#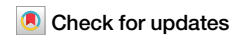 Gut microbiome composition reveals the distinctiveness between the Bengali people and the Indigenous ethnicities in Bangladesh

Check for updates

Ishtiaque Ahammad[1,5], Arittra Bhattacharjee[1,5], Zeshan Mahmud Chowdhury[1,5], Anisur Rahman [1,5], Mohammad Uzzal Hossain[1,5], Gourab Dewan[2], Shiny Talukder[2], Keshob Chandra Das[3], Chaman Ara Keya[4] & Md Salimullah [3] ✉

Ethnicity has a significant role in shaping the composition of the gut microbiome, which has implications in human physiology. This study intends to investigate the gut microbiome of Bengali people as well as several indigenous ethnicities (Chakma, Marma, Khyang, and Tripura) residing in the Chittagong Hill Tracts areas of Bangladesh. Following fecal sample collection from each population, part of the bacterial 16 s rRNA gene was amplified and sequenced using Illumina NovaSeq platform. Our findings indicated that Bangladeshi gut microbiota have a distinct diversity profile when compared to other countries. We also found out that Bangladeshi indigenous communities had a higher *Firmicutes* to *Bacteroidetes* ratio than the Bengali population. The investigation revealed an unclassified bacterium that was differentially abundant in Bengali samples while the genus *Alistipes* was found to be prevalent in Chakma samples. Further research on these bacteria might help understand diseases associated with these populations. Also, the current small sample-sized pilot study hindered the comprehensive understanding of the gut microbial diversity of the Bangladeshi population and its potential health implications. However, our study will help establish a basic understanding of the gut microbiome of the Bangladeshi population.

The human gut microbiome which is composed of an enormous collection of microorganisms plays a crucial role in modulating various physiological processes, metabolic diseases, and immunological dysregulation. The composition of the gut microbiome is influenced by a multitude of factors, including diet, lifestyle, environment, genetics, and ethnicity[1]. They forge a symbiotic relationship with the host in the gastrointestinal tract, where they are monitored by the innate immune system using pattern recognition receptors such as toll-like receptors and Nucleotide-binding oligomerization domain-containing protein (NOD)-like receptors[2]. A number of these bacteria are involved in regulating host metabolism by producing metabolites such as folate, indoles, secondary bile acids, trimethylamine-N-oxide, neurotransmitters (e.g., serotonin,

gamma amino butyric acid), and short-chain fatty acids[3]. Alteration in the host–symbiont relationships in the gut has been observed in diseases like obesity, type 2 diabetes mellitus, non-alcoholic liver disease, cardiometabolic diseases, and so on[4]. Since different ethnic groups have been reported to have variations in genetics, food habits, lifestyle, their alterations can be greatly influenced by ethnicity. As a result, some ethnic groups are predisposed to certain metabolic diseases, while others are inclined to others[5]. For example, high-risk variants of *NOD2* are less prevalent in African Americans which makes their microbiome less prone to Crohn's Disease[6,7]. Similarly, the *lactase* gene varies in different populations[8]. In the Japanese population higher abundance of *Bifidobacterium* has been observed due to their characteristic *LCT*

[1]Bioinformatics Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka 1349, Bangladesh. [2]Rangamati Medical College, Hospital Road, Rangamati-4500 Rangamati, Bangladesh. [3]Molecular Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka 1349, Bangladesh. [4]Department of Biochemistry and Microbiology, North South University, Bashundhara, Dhaka 1229, Bangladesh. [5]These authors contributed equally: Ishtiaque Ahammad, Arittra Bhattacharjee, Zeshan Mahmud Chowdhury, Anisur Rahman, and Mohammad Uzzal Hossain. ✉e-mail: salim2969@gmail.com

genotypes[9,10]. Hence, gut microbiome studies on ethnic groups can be helpful in identifying prognostic microbial biomarkers for various diseases.

The genetic makeup, dietary patterns, and geographic regions shape the gut microbiome of various ethnicities in a distinct pattern[11–14]. Previously, a number of gut microbiome studies in South-East Asia have been conducted. Gut microbiomes of populations living in India and China which are geographically close to Bangladesh, showed an abundance of *Prevotella* and *Bacteroidetes* respectively[5]. In India, gut microbiomes from the Adi, Apatani, and Nyshi tribes of Arunachal Pradesh (a state in Northeastern India) demonstrated the predominance of *Firmicutes*, *Bacteroidates*, *Actinobacteria*, and *Proteobacteria*. On the other hand, Mongoloid and Proto-Australoid tribes of India were dominated by *Prevotella*[15,16]. However, studies from ethnic groups in Bangladesh are still not available. Bangladeshi ethnic groups have diverse dietary habits. Bengali food customs stem from their agrarian culture, which is defined by a profusion of rice, fish, and different vegetables. Dishes like biryani and kebabs are examples of the historical effects of Mughal and British dynasties on the cuisine. Seasonal fruit preparations, street snacks like shingara and samosas, and traditional cakes called "Pitha" are all common[17]. Bangladeshi tribal people use forest resources to eat diversely. They feed on 'bans koral,' bamboo shoots from *Melocanna baccifera* and *Bambusa tulda* during the rainy season. Fresh or dried wild plants commonly accompany staple grain recipes. Acidic leaves are often eaten as salad or chutney. Woody perennials including *Albizia procera*, wild mango, and *Daemonorops jenkinsianus* are vegetables for these ethnic groups. Lentinus, Shizophyllum, and Jew's Ear mushrooms from decomposing wood are eaten. Wild banana inflorescences and leaf sheath soft cores are also edible. During food shortages, banana core cooked with rice or bran is crucial[18].

Bangladesh is a country inhabited by many different ethnicities. Bengalis are the largest among them. Other major groups are Chakma, Marma, Khyang, and Tripura with distinguishable food habits and lifestyles. The Chittagong Hill Tracts in the southeastern parts of Bangladesh are home to the Chakma, Marma, Khyang, and Tripura people. Bengali people (Bengali language speakers) have Indian ancestry with Proto-Australoid, Caucasoid, and Mongoloid genetics[19–21]. On the other hand, Chakma, Marma, Khyang, and Tripura people (Tibeto-Burman language speakers) have predominance of Tibeto-Burman genetics. Tibeto-Burman populations originated from northwestern China and moved to the South. They have interbred with numerous southern tribes for the past 2600 years, creating specific genetic traits among southern Tibeto-Burman populations[22,23]. Hence, Tibeto-Burman speakers from Northeast India are more distinctive than the ones living in the Himalayas[23]. In Bangladesh, the Tibeto-Burman populations (e.g., Chakma, Marma, Khyang, Tripura etc.) have higher similarity with Northeast Indian Tibeto-Burman but they contain more mainland Indian ancestry[24]. In this study, we investigated the gut microbiomes of Bengali, Chakma, Marma, Khyang, and Tripura populations in order to determine whether Bangladeshi Tibeto-Burman populations differ from Bengali people in terms of their gut microbiomes.

Methods
Sample collection
Prior to sample collection, written informed consent was obtained from all participants. Ethical approval and necessary permits were obtained from the National Institute of Biotechnology Ethical Review Committee, Bangladesh (NIBERC2022-01). All ethical regulations relevant to human research participants were followed. Supplementary Data 1 contains detailed sample information regarding age, gender, and cohort.

All of the indigenous volunteers (Chakma, Marma, Khyang, and Tripura) in this study reside in the Rangamati district (22°37'60 N 92°12'0E) of the Chittagong Hill Tracts region. Bengali samples were collected from Dhaka Division (23.9536° N, 90.1495° E). Based on ethnicity, the participants were divided into five groups (Bengali, Chakma, Marma, Khyang, and Tripura). A total of 55 individuals were sampled, of which 13 were Bengali, 15 were Chakma, 10 were Khyang, 6 were

Marma, and 11 were Tripura. Fecal samples of the participants were collected using sterile stool collection tubes. The samples were then transported to the National Institute of Biotechnology using Icebox and subsequently stored at −80 °C temperature. Fecal DNA extraction was executed using the PureLink™ Microbiome DNA Purification Kit (Catalog number: A29790). Specialized beads were used along with a combination of heat, chemical, and mechanical disruption to lyse the microorganisms efficiently. Precipitation with a proprietary cleaning buffer removed inhibitors. After that, the samples were placed in spin columns, and the DNA attached to the column was washed once before elution. DNA concentration and purity were estimated by Thermo Scientific NanoDrop 2000/2000c and then stored at −20 °C.

