

RESEARCH

Open Access



# Nepali oral microbiomes reflect a gradient of lifestyles from traditional to industrialized

Erica P. Ryu<sup>1</sup>, Yoshina Gautam<sup>2</sup>, Diana M. Proctor<sup>3</sup>, Dinesh Bhandari<sup>4,5</sup>, Sarmila Tandukar<sup>4,6</sup>, Meera Gupta<sup>1,7</sup>, Guru Prasad Gautam<sup>8</sup>, David A. Relman<sup>9,10,11</sup>, Ahmed A. Shibl<sup>2,12</sup>, Jeevan Bahadur Sherchand<sup>4</sup>, Aashish R. Jha<sup>2,12\*</sup> and Emily R. Davenport<sup>1,13\*</sup>

## Abstract

**Background** Lifestyle plays an important role in shaping the gut microbiome. However, its contributions to the oral microbiome remain less clear, due to the confounding effects of geography and methodology in investigations of populations studied to date. Furthermore, while the oral microbiome seems to differ between foraging and industrialized populations, we lack insight into whether transitions to and away from agrarian lifestyles shape the oral microbiota. Given the growing interest in so-called “vanishing microbiomes” potentially being a risk factor for increased disease prevalence in industrialized populations, it is important that we distinguish lifestyle from geography in the study of microbiomes across populations.

**Results** Here, we investigate salivary microbiomes of 63 Nepali individuals representing a spectrum of lifestyles: foraging, subsistence farming (individuals that transitioned from foraging to farming within the last 50 years), agriculturalists (individuals that have transitioned to farming for at least 300 years), and industrialists (expatriates that immigrated to the USA within the last 20 years). We characterize the role of lifestyle in microbial diversity, identify microbes that differ between lifestyles, and pinpoint specific lifestyle factors that may be contributing to differences in the microbiomes across populations. Contrary to prevailing views, when geography is controlled for, oral microbiome alpha diversity does not differ significantly across lifestyles. Microbiome composition, however, follows the gradient of lifestyles from foraging through agrarianism to industrialism, supporting the notion that lifestyle indeed plays a role in the oral microbiome. Relative abundances of several individual taxa, including *Streptobacillus* and an unclassified Porphyromonadaceae genus, also mirror lifestyle. Finally, we identify specific lifestyle factors associated with microbiome composition across the gradient of lifestyles, including smoking and grain sources.

**Conclusion** Our findings demonstrate that by studying populations within Nepal, we can isolate an important role of lifestyle in determining oral microbiome composition. In doing so, we highlight the potential contributions of several lifestyle factors, underlining the importance of carefully examining the oral microbiome across lifestyles to improve our understanding of global microbiomes.

**Keywords** Oral microbiome, Oral microbiota, Salivary microbiome, Lifestyle, Nepali populations

\*Correspondence:

Aashish R. Jha

jhaar@nyu.edu

Emily R. Davenport

emily.davenport@psu.edu

Full list of author information is available at the end of the article



© The Author(s) 2024, corrected publication 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### अध्ययन संक्षेप (Abstract in Nepali)

सारांश पृष्ठभूमि मानवको आन्द्रामा ब्याक्टेरिया, भाइरस, फंगस (दुसी), लगायत अरू विभिन्न प्रकारका सूक्ष्मजीवहरूका समुदाय पाईन्छन्, जसलाई माइक्रोबायोम भनिन्छ। आन्द्रामा हुने माइक्रोबायोमले मानवको पाचन, प्रतिरक्षा प्रणाली, र अन्य शारीरिक प्रक्रियाहरूमा योगदान पुर्याउने हुनाले यिनले मानव स्वास्थ्य र रोगको निर्धारण गर्नमा महत्वपूर्ण भूमिका खेल्छन्। त्यसैगरी मानव मुखमा पनि सूक्ष्मजीवहरूका समुदायहरू हुन्छन् जसले दाँत र गिजालाई स्वस्थ राख्नमा योगदान पुर्याउँछन्। जीवनशैलीले आन्द्राको माइक्रोबायोमलाई आकार दिन महत्वपूर्ण भूमिका खेल्छ। साथै, जडलमा निर्भर भएर प्रकृतिक जीवनशैली जिउने र औद्योगिकृत भएका मानव समुदायहरूको बीच मुखको माइक्रोबायोममा धेरै भिन्नता देखिन्छ। जडलमा निर्भर भएर प्रकृतिक जीवनशैली जिउने मानव समुदायहरूमा पाईने धेरै सूक्ष्मजीवहरू औद्योगिक जनसमुदायहरूमा मा पाईने छान्छासकेका छन् र यस्ता 'विलुप्त माइक्रोबायोम' विभिन्न रोगहरूका जोखिमका कारक हुन सक्ने हुँदा चासोका बिषय बन्न पुगेका छन्। तथापि, अहिलेसम्मका अनुसन्धानमा जीवनशैलीको परिवर्तन सँगै मुखको माइक्रोबायोममा के कस्ता परिवर्तनहरू आउँछन् भन्ने कुरा स्पष्ट हुन सकेको छैन। विभिन्न जीवनशैली ब्यथीत गर्ने मानिसहरू बीच जीवनशैलीको साथसाथै भौगोलिक भिन्नता पनि हुने गर्दछ। तसर्थ, मानिसहरूका मुखका माइक्रोबायोममा जीवनशैली र भूगोलको असरलाई छुट्याउन महत्वपूर्ण देखिएको छ।

परिणामहरू यहाँ, हामीले ६३ जना नेपाली व्यक्तिको थुक संकलन गरेर मुखका माइक्रोबायोमको अनुसन्धान गर्यौं। हाम्रो अध्ययनमा संलग्न भएका ब्यक्तिहरूले विभिन्न जीवनशैलीको प्रतिनिधित्व गर्छन्: जडलमा निर्भर भएर प्रकृतिक जीवनशैली जिउने, जडलबाट कृषितर्फ हालसालै परिवर्तन भएका (अन्तिम ५० वर्ष भित्र), कृषक (बिगतका ३०० वर्ष भन्दा बढी समयदेखि कृषिमा संलग्न भएका), र औद्योगिक (अन्तिम २० वर्ष भित्र संयुक्त राज्य अमेरिका मा आप्रवासन गरेका नेपालीहरू)। हामीले यस अध्ययनमा जीवनशैलीले मुखका माइक्रोबायोमको विविधतामा कसरी योगदान पुर्याउँछन् भन्ने कुराको विश्लेषण गरेका छौं, जीवनशैलीहरू बीच फरक पार्ने सूक्ष्मजीवहरूको पहिचान गरेका छौं, र माइक्रोबायोममा फरक पार्ने कारकहरूको पहिचान गरेका छौं।

प्रचलित दृष्टिकोणहरू विपरीत, भूगोललाई नियन्त्रण गर्दा, विभिन्न जीवनशैलीहरू बीच मुखको माइक्रोबायोममा हुने सूक्ष्मजीवहरूको सन्ख्या (अल्फा विविधता) मा धेरै फरक देखिएन, तर मुखको माइक्रोबायोमको संरचना भने जीवनशैलीको क्रमिक चरणसँग मिल्दोजुल्दो देखियो, जसले देखाउँछ कि जीवनशैलीले मुखको माइक्रोबायोममा को संरचना निर्धारण गर्न एउटा प्रमुख भूमिका खेल्दछ। स्ट्रेप्टोब्यासिलस र असम्प्रेषित पोरफिरोमोनाडासिए जीनस सहितका केही सूक्ष्मजीवहरूको सापेक्ष सापेक्षिक प्रचुरता पनि जीवनशैलीसँग मेल खाने हामीले यस अध्ययनबाट थाहा पाएका छौं। अन्ततः, हामीले जीवनशैलीको क्रमिक चरणमा माइक्रोबायोम संरचनासँग सम्बन्धित विशिष्ट जीवनशैलीसँग आबद्ध हुने धूम्रपान र अनाज जस्ता कारकहरूको पहिचान गरेका छौं।

निष्कर्ष हाम्रा निष्कर्षहरूले प्रष्ट पारेका छन् कि भूगोललाई नियन्त्रण गर्दा, हाम्रो जीवनशैलीले हाम्रा मुखको माइक्रोबायोमको संरचना निर्धारण गर्ने महत्वपूर्ण भूमिका खेल्दो रहेछ। यस अध्ययनले मानव समुदायहरूमा मुखको माइक्रोबायोमको अध्ययन गरेमा माइक्रोबायोमबारेको हाम्रो बुझाइमा सुधार ल्याउन सकिने पनि देखाउँछ। बढ्दो शहरीकरण सँगसँगै नेपालीहरूको माइक्रोबायोममा परिवर्तन हुँदै गएको र यस्ता परिवर्तनहरूले नेपालीहरूमा विभिन्न दीर्घकालीन रोगहरूको व्यापकता बढाउन सक्ने हुनाले यस सम्बन्धमा उपयुक्त निकायहरूको ध्यानाकर्षण हुनुपर्ने आवश्यकता पनि यस अध्ययनले प्रष्ट पारेको छ।

### Introduction

Throughout the last 300,000 years, our species experienced continual cultural transformation marked by milestones such as the development and use of tools, specialized division of labor, and urbanization [1]. These cultural shifts profoundly influenced both human societies and biology. One major transition in recent human history was the shift in subsistence strategy from hunting and gathering to agriculture and subsequently to industrialization. Such transitions encompass multifaceted lifestyle changes, including shifts in diet, population density, infectious disease burden, habitat, and other environmental factors [2]. These factors individually play pivotal roles in shaping the human microbiome—the diverse collection of bacteria, archaea, fungi, and other eukaryotes, and viruses that inhabit our bodies [3–6].

Understanding the role of subsistence strategy and accompanying lifestyle transitions has become a major focus of microbiome research [7–10]. Numerous

studies show that the gut microbiome shifts with industrialization. Specifically, industrialized populations generally exhibit lower gut microbiome alpha diversity compared with traditional populations, often lacking microbes commonly found in traditional populations, such as *Prevotella* and *Treponema* [5, 11–13]. These differences are associated with diet, drinking water source, and social structure [13–16]. While considerable progress has been made in understanding how the gut microbiome differs across lifestyles, differences in the oral microbiome across these transitions remain largely uncharacterized. Addressing this gap is crucial considering both the role of the oral microbiome in oral and systemic health [17–19] and also its prominence in the context of ancient DNA research [10, 20]. Therefore, it is important to expand studies of the oral microbiome to encompass diverse global populations practicing different subsistence strategies.

Much of our understanding of the oral microbiome across subsistence strategies comes from a limited

number of studies exploring the oral microbiomes of non-industrialized populations derived from ancient dental calculus [20, 21], although there have been increasing efforts to investigate that of modern-day populations as well [22–27]. Studies of modern-day global populations report that oral microbial diversity decreases with industrialization, with microbiome composition exhibiting differences based on lifestyle as well. More specifically, relative abundances of *Neisseria*, *Haemophilus*, *Prevotella*, and *Streptococcus* are lower in industrialized populations compared with traditional populations from the remote regions of the Philippines, Venezuela, Uganda, Malaysia, and South Africa [23, 24, 26, 28, 29]. However, several potential confounding factors may underlie the reported differences.

First, geography is associated with the microbiome [13, 30–32]. Distinguishing between the potential effects of geography versus lifestyle remains challenging, given the often large geographic variability between populations practicing different lifestyles. More specifically, samples from industrialized populations are generally collected from populations in the USA or Europe, whereas traditional populations might be collected at entirely different latitudes or continents [33]. As a result, it has been difficult to separate the effect of lifestyle and geography.

Second, some studies incorporate publicly available microbiome data from industrialized individuals with newly generated sequencing data from traditional populations [22, 24]. Technical discrepancies in sample collection, processing, and sequencing can influence microbiome study outcomes, raising questions about whether observed variations are due to lifestyle or technical factors [34, 35].

Third, studies that effectively control for technical effects are primarily focused on traditional populations [26, 36]. As a result, we lack an understanding of how the oral microbiome differs across the entire spectrum of lifestyles, from traditional to industrialized, while adequately controlling for temporal, geographical, and technical variation.

Finally, many existing studies make comparisons between distinct lifestyles, comparing traditional foragers/hunters and gatherers with agriculturalists and completely industrialized populations, without evaluating the subtle nature of microbiome shifts to and away from agriculturalist lifestyles [24, 27]. To our knowledge, there is no investigation that has comprehensively examined the entire lifestyle gradient—from hunting and gathering, to subsistence farming, to commercial farming, to early industrialization, and to established industrialization. It is important that we also examine these types of lifestyle transitions in modern-day settings, as they have been demonstrated to play major roles in shaping the

oral microbiome throughout human history via analyses of ancient dental calculus. For example, increasing levels of putatively pathogenic microbes in the oral microbiome are attributed to the Neolithic Revolution and increased starch consumption with the rise of agricultural practices [37, 38]. Whether this pattern holds in modern-day lifestyle transitions remains unclear.

To address these gaps, we characterized the salivary microbiota of Nepali individuals across a spectrum of human subsistence strategies, from traditional foragers to agriculturalists. Importantly, we are able to isolate the role of lifestyle from confounders, as geographical and technical variability are controlled for in the study with the inclusion of multiple lifestyles solely from Nepal. We specifically investigated six ethnically Nepali populations across four lifestyles. The Chepang represent foragers, the Raji and Raute are hunters and gatherers that recently settled and began subsistence farming in the 1980s, the Tharu and Newars living in Nepal are agriculturalists, and the Nepali expatriates living in the USA represent a population that recently transitioned to industrialization. We also include Americans of European descent as representatives of a fully industrialized population, for a total of five lifestyles. We demonstrate that oral microbiome composition differs along a gradient of traditional to industrialized lifestyles but, unlike that of the gut microbiome, differences are relatively subtle. By integrating questionnaire-based data encompassing diverse lifestyle variables such as diet, education, and medical practices, we identify specific lifestyle factors associated with oral microbial compositional changes. Finally, we examine the gut-oral microbiome axis to evaluate whether the degree of intra-individual similarity between the two sites differs across lifestyles. These results demonstrate that like the gut microbiome, the oral microbiome mirrors lifestyle.

