International Journal of Scientific Research in Dental and Medical Sciences

www.ijsrdms.com

Selective Microbial Biomarkers in Type-2 Diabetes with Principal Component Analysis and Receiver-operating Characteristic Curves

Ifeanyi Onyema Oshim[a,*], Nneka Regina Agbakoba[a], Ogonna Celestine Oguejiofor[b]; Kingsley C Anukam[a,c,d]

[a] *Department of Medical Laboratory Science, Faculty of Health Sciences and Technology, Nnamdi, Azikiwe University, Nnewi Campus, Nnewi, Nigeria*
[b] *Department of Medicine, Faculty of Medicine, College of Health Sciences, Nnamdi Azikiwe University,Nnewi Campus, Anambra State, Nigeria*
[c] *Department of Pharmaceutical Microbiology, Faculty of Pharmacy, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria*
[d] *Uzobiogene Genomics, London, Ontario, Canada*

ARTICLE INFO

Article history:
Received 08 January 2021
Received in revised form 24 February 2021
Accepted 09 March 2021
Available online 17 March 2021

Keywords:
Biomarkers
Faecalibacterium
Gut microbiome
Roseburia

ABSTRACT

Background and aim: Gut microbiota dysbiosis has been associated with metabolic disorders, such as obesity and Type-2diabetes Mellitus. This study evaluated the sensitivity, specificity, and diagnostic accuracy of selective biological markers in T2 diabetes.

Materials and methods: Stool samples were collected from 110 confirmed T2DM and ten non-T2DM subjects, and bacterial DNA extracted. The V4 areas of bacterial 16S rRNA were amplified and sequenced using an Illumina NextSeq 500 platform.

Results: There was a strong correlation between the family Streptococcaceae, Sphingobacteriaceae, Alcaligenaceae, Paraprevotellaceae, and Enterobacteriaceae with T2D. The genus-Faecalibacterium and genus-Roseburia demonstrated a negative correlation with T-2D. The Receiver-operating characteristic (ROC) of the Area Under Curve (AUC) value of gut microbiome was in increasing order with family> Genus > Species > Order> Class.Therefore, we classified the diagnostic accuracy as poor ($0.6 < ROC\ AUC \leqslant 0.7$), failed ($ROC\ AUC \leqslant 0.6$), good ($0.8 < ROC\ AUC \leqslant 0.9$), excellent ($0.9 < ROC\ AUC \leqslant 1.0$) and fair ($0.7 < ROC\ AUC \leqslant 0.8$).According to the results, the selected bacterial family/taxa provided fair diagnostic tools followed by genus/taxa, whereas other bacterial genera /taxa failed the diagnostic accuracy.

Conclusion: We could demonstrate the gut microbiome-based classifiers' potential for identifying people suffering from the increased risks for T2D. The findings also revealed that genus-Faecalibacterium, genus-Roseburia, and genus-Phascolarctobacterium were the main discriminants for T2D.

1. Introduction

According to the studies in the field, type II diabetes (T2D) is a major metabolic disease throughout the world with the characteristic of the prolonged high level of blood sugar due to the body's inability to use the generated insulin.[1] Over time, the characterization of the gut microbiome from phyla to species levels in diabetes and detection of gut bacterial markers that may differentiate Type2-diabetes mellitus (T2DM) cases and controls is essential.[2] Moreover, potent differences in the composition of the gut microbiome in conjunction with diabetes can lead to major biomarkers that can be used to diagnose diseases. Currently, the classification of T2DM individuals is mainly based on certain gene clusters and markers.[1] A previous study reveals that meta-genomic profiles may be employed for identifying T2DM individuals with higher precision from a European women cohort.[3] Principal components analysis (PCA) defines the typical patterns of specific illnesses in the correlation between several variables.[4] Often, the first 2 or 3 elements from PCA would be utilized for determining if it is possible to cluster people into two categorization groups with regard to the disease and control groups.[4] Also, PCA has been applied to derive the factors and orthogonal rotation (varimax option) for extracting the non-associated parameters.[5] Such a varimax technique seeks to minimize the number of

* *Corresponding author. Ifeanyi Onyema Oshim*
E-mail address: ifeanyioshim@gmail.com
Department of Medical Laboratory Science, Faculty of Health Sciences and Technology, College of Health Science, Nnamdi Azikiwe University Nnewi Campus, Nnewi, Anambra State, Nigeria
http://doi.org/10.30485/IJSRDMS.2021.272435.1110

indicators with the increased loading on one factor.[6] Therefore, the first extracted factor would include the greatest possible variance in the data-set.

Moreover, the second component independent of the first component would illustrate the maximum possible contribution of the remaining variance so that no components correlated to each other.[7] The application of the ROC model to a Chinese cohort indicated the capability of their model to differentiate T2DM cases from the healthy adults as determined by the gene clusters with the AUC, which operated at the characteristic curve (AUC) of 0.58 for Chinese T2DM cases.[7] Furthermore, the most discriminatory gene clusters showed differences between Chinese and European cohorts, which reflects additional investigation of the T2DM meta-genomic predictive instruments as well as diagnostic biomarkers for certain populations.[2] Furthermore, 16S rRNA sequencing can be one of the more affordable methods to characterize microbiota than whole-genome shotgun sequencing. However, there is no information on using the fecal microbial community structure for diabetes prediction in adults.[2]

Moreover, only a few related studies have considered using Principal Component Analysis (PCA) and an area under the ROC to provide the selective microbial biomarkers for T2DM. A previous study on the metagenomic cluster model revealed that Roseburia and Faecalibacterium prausnitizii were identified as a discriminant for T2D.[8] These findings have supplemented the South-East Nigerian study and African gut microbiome research related to T2DM. However, our research aimed to predict with ROC the relevant gut microbiota related to Type-2 diabetes mellitus.

2. Materials and methods

Ethical Approval

The Nnamdi Azikiwe University Teaching Hospital, Nnewi (NAUTH), the ethical review committee, has verified this research with the reference number of NAUTH/CS/VOL.11/183/2018/121. Notably, the written informed consents and verbal informed consent were received from the educated and non-educated participants before starting the sample collection process.

Research Design

We designed cross-sectional research involving older diabetic and healthy people. A simple random sample was used to obtain samples from the subjects with 110 confirmed T2DM subjects, ranging from 20 to 80 years, to be included in this research. Ten (10) age-matched, healthy people from the control group also participated without any T2D familial history.