Amplicon generation and library preparation for sequencing
Amplicons were generated using the 341 F (5'-CCTAYGGGRBGCAS-CAG-3') and 806 R (5'-GGACTACNNGGGTATCTAAT-3') primers that targeted the V3-V4 region of the bacterial 16 S rRNA gene. All PCR reactions were performed with 15 μL of Phusion® High-Fidelity PCR Master Mix (New England Biolabs), 0.2 μM of forward and reverse primers, and around 10 ng template DNA. For thermal cycling, initial denaturation at 98 °C for 1 min was followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 50 °C for 30 s, and elongation at 72 °C for 30 s. Final elongation was carried out for 5 min at 72 °C. TruSeq® DNA PCR-Free Sample Preparation Kit (Illumina, USA) was used as per the manufacturer's protocol for sequencing library preparation, and index codes were appended. The quality of the library was evaluated by employing the Qubit@2.0 Fluorometer (Thermo Scientific) and the Agilent Bioanalyzer 2100 system. Finally, the library was sequenced on an Illumina NovaSeq platform, resulting in 250 bp paired-end reads.

Data pre-processing and quality control
The paired end sequences were converted into QIIME2 format for data pre-processing, quality control, taxonomic assignment, differential abundance identification, functional analysis using QIIME2 platform (version 2021.4.0)[25]. More specifically, the data pre-processing of paired end sequences were performed using the DADA2 plug-in within QIIME2[26]. DADA2 filtered noisy reads, performed error correction in marginal sequences, removed chimeras and singletons, joined denoised paired-end reads, and also de-replicated the filtered reads. The features produced by DADA2 are denoted as amplicon sequence variants.

Taxonomic assignment
For taxonomic assignment, a pre-trained classifier based on the Naive Bayes machine-learning model was implemented. This model was trained on 99% sequence similarity of Greengenes 13_8 data. This classifier was then deployed for taxonomy assignment of amplicon sequence variants.

Diversity analysis
Several diversity metrics in QIIME2 require a rooted phylogenetic tree generated from the amplicon sequence variants of the sampled data. A reference-based fragment insertion method, using q2-fragment-insertion tool, was applied to construct the rooted tree for this purpose. Greengenes 13_8 data was used as a reference database in the q2-fragment-insertion tool[27,28]. The sequencing depth of the samples was 3525 to observe the richness. This phenomenon was checked with the alpha rarefaction curve generated by the q2-diversity tool. The microbiome within and between samples was calculated by the core-metric-phylogenetic method of the q2-diversity tool. This method computes several alpha (Observed features, Shanon diversity, Faith's phylogenetic diversity, Pielou evenness) and beta (Jaccard distance, Bray–Curtis distance, unweighted UniFrac distance, and weighted UniFrac distance) diversity metrics altogether[29]. Based on each beta diversity metric, this command also performed principal coordinates analysis (PCoA)[30]. To visualize the PCoA plots for every beta diversity metric, EMPeror visualization tool was utilized to generate the figures[31]. Several statistical

tests were conducted during diversity analysis such as Kruskal–Wallis H test, two-way ANOVA, paired *t*-test and PERMANOVA test[32–35]. The boxplot() function in R was used to draw boxplots based on alpha diversity values. This function follows the 1.5 IQR method for detecting outliers which are placed above and below the whiskers on the boxplots. Here, only one sample was identified as outlier (Sample ID: BBT19). The result including the outlier has been provided in Supplementary Fig. 1. After removal of the outlier, the number of samples used for analysis were reduced from 55 – 54.

Differential abundance test

To classify the features that were differentially abundant across various sample groups the analysis of the composition of microbiomes (ANCOM) method was applied by the q2-composition tool[36]. This statistical framework was deployed at the genus level. The minimum sample size for each feature was set to 27 (half of the total samples) because ANCOM fails to manage false discovery rates at sample sizes <10 as well as to remove very lowly abundant features[37]. For linear discriminant analysis by the LEfSe, the feature table was collapsed at the genus level and the minimum sample size was chosen at 27[38]. This tool first performs the non-parametric Kruskal-Wallis (KW) sum-rank test to identify the features which had significant differential abundance across different metadata categories. Finally, LEfSe applies linear discriminant analysis to compute the effect size of each differentially abundant feature and plot the linear discriminant analysis score in the log10 scale. Result of both ANCOM and LEfSe analysis conducted for all samples including the previously mentioned outliers has been provided in Supplementary Fig. 2.

Functional analysis

BURRITO, an interactive visualization web server (http://elbo-spice.cs.tau.ac.il/shiny/burrito/), was utilized to explore the taxa-function relationship within the samples of the study[39]. To acquire gene contents and functional annotations, this tool adopts the PICRUSt and KEGG Orthology databases, respectively[40,41]. At first, features with a sample size of <27 were filtered from the original feature table to remove very low abundant taxon. The q2-vsearch tool was employed for closed-reference clustering of retrieved features at 97% identity based on greengenes 97% OTU IDs as reference[42,43]. Thus the acquired OTU table was then converted to the appropriate table format for BURRITO input.

Comparative analysis with tropical and subtropical countries

Since Bangladesh is a tropical country, we compared the gut microbiome of Bangladeshi samples with several tropical and subtropical countries to see if there is any similarity between them. To do this we took sequence data from publicly available 16 s rRNA amplicon sequence data from the NCBI Sequence Read Archive and MG-RAST databases. Only healthy / control samples were taken from the selected datasets. Countries included in the comparative analysis were Australia, Egypt, India, Indonesia, Malaysia, Mexico, Thailand, and Vietnam. Information about the samples taken from NCBI and MG-RAST has been presented in Supplementary Data 2 and Supplementary Data 3 respectively. The samples downloaded from MG-RAST were processed using q2-deblur while the samples downloaded from NCBI were processed using q2-dada2[34,44]. This is because the samples were taken from different regions and were prepared and sequenced using different methods. Amplicon sequencing using the Illumina MiSeq technology was done on Indian samples, encompassing the V3-V4 region of bacterial 16 S rDNA. On the other hand, the V4 region of bacterial 16 S ribosomal RNA genes from Malawi, Amerindians, and the United States was amplified and sequenced using an Illumina HiSeq 2000. The V1-V3 region of the 16 S ribosomal RNA (rRNA) gene of the Mongolian samples was amplified and sequenced using pyrosequencing on a Roche GS FLX. The samples were merged with the qiime feature-table merge method. All the samples were rarefied to the same depth (3525).

These foreign samples ($n = 181$) were then compared with all Bangladeshi samples ($n = 55$) using the q2-diversity tool via Alpha diversity and Beta Diversity analysis. Afterwards, a phylogenetic rooted tree was generated by the sepp fragment insertion approach. This rooted phylogeny was used to create a Unweighted Pair Group Method with Arithmetic Mean tree based on the Unweighted UniFrac metric by beta-rarefaction command of the q2-diversity tool[45].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Bengali population possessed lower *Firmicutes* to *Bacteroidetes* ratio compared to Indigenous groups

Among the 19 classified phyla, *Firmicutes* and *Bacteroidetes* were the most prevalent (48% and 34% of the total) in the gut microbiome of all Bangladeshi populations. Other phyla with higher abundance were Proteobacteria (14%), Actinobacteria (3%), Tenericutes (0.5%), etc (Supplementary Fig. 3). Interestingly, 100% of taxa from the *Bacteroidetes* phylum belonged to the *Bacteroidia* class. On the other hand, considering *Firmicutes* phylum, 87% of features are *Clostridia* class, 11% *Bacilli*, and 2% are *Erysipelotrichi* (Supplementary Fig. 4). In Supplementary Data 4, the taxonomic classification of each feature id is documented, along with the confidence value.

