Cell. Mol. Biol. (ISSN: 1165-158X)

Journal Homepage: www.cellmolbiol.org

Cellular and Molecular Biology

Original Article

Investigation of bacterial gut microbiome in diverse Egyptian populations “pilot study”

Kareem Talaat Mohamed[1], Sarah Shabayek[2], Nora Fahmy Mahmoud[2,4], Mahmoud Mohamed Tawfick[3], Amro Mohamed Said Hanora[2,5*]

[1] *Department of Microbiology and Immunology, Faculty of Pharmacy, October University for Modern Sciences & Arts (MSA), Giza, Egypt*
[2] *Department of Microbiology and Immunology, Faculty of Pharmacy, Suez Canal University, Ismailia, Egypt*
[3] *Department of Microbiology and Immunology, Faculty of Pharmacy (For Boys), Al-Azhar University, Cairo, Egypt*
[4] *Department of Microbiology and Immunology, Faculty of Pharmacy, Sinai University-Kantara Branch, Ismailia, Egypt*
[5] *Department of Microbiology and Immunology, Faculty of Pharmacy, King Salman International University, Ras Suder, Egypt*

Article Info

Article history:

Received: October 23, 2024
Accepted: December 12, 2024
Published: January 31, 2025

Use your device to scan and read the article online

Abstract

The gut microbiota plays a huge role in human health regarding immunity, metabolism, and nutrient absorption. In this work, the gut microbiota, with its bacterial community structure, is studied using whole genome shotgun (WGS) sequencing for populations from two different geographical regions in Egypt: Cairo (urban) and Ismailia (rural). Fecal samples were obtained from six healthy individuals, three from Cairo and three from Ismailia, of ages ranging from 43 to 52 years. Alpha diversity, measured as Shannon, inverse Simpson, and OTUs, showed no significant differences between the two cities. However, beta diversity analysis by Principal Coordinate Analysis (PCoA) revealed diverse microbial compositions. Thus, only the Ismailia samples contained higher levels of butyrate-producing bacteria involved in maintaining intestinal health, such as *Faecalibacterium prausnitzii* and *Akkermansia muciniphila*. On the other hand, there was a higher prevalence in Cairo of bacteria associated with protein and fat metabolism, like *Bacteroides thetaiotaomicron*. Such findings explain the influence of environmental factors in shaping gut microbiota and show that to get a comprehensive understanding of regional differences, many wider-ranging studies need to be conducted.

Keywords: Gut microbiome, Whole genome sequencing, Dietary influence, Geographic Variation, Egypt.

1. Introduction

The balance of microbes in the gut is directly linked to human health and disease. Compared with other body parts, the gastrointestinal (GI) tract has a very large group of microorganisms, approximately 100 trillion [1]. Many studies have shown the important link between gut bacteria and basic biological processes in humans. For example, research has shown that the microbiome in humans plays a large role in metabolism, immunity, and nutrient acquisition from food [2]. By considering the traits of the gut microbiota, including its wide variety, stability, resilience, and mutually beneficial relationship with the host, we can classify the host and the bacteria that reside within it as “superorganisms" [3]. The gut microbiota can influence biological processes in numerous ways. It plays a vital role in extracting energy and nutrients from food because of harbouring diverse metabolic genes, which encode distinct enzymes and biochemical pathways [4]. Additionally, the production of biologically active compounds such as vitamins, amino acids, and lipids is highly dependent on the gut microbiota [5]. Extensive studies are being conducted to determine the precise characteristics of a "healthy" gut microbiota and its connection to the physiological processes of the host. Nevertheless, the relative distribution of microorganisms is dissimilar among individuals and may vary within the same individual. Notably, the gastrointestinal microbiota of humans may vary due to environmental factors and age [6].

The gut microbiota consists of bacteria, yeasts, and viruses. A healthy microbiota community typically exhibits a substantial range of taxonomic diversity, many microbial genes, and a consistent core microbiota. A healthy microbiota would be a diverse and stable microbial community that contributes to host health and well-being. A typical composition of such is a balance of beneficial bacteria from the genus *Bifidobacterium* and the genus *Faecalibacterium*. These are associated with the maintenance of gut integrity, anti-inflammatory effects, and metabolic func-

* Corresponding author.
E-mail address: a.hanora@pharm.suez.edu.eg (M. S. Hanora).
Doi: http://dx.doi.org/10.14715/cmb/2025.70.1.8

tions. On the other hand, an imbalance in the microbiota, or dysbiosis, is often associated with inflammatory bowel diseases, metabolic disorders, and infections [7]. Bacteria are categorized taxonomically based on their phylum, classes, orders, families, genera, and species. Among all the phyla, only a small number account for over 160 species [8]. Firmicutes (Bacillota), Bacteroidetes, Actinobacteria, Proteobacteria, Fusobacteria, and Verrucomicrobia are the predominant microbial phyla found in the gut. Firmicutes and Bacteroidetes together account for 90% of the gut microbiota [9].

Several factors can change the composition of the gut microbiota, including host genetics, diet and age [10,11]. Other factors such as geographical location, and maternal lifestyle (urban or rural) also shape the gut microbiota composition [12,13]. Diet is one of the most powerful modulators of gut microbiota functions and composition. The table below summarizes the possible effects of different dietary patterns on gut microbiota composition classified by phyla (Table 1)

On the other hand, recent studies have pointed out the important contribution of geographic location to gut microbiota diversity. For example, Sun et al. [21] performed a population-based study in China and found significant regional differences in gut microbiota composition, which indicated that geographic factors should be considered when linking microbiota profiles to human phenotypic variations. Similarly, Porras et al. [22] studied the effect of geographic differences on susceptibility to enteric infections and found that gut microbiota profiles were different in different regions, which shaped immune responses and infection susceptibility. Furthermore, Ying et al. [23] showed how geographic habitats and host species shape gut microbiota diversity in non-human primates, further emphasizing environmental factors as a driver of microbial communities. These findings highlight the importance of geographical and environmental factors in microbiome studies, thus providing a critical context in which to understand the variability observed in gut microbiota composition across different populations and regions.

Another critical factor that has to be taken into consideration, in particular in countries like Egypt, is excessive use of antibiotics in shaping the gut microbiome. In Egypt, antibiotic misuse and overuse have led to a high prevalence of multi-drug resistant (MDR) bacteria among the population [24]. A growing number of studies have demonstrated that antibiotics can have both short-term and long-term effects on intestinal microbial populations [25,26]. Moreover, antibiotic usage has been linked to a decrease in gut microbial diversity [27]. Following antibiotic therapy, the baseline composition of the gastrointestinal microbiota was largely restored within 90 days; however, certain common species remained undetectable [28].

Currently, there are two major methods for analyzing microbial communities with high-throughput sequencing: whole-genome shotgun metagenomics (WGS) and marker gene studies. WGS metagenomics attempts to sequence every genome present in an environmental sample to infer the functional capacities and biodiversity of a microbial community being studied. In this respect, it is possible to characterize the full diversity of habitat archaea, bacteria, eukaryotes, viruses, and plasmids along with their gene content, given that the whole genetic material of a sample is retrieved [29]. WGS metagenomics is primarily advantageous over marker gene sequencing in that it enables the characterization of the genetic and genomic diversity of the analyzed community, as well as the identification of potential and novel functions. Moreover, this approach allows in enabling the analysis of microbial metabolic pathways, resistance genes, and other functional attributes that play critical roles in health and disease. Therefore, we aimed in our study to explore the functional potential of gut microbiota across diverse populations not only the microbial composition. Additionally, the assembly of complete genomes from metagenome data is feasible when an appropriate sequencing depth is employed. This approach allows for acquiring prototype genomes of uncultured organisms and insights into the "genomic diversity" of mi-

Table 1. Effects of different dietary patterns on the composition of gut microbiota classified by phylandance.

Type of diet	Characteristics	Microbial composition	Microbial diversity
Western	Regular intake of saturated fats, animal proteins, refined sugars, and processed foods.	↓ Bacteroidetes (*Prevotella*) [14,15] ↑ Firmicutes ↑ Pseudomonadota (*Proteobacteria, Enterobacteriaceae*) [16]	↓ Decrease[17]
Vegetarian	Plant-based food regular intake of fiber and starch (cereals, legumes, nuts). avoid consumption of any form of seafood or meat	↑ Bacteroidetes (*Prevotella*, *Bacteroides thetaiotaomicron)* ↑ Firmicutes *(Clostridium clostridioforme, Faecalibacterium prausnitzii)* ↑ Pseudomonadota *(Klebsiella pneumoniae)*, [18]	↑ Increase [19]
Omnivorous	Regular intake of a variety of foods, including fish or meat	↓ Bacteroidetes (*Prevotella*) ↓ Firmicutes *(Clostridium clostridioforme, Faecalibacterium prausnitzii)* ↓ Actinobacteria (*Bifidobacteria*) [20]	↓ Decrease [20]

Abbreviations: ↑, increase in bacterial relative abundance; ↓, decrease in bacterial relative abu

crobial ecosystems [30].