Data collection

For data collection, we used an interviewer-administered structured questionnaire (on lifestyle, ethnicity, medical history, educational level, the consumed medicines not less than within the months before the beginning of the research) for obtaining the medical information of the subjects of the study.

DNA Extraction, sequencing, and PCR amplification

The DNA extraction from each subjects' stool sample was done through a QiaAMP mini-stool kit (Qiagen, Valencia, CA: USA). For assessing the variety and composition of the gut bacterial communities, the research followed the same protocols as previously described.[14]

Therefore, PCR has been amplified on the above region with primer pair (515F: GTGCCAGCMGCCGCGGTAA and 806R: GGACTACHVGGGTWTCTAAT) as previously described.[9]

Statistical analysis

This study applied Statistical package of social sciences (SPSS version 20) software to calculate Principal components analysis (PCA) and Receiver-operator characteristic (ROC) while MedCalc statistical software application was used to calculate the Sensitivity, Specificity, and Predicative values. It is widely accepted that PCA is commonly employed to identify the big numbers of microbial species in the humans' fecal samples, demonstrating the higher correlations of several species/ taxonomy.[10-12] Receiver-operating characteristic analysis has been implemented with gut pathogenic bacteria drawn against T2DM. Moreover, the AUC has been expected to assess the predictive power. Ultimately, diagnostic accuracy has been grouped as poor ($0.6 < \text{ROC AUC} \leqslant 0.7$), failed ($\text{ROC AUC} \leqslant 0.6$), good ($0.8 < \text{ROC AUC} \leqslant 0.9$), excellent ($0.9 < \text{ROC AUC} \leqslant 1.0$) and fair ($0.7 < \text{ROC AUC} \leqslant 0.8$).[13]

3. Results

Table 1a revealed that four and five components for T2DM and Control subjects were derived by factor analysis using PCA with varimax rotation for their relative abundance of the bacteria family. The first four components (factors) in the initial solution exhibited an Eigenvalue of more than one, and they accounted for nearly 40% of the observed variations in the bacteria family relative abundance among the T2DM subject. It was observed that the most clustered variables, such as Streptococcaceae, Sphingobacteriaceae, Alcaligenaceae, Paraprevotellaceae, Enterobacteriaceae, Bifidobacteriaceae, Bacteroidaceae, and Porphyromonadaceae, had loadings of 0.967, 0.915, 0.902, 0.886, 0.842, 0.705, 0.700 and 0.621 on factor 1 amongst the T2DM subject. The above factor showed an increase in Streptococcaceae, Sphingobacteriaceae, Alcaligenaceae, Paraprevotellaceae, Enterobacteriaceae, Bifidobacteriaceae, Bacteroidaceae, and Porphyromonadaceae increased the Component 1 value. Factor 2, labeled as Ruminococcaceae, Lachnospiraceae, and Clostridiaceae, exhibited higher loading equal to 0.841, 0.808, and 0.635, which explained 16.0% of the total variations. In addition, out of control, the first five factors explained nearly 25% of the observed variations in the bacteria family relative abundance (Table 1b).

It was noted that the most clustered variables, such as Clostridiaceae, Desulfovibrionaceae, and Paraprevotellaceae, showed loading equal to 0.947,0.858 and 0.597 on factor 1 among the control subject. Factor 2, labeled as Alcaligenaceae and Coriobacteriaceae, exhibited hier high loading equal to 0.896 and 0.836, which explained 21.0% of the total variations. Moreover, factor 3, labeled as Enterobacteriaceae and Bacteroidaceae, exhibited higher loading of 0.930 and 0.651, which explained 16.0% of the total variation. Furthermore, spatial representation shows the relationship of the extracted factors and bacteria family-groups among the T2DM and Control subject (Figs. 1a and 1b). Notably, the symbols – and + after PC numbers illustrate if this PC coefficient is negative or positive in the relative classification model. Finally, the positive coefficient refers to the enhanced probability of the person characterized to Type2-diabetes mellitus by a higher score of PC.

Table 1a. A rotated component matrix for family among T2DM subjects.

Family	Comp1	Comp 2	Comp 3	Comp 4
Streptococcaceae	0.967	0.00	0.00	0.00
Sphingobacteriaceae	0.915	0.00	0.00	0.00
Alcaligenaceae	0.902	0.00	0.00	0.00
Paraprevotellaceae	0.886	0.00	0.00	0.00
Enterobacteriaceae	0.842	0.00	0.00	0.00
Bifidobacteriaceae	0.705	0.00	0.00	0.55
Bacteroidaceae	0.700	0.00	0.586	0.00
Porphyromonadaceae	0.621	0.00	0.444	0.486
Veillonellaceae	0.577	0.00	0.00	0.482
Ruminococcaceae	0.00	0.841	0.00	0.00
Lachnospiraceae	0.00	0.808	0.00	0.00
Clostridiaceae	0.00	0.635	0.392	0.00
Desulfovibrionaceae	0.00	-0.479	0.375	0.373
Prevotellaceae	0.00	0.00	-0.906	0.00
Verrucomicrobiaceae	0.00	-0.448	0.672	0.00
Coriobacteriaceae	0.00	0.00	0.00	0.77
Peptostreptococcaceae	0.00	0.00	0.00	0.00
Initial Eigenvalues (% of variances)	40.998	16.07	11.452	9.034

Extraction method; PCA; Rotation method; Varimax with Kaiser normalization; Rotation converged in seven repetitions. Key 0: No correlation.

Table 1b. A rotated component matrix for family among control subjects.

Family	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5
Streptococcaceae	0.00	0.00	0.00	0.00	0.00
Sphingobacteriaceae	0.00	0.00	0.651	0.936	0.00
Alcaligenaceae	0.00	0.896	0.00	0.00	0.00
Paraprevotellaceae	0.597	0.00	-0.543	0.00	0.00
Enterobacteriaceae	0.00	0.00	0.93	0.00	0.00
Bifidobacteriaceae	-0.787	-0.304	0.00	0.00	0.00
Bacteroidaceae	0.00	0.00	0.00	0.622	0.00
Porphyromonadaceae	0.00	0.00	0.00	0.871	0.00
Veillonellaceae	0.00	0.00	0.00	0.00	0.78
Ruminococcaceae	0.00	-0.814	-0.408	-0.34	0.00
Lachnospiraceae	-0.562	-0.493	-0.323	0.00	-0.475
Clostridiaceae	0.947	0.00	0.00	0.00	0.00
Desulfovibrionaceae	0.858	0.00	0.632	0.00	0.00
Prevotellaceae	0.304	0.00	0.00	-0.386	0.00
Verrucomicrobiaceae	0.00	0.00	0.00	0.00	0.812
Coriobacteriaceae	0.00	0.836	0.00	-0.417	0.00
Peptostreptococcaceae	0.467	0.00	0.731	0.00	0.00

Initial Eigenvalues (% of variances)	25.902	21.805	16.814	13.497	8.286

Extraction method; PCA; Rotation method; Varimax with Kaiser normalization; Rotation converged in seven repetitions. Key 0: No correlation.

