making of the Agta oral microbiome.

Results

Factors influencing Agta oral microbiome composition

The Agta oral microbiome is mostly composed of *Firmicutes* (mean relative abundance = 33.1%, SD = 11.4), *Proteobacteria* ($27 \pm 14.6\%$), *Actinobacteria* ($15.5 \pm 8.3\%$) and *Bacteroidetes* ($14.5 \pm 7.7\%$) (Supplementary Figure S2). To compare and identify the main ecological and social factors contributing to microbiome variation, we performed a constrained log ratio analysis (LRA; see Greenacre, 2018) on relative bacterial genus abundance. Age marginally explains 7.2% of the total log ratio variance ($p < 0.0001$, based on 9999 permutations), sex explains 2.2% ($p = 0.018$) and diet 3.6% ($p = 0.015$). Altogether they explain 13% of the total log ratio variance. We also applied a bipartite stochastic block model (biSBM) approach (Larremore et al., 2014) at the ASV level, where we assigned each bacterium to an individual, and then clustered individuals according to the bacteria they have in common. We restricted the analysis to the core measurable microbiota (CMM), that we define as ASVs present in at least 10% of the Agta to reduce random errors owing to low-prevalence taxa (Supplementary Figure S3). The best model produced two clusters of people and three clusters of bacteria (Figure 1a). While we did not find differences in diet or proportions of sexes between the clusters of individuals (Supplementary Figure S4), they strongly differ in their age distribution and define an adult cluster (mean age = 38 years old) and a young cluster (mean age = 18 years old) (Welch *t*-test, $t = 5.78$, d.f. $= 71.35$, $p < 0.0001$), with 55.48% of ASVs in the CMM being more associated with one of the clusters of individuals. The strong effect of age cannot be attributed to batch effects, as we found no association either between age and sampling time (Pearson $r = -0.1452$, $t = -1.55$, $p = 0.1232$), or between age and source (camp) of sampled water (Pearson $r = 0.05$, $t = 0.58$, $p = 0.5618$; see Methods for details of data collection). Thus, while age, diet and sex influence the composition of
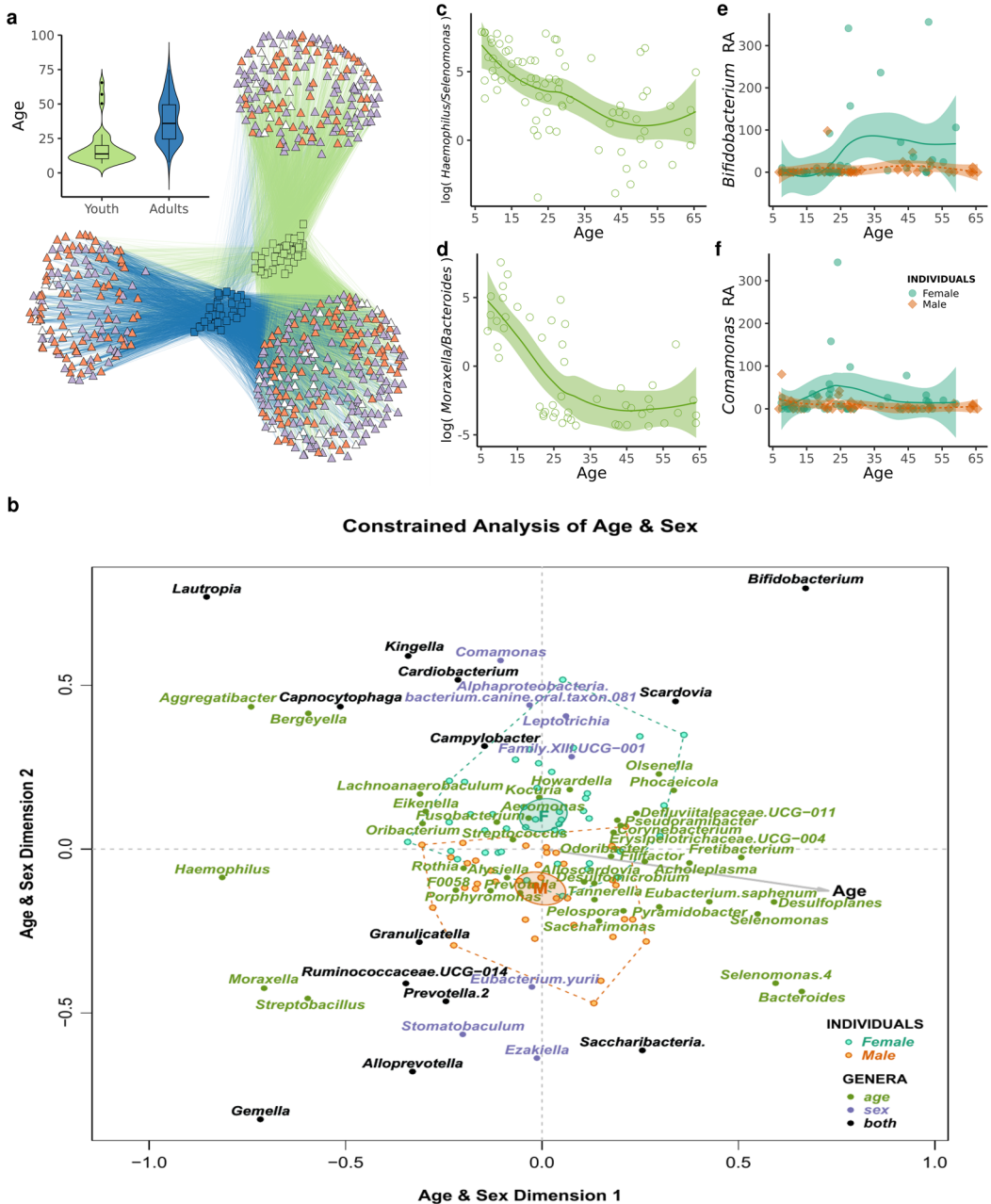
Figure 1. Age- and sex-related effects in the hunter–gatherer oral microbiome. (a) Network representation of the hunter–gatherer core measurable microbiota (CMM). Amplicon sequence variants (ASVs; triangles) are colour-coded as putatively pathogenic (purple), non-pathogenic (orange) or unclassified (white). Inset shows age distribution for the two clusters of individuals (squares). (b) Log ratio analysis constrained to age and sex differences on the bacterial composition at genus level. The effects of diet were partialled out. Only genera statistically significant in at least 20 (for age) or 10 (for sex) log ratios are displayed (*p*-value < 0.05 after Benjamini–Hochberg correction). Dashed lines enclose all individuals (dots) within a sex, with 95% confidence ellipses for their means. Taxa are colour-coded depending on the associated variable: age, sex or both. The starting point of the grey arrow indicates the mean age of the population (30 years old). Log ratio of (c) *Haemophilus* and *Selenomonas* relative abundance and (d) *Moraxella* and *Bacteroides* according to age. Line and shaded area indicate the 95% confidence interval of the mean. Relative abundance of (e) *Bifidobacterium* and (f) *Comamonas* according to age and sex. Lines and shaded areas indicate the 95% confidence interval of the mean for each sex.

Agta microbiome, the biSBM singles out age as the main modulator of the hunter–gatherer core oral microbiome.

Old age is associated with increased frequency of species classified as oral pathogens

To investigate the independent effects of ageing on the oral microbiome, we performed a LRA constrained to age and sex after partialling out the effects of diet. The resulting ordination shows that the effects of age and sex are mostly independent, with only a few genera being affected by both variables (Supplementary Figure S5). As expected, the first dimension is associated with age, while the second dimension splits individuals according to sex (Figure 1b).