The current study aimed to study the structure of the bacterial gut microbiota in populations from two different geographical regions in Egypt through WGS sequencing at a depth of 10 GB data per sample, which may provide a comprehensive understanding of the factors influencing gut health in the region.

2. Materials and methods

2.1. Sample collection and ethics statement

The study was approved and conducted in accordance with the guidelines defined by the ethics committee of the faculty of pharmacy at Suez Canal University (Project number 201809PHDH1). Informed consent was received to disclose the information gathered from the healthy participants and to publish the study while maintaining the confidentiality of their identity.

The strict inclusion and exclusion criteria were done at the participant selection to reduce any potential confounding variables and ensure consistency within the study. The study included participants who self-reported being healthy and had a local omnivorous dietary intake. The exclusion criteria included the use of antibiotics within the six months prior to the collection of samples, as well as any gastrointestinal chronic diseases. Fecal samples were obtained from six individuals (three samples from each region) who were in good health and aged between 43 and 52 years. The samples were collected from two distinct districts in Egypt, Ismailia (30.5965° N, 32.2715° E) and Cairo (30.0444° N, 31.2357° E). Specimens were obtained from Ismailia on September 25, 2021, and from Cairo on October 3, 2021. Study participants were directed to provide fecal samples in a sterile container within 20 minutes after defecation. The fecal samples were immediately preserved on dry ice after being collected from the individuals, then transported to the microbiology and molecular biology laboratory at MSA University and placed in a freezer set at -80 °C until the DNA extraction process. The metadata and informed consent of participants are presented in Table S1 (Supplementary information)

2.2. Microbial metagenomic DNA extraction

Microbial metagenomic DNA was extracted from fecal samples using HiPurA® Stool DNA Purification Kit (HiMedia®, India). A sterilized spatula was used to scrape 250 mg of each frozen fecal sample to prevent freeze-thaw cycles and microbial contamination. After the samples were weighed, lysis buffer was added, and the samples were centrifuged using a Sigma 3-16KL refrigerated benchtop centrifuge at a speed of 8000 xg for 3 minutes at room temperature. The remaining steps of the protocol were carried out as described according to the manufacturer's manual. The quality of the DNA was evaluated through visual inspection of the DNA on a 1% agarose gel electrophoresis [31], whereas the quantification of the DNA was carried out using a Jenway 7415 Nano Scanning Micro-Volume Spectrophotometer (Jenway, United Kingdom).

2.3. Metagenomic sequencing

Metagenomic sequencing of all the samples was performed on the BGISEQ-500 DNBseq platform. High-density DNA Nanochip technology was used to load DNBs into the patterned nanoarray. Finally, the combinatorial Probe-Anchor Synthesis (cPAS) platform was used to obtain paired-end 100-bp reads. All samples library preparation was performed following the BGISEQ-500 whole genome sequencing library preparation protocol Beijing Genomics Institute (BGI), (Shenzhen, China).

2.4. Pre-processing of reads

The BGISEQ-500 sequencer generated paired-end reads for each sample, which went through the following preparation processes: The adapter filtering criteria excluded reads that had a sequence match of 25.0% or higher with the adapter sequence, with a tolerance of up to 3 base mismatches. All readings that met these criteria were completely discarded. In addition, readings with lengths shorter than 150 bp were eliminated using a read length filter. Readings with a nitrogen (N) level of 0.1% or higher were also excluded using the nitrogen removal filter. Reads were quality-checked via FastQC v0.11.5 [32]. The reads were filtered from adaptors, trimmed, and cleaned using Trimmomatic v0.36 [33]. Finally, the quality values of the output reads were modified to Phred+64 to acquire purified readings for subsequent analysis. The completion of this step was carried out using SOAPnuke software, which was developed by BGI [34].

2.5. Diversities

In the present study, alpha diversity, which describes the richness and evenness of microbial species within a single sample, and beta diversity, which compares the differences in microbial composition among different samples or groups, were generated and assessed by QIIME 2 [35] , a very important bioinformatics platform designed for this purpose to analyse microbiomes. For within-community diversity, which refers to species richness in every sample, we estimated some metrics, such as the Shannon diversity index and observed operational taxonomic units (OTUs). These indices provide information on the microbial diversity and evenness of a particular sample. For beta diversity, which quantifies differences in microbial community composition between samples, distance metrics such as the Bray-Curtis and Jaccard indices were used.

2.6. Taxonomic profiling of reads

The taxonomic classification of the shotgun metagenomics data involved the use of two different classifier tools: alignment-based classification, used by Kaiju (https://narrative.kbase.us/#catalog/apps/kb_kaiju/run_kaiju) [36] which is available through KBase [37]. Kaiju yields information about the archaeal and bacterial composition of each sample. The standard Krona [38] plots are supplemented by stacked bar plots with fractional classification in the KBase implementation of Kaiju. Moreover, Kaiju uses the Burrows-Wheeler transform (BWT) to align sequences directly with the NCBI non-redundant protein sequence database (nr). The second classifier tool is the Kraken/kraken2 [39, 40] "k-mer based" taxonomic classification algorithm, which uses a reference database that includes approximately 40,000 bacterial, viral, fungal, and protozoan genomes.

2.7. Metagenomic assembly

Metagenomic assembly was performed via KBase [37]. Each sample was quality-checked again using FastQC [32] and paired reads were merged into one object using the KBase app Merge Reads Libraries v1.0.1 (https://nar-

rative.kbase.us/#catalog/apps/kb_ReadsUtilities/KButil_Merge_MultipleReadsLibs_to_OneLibrary) [37]. Bins from each sample were recovered through coassembly using MetaSPAdes v3.15.3 (https://narrative.kbase.us/#catalog/apps/kb_SPAdes/run_metaSPAdes) [41], MEGAHIT v1.2.9 (https://narrative.kbase.us/#catalog/apps/MEGAHIT/run_megahit) [42], and IDBA-UD v1.1.3 (https://narrative.kbase.us/#catalog/apps/kb_IDBA/run_idba_ud) [43]. "The Compare Assembled Contig Distributions tool v1.1.2 (https://narrative.kbase.us/#catalog/apps/kb_assembly_compare/run_contig_distribution_compare) [37] was used to compare the quality assembly on the basis of the contig lengths and size distributions after coassembly. Longer contigs provide more genomic information making them more reliable for further analysis.

2.8. Metagenomic binning

Contigs were binned and coassembled with MaxBin2 v2.2.4 (https://narrative.kbase.us/#catalog/apps/kb_maxbin/run_maxbin2) [44], MetaBAT2 v1.7 (https://narrative.kbase.us/#catalog/apps/metabat/run_metabat) [45], and CONCOCT v1.1 (https://narrative.kbase.us/#catalog/apps/kb_concoct/run_kb_concoct) [46]. MaxBin2 uses 107 bacterial marker genes with a minimum of 1000-base-pair contigs and a probability threshold of 0.8. Implementation of the Markov Clustering Algorithm by MetaBAT2 clusters contigs into bins based on their sequence and abundance similarities, thereby appropriately binning them into different genomes. CONCOCT uses Bowtie2 for read mapping and a 2500-base-pair contig length threshold. After initial binning, the resulting bins were optimized using DAS Tool v1.1.2. (https://narrative.kbase.us/#catalog/apps/kb_das_tool/run_kb_das_tool) [37] app. DAS Tool bins were quality-checked and refined (parameters: reference tree, full tree; completeness, ≥90%; contamination, ≤5%) using CheckM v1.0.18 app (https://kbase.us/applist/apps/kb_Msuite/run_checkM_lineage_wf) [47]. Consensus-based bin consolidation across samples improves binning by reducing false positives and optimizing bin completeness. These bins were then extracted as assemblies using the KBase app Extract Bins as Assemblies from BinnedContigs v1.0.2 (https://narrative.kbase.us/#catalog/apps/MetagenomeUtils/extract_bins_as_assemblies)[37].

2.9. Genome annotation and functional profiling

Several bioinformatics tools for the annotation of microbial genomes and their functional profiles are available. First, we used RASTk v1.0.7.3 (https://narrative.kbase.us/#catalog/apps/RAST_SDK/annotate_genome_assembly) [48] to precisely annotate the genome by identifying genes and predicting activities against known subsystems. This was followed by one of the very robust functional profiling approaches Distilled and Refined Annotation of Metabolism (DRAM) (https://narrative.kbase.us/#catalog/apps/kb_DRAM/run_kb_dram_annotate) [49] to investigate the metabolic potential and functional diversity of these genomes. Several databases are used by DRAM to annotate metagenomic assembled genomes (MAGs) and subsequently condense the findings to aid in the examination of their functional and structural characteristics. Moreover, it employs functional marker genes to deduce metabolic descriptors of MAGs. This study utilized integrated data management and computing facilities provided by KBase [37] to accelerate data handling and analysis.