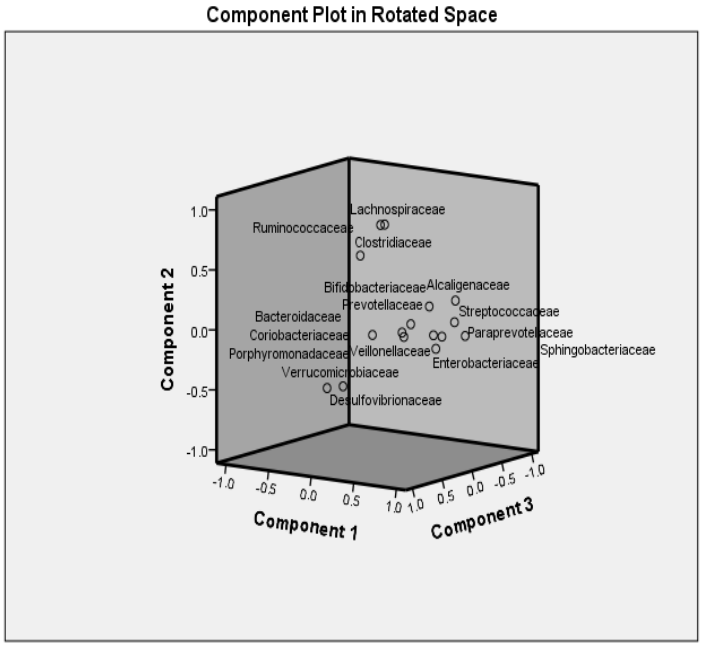


Fig. 1a. Spatial representation of relationships between derived factors and bacteria-family-groups among T2DM.

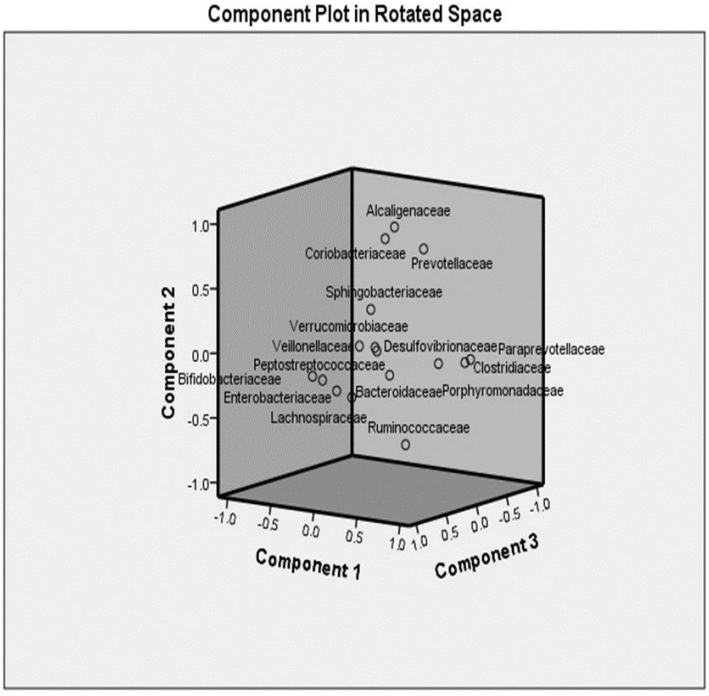


Fig. 1b. Spatial view of the relationship of the extracted factors with bacteria family-groups among Controls.

Table 2a demonstrates the extraction of six and seven components of each T2DM and Control subjects by factor analysis using PCA with varimax rotation for their relative abundance of the bacteria genus. The first six components (factors) in the initial solution, as can be seen, have an Eigenvalue of more than 1, which accounted for nearly 24% of the observed variation in the bacteria genus relative abundance among the T2DM subject. It was observed that the most clustered variables, such as Serratia, Escherichia, Streptococcus, and Blautia, had loadings of 0.901, 0.873, and 0.823 on factor 1 amongst the T2DM subject. The above factor illustrated that an increase in the abundance of Serratia, Escherichia, Streptococcus, and

Blautia increased the Component 1 value and increased the likelihood of diabetes. Factor 2, labeled as Blautia, Coprococcus, and Roseburia, exhibited higher loading equal to 0.808, 0.778, and 0.507, which explained 18.0% of the total variations. The third factor labeled as Bifidobacteria, Dorea, and Parabacteroides had a high loading of 0.924, 0.797, and 0.625 that explained 12.0 % of the total variances. Moreover, out of control, the first seven factors exhibited nearly 30% of the observed variation in the bacteria genus relative abundance (Table 2b). It was noted that the most clustered variables, such as Coprococcus, Faecalibacterium, Bifidobacterium, and Blautia, showed loading equal to 0.905, 0.872, 0.774, and 0.743 on factor 1 among the control subject. Factor 2 labeled as Serratia, Escherichia, Bacteroides, Dorea had a high loading of 0.931, 0.896, 0.871and 0.674which explained 20.0% of the total variation. The third factor labeled as Lactobacillus, Streptococcus, and Roseburia had a high loading of 0.935, 0.894, and 0.849, which explained 12.0 % of the total variances. Furthermore, spatial representation shows the correlation of the extracted components and bacteria genus-groups among the T2DM and Control subject (Figs. 5a and 5b). It should be mentioned that the symbols – or + after the PC numbers represent if this PC coefficient is negative or positive in the relative classification model. Positive coefficient refers to the enhanced probability of the individual being characterized to Type2-diabetes mellitus by a higher PC score. However, genus Faecalibacterium- within column Comp1 in this study shows that Faecalibacterium negatively correlated to Comp 1 (PC1 for T2DM). This suggests that greater Faecalibacterium abundance would decline the Comp 1 value. Since Comp 1 (PC1 for Control subject) positively correlated to being healthy, it suggests that increased Faecalibacterium abundance leads to a decreased likelihood of diabetes.