There is a clear change in the composition and frequency of certain bacteria with age (Figure 1c and d). At a young age, we observe a higher relative abundance of organisms that typically live in mucosa, such as *Haemophilus* and *Moraxella*, which infect the upper and lower respiratory tract but are detected in the oral cavity and saliva as their vehicles of transmission. Other genera found at younger ages include bacteria normally associated with good oral health, such as *Bergeyella* and *Rothia* (Rosier et al., 2020). However, at older ages we observe a marked decline in the relative abundance of those genera and an increase in important pathogens related to periodontitis including the 'red complex' periodontal pathogen *Tannerella*, as well as other periodontitis-related bacteria (*Filifactor*, *Fretibacterium*, *Saccharimonas*, *Selenomonas* and *Phocaeicola*), consistent with a higher incidence of the condition with older age (Kassebaum et al., 2014). We also found organisms associated with cavities (*Olsenella*), dental plaque and dental calculus formation (*Corynebacterium*), pulmonary infections, sepsis or bacteremia, and chronic diseases (*Acholeplasma*) (see Methods for in-depth bacteria pathogenic classification). Another sign of ageing was the presence in the oral cavity of gut bacteria (*Bacteroides*), indicating a potential age-related decline in immunological function and filtering (De Maeyer & Chambers, 2021). However, such changes are not associated with a decrease in the alpha-diversity of the total oral microbiome as measured by the number of bacteria observed or their phylogenetic complexity (Supplementary Figure S6a and b), suggesting that the overall effect of ageing is a replacement of protective and commensal bacteria by putatively pathogenic ones. This is supported by an increase in the number of potential pathogenic bacteria in the CCM in bacterial clusters associated with older ages (Fisher exact test, $p < 0.001$) (Figure 1a).

Sex differences shape composition but not diversity of the Agta oral microbiome

We found no differences in alpha-diversity in the Agta oral microbiome between males and females (Supplementary Figure S6c and d). It is possible that the high levels of equality in dietary composition, social interactions and mobility between men and women in the Agta society (Dyble et al., 2016; Migliano et al., 2017) may be associated with similarities in microbiome diversity between sexes. Nevertheless, the LRA constrained to age and sex shows sex-related differences in the composition of the oral microbiome (Figure 1b). For example, *Stomatobaculum* and *Eubacterium yurii* present in the oral cavity of smokers (Duan et al., 2017) are associated with males, consistent with Agta men chewing tobacco more frequently than women. It should be noticed that betelnut chewing is a widespread habit among the Agta, a practice linked in other populations to oral carcinogenesis (Oslam et al., 2019) and changes in the oral microbiome (Hernandez et al., 2017). Future studies should investigate potential links between betelnut chewing and oral microbiomes in Agta. It is also interesting to mention *Comamonas* (Figure 1f). Even if it has been reported as a possible contaminant in microbiome studies (Einsenhofer et al., 2019), its higher relative abundance in females could be related to the fact that it can degrade progesterone (Liu et al., 2013). This bacterium has been found in subgingival samples, where female hormones could be present either in saliva or in the gingival crevicular fluid.

Age and sex interactions in microbiome composition

Some bacteria are significantly associated with both age- and sex-related differences, such as *Gemella*, which is a prevalent inhabitant of the respiratory mucosa besides *Haemophilus* and

Moraxella, supporting the idea that mucosa-associated or respiratory-tract organisms are more frequently acquired in young individuals, especially males. At older ages, the *Bifidobacterium/Saccharibacteria* ratio differentiates by sex: while *Bifidobacterium* is associated with females, the periodontal pathogen *Saccharibacteria* is associated with males. Thus, the observed trend of an increase in periodontal pathogens with age is stronger in males, in line with the global epidemiology of the disease (Shiau & Reynolds, 2010; Eke et al., 2015). On the other hand, we found an increase in the relative abundance of caries-related pathogens *Scardovia* and *Bifidobacterium* associated with reproductive age females. Caries incidence increases with age and is more prevalent in females (Ferraro & Vieira, 2010), with a more saccharolytic or acidic salivary environment in older women, together with hormonal fluctuations and lower salivary flow (Zaura et al., 2017), possibly facilitating their proliferation. The strong association of *Bifidobacterium* with adult females could also be explained by its presence in breast milk (Fernández et al., 2020). Its proliferation (Figure 1d) also coincides with the start of the reproductive age and increase in childcare (Dyble et al., 2019).

Effect of rice consumption

While the impact of diet on gut microbiomes has been established (Turnbaugh et al., 2009; David et al., 2014; Schnorr et al., 2014; Smits et al., 2017), its role in the oral microbiome is still unclear. Some studies have found little or no effect, whereas others have found associations with specific nutrients (De Filippis et al., 2014; Belstrøm et al., 2014; Weyrich, 2021; Zaura et al., 2017). Variation in rice consumption in our sample allowed us to assess the effects of a hunter–gatherer diet and of the recent introduction of farming products on the Agta oral microbiome. We performed an LRA on relative bacterial genus abundance after partialing out the effects of age and gender (Figure 2). The first dimension of the ordination is related to the transition from a diet where all meals include meat to where most consist of only rice. Agta following a meat-rich diet showed a higher relative abundance of *Actinobacillus*, *Alphaproteobacteria* and *Streptobacillus* and lower abundance of *Selenomonas*, *Atopobium*, *Peptoanaerobacter* and *Pyramidobacter*. The higher relative abundance of *Actinobacillus* in individuals ingesting a protein-rich diet is particularly interesting, given the extraordinary proteolytic potential of *A. actinomycetemcomitans*, a well-known oral pathogen with destructive effects in the gingival tissue and in aggressive forms of periodontitis (Fives-Taylor et al., 1999). At the other extreme, in individuals with a rice-rich diet, there is an increase in the abundance of the highly saccharolytic dental caries pathogen *Scardovia*, of *Treponema*, of gut organisms like *Butyrivibrio* and *Erysipelotrichaceae*, and of *Eggerthia*, a rare organism isolated from dental abscesses. We also ranked ASVs based on whether they are more present than expected in individuals with a high or low proportion of meals with only rice or with meat and fish. We found that the scores associating bacterial species with either rice or meat and fish are negatively correlated (Spearman's $\rho = -0.47$, $p < 0.0001$). This fits with a general separation of oral microorganisms into saccharolytic (caries-related, acidogenic and acidophilic) and proteolytic (gum-disease and halitosis related, alkalophilic and NH_4 generators), as suggested in a metabolome-based study (Zaura et al., 2017). Our results suggest that more settled Agta, who consume more rice, experience a decline in oral health, mirroring a general pattern of health decline owing to a Neolithic-like diet and a more farming-derived lifestyle (Page et al., 2016; Adler et al., 2013; Sabbatani & Fiorino, 2016).