2.10. Data availability

The raw reads from the metagenome sequencing data have been submitted to the NCBI SRA database under accession number PRJNA982264, accession numbers SRR24888589 to SRR24888594. [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA982264]

2.11. Statistical analysis

All statistical analysis was done using the appropriate tools for metagenomic data processing and analysis of diversity indices. In the case of alpha diversity, Shannon, inverse Simpson, and observed OTUs were calculated to assess the richness and evenness of species in the samples. For beta diversity, Bray-Curtis and Jaccard indices were used to compare the microbial composition between samples. Beta diversity patterns were visualized by means of Principal Coordinate Analysis. Statistical significance considerations were made at a *p-value* threshold of <0.05. However, the comparisons of alpha and beta diversities did not show statistical significance among all samples, because all *p-values* were above this threshold. Data pro-

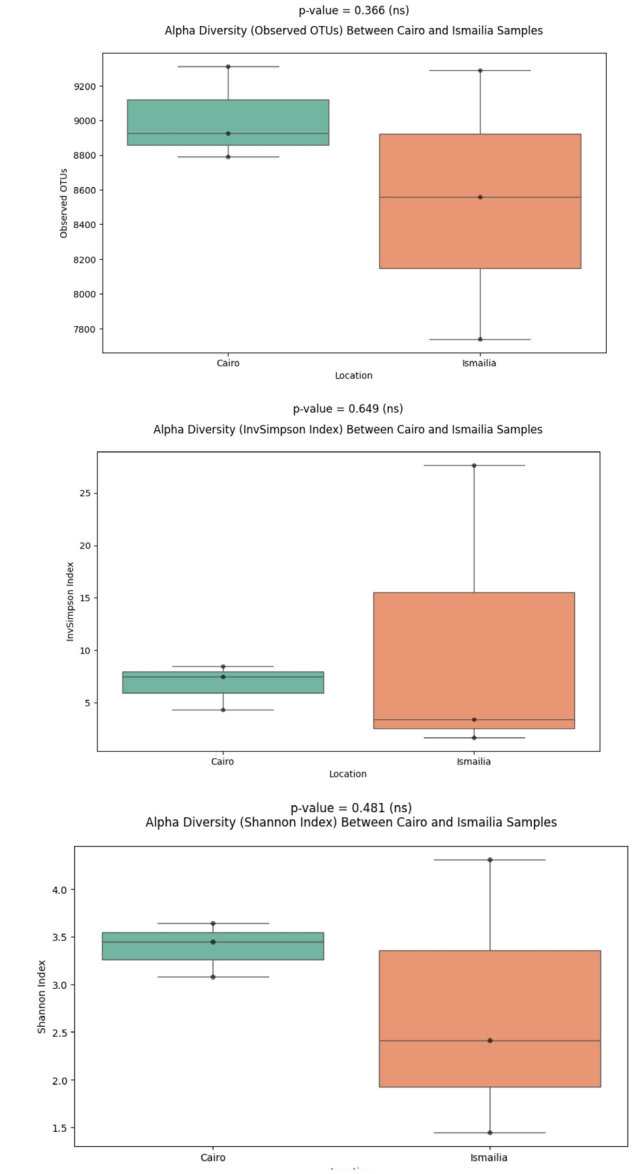

Fig. 1. Three box plots stacked vertically. Each plot compares the alpha diversity indices: (a) Observed OTUs, (b) InvSimpson, and (c) Shannon between Cairo and Ismailia samples, with their respective *p-values* (P ≤ 0.05).

cessing was done in QIIME 2 and visualized with associated bioinformatics to ensure robustness in analysis.

3. Results

3.1. Diversities

The results of the alpha matrices for the microbial diversity between the Cairo and Ismailia samples are presented using three diversity indices: Observed OTUs, the inverse Simpson Index, and the Shannon Index (Figure 1, Table 2). The observed OTUs had a *p-value* of 0.366 for the species richness metric. The inverse Simpson index, which considers species richness and evenness (how the sample is evenly distributed), had a *p-value* of 0.649. Finally, the Shannon index, which also considers richness and evenness with weight on the richness, yielded a *p-value* of 0.481. Sample D4 presented the highest diversity, with the Shannon index equal to 4.31 and the InvSimpson index at 27.66. It may indicate a very diverse and even microbial community. Contrariwise, the lowest diversity has sample D1: Shannon index at 1.45, InvSimpson index-1.69, which corresponds to the low diversification of the microbial community. For beta diversity, axis 1 in the Principal Coordinate Analysis (PCoA) diagram (Figure 2a) of gut microbiome samples accounts for 31.53% of the variation in the data, making it the most significant dimension for differentiating across the samples based on their microbial compositions. Axis 2 accounts for 24.52% of the variation, whereas Axis 3 explains 17.51%. The three axes account for approximately 73.56% of the total variation, providing a substantial summary of the differences in gut microbiome composition between samples. Although the clustering observed indicated notable differences in microbial community structure between Cairo and Ismailia, this difference was not significant at the *p-value* level of 0.05. Such patterns may yet reflect possible microbial community changes due to the different lifestyles in urban and rural areas and/or food intakes. Cairo (A) samples are represented by red dots, whereas Ismailia (D) samples are represented by blue dots. Distinct clustering indicates notable differences in their microbial communities.

Heatmap analysis (Figure 2b) revealed different microbial profiles between samples from Cairo and Ismailia in the similarity of the microbial profiles places the samples in clusters. The color gradient reflects the abundance of a given species. Red and orange shading indicate a relatively high abundance, whereas blue and light shading indicate low abundance. The dendrogram at the top shows the hierarchical clustering of samples, whereas that on the left shows the clustering of microbial species with respect to their co-occurrence patterns. Some microbial species, such as *Segatella hominis* and *Bacteroides thetaiotaomicron*, were highly abundant in the Cairo samples; others, such as *Faecalibacterium prausnitzii* and *Akkermansia muciniphila*, were highly abundant in the Ismailia.

3.2. Taxonomic profiling (phylum level)

The bar plot (Figure 3a, b) revealed that both samples, from Cairo and Ismailia, almost had similar microbial compositions, with minor differences in the relative abundances of specific phyla. In both cases, Bacteroidota and Firmicutes (Bacillota) are the most dominant phyla that constitute almost the vast majority of the microbial com-

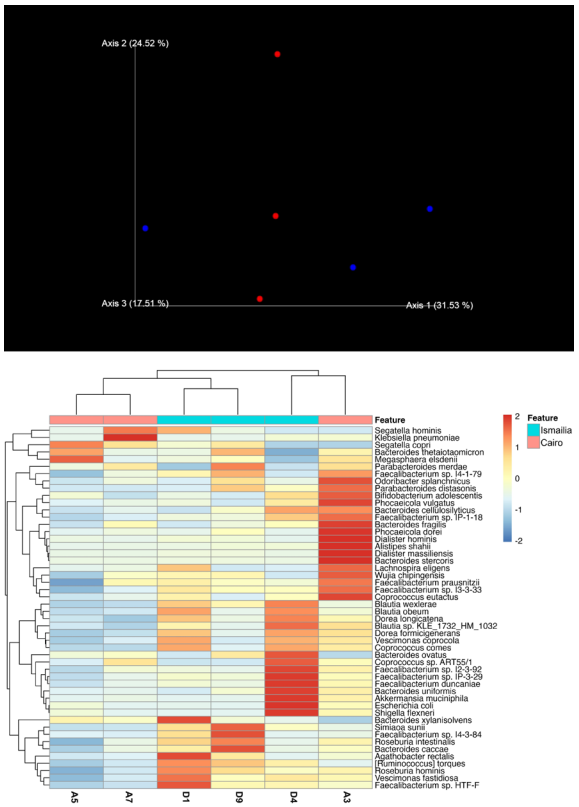

Fig. 2. (a) Principal coordinate analysis (PCoA) plot with Bray-Curtis dissimilarity. Plots illustrating distances between communities in all individual samples (n=6) using the Bray-Curtis distance matrix. The plot shows how different samples cluster based on their microbial community composition. Samples A (Cairo): represented by red dots whereas Samples D (Ismailia): represented by blue dots.. (b) Heatmap depicting the relative abundance of the top 50 microbial species across six samples from Cairo (A3, A5, A7) and Ismailia (D1, D4, D9). The similarity in microbial profiles places the samples in clusters. The color gradient reflects the abundance of a given species. Red and orange shading indicate a relatively high relative abundance, whereas blue and light shading indicate the opposite. The dendrogram at the top shows the hierarchical clustering of samples, whereas that on the left shows the clustering of microbial species with respect to their co-occurrence patterns.