Table 2a. A rotated Component matrix for Genus among T2DM subjects.

Genus	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6
Serratia	0.901	0.00	0.00	0.00	0.00	0.00
Escherichia	0.873	0.00	0.00	0.00	0.00	0.00
Streptococcus	0.823	0.00	0.00	-0.339	0.00	0.00
Blautia	0.00	0.808	0.00	0.00	0.00	0.00
Coprococcus	0.00	0.778	0.41	0.00	0.00	0.00
Faecalibacterium	-0.386	0.749	0.00	-0.328	0.00	0.00
Akkermansia	0.00	-0.635	0.345	0.00	-0.463	0.00
Roseburia	-0.337	0.507	-0.33	-0.375	-0.314	-0.304
Bifidobacterium	0.00	0.00	0.924	0.00	0.00	0.00
Dorea	0.00	0.00	0.797	0.00	0.00	0.00
Parabacteroides	0.00	0.00	0.625	0.00	0.00	0.402
Clostridium	0.00	0.00	0.00	0.912	0.00	0.00
Oscillospira	0.00	0.00	0.00	0.884	0.00	0.00
Ruminococcus	0.415	0.00	0.00	0.719	0.00	0.00
Lactobacillus	0.00	0.00	0.00	0.00	0.824	0.00
Lachnospira	0.00	0.00	-0.386	0.00	0.758	0.364
Bacteroides	0.321	-0.364	0.4	0.00	-0.508	0.368
Succinivibrio	0.00	-0.327	0.00	0.00	0.484	0.00
Sutterella	0.00	0.00	0.00	0.00	0.00	-0.812
Phascolarctobacterium	-0.398	0.00	0.00	0.00	0.00	0.669
Prevotella	-0.467	0.00	-0.32	-0.347	0.375	-0.473
Initial Eigenvalues (% of variances)	24.308	18.821	12.315	10.216	8.099	5.969

Extraction method; PCA; Rotation method; Varimax with Kaiser normalization; Rotation converged in seven repetitions. Key 0: No correlation.

Table 2b. A rotated Component matrix for Genus among Control subjects.

Genus	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7
Serratia	0.00	0.931	0.00	0.00	0.00	0.00	0.00
Escherichia	0.00	0.896	0.346	0.00	0.00	0.00	0.00
Streptococcus	0.00	0.317	0.894	0.00	0.00	0.00	0.00

Blautia	0.743	-0.469	0.00	0.00	0.00	0.316	0.00
Coprococcus	0.905	-0.343	0.00	0.00	0.00	0.00	0.00
Faecalibacterium	0.872	0.00	0.00	0.00	0.00	0.00	0.00
Akkermansia	0.00	0.00	0.00	0.00	0.00	-0.876	0.00
Roseburia	-0.712	0.00	0.849	0.00	0.00	0.00	0.00
Bifidobacterium	0.774	0.00	0.476	0.00	0.00	0.00	0.00
Dorea	0.00	0.674	-0.403	0.00	0.578	0.00	0.00
Parabacteroides	0.00	0.00	0.00	0.935	0.00	0.00	0.00
Clostridium	-0.081	0.00	0.00	-0.345	0.355	0.00	0.00
Oscillospira	-0.673	0.00	-0.48	0.00	0.00	0.00	0.362
Ruminococcus	0.00	-0.4	0.00	0.375	0.00	0.00	0.00
Lactobacillus	0.00	0.00	0.922	0.00	0.00	0.00	0.00
Lachnospira	0.352	0.00	0.469	0.359	0.00	0.622	0.00
Bacteroides	0.00	0.871	0.00	0.408	0.00	0.00	0.00
Succinivibrio	-0.37	0.00	0.00	0.00	0.736	0.00	0.00
Sutterella	0.00	0.00	0.00	0.00	0.00	0.00	-0.947
Phascolarctobacterium	0.00	0.00	-0.317	0.00	-0.807	0.00	0.00
Prevotella	-0.513	-0.428	0.00	-0.533	0.00	0.00	-0.362
Initial Eigenvalues(% of variances)	30.6	20.985	12.546	10.07	8.74	6.206	5.409

Extraction method; PCA; Rotation method; Varimax with Kaiser normalization; Rotation converged in seven repetitions. Key 0: No correlation.

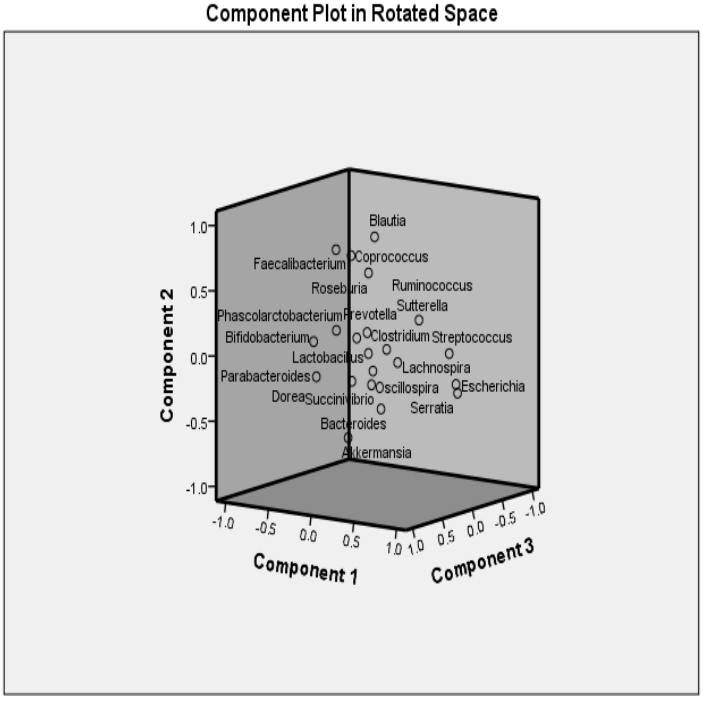


Fig. 2a. Spatial representation of relationships between derived factors and bacteria genus- groups among T2DM.

