Title

Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet

Florent Lassalle[1,2]*, Matteo Spagnoletti[1]*, Matteo Fumagalli[1], Liam Shaw[1], Mark Dyble[3,4], Catherine Walker[1], Mark G. Thomas[1], Andrea Bamberg Migliano[3], Francois Balloux[1].

[1]*University College London, UCL Genetics Institute, Gower Street, London WC1E 6BT, UK;*

[2]*Imperial College London, Department of Infectious Disease Epidemiology, Praed Street, London W2 1NY, UK;*

[3]*University College London, Department of Anthropology, 14 Taviton Street, London WC1H 0BW, UK;*

[4]*University of Cambridge, Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK.*

**These authors contributed equally to this work*

Keywords: metagenomics; hunter-gatherers; oral microbiome; diet; Philippines

Corresponding authors:

Florent Lassalle (f.lassalle@imperial.ac.uk); Francois Balloux (f.balloux@ucl.ac.uk)

Running title

Impact of diet on oral microbiome composition

Abstract (209/200 words)

Maladaptation to modern diets has been implicated in several chronic disorders. Given the higher prevalence of disease such as dental caries and chronic gum diseases in industrialized societies, we sought to investigate the impact of different subsistence strategies on oral health and physiology, as documented by the oral microbiome. To control for confounding variables such as environment and host genetics, we sampled saliva from three pairs of populations of hunter-gatherers and traditional farmers living in close proximity in the Philippines. Deep shotgun sequencing of salivary DNA generated high-coverage microbiomes along with human genomes. Comparing these microbiomes with publicly available data from individuals living on a Western diet revealed that abundance ratios of core species were significantly correlated with subsistence strategy, with hunter-gatherers and Westerners occupying either end of a gradient of *Neisseria* against *Haemophilus*, and traditional farmers falling in between. Species found preferentially in hunter-gatherers included microbes often considered as oral pathogens, despite their hosts' apparent good oral health. Discriminant analysis of gene functions revealed vitamin B5 autotrophy and urease-mediated pH regulation as candidate adaptations of the microbiome to the hunter-gatherer and Western diets, respectively. These results suggest that major transitions in diet selected for different communities of commensals and likely played a role in the emergence of modern oral pathogens.

Introduction

Humans have experienced dramatic changes in diet over the last 10,000 years (Mathieson et al., 2015; Quercia et al., 2014). The Neolithic transition marked the beginning of wide-scale dietary and demographic changes from subsistence by primarily nomadic hunting and gathering to sedentary agriculture (Bocquet-Appel, 2011). A second, equally dramatic nutritional shift occurred with the Industrial Revolution in the mid-19th century, which led to widespread availability of processed flour and sugar (Cordain et al., 2005). These alterations of ancestral diets have been implicated in the emergence of modern chronic disorders, including cardiovascular disease, diabetes, obesity and osteoporosis (Cordain et al., 2005).

The human microbiome, the sum of diverse microbial ecosystems colonizing the various niches offered by the human body, is known to play an important role in human health (Lloyd-Price, Abu-Ali, & Huttenhower, 2016; Yang et al., 2012). In particular, the oral cavity, which is the gateway to the human body for both food and air intake, hosts the oral microbiome (Dewhirst et al., 2010). Shifts in composition of this microbial community have been associated with several oral conditions such as periodontitis (Griffen et al., 2012), which in turn is suspected as a cause of a series of modern chronic disorders, including inflammatory bowel disease, diabetes, cardiovascular disease and some forms of cancer (Kuo, Polson, & Kang, 2008; Li, Kolltveit, Tronstad, & Olsen, 2000; Whitmore & Lamont, 2014). By occupying a major interface between the human body and the external environment, the oral microbiome is shaped both by host variables, such as genetic background, general health and immunity, and by external environmental factors including ecology and diet. The relative abundance of microbes colonizing the mouth changes along the day through growth and regular clearance by swallowing of saliva, but the set of taxa observed over time in an individual's mouth is remarkably stable (Carpenter, 2013; Marsh, Do, Beighton, & Devine, 2016).

Despite its compositional stability on the shorter term, there is strong evidence that oral microbiome composition has been shaped by major sociocultural changes over our recent evolutionary history (Mira *et al.*, 2006; Hunter, 2014). Indeed, analysis of ancient and historic dental calculus samples has identified major shifts in species composition in the oral microbiome coinciding with the Neolithic and Industrial Revolution (Adler et al., 2013; Warinner et al., 2014). As dietary and oral hygiene standard shifts have occurred over a relatively short evolutionary timescale, it has been suggested that modern human microbiomes may be maladapted to current conditions, leading to increased incidence of oral diseases. This would be consistent with the spread of major oral polymicrobial diseases across human populations in recent times (Marsh, 2003; Zaura, Nicu, Krom, & Keijser, 2014). In most industrialized countries 60-90% of children have signs of caries and clinically defined periodontal disease is highly prevalent among adults (Petersen, 2005). Additionally, modern chronic disorders linked to oral disease – inflammatory bowel disease, cardiovascular disease, diabetes and cancer (Kaplan et al., 2017; Kuo, Polson, & Kang, 2008; Whitmore & Lamont, 2014) – all tend to be rare in contemporary hunter-gatherers, whose lifestyle and diet is deemed close to that of ancestral humans (Cordain et al., 2005; Marlowe, 2005). This suggests that the microbiome could act as a coupling link between human lifestyle and health. In particular, we make the hypothesis that recent changes in lifestyle and diet could have impacted the composition of oral microbiomes, which became conducive of modern chronic disorders.

The differences observed between archaeological and modern microbiomes may however not necessarily arise from shifts in subsistence strategies but from many other factors that changed through time. Additionally, direct comparison between modern microbiomes to those generated from ancient, degraded remains is not straightforward. It thus appears that comparison of microbiomes from contemporary populations exposed to similar environments but with contrasted lifestyles may represent the best experimental design to test whether diet is directly shaping the salivary microbiome.

A series of studies have investigated the microbiome composition of modern hunter-gatherers in comparison to neighboring populations of traditional farmers or more distant Western individuals (Clemente et al., 2015; Morton et al., 2015; Nasidze et al., 2011; Obregon-Tito et al., 2015; Schnorr et al., 2014). Notably, a few studies detected an effect of subsistence strategy on the oral microbiome, highlighting composition trends, such as the increased abundance of Fusobacteriaceae, Prevotellaceae, *Veillonella* spp. and *Haemophilus* spp. in hunter-gatherers' oral microbiomes (Clemente et al., 2015; Nasidze et al., 2011). However, these common composition features may be largely coincidental and need to be compared with data from other settings to consider them as diagnostic of subsistence strategy.

In addition, comparisons of microbiomes for a single pair of populations (e.g. hunter-gatherer against a population having adopted a Western diet) are likely to be confounded by additional differences in geographical origin, health, socio-economic status and possibly genetic backgrounds between the populations. To circumvent these problems, we designed our study around three pairs of populations living in close proximity in the Philippines and sharing essentially the same environment: Batak and Tagbanua, Aeta and Zambal and Agta and Casigurani, respectively hunter-gatherers (HGs) and traditional farmers (TFs). This design allowed us to detect systematic differences between all three pairs of populations that are much more likely to be driven by subsistence strategy. We also relied on deep whole genome shotgun (WGS) sequencing rather than the more standard but limited 16S rRNA amplicon-sequencing (Clemente et al., 2015; Nasidze et al., 2011). While the additional costs of the shotgun sequencing limited study sample size, it comes with an increased ability to resolve microbial species composition – in particular for populations whose microbiomes have not been well characterized to date – and also opens the door to direct investigation of the biological functions involved in their adaptation. This WGS approach also allowed us to generate human genomes (2-20x depth), which we used to control for a possible effect of the host genetic make-up.

The high-coverage oral microbiomes we generated were combined with previous datasets obtained with a similar protocol from individuals from the USA subsisting on a Western diet (Hasan et al., 2014; The Human Microbiome Project Consortium, 2012). These data were processed with state-of-the-art taxonomic assignation and phylogenetic diversity analyses to tease apart the effect of diet, environment and human genetic make-up in shaping the composition of the oral microbiome.

Materials and Methods

Study design, subject enrollment and DNA collection

The study included 24 samples selected from a large collection of saliva samples (>350) collected during a long-term fieldwork project in the Philippines under the supervision of Dr. Andrea Migliano (Hunter-Gatherers' resilience Project). Aeta live in the mountain forests from the western part of Luzon island, and Agta are from the East of Luzon, close to the coast. Batak live in the mountain forests in the central part of the Palawan island; the TF groups (Zambal, Casigurani and Tagbanua) live in close geographical proximity (1-10km) to each of the respective neighboring HG groups (Fig. 1). Saliva samples were collected in 2007, 2008 and 2009 (Table S1). Using the Oragene DNA OG-500 collection kit (DNA Genotek, Kanata, Canada), participants were asked to wash their mouth with water and then to spit into the vial until it is half full. All the samples were transported to London UK, where they were stored at the UCL department of Anthropology at -20°C.

The protocol was in accordance with the Helsinki Declaration, and was approved by the Ethics Commission of the University College London, London UK. We further obtained ethical clearance from the National Commission on Indigenous Peoples (NCIP) (Cariño, 2012). Approval was also obtained at the local community level, from the elders' committee in each of the locations, and informed consent was obtained from all participants (written in their own languages) after a

presentation of the research objectives in Tagalog for the Philippine populations; a copy of the Participant Information Sheet and an English version of the Participant Consent Form (blank copy) are available in Sup. File S1.

Sample selection and DNA extraction.

We selected four samples (from individual aged between 20 and 40 and in good oral health) for each group of hunter-gatherers and their neighboring farmers, generating three geographical groups of eight samples each, for a total of 24 samples. We randomly selected two males and two females, under the constraint of individuals being unrelated (without known family relationships, based on using the anthropological information collected during the field work). All the samples have been anonymized. DNA was purified from saliva employing the Oragene DNA isolation kit (DNA Genotek, Kanata, Canada), following the manufacturer's recommended instructions. DNA quantification and quality controls were accomplished using Qubit 2.0 fluorimeter (Thermo Fisher Scientific, Waltham, USA) and Agilent 2100 Bioanalyzer DNA chips (Agilent technologies, Santa Clara, USA).