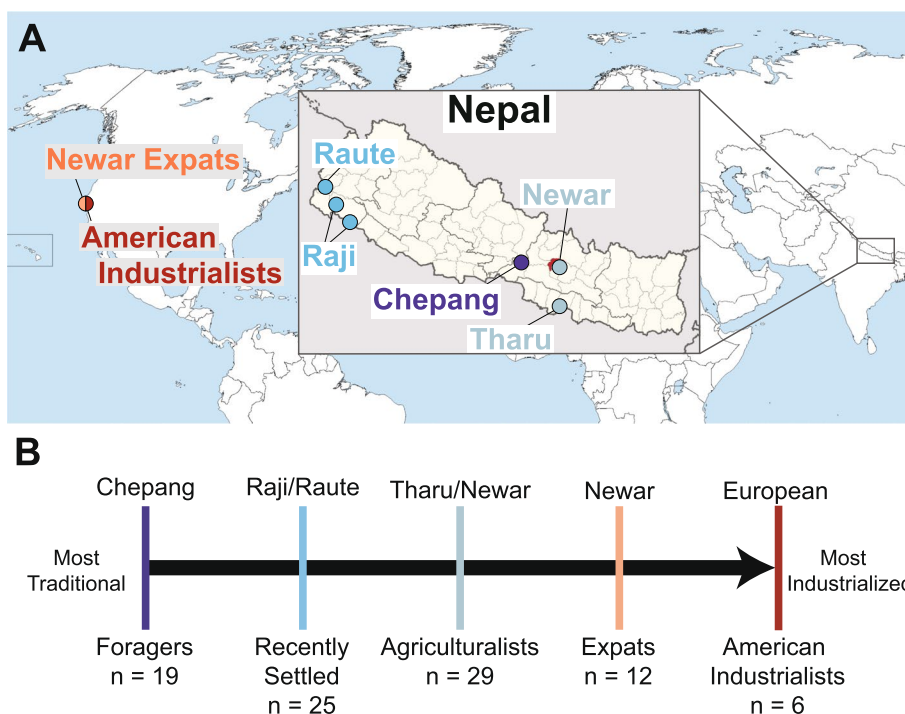
## Results

### Description of populations

We investigated the oral microbiome in diverse Nepali populations practicing a spectrum of lifestyles and an American population representing an industrialized lifestyle. In total, we examined individuals practicing five lifestyles: foragers, recently settled, agriculturalists, expats, and American industrialists. These lifestyles were defined a priori by subsistence strategy. However, a number of specific lifestyle factors differ across groups in addition to subsistence strategies, such as household size, diet, and smoking habits. Briefly, the Nepali individuals in this study belonged to five ethnic groups native to Nepal—Chepang ( $n=18$ ), Raji ( $n=11$ ), Raute ( $n=14$ ), Tharu ( $n=20$ ), and Newar ( $n=8$ ) (Fig. 1, Table S1). We also included expatriate Newar ( $n=12$ ) and European-Americans ( $n=6$ ), both of whom reside in the San

Francisco Bay Area. The Chepang, numbering 84,400 [39], are foragers who primarily reside in small communities of remote, isolated villages within the hills of the lower Himalaya in central Nepal. The Chepang village in this study lacks modern amenities such as electricity, running water, and other indicators of urbanization. While they supplement their diets with food grown via slash-and-burn agriculture, the low productivity of the hilly terrain compels them to heavily depend on foraged jungle foods such as undomesticated tubers and wild nettle (*sisnu*). The Raji and Raute, previously nomadic foragers, transitioned to subsistence agriculture within the past 50 years. The Raute and Raji are among the smallest ethnic groups in Nepal. The Raute, with roughly 550 individuals [39], reside in the far western hills of Nepal, whereas the Raji, numbering 5100 [39], inhabit the neighboring Terai plains. These two populations will be referred to as “recently settled” in this study, due to the transitional nature of their lifestyle. By contrast, the Tharu and Newar, two of the largest ethnic groups in Nepal numbering 1.8 and 1.3 million, respectively [39], practice agriculture. The Tharu, hailing from the Terai plains in Southern Nepal, fully transitioned to agriculture

about 300 years ago. The Newar originate from Kathmandu valley and are renowned for their cultural and economic contributions to Nepal. Although increasing urbanization of Kathmandu valley has afforded some Newar individuals access to industrialized comforts, the population in this study resides in a relatively rural village on the outskirts of Kathmandu valley and primarily engages in agriculture. For those reasons, in this study, both the Tharu and Newar will be referred to as the established agriculturalists. The expatriate Newar in the USA (“expats”) also originated from the Kathmandu valley and emigrated within the past 20 years, settling in the USA in their mid-30s. The Chepang, Raji, Raute, and Tharu individuals sampled here largely overlap with a previous study of Nepali gut microbiomes [15]. Both the fecal samples in this previous study and the saliva samples in this current study were collected concurrently. By focusing on individuals across a range of lifestyles within a confined geographic region, our study aims to discern oral microbial signatures of lifestyle without the confounding effects of geography, climate, and technical factors.



**Fig. 1** Sampling locations of all Nepal- and USA-based populations. **A** Locations of the populations sampled in Nepal and the USA. The US populations are specifically from the Northern California region. Location of Kathmandu is indicated in red on the Nepal map. Colors correspond to lifestyle groupings as described in **B**. **B** Oral microbiome samples were collected from individuals that span a spectrum of lifestyles, from nomadic foraging populations (dark blue), to populations that recently transitioned from foraging to small scale agriculture (teal), to established small scale agriculturalists (sky blue), to Nepali expats residing in the US practicing an industrialized lifestyle (peach), and to American industrialists (red). Sample sizes for each lifestyle category are indicated



### Lifestyle factors differ across populations

Given that multiple dietary, environmental, and socio-economic factors differ across populations in Nepal, we administered a survey to capture the specific lifestyle factors that differ across the Nepali populations (Table S1). We used Random Forests to evaluate the ability of the 37 lifestyle factors obtained from the survey questions to classify individuals according to their respective lifestyle groups. Nepali individuals are correctly assigned to their respective lifestyle categories with 92.06% overall accuracy (Fig. S1), suggesting that the populations in our study are highly distinguishable based on these factors.

We then performed correspondence analysis (CA) to determine which specific variables distinguish the lifestyles. Our analysis reveals a significant dependence between the samples and lifestyle factors ( $p=0.004$ ; chi-square test of independence). Individuals cluster closely in the top two dimensions of CA based on their lifestyle group, as do similar specific lifestyle variables (i.e., education and literacy, Fig. S2 A). Specifically, CA axis 1 follows the lifestyle trend ( $p=3.26 \times 10^{-14}$ ; Jonckheere-Terpstra test), whereas CA axis 2 does not ( $p=0.21$ ; Jonckheere-Terpstra test, Fig. S2 B-C). The top 10 contributing lifestyle factors to CA1 include *sisnu* consumption, fuel source, and literacy (Fig. S2 D). For example, the Chepang consume *sisnu* most frequently, use solid fuel for cooking, and have low literacy rates, while the Expats do not consume *sisnu*, use gas or electricity for cooking, and are literate. By contrast, the top contributing factors to CA2 include behavioral factors like smoking and alcohol consumption (Fig. S2 E). The high predictive capabilities of the Random Forest analysis and discrimination between lifestyle categories based on specific lifestyle factors suggest that the lifestyles are both well-defined by subsistence strategy and effectively described by the survey metadata.

### Oral microbiome diversity does not differ across Nepali populations

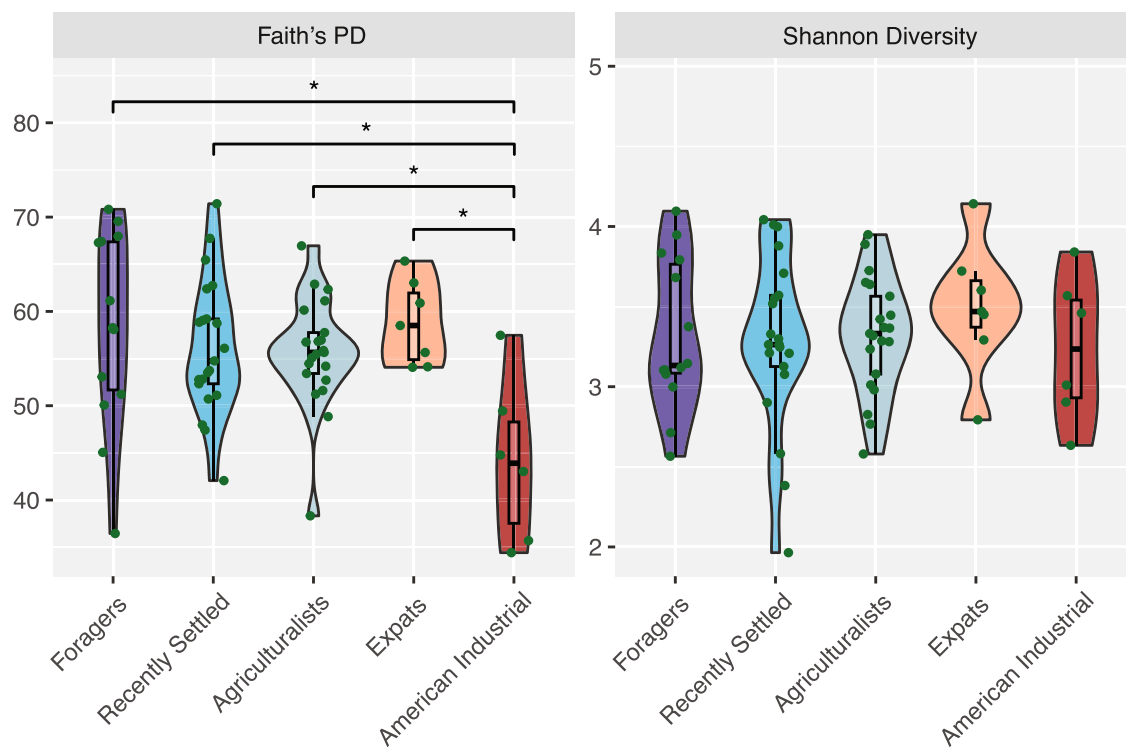
Fecal samples concurrently collected with the saliva from the Chepang, Raji, Raute, and Tharu individuals previously revealed pronounced gut microbiome compositional differences across the gradient of lifestyles from foraging to industrialized [15]. To evaluate whether oral microbiome compositional differences align with the continuum of lifestyles, we initially characterized the oral microbiome via saliva samples collected from 89 individuals across five lifestyles (Table S1). Recognizing the potential for DNA extraction methodology to introduce variability in microbiome studies [34, 35], we performed DNA extraction using two different kits, Qiagen QIAamp MinElute Virus Spin kit and MO BIO PowerSoil DNA, to ensure the robustness of our conclusions. Overall

microbiome composition and diversity are consistent between the two kits (Fig. S3). Consequently, all subsequent analyses were performed using the data obtained from the Qiagen kit for simplicity, which resulted in 69 individuals that passed quality control steps (see the “Methods” section).

We first evaluated whether overall microbiome diversity differed across lifestyle groups, as decreasing diversity is typically thought of as a hallmark of traditional to industrialized lifestyle transitions [12, 24, 26]. We observe no significant difference in Shannon diversity across the lifestyle groups ( $p>0.05$ , Kruskal–Wallis; Fig. 2), but there is a significant difference in Faith’s phylogenetic diversity between the lifestyles ( $p=0.028$ ; Kruskal–Wallis, Fig. 2). A post hoc pairwise comparison demonstrates that the American industrialists are driving the differences in Faith’s phylogenetic diversity (American Industrialists vs. other lifestyles:  $p<0.05$ , Dunn’s post hoc test). Notably, there is no significant difference in Faith’s phylogenetic diversity between the four Nepali populations ( $p>0.05$ ; Kruskal–Wallis). These findings remain largely consistent with other alpha diversity metrics across both extraction kits (see the “Methods” section, Fig. S4–S5). Aligning with the observations in the gut microbiomes of these individuals [15], our results indicate that oral microbiome diversity does not correlate with lifestyle differences within Nepal when geography is controlled for. Notably, our sample sizes, while modest, are larger than most other oral microbiome studies examining traditional lifestyles [22–24, 26], underscoring that the lack of signal is not due to insufficient power. It is also important to note that the difference in microbiome diversity is observed using Faith’s phylogenetic diversity and not using Shannon’s diversity. While both within-sample diversity metrics, Shannon’s diversity accounts for species abundance and evenness without consideration of underlying phylogenetic structure, whereas Faith’s phylogenetic diversity accounts for the total phylogenetic breadth of the residing taxa. As a result, our results suggest that while the microbes residing in the Nepali individuals cover a larger extent of the phylogenetic tree than the American industrialists, the relative evenness is fairly equal between the populations.

### Oral microbiome composition differs across lifestyles

Unlike microbiome diversity, microbiome composition varies across lifestyles. We calculated between-sample Bray–Curtis distances to measure beta diversity [146], revealing that oral microbiome composition varies significantly with lifestyle ( $p=2.3 \times 10^{-4}$ ; PERMANOVA, Fig. 3A). This relationship remains significant even when accounting for sex across all individuals and both sex and age across the Nepali individuals ( $p=1.5 \times 10^{-4}$ ,

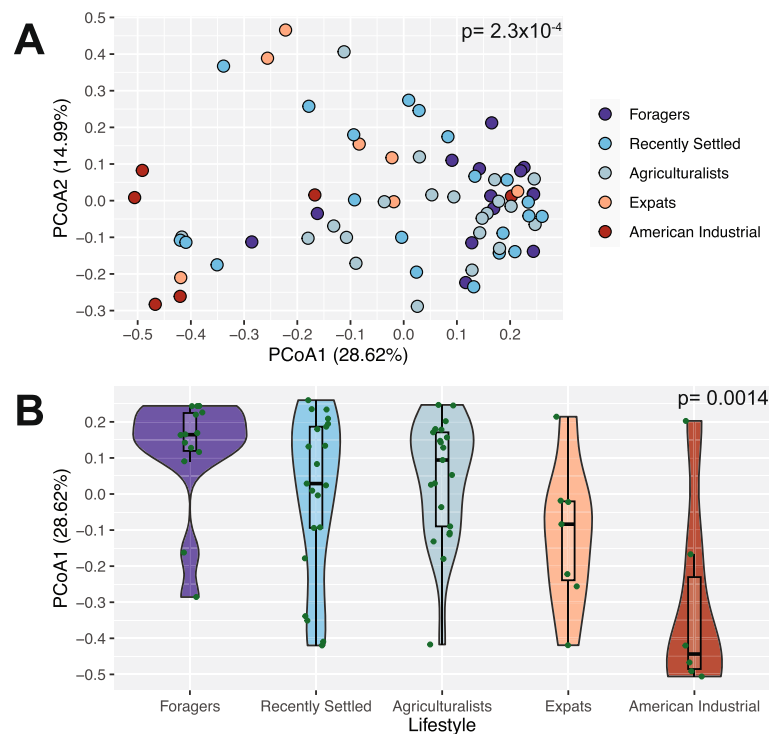


**Fig. 2** Alpha diversity does not significantly differ by lifestyle in Nepal. Faith's phylogenetic diversity (Faith's PD—left) and Shannon alpha diversity (Shannon Diversity—right) shown for all individuals, grouped by lifestyle. Lifestyles are ordered from most traditional (foragers) to most industrialized (American industrialists), left to right. No significant difference detected across lifestyles for Shannon alpha diversity ( $p=0.8$ , Kruskal–Wallis), but a marginally significant difference detected for Faith's phylogenetic diversity ( $p=0.028$ , Kruskal–Wallis). Notably, those significant differences only occur between the American industrialists and other lifestyle groups, not between Nepali individuals residing in Nepal or the USA. Significant differences ( $p < 0.05$ ) between specific populations are indicated (\*)

$p=0.025$ ; respectively, PERMANOVA). Subsequent pairwise comparisons demonstrate multiple significant differences between specific pairs of groups (foragers vs. American industrialists, foragers vs. expats, agriculturalists vs. American industrialists:  $p < 0.05$ , pairwise PERMANOVA). When visualized via principal coordinate analysis (PCoA), the first axis (PCoA1, explaining 28.62% of microbiome variation) follows the lifestyle gradient ( $p=0.0014$ ; Jonckheere–Terpstra test, Fig. 3B), with the expatriates and Americans being most different from the traditional Nepali populations. This pattern is consistent across data from both extraction kits (Fig. S6 B–D) and with UniFrac metrics (Figs. S7–S8) [40]. These compositional differences across groups, however, are fairly subtle, with the classification of microbiomes into lifestyle groupings via Random Forests not being better than expected by random chance (43.48% accuracy, Fig. S9).