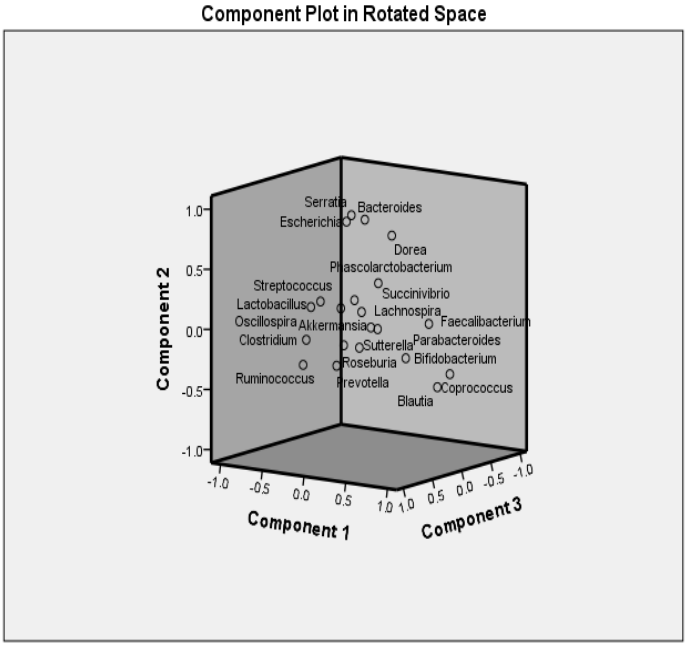
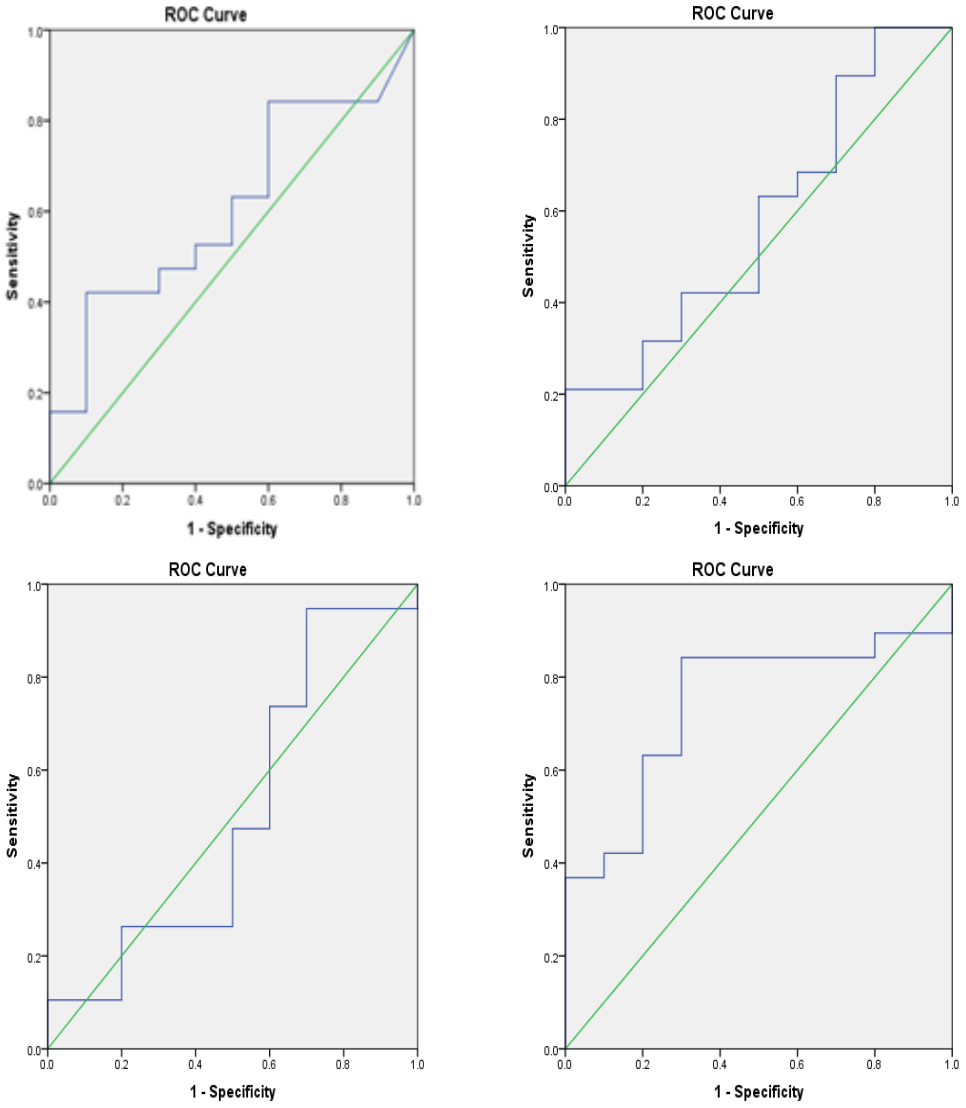


Fig. 2b. Spatial view of the relationship between the extracted factors and bacteria genus-groups among control.

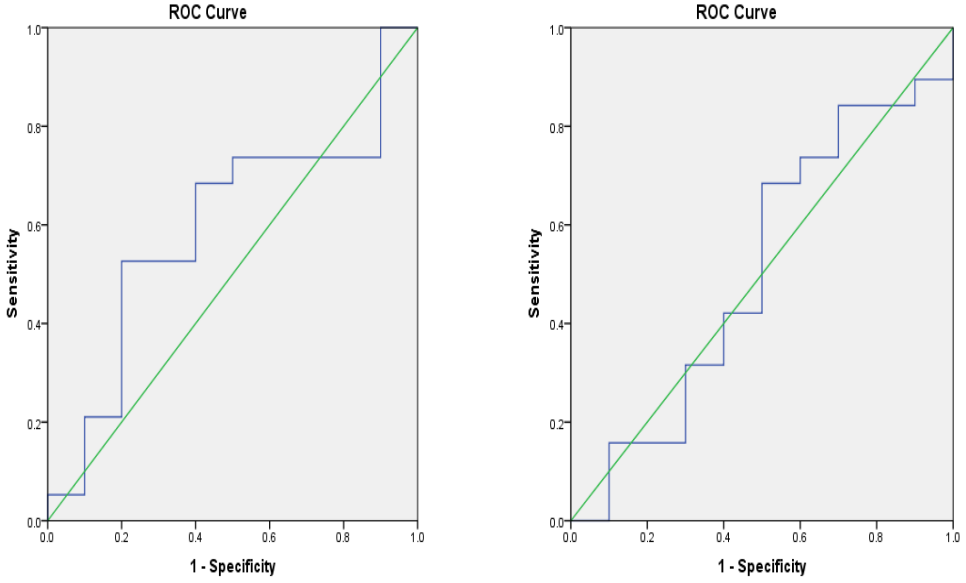
The Receiver operating characteristic analysis has been done for evaluating the predictive power of the gut microbiome family for T2DM. Therefore, according to the results, the family and T2DM have been compared, and AUC has been more than 0.500. Moreover, if the AUC is closer to 1.0, a more acceptable prediction of the bacterial counts predict T2DM will be presented. When a comparison was made between family Paraprevotellaceae/ Porhyromonadaceae and T2DM, the AUC were 0.608 and 0.742 ($P<0.01$), which showed that family Paraprevotellaceae/ Porhyromonadaceae had more predictive power for T2DM than other families. However, family Porhyromonadaceae had the best diagnostic accuracy with 95% sensitivity and 100% specificity, respectively, while in a comparison between Enterobacteriaceae, Bacteriodaceae, Veillonellaceae, Lachnospiraceae, Clostrodiaceae, Coriobacteriaceae, and T2DM, the AUC value was higher than 0.500 with excellent diagnostic accuracy (Table 3). The ROC curves for family Porhyromonadaceae molecular counts employed for predicting T2DM were the best (Figs. 3a and 3b).