The normalized abundance of the top ten genera across different cohorts is represented in cpm (counts per million) (Fig. 1a). The *Prevotella* genus abundance was relatively similar in all cohorts. Nevertheless, the prevalence of the *Bacteroides* genus was drastically higher in the Chakma population and very low in Marma and Tripura samples. Moreover, Chakma tribal group also contained *Faecalibacterium*, *Roseburia*, and uncharacterized genera from the *Ruminococcaceae* family and *Bacteroidales* order in a relatively higher amount. On the other hand, *Streptococcus* and *Megasphaera* genera were highly abundant in the Bengali and Tripura populations and the Marma cohort contained these genera in relatively lower quantities. The Khyang group contained the genus *Dialister* and uncharacterized genera from *Enterobacteriaceae* and *Ruminococcaceae* families. About 0.05% amplicon sequence variants were classified as archaea of the *Methanobacteriaceae* family, of which only 0.1% were from the *Methanosphaera* genus, and the rest of the features belonged to the *Methanobrevibacter* genus (Fig. 1b). Twenty-four families of features were found as core features and the frequency heatmap of these features revealed that there were two main clusters, one with higher frequency and another with lower frequency (Fig. 1c). The *Streptococcaceae*, *Prevotellaceae*, *Veillonellaceae*, *Lachnospiraceae*, *Ruminococcaceae*, *Enterobacteriaceae*, and *Bacteroidaceae* families were in the higher frequency cluster. Most of the samples from the Bengali population clustered together and contained these higher frequency microbial families in relatively lower amounts. The median value of the *Firmicutes* to *Bacteroidetes* ratio was highest (2.018) in the case of Tripura samples, while the Bengali population had the lowest (0.877) median ratio (Fig. 1d). All other cohorts (Chakma, Khyang, Tripura) had a median *Firmicutes* to *Bacteroidetes* ratio of more than one.

The Bengali population showed inferior microbiome richness than other

Ethnic cohorts. Alpha diversity depicts the diversity within the sample. Alpha diversity is measured via Shannon diversity, Observed features, Faith pd and Pielou evenness parameters. Shannon diversity calculates both the number of species in a community and their relative abundance. The median of Shannon diversity was relatively lower in the Bengali group compared to others (Fig. 2). For ethnicity based cohorts, Kruskal-Wallis test for all groups had a *p*-value of 0.0033 (Supplementary Data 5). Observed features counted the number of distinct features present in the cohorts (Fig. 2). Chakma, Marma, Khyang and Tripura populations had higher species richness than the Bengalis (*p*-value = 0.0003) (Supplementary

Data 6). Faith pd incorporated information of the evolutionary relationships between different bacterial species. Bengali samples had lower Fatih *p*-values than indigenous populations (Fig. 3). However, the difference in Faith pd score was statistically insignificant (*p*-value = 0.3816)

(Supplementary Data 7). Pielou evenness measured how evenly different species' abundances were distributed within a community (Fig. 3). There was no significant difference in Pielou evenness profile between indigenous populations (*p*-value = 0.0631) (Supplementary Data 8).
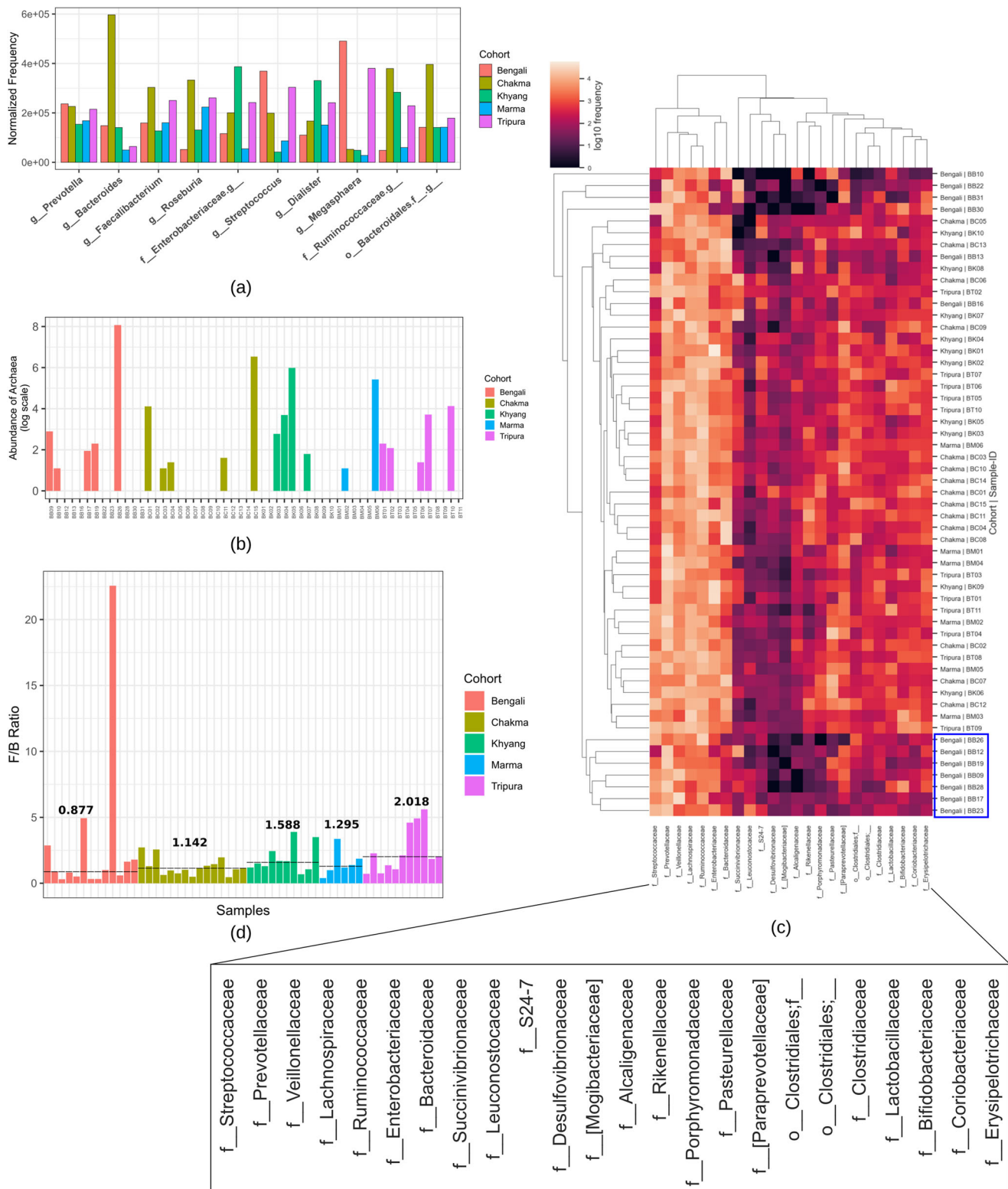
Fig. 1 | The abundance of the top prevailing bacterial genera, archaeal features, core features, and *Firmicutes* to *Bacteroidetes* ratio in various Bangladeshi cohorts. **a** The bar plot represents the normalized frequency of the top genera among different cohorts (**b**) Several archaeal species appeared to be present in different samples from each cohort. **c** The frequency of core features resulted in two primary clusters, with somewhat varied abundances in some Bengali samples. **d** The median ratio of *Firmicutes* to *Bacteroidetes*.
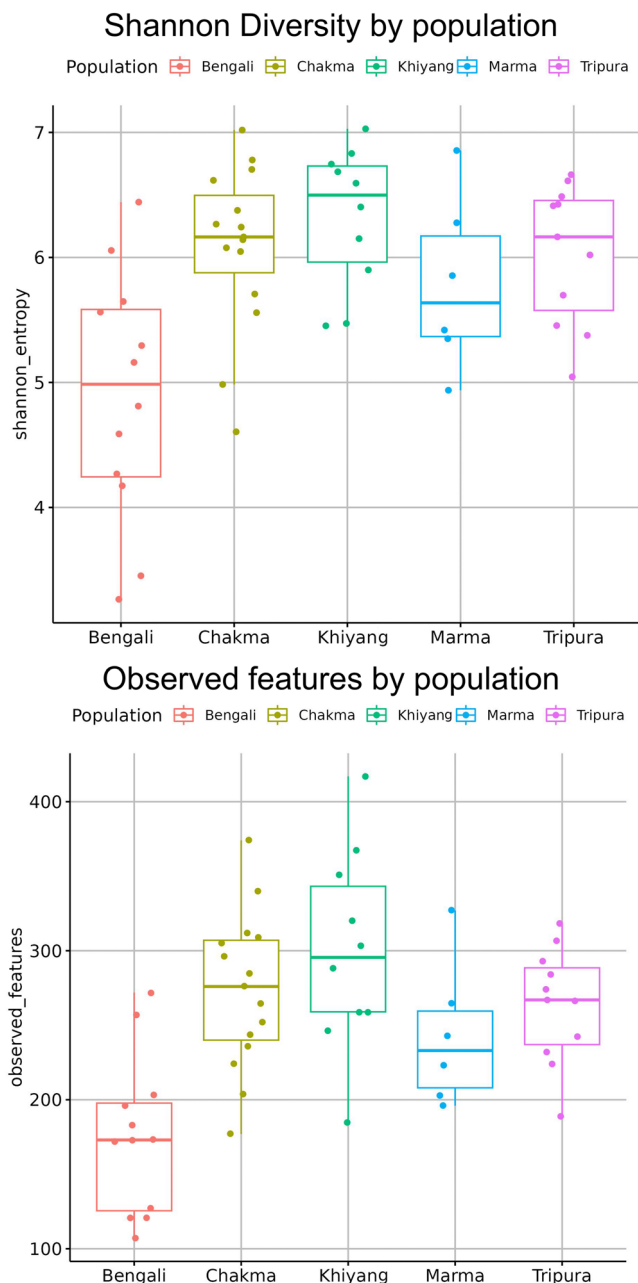
Fig. 2 | Distribution of alpha diversity in terms of Shannon diversity and observed features among various Bangladeshi ethnic cohorts. The Kruskal–Wallis test for all groups had a *p*-value of 0.0033. Chakma, Marma, Khyang and Tripura populations had higher species richness than the Bengalis (*p*-value = 0.0003) (*n* = 54 biologically independent samples).