Table 2. Alpha Diversity Metrics of the gut microbiome samples showing observed values, Shannon index, and inverse Simpson index of different sample IDs.

Sample-ID	Observed taxonomic units (OTUs)	Shannon	InvSimpson
A3	8792	3.444615	7.467138
A5	9312	3.639237	8.463820
A7	8927	3.077853	4.359834
D1	7740	1.450012	1.694001
D4	9288	4.309622	27.658565
D9	8557	2.410568	3.429882

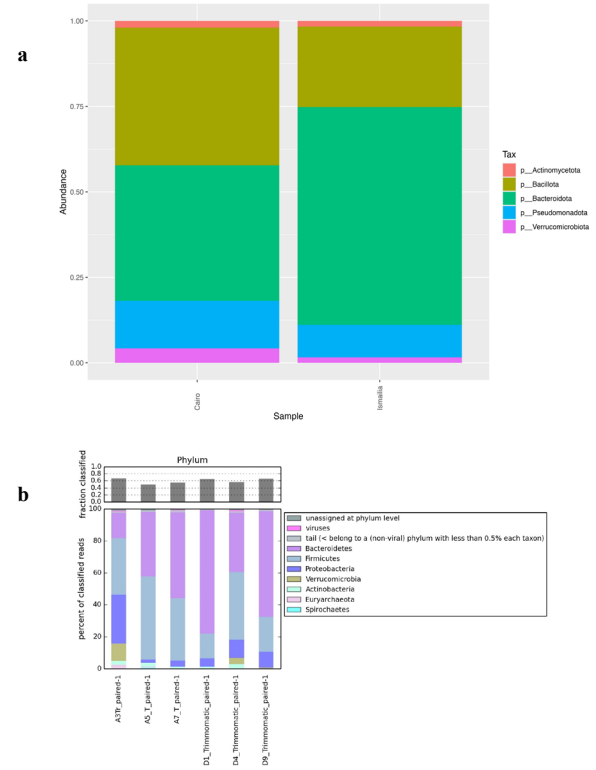

Fig. 3. (a) Bar plot comparing the relative abundance of bacterial phyla in samples from Cairo and Ismailia. The five most abundant phyla that can be observed in the plot are Actinomycetota, Bacillota (Firmicutes), Bacteroidota, Pseudomonadota, and Verrucomicrobiota. The bars represent the microbial composition of each sample; the different colored sections within each bar reflect the relative abundances of the different phyla. (b) Kaiju data (phylum level) stacked-bar plot of the classification of reads at the phylum level across different samples. Each bar represents a sample, and the colors within the bar represent different bacterial phyla. These phyla include Bacteroidetes, Firmicutes, Proteobacteria, Verrucomicrobia, Actinobacteria, Euryarchaeota, Spirochaetes, some that remain unassigned at the phylum level, viruses, and a long tail of phyla with less than 0.5% each. The top subplot shows the fraction of classified reads, whereas the bottom subplot shows their percentage distribution across the different phyla.

munity. While Bacteroidota was more abundant in the Ismailia samples (~60–65%), the relative abundance of Bacillota was greater in the Cairo samples (~35–40%). Other phyla, such as Pseudomonadota, Verrucomicrobiota, and Actinomycetota, presented relatively low abundance at both locations, with the first demonstrating a slightly relatively high relative abundance in the Ismailia samples.

3.3. Assembly

The assemblies with the best values in terms of N50 in this regard were chosen for binning and further analysis (Table 3), which corresponded to metaSPAdes [41] in both the Ismailia and Cairo samples with N50 values of: 12,768 and 12,536, respectively. The quality of the assemblies was assessed by N50, which is a median contig length that covers 50% of the total base content in the assembly. The better this value is, the stronger an assembly is for doing subsequent analyses, including the correct binning.The metagenomic binning of samples from Cairo and Ismailia was subsequentially conducted by MaxBin2 [44], MetaBAT2 [45], and CONCOCT [46].

3.4. Binning

Each of these tools generated different results. MaxBin2 generated 206 bins from the input 85,538 contigs of the Cairo samples; it could bin approximately 79.3% of them with a total binned contig length of 515,308,190 bp, accounting for 85.0%. MetaBAT2 generated 309 bins and CONCOCT generated 237 bins for the same set of samples. Binning analysis yielded a varying number of bins, depending on the tool. MaxBin2 yielded the highest number of bins, likely due to the fact that it relies on marker gene-based approaches. The two clustering algorithms focusing on sequence composition and contig connectivity MetaBAT2 and CONCOCT predicted fewer bins. The results of binning were then optimized via DAS Tool v1.1.2. Improved bin accuracy by integrating outputs from multiple binning tools and producing a more optimized and non-redundant set of bins, optimizing their accuracy through a dereplication, aggregation, and scoring strategy that ensures greater reliability for downstream analysis [37], maximizing the outputs from all three tools and compressing the 20,654 binned contigs to 115 high-quality bins, which significantly improved the accuracy

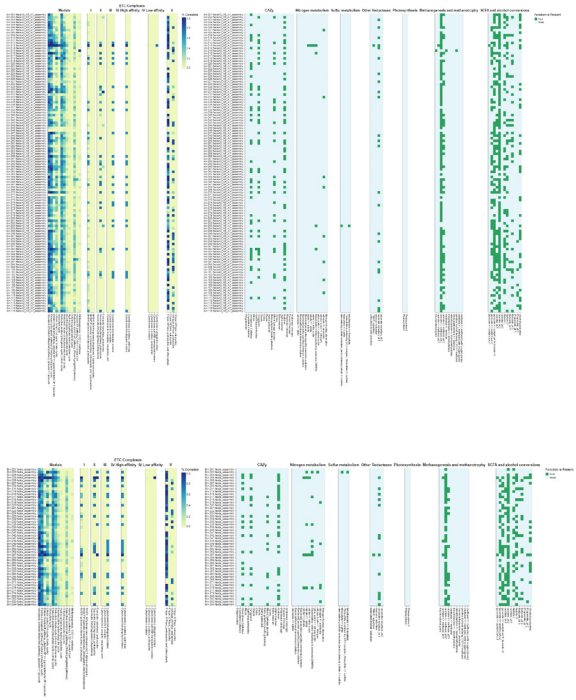

Fig. 4. (a) Functional Gene Distribution using DRAM in gut microbiome samples from Cairo. Heatmap representing the functional gene distribution of gut microbiome isolates from Cairo. A broader distribution of nitrogen metabolism genes indicates prominent nitrogen cycling and may be attributed to a diet rich in proteins. (b) Functional Gene Distribution in gut microbiome samples from Ismailia. This heatmap highlights the presence and completeness of different functional genes in gut microbiome samples from Ismailia. Notably, many genes related to methanogenesis and methanotrophy were observed, which highlights the existence of extensive methane-related processes. In addition to genes implicated in SCFA and alcohol conversions, strong representation has been observed, revealing the main metabolic activities in the gut ecosystem of this population. Like those of Cairo, genes for SCFA and alcohol conversions are well-represented, reflecting their important roles in gut metabolic processes. These differences in gene distribution reflect dietary and other unique environmental influences on the gut microbiomes of Ismailia residents.

and reliability of the metagenomic analysis. For the Ismailia samples, MaxBin2 produced 153 bins from 56,612 input contigs, binning 78.4% of them (total length of 354,042,378 bp, 83.4%). This corresponds to 218 bins for MetaBAT2, whereas CONCOCT delivers 184 bins. The DAS Tool v1.1.2 refinement yielded 14,446 binned contigs, condensing into 50 high-quality bins and improving the accuracy and reliability of the binning process further, enabling better characterization of microbial communities in both regions.

3.5. Annotation and metabolism

Following Rastk annotation, functional profiling was carried out using DRAM and GTDB. The analysis revealed 4310 genes common to both Cairo and Ismailia, indicating a significantly high degree of overlap in the microbial gene pool. Cairo had 775 unique genes, whereas Ismailia had 384 unique genes. This overlap in genes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways from both cities indicates a core of the microbial community and the functional potential of participants from both cities. Some of the shared functions include basic metabolic and biosynthetic processes: central carbon metabolism, amino acid biosynthesis, and lipid metabolism. Common bins underline fundamental metabolic functions conserved across populations and required for microbial survival and growth in both populations (Figures 4a, b). The table below summarizes selected common and unique bins between regions (Table 3). It is important to highlight that KEGG pathway analysis also identified unique pathways in Cairo associated with antibiotic resistance and pathogenicity, which may reflect the increased urban environmental exposure to antibiotics.