DNA library preparation, sequencing and quality control

Aliquots of 1µg DNA per sample were used to create sequencing libraries. First, genomic DNA was fragmented using a Covaris S2 sonicator (Covaris Inc., Woburn, USA) to approximately 300bp. Fragmented DNA was quantified and used to synthesize shotgun libraries with the NebNext Ultra DNA library preparation kit for Illumina (New England Biolabs, Ipswich, USA), according to manufacturer's instructions. PCR cycling conditions were set to a minimum of 4 cycles for annealing/extension to minimize PCR duplicates. NEBNext Singleplex Oligos for Illumina were used for indexing samples without multiplexing. All the samples have been sequenced at the UCL Institute of Neurology using 100bp paired-end chemistry and the Illumina HiSeq 2500 system (Illumina, San Diego, USA). Three libraries were prepared, each grouping eight individuals: library #1 (4 Aeta HGs

154 + 4 Zambal neighboring TFs), library #2 (4 Batak HGs + 4 Tagbanua neighboring TFs) and library #3
155 (4 Agta HGs + 4 Casigurani neighboring TFs).
156 The libraries #1 and #2 were sequenced on one Illumina flow-cell each (8 lanes, one per individual),
157 while the library #3 was sequenced in two rounds, using two flow-cells (16 lanes, two per individual).
158 The whole sequencing process produced 21,362,688,072 reads (>870GB of data) passing filters
159 (Illumina CASAVA 1.8.0, default settings). Raw reads were processed using the first step of the
160 MOCAT pipeline (version 1.3) (Kultima et al., 2012) with standard settings (options "-identity 97 -
161 length 45 -soapmaxmm 5"): reads were quality trimmed, adapters were removed, and so were reads
162 matching human when mapping to reference hg19 (Genome Reference Consortium Human Reference
163 [GRCh] 37) using SOAPAligner2 (Li et al., 2009) version 2.21 with options "-r 2 -M 4 -l 30 -v 5 -p
164 4". This reduced the dataset to a total of 1.13 billion reads, with 8.3—147.7 million reads per
165 individual. These read sets were submitted to the ENA (www.ebi.ac.uk/ena) under the BioSample
166 accessions ERS1202862—ERS1202885. Human-mapped reads were further used to analyze the
167 genetic diversity of the sampled individuals (see Supplementary Methods).

168 **Kraken reference database**

169 We built a custom Kraken database (Wood & Salzberg, 2014) made from all available RefSeq
170 genomes for bacteria (94,803), archaea (676), viruses (7,497), protozoa (79) and fungi (238) using the
171 ncbi-genome-download application (https://github.com/kblin/ncbi-genome-download), as well as all
172 available RefSeq plasmids (10,842) directly from the NCBI FTP server
173 (ftp://ftp.ncbi.nih.gov/refseq/release/plasmid) as of September 19th 2017. We added the GRCh38,
174 HuRef and YanHuang human genome reference sequences (International Human Genome Sequencing
175 Consortium, 2004; Levy et al., 2007; Wang et al., 2008). The database was indexed for the distribution
176 of 31-mers in reference genomes, using 15-bp minimizers (Wood & Salzberg, 2014). The full database

had a final size of 539 Gb; this was shrunk to a 'Mini-kraken' indexed database of 193 Gb, covering 38,190 different taxa (with distinct NCBI taxon id).

Estimation of microbial taxonomic abundances

The 24 metagenomes generated in this study and nine additional Western metagenomes from other studies (see 'Public microbiome data' section in Supplementary text) were analyzed as follows. Reads were classified in terms of taxonomic origin using Kraken (Wood & Salzberg, 2014) version 0.10.6. This software searches k-mers in sequencing reads that match a custom database of reference genomes. Inclusion of the human genome in the reference database allowed to screen for remaining reads that were not identified at the previous filtering step by mapping. Reads assigned to human were removed from later steps of the analyses using a custom Python script (http://github.com/flass/microbiomes/kraken/parseKronaGetReadsByTaxid.py, option '--exclude.taxa 9606').

Kraken assigns reads to all taxonomic levels in a cumulative manner, and relative abundance of taxa can be computed using the ratio of read counts at one specific level over the total. Read counts were computed 1) with a conservative filter on read confidence scores, i.e. keeping only reads with more than 20% k-mers assigned to congruent taxa (using kraken-filter executable with option "--thresh 0.20"); and 2) in a sensitive mode, i.e. without confidence score filtering. Relative abundances were computed at the species and genus level. Distribution of relative species abundances per sample (from sensitive mode) showed significant bias relative to sequencing depth for values under 10^{-12}, with low-depth samples being depleted in rare species (Fig. S1), so the dataset was truncated to species relative abundance values above 10^{-12}, decreasing the number of represented species from 8,226 to 5,323. We used linear discriminant analysis (LDA) effect size (LEfSe) (Segata et al., 2011) to detect taxa that significantly differentiate groups of samples based on their subsistence strategy (accounting for the underlying grouping by population). We then used a simple LDA, as implemented in the ade4 R

package (Dufour & Dray, 2007), to identify the species that specifically differentiate microbiomes along the human lifestyle gradient opposing HGs to Western controls (WCs); significance was assessed with pairwise t-tests, Wilcoxon rank-sum tests (using Benjamini-Hochberg false discovery rate [FDR] correction procedure for multiple testing) and ANCOM test (Mandal et al., 2015), with low stringency multiple testing correction (option 'multcorr=2'). Abundance tables and complete reports of statistical analyses for filtered and unfiltered dataset, at species and genus levels, are available on Figshare at: https://doi.org/10.6084/m9.figshare.5213158.v1. Kraken taxonomic assignation makes use of the entire WGS dataset, allowing to characterize the presence of low-abundance organisms, but is biased towards taxa closely related to organisms represented in the reference database where exact sequence matches are possible, and does therefore not account for the phylogenetic sampling bias in the database.

We thus used Phylosift (Darling et al., 2014) (version 1.0.1) to characterize relative abundances of lineages in a phylogenetic placement framework that naturally allows for a robust assignment of taxonomic identity to sequences that are highly divergent with respect to the reference database. Briefly, a database of 33 highly conserved marker genes (Phylosift default built-in database, version 1395376975, available at http://edhar.genomecenter.ucdavis.edu/~koadman/phylosift_markers/) was searched for similarity with all reads. Those reads that matched (roughly 0.5-1% of the total dataset) were then assigned to a branch of a species tree built from the concatenation of the marker genes' reference alignments, using a phylogenetic placement algorithm (Matsen, Kodner, & Armbrust, 2010).

This procedure yielded a table of the density of placed reads per branch of the reference species tree, which can be used to compute relative abundances of a clade by summing the placement densities of all branches of the corresponding subtree. These can be translated into robust relative abundance estimates of named organisms at any taxonomic level using the taxonomic labelling of the branches of the tree provided with the Phylosift package, typically with a resolution of 10^{-3} for frequencies of

named species. The structure of the diversity of microbiome composition among samples can be conveniently explored using principal component analysis (PCA) of the difference of placement densities between reference tree edges, hereafter referred as 'edge PCA'. Custom Python and R scripts using the ade4 package (Dufour & Dray, 2007) (http://github.com/flass/microbiomes/tree/master/scripts/phylosift) were used to select representative eigenvectors in the edge PCA (corresponding to branches of the reference tree) for graphical representation in a 2-D plane: among the set of all eigenvectors directed in the same quadrant of the plane that correspond to branches of a same clade in the species tree, the eigenvectors with the longest norm were selected.

Alpha diversities were computed using phylodiversity metrics (McCoy & Matsen, 2013). The effect of variation in sequencing depth between samples was controlled for by taking average diversity estimates from 100 rarefying draws of the marker gene-matching reads. For each draw, 9,000 reads were considered, which corresponds to the lowest marker gene-matching read count among all samples.

Functional annotation of shotgun metagenomes

We used the metagenomic pipeline of the EBI (Mitchell et al., 2016) to scan reads from the metagenomes with the InterProScan tool for functional protein domain annotation (Mitchell et al., 2014). This analysis was repeated on contigs obtained with Ray assembler (Boisvert, Laviolette, & Corbeil, 2010); as too large a share of the read data was not assembled, we chose to use only the read-based results. Only seven out of the nine Western control datasets were amenable to this analysis as the two samples from (Hasan et al., 2014) have not been publicly released and notably lacked sequencing read quality data. Results are accessible by searching the BioProject accession ERP016024 on the EBI Metagenomics website (https://www.ebi.ac.uk/metagenomics/). We then performed LDA based on the relative abundances of the InterPro terms (normalized by each sample's total annotated

read count [Table S1]) to compare HGs to Western controls (Table S5), using a custom R script (http://github.com/flass/microbiomes/tree/master/scripts/interpro/lda_functional.r).

To assess the enrichment of particular biological systems or processes in the different subsistence strategy groups, biological processes that were represented by best-ranking functional terms in the LDA, including pantothenate (vitamin B5) biosynthesis, Coenzyme A related metabolism and urease activity (listed Table S6), had the LDA scores of all their dependent terms compared to those of a control high-ranking process (ribosome). Presence of pantothenate biosynthesis pathway in *Heamophilus* spp. reference genomes was investigated by browsing the Interpro database (www.ebi.ac.uk/interpro, last accessed 11 October 2016).

Results and Discussion

Study design

To circumvent the problem of lack of replication in previous studies on the oral microbiome of hunter-gatherers (HG), we set up a design analyzing three pairs of HG populations and their traditional farmer (TF) neighbors. The three HG populations are the Batak, Aeta and Agta, all members of the 'Negrito' group who are believed to be predominantly descended from the first humans to have settled in the Philippines (Lipson et al., 2014). They live in close proximity with the TF populations, Tagbanua, Zambal and Casigurani, respectively, who are all descendants of a later wave of settlement (Cariño, 2012). The geographic distances between the locations occupied by the pairs of populations range from 1 to 10 km, (Fig. 1; Table S1).