Because overall microbiome composition differed based on lifestyle, we next determined which specific taxa differed across the lifestyles. To do this, we conducted

differential abundance analysis using ALDEx2, which accounts for compositionality in its application [41]. We observe that 2 of the 111 oral genera (1.8%) were significantly differentially abundant after accounting for multiple tests, namely *Streptobacillus* and an unclassified Porphyromonadaceae genus ( $\text{padj}=0.011$  and 0.021, respectively; Kruskal–Wallis) (Table S2). These results are robust to the inclusion of sex and age as covariates (Table S3). We also implemented an alternative approach for identifying which taxa were following the lifestyle gradient by performing the Jonckheere–Terpstra test for all genera and correcting for multiple tests. We find that nine genera significantly followed the lifestyle gradient, including the two that were also identified using ALDEx2 (Fig. 4, Table S4). Eight genera—*Streptobacillus*, *Porphyromonadaceae\_unclassified*, *Granulicatella*, *Moraxella*, *Simonsiella*, *Neisseria*, *Bacteroidetes\_unclassified*, and *Brachymonas*—show decreasing abundance with industrialization, consistent with the trend observed in Faith's phylogenetic diversity. The only exception is *Atopobium*, which shows the opposite trend of increasing abundance



**Fig. 3** Oral microbiome composition significantly differs based on lifestyle. **A** Microbiome composition varies significantly with lifestyle ( $p = 2.3 \times 10^{-4}$ , PERMANOVA). The PCoA plot shows individuals ordinated based on Bray–Curtis distance and colored by lifestyle. **B** The distribution of individuals along PCoA axis 1 follows the lifestyle gradient, from traditional to industrial ( $p = 0.0014$ , Jonckheere–Terpstra test). Lifestyles are ordered from most traditional (foragers) to most industrialized (American industrialists), left to right

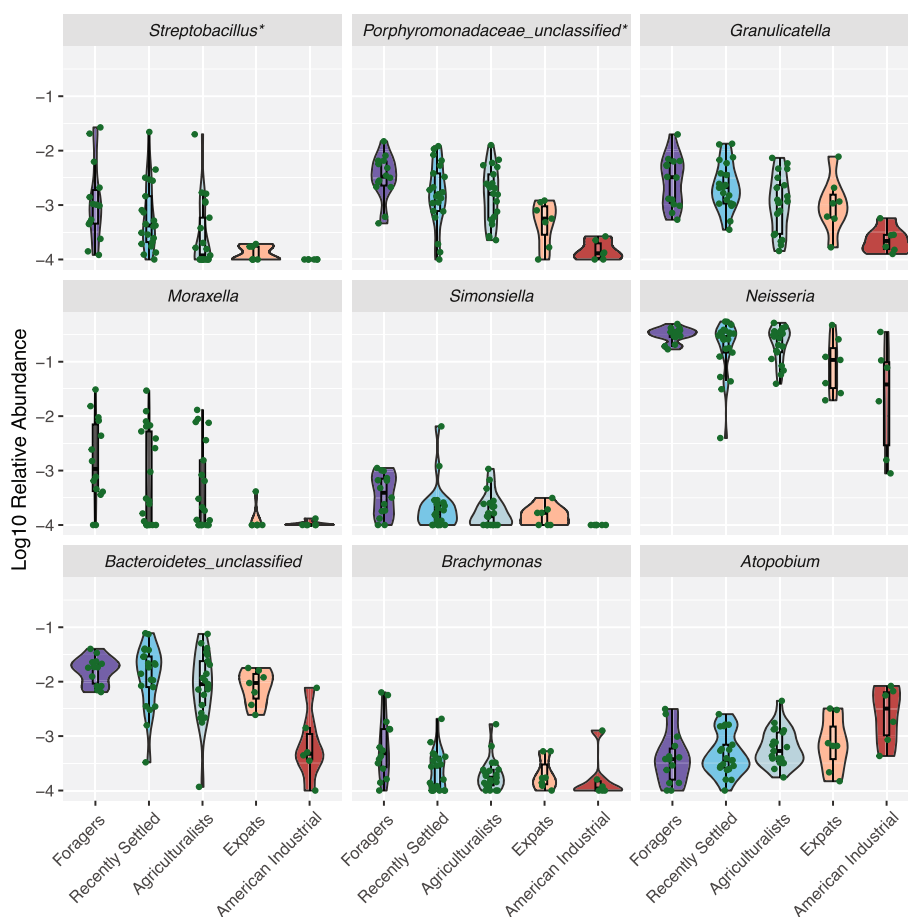
with industrialization (Fig. 4). To identify whether these trends were driven primarily by the industrialists, we analyzed the lifestyle gradient for only the Nepali groups. We find that only *Streptobacillus* significantly follows the lifestyle gradient within Nepali individuals after correcting for multiple tests. Interestingly, when applying the same approach to all individuals except the foragers, we also find that, again, only *Streptobacillus* significantly follows the lifestyle gradient. These results confirm that the extreme ends of the lifestyle gradient have substantially different microbial relative abundances, with the intermediate stages exhibiting more subtle differences in relative abundance.

#### Grain type is associated with microbiome differences across lifestyles

Given that overall microbiome composition mirrored the transition of lifestyles within Nepal, we sought to identify lifestyle factors that potentially underlie these differences. To do this, we first started at a broad scale by comparing the major axes defining oral microbiome composition and the lifestyle variables from PCoA and CA, respectively. We calculated the correlation between the first three CA axes, which cumulatively captured 37.14%

of the variation in the lifestyle survey data, and the first three PCoA axes, which cumulatively captured 49.35% of the variation in the microbiome data. We observe a significant correlation between PCoA2 and CA2, which is primarily comprised of behavioral lifestyle factors like tobacco and alcohol use and distinguishes the recently settled populations from foragers and agriculturalists ( $p = 0.03$ ;  $\rho = -0.27$ ; Spearman correlation, Fig. S10 A). No significant correlation is observed between the CA axes and alpha diversity (Fig. S10 B).

We then identified which specific lifestyle factors are associated with the observed differences in microbiome composition. As testing all 37 measured lifestyle factors would be prohibitive due to multicollinearity, we selected the top 15 key lifestyle distinguishing factors based on their contributions to the first two CA axes (Fig. S2). We used these factors to perform canonical correspondence analysis (CCA) to determine which variables are associated with shifts in the microbiome. We find significant associations between these lifestyle distinguishing factors and the oral microbiome composition among the Nepalis ( $p = 0.013$ ; ANOVA); with the top factors being alcohol consumption, smoking habits, location, *sisnu* consumption, and grain type ( $p = 0.044, 0.001, 0.003, 0.003, \text{ and } 0.027$ , respectively, ANOVA) (Fig. 5A).



**Fig. 4** Abundances of nine genera significantly follow the lifestyle gradient. The relative abundances of nine genera significantly follow the lifestyle gradient via a Jonckheere-Terpstra test followed by Benjamini–Hochberg correction (adjusted  $p < 0.05$ ). Lifestyles are ordered from most traditional (foragers) to most industrialized (American industrialists), left to right. All taxa have been log<sub>10</sub>-transformed for visualization purposes. Taxa marked with \* are also significantly differentially abundant across lifestyles based on ALDEx2. Most taxa tend to decrease in relative abundance as the lifestyles transition from more traditional to industrial, while the abundance of *Atopobium* increases.

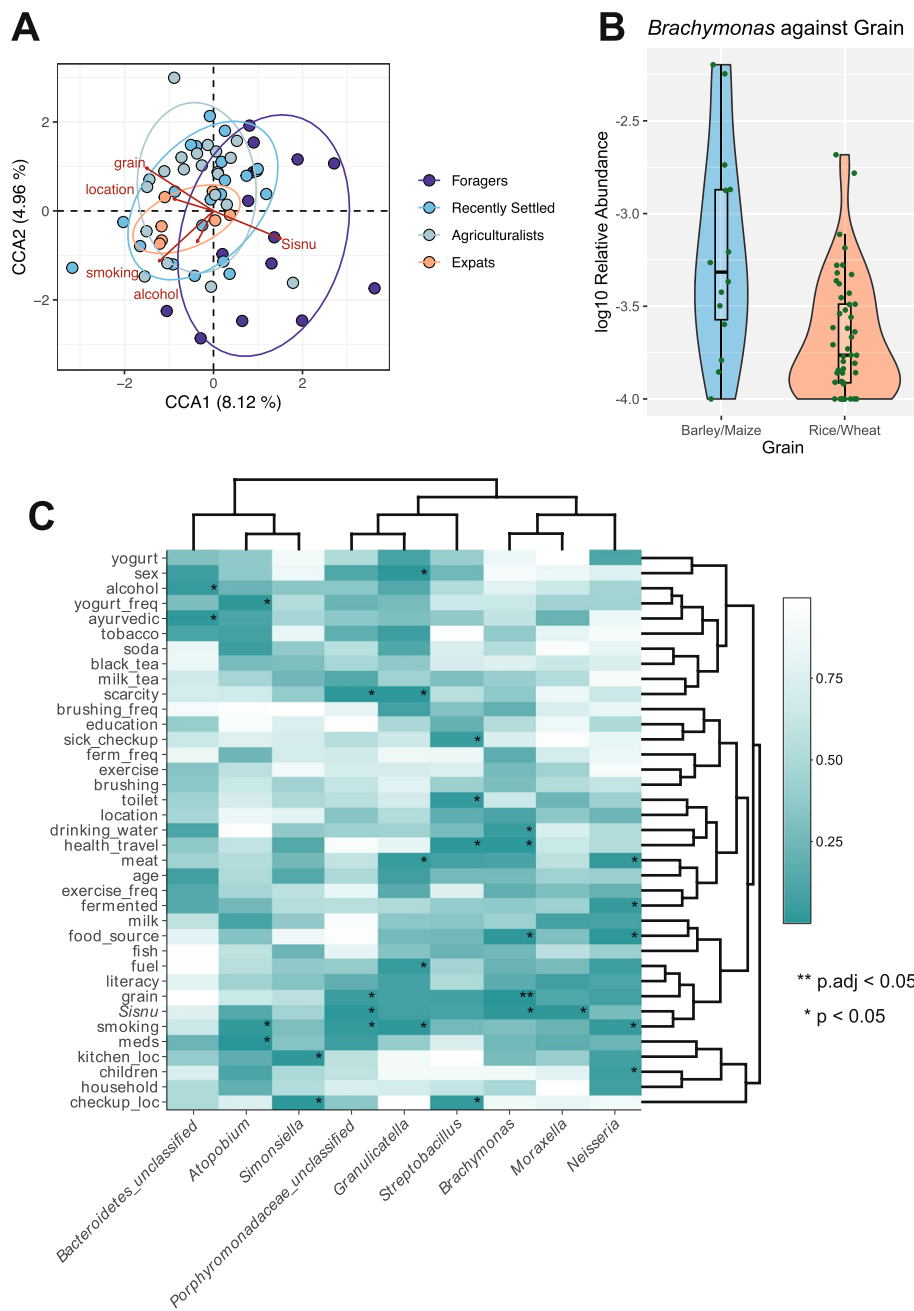
Finally, we identified associations between the nine taxa that are differentially abundant across lifestyle groupings (Table S4) and the 37 lifestyle factors included in our survey via linear models. Out of the 333 associations tested, we find that *Brachymonas* is significantly associated with grain consumption at an adjusted  $p$ -value, with a higher abundance of this taxon observed in individuals who primarily consume barley and maize compared with those who primarily consume rice and wheat ( $p_{adj}=0.022$ ; Fig. 5B). Furthermore, we observe an additional 27 associations that are significant at a nominal  $p$ -value  $< 0.05$  (Fig. 5C, Table S5). For example, we observe that the relative abundances of *Granulicatella*, *Neisseria*, and *Porphyromonadaceae\_unclassified* are higher in non-smokers, whereas *Atopobium* relative abundance is lower in non-smokers ( $p=0.006$ ,  $p=0.032$ ,  $p=0.033$ , and  $p=0.023$ , respectively, Fig. S11 A). Similarly, the relative

abundances of *Brachymonas*, *Moraxella*, and *Porphyromonadaceae\_unclassified* are higher in individuals that consume *sisnu* ( $p=0.014$ ,  $p=0.019$ , and  $p=0.012$ , respectively, Fig. S11 B). Notably, while both smoking and alcohol are associated with the oral microbiome and both factors are top contributors to lifestyle, neither of these factors significantly follow the lifestyle gradient from traditional to expatriate ( $p > 0.05$ ; Cochran-Armitage test, Fig. S12). These results demonstrate that a variety of lifestyle factors potentially underlie the differences in oral microbiome composition observed between lifestyles within Nepal.

#### Predicted metabolism pathways are differentially abundant across lifestyles

In addition to the taxonomic differences observed between lifestyle groups, the predicted functional





**Fig. 5** Alcohol, smoking, location, *sisnu*, and grain type are associated with the oral microbiome. **A** There are significant associations between lifestyle factors and the microbiome as observed via CCA. Points represent individuals and color represents corresponding lifestyle, with ellipses around each population. Red arrows represent the lifestyle factors significantly associated with the microbiome. A total of 15 lifestyle factors were inputted into the CCA model based on contribution to each CA axis. **B** *Brachymonas* is significantly associated with grain type consumed ( $p_{adj}=0.022$ ). Specifically, relative abundance is higher in individuals who report barley and maize consumption compared with rice and wheat. Taxa were log<sub>10</sub> transformed for visualization. **C** Several specific lifestyle factors are associated with individual oral genera. Significant associations based on a nominal  $p$ -value threshold are indicated with \*. The significant association based on an adjusted  $p$ -value threshold is indicated with \*\*

potential of the microbiome significantly differs as well. Based on the use of PICRUSt2 [42], predicted functional abundance significantly varies with lifestyle ( $p=0.0036$ ; PERMANOVA, Fig. S13 A). The top two PCA axes both significantly follow the lifestyle gradient (PC1:  $p=0.049$ , PC2:  $p=0.0064$ ; Jonckheere-Terpstra test, Fig. S13 B). We note that while the taxa present in Nepali oral microbiomes seem to be fairly closely represented in the reference taxonomy used to predict gene content (average NSTI=0.037), there are many caveats to predictive methods such as PICRUSt2 and these results should be viewed as hypothesis-generating [43].

To identify specific potential functional differences across lifestyles, we conducted differential abundance testing with ALDEx2 using the predicted abundances of 109 pathways. Although none are significant after multiple test corrections, 22 pathways are significant at a nominal  $p$ -value of  $p < 0.05$  (Table S6), 13 of which are classified as metabolism pathways (Fig. S14A). These metabolism pathways can be categorized into 7 classes, some of which increase in abundance with increasing industrialization, including lipid metabolism and glycan biosynthesis, while others decrease, such as xenobiotics degradation and microbial metabolism in diverse environments (Fig. S14B). General transporter proteins (ATP-binding cassette transporters— $\text{padj}=0.0013$ , phosphotransferase system— $\text{padj}=0.0013$ ) and degradation pathways (aminobenzoate degradation— $\text{padj}=0.0028$ ) are significantly enriched via enrichment analysis (Fig. S15). Finally, the top 10 most significant pathways from ALDEx2 were further examined to identify the top contributing microbes and whether they differ by lifestyle. *Fusobacterium* is one of the top taxa contributing to platinum resistance (Fig. S16 A). There is a significant enrichment of *Fusobacterium* in the traditional Nepali populations compared with industrialized populations ( $p=0.0037$ ; Kruskal–Wallis test, Fig. S16 B). Overall, predicted metabolism pathways significantly differ across lifestyles, mirroring the taxonomic gradient across lifestyles in Nepal.