With 115 extracted genomes from the Cairo samples and 50 genomes from the Ismailia samples, 165 genomes were classified into the taxonomic class at the minimum

Table 3. Selected common and unique bins between regions.

Bin ID	Region	Annotation Role	Metabolism Role
		Common Bins	
Bin.001.fasta_assembly	Common	Fundamental cellular processes (replication, transcription, translation)	Central metabolic pathways (glycolysis, TCA cycle)
Bin.029.fasta_assembly	Common	Stress response, DNA repair mechanisms, amino acid metabolism, nutrient uptake	Nutrient acquisition and stress response
Bin.075.fasta_assembly	Common	Lipid metabolism (beta-oxidation, phospholipid synthesis)	Lipid biosynthesis and degradation
Bin.018.fasta_assembly	Common	Oxidative stress response (superoxide dismutase, catalase)	Detoxifying reactive oxygen species
Bin.002.fasta_assembly	Common	Sulfur metabolism (sulfate transporters, reductases)	Assimilatory sulfate reduction and sulfur amino acid biosynthesis
Bin.025.fasta_assembly	Common	Motility and chemotaxis pathways (flagellar assembly, motor proteins)	Motility and environmental response
		Unique Bins Cairo	
Bin.079.fasta_assembly	Cairo	Secondary metabolite production (antibiotics, pigments)	Secondary metabolite biosynthesis (antibiotic and pigment production)
Bin.113.fasta_assembly	Cairo	Heavy metal resistance (mercury reductase, cadmium efflux pumps)	Heavy metal detoxification
Bin.043.fasta_assembly	Cairo	Xenobiotic degradation (aromatic ring-hydroxylating dioxygenases)	Degradation of complex organic pollutants
Bin.027.fasta_assembly	Cairo	Antibiotic resistance (beta-lactamases, efflux pumps)	degrading and neutralizing antibiotics
Bin.045.fasta_assembly	Cairo	Pathogenicity (toxins, secretion systems)	Virulence factors and pathogenic interactions
Bin.031.fasta_assembly	Cairo	Aromatic compound degradation (dioxygenases, monooxygenases)	Degradation of polycyclic aromatic hydrocarbons (PAHs)
		Unique Bins Ismailia	
Bin.084.fasta_assembly	Ismailia	Carbohydrate metabolism (CAZymes, plant polysaccharide degradation)	Degrading complex carbohydrates
Bin.077.fasta_assembly	Ismailia	Nitrogen fixation (nitrogenase, ammonia production)	Converting atmospheric nitrogen to ammonia
Bin.003.fasta_assembly	Ismailia	Phosphate metabolism (solubilization, transport)	Phosphorus acquisition and utilization
Bin.059.fasta_assembly	Ismailia	Lignin degradation (laccases, peroxidases)	Breaking down lignin
Bin.072.fasta_assembly	Ismailia	Sulfur oxidation (sulfide:quinone oxidoreductase, adenylylsulfate reductase)	Sulfur metabolism pathways
Bin.091.fasta_assembly	Ismailia	Vitamin B12 biosynthesis (cobalamin production)	Cobalamin (vitamin B12) production

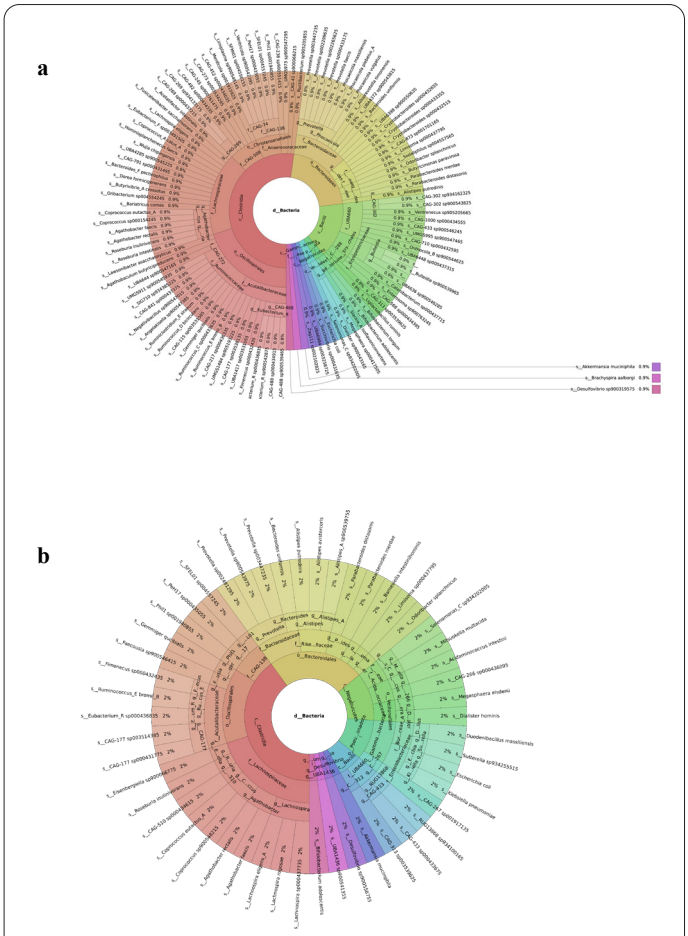

Fig. 5. (a) Krona circular view data of taxonomic genomic annotation for Cairo Ismailia samples (b) Krona circular view data of taxonomic genomic annotation for Ismailia samples.

level using Classify Microbes with GTDB-Tk - v2.3.2 [37]. At the lower taxonomic ranks, the numbers of genomes classified at the order, family, genus, and species levels were 153, 134, 130, and 108 respectively. According to the ANI analysis, all the genomes had no less than 95% average nucleotide identity (ANI) with the reference genomes of the GTDB (Figure 5a, b).

4. Discussion

Studies of alpha and beta diversities in the gut microbiome enable understanding of the complexity and health of microbial communities from a part of the gastrointestinal tract. Alpha diversity measures the species richness within the sample and indicates the diversity and evenness balance of microbes. All three indices used in the comparison of alpha diversity in the study showed no statistically significant differences in microbial diversity between metagenomic samples originating from Cairo and those from Ismailia. Regarding OTUs, there is no indication that the species richness differs between the Cairo and Ismailia samples with a *p-value* of 0.366. Thus, in other words, it means that the incidence of various microbial species between the two locations is almost the same. The inverse Simpson index, which weights species richness and evenness equally, also showed no significant difference between the two locations with a *p-value* of 0.649. This, therefore, means that the overall diversity, including the dominant structure of the communities, is likely to be very similar between Cairo and Ismailia. Similarly, the Shannon index, which considers richness and evenness, does not differ significantly between Cairo and Ismailia, with a *p-value* of 0.481 which is above the threshold of $p < 0.05$ However, it should be taken into consideration that the nonsignificant differences obtained might be linked to the small sample size, which reduces the statistical power of the study. This may limit the detection of small yet biologically important differences in microbial diversity between the two locations.

The PCoA plot indicates that, although species richness and diversity (alpha diversity) may be similar between the two regions, there are differences in the types of species present that make up these communities. Although alpha diversity seems relatively constant across most environments, it is often beta diversity that captures the ecological processes underpinning community assembly in reaction to local conditions [50]. The nonsignificant differences in alpha diversity suggest that the microbial communities in Cairo and Ismailia exhibit similar levels of species richness and diversity. Nevertheless, the PCoA map clearly illustrates considerable differences in species compositions between these locations, which may be accurately measured using beta diversity. This implies that although the overall diversity is similar, the specific taxa and their relative abundances differ between the two regions, most likely due to environmental or ecological factors.

High microbial diversity in the gut is regarded as a critical marker of health; hence, higher values are usually related to better resistance to diseases and good metabolic health [51]. Additionally, geographic location is known to affect microbial diversity, and individuals from different regions mostly present comparatively different microbial profiles [52]. However, the results of this study did not yield large differences between Cairo and Ismailia, suggesting that the environmental and dietary influences may be more similar between the groups under study, resulting in more similar microbial diversity. Both the Shannon and inverse Simpson indices are relevant for health since they indicate the number of species present and how even the population is among those species. These findings imply that, compared with cohorts from different geographical regions in this study, individuals from both cohorts are relatively similar concerning gut microbial diversity. This might indicate a shared dietary pattern, close geographical location, way of life, or other common environmental factors between Cairo and Ismailia, predisposing them to a stable and comparable microbiome. These results agree with some global studies where geographic location alone does not always translate into large differences in microbial diversity, especially in close regions with similar socioeconomic statuses and diets [53].

It should, however, be noted that sample size and the choice of indices may influence the ability to detect significant differences. Samples of a larger size, with more additional indices measured in more extensive studies, may capture subtle differences that were perhaps not detected within this analysis.