Food exchange between HGs and TFs is common, with up to 50% of the HGs' meals nowadays including rice (Page et al., 2016). Despite this, the two populations maintain distinct diets. HGs are foragers, i.e. still largely relying on fishing, hunting, and gathering (honey, leaves and wild fruits,

seeds and tubers; detailed records for Agta Table S7), whereas TFs rely on a traditional farming subsistence strategy, which in the Philippines is mainly based on cultivated rice and vegetables and excludes forest products (Bamberg-Migliano, personal observations). Pairs of HG and TF populations live in close enough proximity to likely be exposed to similar environmental sources of microbes, but their lifestyles differ substantially: the HGs usually live in lean-tos (without walls), while TFs live in houses with walls; HGs do not brush their teeth, while TFs go to school and receive education relating to oral prophylaxis, and usually have access to toothpaste and a brush (Bamberg-Migliano, personal observations). This setting offers an experimental design of three independent replicates with a comparable level of genetic and ecological differentiation between the populations within pairs, so that any systematic difference in the microbiome species composition between the two groups of populations should be primarily driven by the difference in subsistence strategies and lifestyle.

For each of those populations, we sampled saliva from four individuals in good oral health from which we deep-sequenced the whole extracted DNA, yielding between 167 to 662 million reads per sample, of which 77.0 to 94.7% could be assigned to the human genome (resulting in 2x-20x depth) (Table S1).

Genetic differentiation and admixture between human populations

We explored the genetic structure of the human populations using a robust probabilistic framework suitable for low and variable sequencing depth (Fumagalli, 2013; Skotte, Korneliussen, & Albrechtsen, 2013). Principal component analysis (PCA) and admixture analysis cluster together all individuals from the three TF populations, in accordance with their recent common ancestry (Fig. S2 and S3). However, individuals from the foraging Batak population cluster together with the TFs, while the two other foraging populations form clusters of their own. Only when considering the fourth principal component (PC) of the PCA, or an admixture model with at least four clusters, do the Batak

form a cluster of their own that also includes one Tagbanua individual (Fig. S2 and S3). This inference is in line with previous findings based on SNP chips and larger samples (Migliano et al., 2013) and might be explained by the drastic reduction in population size of the Batak down to 300 individuals in recent times (Scholes et al., 2011). Following this reduction in population size, they developed closer contact with their Tagbanua neighbors, including food trading and occasional marriages (Cariño, 2012), which might have mediated sufficient genetic admixture for them to become more closely related to the farmers.

To ensure that uneven sequencing depth across samples did not affect our estimates of genetic relatedness, we additionally performed a PCA on a sample of the data chosen to equalize the individual population depth. This analysis on the resampled data led to similar patterns of population structure (data not shown) and the principal components did not differ statistically from those obtained from the entire dataset (Procrustes analysis, permutation test, $p \leq 0.0001$) (see Supplementary Methods).

Relative abundances of core oral microbial taxa

The remainder of the reads not mapping to the human genome were used to characterize the composition of oral microbiomes. As an external reference, we included nine methodologically comparable WGS metagenomes of saliva-derived microbiomes from the Human Microbiome Project (HMP) (The Human Microbiome Project Consortium, 2012) and another initiative (Hasan et al., 2014) generated from North-American individuals, hereafter referred to as Westerners. We also attempted to incorporate salivary microbiomes extracted from exome sequencing of South-African HGs (Kidd et al., 2014), but the read depth of those samples was too low to justify their inclusion in the analyses. While we acknowledge that differences in sample collection procedures and sequencing batches may bias the comparison of metagenomes from different datasets, previous studies using datasets of

different origins found consistently more similar community compositions between HG groups than between HG and Westerner populations (Clemente et al., 2015; Schnorr et al., 2014), suggesting that batch effects are less important than effects of lifestyle.

We first characterized the microbial composition of all samples using Phylosift (Darling et al., 2014), a pipeline robustly estimating the relative abundances of all lineages of the Tree of Life based on reads matching a dataset of 33 universally conserved marker genes and using a phylogenetic placement framework (Matsen et al., 2010), which accounts evenly for well-characterized clades and deep lineages with few known representatives.

Comparing the Phylosift profiles of our 24 samples of paired HG and TF populations with edge PCA, we found that the PCs of microbiome composition variation in this dataset were driven by differential abundances in widespread oral taxa, including *Veillonella, Streptococcus, Haemophilus, Neisseria, Prevotella*, and various lineages of Actinobacteria. The largest fraction of inter-individual variance in the relative abundance of these taxa (PC1 and PC2 accounting for 52% and 21% of variance, respectively) does not segregate individuals by population or subsistence strategy (Fig. S4 A, B). This suggests that individual factors dominate the major source of variation in oral microbiome composition. This individual noise could reflect the high variation in the oral community within a single host over time, due to regular clearance of microbes by salivary proteins and swallowing of saliva (Carpenter, 2013). Alternatively, individual differences in nutrition or social relationships involving close physical contact (Kort et al., 2014) may participate in shaping individual microbiomes. However, the following principal components (PC3 and PC4, accounting for 12% and 8% of variance, respectively) result in the separation of the populations by geographic location and subsistence strategy (Fig. S4 C, D). This microbiome composition gradient becomes even more evident when the group of individuals with a Western diet is included in the analysis, as TF populations appear as intermediates between HGs and Western Controls (WCs) (Fig. 2 A, B). The addition of Phylosift

342 estimates of microbiome composition from WC samples to the analysis results in very similar edge
343 PCA plans, regarding the distribution of samples in relation to the vectors of differential abundance
344 (Fig. 2 C; Fig. S4 E, F; Fig. S5 C), and regarding the amount of variance they represent (PC 1-4
345 respectively account for 47, 18, 14 and 8% of the variance for the 33-sample dataset). Thus, we chose
346 to present the following results in the context of the 33-sample meta-analysis that includes the WC
347 samples; all corresponding graphics and numerical results for the 24-sample analysis – all qualitatively
348 equivalent to those presented below – are available online on the Figshare website at:
349 https://doi.org/10.6084/m9.figshare.5213041.v1.

350 **Taxonomic gradients reflect geography, host genetics, and subsistence strategy**

351 A first gradient opposing enrichment in *Prevotella* and *Streptococcus* against *Veillonella* and
352 Actinobacteria separates the microbiomes of the three HG populations along the third PC axis (Fig. 2
353 A, C, E). This could indicate an effect of the local environment, or be linked to genetic differences in
354 the hosts. To distinguish between these hypotheses, we used the reconstructed host genomes
355 associated to the microbiomes to test for correlation of host genetic background and microbiome
356 composition. We computed inter-individual distances in three ways: based on host genotypes (see Sup.
357 Methods); based on their geographic location at time of sampling; and based on the multivariate space
358 depicting their microbiome composition variation. No correlation was observed between Euclidean
359 distances in the full microbiome space and either the host genetic distances or the geographic distances
360 (Mantel test, p-values of 0.53 and 0.62, respectively). However, when considering projections of this
361 multivariate microbiome space on each of its PC, we recovered a trend for an association between
362 partial microbiome distances from the PC3 projection and host genetic distances (Mantel test, $p = 0.078$), as well as a strongly significant correlation with geographic distances (Mantel test, $p = 0.001$).
364 No other PC-projected distances correlated significantly with this factor (Table S2). Host genetics and
365 geography are largely collinear (Mantel test, $p = 0.02$); and after controlling for geography, the

correlation between genetic distances and the microbiome-derived PC3 does not remain statistically significant (partial Mantel test, $p = 0.140$). However, after controlling for host genetics, the correlation between geography and microbiome-derived PC3 only slightly decreases (partial Mantel test, $p = 0.002$). Taken together, this suggests that host genetic variation cannot explain microbiome variation on its own, but that the association between geography and microbiome make-up is more robust and likely causal i.e. the local environment, but not host genetics, is likely shaping the composition of the oral microbiome.

This is illustrated by the clustering pattern of samples by geographic origin on PC3 (Fig. 2F). Two out of the three pairs of populations of HGs and farmers share the same mean coordinate on PC3 (Fig. 2C) (Batak vs. Tagbanua and Aeta vs. Zambal; t-tests p-values of 0.91 and 0.52, respectively). The last pair, however, Agta and Casigurani, shows a marked differentiation on this axis (t-test, $p < 0.005$). This could be explained by the fact that while Agta and Casigurani live in close geographic proximity, their respective villages are separated by an inlet of the sea (Fig. 1), which may contribute to differences in the environments experienced by the two populations. Conversely, the very similar pattern of enrichment in streptococci and *Prevotella* observed for the Batak and Tagbanua could reflect that they often live in the actual same village (Fig. 1), and engage in far more frequent social and genetic exchanges than the other two pairs of populations (Cariño, 2012).

Another composition gradient following the fourth PC axis segregates the samples according to their subsistence strategy (Fig. 2D), with forager populations enriched in *Neisseria* spp. of the *N. lactamica* / *N. meningitidis* / *N. cinerea* group and farmers enriched in *Haemophilus* spp. of the *H. influenzae* / *H. haemolyticus* / *H. aegyptus* group (Fig. 2B, C). This gradient along PC4 appears to be the best way to segregate the subsistence strategies in our sample, as it constitutes the main contributing vector in a linear discriminant analysis of the principal components (DAPC) (Jombart et al., 2010), which results

389 in a very similar projection (Fig. S6), with significant separation between subsistence strategies (Pilai
390 test, $p < 0.030$).

391 The apparent enrichment of opportunistic pathogens including *N. meningitidis* and *H. influenzae* could
392 be interpreted as being indicative of poor health. However, these species are ubiquitous in the healthy
393 human oral cavity (Costalonga & Herzberg, 2014) and those detected here were likely commensal
394 strains. To further examine this hypothesis, we searched for genes specifically encoding *N.*
395 *meningitidis* and *H. influenzae* capsular polysaccharides, which are required for virulence. We found
396 limited evidence of their presence in any of the metagenomic assemblies (Supplementary Methods;
397 Table S3, Table S4). This suggests only commensal *Neisseria* spp. and *Haemophilus* spp. that lack
398 established virulence factors colonized the oral cavities of the studied individuals.