#### Microbial network structure varies across lifestyles

We then investigated network structure to determine whether community structure differs across lifestyles. We used the SparCC module in the SpiecEasi package [44, 45] to generate a network from all 111 genera observed in this study. The resulting network consists of 37 nodes with at least one edge and 6 co-abundance groups, with a modularity of 0.45 (Fig. 6A). Among the taxa identified as following the lifestyle gradient, 5 out of the 9 are connected to at least one other taxon, with *Porphyromonadaceae\_unclassified*, *Neisseria*, *Bacteroidetes\_unclassified*, and *Granulicatella* being in the same

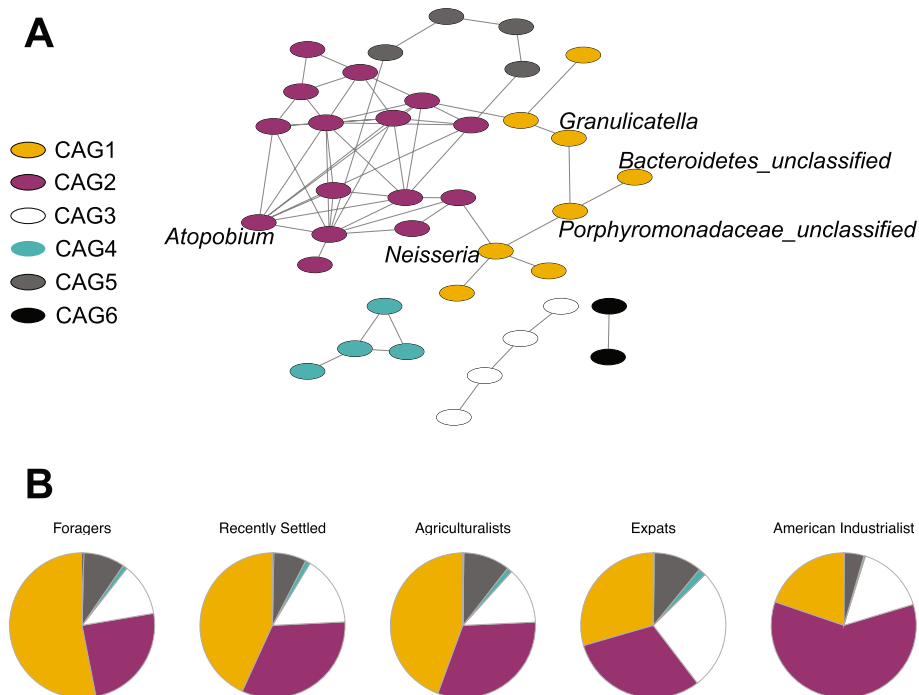
co-abundance group (CAG1), whereas *Atopobium* is in a separate co-abundance group (CAG2). Interestingly, the proportions of CAGs differ across lifestyles, with CAG1 decreasing with industrialization and CAG2 increasing with industrialization (Fig. 6B). These results demonstrate that community network structure differs along the lifestyle gradient, in addition to individual microbial taxa and predicted functional potential.

#### Oral-gut microbiome distance decreases with agrarianism

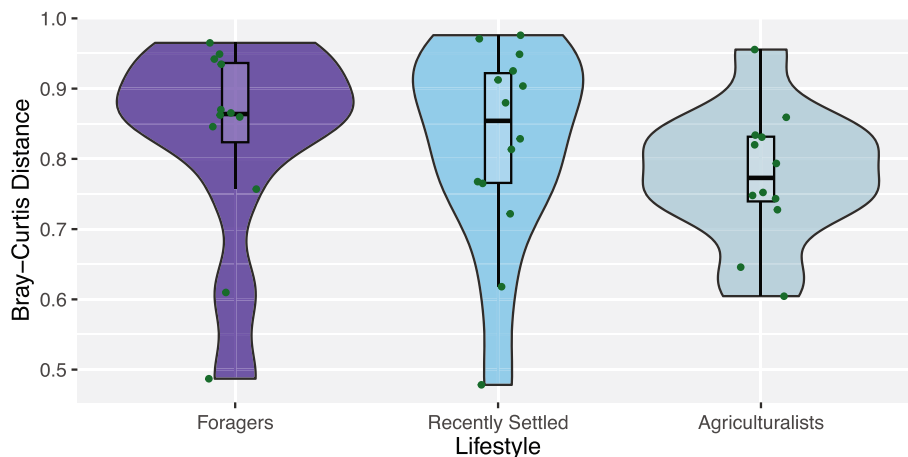
Finally, we examined the role of lifestyle along the oral-gut microbiome axis. While different lifestyle factors independently associated with the oral and gut microbiomes, the compositional similarity between the two sites within an individual increases with the extent of urbanization [25]. Thus, we were interested in assessing whether there was a similar association across lifestyles, in which we might expect to see increasing intra-individual similarity across the gradient of lifestyles from traditional to agrarian. To do this, we examined individuals for whom both oral and gut microbiome data were collected concurrently, for a total of 12 foragers, 14 recently settled individuals, and 12 agriculturalists (Table S7), and calculated Bray–Curtis dissimilarity between the oral and gut microbiomes. We find that intra-individual oral-gut microbiome dissimilarity decreases, and therefore similarity increases, across the gradient of traditional to agrarian lifestyle as predicted, although this trend is not statistically significant ( $p=0.11$ ; Jonckheere-Terpstra test, Fig. 7). We, however, do observe significant similarities in composition between the two body sites across individuals ( $p=0.013$ ,  $\rho=-0.4$ ; Spearman correlation, Fig. S17), suggesting that we are perhaps underpowered to detect the significance of intra-individual dissimilarity at our current sample size. To verify this, we conducted a power analysis and determined that we only have 4.1% power given our measured effect size differences between groups and would require at least 62 individuals per group to detect a significant trend with 80% power (Fig. S18).

#### Discussion

Much of human microbiome research is weighted towards populations living in North America and Europe, with South Asia being particularly underrepresented [46]. As a consequence, our understanding of how the oral microbiome varies across human lifestyles is extremely limited, especially with regard to non-industrialized lifestyles including foraging, hunting and gathering, and small-scale agriculturalism. Some existing studies identify differences in the oral microbiome between hunter-gatherers and farmers [24, 26]. Others investigate the role of a few specific lifestyle factors; like



**Fig. 6** Differentially abundant taxa are highly connected in the oral microbiome co-occurrence network. **A** The SparCC module in the SpiecEasi package was used to generate a network from 111 genera. Network of 37 nodes with at least one significant edge is shown, with 6 co-abundance groups (CAGs) indicated by node color. Labeled nodes indicate genera that were identified as significantly differentially abundant across lifestyles. **B** Proportions of CAGs vary across lifestyle. Specifically, CAG1 decreases with industrialization, whereas CAG2 increases with industrialization



**Fig. 7** Correlation between the oral and gut microbiomes within an individual strengthens with agrarianism. Microbiome dissimilarity (as measured by Bray–Curtis dissimilarity) between the gut and oral microbiomes within an individual decreases across the gradient of lifestyles from traditional to agrarian for individuals residing in Nepal, albeit not significantly ( $p=0.11$ ; Jonckheere-Terpstra test)

smoking, alcohol consumption, and diet; but only in a single, usually industrialized, population [36, 47–49]. Our study is the first comprehensive examination of the oral microbiome across a gradient of lifestyles, including transitional lifestyles—recently settled nomadic foraging individuals now practicing small-scale agriculture

and expatriates transitioning from farming to industrialization—and numerous lifestyle factors characterizing each lifestyle. Importantly, we are able to isolate the role of lifestyle, as geography and technical factors are controlled for. We observe that even when controlling for geography, microbiome composition mirrors lifestyle

transitions, likely due to differences in dietary and behavioral habits.

In general, populations in industrialized countries have lower gut microbiome diversity than those practicing more traditional lifestyles [5, 11–14, 16, 50, 51]. This has led to an active discussion regarding whether we should intervene to improve human health and reduce microbial loss in industrialized countries [9, 52, 53]. While it is tempting to attribute the observed diversity differences across populations to lifestyle, controlling for geography eliminates those differences [14, 54, 55]. In fact, when examining the gut microbiomes of the individuals included in this study, there was no significant difference in within-sample alpha diversity across lifestyles, even when including American industrialists [15]. Fewer studies examine oral microbiome diversity across human lifestyles, and most of them observe a decrease in diversity across lifestyles [24, 26]. That said, it is important to highlight that geography was not closely controlled for and might have confounded the results, as examining lifestyle while controlling for geography results in no difference in diversity [22].

Here, we demonstrate that alpha diversity of the oral microbiome does not significantly differ across lifestyles between Nepali individuals, aligning with findings from oral microbiome studies that control for geography [22]. Interestingly, unlike the gut microbiomes of the same individuals [15], we observe a significant decrease in Faith's phylogenetic diversity in the American industrial population compared with Nepali individuals, including the Nepali expats. As the Nepali expats and American industrialists currently reside within the same metro area, geography is effectively controlled. There are several possible explanations for why these differences are observed between the Nepali expats and American industrialists, but not the populations within Nepal. First, the main differentiating factor may not be geography, as one might assume if only comparing the populations residing in Nepal with American industrialists, but rather lifestyle factors that have more extreme effects between industrialists and other lifestyles than between the traditional and agrarian populations within Nepal. For example, the Nepali expats included in this study tend to retain their traditional cuisine, which differs greatly from a standard American diet. Although recipes are modified to account for local ingredient availability, the main dietary components remain consistent across the ethnically Nepali populations, regardless of geography. Another possibility is that geography does drive oral microbial diversity, but can only do so during critical windows earlier in life in which oral microbiomes are malleable [56]. While gut microbiota appear to be malleable even with immigration well into adulthood [57], it is not clear if the same

is true for the oral microbiome. Once a stable microbial community or host immune repertoire is established, moving to another geographic region may not result in diversity changes. The Nepali expatriates included in this study immigrated around their mid-30s, so they may have missed this window. Further investigation would be needed to tease apart these and other possible explanations for the differences in diversity we observed between Nepali individuals and American industrialists.

Similar to the gut microbiomes in the same individuals [15], oral microbiome composition mirrors lifestyle. Specifically, we observe a consistent compositional gradient when comparing individuals from traditional foraging populations (Chepang) to recently settled populations (Raji and Raute), to small-scale agriculturalists (Newar and Tharu), to immigrants (Newar) and industrialized Americans. Compared with the gut, which we reanalyzed using ALDEx2 to ensure comparability of results (Table S8), differences across lifestyles are more muted in the oral microbiome. More genera in the gut are significantly differentially abundant across lifestyles (27% in gut, 1.8% in oral), even though there was slightly more power with the oral samples due to increased sample size. This finding may be due to the increased resiliency of the salivary microbiome compared with the gut microbiome, thus resulting in fewer differences in the microbiome across lifestyles [58–62]. It is important to note that these considerations are specific to the salivary microbiome, as the microbiomes of other oral anatomical sites, including dental calculus, differ significantly from that of saliva [63–65].

When examining microbial abundance across lifestyles, most of the differentially abundant microbes decrease across the gradient of traditional to industrialized. Many of these taxa co-occur and lie in the same co-abundance group when considering the full oral microbiome network (CAG1). One such taxon is *Neisseria*, which decreases in abundance with industrialization in other lifestyle studies [24, 28]. *Neisseria* plays a beneficial role in periodontal health, possibly by preventing the colonization of pathogenic microbes [66, 67]. Its decreasing abundance aligns with the hypothesis that the loss of crucial microbes is associated with the emergence of disease in industrialized society [68]. Other previously identified lifestyle-associated oral microbes are not significantly associated with lifestyle in our study, such as *Haemophilus*, *Prevotella*, and *Streptococcus* [23, 24, 26, 29]. Instead, we observe decreasing levels of an unclassified Porphyromonadaceae genus. *Porphyromonas gingivalis* is a member of the Porphyromonadaceae family, well-established as a pathogenic oral microbe contributing to periodontal disease [69]. Further investigation is needed of this microbe in our dataset, as this would

provide insight into Porphyromonadaceae as a potential oral pathogen at a global scale. *Streptobacillus* is another microbe that decreases in abundance with increasing industrialization. In North America, *Streptobacillus* is most well-known as *Streptobacillus moniliformis* for its presence in rat oral microbiomes and its role in rat bite fever [70]. By contrast, *Streptobacillus* appears as a commensal member in the human oral microbiomes of Asian populations. More specifically, *Streptobacillus hongkongensis* resides in the oral cavity of populations from Hong Kong and the United Arab Emirates [71, 72]. *Streptobacillus* has also been found in the Agta hunter-gatherers in the Philippines [73], thus suggesting that *Streptobacillus* presence in the human oral microbiome may be regionally limited to Asia. This hypothesis aligns with our study, as *Streptobacillus* is not observed in the American industrialists and is also less abundant in the expats. However, we note that our American industrialist sample size is limited, and larger sample sizes would be needed to confirm the low prevalence of this taxa in non-Asian populations. The relative abundance of differentially abundant microbes decreasing with industrialization is consistent with prior observations of decreased microbiome diversity in industrialized populations.

By contrast, the one genus that significantly increases in relative abundance with industrialization is *Atopobium*. High levels of oral *Atopobium* relative abundance are associated with a variety of negative health outcomes, many of which are more common in industrialized societies. Highly abundant in individuals with dental caries, *Atopobium* is believed to contribute to the development of dental caries as an accessory to *Streptococcus mutans*, a leading microbial cause of caries [74–76]. Oral *Atopobium* carriage is also enriched in individuals with hypertension [77], Sjögren's Syndrome [78], and patients with severe oral mucositis, a toxicity occurring from cancer treatments [79]. Notably, we observe that *Atopobium* belongs to a separate co-abundance group from the other differentially abundant taxa, CAG2. CAG2 also contains *Veillonella*, a microbe that may be an accessory for oral pathogen colonization [80]. Future research is needed to disentangle whether *Atopobium* and other potentially pathogenic microbes play causal roles in the development of oral conditions.

Multiple specific lifestyle factors are believed to play a role in shaping the oral microbiome, such as smoking, dietary fiber, and carbohydrate sources [47, 48, 81]. Most of these factors, however, were characterized in industrialized populations and it remains unclear whether the same factors play a role in the oral microbiota of traditional populations. Using extensive survey metadata, we observe significant associations between the microbiome

and 15 lifestyle factors, with smoking, grain type, and *sisnu* consumption most strongly associated.