On the other hand, beta diversity compares differences in microbial composition between various samples or populations, thereby helping to explain the variation pattern influenced by environmental factors, diet, or diseased conditions in the microbiome. The observed microbial compositions suggest that although the overall microbial communities in Cairo and Ismailia are similarly structured, there is a slight variation in the relative abundance of some phyla. The predominance of the phyla: Bacteroidota and

Bacillota (Firmicutes) from both locations is typical of gut microbiota compositions, where they play very important roles in the digestion of complex carbohydrates and maintenance of gut health. The slight enrichment of Bacteroidota in the Ismailia samples indicates that it is enriched with more fiber and plant-based food components (characteristic of rural lifestyles), which favor their growth [54]. In contrast, the greater abundance of Bacillota in Cairo may be related to a diet containing more proteins and fats, since members of this phylum are known to grow well in such environments [55].

The lower abundances of the Actinomycetota, Pseudomonadota, and Verrucomicrobiota phyla are in accordance with the results of several studies on the gastrointestinal microbiome [56]. Actinomycetota, Pseudomonadota, and Verrucomicrobiota are not as prevalent as Bacteroidota and Bacillota however, their presence remains crucial for the overall health of the colon. In particular, Verrucomicrobiota, mainly *Akkermansia muciniphila*, has recently been increasingly recognized for its beneficial effects on gut integrity and metabolic health [11].

At the species level, in the case of Cairo, there was a high abundance of *Bacteroides thetaiotaomicron*, which is among the most highly abundant bacterial gut symbionts and is implicated in the degradation of complex polysaccharides and the maturation of the host immune system [57]. This could mean that the increased abundance of this bacterium in the samples from Cairo is related to a diet rich in complex carbohydrates, which this bacterium uses as a primary energy source. In urban diets, which often include a variety of plant-based foods as well as processed foods fortified with fibers, *Bacteroides thetaiotaomicron* thrives because of its ability to break down these complex carbohydrates [58]. The relatively high levels of this species in the Cairo samples might also indicate a gut environment that favors efficient energy extraction from food, a common trait in populations with access to diverse diets. Another most abundant species in Cairo is *Megasphaera elsdenii*. *M. elsdenii* is rarely present in the human intestinal microbiota [59]. These bacteria produce butyrate, one of the key short-chain fatty acids important for colon health [60]. It may have a greater incidence in Cairo due to diets that contain greater amounts of fermentable fiber or specific carbohydrates, which favor the proliferation of SCFA-producing bacteria. This might indicate dietary influences that favor the fermentation of fiber and, therefore, may signal some form of balanced fiber intake even within an urban diet.

On the other hand, *Faecalibacterium prausnitzii* was more abundant in the Ismailia samples. This bacterium is one of the largest producers of butyrate, an SCFA [61] with anti-inflammatory activity through the production of anti-inflammatory molecules [62] such as shikimic and salicylic acids [63] and a role in maintaining gut health [64]. This result may indicate that people living in Ismailia have a relatively high intake of dietary fiber [65], especially from fruits, vegetables, and whole grains, which are related to the proliferation of butyrate producers. The abundance of this microorganism is a signature of a healthy gut microbiome and has been implicated in protective effects against inflammatory Crohn's disease and ulcerative colitis [66]. Another species that was more abundant in the Ismailia samples was *Akkermansia muciniphila*. This bacterium has already been reported to play a role in degrading mucin, a glycoprotein that forms a protective layer in the gut [67]. The presence of *Akkermansia muciniphila* is related to a leaner body phenotype and might have a role in improving metabolic health [68], thus protecting against metabolic disorders such as obesity and type 2 diabetes [69]. The relatively high levels of this bacterium in Ismailia may suggest that this population has a diet that is favorable for gut health, and rich in polyphenols and fibers, which already has the ability to favor the growth of *Akkermansia* [70]. This species takes part in the maintenance of gut barrier integrity, which plays a crucial role in protection from inflammation, and hence in gut health. The high abundance of some beneficial species, such as *Faecalibacterium prausnitzii* and *Akkermansia muciniphila*, in Ismailia, may indicate that dietary habits and lifestyles may be more conducive to maintaining a healthy gut microbiome. This finding could be indicative of a diet richer in whole foods, fibers, and probably less processed foods compared with those from Cairo. This may indicate, on the other hand, that at a higher level, *Bacteroides thetaiotaomicron* could be an indication of a diet that is varied but perhaps replete with more processed foods typical of urban settings such as Cairo.

By contrast, the lower abundance of such beneficial bacteria as *Faecalibacterium prausnitzii* and *Akkermansia muciniphila* in Cairo samples, together with a high abundance of *Bacteroides thetaiotaomicron*, indicates a gut microbiota shaped by the protein- and fat-rich urban diet. Such a microbial profile would indeed imply a higher risk of metabolic and inflammatory disorders, including obesity, type 2 diabetes, and gut inflammation in urban populations of Cairo.

These findings strongly underscore the profound impact of dietary and lifestyle differences between regions on gut microbiota composition. These data indicate the urgent need for targeted dietary interventions in urban settings, promoting enhanced fiber and polyphenol intake to increase the beneficial bacteria of the gut and help reduce health risks associated with urban diets. Such interventions have the potential to transform public health by way of reducing diet-related burdens of chronic diseases in such populations.

Coassembly in metagenomics combines the sequencing reads of various samples into one composite assembly. It increases the length of contigs, thus increasing coverage to detect low-abundance species. This approach can improve the quality of the assembly by making full use of combined sequencing depth, thus identifying microbial genomes and genes shared across samples [71]. Hence, samples were coassembled according to their geographical location. Following Rast-k annotation, functional profiling was carried out using DRAM (figure 3a, b). Comparative analysis revealed differences in functional gene structure, probably driven by regional dietary habits and local environmental factors.

Increasing the representation of methanogenesis and methanotrophy-related genes in Ismailia marks methane-related processes within its gut microbiome. This influences methanogenic or methane-producing archaea, which can be found in a wide range of habitats, including the gastrointestinal tracts (GIT) of animals and humans [72]. These genes are so widespread to indicate that methane production and consumption are two relevant processes in the gut microbial ecosystem of this population. Such

functional divergence underlines the role of local ecological and dietary factors on gut microbiota. Enrichment in methane-related genes in Ismailia participants might indicate a diet rich in plant-based and high-fiber foods, which support methanogenic processes. In contrast, while the gut microbiomes from Cairo have a wider distribution of nitrogen metabolism genes, this would point to a more relevant role of nitrogen cycling in these microbial communities. This may be driven by dietary practices, including increased intake of proteins that increase demands for nitrogen transformation processes such as ammonification, nitrification, and denitrification [73].

In both Ismailia and Cairo, there are many genes involved in short-chain fatty acid conversions and alcohol conversions, highlighting the importance of these metabolic pathways in diverse gut environments. These pathways are important in the digestion and absorption of dietary fibers and other complex carbohydrates, resulting in SCFAs that play important roles in gut health [61,62].

Such differences in gut microbiota composition between Cairo and Ismailia could be related to several environmental factors. Inhabitants of Ismailia would more often eat plant-based and high-fiber diets because of fresh agricultural production, which corresponds with the higher abundances of butyrate-producing bacteria *Faecalibacterium prausnitzii* and *Akkermansia muciniphila*. Whereas in Cairo, probably the urban population has bigger access to processed and protein-rich food-mix, reflected in bacterial abundances like *Bacteroides thetaiotaomicron*.

This microbial profile suggests potential implications for metabolic health in Cairo's population. Diets high in protein and fats, particularly those derived from processed foods, have been associated with an increased risk of metabolic conditions such as obesity, type 2 diabetes, and cardiovascular diseases. The higher abundance of bacteria like Bacteroides thetaiotaomicron may indicate an adaptation to these dietary patterns but may also be linked to an altered gut environment with pro-inflammatory abilities. Together with the lower prevalence of butyrate-producing bacteria, such as *Faecalibacterium prausnitzii* may indicate a reduced anti-inflammatory capacity in this population.

Differences in lifestyle are another important factor. For Cairo urban living, the stress associated with it, a trend of eating irregularly, and less physical activity could negatively influence the diversity of gut microbiota. Further, exposure to industrial pollutants and higher rates of antibiotic misuse in urban settings could add to these microbial differences.

On the other hand, genome taxonomic annotation with the GTDB [37] tree was obviously dominated by bacterial genomes from three phyla: Firmicutes, (101); Bacteroidetes, (36), which accounted for nearly 80% of all MAGs. With fractional classification according to standard Krona [38] plots (Fig. 5a, b), most of the genomes within the Firmicutes cluster were further classified into two largest clades, namely Clostridia (80), and Bacilli (21). Firmicutes were dominated by four clostridial populations at the family level, including Lachnospiraceae (30), Oscillospiraceae (26), CAG 508 (7), Acutalibacteraceae (5), and Anaerovoracaceae (3). In the same order, in Bacteroidales, the most abundant family was Bacteroidaceae (26).