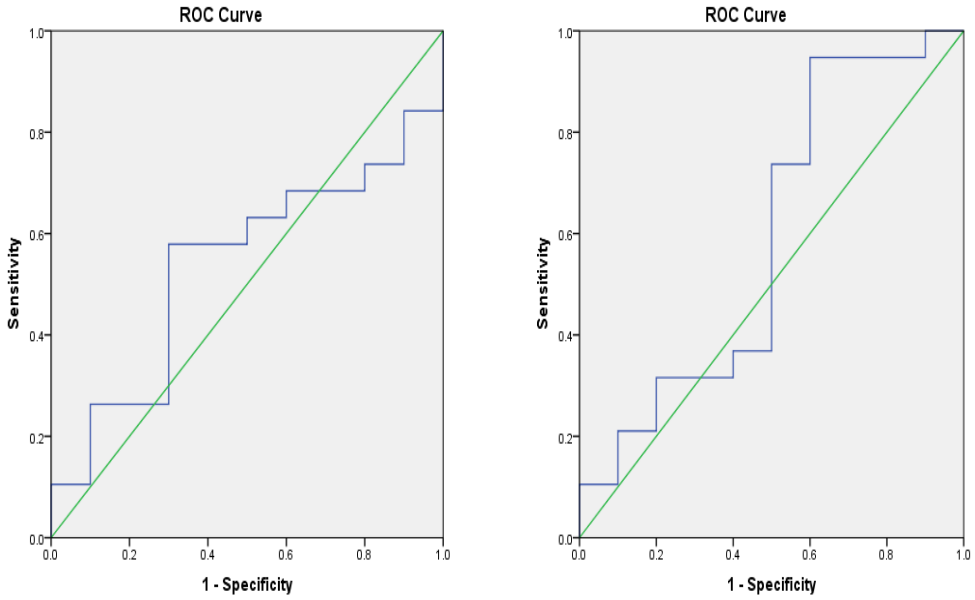
Table 3. Sensitivity, Specificity, and predictive values of the gut microbiome and family for the prediction of T2DM.

Family (Cut off point) (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	95% CL	AUC
Paraprevotellaceae (0.0001870)	84	90	28.82	99.15	97.6-99.7	0.608
Enterobacteriaceae (0.0589970)	100	80	19.43	100	6.5-45.4	0.579
Bacteriodaceae (4.4156480)	95	100	100	99.76	98.3-100	0.505
Porhyromonadaceae (0.1906)	95	100	100	99.76	98.3-100	0.742
Veillonellaceae (0.1035)	100	90	32.53	100	6.9-75.6	0.595
Lachnospiraceae (10.2712)	90	100	100	99.52	98.2-100	0.505
Clostrodiaceae (0.6211)	84	100	100	99.23	97.8-99.7	0.537
Coriobacteriaceae (0.1497)	100	90	32.53	100	6.9-75.6	0.589



Figs. 3a. Comparative of Sensitivity and Specificity of the Family/taxa to predict T2DM, (A) Specificity and Sensitivity of the Paraprevotellaceae to predict T2DM, (B) Specificity and Sensitivity of the Enterobacteriaceae to predict T2DM, (C) Specificity and Sensitivity of the Bacteriodaceae to predict T2DM, (D) Specificity and Sensitivity of the Porhyromonadaceaeee for the Prediction of T2DM.