Smoking consistently associates with oral microbiome composition across numerous industrialized populations, with increases in *Atopobium* and decreases in *Neisseria* and family-level Porphyromonadaceae being trademark indicators [48, 82–85]. We identify similar associations with smoking in our study as well. Specifically, *Atopobium* abundance is increased in smokers, whereas *Granulicatella*, *Neisseria*, and *Porphyromonadaceae\_unclassified* abundances are decreased in smokers. The taxa with increased abundances in smokers cluster in the same co-abundance group, whereas *Atopobium* is found in a separate co-abundance group. Notably, while smoking prevalence rates differ across lifestyle groups, they do not follow the gradient of lifestyles from traditional to agricultural. This suggests that the association of these taxa with smoking is independent of subsistence strategy. These findings suggest that smoking habits may play an important, consistent role in defining oral microbiome community dynamics across lifestyles and highlight the importance of accounting for smoking as a factor in future studies of lifestyle and the oral microbiome.

The associations we observe between grain type and the oral microbiome in Nepali populations are of particular interest, given the proposed importance of starch-rich foods in shaping oral microbiomes along the primate lineage [86]. Carbohydrates are associated with a myriad of oral microbes, either due to their role in starch digestion or pathogenicity [87, 88]. In this study specifically, populations reported primarily consuming either barley and maize, or rice and wheat. Barley and maize differ substantially from rice and wheat in terms of phenol content, digestibility, fiber content, and glycemic index [89, 90]. More specifically, barley and maize contain higher levels of phenols, which positively associated with gut microbiome health, oral microbiome health, and overall systemic health [91, 92]. In addition, rice and wheat are digested faster than barley and maize [90]. This higher digestibility might be attributed to differences in cell structure, like thinner cell walls [93]. More starch consumption and more salivary amylase, the first step of starch digestion, may also be contributing factors. Higher salivary amylase copy number has been observed in individuals with high starch diets [94] and is associated with oral microbiome composition [95]. Furthermore, barley has higher fiber content compared with other refined grains [96], which is also associated with improved health outcomes [97]. Finally, barley is reported to have a lower glycemic index compared with rice and wheat [89], which is generally associated with more positive health outcomes [98]. We observe a significant association between grain type and *Brachymonas*, which has not been previously found to be



associated with grains in human oral microbiomes. Little is known about the role of this microbe in humans, beyond its presence in healthy oral microbiomes [99–102]. Instead, this microbe has been demonstrated to decrease in abundance in the rumen microbiome of cows fed a high-grain diet [103], so further investigation is needed to fully understand this potential relationship.

We also identify a possible relationship between *sisnu* and the oral microbiome. *Sisnu*, also referred to as nettle, is a fibrous plant known for its medicinal benefits and primarily consumed by the Chepang foragers [104]. *Sisnu* consumption is a major differentiating lifestyle factor in this study, as demonstrated by correspondence analysis (Fig. S2). Although widely used in traditional medicine, little is known about its role in the microbiome. *Sisnu* shows strong antimicrobial properties against a variety of gram-positive and negative bacteria in vitro [105], although evidence is mixed [106]. Here, *Porphyromonadaceae\_unclassified* relative abundance increases with *sisnu* consumption. Several mechanisms could explain this association. First, ingesting *sisnu* may result in lowered absolute abundance of the oral microbiome overall, with *Porphyromonadaceae\_unclassified* being more resilient than other microbes. Second, the antimicrobial effects could create an expanded niche for this microbe to thrive, without impacting absolute abundance levels across the oral microbiome. Finally, the association may be unrelated to the potential antimicrobial properties of the plant, but rather nutrition. A finer-grain investigation would be needed to fully establish the underlying causes of this association, especially considering the prominence of *sisnu* as a therapeutic agent. Overall, these analyses of specific lifestyle factors associated with the oral microbiome in Nepali populations provide new insight into the role of specific dietary components and environmental factors in the oral microbiomes of non-industrialized, non-equatorial populations.

One unexpected result was the lack of an observed association with teeth brushing or flossing. Oral hygiene practices are associated with the microbiome, as the mechanical actions of brushing and flossing disrupt plaques and antimicrobial toothpastes also chemically break down biofilms. As a result, even subtle differences in types of toothpaste and brushing frequency result in changes in the plaque and salivary microbiota [107–110]. In traditional populations, miswak, also referred to as a chewing stick, is often utilized for mechanically and chemically cleaning teeth and has been demonstrated to inhibit common oral pathogens [111–115]. Finally, charcoal and ash have also been used historically for oral hygiene, although their effectiveness is highly disputed [116]. Our study included individuals who reportedly do not brush their teeth, those who use traditional

methods of miswak or charcoal, and those who brush with a toothbrush and toothpaste. Unexpectedly, we observed no association with reported tooth brushing, let alone the brushing method or frequency. This may be due to undocumented variance in “brushers,” as important aspects of oral hygiene such as differing toothpaste amounts, brushing times, and toothpaste type, including whether toothpaste included antibiotics like triclosan, were not fully captured in the administered survey. These factors may have confounded our ability to appropriately identify an association between teeth brushing and the microbiome. In a similar vein, we are also missing insight into oral health, such as disease status or symptoms. Gathering this type of information can only be effectively executed by a dentist, which entails identifying a dental professional willing to go into the field to do so. While challenging, additional data about oral hygiene and oral health would provide insight into how these factors shift with the microbiomes of traditional populations.

Finally, we examined the oral-gut microbiome axis across traditional lifestyles. The oral and gut niches are linked by a consistent one-way flow of the saliva, food, and microbes from the mouth through to the colon. Although each niche hosts a distinct microbial community of locally adapted strains [117], the microbes originating in the oral cavity can colonize distal sites in the gut and the two communities have been demonstrated to be predictive of one another [118–120]. An estimated one third of oral microbes are able to colonize the guts of healthy individuals [121], with increasing rates of colonization in individuals with diseases such as bowel cancer [122, 123], rheumatoid arthritis [124], and inflammatory bowel diseases [125]. Given the higher incidence of these diseases in industrialized countries [126–128], an open question remains whether rates of translocation along the oral-gut axis vary with lifestyle, potentially being a risk factor for disease. While our 16S rRNA sequence data lacks the resolution needed to detect translocation specifically, we evaluated whether intra-individual similarity between the oral and gut microbiomes increased along the lifestyle gradient from traditional to agrarian. As expected under this hypothesis, there is a decrease in intra-individual Bray–Curtis distances along the gradient of traditional to agrarian populations within Nepal, although not statistically significant. This finding is similar to results from Cameroonian populations, in which the similarity of the oral-gut axis increased across a gradient of rural to urban populations, albeit not significantly [25]. Power analysis shows that given the effect size estimates at our current sample size, we are quite underpowered to detect a significant relationship, which may explain why there is a visible trend but no statistically significant trend. It is worth noting that we did not

include Nepali expats or American industrialists in this analysis as we did not have paired oral-gut samples for those participants. Other analyses have suggested that the lifestyle gradient is subtle, meaning that detecting the trend between similar lifestyle stages is more challenging. As a result, without the inclusion of the industrialized populations, our ability to observe a significant trend may be dampened. We also note that the sequencing approach used only provides resolution to the genus level, so we were not able to distinguish whether similarity is a result of the translocation of oral strains to the gut versus a homogenization towards similar taxonomic profiles between the niches. Regardless, these results point towards an intriguing hypothesis: oral-gut microbial translocation increases with industrialization, potentially being a risk factor for disease. Future work would entail both increasing sample sizes of individuals across lifestyles as well as using methodologies that reliably track strain sharing across the two body sites.

While we identify new insights into the role of lifestyle in shaping the oral microbiome, there are limitations in our investigation. First, our sample sizes may seem low. While our sample size is sufficient to capture broad compositional differences across subsistence strategy groupings and differential abundance of common taxa, we may not have the ability to detect significant differences in lowly abundant or rare taxa. In practice, however, achieving large sample sizes in studies of traditional human lifestyles can be challenging. For example, the total population size of the Raute is only 550 individuals. We limited our sampling to unrelated individuals, further limiting the sample population. Similarly, the Chepang and Raji reside in small, remote villages scattered throughout Himalaya. As each village only consists of a few hundred individuals, it is very difficult to both reach these villages and also identify unrelated individuals. The industrialist sample size is also low; however, we note that the transition from traditional to agricultural lifestyles was the main focus of this study and that including publicly available data has the potential to introduce geographical and technical confounders. Ultimately, the purpose of this investigation is to examine the transitions across lifestyles, and as it stands, our study has sufficient sample size to identify novel statistical trends and address key gaps within the field. Additional sampling would allow us to dive deeper into these trends and better understand the specific contributors to this relationship.

Second, our investigation utilizes 16S rRNA sequencing, which only identifies taxa at genus-level resolution. As a result, we are unable to compare species or strain level differences in the microbiome across groups [129]. In addition, we currently can only indirectly infer the

functional capabilities of specific microbes. While PICRUSt2 has been demonstrated to be more effective than other predictive functional potential methods, these tools are bound by their respective reference databases and have been demonstrated to miss microbial functional genes, highlighting the importance of validating such findings with metagenomic data [130, 131]. The collection of metagenomic data would allow for a more granular investigation of the microbial strains and functions that differ across lifestyles.

Third, our selection process of the lifestyle factors included in the CCA model may be biased. Our primary considerations were minimizing the number of tests performed to preserve power and minimizing multicollinearity, which is the correlation between predictor variables. Keeping collinear predictors can result in difficulty assessing the relationship between the predictors and outcome and therefore inaccurate estimations of model coefficients. Our method of predictor removal involved selecting variables based on contribution to lifestyle as our goal was to identify which lifestyle factors were associated with the microbiome; however, in doing so, we may have inadvertently removed any factors that are associated with the microbiome by means other than lifestyle. As a result, our CCA model may be inherently biased towards factors that follow the lifestyle trend and not fully capture all possible factors that influence the microbiome. It is for this reason that we also investigated the relationships of all lifestyle factors with the nine differentially abundant taxa to confirm our CCA findings. Future work should entail a more targeted investigation of the lifestyle factors with larger sample sizes to mitigate the concerns of loss of power due to multiple testing and multicollinearity.

## Conclusions

Our investigation of Nepali populations across a variety of human lifestyles expands our understanding of the oral microbiome at a global scale. In conjunction with gut microbiomes collected from the same individuals [15], we find that lifestyle is associated with the composition of both the gut and oral microbiomes, albeit to differing degrees. Metagenomic sequencing would provide finer-scale microbiome data that would allow us to more effectively identify the taxa and functional potential associated with lifestyle. In addition, studies in which industrialization is decoupled from background genetics, geography, and latitude will be essential for identifying the specific factors that result in microbiome differences across populations. Future work will reveal the extent to which oral microbiomes vary around the globe and refine our understanding of the environmental factors involved.

## Methods

### Ethics approval

This work was approved by the Ethical Review Board of the Nepal Health Research Council (NHRC) as well as by the Stanford University Institutional Review Board. Samples were collected between March and April 2016 with informed consent from all participants.

### Sample collection

Saliva samples were collected under informed consent using the DNA Genotek Oral Saliva kit (DNA Genotek, Stittsville, CA, USA) from five populations in Nepal: the Chepang (foragers,  $n=19$ ), Raji (recently settled/transitioned to agriculture,  $n=11$ ), Raute (recently settled/transitioned to agriculture,  $n=14$ ), Tharu (established agriculturalist,  $n=21$ ), and Newar ethnic groups (established agriculturalist,  $n=8$ ). Samples were collected in the winter of 2016 in March and April. Approximately 1 mL of unstimulated saliva samples were stored in the stabilizing buffer provided in the DNA Genotek kit and transported over 7–10 days at room temperature, after which they were stored at  $-20\text{ }^{\circ}\text{C}$  for 2–3 months at the Institute of Medicine in Tribhuvan University, Kathmandu, Nepal.

At the time of saliva sample collection, detailed meta-data was also collected from all participants. These data include demographic, anthropometric measurements, environmental, and dietary data using a survey questionnaire specifically designed for this study (Table S1). Participants' responses to the survey data questionnaires were cleaned and standardized prior to analysis (Table S1).

Following the field work in Nepal, saliva samples were also collected from two populations in the USA with the same kits: the Newar population (expats emigrated from Nepal to the USA,  $n=12$ ) and European Americans (industrialist,  $n=6$ ) (Stanford IRB: 35,580). USA-based samples were collected in the winter of 2016 (November and December). All sampled individuals were unrelated and over 18 years old. Detailed survey data are available for the expat Newar population, but not the European American Industrialists (Table S1). Overall, our cohort included foragers (Chepang,  $n=19$ ), recently settled individuals (total,  $n=25$ ; Raji,  $n=11$ ; Raute,  $n=14$ ), established agriculturalists (total,  $n=29$ ; Tharu,  $n=21$ ; Newar,  $n=8$ ), expats (Newar,  $n=12$ ), and industrialists (European-Americans,  $n=6$ ).

### DNA extraction and 16S rRNA Amplicon sequencing

For samples collected in Nepal, total DNA was extracted at the Institute of Medicine (IOM) in Kathmandu using the Qiagen QIAamp MinElute Virus Spin kit (Qiagen, Germantown, MD, USA) according to the manufacturer's

protocol. Both the remaining original saliva samples and the Qiagen kit-extracted DNA were shipped to Stanford University on dry ice and then stored at either  $-20\text{ }^{\circ}\text{C}$  until sequencing (extracted DNA) or at  $-80\text{ }^{\circ}\text{C}$  (remaining saliva sample). Total DNA was again extracted from saliva samples using the MO BIO PowerSoil DNA Isolation kit (MO BIO, Carlsbad, CA, USA) following the manufacturer's recommended protocol. For samples collected in the USA, they were extracted with both the Qiagen QIAamp MinElute Virus Spin kit and MO BIO PowerSoil DNA Isolation kit in the US. Extraction-negative controls were included in all extractions to evaluate contamination during analysis.

The V4 hypervariable region of the 16S rRNA gene was amplified for all DNA extracts and PCR-negative controls using established primers and protocols [132]. The sample and negative control libraries were multiplexed and sequenced 250 bp single-end on the Illumina MiSeq platform at Stanford University, targeting a minimum of 25,000 reads per sample for accurate relative abundance quantification (Table S1).

### Sequencing data quality control and cleaning

All bioinformatic analyses were conducted in R version 4.1.2, unless otherwise stated. Single-end sequences were cleaned and processed using DADA2 v.1.22.0 [133]. First, reads were trimmed at 150 bp to remove low-quality bases, and then filtered to remove any reads with  $N$  nucleotides or more than two expected errors ( $\text{maxN}=0$ ,  $\text{maxEE}=2$ ,  $\text{truncQ}=2$ , Table S9). Next, sequence variants were inferred by pooling reads across samples ( $\text{pool}=\text{TRUE}$ ). More specifically, 173 samples were pooled using 8,435,257 reads across 674,656 unique sequences. A sequence table was generated, consisting of 173 samples and 5068 amplicon sequence variants (ASV). Eleven percent of the reads were removed as chimeric, resulting in 7,490,294 reads and 1424 ASVs remaining (Table S9). ASVs were classified using the RDP v14 training set [134]. Finally, a phylogenetic tree was generated by performing multiple sequence alignment using DECIPHER v.2.22.0 [135] and then constructing the tree by using a neighbor-joining tree as a starting point via the package phangorn v.2.11.1 [136]. ASVs were then handed off to phyloseq v.1.38.0 for additional cleaning and downstream analysis [137].