## Pathogenic oral bacteria are associated with host collagen genes

The interaction between host genetic makeup and microbiome composition differs across body sites (Kolde et al., 2018; Blekhman et al., 2015), and seems especially weak in the oral cavity (Shaw et al., 2017), making it difficult to assess the co-evolution of our genome and the oral microbiome. We therefore performed a genome-wide association study (GWAS) using a mixed model approach in a hunter–gatherer population without the confounding influence of antibiotics or brushing. We treated the relative abundance of each bacterium as an independent trait, adding age, sex and household as
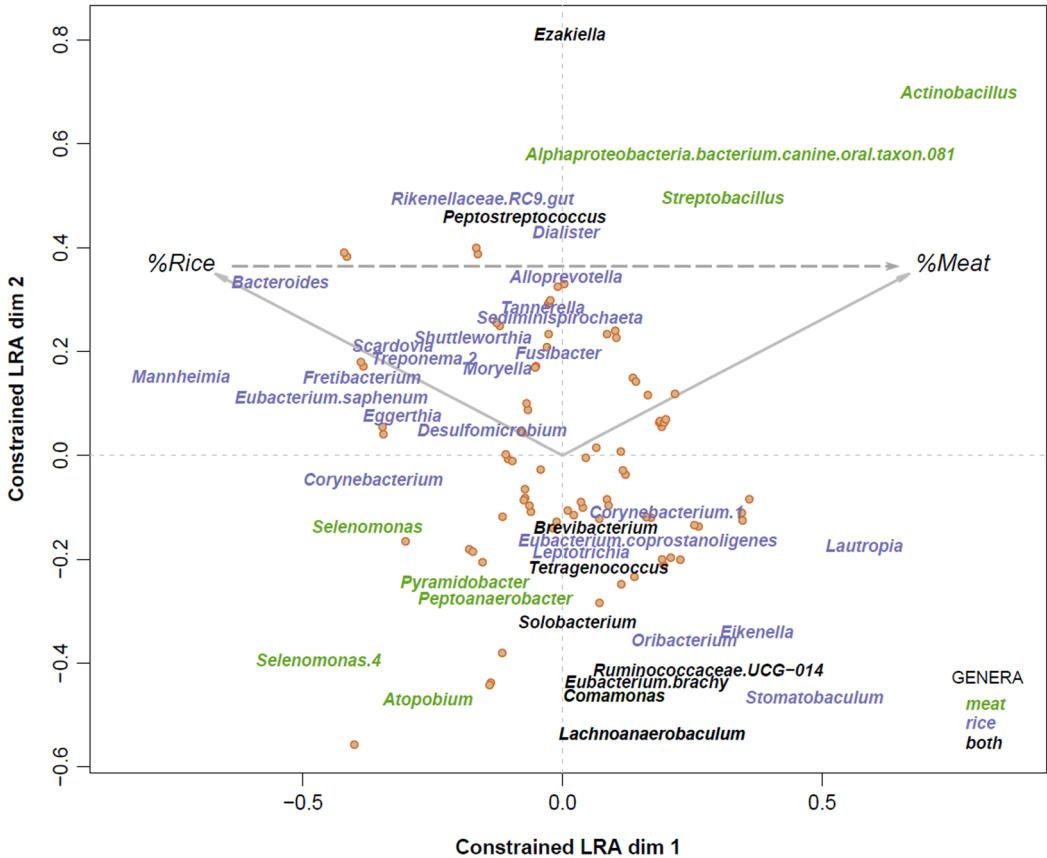
Figure 2. Effect of diet on the oral microbiome in the Agta. Log ratio analysis constrained to diet differences on bacterial composition at genus level. The effects of age and sex were partialled out. Only genera statistically significant in more than five (for rice) or three (for meat) log ratios are displayed (*p*-value < 0.05 after Benjamini–Hochberg correction). Taxa are colour-coded based on the variable they are associated with proportion of meals with meat (%Meat), proportion of meals with only rice (%Rice) or both. The original plot was slightly rotated without any change in explained variance, so that the dashed vector indicating the difference between %Meat and %Rice was horizontal.

covariates and kinship as a random effect. Household membership was used as proxy for the strength of social interactions between individuals, as social interactions predict microbiome sharing; for an extended analysis of the effects of sociality on the oral microbiome of the Agta population, see the companion article by Musciotto et al. (2023) in this issue.

Analyses were performed using the CMM, and then using 92 genera present in at least 10 Agta. All bacteria identified in the Agta (Supplementary Table S2 and Supplementary Figure S7) overlap with those of other oral microbiome GWAS (Blekhman et al., 2015; Gomez et al., 2017; Kolde et al., 2018), pointing to a small subset of oral bacteria influenced by the human genome (Figure 3). A pathway enrichment analysis linked this subset to several biological host functions (body fat metabolism, wound healing and collagen trimmers; Supplementary Table S3). Of relevance is an association between the bacterial genera classified as potentially pathogenic (*Aggregatibacter* and *Selenomonas*) with genetic variation in collagen genes. The ability to bind collagen is a vital feature in the oral cavity, as many oral bacteria require collagen-binding proteins to attach to oral tissues (Mira et al., 2017) suggesting a genetic basis for the predisposition to bacterial biofilm formation. We further tested whether we could detect signatures of recent positive selection in the host genomic regions associated with the oral microbiome, but found no signal.
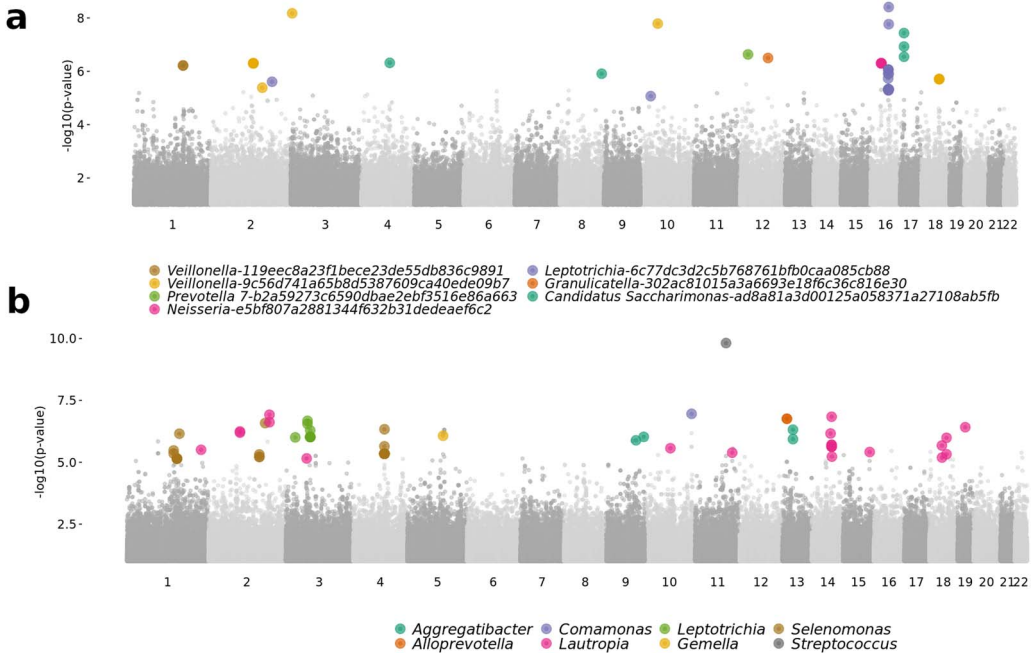
Figure 3. Genome-wide association study of bacterial abundance. Aggregated Manhattan plot of the GWAS results of (a) seven ASV and (b) eight genera with non-zero PVE ('chip heritability') estimates with at least one significant genetic association. Each dot is a single nucleotide polymorphism (SNP), and significant SNP–bacteria associations ($q < 0.1$) are colour-coded.