5. Conclusion

This study provides a comprehensive analysis of the gut microbiome in two distinct geographic regions of Egypt, using advanced metagenomic sequencing techniques. The findings reveal that although microbial diversity is generally similar between the two regions, notable differences can be observed in the relative abundance of particular microbial species and functional genes. Regional dietary habits and environmental factors might influence these variations. Ismailia showed a higher prevalence of butyrate-producing bacteria, such as *Faecalibacterium prausnitzii* and *Akkermansia muciniphila*, which is associated with gut health and metabolic benefits. On the other hand, in Cairo, there were higher levels of bacteria related to protein and fat digestion, a fingerprint of the urban dietary pattern. Functional gene analysis also showed these differences related to methane and nitrogen metabolism between these two regions. However, this study had a limited sample size, and thus these results could not provide the general basis; therefore, larger studies on these observations are further required. However, the findings from this study will set a base for future studies in which hypotheses on dietary and environmental factors affecting the abundance of *Faecalibacterium prausnitzii* and *Akkermansia muciniphila* including their functions, are tested. Furthermore, urban dietary habits will be analyzed regarding the presence of protein- and fat-metabolizing bacteria, such as *Bacteroides thetaiotaomicron*, in order to better understand their potential relationship with metabolic and inflammatory diseases. These trials will be targeting larger, more diverse populations in order to give a clearer understanding of the health consequences of regional variations in gut microbiota.

Acknowledgements

The authors thank the members of the Beijing Genomics Institute (BGI) in China for conducting the sequencing step.

Author contributions

Amro Hanora led the study design. Kareem Talaat Mohamed contributed to the sample collection, executed the study, performed the data analysis, and wrote the manuscript, while Amro Hanora, Nora Fahmy, Sarah Shabayek and Mahmoud Mohamed Tawfick reviewed the manuscript. As part of our scientific community, all authors approved the final manuscript version.

Statements and declarations

Funding

No funding was received to conduct this study.

Competing interests

The authors declare no competing interests.

References

1. Ley RE, Turnbaugh PJ, Klein S, et al (2006) Human gut microbes associated with obesity. Nat 444:1022–1023. https://doi.org/10.1038/4441022a
2. Bouskra D, Br´ezillon C, B´erard M, et al (2008) Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. Nat 456:507–510. https://doi.org/10.1038/nature07450
3. Gill SR, Pop M, DeBoy RT, et al (2006) Metagenomic analysis of the human distal gut microbiome. Sci 312(5778):1355–1359.

https://doi.org/10.1126/science.1124234
4. Turnbaugh PJ, Ley RE, Mahowald MA, et al (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. Nat 444(7122):1027– 1031. https://doi.org/10.1038/nature05414
5. Roberfroid M, Bornet F, Bouley Ce, et al (1995) Colonic microflora: nutrition and health. Summary and conclusions of an International Life Sciences Institute (ILSI) [Europe] workshop held in Barcelona, Spain. Nutr Rev 53(5):127–130 https://doi.org/10.1111/j.1753-4887.1995.tb01535.x
6. Fan Y, Pedersen O (2021) Gut microbiota in human metabolic health and disease. Nat Rev Microbiology 19:55–71. https://doi.org/10.1038/s41579-020-0433-9
7. Flint HJ, Scott KP, Louis P, et al (2012) The role of the gut microbiota in nutrition and health. Nat Rev Gastroenterol Hepatol 9:577–589. https://doi.org/10.1038/nrgastro.2012.156
8. Laterza L, Rizzatti G, Gaetani E, et al (2016) The gut microbiota and immune system relationship in human graft-versus-host diseaseMediterr J Hematol Infect Dis 8(1):e2016025. https://doi.org/10.4084/MJHID.2016.025
9. Arumugam M, Raes J, Pelletier E, et al (2011) Enterotypes of the human gut microbiome. Nat 473(7346):174–180 https://doi.org/10.1038/nature09944
10. Odamaki T, Kato K, Sugahara H, et al (2016) Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. BMC microbiol 16:1–12. https://doi.org/10.1186/s12866-016-0708-5
11. Jandhyala SM, Talukdar R, Subramanyam C, et al (2015) Role of the normal gut microbiota. World J Gastroenterol 21(29):8787-8803. https://doi.org/10.3748/wjg.v21.i29.8787
12. Thursby E, Juge N (2017) Introduction to the human gut microbiota. Biochem J 474(11):1823–1836. https://doi.org/10.1042/BCJ20160510
13. Chong CYL, Bloomfield FH, O'Sullivan JM (2018) Factors affecting gastrointestinal microbiome development in neonates. Nutrients 10(3):274. https://doi.org/10.3390/nu10030274
14. Yatsunenko T, Rey F, Manary M, et al (2012) Human gut microbiome viewed across age and geography. Nat 486:222–227. https://doi.org/10.1038/nature11053
15. De Filippo C, Cavalieri D, Di Paola M, et al (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci 107(33):14691–14696. https://doi.org/10.1073/pnas.1005963107
16. Feng Z, Long W, Hao B, et al (2017) A human stool-derived Bilophila Wadsworthia strain caused systemic inflammation in specific-pathogen-free mice. Gut Pathog 9:1–10. https://doi.org/10.1186/s13099-017-0208-7
17. Fouesnard M, Zoppi J, Petera M, et al (2021) Dietary switch to Western diet induces hypothalamic adaptation associated with gut microbiota dysbiosis in rats. Int J Obes 45(6):1271–1283. https://doi.org/10.1038/s41366-021-00796-4
18. Ruengsomwong S, La-Ongkham O, Jiang J, et al (2016) Microbial Community of Healthy Thai Vegetarians and Non-Vegetarians Their Core Gut Microbiota, and Pathogen Risk. J Microbiol Biotechnol 26(10):1723–1735. https://doi.org/10.4014/jmb.1603.03057
19. Sidhu S, Kok C, Kunasegaran T, et al (2023 Effect of Plant-Based Diets on Gut Microbiota: A Systematic Review of Interventional Studies. Nutrients 15 (6), 1510. https://doi.org/10.3390/nu15061510
20. [20] Xiao W, Zhang Q, Yu L, et al (2022) Effects of vegetarian diet-associated nutrients on gut microbiota and intestinal physiology. Food Sci Hum Wellness 11(2):208–217. https://doi.org/10.1016/j.fshw.2021.11.002
21. Sun, S., Wang, H., Tsilimigras, M. C. B., Howard, A. G., Sha, W., Zhang, J., Su, C., Wang, Z., Du, S., Sioda, M., Fouladi, F., Fodor, A., Gordon-Larsen, P., & Zhang, B. (2020). Does geographical variation confound the relationship between host factors and the human gut microbiota: a population-based study in China. BMJ Open 10(11), e038163. https://doi.org/10.1136/bmjopen-2020-038163
22. Porras, A. M., Shi, Q., Zhou, H., Callahan, R., Montenegro-Bethancourt, G., Solomons, N., & Brito, I. L. (2021). Geographic differences in gut microbiota composition impact susceptibility to enteric infection. Cell Rep 36(4), 109457. https://doi.org/10.1016/j.celrep.2021.109457
23. Ying, C., Zhao, X., Wang, X., Zhang, Y., & Zhang, Z. (2022). Gut microbiota reveals the environmental adaptation in gastro-intestinal morphology of wild boars in karst region. Ann Microbiol 72(1), 5. https://doi.org/10.1186/s13213-022-01669-5
24. Hashem RA, Yassin AS, Zedan HH, et al (2013) Fluoroquinolone resistant mechanisms in methicillin-resistant Staphylococcus aureus clinical isolates in Cairo, Egypt. J Infect Dev Ctries 7(11):796–803. https://doi.org/10.3855/jidc.3105
25. Yang L, Bajinka O, Jarju PO, et al (2021) The varying effects of antibiotics on gut microbiota. AMB Express 11(1), 116. https://doi.org/10.1186/s13568-021-01274-w
26. Yang L, Sakandar HA, Sun Z, et al (2021) Recent advances of intestinal microbiota transmission from mother to infant. J Funct Foods 87:104719. https://doi.org/10.1016/j.jff.2021.104719
27. Yassour M, Vatanen T, Siljander H, et al (2016) Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci Transl Med 8(343):343ra81. https://doi.org/10.1126/scitranslmed.aad0917
28. Palleja A, Mikkelsen KH, Forslund SK, et al (2018) Recovery of gut microbiota of healthy adults following antibiotic exposure. Nat Microbiol 3(11):1255– 1265. https://doi.org/10.1038/s41564-018-0257-9
29. Case R, Boucher Y, Dahllof I, et al (2007) Use of 16S rRBA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 73(1):278–288. https://doi.org/10.1128/AEM.01177-06
30. Wilkins LG, Ettinger CL, Jospin G, et al (2019) Metagenome-assembled genomes provide new insight into the microbial diversity of two thermal pools in Kamchatka, Russia. Sci Rep 9(1):3059. https://doi.org/10.1038/s41598-019-39576-6
31. VENNISON S (2009) Laboratory manual for genetic engineering. PHI Learn Pvt Ltd.
32. Andrews S, et al (2010) Fastqc: a quality control tool for high throughput sequence data
33. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform 30(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170
34. Chen Y, Chen Y, Shi C, et al (2018) SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigasc 7(1):1-6. https://doi.org/10.1093/gigascience/gix120
35. Bolyen E, Rideout J, Dillon M, et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37(8):852–857. https://doi.org/10.1038/s41587-019-0209-9
36. Menzel P, Ng K, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat commun 7: 11257. https://doi.org/10.1038/ncomms11257
37. Arkin AP, Cottingham RW, Henry CS, et al (2018) KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36(7):566–569. https://doi.org/10.1038/nbt.4163