Next, predicted contaminants were identified and removed via decontam v.1.14.0 using both the frequency and prevalence methods, thereby removing 19 ASVs [138]. Singletons and any taxa that do not appear at least five times across at least two samples were removed to account for spurious taxa stemming from sequencing errors (Table S9). For alpha diversity analyses, samples were rarefied using the `rarefy_even_depth` function

from the phyloseq package by subsampling to 26,923 sequences (the minimum number of sequences across non-control samples) and then calculating alpha diversity from the rarefied samples (Fig. S19–S20). This was repeated 1000 times to account for randomness in rarefaction, and the mean alpha diversity value was calculated. Non-rarefied counts were used for tools with in-built compositionally aware transformation methods (i.e., centered log-ratio transformation). For all other analyses, counts were transformed to relative abundances using total-sum scaling. Log transformations were conducted with log10 and a pseudocount of 0.0001. Finally, individuals currently taking antibiotics were removed ( $n=2$ ) (Fig. S21), because antibiotics have previously been associated with changes in the oral microbiome [139]. Furthermore, current antibiotic use was found to be marginally significant when compared across Shannon alpha diversity for samples extracted with the Qiagen kit ( $p=0.048$ , Kruskal–Wallis). The post-QC result sample sizes are 60 samples extracted by the PowerSoil kit and 69 samples extracted by the Qiagen kit (Fig. S22). The total number of ASVs across all samples extracted via the Qiagen and PowerSoil kits is 1000 ASVs.

#### Random Forests classifier

Random Forests was first conducted by building 500 trees with 63 samples and 37 categorical variables from the survey data using the R package randomForest v.4.7–1.1 [140]. Random Forests models were subsequently conducted with 500 trees using microbiome data agglomerated to the genus level and transformed to relative abundances. Random Forests models were evaluated via out-of-bag error estimates, as well as assessing confusion matrices. Improvement beyond random chance was assessed using the R package verification v.1.42 [141].

#### Diversity analyses

Five metrics were used to assess alpha diversity—Faith's phylogenetic distance, Fisher's alpha, Shannon alpha, Simpson alpha, and species richness [142–145]. Alpha diversity was calculated after rarefying sample counts to 26,923 counts 1000 times and taking the mean value. Kruskal–Wallis tests were performed to assess significant differences in alpha diversity between the ethnic groups. Dunn's post hoc test was used to identify the group that was driving the differences.

Beta diversity was calculated from relative abundance counts using Bray–Curtis distance, unweighted Unifrac, and weighted Unifrac [40, 146]. The resulting distances were ordinated using PCoA as implemented in phyloseq. PERMANOVA was performed to assess dissimilarity between ethnic groups using vegan v.2.6–4, permuted

99,999 times [147, 148]. Pairwise PERMANOVA comparisons were conducted using pairwiseAdonis v.0.4.1 [149]. Jonckheere–Terpstra tests were used to assess whether the individual PCoA axes followed the lifestyle trend [150].

#### Extraction kit comparison

The Qiagen QIAamp MinElute Virus Spin and MO BIO PowerSoil DNA isolation extraction kits were compared based on overall microbiome beta diversity. Beta diversity was calculated, and PERMANOVA was performed as described above. Comparisons suggest that while there is some qualitative difference between the two kits, there is no statistically significant difference (PERMANOVA,  $p>0.05$ ) (Fig. S3A), and PCoA axes 1 and 2 are highly correlated across kit (PCoA1  $\rho=0.96$ ,  $p<2.2*10^{-16}$ ; PCoA2  $\rho=0.89$ ,  $p<2.2*10^{-16}$ ) (Fig. S3B). For fidelity, diversity analyses were conducted using both extraction kits, but other analyses were conducted using only the dataset extracted via the Qiagen kit, selected for its larger sample size (Fig. S3).

#### Differential abundance analysis

Differential abundance analysis was conducted with microbiome count data agglomerated to the genus level using the ALDEx2 v.1.29.2.1. The Kruskal–Wallis module was used to identify microbes differentially abundant across all lifestyles, whereas the standard  $t$  test module was used for assessing microbes differentially abundant across two conditions. To account for multiple tests, the Benjamini–Hochberg method for  $p$ -value correction was applied [151]. Effect sizes  $>1$  and adjusted  $p$ -values  $<0.05$  were considered significant.

To validate that our findings are robust to the inclusion of sex and age as covariates, the GLM module was used. Because the outputs for the two modules differ, the data was reanalyzed with and without covariates for comparison purposes. Furthermore, because we are missing age data for the American industrialists, this comparison was conducted twice: once with all individuals and sex as the only covariate, and once with the Nepali individuals and both sex and age as the covariates. To account for multiple tests, the Benjamini–Hochberg method for  $p$ -value correction was applied [151]. Adjusted  $p$ -values  $<0.05$  were considered significant.

To identify which microbes followed the lifestyle trend, Jonckheere–Terpstra tests were performed for each genus using microbiome data agglomerated to the genus level and transformed to relative abundances [150]. All  $p$ -values stemming from the Jonckheere–Terpstra tests were corrected for multiple tests using the Benjamini–Hochberg method [151], and adjusted  $p$ -values  $<0.05$  were considered significant.



To validate that the trend observed is not an artifact of sequencing depth, read depth was examined across lifestyles. Sequencing depth was not found to be associated with lifestyle ( $p > 0.05$ , Kruskal-Wallis; Fig. S23 A). For additional confirmation, *Brachymonas* relative abundance was assessed against read depth, and no correlation was found ( $p > 0.05$ ,  $\rho = -0.11$ , Spearman correlation; Fig. S23 B).

#### Associations between lifestyle factors, population, and the microbiome

Correspondence analysis (CA) was performed on the survey metadata using FactoMineR v.2.9 [152]. To determine whether the samples (rows) and lifestyle factors (columns) were significantly associated, a chi-square test of independence was performed. The top 10 most contributing factors for CA1 and CA2 were identified. Alpha and beta diversity values were calculated as described above, and correlations to the CA axes were calculated using Spearman correlation. Canonical correspondence analysis (CCA) was performed with microbiome data agglomerated to the genus level, transformed to relative abundances, and then log-transformed using vegan v.2.6–4 [148]. As there was multicollinearity between the 37 metadata variables (as established by variance inflation factor analysis using the function `vif.cca`,  $VIF > 10$ ), we considered the top 10 categorical lifestyle factors contributing to CA1 or CA2 only to reduce this burden. As some factors were top contributors to both CA1 and CA2, a total of 15 lifestyle factors were used for the CCA model. These 15 factors did not demonstrate collinearity ( $VIF < 10$ ). Model significance and significant lifestyle factors were identified and assessed using the function `anova.cca` from the vegan package, which performs an ANOVA-like permutation test, permuted 999 times.

To identify which lifestyle factors were specifically associated with a particular taxon, a linear model was generated for the relationship between lifestyle variables and a selected taxon and then tested for significance. Taxa were agglomerated to the genus level, transformed to relative abundance, and then log-transformed for visualization. All  $p$ -values stemming from linear models were corrected for multiple tests using the Benjamini–Hochberg method [151], and adjusted  $p$ -values  $< 0.05$  were considered significant. To test whether alcohol use and smoking follow the lifestyle gradient from traditional to expatriate, we performed the Cochran–Armitage test using the function `prop_trend_test` from the package `rstatix` 0.7.2 [153].

#### PICRUSt2 analysis

PICRUSt2 v2.5.2 was conducted using the “`-per_sequence_contrib`” option to predict pathway abundances

for each ASV [42]. Pathway abundances were predicted using the KEGG database FTP release 2022–11–07 [154]. Differential abundance analysis and examination of functions following the lifestyle trend were conducted as previously described using the pathway abundance output. To investigate microbiome functional enrichment, differential abundance analysis was conducted with the KEGG Orthologs (KO) and any significant K genes (prior to multiple test correction) were inputted into Microbiome-Profiler v1.0.0 [155].

#### Network analysis

Network analysis was conducted using the SparCC module in SpiecEasi v.1.1.2 [44, 45]. Microbiome count data agglomerated to the genus level was inputted to generate correlations between taxa. Networks were analyzed for centrality, modularity, and degree distribution using `igraph` v.1.6.0 [156]. Co-abundance groups (CAGs) were generated using the `cluster_fast_greedy` function in the `igraph` package [157]. Networks were plotted in Cytoscape v.3.8.2 [156, 158]. Nodes with 0 edges were removed for visualization and generating CAGs.

#### Comparison of the gut and oral microbiomes

The gut and oral microbiomes were only compared between the same individuals overlapping across both studies (Table S7). PCoA axes were generated for the microbiome datasets for each location as previously described and Spearman correlations were calculated between axes. To compare intra-individual oral-gut microbiome dissimilarity across lifestyles, Bray–Curtis distance was calculated between the gut and oral microbiomes within each individual and then compared across all individuals for differences across lifestyle, with significance determined by a Jonckheere–Terpstra test. To conduct the power analysis, Cohen’s D was calculated between each group and used as effect sizes. Power was subsequently calculated using the package `clinfun` v.1.1.5 [159].

#### Abbreviations

CA	Correspondence analysis
PERMANOVA	Permutational multivariate analysis of variance using distance matrices
PCoA	Principal coordinate analysis
CCA	Canonical correspondence analysis
ANOVA	Analysis of variance
<code>padj</code>	$p$ -Values adjusted for multiple test correction
CAG	Co-abundance group
USA	United States of America
IRB	Institutional Review Board
bp	Base pair
ASV	Amplicon sequence variants
KO	KEGG Orthologs



## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-024-01941-7>.

Additional file 1: S1 Figure: Confusion matrix of random forest classification based on survey data. S2 Figure: Correspondence analysis based on survey data. S3 Figure: Microbiome composition within sample is similar across the two extraction kits tested. S4 Figure: All alpha diversity metrics across lifestyle groups with DNA extracted using the Qiagen kit. S5 Figure: Alpha diversity metrics with DNA extracted using the PowerSoil kit. S6 Figure: Oral microbiome composition extended figures—Axis 2 from Qiagen kit extraction, and ordinations on samples extracted using the PowerSoil kit. S7 Figure: Oral microbiome composition calculated with unweighted Unifrac distance. S8 Figure: Oral microbiome composition calculated with weighted Unifrac distance. S9 Figure: Confusion matrix of Random Forest classification based on microbiome data. S10 Figure: Correlations between diversity metrics and CA axes. S11 Figure: Smoking and *sinu* are associated with several differentially abundant taxa. S12 Figure: Alcohol and Smoking by lifestyle. S13 Figure: Predicted functional abundance significantly differs by lifestyle. S14 Figure: Metabolism pathways form the majority of the significantly differentially abundant predicted functions from PICRUSt2. S15 Figure: Microbiome functional enrichment analysis. S16 Figure: *Fusobacterium* contributes to predicted platinum resistance, which varies by lifestyle. S17 Figure: Correlations between the top three oral and gut microbiome PCoA axes. S18 Figure: Power analysis for the Jonckheere-Terpstra test used in the oral-gut dissimilarity analysis. S19 Figure: Rarefaction curves. S20 Figure: Read depth across samples after read QC. S21 Figure: Comparison of antibiotic use on oral microbiome diversity. S22 Figure: Final sample size per lifestyle per extraction kit. S23 Figure: Read depth against lifestyle and *Brachymonas* relative abundance.

Additional file 2: S1 Table—Sequence, survey, population, and questionnaire info of the sampled individuals. Table S1 describes survey and sequence metadata data collected. Column abbreviations and responses are explained in greater detail in Tab 3. Column names that end with “2” contain responses that were categorized and transformed to a scale of 0–3, in which possible values for binary variables are 0 or 3 (ie. sex) and possible values for continuous variables are 0, 1, 2, or 3 (ie. fuel source). No survey data was collected for the American Industrialists. Tab 2 describes the lifestyle pertaining to each population and their sample sizes. Tab 3 contains the survey questionnaire, including the codes pertaining to each question asked and list of possible responses.

Additional 3: S2 Table—Oral microbiome differential abundance results from ALDEx2. Oral microbiome genera tested for differential abundance across lifestyle. Overall, 2/111 genera were identified as significantly differentially abundant. Kruskal-Wallis module was utilized and *p*-value correction was applied using the Benjamini-Hochberg method. Both unadjusted (kw.ep) and adjusted *p*-values (kw.eBH) are shown, and adjusted *p*-value < 0.05 is the threshold for significance.

Additional file 4: S3 Table—Oral microbiome differential abundance results from ALDEx2 with covariates age and sex. Oral microbiome genera tested for differential abundance across lifestyle with age and sex as covariates. Due to the lack of age data for the American Industrialists, only sex was included in the model tested with all individuals, whereas both sex and age were included in the model tested across the Nepali individuals. The glm module was utilized and *p*-value correction was applied using the Benjamini-Hochberg method. Both unadjusted (kw.ep) and adjusted *p*-values (kw.eBH) are shown, and adjusted *p*-value < 0.05 is the threshold for significance. Tab 1 shows the output for the glm model with no covariates and all individuals. Tab 2 shows the output for the glm model with sex as a covariate and all individuals. Tab 3 shows the output for the glm model with no covariates and only Nepali individuals. Tab 4 shows the output for the glm model with both sex and age as covariates and only Nepali individuals.

Additional file 5: S4 Table—Results of genera tested for following the lifestyle gradient. All genera tested for following the lifestyle gradient using the Jonckheere-Terpstra test followed by the Benjamini-Hochberg method to correct for multiple tests (BHadj\_*p*\_value). Adjusted *p*-value < 0.05 is the threshold for significance. Nine genera significantly follow the lifestyle gradient.

Additional file 6: S5 Table—Associations between differentially abundant microbes and lifestyle factors. Associations between differentially abundant microbes from the Jonckheere-Terpstra test and lifestyle factors were tested via linear models. Linear models were generated between each microbe and each lifestyle factor and then tested for significance, for a total of 333 tested associations. *P*-value correction was applied using the Benjamini-Hochberg method. Both unadjusted and adjusted *p*-values are shown. Adjusted *p*-value < 0.05 is the threshold for significance.

Additional file 7: S6 Table—Predicted functional potential differential abundance results. PICRUSt2 predicted functions were analyzed for differential abundance based on lifestyle. None of the 107 tested functions were found to be significant after multiple test correction, but 21/107 pathways were significant prior to correction. Kruskal-Wallis module in ALDEx2 was utilized and *p*-value correction was applied using the Benjamini-Hochberg method. Both unadjusted (kw.ep) and adjusted *p*-values (kw.eBH) are shown. Adjusted *p*-value < 0.05 is the threshold for significance.

Additional file 8: S7 Table—Samples overlapping between the gut and oral microbiome studies. List of samples overlapping between the gut and oral microbiome studies, along with the samples unique to each microbiome study. The first column “both” lists sample IDs that are associated with both gut and oral samples. The second column “gut\_only” lists sample IDs that are associated with only gut samples. The third column “oral\_only” lists sample IDs that are associated with only oral samples.

Additional file 9: S8 Table—Gut microbiome differential abundance results from ALDEx2. Gut microbiome genera analyzed for differential abundance based on lifestyle via ALDEx2. Overall, 37/136 genera were identified as significantly differentially abundant. Kruskal-Wallis module was utilized and *p*-value correction was applied using the Benjamini-Hochberg method. Both unadjusted (kw.ep) and adjusted *p*-values (kw.eBH) are shown. Adjusted *p*-value < 0.05 is the threshold for significance.