399 **Species considered pathogenic discriminate between subsistence strategies**

400 Because increases in the abundance of a few key species can lead to disease (Chen et al., 2015), we
401 also examined fine variation in abundances for all taxa, including rare ones. To do so, we used an
402 alternative method of classification of metagenomic sequences, Kraken, which relies on the finding of
403 exact matches between the metagenomic reads and a large database of complete genomes (Wood &
404 Salzberg, 2014). This method not only provides highly accurate classification of reads, but also makes
405 use of the total information from the WGS dataset and thus provides the best possible estimate of the
406 relative abundances of taxa. We used a linear discriminant analysis (LDA) and LDA effect size
407 (LefSe) (Segata et al., 2011) to find the species with most markedly contrasting abundances across the
408 three lifestyles. The top discriminating taxa included a number of species previously associated with
409 periodontal disease (Chen et al., 2015; Torrungruang et al. 2015): *Prevotella intermedia,*
410 *Porphyromonas gingivalis, Treponema denticola, Tannerella forsythia, Aggregatibacter*
411 *actinomycetemcomitans* and *Eubacterium nodatum* were associated with foraging and farming

412 subsistence strategies (Fig. 3). Intriguingly, despite carrying taxa associated with periodontal disease
413 at higher rates than the traditional farmers, the HGs in the Philippines are seemingly in far better oral
414 health (Bamberg-Migliano, pers. obs.).

415 This apparent lack of a negative effect of taxa previously associated with periodontal disease in
416 developed countries on the HGs' oral health suggests these might behave as commensals in HGs and
417 participate in the processing of foods specific to the foragers' diet. This has been previously
418 hypothesized for *Treponema* species in the gut of African and American HGs, which supposedly help
419 degradation of ligneous plant materials (Obregon-Tito et al., 2015; Schnorr et al., 2014). Such
420 commensals may have been present in the ancestral human oral cavity and secondarily lost in
421 populations with increased sanitation and lack of exposure to environmental sources. The only species
422 identified at a markedly higher prevalence in Westerners relative to the other groups is *Cutibacterium*
423 (formerly *Propionibacterium*) *acnes*, an organism mostly associated to skin follicles, but is also found
424 in the digestive tract. At the genus level *Bacteroides, Cutibacterium* and *Campylobacter* also show
425 enrichment in Westerners (significant under ANCOM tests), with the latter genus notably represented
426 by *C. concisus*, a species which abundance is up to 5% of a sample (Fig. 3). *C. concisus* has been
427 hypothesized to be associated with Crohn's disease (Kaakoush et al., 2014), an inflammatory bowel
428 disease with landmark high incidence in the developed world.

429 **Global shifts in species composition**

430 The gradient pattern of *Neisseria* spp. abundances (gradually higher in HGs than in TFs and WCs),
431 seen in the Phylosift-based edge PCA (Fig. 2E, G), is confirmed by Kraken analysis in several species
432 (*N. sicca*, *N. flavescens*, *N. gonorrhoeae,* FDR-corrected Wilcoxon rank-sum test p-value < 0.05) (Fig.
433 3). In contrast, the opposite gradient of *Heamophilus* spp. is not recovered by the Kraken analysis,
434 possibly due to the high variance of estimated abundances for WC samples, ranging between 0-18% of

435 the microbiome composition; the higher prevalence of the *Heamophilus* genus in TFs than in HGs is
436 however confirmed by the Kraken analysis (supplemental data online at:
437 https://doi.org/10.6084/m9.figshare.5213158.v1), indicating that the depletion of this taxon in HGs is a
438 robust feature.

439 The enrichment in *Neisseria* spp. in the oral microbiota of HGs versus Westerners was also observed
440 in a comparison between Westerners and South African HGs (Kidd et al., 2014), but Neisseriaceae did
441 not discriminate central African HGs from TFs (Nasidze et al., 2011), and were found depleted in
442 Amerindian HGs relative to Westerners (Clemente et al., 2015). Moreover, all three studies found an
443 enrichment of *Haemophilus* spp. in HGs' saliva. This suggests that the balance between these
444 proteobacterial lineages is an important feature discriminating subsistence strategies, but their relative
445 abundance may still be impacted by additional variables specific to each population. Similarly, an
446 enrichment in Prevotellaceae in HGs, as opposed to an enrichment in *Veillonella* in Westerners, was
447 previously reported (Clemente et al., 2015; Kidd et al., 2014); a similar contrasting microbial
448 enrichment is also observed in our marker gene-based (Phylosift) analysis, but is largely independent
449 of the foragers vs. Westerners divide, and rather characterizes the genetic diversity or geographical
450 location of populations on PC3 (Fig. 2C, E, G). This highlights the importance of controlling for such
451 confounding variables when identifying subsistence strategy-associated oral microbes. At the finer
452 level, as revealed by our WGS-based (Kraken) analysis, some species of *Prevotella* are indeed
453 enriched in foragers (*P. intermedia* and *P. shahii*), but another lineage (*P.* sp. HMSC077E09) is
454 enriched in Westerners, explaining the absence of lifestyle-discriminating signal at higher taxonomic
455 ranks.

456 **Increased diversity in the oral microbiomes of Hunter-Gatherers**

Using the Phylosift framework, we measured the diversity of microbes present in the salivary samples. This revealed a significantly larger phylogenetic diversity (PD) in HGs than in Westerners (t-test, $p < 0.02$), with the Filipino farmers occupying intermediate values (Fig. 4 A), mirroring the gradient observed in relative abundances of core oral taxa (Fig. 2 D). This difference remains significant (t-test, $p < 0.04$), when using balance-weighted PD (BPWD), a measure of diversity partially weighted by lineage abundance (scaling parameter $\Theta = 0.25$) that has been shown to be robust to variation in sampling depth (McCoy & Matsen, 2013). The trend in diversity is also maintained when rarefying all samples to the lowest depth in the dataset (9,000 marker gene-mapped reads), but at this point it loses statistical significance (t-test, $p = 0.07$). Interestingly, this trend emerges from a systematic increase in mean diversity between population of HGs relative to their paired TF population (Fig. 4 C), notwithstanding variations between geographical groups (Fig. 4 B).

An increased diversity in the oral microbiota is generally interpreted as evidence of poorer oral health (Costalonga & Herzberg, 2014). The mouth ecosystem is regularly cleared and re-colonized, and the opening of new niches in gingival crevices and cavities, as well as presence of carbohydrates, can lead to colonization by opportunist microbes and over-growth of commensal taxa into invasive ones (Costalonga & Herzberg, 2014). However, these observations concern individuals living a modern Western lifestyle. In the context of Philippines' HGs, an alternative hypothesis would be that higher diversity is linked to an extended commensal microbiota, possibly leading to gains of function. We therefore investigated in more details what differentiates the taxonomic and functional structures of the microbiomes of each subsistence strategy group.

Functional analysis reveals potential adaptations to diet

Species classification may prove a limited predictor of microbial community function due to phenotypic diversity of bacterial strains within the same species (Zhu, Sunagawa, Mende, & Bork,

480 2015). We thus used InterProScan to directly annotate metagenomic reads with biochemical functions.
481 From the total of 701,201,172 submitted reads (559,190,367 from the 24 Philippines samples),
482 242,348,233 coding sequences (136,537,593) were predicted, out of which 76,272,138 (50,384,528)
483 had a functional signature match to InterPro, covering together 11,307 unique functional terms
484 (detailed results accessible at https://www.ebi.ac.uk/metagenomics/projects/ERP016024). We first
485 applied PCA to explore the structure of the functional variation within our dataset. We observed that
486 neither subsistence strategy groups nor populations are well separated along the first six principal
487 components (together accounting to 80% of total variance), indicating that there is no marked
488 functional differentiation between those groups (Fig. S7). However, only a few functions with
489 significant differences abundant functions could still result in relevant ecological differences.

490 We thus applied LDA to this functional profile, searching for terms that discriminated HGs from
491 Westerners (Table S5). Amongst the top 95 (top 1%) discriminant annotations, we found several terms
492 relating to a few pathways: ribosome structure (11 in the top 1% out of 135 annotations), urease
493 activity (2/9 in top 1%), pantothenate (vitamin B5) and coenzyme A (CoA) biosynthesis (3/9 in top
494 1%) and CoA-dependent lipid metabolism (5/60 in top 1%) (Table S6). The directions of the
495 imbalances of ribosomal protein-coding sequences were randomly distributed (6 enriched in foragers,
496 5 enriched in Westerners in the top 1%; 60 and 75 in total), indicating that, when considered globally,
497 ribosome function is evenly distributed, as expected. In contrast, urease annotations were consistently
498 enriched in Westerners (8/9 of all annotations), and CoA-related annotations were consistently
499 enriched in foragers (all of biosynthesis-related annotations and 40/60 of the lipid metabolism-related
500 annotations, including all in the top 1% discriminant ones).

501 Microbiomes from Westerners, and TFs to a lesser extent, were found to be enriched in metagenomic
502 reads associated with urease function. Reads annotated for this function were mostly assigned to the
503 *Haemophilus* genus (Sup File S2), consistent with our taxononomic abundance-based analysis and

504 with the established ureolytic function of this lineage (Burne & Marquis, 2000). This enzymatic
505 pathway leads to the alkalinogenic release of ammonia, a reaction known to help buffer dental
506 biofilms against acidification. A drop in pH typically occurs when saccharolytic bacteria rapidly
507 degrade free sugars into acidic compounds, promoting tooth demineralization and favoring the growth
508 of cariogenic bacteria (Liu, Nascimento, & Burne, 2012; Reyes et al., 2014). The reduced abundance
509 of *Haemophilus* in HGs' saliva might therefore be expected to lead to dental plaque acidification, and
510 the development of oral diseases like caries. However, this would also require the presence of
511 acidogenic bacteria and more crucially their sugar substrates, which the hunter-gatherer diet is unlikely
512 to provide, as can be seen for the Agta, for whom extensive diet data have been collected (Table S7).
513 This is more likely to happen to the TFs and Westerners, whose diets are richer in starch and
514 processed sugars (Britten et al., 2012). It has been shown in synthetic oral communities repeatedly
515 exposed to pH drops that aciduric species including *Veillonella* spp. increased in frequency and
516 excluded *Neisseria* spp. (Bradshaw & Marsh, 1998), a pattern reminiscent of our observations (Figure
517 2).