Fig. 3 | Distribution of alpha diversity in terms of Faith phylogenetic diversity and Pielou evenness among various Bangladeshi ethnic cohorts. The difference in Faith pd score was statistically insignificant (*p*-value = 0.3816). There was no significant difference in Pielou evenness profile between indigenous populations (*p*-value = 0.0631) (*n* = 54 biologically independent samples).

Presence and absence of various species Made Bengali microbiome distinct. The PCoA plot revealed that the Bengali cohort was distinct from others in terms of Jaccard distance based on ethnicity (Fig. 4b) Jaccard distance is a binary distance matrix that only considers whether a taxon is present or not.

In other metrics, such as Bray–Curtis, weighted and unweighted UniFrac distances calculate presence and relative abundances of microbial taxa but no separated clusters were observed there (Figs. 4a, c, and d).

Alistipes genus is differentially abundant in Chakma population. ANCOM analysis revealed the microbial community compositions between the various Bangladeshi ethnic groups. It used a log-ratio test to compare the relative abundance using non-parametric methods. Here,
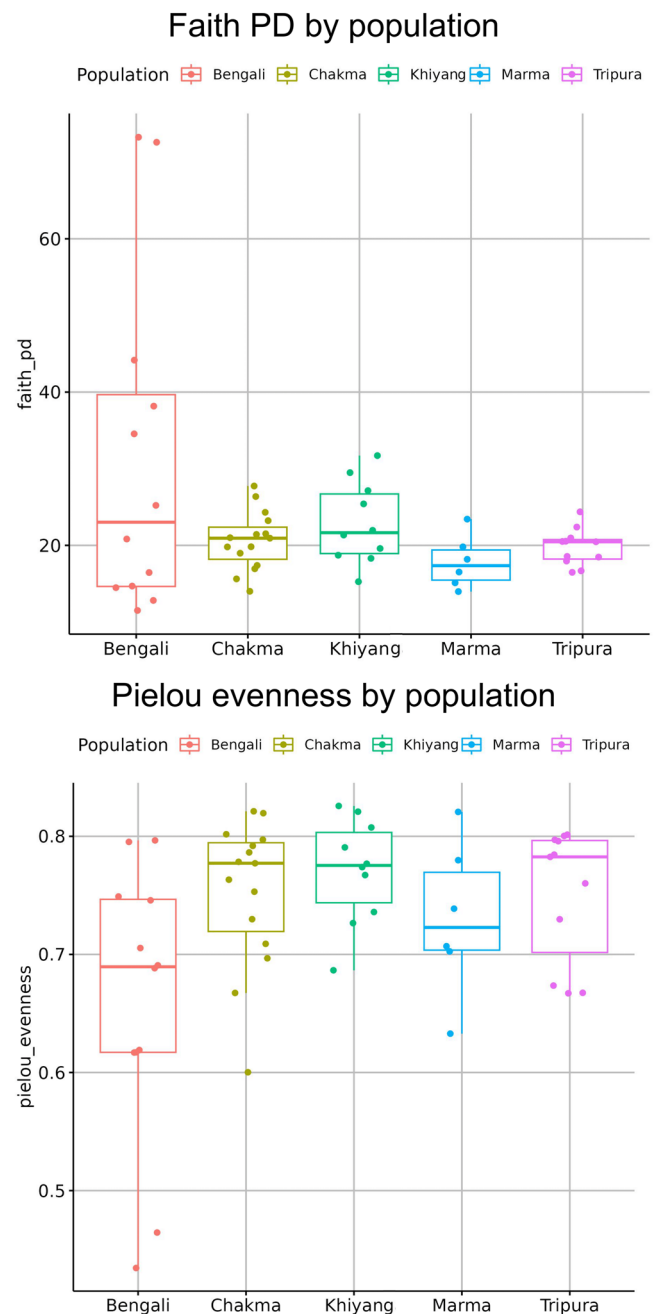
only the Bengali vs. Non-Bengali and the Chakma vs. Non-Chakma comparisons demonstrated differentially abundant taxa between the groups. In terms of ANCOM, the other groups (Marma, Khyang, and Tripura) did not reveal any differentially abundant bacteria. In the Bengali group, an unclassified bacteria showed a negative centered log ratio (clr) with higher W score (Fig. 5a). Whereas, the Chakma population depicted abundance for *Alistipes* and *Odoribacter* (Fig. 5b). Another tool, LEfSe implemented statistical methods to identify microbial features that were differentially abundant. LEfSe unveiled differentially abundant genus *Paraprevotella*, *Lactonifactor*, *Barnesiella*, *Bacteroides*, *Alistipes*, and *Ruminococcus* within the Chakma population (Fig. 5c). The results from ANCOM and LEfSe analysis complemented
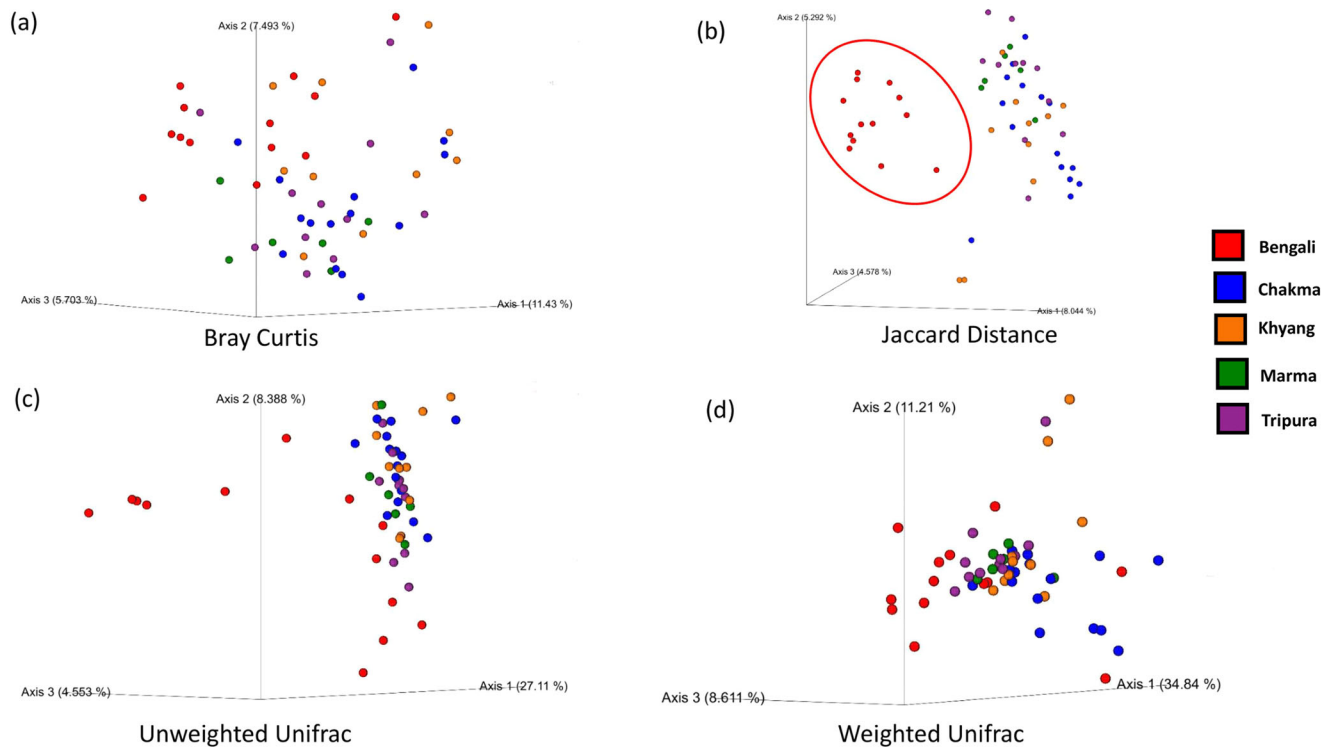
Fig. 4 | Beta diversity clustering between various Bangladeshi ethnic cohorts. The diversity was measured using (**a**) Bray–Curtis matrix (**b**) Jaccard distance matrix (**c**) Unweighted unifrac distance (**d**) Unweighted unifrac distance.

