BMC Genomic Data

DATA NOTE Open Access

A pilot study revealed the gut microbiota based on 16S rRNA metagenomics in gestational diabetes

Rizwana Kousar[1*], Sadia Latif[1], Mahrukh Zahoor[2] and Sidra Tabassum[3]

Abstract

Objective Microbiome being a potential biomarker holds the future hope for insight into GD pathogenesis, diagnosis and cure. Many factors such as diet, lifestyle, environment, host genetics shape the diversity and composition of human microbiome. Although recent studies have indicated that gut microbiome dysbiosis was significantly associated with the onset of gestational diabetes mellitus (GDM). Information on the alteration of gut microbiota composition in Pakistani women is limited. Therefore, present study was designed to elucidate gut microbiota taxonomic composition and relative abundance of taxa in local GD women.

Data description 16S Metagenomics data revealed variation in bacterial community structure in gestational diabetic (GD), pregnant non-diabetic (PND), non-pregnant diabetic (NPD) and non-pregnant non-diabetic (NPND) women. Predominant bacterial phyla residing faecal sample of GD, NPD, PND included Firmicutes, Bacteriodota, Proteobacteria accounting for 95.07%, 97.1% and 97.04%, of relative abundance respectively. While Predominant phyla inhabiting faecal sample of NPND included firmicutes, Bacteriodota, and Actinobacteriodota accounting for 98.4%. Simpson's Reciprocal Index showed great variability, the value ranged between 11.655 and 16.17. The distance matrix depicted dissimilarity between samples with dissimilarity coefficient values ranging between 0.381 and 0.588. These findings would pave the way for future studies, which will aid in early diagnosis, management and treatment.

Clinical trial number Not applicable.

Keywords Gestational diabetes, Microbiome, 16S rRNA, Metagenomics, Gut microbiota

Objective

Diabetes Mellitus (DM), a chronic metabolic disorder emerged as a global public health concern with global prevalence of 7.5% (374 million people) in 2019 and is expected to rise to 8.0% (454 million) by 2030 and 8.6% (548 million) by 2045 [1]. Gestational diabetes mellitus (GDM) is predominant medical complication of gestation posing greater health risk for both mother and developing fetus [2]. Multiple risk factors including obesity, maternal age, ethnicity, microbiome composition and family history of T2D contribute to the development of GDM [3, 4]. Therefore, it is mandatory to investigate

*Correspondence:
Rizwana Kousar
rizwana.kousar@aiou.edu.pk
[1]Department of Biology, Allama Iqbal Open University, Islamabad, Pakistan
[2]Capital Hospital CDA, Islamabad, Pakistan
[3]Federal Medical College, Islamabad, Pakistan

underlying factors and mechanisms involved in prevention and treatment of GDM. Until recently, limited microbiome knowledge existed due to laborious, time-consuming and anaerobic-dependent culture techniques. Currently, with the advent of high-throughput molecular techniques involving amplicon sequencing targeting 16s rRNA gene is performed efficiently to comprehend taxonomic composition of microbial communities residing human gut.

In the present era, the composition and diversity of intestinal microbiota is gaining ample importance due to association with human health and disease states including gestational diabetes mellitus [5]. Keeping in view the limited reports on gut microbiome composition in Pakistani GD women, current study was designed to unravel gut microbiota taxonomic composition and relative abundance of taxa in Pakistani pregnant and non-pregnant women with and with-out diabetes. Stool samples were collected from four groups including gestational diabetic (GD), pregnant non-diabetic (PND), non-pregnant diabetic (NPD) and non-pregnant non-diabetic (NPND) women. DNA extraction was performed using Invitrogen™ PureLink™ Microbiome DNA Purification Kit. The isolated DNA was subjected to 16S metagenomics employing Ilmunina MiSeq. The objective of this pilot study is to capture group level microbial signatures comprising gut microbiome in local gestational diabetic women.