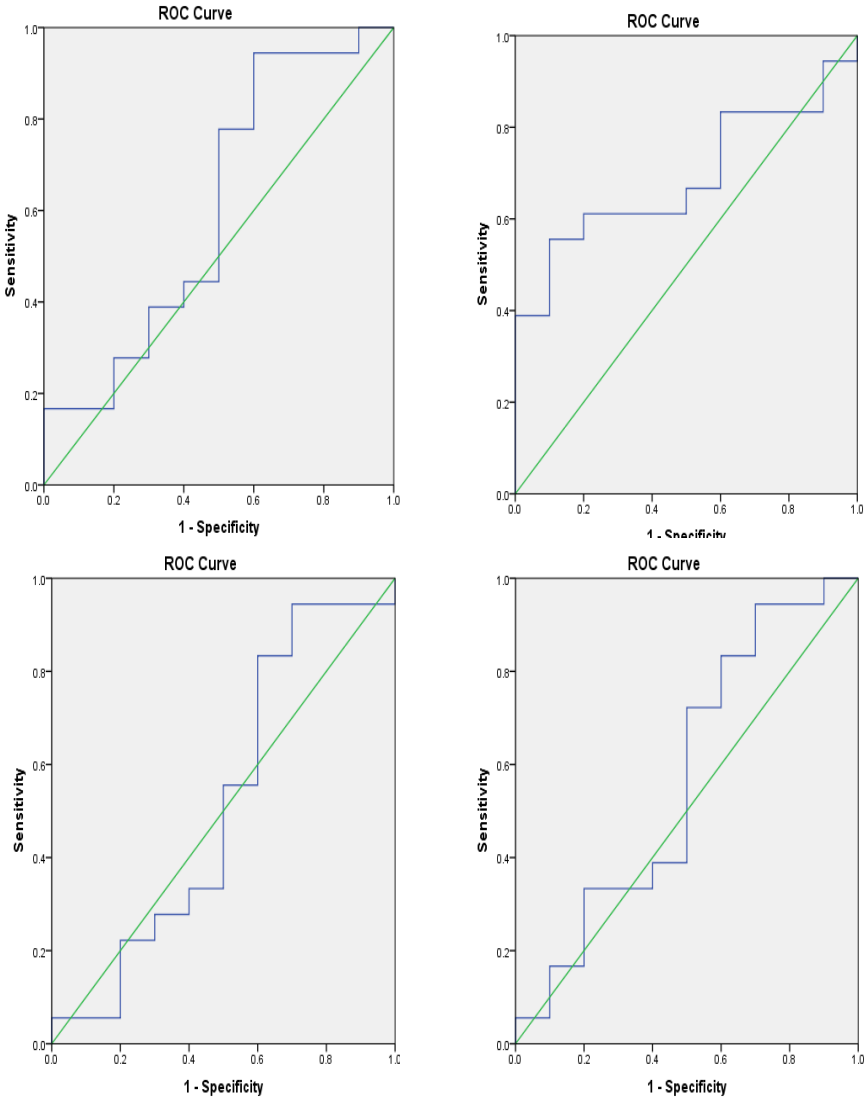


Figs. 3b. Comparative of Sensitivity and Specificity of the Family/taxa to predict T2DM, (A) Specificity and Sensitivity of the Veillonellaceae to predict T2DM, (B) Specificity and Sensitivity of the Lachnospiraceae to predict T2DM, (C) Specificity and Sensitivity of the Clostrodiaceae to predict T2DM, (D) Specificity and Sensitivity of the Coriobacteriaceae for the Prediction of T2DM.

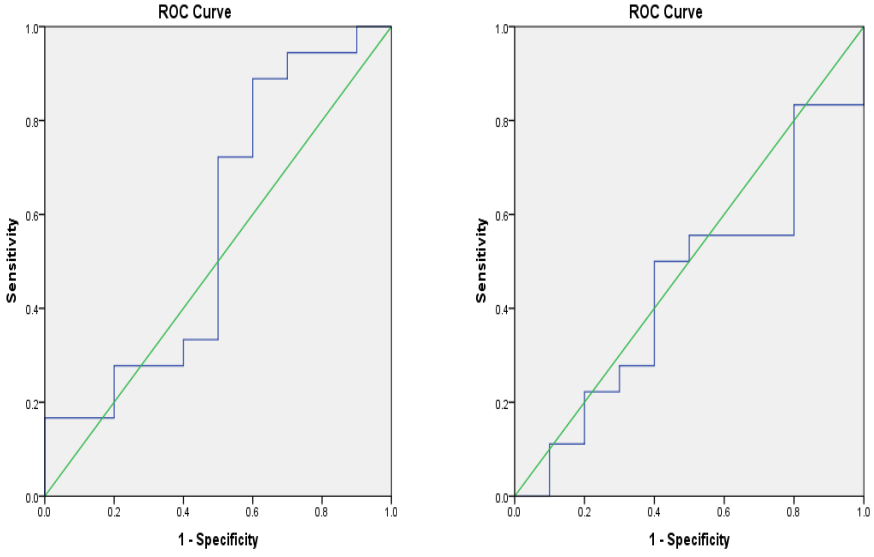
The ROC analysis has been implemented for evaluating the predictive power of the gut microbiome genus for T2DM. If AUC is closer to 1.0, the bacterial counts predict T2DM better. When a comparison was made between genus Escherichia, Streptococcus, Parabacteriode, and T2DM, the AUC were 0.606, 0.689, and 0.694 respectively (P<0.01), which showed that genus Escherichia, Streptococcus, Parabacteriodes had more predictive power for T2DM than another genus together with the best diagnostic accuracy. In contrast, in comparing Ruminococcus, Bacteriodes, Serratia, and T2DM, the AUC value was higher than 0.500 with good diagnostic accuracy. Also, the genus Bifidobacteria and Coprococcus recorded poor predictive power for T2DM (Table 4). The ROC curves for genus Escherichia, Streptococcus, Parabacteriodes molecular counts used to predict T2DM were excellent (Figs. 4a and 4b).

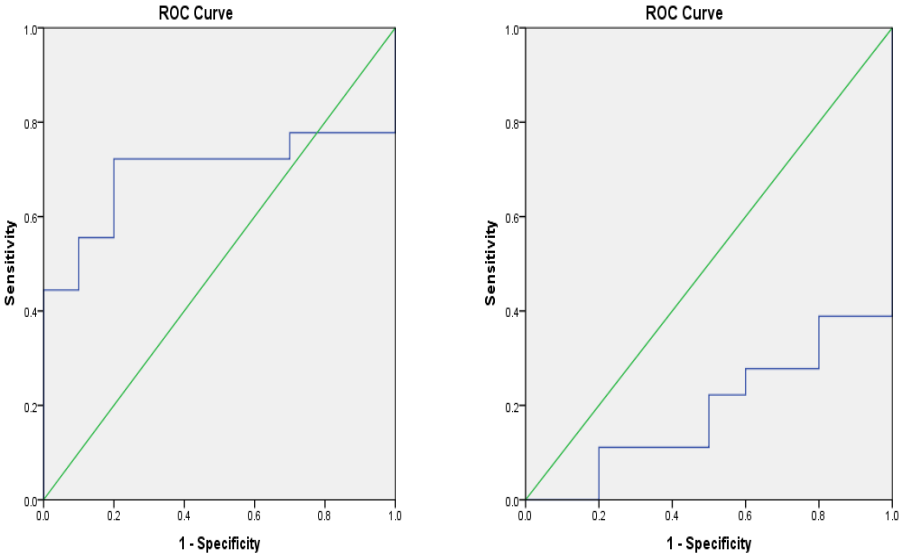
Table 4. Sensitivity, Specificity, and predictive values of the gut microbiome and genus for the prediction of T2DM.