each other as they both unveiled that the genus *Alistipes* is differentially abundant in the Chakma population.

Bengali gut microbiome contributed to markedly diverse biological pathways. Relationships between microbiome composition and its effect on different biological functions were also explored in this study. Among all metadata categories, only the Bengali, Chakma, and Marma groups showed differential enrichment of function based on their gut microbiome composition (Fig. 6). All differentially abundant functions along with related contributing taxa and BH (Benjamini and Hochberg) FDR-adjusted *p*-values are tabulated in Supplementary Data 9. Thirty-three pathways were differentially abundant in Bengali samples. Several of them were highly enriched in Bengali samples such as peptidases, Ribosomes, Purine, and Pyrimidine metabolism (Fig. 6a). The genus *Bacteroides*, *Blautia*, *Collinsella*, *Coprococcus*, *Dorea*, *Parabacteroides*, SMB53, and *Slackia* were the top significantly contributing taxa in the enrichment of these functions. Chakma samples seemed to be enriched with histidine metabolism, arginine biosynthesis, and transcription machinery (Fig. 6b). But, in the case of the Marma samples histidine metabolism, arginine biosynthesis functions were in relatively lower abundance than other Non-Marma samples, and transcription machinery function was in higher frequency (Fig. 6c). For both categories, the genus *Bacteroides*, *Blautia*, and *Parabacteroides* were the most contributing taxa for function abundance. Detailed statistics for significant pathways and taxa contributions for each pathway are documented in Supplementary Data 10.

Bangladeshi gut microbiome has a distinct composition compared to tropical and subtropical populations. When compared with the gut microbiome of several other countries, the Bangladeshi samples formed a vividly distinct cluster based on bacterial composition and abundance (Fig. 7a, b). Cluster for Bray–Curtis and Jaccard distance indicates that bacterial composition and their abundances are unique for Bangladeshi populations. Their species richness was higher than other countries

(Fig. 7f). The unweighted UniFrac, which incorporates phylogenetic information, demonstrated that most of the samples from the Bangladeshi population clustered closer to the Indian samples (Fig. 7c). All numerical source data of Fig. 7 can be found in Supplementary Data 11a–h. Furthermore, the Unweighted Pair Group Method with Arithmetic Mean based tree showed that most of the indigenous samples of Bangladesh have clustered with different Indian indigenous groups while the Bangladeshi samples as a whole were scattered across different phylogenetic clades (Fig. 8).

Discussion

The human gut microbiome is an integral part of human growth and development and it is highly influenced by ethnicity[14,46]. Diverse ethnic populations reside in Bangladesh. The majority of the population is Bengali whereas large indigenous communities such as the Chakma, Marma, Tripura, and Khyang live in Chittagong Hill Tracts region of the country. The eating habits of these communities substantially differ from Bengali people. Their gut microbiomes, which may have an impact on important public health issues, can be influenced by their genetic make-up, geographic location, and lifestyle. In this study, we conducted a pilot survey to shed light of these areas.

Here, we have performed a 16 s rRNA gene amplicon sequencing to identify the gut microbiome composition. To execute this, we have collected fecal samples from Bengali, Chakma, Marma, Khyang, and Tripura populations. After collecting the microbial DNA from fecal samples, parts of the 16 S rRNA genes of the bacterial species were amplified. The amplified products were then sequenced. The sequence data was then processed and analyzed to draw interpretations. After all the analysis, we found that at the phylum level, Bengali population has lower *Firmicutes* to *Bacteroidetes* ratio compared to indigenous groups. Among *Firmicutes* phylum, *Dialister* and *Faecalibacterium* are highly abundant in Bangladeshi microbiomes (Fig. 1a). On the other hand, *Prevotella*, a member of the *Bacteroidetes* phylum, was distributed evenly across each cohort. People from the Indian subcontinent usually have *Prevotella* abundance
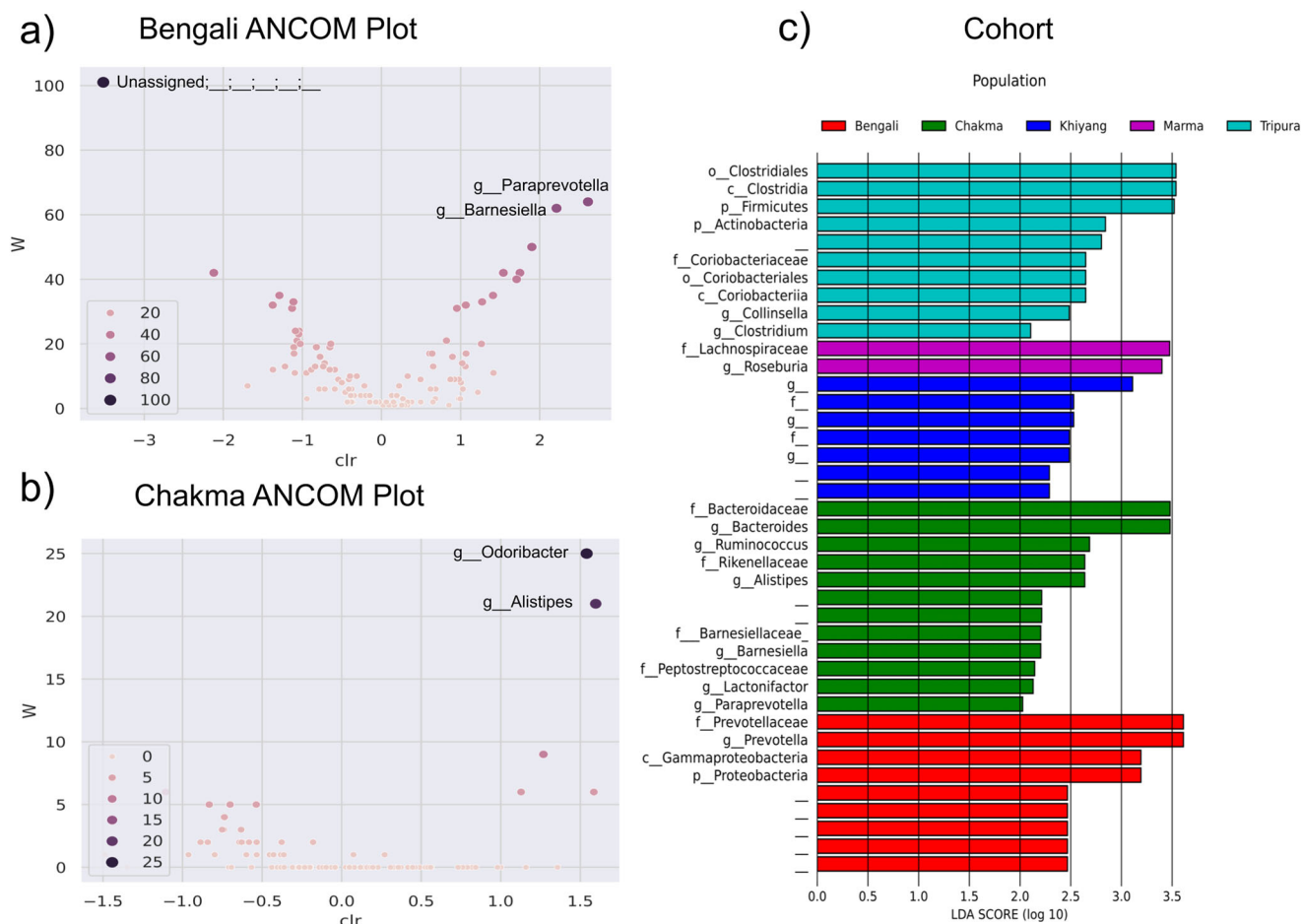
Fig. 5 | Differential abundance of gut microbes of various ethnic groups in Bangladesh. a ANCOM analysis of Bengali people (**b**) ANCOM analysis of Chakma people (**c**) LEfSe analysis of Bengali, Chakma, Marma, Khyang, and Tripura people (n = 54 biologically independent samples).