Discussion

The Agta oral microbiome is influenced by external factors such as social interactions (Musciotto et al., 2023), as well as intrinsic and ecological factors including age, sex, diet and host genetics. Among the latter we identified age as the factor with the strongest effect on microbiome composition, with commensal or beneficial microbiota being replaced by potentially pathogenic ones with ageing. The proliferation of oral bacteria likely to be pathogenic exhibits sexual dimorphism, with caries-related (in females) and periodontitis-related (in males) bacteria increasing with age, possibly associated with immunosenescence (Preshaw et al., 2017) and a sex-specific oral environment resulting from bio-logical and cultural factors. The increase in farming-derived novel foods such as rice also seems to have an effect on microbiome composition and host health. Overall, the relatively small subset of bacteria linked to the host genome suggests that the Agta oral microbiome is mainly affected by environmental (diet) and intrinsic factors (age), with little influence of individual host genetic variation (Mukherjee et al., 2021; see our Supplementary Figure S5). We conclude that the Agta Palanan oral microbiome is multifactorial with distinct subsets of bacteria shaped by specific ecological and social factors, reflecting multiple adaptations in the domains of life history, sociality and diet.

Methods

Ethics approval

This study was approved by the UCL Ethics Committee (UCL ethics code 3086/003) and carried out with permission from local government in the Philippines and Agta community members in Palanan. Informed consent was obtained from all participants, after group and individual explanation of research objectives in the indigenous language. A small compensation (usually a thermal bottle or cooking utensils) was given to each participant. The National Commission for Indigenous Peoples (NCIP), advised us that the process of Free Prior Informed Consent had to be obtained from

community leaders, youth and elders under the supervision and validation of the NCIP. This was done in 2017 with the presence of all community leaders, elders and youth representatives at the NCIP regional office, with the mediation of the regional officer and the NCIP Attorney. The validation process was approved unanimously by the community leaders and the NCIP, and validated the full five years of prior data collection.

Saliva sample collection

Saliva samples from 155 Palanan Agta were collected over two field seasons: April–June 2013 and February–October 2014 during the dry season. For comparative genetic studies we also used saliva samples from 21 Mbendjele BaYaka, an African hunter–gatherer population, collected in 2014, and 14 Palanan farmers collected in 2007–2009. In all cases saliva was collected using the Oragene·DNA/saliva kit and participants were asked to rinse their mouth with water and spit into a vial until half full. After collection and transportation, saliva samples were stored at the UCL Department of Anthropology, London, UK at −20°C.

Microbial DNA extraction and 16S rRNA gene sequencing

DNA was extracted from 155 Agta saliva samples following the protocol for manual purification of DNA for Oragene·DNA/saliva sample. The 16S rRNA gene V3-V4 region was amplified by PCR with primers containing Illumina adapter overhang nucleotide sequences following the 16S Metagenomic Sequencing Library Preparation protocol for the Illumina MiSeq System. All PCR products were validated through an agarose gel and purified with magnetic beads. Index PCR was then performed to create the final library also validated through an agarose gel. All samples were pooled together at equimolar proportions and the final pool was qPCR quantified prior to the MiSeq loading. Raw Illumina paired-end sequence data were demultiplexed, quality filtered and denoised with QIIME 2 2019.1 (Bolyen et al., 2019) and DADA2 (Callahan et al., 2016). DADA2 generates single nucleotide exact amplicon sequence variants (ASVs). ASVs are biological meaningful entities as they identify a specific DNA sequence and allow for higher resolution than using operational taxonomic units (see Callahan et al., 2017). Taxonomic information was assigned to ASVs using a naive Bayes taxonomy classifier against SILVA database release 132 with a 99% identity sequence (Quast et al., 2013). ASVs that did not belong to the kingdom Bacteria were classified as mitochondrial or chloroplast sequences, and samples with an extremely low number of sequences (8000) were excluded from further analyses. ASVs were aligned with MAFFT (Katoh, 2002) and a rooted phylogenetic tree was constructed with FastTree2 (Price et al., 2010) using default settings via QIIME 2. This resulted in a total of 5430 ASVs and 138 Agta (67 women and 71 men). We generated a rarefaction curve with R package *vegan* version 2.5-7 (Oksanen et al., 2020) to determine that the richness of samples had been fully observed (Supplementary Figure S8). The number of observed ASVs and Shannon Diversity indexes were calculated with R package Phyloseq version 1.30.0 (McMurdie & Holmes, 2013). Faith's Phylogenetic Diversity index (Faith, 1992) was calculated with R package *picante* version 1.8.2 (Kembel et al., 2010) using the rooted phylogenetic tree generated in R (R Core Team, 2020). To determine the set of microbial traits to be included in the analyses, we selected ASVs with at least 10 reads in at least two individuals ($n = 1980$), then aggregated those with a taxonomic assignment at a genus level, resulting in 110 genera. At the ASV level we also defined a CMM, consisting of ASVs that appeared in at least 10% of the Agta individuals (14 or more) resulting in 575 ASVs (out of 1980) representing 90% of the composition of the Agta oral microbiome.

Genotype data

A total of 190 saliva samples were genotyped with the Affymetrix Axiom Genome-Wide Human Origins 1 array. DNA extraction was carried out following the protocol for manual purification of

DNA for Oragene·DNA/saliva samples in the same laboratory that sequenced the 16S rRNA data. Samples were analysed with Axiom Analysis Suite v4.0 following the Axiom genotyping best-practices workflow for saliva samples. 618,810 markers and 177 samples passed initial quality control. Single nucleotide polymorphisms (SNPs) with less than 95% genotyping rate and samples where the estimated gender from the genotypes did not match the recorded gender were excluded. Duplicated samples were identified with KING (Manichaikul et al., 2010) and removed. This resulted in a total of 617,063 markers and 174 samples: 141 Agta, 19 BaYaka and 14 Palanan farmers.

Ethnographic data collection

Ethnographic data collection occurred over two field seasons from April to June 2013 and from February to October 2014. In the first season we censused 915 Agta individuals (54.7% male) across 20 camps, collecting basic information on household composition, sex and estimated ages. Following relative ageing protocols (Diekmann et al., 2017), accurate ages were established for all individuals after data collection.

Diet data collection

Dietary recall data were collected at the household level over a period of 10 days. We asked the mother and father in the afternoon (between 17:00 and 18:00) which foods they had eaten on that day, including agricultural products traded with nearby farmers. We counted the total amount of meals recorded for a household and established what proportion of these consisted of meat, vegetables, fruits, honey and rice. Therefore, this is only a rough guide to dietary composition and does not take calorific intake or absolute weights of the different food types into account. Dietary data for 80 individuals (37 males and 43 females) were annotated based on the proportion of meals that consisted of only rice, and the proportion of meals that included meat (primarily fish and other marine resources and game).