518 Conversely, we observed an opposite gradient with highest prevalence in foragers of vitamin B5
519 biosynthetic pathway-associated metagenomic reads. This indicates that microbes autotrophic for this
520 vitamin are more successful at colonizing the mouths of HGs and to a lesser extent TFs. This same
521 trait has been previously observed as the most marked genomic difference between *Campylobacter*
522 spp. colonizing guts of cattle versus poultry. The frequent absence of genes in the vitamin B5
523 biosynthesis pathway in chicken-associated strains was suggested to reflect their diet of vitamin B5-
524 rich cereals and grains, as opposed to the grass-based cattle diet (Sheppard et al., 2013). Similarly, the
525 difference we observe may reflect the abundance of this essential nutrient in processed food present in
526 the Western diet, as opposed to its scarcity in food consumed by populations from the Philippines.
527 According to daily records of food consumed in Agta camps and in the general American population

528 (see Supplementary Methods), Americans and Agta eat food with globally comparable concentrations
529 of vitamin B5, but Americans have much larger daily portions (Table S8), and hence Westerners
530 consume greater quantities of vitamin B5. This discrepancy in daily ingested quantities of vitamin B5
531 may result in a different availability of this vitamin in the saliva of each group, which could have
532 impacted the profile of microbes colonizing their oral cavities.

533 The relative lack of vitamin B5 in the Philippines foragers' diet could select for microbes that are able
534 to synthesize it *de novo*. Such selective pressure on the oral microbiome may explain some of the
535 taxonomic signatures we found to be associated with subsistence strategies. Notably, *Haemophilus*
536 spp., which we showed to be the main bacterial lineage depleted in the HGs' microbiomes (Fig. 2),
537 have genomes devoid of the relevant vitamin B5 biosynthesis pathway (see Supplementary Methods),
538 suggesting *Haemophilus* spp. are counter-selected in the HGs' saliva. Conversely, the diet of
539 Westerners, which provides them with a greater intake of vitamin B5, may have allowed the
540 colonization of the mouth of certain individuals by bacteria auxotrophic for this nutrient, such as
541 *Haemophilus* spp.. An increased abundance of this particular lineage, with its urease activity able to
542 counter acidic bursts, could in turn have geared the microbiome towards an adaptive response to the
543 Westerner's acidogenic sugar-rich diet.

544

545 **Conclusion**

546 Despite high inter-individual variability and a strong impact of the geographic location of host
547 populations on oral microbiome composition, we were able to recover consistent differentiation
548 associated to subsistence strategies thanks to the replicated design of the study. Key signatures of
549 subsistence strategies include shifts in species distribution including relative abundance of core species
550 such as *Neisseria* spp. vs. *Haemophilus* spp.. This suggests that the hunter-gatherer and traditional

551 farmer diets in themselves, or closely associated ecological or socio-economic factors, are significant
552 drivers of differentiation in saliva.

553 Our results paint an interesting picture of the oral microbiome in HGs in terms of health and disease.
554 Oral microbiomes from HGs were significantly more diverse than those from TFs or Westerners, as
555 was found previously in distant hunter-gatherer populations (Clemente et al., 2015; Nasidze et al.,
556 2011). While high diversity of microbiomes in the oral cavity has been associated with disease
557 (Griffen et al., 2012), some of this diversity is likely to be adaptive to their forager diet as possibly
558 illustrated by the presence of species involved in the degradation of ligneous material such as
559 *Treponema* spp. (Obregon-Tito et al., 2015; Schnorr et al., 2014). While the HG microbiomes
560 comprise an excess of species that have been shown to be associated to oral disease, it is unclear to
561 what extent these species cause disease in HGs. Indeed, all subjects enrolled in this study were
562 apparently in good oral health and HGs in the Philippines tend to have fewer caries than the TFs
563 (Bamberg-Migliano, pers. obs.). It is possible that the species complex associated to gingivitis and
564 periodontitis might be part of the healthy microbiota of the HGs' buccal cavity, with pathogenic
565 strains only selected in populations subsisting on a diet richer in starch and refined sugar.

566

Acknowledgements

The authors acknowledge support from the European Research Council (ERC) (grant ERC260801 – BIG_IDEA), and the National Institute for Health Research University College London Hospitals Biomedical Research Centre. FL was additionally supported by the UK Medical Research Council (grant 412 MR/N010760/1). MS was supported by an Institut Pasteur-Cenci Bolognetti fellowship. MF was supported by a Human Frontiers (HFSP) fellowship. L.S. was supported by a PhD scholarship from EPSRC (EP/F500351/1) and the Reuben Centre for Paediatric Virology and Metagenomics. MD was supported by Leverhulme Trust (grant RP2011-R-045 to ABM), and by French National Research

Agency (ANR) grant Labex IAST. We are grateful for comments on the manuscript by Mark Achtman. Finally, we wish to thank the National Commission on Indigenous Peoples (NCIP) and the anonymous donors of saliva samples that were used in the study.

Author Contributions

MS, MGT and FB designed the study; ABM provided samples; MS did the molecular analyses; MS, FL, MF, LS and FB performed the bioinformatics and computational analyses; MS, FL, ABM and FB wrote the paper; ABM, MD, CW and MGT collected and modeled diet information; all authors read and commented on the manuscript.

Conflict of Interests statement

We declare no conflict of interest.

Data availability

- Microbial metagenomic read datasets (after trimming, quality filtering and removal of Kraken-assigned human reads): ENA (www.ebi.ac.uk/ena), BioSample accessions ERS1202862—ERS1202885.

- Results of the EBI metagenome analysis pipeline: EBI Metagenomics (https://www.ebi.ac.uk/metagenomics), project accession ERP016024.

- Output of Kraken (tables of taxon abundances): Figshare, doi: 10.6084/m9.figshare.5213158

- Output of guppy (placement edge difference data matrix, phylodiversity estimates and edgePCA projections): Figshare,

582 doi: 10.6084/m9.figshare.5213041

583 - Output of LDA and PCA on Interproscan functional classification: Figshare

584 doi: 0.6084/m9.figshare.5464423

585 - Sup. Files S1 and S2: Figshare,

586 doi: 10.6084/m9.figshare.5545195

587 - Output of Phylosift (placement and summary files): DRYAD,

588 doi:10.5061/dryad.sn00q.

589 **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. doi:10.1038/ng.2536

Bocquet-Appel, J.-P. (2011). When the world's population took off: the springboard of the Neolithic Demographic Transition. *Science (New York, N.Y.)*, *333*(6042), 560–561. doi:10.1126/science.1208880

Boisvert, S., Laviolette, F., & Corbeil, J. (2010). Ray: Simultaneous Assembly of Reads from a Mix of High-Throughput Sequencing Technologies. *Journal of Computational Biology*, *17*(11), 1519–1533. doi:10.1089/cmb.2009.0238

Bradshaw, D. J., & Marsh, P. D. (1998). Analysis of pH-driven disruption of oral microbial communities in vitro. *Caries Research*, *32*(6), 456–462.

Britten, P., Cleveland, L. E., Koegel, K. L., Kuczynski, K. J., & Nickols-Richardson, S. M. (2012). Impact of typical rather than nutrient-dense food choices in the US Department of Agriculture Food Patterns. *Journal of the Academy of Nutrition and Dietetics*, *112*(10), 1560–1569. doi:10.1016/j.jand.2012.06.360

Burne, R. A., & Marquis, R. E. (2000). Alkali production by oral bacteria and protection against dental caries. *FEMS Microbiology Letters*, *193*(1), 1–6. doi:10.1111/j.1574-6968.2000.tb09393.x

Cariño, J. K. (2012). Country Technical Notes on Indigenous People's Issues: Republic of the Philippines. Retrieved from https://www.ifad.org/documents/10180/0c348367-f9e9-42ec-89e9-3ddbea5a14ac

Carpenter, G. H. (2013). The Secretion, Components, and Properties of Saliva. *Annual Review of Food Science and Technology*, *4*(1), 267–276. doi:10.1146/annurev-food-030212-182700

Chen, H., Liu, Y., Zhang, M., Wang, G., Qi, Z., Bridgewater, L., … Pang, X. (2015). A Filifactor alocis-centered co-occurrence group associates with periodontitis across different oral habitats. *Scientific Reports*, *5*. doi:10.1038/srep09053

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. *Science Advances*, *1*(3), e1500183–e1500183. doi:10.1126/sciadv.1500183

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. *The American Journal of Clinical Nutrition*, *81*(2), 341–354.