in their gut[47]. Here, we have explored the alpha and beta diversity of the gut microbiome. Alpha diversity depicted within-group diversity and beta diversity showed different diversity ratios between the groups. The alpha diversity of the Bengali population was lower compared to others according to the Shannon diversity and species richness (Observed features) (Fig. 2). The overall species evenness was enriched for indigenous populations (Fig. 3). Based on the phylogenetic diversity of the gut microbes, the Bengali and the people of the Chittagong Hill Tracts were found not to be very distant. However, our findings indicate that the overall gut microbial population and abundance is unique for Bangladesh when compared with international datasets. However, most of the indigenous samples from Bangladesh (Chakma, Marma, Khyang, Tripura) clustered separately with a common origin with the Indian indigenous population whereas Bengali samples were spread across various clades (Fig. 8). Some bacterial species were found to be abundant in indigenous groups that were not enriched in Bengali samples or vice versa. To identify this differential abundance, we employed several statistical approaches using LEfSe and ANCOM. Both of these tools identified an unclassified bacteria, which can be studied further, that is significantly abundant in Bengali samples whereas Chakma samples had enriched levels of *Alistipes* species. *Alistipes* is a newly discovered genus under *Bacteroidetes* phylum.

The gut microbial pathways play a critical role in health and diseases[48]. In Bengali samples DNA repair system, pyrimidine and purine metabolism, lipopolysaccharide (LPS) biosynthesis, peptidases, ribosome, etc functions were enriched. On the other hand, Chakma samples were enriched with histidine metabolism and arginine biosynthesis (Fig. 6b) whereas in the Marma group abundance of these pathways were lower (Fig. 6c).

Furthermore, Bangladeshi gut microbes possessed a distinctive diversity profile. Further research is needed to investigate the specific differences in responses and underlying mechanisms associated with distinct gut microbiota profiles.

Limitations of the study

The primary limitation of the study is small sample size, which exhibits significant variability among individuals sampled. With a sample size of n = 55 divided into five ethnic groups, two genders, and three large age groups (20–40, 40–60, and 60–80 years), the representativeness for each combination is low. The gut microbiome (GM) is well-known to be highly influenced by factors such as sex, age, current health status and so on. In our study, the representativeness of each of these groups is limited. A higher sample size would certainly enhance the credibility of the differences between these groups and also help draw more meaningful conclusions. However, the primary goal of our study was to explore and characterize the gut microbiome of Bangladeshi population for the first time and thus establish the baseline data for further gut microbiome research in Bangladesh. During the course of the study, as a secondary goal, we also wanted to see if there is any difference in the gut microbiome of the various ethnicities. Since our main priority was to characterize the Bangladeshi population as a whole, the difference between various subgroups based on age, sex, and health status was given less importance.

We want to reiterate that the number of participants in the current study was not sufficient to provide a comprehensive representation of the overall gut microbial diversity within neither the Bangladeshi population as a whole nor the ethnicities studied (Bengali, Chakma, Marma, Khyang, and
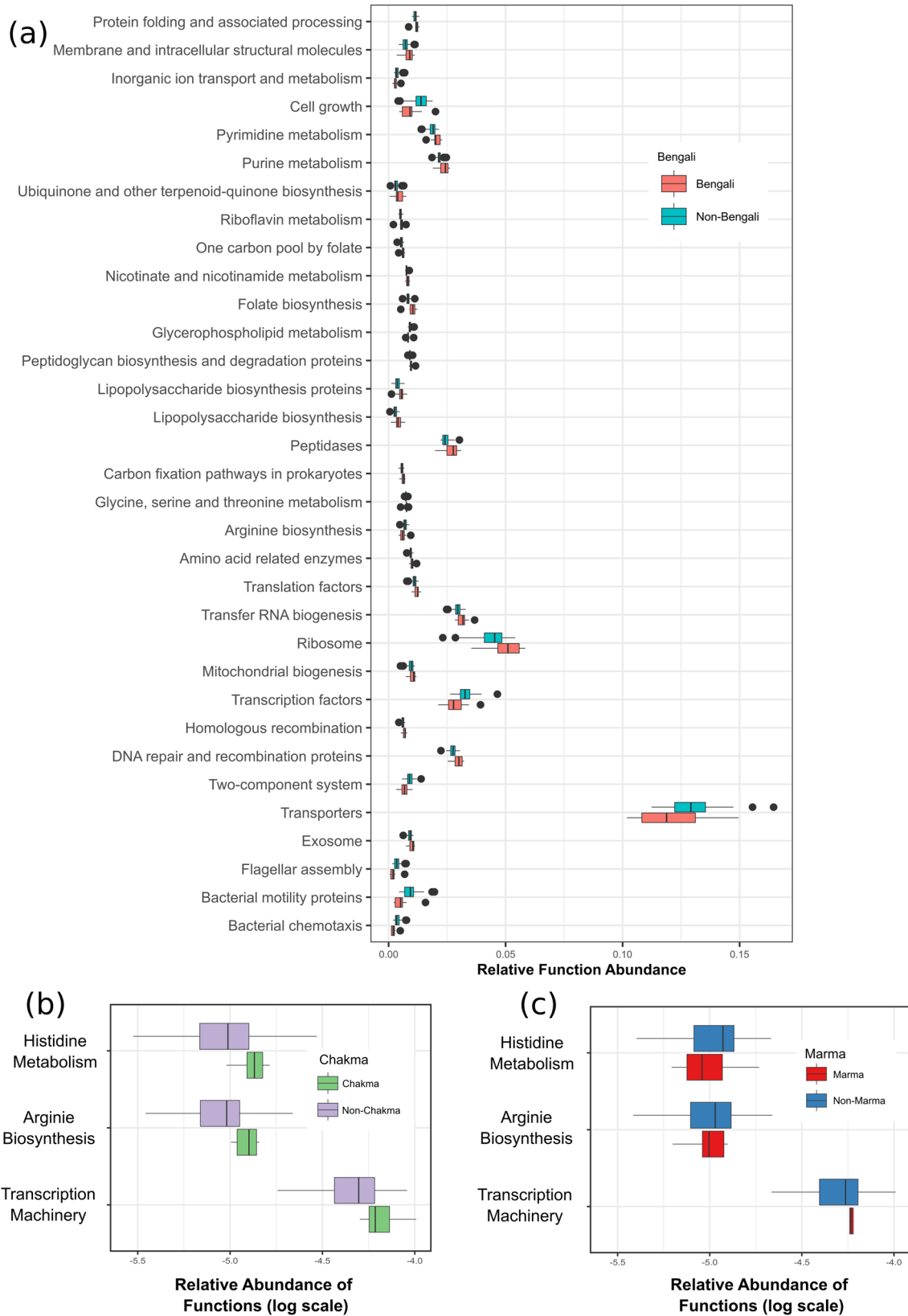
Fig. 6 | Functional analysis of enriched pathways in Bengali, Chakma, and Marma population. a Bengali (**b**) Chakma, and (**c**) Marma groups showed differential enrichment of function.
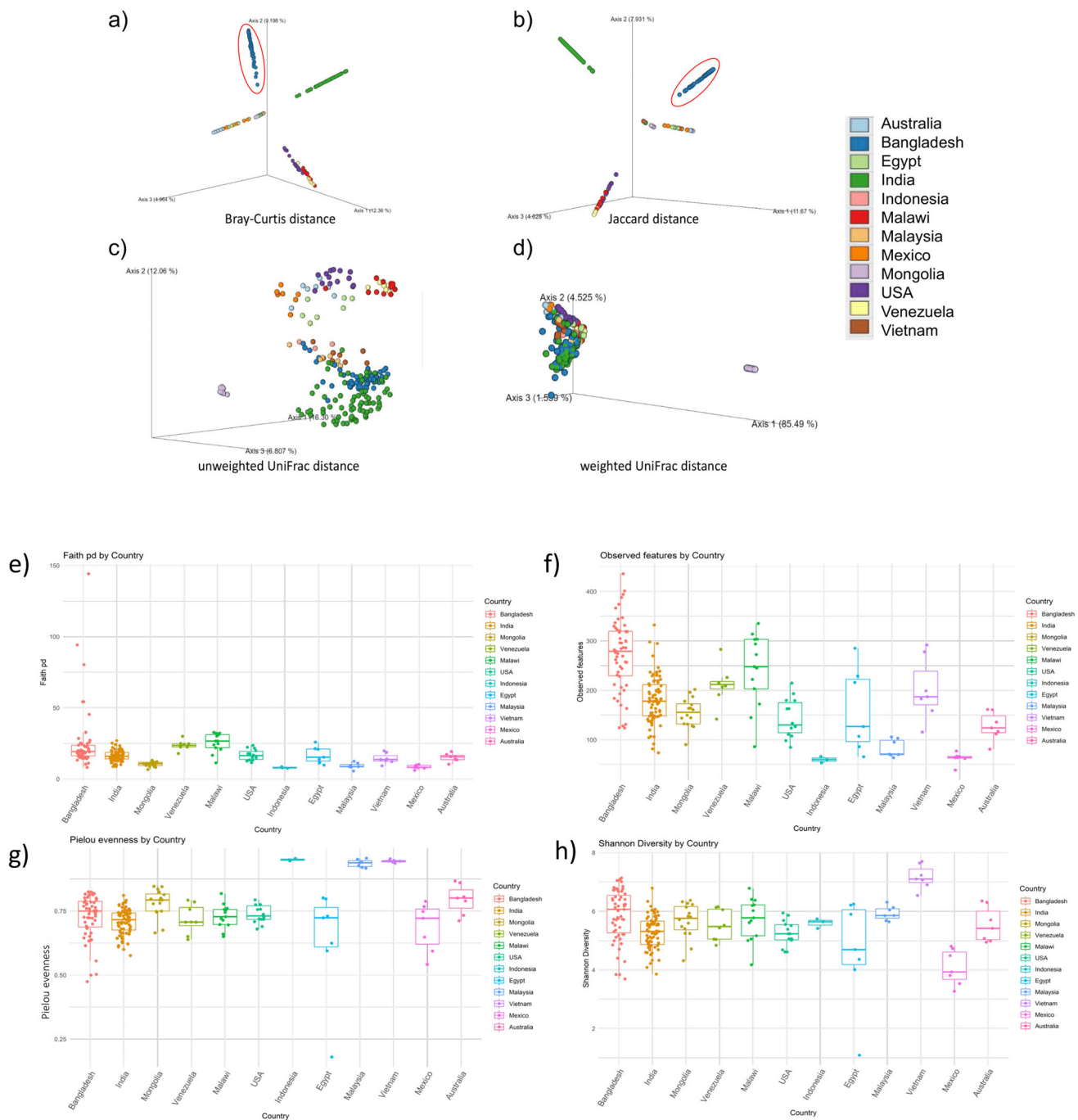
Fig. 7 | Comparative diversity analysis between Bangladesh and several other tropical and subtropical countries. a Bray–Curtis matrix (Red circle encircles Bangladeshi samples) (**b**) Jaccard distance matrix (Red circle encircles Bangladeshi samples) (**c**) Unweighted unifrac distance (**d**) Unweighted unifrac distance (**e**) Faith phylogenetic diversity (*p*-value = 2.3238) (**f**) Observed features (*p*-value = 1.6253) (**g**) Pielou evenness (*p*-value = 9.3489) (**h**) Shannon diversity (*p*-value = 6.0806) (*n* = 55 biologically independent samples).