Classification of oral bacteria as pathogens

Bacteria were classified as potential oral pathogens if they have been reported as etiological agents of periodontitis or dental caries. Assignment as a periodontal pathogen was performed according to the systematic review of Pérez-Chaparro et al. (2014) and Socransky et al. (1998), or if they have been previously associated with gum disease (Camelo-Castillo et al., 2015; Khemwong et al., 2019). Bacteria were classified as caries pathogens if they were described in transcriptomic studies of human cavities, according to Sómon-Soro et al. (2014) and Sómon-Soro and Mira (2015), previously associated with caries (Tanner, 2015; Kressirer, 2017), cavities (Wolff et al., 2019) or dental plaque and dental calculus formation (Mark Welch et al., 2016; Ferrer & Mira, 2016). Bacteria reported as etiological agents of respiratory infections and biofilm-mediated infections were also considered pathogens, including organisms present in healthy carriers. These included species described in Leung et al. (2018), Bellussi et al. (2019) and Natsis and Cohen (2018). Bacteria causing urinary tract infection or sexually transmitted diseases transiently found in the oral cavity were also considered as potential pathogens and included microorganisms described in Lanao and Pearson-Shaver (2020) and Jung et al. (2017). Common oral commensals potentially causing endocarditis or systemic infections in immunocompromised patients were not considered pathogens. If a bacterium was isolated from the oral cavity of another animal species, it was considered oral for the sake of our classification. If taxonomic classification in our dataset could be assigned at the genus level only, it was considered potentially pathogenic if: (i) >90% of species within the genus were pathogenic; or (ii) it included a major pathogenic species but the other species within the genus were not oral according to the Human Oral Microbiome Database (http://www.ehomd.org/) (Chen et al., 2010; Escapa et al., 2018). Cases where taxonomic classification of the ASV was only possible at the family level or higher (order, class) or ASVs with a top hit to a sequence classified as 'Oral taxa' in databases but without a species assignment were

not included in this classification (Supplementary Table S4). We cannot rule out the presence of relatively new genera not included in the database at the time of this study, such as *Schallia*.

The alternative classification of certain bacteria as 'pathobionts' rather than pathogenic has been proposed (Simón-Soro & Mira, 2015) to account for the fact that while some bacteria are associated with oral pathologies when present in high levels, they can also be found at low levels in healthy individuals. However, this does not prevent the identification of potential pathogens where there is strong evidence of the role of bacteria in creating the conditions for the development of oral disease. A clear example is provided by the role of the extremely saccharolytic, acidogenic and acidophilic *Streptococcus mutans* or *Scardovia wiggsiae* bacteria that proliferate in an acidic environment (either by causing it or being able to tolerate it), a known condition for the development of cavities. For this reason, we decided to rely on the categories of potentially pathogenic and non-pathogenic in our classification scheme.

Multivariate compositional data analysis on microbial composition

We performed a constrained LRA using the R package *easyCODA* (Greenacre, 2018) on the Agta oral microbiome at the genus level using as constraining covariates age (continuous variable), sex (male and female) and diet (both proportion of meals with meat and proportion of meals with only rice). Microbiome abundance counts of each Agta individual were treated as compositional data (Gloor et al., 2017) and converted to logarithms of ratios (log ratios). Constrained LRA is a special case of redundancy analysis (Greenacre, 2018) where total log ratio variance is decomposed into fractions explained by covariates (the 'constrained variance') and a residual fraction. Then, ordination resulting from LRA explains a maximum of the constrained variance in a reduced two-dimensional solution. Statistical significance of the three covariates was assessed using a multivariate permutation test (999 permutations) in the R package *vegan*. There is no correlation between the three covariates, except within the diet covariate, where the two variables are negatively correlated (Spearman's $\rho = -0.54$, $p < 0.0001$; Supplementary Figure S1). To focus on the genera affected only by intrinsic factors (age and sex), we performed a constrained LRA on the microbial composition after partialling out the effects of diet. Similarly, to identify genera affected exclusively by diet, we performed a constrained LRA after partialling out the effects of age and sex. Taxon-covariate association was ranked by counting the number of significant log ratios for each of the taxa, with p-value < 0.05 controlling for false discovery rate (FDR) at level $\alpha = 0.05$.

Community detection

To model the relationship between the Agta and the CMM, we used a stochastic block model (SBM) approach specifically suited for bipartite networks (Larremore et al., 2014). SBM infers the community structure (Fortunato, 2010) that better fits the existing graph, by building a prior distribution for edges that holds no information on real data and using it in the framework of Bayesian inference (biSBM) to find a partition of the two types of nodes whose associated entropy is maximal. In this framework, the absence of links between nodes of the same type or set is not considered informative, as expected given the bipartite nature of the graph, different from the general version of SBM. We selected the number of clusters in the two sets that minimised description length (Peixoto, 2017). Robustness of the clustering was assessed by calculating the average adjusted Rand index (ARI) between iterations ($n = 100$), finding a mean ARI on Agta = 0.90 and a mean ARI on ASV = 0.70. The ARI measures the similarity of two partitions against a null hypothesis of random assignment maintaining the size of the different clusters; the closer to 1 it is, the more robust is the classification (Hubert & Arabie, 1985). The resulting clusters were plotted with *graph-tool* (Tiago, 2014).

Ranking of bacteria associated with diet

ASVs were ranked from −1 to 1 based on whether they were more present than expected in individuals from a given category: low proportion vs. high proportion of meals with only rice, and low proportion

vs. high proportion of meals with meat, based on the median value of the population for each variable. For example, a meat-associated score close to 1 indicates that an ASV is present above the expected in individuals with high proportion of meals with meat (above the median of the population). A rice-associated score near 1 indicates that an ASV is present above the expected in individuals with high proportion of meals that consist of only rice.

Microbiome genome-wide association studies

We used a GWAS approach to identify specific SNPs associated with microbial abundance in the Agta using *GEMMA* version 0.94 (Zhou & Stephens, 2012). GWAS were performed using the relative abundance of a given taxon as a phenotype trait, adding as covariates age, sex and household (as a proxy for diet and shared environment). A kinship matrix calculated by KING using identical by descent segment inference (Manichaikul et al., 2010) was included as random effects. For the GWAS analyses, we applied the following quality control steps. First, to detect ancestry outliers, we filtered samples to keep only bi-allelic autosomal SNPs with minor allele frequency (MAF) > 5% and without missing data with *PLINK* 1.9 (Chang et al., 2015). This dataset was pruned for linkage disequilibrium using --indep-pairwise 50 5 0.2, and we performed a principal component analysis with *EIGENSOFT* version 7.2.1 (Patterson et al., 2006) to identify ancestry outliers and exclude them (Supplementary Figure S9). Second, per sample heterozygosity was calculated with *PLINK* and samples with overall increased or decreased heterozygosity rates (±3 SD from the mean of the population) were removed. A total of 129 Agta samples passed microbiome and genotype data quality controls and were included in the microbiome GWAS analyses. Analyses were done at genus level and on the CMM. The number of individuals tested ranged from 10 to 129 and the number of SNPs tested ranged from 270,569 to 313,198 markers depending on the taxon, as we included only samples with non-zero abundance, excluding SNPs with MAF < 10% and with more than 5% missing data. When we performed the GWAS at the genus level, we only included in the analysis the 92 genera present in at least 10 Agta individuals. *p*-Values were adjusted for multiple testing by FDR within each taxa (92 genera and 575 ASV), and SNP-taxa associations were considered significant at *q*-value < 0.1 when the proportion of variance in bacterial abundance explained by the genotypes (PVE or 'chip heritability') was non-zero. The proportion of variance in the phenotype (bacterial abundance) explained by the genotypes tested (PVE or 'chip heritability') was estimated for each taxon and was considered non-zero if the standard error measurements did not intersect zero. We applied genomic control to correct for cryptic relatedness and population stratification and minimise false positives induced by inflated association test statistics (Devlin & Roeder, 1999). To do so, we estimated the genomic inflation factor as the median value of the likelihood ratio test (LRT) values divided by 0.456 (median of a $\chi^2(1)$ distribution) and recalculated the *p*-values after dividing LRT values by the genomic inflation factor (Bacanu et al., 2000). The threshold of significance was set at FDR 10%, and only genomic positions having at least three samples for the major homozygous genotype and for the heterozygous genotype were considered. SNPs were annotated with *ANNOVAR* (Wang et al., 2010) in *GRCh37* (hg19) using *RefSeqGene* and *dbSNP* 147. For the enrichment analyses, we extracted genes associated with all non-intergenic SNPs and classified genes in the background set (consisting of all genes in the Axiom Human origin array) and the set to test (all genes with a non-intergenic SNP significantly associated with an ASV or a genus). We performed a gene ontology enrichment analysis with *ViSEAGO* (Brionne et al., 2019) and *topGo* (Alexa and Rahnenfuhrer, 2019) R packages (Fisher exact test) and *FUMA GENE2FUNCTION* module (Functional Mapping and Annotation of Genome-Wide Association Studies; Watanabe et al., 2017) to perform pathway enrichment analysis (hypergeometric test) with FDR 5%.