Data description

The present cross-sectional study recruited twelve participants, equally distributed into 4 groups; (1) Gestational diabetic (GD), (2) Pregnant non-diabetic (PND), (3) Non-pregnant diabetic (NPD) and (4) Non-pregnant non-diabetic (NPND). Gestational diabetic and pregnant non-diabetic were in 3rd trimester (24th to 28th weeks of gestation). None of the participant was suffering from urinary tract infection (UTI), kidney diseases or any other chronic disease and those taking medications, supplements, antibiotics or probiotics with in three weeks prior to sampling were excluded from the study. Relevant parameters of all study subjects were recorded (*Data file* 1). Microbial DNA was extracted from stool sample (0.2 gm) of each participant using the Invitrogen™ PureLink™ Microbiome DNA Purification Kit following manufacturer's instructions. Extracted DNA from three participants of each group was pooled into a single composite sample to capture group level microbial signatures while maintaining cost effectiveness and then subjected to amplification of the V3 and V4 regions of 16 S rDNA using proprietary primers. Sequencing libraries were generated using NEB Next Ultra DNA Library Pre ® Kit for Illumina, following manufacturer's instructions and index codes were provided. The quality of libraries was evaluated and high quality libraries were subjected to sequencing using Illumina Mi seq platform deploying paired-end configuration. Subsequent 16 S rRNA sequencing data (Data set 1, Data set 2, Data set 3, Data set 4) was analysed using Kraken2 [6] and Bracken.

The total high-quality tags obtained from the faecal samples of GD, NPD, NPND and PND for the V3-V4 region of 16 S rRNA were 276,474, 242,225, 358,997 and 270,989 respectively. These were comprised of classified and unclassified tags (*Data file* 2). Total high quality bacterial reads obtained from each group were assigned to phylum, class, order, family and genus rank, respectively (*Data file* 3). High quality reads were clustered into operational taxonomic units at 97% similarity. OTUs found in the faecal sample of GD and NPD subjects were 286 anvd 352, respectively while 277 and 286 OTUs were identified in NPND and PND subjects respectively having an abundance threshold of over 10 reads. OTUs obtained from faecal sample of each group were classified into phyla, classes, orders, families and genera (*Data file* 4).

Dominant bacterial taxa forming core bacterial composition are summarized in figure (*Data file* 5). Relative abundance of each taxon has been depicted in *Data file* 6. Alpha diversity was assessed using shannon's diversity, simpson's index, simpson's reciprocal index, berger parker's diversity and fisher index. Shannon, fisher and simpsons reflected species diversity accounting for species richness and species evenness. Berger Parker showed proportional abundant species. The data depicted variability in bacterial communities residing the faecal samples of under study groups (*Data file* 7).

Distance matrix generated using python script indicated the dissimilarities between samples analysed pairwise. The dissimilarity coefficients ranged from 0.381 to 0.588. Maximum variability was depicted between gestational diabetic and non-pregnant diabetic women with a dissimilarity coefficient of 0.588. Dissimilarity coefficient of 0.583 was observed between non-pregnant diabetic and non-pregnant non-diabetic women (*Data file* 8) (see Table 1).

Limitations

The current pilot study utilized 16S rRNA metagenomics to elucidate the gut microbiota profiles of four distinct metabolic groups including GD, PND, NPD and NPND. The data revealed notable perturbation in the relative abundance of dominant taxa across four study groups. Firmicutes is the predominant phylum overall in under study samples followed by Bacteroidota in GD, PND and NPND while Proteobacteria was found to be second dominant phyla in NPD. However, current study was preliminary designed to address potential difference in bacterial composition across four metabolic groups comprised of twelve participants in total. Due to limited

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Demographic and clinical parameters of study subjects	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978088) [7]
Data file 2	Tags statistics obtained from faecal samples of under study groups	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978127) [8]
Data file 3	Classified Tags obtained from faecal samples of under study groups	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978133) [9]
Data file 4	Crona exhibiting the bacterial community structure of faecal samples. (a) GD, (b) NPD, (c) NPND, (d) PND Krona depicted abundance and hierarchy simultaneously. Circles from inside to outside stand for different taxonomic ranks and the area of sector means respective proportion of different OTUs	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978160) [10]
Data file 5	Bacterial composition and read abundance visualization in faecal samples of understudy groups [(a) GD, (b) NPD, (c) NPND, (d) PND] at various taxa level by Sankey diagrams	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978178) [11]
Data file 6	The relative abundance of major phylum, class, order, family and genus of bacteria in different study groups	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978181) [12]
Data file 7	Alpha diversity of bacterial community inhabiting gut in the under-study samples	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978184) [13]
Data file 8	Beta-diversity represents dissimilarity between the under-study samples	*MS word document (.docx)*	*Figshare* (https://doi.org/10.6084/m9.figshare.30978187) [14]
Data set 1	16S metagenomics from stool sample of gestational diabetic	*Fasta*	*NCBI Sequence Read Archive*https://identifiers.org/ncbi/insdc.sra:SRX26112933 [15]
Data set 2	16S metagenomics from stool sample of non pregnant diabetic	*Fasta*	*NCBI Sequence Read Archive*https://identifiers.org/ncbi/insdc.sra:SRX26112934 [16]
Data set 3	16S metagenomics from stool sample of non pregnant non diabetic	*Fasta*	*NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX26112935 [17]
Data set 4	16S metagenomics from stool sample of pregnant non diabetic	*Fasta*	*NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX26112936 [18]