Tripura). As a result, the findings from this study may lack statistical power and robustness, leading to potential biases and limited generalizability of the results. Moreover, the small sample size may hinder the identification of subtle microbial variations that could be crucial in understanding complex interactions between the gut microbiota and various health conditions. Consequently, caution must be exercised when interpreting the results of this gut microbiome study due to its limited sample size. Further research with larger cohorts is absolutely necessary to draw more definitive and reliable conclusions about the role of gut microbiota in the Bangladeshi population.

In conclusion, the current study indicates that the indigenous gut microbiome was more diverse and distinct from the Bengali population. Our study will help establish the baseline data for the gut microbiome of the Bangladeshi population.
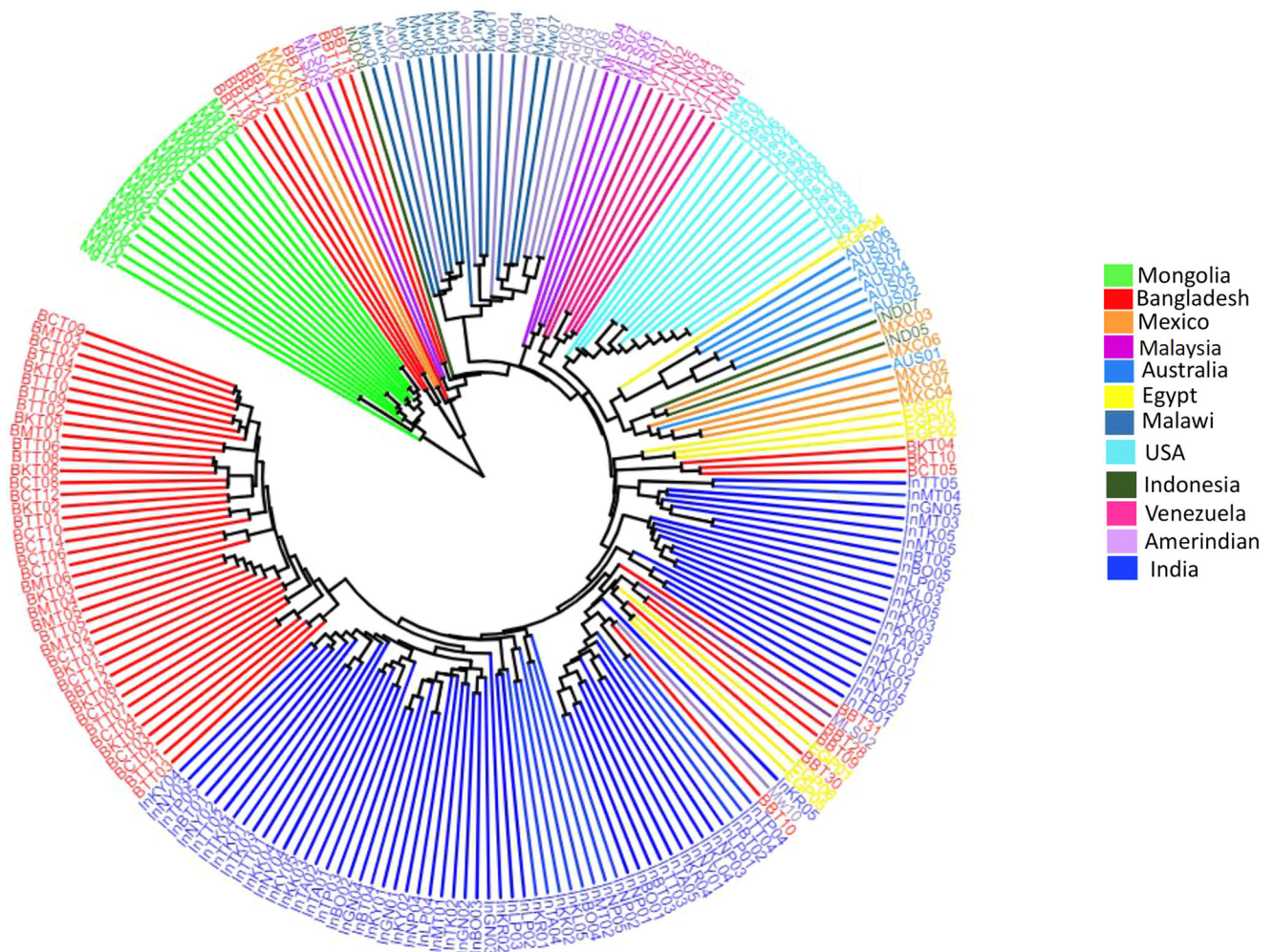
Data availability

Fig. 8 | Unweighted pair group method with arithmetic mean based gut microbiome clustering of Bangladeshi and several other tropical and subtropical countries. Majority of the Bangladeshi indigenous samples clustered with Indian indigenous groups.