Costalonga, M., & Herzberg, M. C. (2014). The oral microbiome and the immunobiology of periodontal disease and caries. *Immunology Letters*, *162*(2, Part A), 22–38. doi:10.1016/j.imlet.2014.08.017

Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A., Bik, H. M., & Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*, *2*, e243. doi:10.7717/peerj.243

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H., … Wade, W. G. (2010). The Human Oral Microbiome. *Journal of Bacteriology*, *192*(19), 5002–5017. doi:10.1128/JB.00542-10

Dufour, A.-B., & Dray, S. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, *22*(i04). Retrieved from https://www.jstatsoft.org/article/view/v022i04

Fumagalli, M. (2013). Assessing the Effect of Sequencing Depth and Sample Size in Population Genetics Inferences. *PLoS ONE*, *8*(11), e79667. doi:10.1371/journal.pone.0079667

Griffen, A. L., Beall, C. J., Campbell, J. H., Firestone, N. D., Kumar, P. S., Yang, Z. K., … Leys, E. J. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *The ISME Journal*, *6*(6), 1176–1185. doi:10.1038/ismej.2011.191

Hasan, N. A., Young, B. A., Minard-Smith, A. T., Saeed, K., Li, H., Heizer, E. M., … Colwell, R. R. (2014). Microbial Community Profiling of Human Saliva Using Shotgun Metagenomic Sequencing. *PLoS ONE*, *9*(5), e97699. doi:10.1371/journal.pone.0097699

Hunter, P. (2014). Pulling teeth from history: DNA from ancient teeth can help to yield information about our ancestors' health, diet and diseases. *EMBO Reports*, *15*(9), 923–925. doi:10.15252/embr.201439353

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. doi:10.1038/nature03001

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. doi:10.1186/1471-2156-11-94

Kaakoush, N. O., Castaño-Rodríguez, N., Day, A. S., Lemberg, D. A., Leach, S. T., & Mitchell, H. M. (2014). Campylobacter concisus and exotoxin 9 levels in paediatric patients with Crohn's disease and their association with the intestinal microbiota. *Journal of Medical Microbiology*, *63*(1), 99–105. doi:10.1099/jmm.0.067231-0

Kaplan, H., Thompson, R. C., Trumble, B. C., Wann, L. S., Allam, A. H., Beheim, B., … Thomas, G. S. (2017). Coronary atherosclerosis in indigenous South American Tsimane: a cross-sectional cohort study. *Lancet (London, England)*, *389*(10080), 1730–1739. doi:10.1016/S0140-6736(17)30752-3

Kidd, J. M., Sharpton, T. J., Bobo, D., Norman, P. J., Martin, A. R., Carpenter, M. L., … Henn, B. M. (2014). Exome capture from saliva produces high quality genomic and metagenomic data. *BMC Genomics*, *15*(1), 262. doi:10.1186/1471-2164-15-262

Kort, R., Caspers, M., Graaf, A. van de, Egmond, W. van, Keijser, B., & Roeselers, G. (2014). Shaping the oral microbiota through intimate kissing. *Microbiome*, *2*(1), 41. doi:10.1186/2049-2618-2-41

Kultima, J. R., Sunagawa, S., Li, J., Chen, W., Chen, H., Mende, D. R., … Bork, P. (2012). MOCAT: A Metagenomics Assembly and Gene Prediction Toolkit. *PLoS ONE*, *7*(10), e47656. doi:10.1371/journal.pone.0047656

Kuo, L.-C., Polson, A. M., & Kang, T. (2008). Associations between periodontal diseases and systemic diseases: a review of the inter-relationships and interactions with diabetes, respiratory diseases, cardiovascular diseases and osteoporosis. *Public Health*, *122*(4), 417–433. doi:10.1016/j.puhe.2007.07.004

Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., … Venter, J. C. (2007). The Diploid Genome Sequence of an Individual Human. *PLoS Biol*, *5*(10), e254. doi:10.1371/journal.pbio.0050254

Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen, K., & Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, *25*(15), 1966–1967. doi:10.1093/bioinformatics/btp336

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. *Nature Communications*, *5*, 4689. doi:10.1038/ncomms5689

Liu, Y.-L., Nascimento, M., & Burne, R. A. (2012). Progress toward understanding the contribution of alkali generation in dental biofilms to inhibition of dental caries. *International Journal of Oral Science*, *4*(3), 135–140. doi:10.1038/ijos.2012.54

Lloyd-Price, J., Abu-Ali, G., & Huttenhower, C. (2016). The healthy human microbiome. *Genome Medicine*, *8*, 51. doi:10.1186/s13073-016-0307-y

Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial Ecology in Health and Disease*, *26*. doi:10.3402/mehd.v26.27663

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, *14*(2), 54–67. doi:10.1002/evan.20046

Marsh, P. D. (2003). Are dental diseases examples of ecological catastrophes? *Microbiology (Reading, England)*, *149*(Pt 2), 279–294. doi:10.1099/mic.0.26082-0

Marsh, P. D., Do, T., Beighton, D., & Devine, D. A. (2016). Influence of saliva on the oral microbiota. *Periodontology 2000*, *70*(1), 80–92. doi:10.1111/prd.12098

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., … Reich, D. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503. doi:10.1038/nature16152

Matsen, F. A., Kodner, R. B., & Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, *11*(1), 538. doi:10.1186/1471-2105-11-538

McCoy, C. O., & Matsen, F. A. (2013). Abundance-weighted phylogenetic diversity measures distinguish microbial community states and are robust to sampling depth. *PeerJ*, *1*, e157. doi:10.7717/peerj.157

Migliano, A. B., Romero, I. G., Metspalu, M., Leavesley, M., Pagani, L., Antao, T., … Kivisild, T. (2013). Evolution of the Pygmy Phenotype: Evidence of Positive Selection from Genome-wide Scans in African, Asian, and Melanesian Pygmies. *Human Biology*, *85*(1–3), 251–284. doi:10.3378/027.085.0313

Mira, A., Pushker, R., & Rodríguez-Valera, F. (2006). The Neolithic revolution of bacterial genomes. *Trends in Microbiology*, *14*(5), 200–206. doi:10.1016/j.tim.2006.03.001

Mitchell, A., Bucchini, F., Cochrane, G., Denise, H., Hoopen, P. ten, Fraser, M., … Finn, R. D. (2016). EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Research*, *44*(D1), D595–D603. doi:10.1093/nar/gkv1195

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., … Finn, R. D. (2014). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research*, *43*(D1), D213–D221. doi:10.1093/nar/gku1243

Morton, E. R., Lynch, J., Froment, A., Lafosse, S., Heyer, E., Przeworski, M., … Ségurel, L. (2015). Variation in Rural African Gut Microbiota Is Strongly Correlated with Colonization by Entamoeba and Subsistence. *PLoS Genetics*, *11*(11), e1005658. doi:10.1371/journal.pgen.1005658

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

Obregon-Tito, A. J., Tito, R. Y., Metcalf, J., Sankaranarayanan, K., Clemente, J. C., Ursell, L. K., … Lewis, C. M. (2015). Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Communications*, *6*, 6505. doi:10.1038/ncomms7505

Page, A. E., Viguier, S., Dyble, M., Smith, D., 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 Sciences*, *113*(17), 4694–4699. doi:10.1073/pnas.1524031113

Petersen, P. E. (2005). Priorities for research for oral health in the 21st century--the approach of the WHO Global Oral Health Programme. *Community Dental Health*, *22*(2), 71–74.

Quercia, S., Candela, M., Giuliani, C., Turroni, S., Luiselli, D., Rampelli, S., … Pirazzini, C. (2014). From lifetime to evolution: timescales of human gut microbiota adaptation. *Frontiers in Microbiology*, *5*, 587. doi:10.3389/fmicb.2014.00587

Reyes, E., Martin, J., Moncada, G., Neira, M., Palma, P., Gordan, V., … Yevenes, I. (2014). Caries-free subjects have high levels of urease and arginine deiminase activity. *Journal of Applied Oral Science: Revista FOB*, *22*(3), 235–240.

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*. doi:10.1038/ncomms4654

Scholes, C., Siddle, K., Ducourneau, A., Crivellaro, F., Järve, M., Rootsi, S., … Migliano, A. B. (2011). Genetic diversity and evidence for population admixture in Batak Negritos from Palawan. *American Journal of Physical Anthropology*, *146*(1), 62–72. doi:10.1002/ajpa.21544

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*, R60. doi:10.1186/gb-2011-12-6-r60

Sheppard, S. K., Didelot, X., Meric, G., Torralbo, A., Jolley, K. A., Kelly, D. J., … Falush, D. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. *Proceedings of the National Academy of Sciences*, *110*(29), 11923–11927. doi:10.1073/pnas.1305559110

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702. doi:10.1534/genetics.113.154138

The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. doi:10.1038/nature11234

Torrungruang, K., Jitpakdeebordin, S., Charatkulangkun, O., & Gleebbua, Y. (2015). Porphyromonas gingivalis, Aggregatibacter actinomycetemcomitans, and Treponema denticola / Prevotella intermedia Co-Infection Are Associated with Severe Periodontitis in a Thai Population. *PLOS ONE*, *10*(8), e0136646. doi:10.1371/journal.pone.0136646

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., … Wang, J. (2008). The diploid genome sequence of an Asian individual. *Nature*, *456*(7218), 60–65. doi:10.1038/nature07484

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., … Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344. doi:10.1038/ng.2906

Whitmore, S. E., & Lamont, R. J. (2014). Oral bacteria and cancer. *PLoS Pathogens*, *10*(3), e1003933. doi:10.1371/journal.ppat.1003933

Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), R46. doi:10.1186/gb-2014-15-3-r46