38. Ondov BD, Bergman NH, Phillippy AM (2011) Interactive metagenomic visualization in a web browser. BMC Bioinform 12:1–10. https://doi.org/10.1186/1471-2105-12-385
39. Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:1–12. https://doi.org/10.1186/gb-2014-15-3-r46
40. Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with kraken2. Genome biol 20:1–13. https://doi.org/10.1186/s13059-019-1891-0
41. Nurk S, Meleshko D, Korobeynikov A, et al (2017) metaSPAdes: a new versatile metagenomic assembler. Genome Res 27(5):824–834. https://doi.org/10.1101/gr.213959.116
42. Li D, Liu CM, Luo R, et al (2015) EGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10):1674–1676. https://doi.org/10.1093/bioinformatics/btv033
43. Peng Y, Leung HC, Yiu SM, et al (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28(11):1420–1428. https://doi.org/10.1093/bioinformatics/bts174
44. Wu YW, Simmons BA, Singer SW (2016) MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinform 32(4):605–607. https://doi.org/10.1093/bioinformatics/btv638
45. Kang D, Froula J, Egan R, et al (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3: e1165. https://doi.org/10.7717/peerj.1165
46. Alneberg J, Bjarnason BS, De Bruijn I, et al (2014) Binning metagenomic contigs by coverage and composition. Nat Methods 11(11):1144–1146. https://doi.org/10.1038/nmeth.3103
47. Parks DH, Imelfort M, Skennerton CT, et al (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25(7):1043–1055. https://doi.org/10.1101/gr.186072.114
48. Brettin T, Davis JJ, Disz T, et al (2015) RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 5(1):1–6. https://doi.org/10.1038/srep08365
49. Shaffer M, Borton MA, McGivern BB, et al (2020) Dram for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 48(16):8883–8900. https://doi.org/10.1093/nar/gkaa621
50. Walters KE, Martiny JB (2020) Alpha-, beta-, and gamma-diversity of bacteria varies across habitats. PLoS One 15(9): e0233872. https://doi.org/10.1371/journal.pone.0233872
51. Manor O, Dai CL, Kornilov SA, et al (2020) Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun 11(1):5206. https://doi.org/10.1038/s41467-020-18871-1
52. Kumar G, Bhadury P (2023) Exploring the influences of geographical variation on sequence signatures in the human gut microbiome. J Genet 102(2):51.
53. Girard C, Tromas N, Amyot M, et al (2017) Gut Microbiome of the Canadian Arctic Inuit. mSphere 2(1):10 1128. https://doi.org/10.1128/mSphere.00297-16
54. Tsigalou C, Paraschaki A, Karvelas A, et al (2021) Gut microbiome and mediterranean diet in the context of obesity. current knowledge, perspectives and potential therapeutic targets. Metab Open 9:100081. https://doi.org/10.1016/j.metop.2021.100081
55. Plaza-D´ıaz J, Manzano M, Ruiz-Ojeda FJ, et al (2022) Intake of slow-digesting carbohydrates is related to changes in the microbiome and its functional pathways in growing rats with obesity induced by diet. Front Nutr 9:992682. https://doi.org/10.3389/fnut.2022.992682
56. Syromyatnikov M, Nesterova E, Gladkikh M, et al (2022) Characteristics of the gut bacterial composition in people of different nationalities and religions. Microorganisms 10(9):1866. https://doi.org/10.3390/microorganisms10091866
57. Martin-Gallausiaux C, Marinelli L, Blotti`ere HM, et al (2021) SCFA: mechanisms and functional importance in the gut. Proc Nutr Soc 80(1):37–49. https://doi.org/10.1017/S0029665120006916
58. Tap J, Mondot S, Levenez F, et al (2009) Towards the human intestinal microbiota phylogenetic core. Environ Microbiol 11(10):2574–2584. https://doi.org/10.1111/j.1462-2920.2009.01982.x
59. Verhoog S, Taneri PE, Roa D´ıaz ZM, et al (2019) Dietary Factors and Modulation of Bacteria Strains of Akkermansia muciniphila and Faecalibacterium prausnitzii: A Systematic Review. Nutrients 11(7):1565. https://doi.org/10.3390/nu11071565
60. Mutuyemungu E, Motta-Romero HA, Yang Q, et al (2024) Megasphaera elsdenii, a commensal member of the gut microbiota, is associated with elevated gas production during in vitro fermentation. Gut Microbiome 5:e. https://doi.org/10.1017/gmb.2023.18
61. Miquel S, Martin R, Rossi O, et al (2013) Faecalibacterium prausnitzii and human intestinal health. Curr Opin Microbiol 16(3):255–26. https://doi.org/10.1016/j.mib.2013.06.003
62. Leylabadlo HE, Ghotaslou R, Feizabadi MM, et al (2020) The critical role of Faecalibacterium prausnitzii in human health: An overview. Microb Pathog 149:104344. https://doi.org/10.1016/j.micpath.2020.104344
63. Miquel S, Leclerc M, Martin R, et al (2015) Identification of metabolic signatures linked to anti-inflammatory effects of Faecalibacterium prausnitzii. mBio 6(2):10– 1128. https://doi.org/10.1128/mBio.00300-15
64. Ferreira-Halder CV, Faria AVdS, Andrade SS (2017) Action and function of Faecalibacterium prausnitzii in health and disease. Best Pract Res Clin Gastroenterol 31(6):643–648. https://doi.org/10.1016/j.bpg.2017.09.011
65. Lin D, Peters BA, Friedlander C, et al (2018) Association of dietary fibre intake and gut microbiota in adults. Br J Nutr 120(9):1014–1022. https://doi.org/10.1017/S0007114518002465
66. Mart´ın R, Chain F, Miquel S, et al (2014) The commensal bacterium faecalibacterium prausnitzii is protective in dnbs-induced chronic moderate and severe colitis models. Inflamm Bowel Dis 20(3):417 430. https://doi.org/10.1097/01.MIB.0000440815.76627.64
67. Kim S, Shin YC, Kim TY, et al (2021) Mucin degrader Akkermansia muciniphila accelerates intestinal stem cell-mediated epithelial development. Gut microbes 13(1):1892441. https://doi.org/10.1080/19490976.2021.1892441
68. Ganesh BP, Klopfleisch R, Loh G, et al (2013) Commensal Akkermansia muciniphila exacerbates gut inflammation in salmonella typhimurium-infected gnotobiotic mice. PloS one 8(9):e74963. https://doi.org/10.1371/journal.pone.0074963
69. Niu H, Zhou M, Zogona D, et al (2024) Akkermansia muciniphila: a potential candidate for ameliorating metabolic diseases. PLoS One 15:1370658. https://doi.org/10.3389/fimmu.2024.1370658
70. Anhˆe FF, Pilon G, Roy D, et al (2016) Triggering Akkermansia with dietary polyphenols: A new weapon to combat the metabolic syndrome? Gut microbes 7(2):146–153. https://doi.org/10.1080/19490976.2016.1142036
71. Delgado LF, Andersson AF (2022) Evaluating metagenomic assembly approaches for biome-specific gene catalogues. Microbiome 10(1):72. https://doi.org/10.1186/s40168-022-01259-2
72. Mafra D, Ribeiro M, Fonseca L, et al (2022) Archaea from the gut microbiota of humans: Could be linked to chronic diseases? Anaerobe 77:102629. https://doi.org/10.1016/j.anaerobe.2022.102629

73. Zeng X, Xing X, Gupta M, et al (2022) Gut bacterial nutrient preferences quantified in vivo. Cell 185(18):3441–3456. https://doi.org/10.1016/j.cell.2022.07.020