Selection analyses

To test whether GWAS SNPs showed any signal of recent positive selection we performed a genome-wide selection scan. We phased the Agta and Palanan farmer populations independently. For each

population, samples identified as ancestry outliers by a principal component analysis or with overall increased or decreased heterozygosity rates (±3 SD from the population mean) were excluded. A total of 138 Agta samples were phased using *SHAPEIT2* version v2 (r900) (O'Connell et al., 2014) with the duoHMM method to improve phasing by integrating known pedigree information. SNPs with missing data were removed and window size was set to 5Mb. Due to the small sample size, to phase the 14 unrelated Palanan farmers we used SHAPEIT2 with default parameters and the 1,000 Genomes Phase 3 panel of haplotypes (Auton et al., 2015) as a reference dataset. SNPs with missing data were removed. For the selection analyses, we excluded one of each pair of related individuals by removing the individual with the lowest call rate in the Agta phased dataset. This resulted in 38 unrelated Agta individuals and 14 unrelated Palanan farmers. We ran the Integrated Haplotype Score (iHS) (Voight et al., 2006) on the Agta phased dataset, and the Cross-population Extended Haplotype Homozygosity (XP-EHH) test (Sabeti et al., 2007) comparing the Agta against Palanan farmers as implemented in *selscan* version v1.3.0 (Szpiech & Hernandez, 2014) to identify signals of positive selection in GWAS SNPs. Both tests were run with default parameters and with the genetic map provided by the 1,000 Genomes Phase 3 (Auton et al. 2015). To identify regions under selection, for each test we selected markers with scores in the 95th percentile that had at least three markers in the 99th percentile in the surrounding area (± 10 kb). For iHS we used absolute values, while only positive scores were analysed for XP-EHH.

References

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., … Cooper, A. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nature Genetics*, *45*(4), 450–455. http://dx.doi.org/10.1038/ng.2536

Alexa, A. & Rahnenfuhrer, J. (2019). topGO: Enrichment Analysis for Gene Ontology. R package version 2.52.0.

Auton, A., Abecasis, G. R., Altshuler, D., Durbin, R., Bentley, D., Chakravarti, A., … Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. http://dx.doi.org/10.1038/nature15393

Bacanu, S. A., Devlin, B., & Roeder, K. (2000). The power of genomic control. *American Journal of Human Genetics 66* (6), 1933–1944.

Bellussi, L., Passali, F. M., Ralli, M., De Vincentiis, M., Greco, A. V., & Passali, D. (2019). An overview on upper respiratory tract infections and bacteriotherapy as innovative therapeutic strategy. *European Reviews on Medical and Pharmacological Science*, *23*, 27–38. https://doi.org/10.26355/eurrev_201903_17345

Belstrøm, D., Holmstrup, P., Nielsen, C. J., Kirkby, N., Twetman, S., Heitmann, B. L, … Fiehn, N. (2014). Bacterial profiles of saliva in relation to diet, lifestyle factors, and socioeconomic status. *Journal of Oral Microbiology*, *6*(1), 23609. https://doi.org/10.3402/jom.v6.23609

Blekhman, R., Goodrich, J. K., Huang, K. H., Hu, F. B., Bukowski, R., Bell, J. T., … Clark, A. G. (2015). Host genetic variation impacts microbiome composition across human body sites. *Genome Biology*, *16*(1), 191. https://doi.org/10.1186/s13059-015-0759-1

Bolyen, E., Rideout, J. R., Dillon, M., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., … Walters, W. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Brionne, A., Juanchich, A., & Hennequet-Antier, C. (2019). ViSEAGO: A bioconductor package for clustering biological functions using Gene Ontology and semantic similarity. *BioData Mining*, *12*(1), 16. https://doi.org/10.1186/s13040-019-0204-1

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, *11*(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. J., Johnson, A. J., & Holmes, S. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camelo-Castillo, A., Balsa-Castro, L. N., Balsa-Castro, C., Blanco, J., Mira, A., & Tomás, I. (2015). Relationship between periodontitis-associated subgingival microbiota and clinical inflammation by 16S pyrosequencing. *Journal of Clinical Periodontology*, *42*(12), 1074–1082. https://doi.org/10.1111/jcpe.12470

Chang, C. J., Chow, C. C, Tellier, L. C. A. M., Vattikuti, S., Purcell, S., & Lee, J. D. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, T., Yu, W, Izard, J., Baranova, O. V., Lakshmanan, A., & Dewhirst, F. E. (2010). The Human Oral Microbiome Database: A web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)*, *2010*(0), baq013. https://doi.org/10.1093/database/baq013

Cornejo Ulloa, P., van der Veen, M. H., & Krom, B. P. (2019). Review: modulation of the oral microbiome by the host to promote ecological balance. *Odontology*, *107*(4), 437–448. https://doi.org/10.1007/s10266-019-00413-x

Costello, E. K., Lauber, C. L, Hamady, M., Fierer, N., Gordon, J. I., & Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science*, *326*(5960), 1694–1697. https://doi.org/10.1126/science.1177486

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., … Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559–563. https://doi.org/10.1038/nature12820

De Filippis, F., Vannini, L., La Storia, A., Laghi, L., Piombino, P., Stellato, G., … Gobbetti, M. (2014). The same microbiota and a potentially discriminant metabolome in the saliva of omnivore, ovo-lacto-vegetarian and vegan individuals. *PLoS One*, *9*(11), e112373. https://doi.org/10.1371/journal.pone.0112373

De Maeyer, R. P. H., & Chambers, E. S. (2021). The impact of ageing on monocytes and macrophages. *Immunology Letters*, *230*, 1–10. https://doi.org/10.1016/j.imlet.2020.12.003

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997–1004. https://doi.org/10.1111/j.0006-341x.1999.00997.x

Diekmann, Y., Smith, D. J., Gerbault, P., Dyble, M., Page, A. E., Chaudhary, N., … Thomas, M. G. (2017). Accurate age estimation in small-scale societies. *Proceedings of the National Academy of Science*, *114*(31), 8205–8210. https://doi.org/10.1073/pnas.1619583114

Duan, X., Wu, T., Xu, X., Chen, D., Mo, A., Lei, Y., … Yuan, Q. (2017). Smoking may lead to marginal bone loss around non-submerged implants during bone healing by altering salivary microbiome: A prospective study. *Journal of Periodontology*, *88*(12), 1297–1308. https://doi.org/10.1902/jop.2017.160808

Dyble, M., Thompson, J. C., Smith, D. J., Salali, G. D., Chaudhary, N., Page, A. E., … Migliano, A. B. (2016). Networks of food sharing reveal the functional significance of multilevel sociality in two hunter–gatherer groups. *Current Biology*, *26*(15), 2017–2021. https://doi.org/10.1016/j.cub.2016.05.064