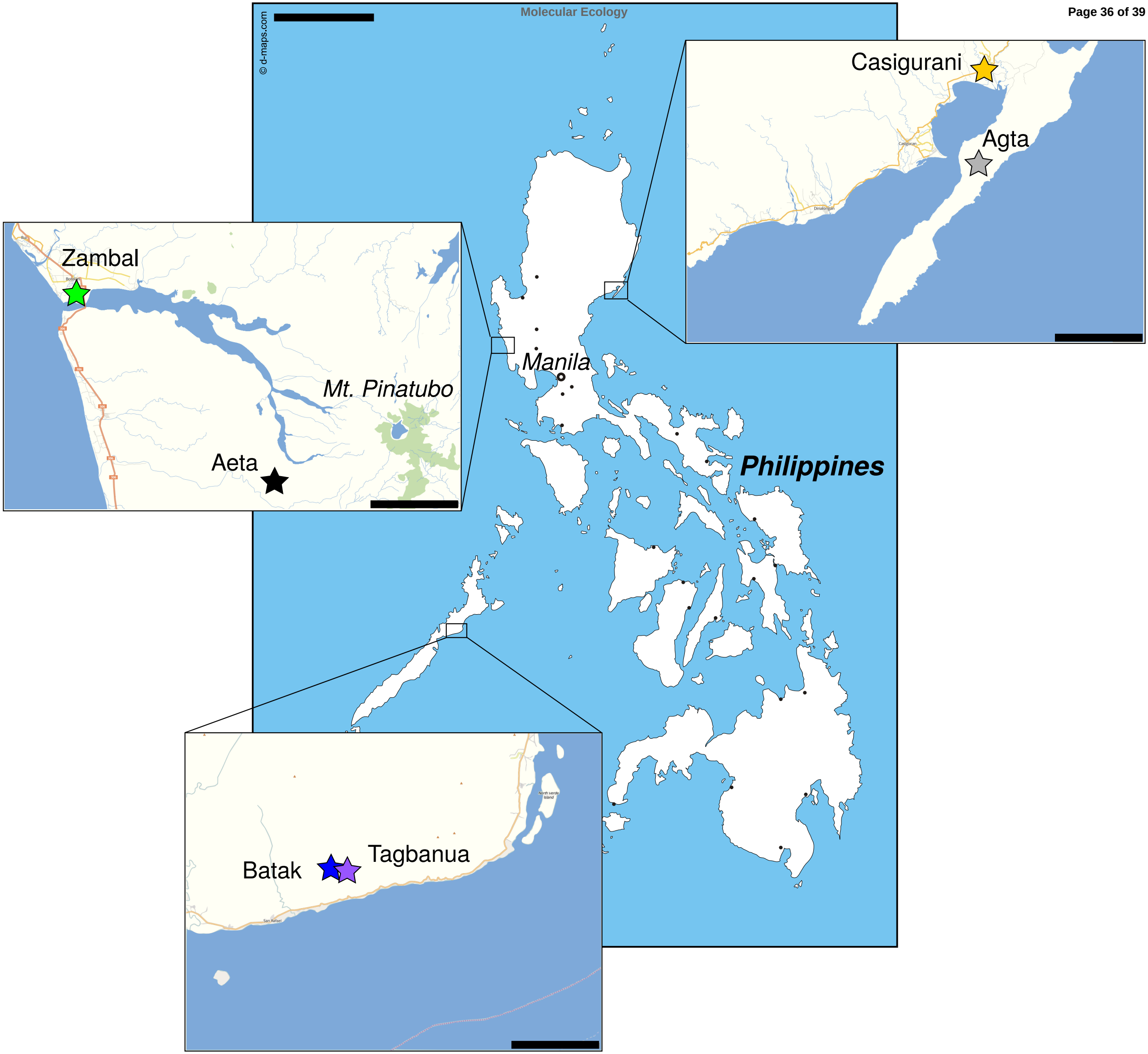
Yang, F., Zeng, X., Ning, K., Liu, K.-L., Lo, C.-C., Wang, W., … Xu, J. (2012). Saliva microbiomes distinguish caries-active from healthy human populations. *The ISME Journal*, *6*(1), 1–10. doi:10.1038/ismej.2011.71

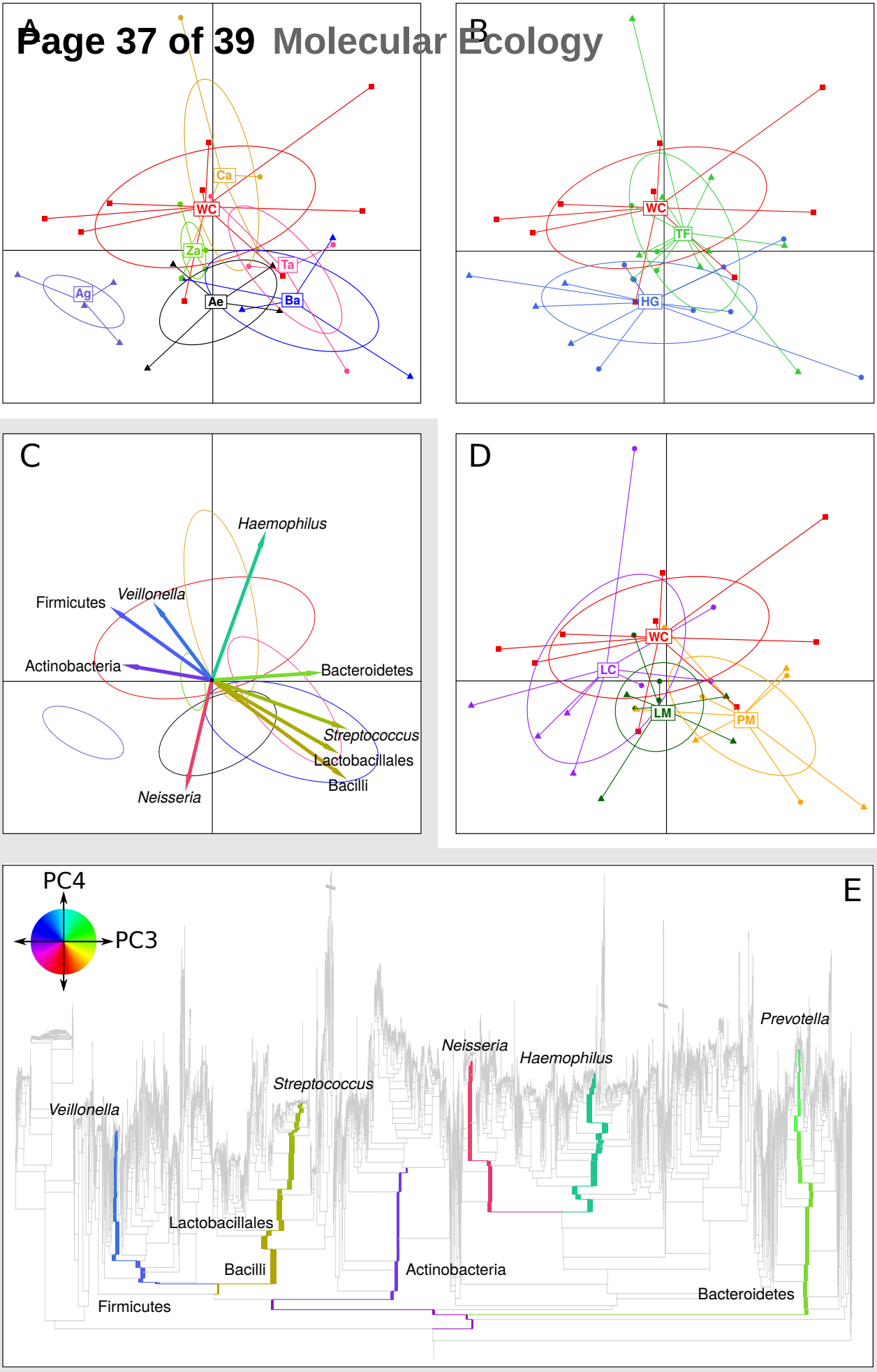
Zaura, E., Nicu, E. A., Krom, B. P., & Keijser, B. J. F. (2014). Acquiring and maintaining a normal oral microbiome: current perspective. *Frontiers in Cellular and Infection Microbiology*, *4*, 85. doi:10.3389/fcimb.2014.00085

Zhu, A., Sunagawa, S., Mende, D. R., & Bork, P. (2015). Inter-individual differences in the gene content of human gut bacterial species. *Genome Biology*, *16*, 82. doi:10.1186/s13059-015-0646-9