Additional file 10: S9 Table—Read counts through each sequence processing step. Read counts at each step of sequence processing, starting with raw demultiplexed reads and all the way through DADA2, merging, and chimera removal. “Input” column refers to the number of raw reads obtained per sample after sequencing, “filtered” refers to the number of reads remaining after initial read QC, “denoised” refers to the number of reads remaining after denoising in DADA2, “nochim” refers to the number of reads remaining after chimeric sequences were removed, and “retained\_overall” is the total proportion of reads retained following all QC steps from the input amount. Table does not include the samples that failed to pass initial read QC.

### Acknowledgements

First and foremost, we would like to thank all of the participants of this study. We would also like to thank the Nepal Health Research Council (NHRC) under the Government of the Nepal Ministry of Health for providing research permits to conduct our work in Nepal, as well as Mr. Biswash Chepang and Dil Bhakta Maharjan who acted as community liaisons for the Chepang and Newar communities.

### Authors' contributions

The study was designed by ARJ and ERD. Samples were collected by YG, ARJ, and GPG. Data generation was performed by DB, ST, and DMP and supervised by JBS and DAR. EPR, AAS, MG, ARJ, and ERD developed the data analysis plan and interpreted results. Computational and statistical analyses were performed by EPR and MG. Finally, EPR, ARJ, and ERD drafted the manuscript. All authors contributed to its revision and approved the final version.

### Funding

PSU-NIH funded the Computation, Bioinformatics, and Statistics training grant (T32 GM102057) to EPR, NIH grant R35GM146980 to ERD, Seed Award from Center for Human and Evolutionary Genomics (CEHG, <https://cehg.stanford.edu/>), and Faculty Research Fund from New York University Abu Dhabi (NYUAD, <https://nyuad.nyu.edu/>) to ARJ. The Thomas C. and Joan M. Merigan Endowment at Stanford University to DAR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Data availability**

Sequence data can be retrieved from the Sequence Read Archive (SRA) under BioProject number PRJNA1098228. All scripts used in this analysis are available at [http://github.com/davenport-lab/Nepali\\_oral\\_microbiomes](http://github.com/davenport-lab/Nepali_oral_microbiomes).

**Declarations****Ethics approval and consent to participate**

This work was approved by the Ethical Review Board of the Nepal Health Research Council (NHRC) as well as by the Stanford University Institutional Review Board. Samples were collected between March and April 2016 with informed consent from all participants.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare no competing interests.

**Author details**

<sup>1</sup>Department of Biology, Pennsylvania State University, University Park, PA, USA. <sup>2</sup>Genetic Heritage Group, Program in Biology, New York University Abu Dhabi, Abu Dhabi, UAE. <sup>3</sup>Department of Microbiology and Molecular Genetics, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA. <sup>4</sup>Public Health Research Laboratory, Institute of Medicine, Maharajgunj, Kathmandu, Nepal. <sup>5</sup>School of Public Health, University of Adelaide, Adelaide, SA, Australia. <sup>6</sup>Organization for Public Health and Environment Management, Lalitpur, Bagmati, Nepal. <sup>7</sup>Sidney Kimmel Medical College, Philadelphia, PA, USA. <sup>8</sup>Department of Geography, Tribhuvan University, Nepalgunj, Nepal. <sup>9</sup>Department of Medicine, Stanford University, Stanford, CA, USA. <sup>10</sup>Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA. <sup>11</sup>Section of Infectious Diseases, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA. <sup>12</sup>Center for Genomics and Systems Biology, and Public Health Research Center, New York University Abu Dhabi, Abu Dhabi, UAE. <sup>13</sup>Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA.

Received: 2 July 2024 Accepted: 27 September 2024

Published: 4 November 2024

**References**

- Hublin J-J, Ben-Ncer A, Bailey SE, Freidline SE, Neubauer S, Skinner MM, et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature*. 2017;546:289–92.
- Alt KW, Al-Ahmad A, Woelber JP. Nutrition and health in human evolution—past to present. *Nutrients*. 2022;14:3594.
- Clarke SF, Murphy EF, O'Sullivan O, Lucey AJ, Humphreys M, Hogan A, et al. Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*. 2014;63:1913–20.
- David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol*. 2014;15: R89.
- Filippo CD, Cavalieri D, Paola MD, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *PNAS*. 2010;107:14691–6.
- Kim S, Covington A, Pamer EG. The intestinal microbiota: antibiotics, colonization resistance, and enteric pathogens. *Immunol Rev*. 2017;279:90–105.
- Quercia S, Candela M, Giuliani C, Turrone S, Luiselli D, Rampelli S, et al. From lifetime to evolution: timescales of human gut microbiota adaptation. *Front Microbiol*. 2014;5:587.
- Schnorr SL, Sankaranarayanan K, Lewis CM, Warinner C. Insights into human evolution from ancient and contemporary microbiome studies. *Curr Opin Genet Dev*. 2016;41:14–26.
- Sonnenburg ED, Sonnenburg JL. The ancestral and industrialized gut microbiota and implications for human health. *Nat Rev Microbiol*. 2019;17:383–90.
- Weyrich LS. The evolutionary history of the human oral microbiota and its implications for modern health. *Periodontol*. 2000;2021(85):90–100.
- Carter MM, Olm MR, Merrill BD, Dahan D, Tripathi S, Spencer SP, et al. Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. *Cell*. 2023;186:3111–3124.e13.
- Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, et al. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun*. 2014;5:3654.
- Yatsunenkov T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486:222–7.
- Gomez A, Petzelkova KJ, Burns MB, Yeoman CJ, Amato KR, Vlckova K, et al. Gut microbiome of coexisting baaka pygmies and bantu reflects gradients of traditional subsistence patterns. *Cell Rep*. 2016;14:2142–53.
- Jha AR, Davenport ER, Gautam Y, Bhandari D, Tandukar S, Ng KM, et al. Gut microbiome transition across a lifestyle gradient in Himalaya. *PLoS Biol*. 2018;16:e2005396.
- Martínez I, Stegen JC, Maldonado-Gómez MX, Eren AM, Siba PM, Greenhill AR, et al. The gut microbiota of rural Papua New Guineans: composition, diversity patterns, and ecological processes. *Cell Rep*. 2015;11:527–38.
- Baker JL, Mark Welch JL, Kauffman KM, McLean JS, He X. The oral microbiome: diversity, biogeography and human health. *Nat Rev Microbiol*. 2024;22:89–104.
- Radaic A, Kapila YL. The oralome and its dysbiosis: new insights into oral microbiome-host interactions. *Comput Struct Biotechnol J*. 2021;19:1335–60.
- Willis JR, Gabaldón T. The human oral microbiome in health and disease: from sequences to ecosystems. *Microorganisms*. 2020;8: 308.
- Weyrich LS, Dobney K, Cooper A. Ancient DNA analysis of dental calculus. *J Hum Evol*. 2015;79:119–24.
- Warinner C. Dental calculus and the evolution of the human oral microbiome. *J Calif Dent Assoc*. 2016;44:411–20.
- Araújo V, Fehn A-M, Phiri A, Wills J, Rocha J, Gayà-Vidal M. Oral microbiome homogeneity across diverse human groups from southern Africa: first results from southwestern Angola and Zimbabwe. *BMC Microbiol*. 2023;23:226.
- Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, et al. The microbiome of uncontacted Amerindians. *Sci Adv*. 2015;1: e1500183.
- Lassalle F, Spagnoletti M, Fumagalli M, Shaw L, Dyble M, Walker C, et al. Oral microbiomes from hunter-gatherers and traditional farmers reveal shifts in commensal balance and pathogen load linked to diet. *Mol Ecol*. 2018;27:182–95.
- Lokmer A, Aflalo S, Amougou N, Lafosse S, Froment A, Tabe FE, et al. Response of the human gut and saliva microbiome to urbanization in Cameroon. *Sci Rep*. 2020;10:2856.
- Nasidze I, Li J, Schroeder R, Creasey JL, Li M, Stoneking M. High diversity of the saliva microbiome in Batwa pygmies. *PLoS ONE*. 2011;6:e23352.
- Sprockett DD, Martin M, Costello EK, Burns AR, Holmes SP, Gurven MD, et al. Microbiota assembly, structure, and dynamics among Tsimane horticulturalists of the Bolivian Amazon. *Nat Commun*. 2020;11:3772.
- Kidd JM, Sharpton TJ, Bobo D, Norman PJ, Martin AR, Carpenter ML, et al. Exome capture from saliva produces high quality genomic and metagenomic data. *BMC Genomics*. 2014;15:262.
- Yeo L-F, Lee SC, Palanisamy UD, Khalid BAK, Ayub Q, Lim SY, et al. The oral, gut microbiota and cardiometabolic health of indigenous Orang Asli communities. *Front Cell Infect Microbiol*. 2022;12:812345.
- Li J, Quinke D, Horz H-P, Li M, Rzhetskaya M, Raff JA, et al. Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. *BMC Microbiol*. 2014;14:316.
- Ruan X, Luo J, Zhang P, Howell K. The salivary microbiome shows a high prevalence of core bacterial members yet variability across human populations. *NPJ Biofilms Microbiomes*. 2022;8:1–14.
- Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 2017;551:457–63.
- Dikongué E, Ségurel L. Latitude as a co-driver of human gut microbial diversity? *Bioessays*. 2017;39:1600145.

34. Fachrul M, Méric G, Inouye M, Pamp SJ, Salim A. Assessing and removing the effect of unwanted technical variations in microbiome data. *Sci Rep.* 2022;12:22236.
35. Lim MY, Song E-J, Kim SH, Lee J, Nam Y-D. Comparison of DNA extraction methods for human gut microbial community profiling. *Syst Appl Microbiol.* 2018;41:151–7.
36. Handsley-Davis M, Skelly E, Johnson NW, Kapellas K, Lalloo R, Kroon J, et al. Biocultural drivers of salivary microbiota in Australian Aboriginal and Torres Strait Islander children. *Front Oral Health.* 2021;2:641328.
37. Adler CJ, Dobney K, Weyrich LS, Kaidonis J, Walker AW, Haak W, Bradshaw CJA, Townsend G, Soltysiak A, Alt KW, Parkhill J, Cooper A. Cooper Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nat Genet.* 2013;45(4):450–5 <https://doi.org/10.1038/ng.2536>.
38. Quagliariello A, Modi A, Innocenti G, Zaro V, Conati Barbaro C, Ronchitelli A, Boschini F, Cavazzuti C, Dellù E, Radina F, Sperduti A, Bondioli L, Ricci S, Lognoli M, Belcastro MG, Mariotti V, Caramelli D, Mariotti Lippi M, Cristiani E, Martino ME, Muntoni IM, Lari M. Ancient oral microbiomes support gradual Neolithic dietary shifts towards agriculture. *Nat Commun.* 2022;13(1):6927. <https://doi.org/10.1038/s41467-022-34416-0>.
39. Government of Nepal CB of S. National population and housing census 2021. 2021.
40. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol.* 2005;71:8228–35.
41. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome.* 2014;2:15.
42. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol.* 2020;38:685–8.
43. Langille MGI. Exploring linkages between taxonomic and functional profiles of the human microbiome. *mSystems.* 2018;3:e00163-17.
44. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol.* 2012;8:e1002687.
45. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol.* 2015;11:e1004226.
46. Abdill RJ, Adamowicz EM, Blekhman R. Public human microbiome data are dominated by highly developed countries. *PLoS Biol.* 2022;20:e3001536.
47. Hansen TH, Kern T, Bak EG, Kashani A, Allin KH, Nielsen T, et al. Impact of a vegan diet on the human salivary microbiota. *Sci Rep.* 2018;8:5847.
48. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, et al. Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J.* 2016;10:2435–46.
49. Yu K-M, Cho H-S, Lee A-M, Lee J-W, Lim S-K. Analysis of the influence of host lifestyle (coffee consumption, drinking, and smoking) on Korean oral microbiome. *Forensic Sci Int Genet.* 2024;68:102942.
50. Rosas-Plaza S, Hernández-Terán A, Navarro-Díaz M, Escalante AE, Morales-Espinosa R, Cerritos R. Human gut microbiome across different lifestyles: from hunter-gatherers to urban populations. *Front Microbiol.* 2022;13:843170.
51. Smits SA, Leach J, Sonnenburg ED, Gonzalez CG, Lichtman JS, Reid G, et al. Seasonal cycling in the gut microbiome of the Hadza Hunter-Gatherers of Tanzania. *Science.* 2017;357:802–6.
52. Blaser MJ. The past and future biology of the human microbiome in an age of extinctions. *Cell.* 2018;172:1173–7.
53. Carmody RN, Sarkar A, Reese AT. Gut microbiota through an evolutionary lens. *Science.* 2021;372:462–3.
54. Ayeni FA, Biagi E, Rampelli S, Fiori J, Soverini M, Audu HJ, et al. Infant and adult gut microbiome and metabolome in rural Bassa and urban settlers from Nigeria. *Cell Rep.* 2018;23:3056–67.
55. Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell LK, et al. Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun.* 2015;6:6505.
56. Metcalf CJE, Tepekule B, Bruijning M, Koskella B. Hosts, microbiomes, and the evolution of critical windows. *Evolution Letters.* 2022;6:412–25.
57. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM, et al. US immigration Westernizes the human gut microbiome. *Cell.* 2018;175:962–972.e10.
58. Blekhman R, Goodrich JK, Huang K, Sun Q, Bukowski R, Bell JT, et al. Host genetic variation impacts microbiome composition across human body sites. *Genome Biol.* 2015;16:191.
59. Cameron SJS, Huws SA, Hegarty MJ, Smith DPM, Mur LAJ. The human salivary microbiome exhibits temporal stability in bacterial diversity. *FEMS Microbiol Ecol.* 2015;91:fiv091.
60. Galloway-Peña JR, Smith DP, Sahasrabhojane P, Wadsworth WD, Fellman BM, Ajami NJ, et al. Characterization of oral and gut microbiome temporal variability in hospitalized cancer patients. *Genome Medicine.* 2017;9:21.
61. Marotz C, Morton JT, Navarro P, Coker J, Belda-Ferre P, Knight R, et al. Quantifying live microbial load in human saliva samples over time reveals stable composition and dynamic load. *mSystems.* 2021;6:e01182-20.
62. Zaura E, Brandt BW, de Teixeira Mattos MJ, Buijs MJ, Caspers MPM, Rashid M-U, et al. Same exposure but two radically different responses to antibiotics: resilience of the salivary microbiome versus long-term microbial shifts in feces. *mBio.* 2015;6:e01693-15.
63. Hall MW, Singh N, Ng KF, Lam DK, Goldberg MB, Tenenbaum HC, et al. Inter-personal diversity and temporal dynamics of dental, tongue, and salivary microbiota in the healthy oral cavity. *NPJ Biofilms Microbiomes.* 2017;3:1–7.
64. Shi W, Tian J, Xu H, Zhou Q, Qin M. Distinctions and associations between the microbiota of saliva and supragingival plaque of permanent and deciduous teeth. *PLoS ONE.* 2018;13:e0200337.
65. Yang X, He L, Yan S, Chen X, Que G. The impact of caries status on supragingival plaque and salivary microbiome in children with mixed dentition: a cross-sectional survey. *BMC Oral Health.* 2021;21:319.
66. Demirci M. Could Neisseria in oral microbiota modulate the inflammatory response of COVID-19? *Oral Dis.* 2021; <https://doi.org/10.1111/odi.14082>.
67. Yamashita Y, Takeshita T. The oral microbiome and human health. *J Oral Sci.* 2017;59:201–6.
68. Broussard JL, Devkota S. The changing microbial landscape of Western society: Diet, dwellings and discordance. *Mol Metab.* 2016;5:737–42.
69. How KY, Song KP, Chan KG. Porphyromonas gingivalis: an overview of periodontopathic pathogen below the gum line. *Front Microbiol.* 2016;7:53.
70. Elliott SP. Rat bite fever and Streptobacillus moniliformis. *Clin Microbiol Rev.* 2007;20:13–22.
71. Al Bataineh MT, Dash NR, Elkhazendar M, Alnusairat DMH, Darwish IMI, Al-Hajjaj MS, et al. Revealing oral microbiota composition and functionality associated with heavy cigarette smoking. *J Transl Med.* 2020;18:421.
72. Lau SKP, Chan JFW, Tsang C-C, Chan S-M, Ho M-L, Que T-L, et al. Human oropharynx as natural reservoir of Streptobacillus hongkongensis. *Sci Rep.* 2016;6:24419.
73. Dobon B, Musciotto F, Mira A, Greenacre M, Schlaepfer R, Aguilera G, et al. The making of the oral microbiome in Agta hunter-gatherers. *Evol Hum Sci.* 2023;5:e13.
74. Aas JA, Griffen AL, Dardis SR, Lee AM, Olsen I, Dewhirst FE, et al. Bacteria of dental caries in primary and permanent teeth in children and young adults. *J Clin Microbiol.* 2008;46:1407–17.
75. Holgerson PL, Öhman C, Rönnlund A, Johansson I. Maturation of oral microbiota in children with or without dental caries. *PLoS ONE.* 2015;10:e0128534.
76. Obata J, Takeshita T, Shibata Y, Yamanaka W, Unemori M, Akamine A, et al. Identification of the microbiota in carious dentin lesions using 16S rRNA gene sequencing. *PLoS ONE.* 2014;9:e103712.
77. Murugesan S, Al KS. Salivary microbiome and hypertension in the Qatari population. *J Transl Med.* 2023;21:454.
78. Alam J, Lee A, Lee J, Kwon DI, Park HK, Park J-H, et al. Dysbiotic oral microbiota and infected salivary glands in Sjögren's syndrome. *PLoS ONE.* 2020;15:e0230667.
79. Shouval R, Eshel A, Dubovski B, Kuperman AA, Danylesko I, Fein JA, et al. Patterns of salivary microbiota injury and oral mucositis in recipients of allogeneic hematopoietic stem cell transplantation. *Blood Adv.* 2020;4:2912–7.