Dyble, M., Thorley, J., Page, A. E., Smith, D., & Migliano, A. B. (2019). Engagement in agricultural work is associated with reduced leisure time among Agta hunter–gatherers. *Nature Human Behaviour*, *3*(8), 792–796. https://doi.org/10.1038/s41562-019-0614-6

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in Microbiology*, *27*(2), 105–117. https://doi.org/10.1016/j.tim.2018.11.003

Eke, P. I., Dye, B. A., Wei, L., Slade, G. D., Thornton-Evans, G., Borgnakke, W. S., … Genco, R. J. (2015). Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012. *Journal of Periodontology*, *86*(5), 611–622. https://doi.org/10.1902/jop.2015.140520

Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E., & Lemon, K. N. (2018). New insights into human nostril microbiome from the Expanded Human Oral Microbiome Database (eHOMD): A resource for the microbiome of the human aerodigestive tract. *mSystems*, *3*(6), e00187. https://doi.org/10.1128/msystems.00187-18

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biology Conservation*, *61*(1), 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Fernández, L., Pannaraj, P. S., Rautava, S., & Rodríguez, J. M. (2020). The microbiota of the human mammary ecosystem. *Frontiers in Cellular and Infection Microbiology*, *10*, 586667. https://doi.org/10.3389/fcimb.2020.586667

Ferraro, M., & Vieira, A. R. (2010). Explaining gender differences in caries: A multifactorial approach to a multifactorial disease. *International Journal of Dentistry*, 2010, 1–5. https://doi.org/10.1155/2010/649643

Ferrer, M. D., & Mira, A. (2016). Oral biofilm architecture at the microbial scale. *Trends in Microbiology*, 24(4), 246–248. https://doi.org/10.1016/j.tim.2016.02.013

Fives-Taylor, P. M., Meyer, D. H., Mintz, K. P., & Brissette, C. A. (1999). Virulence factors of *Actinobacillus actinomycetem-comitans*. *Periodontology* 2000, 20(1), 136–167. https://doi.org/10.1111/j.1600-0757.1999.tb00161.x

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, 8, 2224. https://doi.org/10.3389/fmicb.2017.02224

Gomez, A., Espinoza, J. L., Harkins, D. M., Leong, P., Saffery, R., Bockmann, M., … Nelson, K. E. (2017). Host genetic control of the oral microbiome in health and disease. *Cell Host Microbe*, 22(3), 269–278. https://doi.org/10.1016/j.chom.2017.08.013

Greenacre, M. (2018).*Compositional Data Analysis in Practice*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429455537

Hernandez, B. Y., Zhu, X., Goodman, M. T., Gatewood, R. D., Mendiola, P., Quinata, K., & Paulino, Y. C. (2017). Betel nut chewing, oral premalignant lesions, and the oral microbiome. *PLOS One*, 12(2), e0172196. https://doi.org/10.1371/journal.pone.0172196

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218. https://doi.org/10.1007/bf01908075

Islam, S., Muthumala, M., Matsuoka, H., Uehara, O., Kuramitsu, Y., Chiba, I., & Abiko, Y. (2019). How each component of betel quid is involved in oral carcinogenesis: Mutual interactions and synergistic effects with other carcinogens – a review article. *Current Oncology Reports*, 21(6), 53. https://doi.org/10.1007/s11912-019-0800-8

Jung, H. S., Ehlers, M. M., Lombaard, H., Redelinghuys, M. J., & Kock, M. M. (2017). Etiology of bacterial vaginosis and polymicrobial biofilm formation. *Critical Reviews in Microbiology*, 43(6), 651–667. https://doi.org/10.1080/1040841x.2017.1291579

Kassebaum, N. J., Bernabé, E., Dahiya, M, Bhandari, B., Murray, C. J. L., & Marcenes, W. (2014). Global Burden of Severe Periodontitis in 1990-2010. *Journal of Dental Research*, 93(11), 1045–1053. http://dx.doi.org/10.1177/0022034514552491

Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kembel, S. W, Cowan, P. J., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., … Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26(11), 1463–1464. http://dx.doi.org/10.1093/bioinformatics/btq166

Khemwong, T., Kobayashi, H., Ikeda, Y., Matsuura, T., Sudo, T., Kano, C., … Izumi, Y. (2019). *Fretibacterium* sp. human oral taxon 360 is a novel biomarker for periodontitis screening in the Japanese population. *PLoS One*, 14(6), e0218266. https://doi.org/10.1371/journal.pone.0218266

Kolde, R., Franzosa, E. A., Rahnavard, G., Hall, A. J., Vlamakis, H., Stevens, C., … Huttenhower, C. (2018). Host genetic variation and its microbiome interactions within the Human Microbiome Project. *Genome Medicine*, 10(1), 207. http://dx.doi.org/10.1186/s13073-018-0515-8

Kressirer, C. A., Smith, D. J., King, W. P., Dobeck, J. M., Starr, J. R., & Tanner, A. C. (2017). Scardovia wiggsiae and its potential role as a caries pathogen. *Journal of Oral Bioscience*, 59(3), 135–141. https://doi.org/10.1016/j.job.2017.05.002

Lanao, A. E., & Pearson-Shaver, A. L. (2020). Mycoplasma infections. *StatPearls [Internet]*.

Larremore, D. B., Clauset, A., & Jacobs, A. Z. (2014). Efficiently inferring community structure in bipartite networks. *Physics Reviews E*, 90(1), 012805. https://doi.org/10.1103/physreve.90.012805

Leung, A. K. C., Wong, A. H. C., & Hon, K. L. (2018). Community-Acquired Pneumonia in Children. *Recent Patents on Inflammation & Allergy Drug Discovery*, 12(2), 136–144. https://doi.org/10.2174/1872213x12666180621163821

Liu, S., Ying, G. G., Liu, Y. S., Peng, F. Q., & He, L. Y. (2013). Degradation of norgestrel by bacteria from activated sludge: Comparison to progesterone. *Environmental Science and Technology*, 47, 130829113920003. https://doi.org/10.1021/es304688g

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M. M., & Chen, W. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22), 2867–2873. https://doi.org/10.1093/bioinformatics/btq559

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. https://doi.org/10.1371/journal.pone.0061217

Migliano, A. B., Page, A. E., Gómez-Gardeñes, J., Salali, G. D., Viguier, S, Dyble, M., … Vinicius, L. (2017). Characterization of hunter–gatherer networks and implications for cumulative culture. *Nature Human Behaviour*, 1(2), 0043. https://doi.org/10.1038/s41562-016-0043

Minter, T. (2010) *The Agta of the Northern Sierra Madre: Livelihood strategies and resilience among Philippine hunter–gatherers*. Faculty of Social Sciences, Leiden University.