Code availability
Code can be accessed from the corresponding author on reasonable request. Versions of software used to process the current dataset: QIIME2 (version 2021.4.0).

References

1. Bull, M. J. & Plummer, N. T. Part 1: The human gut microbiome in health and disease. *Integr. Med. Clin. J.* **13**, 17–22 (2014).
2. Chassaing, B., Kumar, M., Baker, M. T., Singh, V. & Vijay-Kumar, M. Mammalian gut immunity. *Biomed. J.* **37**, 246–258 (2014).
3. Cani, P. D. Human gut microbiome: hopes, threats and promises. *Gut* **67**, 1716–1725 (2018).
4. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
5. Gupta, V. K., Paul, S. & Dutta, C. Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front. Microbiol.* **8**, 1162 (2017).
6. Adeyanju, O. et al. Common NOD2 risk variants in African Americans with Crohn's disease are due exclusively to recent caucasian admixture. *Inflamm. Bowel Dis.* **18**, 2357–2359 (2012).
7. Lauro, M. L., Burch, J. M. & Grimes, C. L. The effect of NOD2 on the microbiota in Crohn's disease. *Curr. Opin. Biotechnol.* **40**, 97–102 (2016).
8. Charati, H. et al. The evolutionary genetics of lactase persistence in seven ethnic groups across the Iranian plateau. *Hum. Genomics* **13**, 7 (2019).
9. Hall, A. B., Tolonen, A. C. & Xavier, R. J. Human genetic variation and the gut microbiome in disease. *Nat. Rev. Genet.* **18**, 690–699 (2017).
10. Kato, K. et al. Association between functional lactase variants and a high abundance of bifidobacterium in the gut of healthy Japanese people. *PloS One* **13**, e0206189 (2018).
11. Huang, T., Shu, Y. & Cai, Y.-D. Genetic differences among ethnic groups. *BMC Genomics* **16**, 1093 (2015).
12. Bennett, G., Bardon, L. A. & Gibney, E. R. A comparison of dietary patterns and factors influencing food choice among ethnic groups living in one locality: a systematic review. *Nutrients* **14**, 941 (2022).
13. Manica, A., Prugnolle, F. & Balloux, F. Geography is a better determinant of human genetic differentiation than ethnicity. *Hum. Genet.* **118**, 366–371 (2005).
14. Dwiyanto, J. et al. Ethnicity influences the gut microbiota of individuals sharing a geographical location: a cross-sectional study from a middle-income country. *Sci. Rep.* **11**, 2618 (2021).
15. Dehingia, M. et al. Gut bacterial diversity of the tribes of India and comparison with the worldwide data. *Sci. Rep.* **5**, 18563 (2015).
16. Hazarika, P., Chattopadhyay, I., Umpo, M., Choudhury, Y. & Sharma, I. Elucidating the gut microbiome alterations of tribal community of Arunachal Pradesh: perspectives on their lifestyle or food habits. *Sci. Rep.* **12**, 18296 (2022).
17. Alam, S. M. N. & Naser, M. N. Chapter 2—Role of traditional foods of Bangladesh in reaching-out of nutrition. In *Nutritional and Health*

Aspects of Food in South Asian Countries (eds. Prakash, J., Waisundara, V. & Prakash, V.) 217–235 (Academic Press, 2020) https://doi.org/10.1016/B978-0-12-820011-7.00025-3.

18. Haque, M. M. Dietary practice among mainstream Bengali population and ethnic communities in Bangladesh. *Arch. Nutr. Food Sci*. **1**, 5–7 (2020).

19. Chakraborty, R. et al. Gene differentiation among ten endogamous groups of West Bengal, India. *Am. J. Phys. Anthropol.* **71**, 295–309 (1986).

20. Saha, N. Blood genetic markers in Bengali muslims of Bangladesh. *Hum. Hered.* **37**, 86–93 (1987).

21. Hasan, M. M. et al. Phylogenetic and forensic studies of the Bangladeshi population using next-generation powerPlex® Y23 STR marker system. *Int. J. Leg. Med.* **130**, 1493–1495 (2016).

22. Wen, B. et al. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in Southern Tibeto-Burmans. *Am. J. Hum. Genet.* **74**, 856–865 (2004).

23. Gayden, T. et al. Genetic insights into the origins of Tibeto-Burman populations in the himalayas. *J. Hum. Genet.* **54**, 216–223 (2009).

24. Gazi, N. N. et al. Genetic structure of Tibeto-Burman populations of Bangladesh: evaluating the gene flow along the sides of Bay-of-Bengal. *PLoS One* **8**, e75064 (2013).

25. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

26. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

27. Janssen, S. et al. Phylogenetic placement of exact amplicon sequences improves associations with clinical information. *mSystems* **3**, e00021–18 (2018).

28. Matsen, F. A., Hoffman, N. G., Gallagher, A. & Stamatakis, A. A format for phylogenetic placements. *PLoS One* **7**, e31009 (2012).

29. Weiss, S. et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).

30. Halko, N., Martinsson, P.-G., Shkolnisky, Y. & Tygert, M. An algorithm for the principal component analysis of large data sets. *SIAM J. Sci. Comput.* **33**, 2580–2594 (2011).

31. Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience* **2**, 16 (2013).

32. Kruskal, W. H. & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).

33. McKinney, W. Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf.* 56–61 (2010).

34. Bokulich, N. A. et al. q2-longitudinal: longitudinal and paired-sample analyses of microbiome data. *mSystems* **3**, e00219–18 (2018).

35. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).

36. Mandal, S. et al. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* **26**, 27663 (2015).

37. Lin, H. & Peddada, S. D. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms. Microbiomes.* **6**, 1–13 (2020).

38. Segata, N. et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).

39. McNally, C. P., Eng, A., Noecker, C., Gagne-Maynard, W. C. & Borenstein, E. BURRITO: An interactive multi-omic tool for visualizing taxa–function relationships in microbiome data. *Front. Microbiol.* **9**, 365 (2018).

40. Langille, M. G. I. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–821 (2013).

41. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).

42. DeSantis, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).

43. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

44. Amir, A. et al. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* **2**, e00191–16 (2017).

45. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).

46. Guinane, C. M. & Cotter, P. D. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Ther. Adv. Gastroenterol.* **6**, 295–308 (2013).

47. Prasoodanan, P. K. V. et al. Western and non-western gut microbiomes reveal new roles of prevotella in carbohydrate metabolism and mouth–gut axis. *NPJ Biofilms. Microbiomes.* **7**, 1–17 (2021).

48. Hou, K. et al. Microbiota in health and diseases. *Signal Transduct. Target. Ther.* **7**, 1–28 (2022).

Acknowledgements

Author contributions

Conceptualization: Mohammad Uzzal Hossain, Keshob Chandra Das, Chaman Ara Keya, Md Salimullah, Analysis: Ishtiaque Ahammad, Arittra Bhattacharjee, Zeshan Mahmud Chowdhury, Anisur Rahman, Resources: Gourab Dewan, Shiny Talukder, Keshob Chandra Das, Md Salimullah. Writing—Original Draft: Ishtiaque Ahammad, Arittra Bhattacharjee, Zeshan Mahmud Chowdhury, Anisur Rahman, Writing—Review & Editing: Ishtiaque Ahammad, Arittra Bhattacharjee, Zeshan Mahmud Chowdhury, Anisur Rahman, Mohammad Uzzal Hossain, Chaman Ara Keya, Md Salimullah. Supervision: Keshob Chandra Das, Chaman Ara Keya, Md Salimullah. Funding acquisition: Md Salimullah.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06191-9.

Correspondence and requests for materials should be addressed to Md Salimullah.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Kevin Theis and Tobias Goris.

Reprints and permissions information is available at http://www.nature.com/reprints

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