80. Zhou P, Manoil D, Belibasakis GN, Kotsakis GA. Veillonellae: beyond bridging species in oral biofilm ecology. *Front Oral Health*. 2021;2:774115.
81. Santonocito S, Giudice A, Polizzi A, Troiano G, Merlo EM, Sclafani R, et al. A cross-talk between diet and the oral microbiome: Balance of Nutrition on Inflammation and Immune System's Response during Periodontitis. *Nutrients*. 2022;14:2426.
82. Maki KA, Ganesan SM, Meeks B, Farmer N, Kazmi N, Barb JJ, et al. The role of the oral microbiome in smoking-related cardiovascular risk: a review of the literature exploring mechanisms and pathways. *J Transl Med*. 2022;20:584.
83. Shapiro H, Goldenberg K, Ratiner K, Elinav E. Smoking-induced microbial dysbiosis in health and disease. *Clin Sci*. 2022;136:1371–87.
84. Suzuki N, Nakano Y, Yoneda M, Hirofujii T, Hanioka T. The effects of cigarette smoking on the salivary and tongue microbiome. *Clin Exp Dent Res*. 2022;8:449–56.
85. Wirth R, Maróti G, Mihók R, Simon-Fiala D, Antal M, Pap B, et al. A case study of salivary microbiome in smokers and non-smokers in Hungary: analysis by shotgun metagenome sequencing. *J Oral Microbiol*. 2020;12:1773067.
86. Yates JAF, Velsko IM, Aron F, Posth C, Hofman CA, Austin RM, et al. The evolution and changing ecology of the African hominid oral microbiome. *PNAS*. 2021;118:e202165118.
87. Klein MI, Duarte S, Xiao J, Mitra S, Foster TH, Koo H. Structural and molecular basis of the role of starch and sucrose in *Streptococcus mutans* biofilm development. *Appl Environ Microbiol*. 2009;75:837–41.
88. Saito S, Aoki Y, Tamahara T, Goto M, Matsui H, Kawashima J, et al. Oral microbiome analysis in prospective genome cohort studies of the Tohoku Medical Megabank project. *Front Cell Infect Microbiol*. 2021;10:604596.
89. Atkinson FS, Foster-Powell K, Brand-Miller JC. International tables of glycemic index and glycemic load values: 2008. *Diabetes Care*. 2008;31:2281–3.
90. Soong YY, Tan SP, Leong LP, Henry JK. Total antioxidant capacity and starch digestibility of muffins baked with rice, wheat, oat, corn and barley flour. *Food Chem*. 2014;164:462–9.
91. Esteban-Fernández A, Zorraquín-Peña I, González de Llano D, Bartolomé B, Moreno-Arribas MV. The role of wine and food polyphenols in oral health. *Trends Food Sci Technology*. 2017;69:118–30.
92. Narduzzi L, Agulló V, Favarì C, Tosi N, Mignogna C, Crozier A, et al. (Poly) phenolic compounds and gut microbiome: new opportunities for personalized nutrition. *Microbiome Res Rep*. 2022;1:16.
93. Tamura M, Imaizumi R, Saito T, Watanabe T, Okamoto T. Studies of the texture, functional components and *in vitro* starch digestibility of rolled barley. *Food Chem*. 2019;274:672–8.
94. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. *Nat Genet*. 2007;39:1256–60.
95. Poole AC, Goodrich JK, Youngblut ND, Luque GG, Ruard A, Sutter JL, et al. Human salivary amylase gene copy number impacts oral and gut microbiomes. *Cell Host Microbe*. 2019;25:553–564.e7.
96. Wursch P, Pi-Sunyer FX. The role of viscous soluble fiber in the metabolic control of diabetes. *Diabetes Care*. 1997;20:1774–80.
97. Li J, Kaneko T, Qin L-Q, Wang J, Wang Y, Sato A. Long-term effects of high dietary fiber intake on glucose tolerance and lipid metabolism in GK rats: comparison among barley, rice, and cornstarch. *Metabolism*. 2003;52:1206–10.
98. Jenkins DJ, Kendall CW, Augustin LS, Franceschi S, Hamidi M, Marchie A, et al. Glycemic index: overview of implications in health and disease. *Am J Clin Nutr*. 2002;76:2665–2735.
99. Liu S, Xie G, Chen M, He Y, Yu W, Chen X, et al. Oral microbial dysbiosis in patients with periodontitis and chronic obstructive pulmonary disease. *Front Cell Infect Microbiol*. 2023;13:13.
100. Liu Y, Liu H, Rong Y, Shi Q, Yang Q, Li H, et al. Alterations of oral microbiota are associated with the development and severity of acute pancreatitis. *J Oral Microbiol*. 2023;15:2264619.
101. Manzoor M, Lommi S, Furuholm J, Sarkkola C, Engberg E, Raju S, et al. High abundance of sugar metabolisers in saliva of children with caries. *Sci Rep*. 2021;11:4424.
102. Raju SC, Lagström S, Ellonen P, de Vos WM, Eriksson JG, Weiderpass E, et al. Gender-specific associations between saliva microbiota and body size. *Front Microbiol*. 2019;10:767.
103. Petri RM, Wetzels SU, Kumar M, Khaioa-ard R, Zebeli Q. Adaptive responses in short-chain fatty acid absorption, gene expression, and bacterial community of the bovine rumen epithelium recovered from a continuous or transient high-grain feeding. *J Dairy Sci*. 2019;102:5361–78.
104. Sapkota S, Shrestha S. An explorative survey on Sisnu: a wonder but highly underutilized crop of Nepal. *J Pharmacogn Phytochem*. 2018;7:832–3.
105. Kregiel D, Pawlikowska E, Antolak H. *Urtica* spp.: Ordinary plants with extraordinary properties. *Molecules*. 2018;23:1664.
106. Harrison F, Furner-Pardoe J, Connelly E. An assessment of the evidence for antibacterial activity of stinging nettle (*Urtica dioica*) extracts. *Access Microbiol*. 2022;4:000336.
107. Carda-Diéguez M, Moazzez R, Mira A. Functional changes in the oral microbiome after use of fluoride and arginine containing dentifrices: a metagenomic and metatranscriptomic study. *Microbiome*. 2022;10:159.
108. Adams SE, Arnold D, Murphy B, Carroll P, Green AK, Smith AM, et al. A randomised clinical study to determine the effect of a toothpaste containing enzymes and proteins on plaque oral microbiome ecology. *Sci Rep*. 2017;7:43344.
109. Kong J, Zhang G, Xia K, Diao C, Yang X, Zuo X, et al. Tooth brushing using toothpaste containing theaflavins reduces the oral pathogenic bacteria in healthy adults. *3 Biotech*. 2021;11:150.
110. Calderon SJ, Chung SY, Fields CJ, Mortimer NT. Children tooth brushing behavior and oral microbiota: a pilot study. *Oral*. 2021;1:112–21.
111. Sukkarwalla A, Ali SM, Lundberg P, Tanwir F. Efficacy of Miswak on oral pathogens. *Dent Res J (Isfahan)*. 2013;10:314–20.
112. Sofrata AH, Claesson RLK, Lingström PK, Gustafsson AK. Strong antibacterial effect of miswak against oral microorganisms associated with periodontitis and caries. *J Periodontol*. 2008;79:1474–9.
113. Abhary M, Al-Hazmi A-A. Antibacterial activity of Miswak (*Salvadora persica* L.) extracts on oral hygiene. *J Taibah Univ Sci*. 2016;10:513–20.
114. Rifaey N, AlAdwani M, Karched M, Baskaradoss JK. A clinical investigation into the efficacy of miswak chewing sticks as an oral hygiene aid: a crossover randomized trial. *Int J Dental Hygiene*. 2021;19:223–30.
115. Sofrata A, Brito F, Al-Otaibi M, Gustafsson A. Short term clinical effect of active and inactive *Salvadora persica* miswak on dental plaque and gingivitis. *J Ethnopharmacol*. 2011;137:1130–4.
116. Brooks JK, Bashirelahi N, Reynolds MA. Charcoal and charcoal-based dentifrices: a literature review. *J Am Dent Assoc*. 2017;148:661–70.
117. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017;550:61–6.
118. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. *Nature*. 2014;509:357–60.
119. Kageyama S, Sakata S, Ma J, Asakawa M, Takeshita T, Furuta M, et al. High-resolution detection of translocation of oral bacteria to the gut. *J Dent Res*. 2023;102:752–8.
120. Liao C, Rolling T, Djukovic A, Fei T, Mishra V, Liu H, et al. Oral bacteria relative abundance in faeces increases due to gut microbiota depletion and is linked with patient outcomes. *Nat Microbiol*. 2024;9:1555–65.
121. Schmidt TS, Hayward MR, Coelho LP, Li SS, Costea PI, Voigt AY, et al. Extensive transmission of microbes along the gastrointestinal tract. *eLife*. 2019;8:e42693.
122. Flynn KJ, Baxter NT, Schloss PD. Metabolic and community synergy of oral bacteria in colorectal cancer. *mSphere*. 2016;1: <https://doi.org/10.1128/msphere.00102-16>.
123. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol*. 2014;10:766.
124. Zhang X, Zhang D, Jia H, Feng Q, Wang D, Liang D, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat Med*. 2015;21:895–905.
125. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in New-Onset Crohn's disease. *Cell Host Microbe*. 2014;15:382–92.

126. Arnaout AY, Nerabani Y, Douba Z, Kassem LH, Arnaout K, Shabouk MB, et al. The prevalence and risk factors of irritable bowel syndrome (PRIBS study) among adults in low- and middle-income countries: a multicenter cross-sectional study. *Health Sci Rep.* 2023;6:e1592.
127. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68:394–424.
128. GBD 2021 Rheumatoid Arthritis Collaborators. Global, regional, and national burden of rheumatoid arthritis, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol.* 2023;5:e594–610.
129. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep.* 2021;11:3030.
130. Sun S, Jones RB, Fodor AA. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome.* 2020;8:46.
131. Matchado MS, Rühlemann M, Reitmeier S, Kacprowski T, Frost F, Haller D, et al. On the limits of 16S rRNA gene-based metagenome prediction and functional profiling. *Microb Genom.* 2024;10:001203.
132. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 2012;6:1621–4.
133. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13:581–3.
134. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The ribosomal database project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 2009;37 suppl\_1:D141–5.
135. Wright ES. Using DECIPHER v2.0 to analyze big biological sequence data in R. *The R J.* 2016;8:352.
136. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics.* 2011;27:592–3.
137. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE.* 2013;8:e61217.
138. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome.* 2018;6:226.
139. Cheng X, He F, Si M, Sun P, Chen Q. Effects of antibiotic use on saliva antibody content and oral microbiota in sprague dawley rats. *Front Cell Infect Microbiol.* 2022;12:721691.
140. Liaw A, Wiener M. Classification and regression by random forest. *Sci Rep.* 2002;2:13458.
141. Laboratory N-RA. Verification: weather forecast verification utilities. 2015.
142. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Cons.* 1992;61:1–10.
143. Fisher RA, Corbet AS, Williams CB. The relation between the number of species and the number of individuals in a random sample of an animal population. *J Anim Ecol.* 1943;12:42–58.
144. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* 1948;27:379–423.
145. Simpson EH. Measurement of diversity. *Nature.* 1949;163:688–688.
146. Bray JR, Curtis JT. An ordination of the upland forest communities of Southern Wisconsin. *Ecol Monogr.* 1957;27:325–49.
147. Anderson MJ. Permutational multivariate analysis of variance (PERMANOVA). In: *Wiley StatsRef: Statistics Reference Online.* Hoboken: Wiley; 2017. p. 1–15.
148. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
149. Arbizu PM. pmartinezarbizu/pairwiseAdonis. 2024.
150. Lunneborg CE. Jonckheere–Terpstra Test. In: *Wiley StatsRef: Statistics Reference Online.* Hoboken: Wiley; 2014.
151. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B (Methodol).* 1995;57:289–300.
152. Lê S, Josse J, Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw.* 2008;25:1–18.
153. Kassambara A. rstatix: pipe-friendly framework for basic statistical tests. 2023.
154. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28:27–30.
155. Chen M, Yu G. MicrobiomeProfiler: an R/shiny package for microbiome functional enrichment analysis. 2023.
156. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal.* 2005;Complex Systems:1695.
157. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. *Phys Rev E.* 2004;70: 066111.
158. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
159. Seshan VE, Whiting K. clinfun: clinical trial design and data analysis functions. 2023.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.