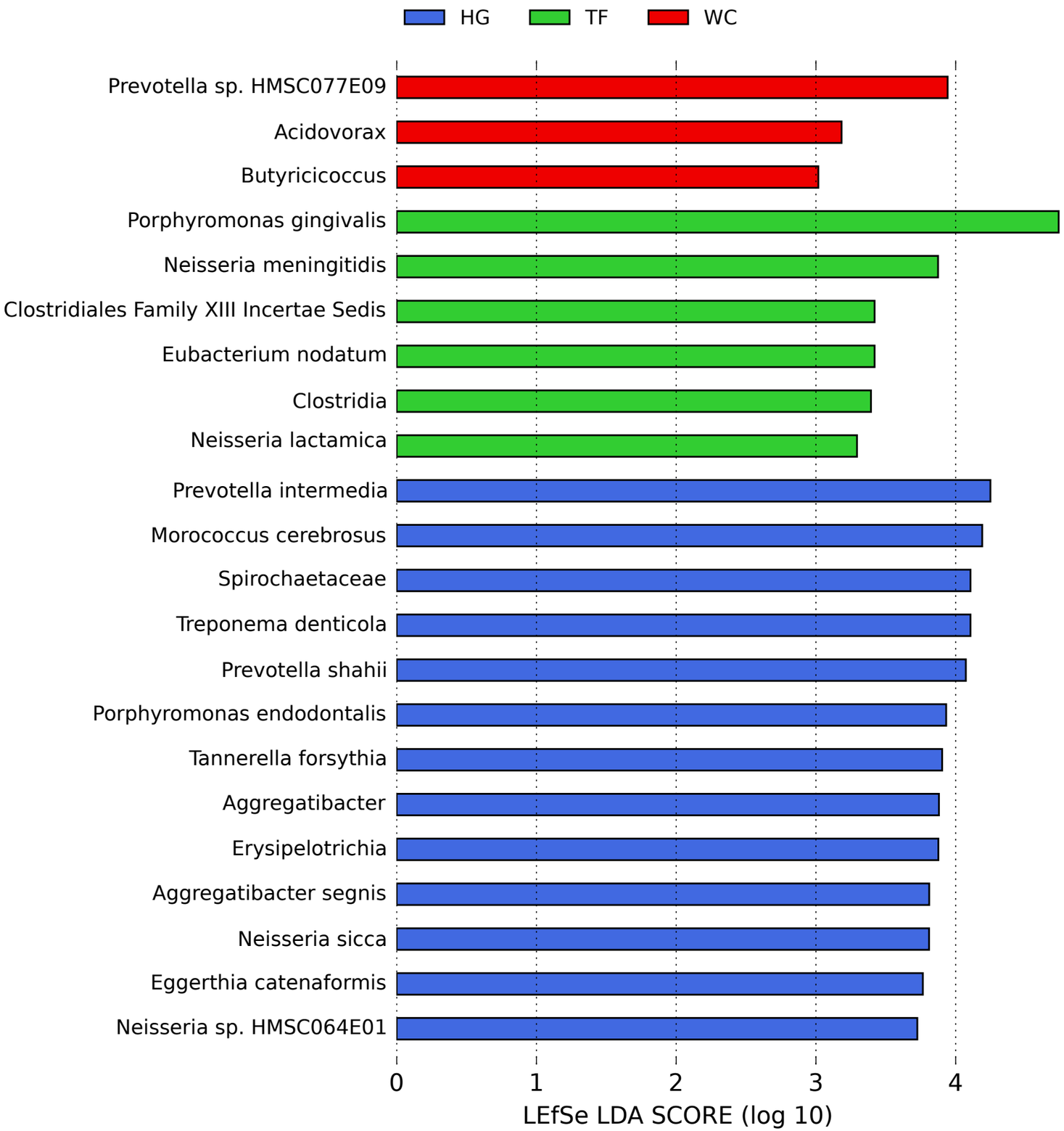
590

591

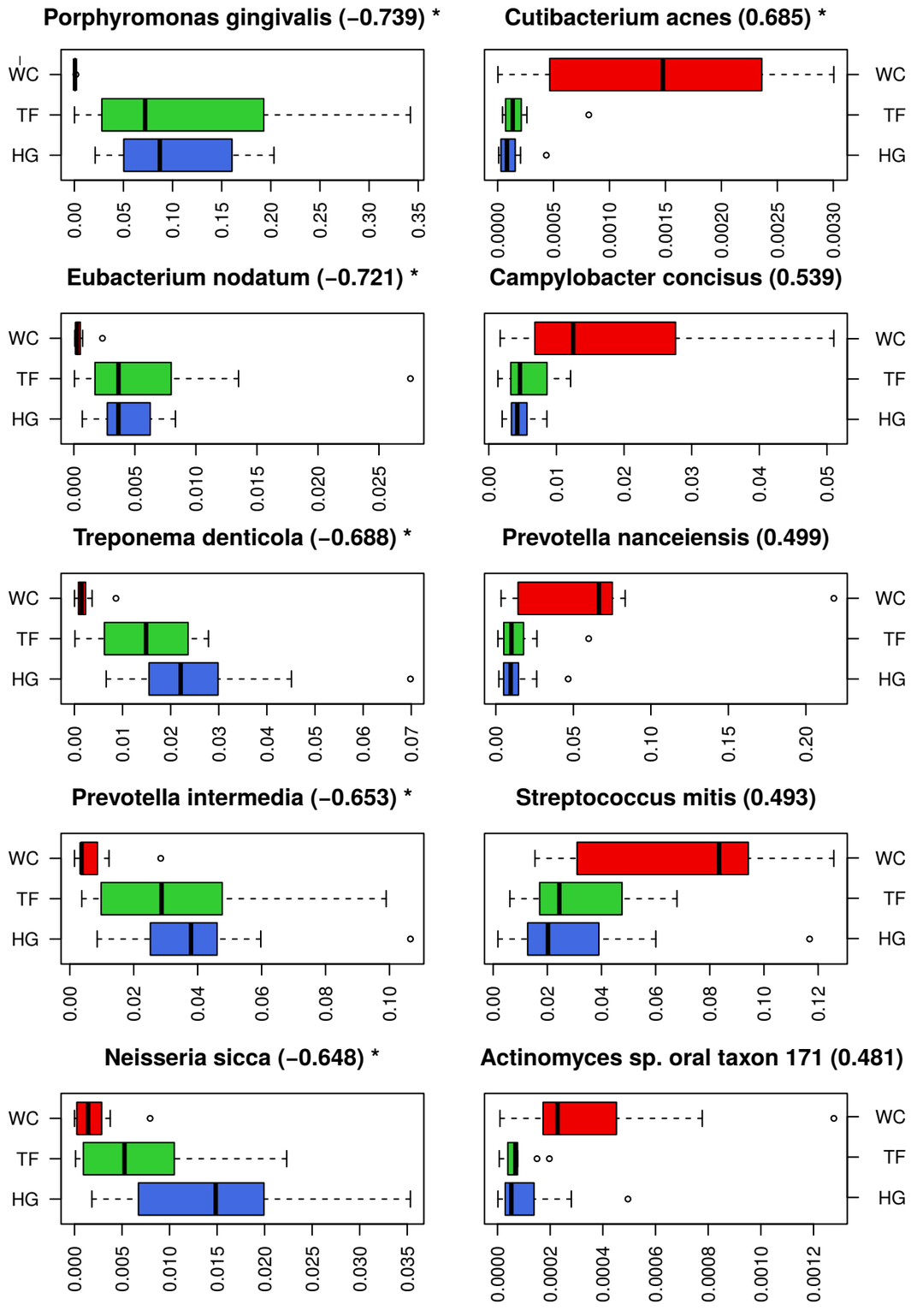

© d-maps.com
Casigurani
Agta
Zambal
Mt. Pinatubo
Aeta
Manila
Philippines
Batak
Tagbanua

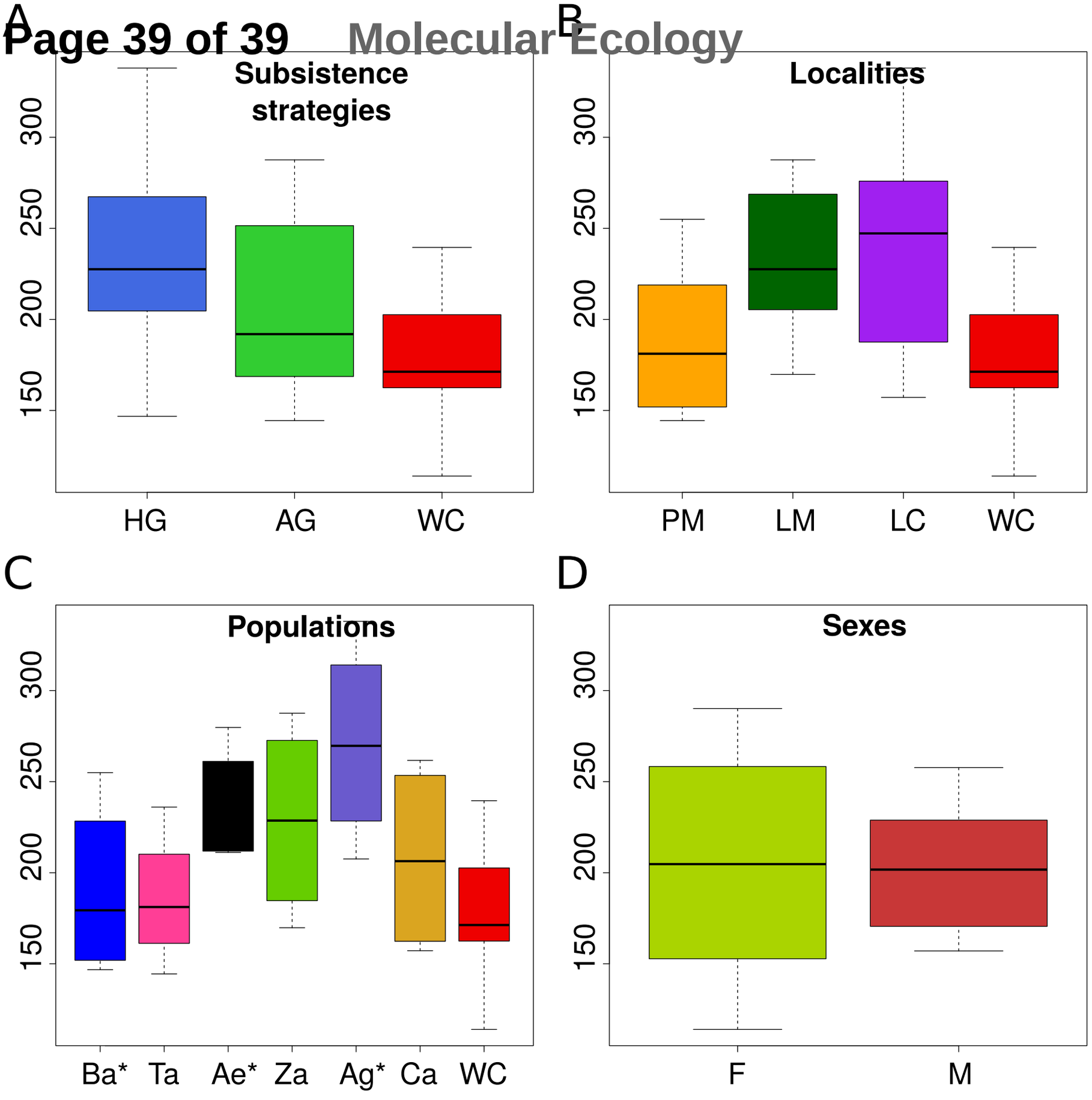

A
Ca
WC
Za
Ta
Ag
Ae
Ba
B
WC
TF
HG
C
Haemophilus
Firmicutes
Veillonella
Actinobacteria
Bacteroidetes
Streptococcus
Lactobacillales
Bacilli
Neisseria
D
WC
LC
LM
PM
E
PC4
PC3
Prevotella
Neisseria
Haemophilus
Streptococcus
Veillonella
Lactobacillales
Bacilli
Actinobacteria
Bacteroidetes
Firmicutes

A


HG
TF
WC
Prevotella sp. HMSC077E09
Acidovorax
Butyricicoccus
Porphyromonas gingivalis
Neisseria meningitidis
Clostridiales Family XIII Incertae Sedis
Eubacterium nodatum
Clostridia
Neisseria lactamica
Prevotella intermedia
Morococcus cerebrosus
Spirochaetaceae
Treponema denticola
Prevotella shahii
Porphyromonas endodontalis
Tannerella forsythia
Aggregatibacter
Erysipelotrichia
Aggregatibacter segnis
Neisseria sicca
Eggerthia catenaformis
Neisseria sp. HMSC064E01
0
1
2
3
4
LEfSe LDA SCORE (log 10)

B


Porphyromonas gingivalis (−0.739) *
Cutibacterium acnes (0.685) *
Eubacterium nodatum (−0.721) *
Campylobacter concisus (0.539)
Treponema denticola (−0.688) *
Prevotella nanceiensis (0.499)
Prevotella intermedia (−0.653) *
Streptococcus mitis (0.493)
Neisseria sicca (−0.648) *
Actinomyces sp. oral taxon 171 (0.481)
WC
TF
HG


A
Subsistence strategies
150
200
250
300
HG
AG
WC
B
Localities
150
200
250
300
PM
LM
LC
WC
C
Populations
150
200
250
300
Ba*
Ta
Ae*
Za
Ag*
Ca
WC
D
Sexes
150
200
250
300
F
M