Mira, A., Artacho, A., Camelo-Castillo, A., Garcia-Esteban, S., & Simon-Soro, A. (2017). Salivary immune and metabolic marker analysis (SIMMA): A diagnostic test to predict caries risk. *Diagnostics*, 7(3), 38. https://doi.org/10.3390/diagnostics7030038

Mukherjee, C., Moyer, C. O., Steinkamp, H. M., Hashmi, S. B., Beall, C. J., Guo, X., … Griffen, A. L. (2021). Acquisition of oral microbiota is driven by environment, not host genetics. *Microbiome*, 9(1), 54. https://doi.org/10.1186/s40168-020-00986-8

Musciotto, F. Dobon, B., Greenacre, M., Mira, A., Chaudhary, N., Salali, G. D., … Migliano, A. B. (2023). Agta hunter–gatherer oral microbiomes are shaped by contact network structure. *Evolutionary Human Sciences*, 5, E9. https://doi.org/10.1017/ehs.2023.4

Natsis, N. E., & Cohen, P. R. (2018). Coagulase-negative *Staphylococcus* skin and soft tissue infections. *American Journal of Clinical Dermatology*, 19(5), 671–677. https://doi.org/10.1007/s40257-018-0362-9

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., … Marchini, J. (2014). A General approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genetics*, 10(4), e1004234. https://doi.org/10.1371/journal.pgen.1004234.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2020). Vegan: Community ecology package. Ordination methods, diversity analysis and other functions for community and vegetation ecologists. R Package version 2.5-7.

Page, A. E., Minter, T., Viguier, S., & Migliano, A. B. (2018). Hunter–gatherer health and development policy: How the promotion of sedentism worsens the Agta's health outcomes. *Social Science Medicine*, 197, 39–48. https://doi.org/10.1016/j.socscimed.2017.12.002

Page, A. E., Viguier, S., Dyble, M., Smith, D. J., Chaudhary, N., Salali, G. D., … Migliano, A. B. (2016). Reproductive trade-offs in extant hunter–gatherers suggest adaptive mechanism for the Neolithic expansion. *Proceedings of the National Academy of Science*, 113(17), 4694–4699. https://doi.org/10.1073/pnas.1524031113

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Peixoto, T. P. (2014). The graph-tool python library. figshare. doi: 10.6084/m9.figshare.1164194.

Peixoto, T. P. (2017). Nonparametric Bayesian inference of the microcanonical stochastic block model. *Physics Review E*, 95(1), 012317. https://doi.org/10.1103/physreve.95.012317

Pérez-Chaparro, P. J., Gonçalves, C. R., Figueiredo, L. C., Faveri, M., Lobão, E., Tamashiro, N., … Feres, M. (2014). Newly identified pathogens associated with periodontitis. *Journal of Dental Research*, 93(9), 846–858. https://doi.org/10.1177/0022034514542468

Preshaw, P. M., Henne, K., Taylor, J. J., Valentine, R. A. & Conrads, G. (2017). Age-related changes in immune function (immune senescence) in caries and periodontal diseases: A systematic review. *Journal of Clinical Periodontology*, 44, S153–S177. https://doi.org/10.1111/jcpe.12675

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J. E., Schweer, T., Yarza, P., … Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org

Rosier, B. T., Moya-Gonzalvez, E. M., Corell-Escuin, P., & Mira, A. (2020). Isolation and characterization of nitrate-reducing bacteria as potential probiotics for oral and systemic health. *Frontiers in Microbiology*, 11, 2261. https://doi.org/10.3389/fmicb.2020.555465

Sabbatani, S., & Fiorino, S. (2016). Dental worm disease. *Le Infezioni in Medicina*, 24(4), 349–358.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Pawlikowska, L. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, 449(7164), 913–918. https://doi.org/10.1038/nature06250

Schnorr, S. L., Candela, M., Rampelli, S, Centanni, M., Consolandi, C., Basaglia, G., … Crittenden, A. N. (2014). Gut microbiome of the Hadza hunter–gatherers. *Nature Communications*, 5(1), 3654. https://doi.org/10.1038/ncomms4654

Shaw, L. P., Ribeiro, A. S., Levine, A. C., Pontikos, N., Balloux, F., Segal, A. W., … Smith, A. M. (2017). The human salivary microbiome is shaped by shared environment rather than genetics: Evidence from a large family of closely related individuals. *MBio*, 8(5), e01237. https://doi.org/10.1128/mbio.01237-17

Shiau, H. J., & Reynolds, M. A. (2010). Sex differences in destructive periodontal disease: A systematic review. *Journal of Periodontology*, 81(10), 1379–1389. https://doi.org/10.1902/jop.2010.100044

Simón-Soro, A., Guillen-Navarro, M., & Mira, A. (2014). Metatranscriptomics reveals overall active bacterial composition in caries lesions. *Journal of Oral Microbiology*, 6(1). https://doi.org/10.3402/jom.v6.25443

Simón-Soro, A., & Mira, A. (2015). Solving the etiology of dental caries. *Trends in Microbiology*, 23(2), 76–82. https://doi.org/10.1016/j.tim.2014.10.010

Smits, S. A., Leach, J. D., Sonnenburg, E. D., González, C., Lichtman, J. S., Reid, G., … Sonnenburg, J. L. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter–gatherers of Tanzania. *Science*, 357(6353), 802–806. https://doi.org/10.1126/science.aan4834

Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., & Kent Jr, R. L. (1998). Microbial complexes in subgingival plaque. *Journal of Clinical Periodontology*, 25(2), 134–144. https://doi.org/10.1111/j.1600-051x.1998.tb02419.x

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mology Biology Evology*, *31*(10), 2824–2827. https://doi.org/10.1093/molbev/msu211

Tanner, A. C. R. (2015). Anaerobic culture to detect periodontal and caries pathogens. *Journal of Oral Bioscience*, *57*(1), 18–26. https://doi.org/10.1016/j.job.2014.08.001

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., & Gordon, J. I. (2009). The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine*, *1*(6). https://doi.org/10.1126/scitranslmed.3000322

Velsko, I. M., Yates, J. W., Aron, F., Hagan, R. W., Frantz, L. A. F., Loe, L., … Warinner, C. (2019). Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. *Microbiome*, *7*(1), 102. https://doi.org/10.1186/s40168-019-0717-3

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006) A map of recent positive selection in the human genome. *PLoS Biology*, *4*(3). https://doi.org/10.1371/journal.pbio.0040072

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164–e164. https://doi.org/10.1093/nar/gkq603

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, *8*(1), 1826. https://doi.org/10.1038/s41467-017-01261-5

Welch, J. L. M., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E., & Borisy, G. G. (2016). Biogeography of a human oral microbiome at the micron scale. *Proceedings of the National Academy of Science*, *113*(6), E791–E800. https://doi.org/10.1073/pnas.1522149113

Weyrich, L. S. (2021). The evolutionary history of the human oral microbiota and its implications for modern health. *Periodontology* 2000, *85*(1), 90–100. https://doi.org/10.1111/prd.12353

Willis, J. R., González-Torres, P., Pittis, A. A., Bejarano, L., Cozzuto, L., Andreu-Somavilla, N., … Gabaldón, T. (2018). Citizen science charts two major 'stomatotypes' in the oral microbiome of adolescents and reveals links with habits and drinking water composition. *Microbiome*, *6*(1), 218. https://doi.org/10.1186/s40168-018-0592-3

Wolff, D., Frese, C., Schoilew, K., Dalpke, A, Wolff, B., & Boutin, S. (2019). Amplicon-based microbiome study highlights the loss of diversity and the establishment of a set of species in patients with dentin caries. *PLoS One*, *14*(7), e0219714. https://doi.org/10.1371/journal.pone.0219714

Zaura, E., Brandt, B. W., Prodan, A., De Mattos, M. J. T., Imangaliyev, S., Kool, J., … Keijser, B. J. F. (2017). On the ecosystemic network of saliva in healthy young adults. *ISME Journal*, *11*(5), 1218–1231. https://doi.org/10.1038/ismej.2016.199

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310