sample size and pooling strategy adopted to minimize sequencing cost, the findings of present investigation are best suited for exploratory comparisons or for inclusion in meta-analyses where pooled data are acceptable. Furthermore, these findings should not be used in meta-analysis to interpret individual sample level taxonomic variability. Despite this constraint, the data revealed intriguing preliminary insights.

Future studies recruiting large sample size will better comprehend and compare bacterial community structure and composition across four groups including gestational diabetic, pregnant non-diabetic, non-pregnant diabetic and non-pregnant non-diabetic women taking specific diet and practicing particular lifestyle.

Abbreviations

GD Gestational diabetes
PND Pregnant non-diabetic
NPD Non-pregnant diabetic
NPND Non-pregnant non-diabetic

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-026-01413-x.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
We acknowledge Bibi Humna (BS student) and Areej Arshad (BS student) and all the subjects who participated in this study.

Author contributions
Rizwana Kousar conceived the idea, Designed and supervise the study, wrote draft. Sadia Latif performed wet lab and analyzed the data, reviewed the final draft. Mahrukh Zahoor reviewed the final draft. Sidra Tabassum performed sampling, analyzed the data.

Funding
There is no funding source available.

Data availability
The data described in this Data note can be freely and openly accessed on [NCBI] under [SRX26112933, SRX26112934, SRX26112935, SRX26112936]. Please see Table 1 for details and links to the data.

Declarations

Ethics approval and consent to participate
The current study involving human participants was conducted in accordance with the Declaration of Helsinki and approved by the ethical committee, Department of Biology, AIOU. Written informed consent was also obtained from all participants before sample collection.

Consent for publication
Written consent to publish the data was also taken all participants.

Competing interests
The authors declare no competing interests.

Received: 16 August 2025 / Accepted: 26 February 2026
Published online: 28 March 2026

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, Colagiuri S, Guariguata L, Motala AA, Ogurtsova K, Shaw JE, Bright D, Williams R, IDF Diabetes Atlas Committee. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Res Clin Pract. 2019;157:107843. https://doi.org/10.1016/j.diabres.2019.107843. Epub 2019 Sep 10. PMID: 31518657.
2. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS, Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med. 2005;352(24):2477–86. https://doi.org/10.1056/NEJMoa042973. Epub 2005 Jun 12. PMID: 15951574.
3. Li WZ, Stirling K, Yang JJ, Zhang L. Gut microbiota and diabetes: From correlation to causality and mechanism. World J Diabetes. 2020;11(7):293–308. https://doi.org/10.4239/wjd.v11.i7.293. PMID: 32843932; PMCID: PMC7415231.
4. Lu Y, Huang J, Yan J, Wei Q, He M, Yuan C, Long Y. Meta-analysis of risk factors for recurrent gestational diabetes mellitus. BMC Pregnancy Childbirth. 2025;25(1):257. https://doi.org/10.1186/s12884-025-07367-9. PMID: 40057727; PMCID: PMC11889910.
5. Hou K, Wu ZX, Chen XY, et al. Microbiota in health and diseases. Sig Transduct Target Ther. 2022;7:135. https://doi.org/10.1038/s41392-022-00974-4.
6. Lu J. and Steven L. Salzberg. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome* 8.1 (2020): 124.
7. Figshare. https://doi.org/10.6084/m9.figshare.30978088.
8. Figshare. https://doi.org/10.6084/m9.figshare.30978127.
9. Figshare. https://doi.org/10.6084/m9.figshare.30978133.
10. Figshare. https://doi.org/10.6084/m9.figshare.30978160.
11. Figshare. https://doi.org/10.6084/m9.figshare.30978178.
12. Figshare. https://doi.org/10.6084/m9.figshare.30978181
13. Figshare. https://doi.org/10.6084/m9.figshare.30978184
14. Figshare. https://doi.org/10.6084/m9.figshare.30978187
15. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX26112933
16. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX26112934
17. NCBI Sequence Read. *Archive* https://identifiers.org/ncbi/insdc.sra:SRX26112935
18. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX26112936

Publisher's note