Genus (Cut off point) (%)	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)	95% CL	AUC
Escherichia (0.0026925)	100	90	23.62	100	4.6-66.5	0.606
Streptococcus (0.0093065)	94	100	100	99.81	98.9-100	0.689
Bacteroides (4.4133905)	94	100	100	99.81	98.9-100	0.517
Ruminococcus (3.1916040)	100	90	23.62	100	4.6-66.5	0.572
Serratia (0.0279895)	94	70	8.83	99.74	98.3-100	0.570
Coprococcus (0.9270735)	83	80	11.38	99.35	98.1-99.8	0.444
Parabacteriodes (0.3646430)	78	70	7.44	99.04	97.5-99.6	0.694
Bifidobacteria (0.0670000)	44	100	100	98.3	97.5-98.9	0.189



Figs. 4a. Comparative of Sensitivity and Specificity of the genus/taxa to predict T2DM, (A) Specificity and Sensitivity of the Escherichia to predict T2DM, (B) Specificity and Sensitivity of the Streptococcus to predict T2DM, (C) Specificity and Sensitivity of the Bacteroides to predict T2DM, (D) Specificity and Sensitivity of the Ruminococcus for the Prediction of T2DM.





Figs. 4b. Comparative of Sensitivity and Specificity of the genus/taxa for to predict T2DM, (A) Specificity and Sensitivity of the Serratia to predict T2DM, (B) Specificity and Sensitivity of the Coprococcus to predict T2DM, (C) Specificity and Sensitivity of the Parabacteriodes to predict T2DM, (D) Specificity and Sensitivity of the Bifidobacteria for the Prediction of T2DM.

4. Discussion

Understanding the stability of the microbiome within an individual over time is an important step to predict diseases and develop therapies to correct microbial community imbalance. These findings have supplemented the South-East Nigerian study and African gut microbiome research related to T2DM. However, We evaluated the predictive power of the possible microbial communities associated with T2 diabetes. In this present study, there was also a strong association between the family Streptococcaceae, Sphingobacteriaceae, Alcaligenaceae, Paraprevotellaceae, and Enterobacteriaceae with T2D, followed by genus Serratia, Escherichia, and Streptococcus, respectively. Thus, the increase in gut microbiota-related T2D score of this Principal component1 (PC1) would enhance the individuals' probability been characterized as T2D. For instance, genus-Faecalibacterium-within column Comp1 and Comp 4 demonstrated that Faecalibacterium was negatively correlated with T2DM Faecalibacterium-within column Comp 2 for T2D and Comp1 for Healthy controls showed that Faecalibacterium was positively correlated. This observation suggests that higher Faecalibacterium abundance would increase the Comp 2 value and reduce the likelihood of T2D. This finding corroborates a previous study in obese and healthy control associated with Faecalibacterium, whereby a principal component is often negatively correlated.[4] A similar study reported that xylan alpha-glucuronosyltransferase involved in the metabolism of non-digestible fiber was switched off in T2DM patients.[14] It might be contributing to a decrease in Faecalibacterium, Roseburia, and Phascolarctobacterium in T2D patients. Such microbiota has been reported to improve diabetic control and insulin sensitivity.[16] This study showed the validity of these chosen gut microbial markers' discriminatory power ranging from Family /taxa to Genus/ taxa. When applying a ROC and predictive model, the study showed that at the family-levels, such as those of the Paraprevotellaceae, Enterobacteriaceae, Bacteriodaceae, Porhyromonadaceae, Veillonellaceae, Lachnospiraceae, Clostrodiaceae, Coriobacteriaceae demonstrated the best sensitivity and specificity with excellent diagnostic accuracy for T2DM followed by the genus levels, like; Escherichia, Streptococcus, Bacteroides Ruminococcus, Serratia, Parabacteriodes which recorded the high rate of sensitivity and specificity with better predictive values. This finding supports the previous study done by Shiheng et al.,[2] who revealed that predictive models might be employed for discriminating T2DM people from healthy ones. This finding is consistent with a previous study done by Larsen et al.,[15] showing that the abundance of Escherichia coli in Denmark type-2 diabetic patients (belonging to Phyla: Proteobacteria; Class: Gammaproteobacteria, Order: Entero-bacterial, Family: Enterobacteriaceae, Genus: Escherichia) and supported by Karlsson et al.,[3] study that reported the same increase in Chinese type-2 diabetic patients.

5. Conclusion

The study demonstrated for the first time in the south-East of Nigeria in diabetic patients that showed a higher abundance of Paraprevotellaceae, Enterobacteriaceae (Escherichia;-Serratia), Bacteriodaceae(Bacteroides), Porhyromonadaceae. Parabacteriodes,Veillonellaceae, Lachnospiraceae, Clostrodiaceae followed by Streptococcus, and Ruminococcus could be used as biomarkers for the prognosis of T2D. The study validity has been verified by the discriminatory power of gut microbiome-based T2D classifiers to identify the increased risks for T2D. These findings revealed that changes in the gut microbiome, especially at the family level followed by genus level, might be employed for identifying people at the increased risks for T2D. The findings also revealed that the cluster of gut microbiota correlated positively in Comp1 for T2DM and are a potential risk factor associated with T2D. On the other hand, those gut microbiota, especially Faecalibacterium, Roseburia, and Phascolarctobacterium, correlated negatively in Comp1 for T2D, may provide health benefits for T2D.

Conflict of Interest

The authors declared that there is no conflict of interest.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

[1] Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490(7418):55-60. https://doi.org/10.1038/nature11450.

[2] Li Q, Chang Y, Zhang K, Chen H, Tao S, Zhang Z. Implication of the gut microbiome composition of type 2 diabetic patients from northern China. Scientific reports. 2020;10(1):1-8. https://doi.org/10.1038/s41598-020-62224-3.

[3] Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498(7452):99-103. https://doi.org/10.1038/nature12198.

[4] Zhang P, West NP, Chen PY, Thang MW, Price G, Cripps AW, et al. Selection of microbial biomarkers with genetic algorithm and principal component analysis. BMC bioinformatics. 2019;20(6):1-8. https://doi.org/10.1186/s12859-019-3001-4.

[5] Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. Multivariate data analysis (4th Ed.). Upper Saddle River, N J: Prentice Hall. 1995.

[6] Mardia KV, Kent JM, Bibby JM. Multivariate analysis. London: Academic press. 521.1980.

[7] Newby PK, Muller D, Hallfrisch J, Andres R, Tucker KL. Food patterns measured by factor analysis and anthropometric changes in adults. The American journal of clinical nutrition. 2004;80(2):504-13. https://doi.org/10.1093/ajcn/80.2.504.

[8] Harsch IA, Konturek PC. The role of gut microbiota in obesity and type 2 and type 1 diabetes mellitus: new insights into “old” diseases. Medical sciences. 2018;6(2):32.

[9] Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the national academy of sciences. 2011;108(Supplement 1):4516-22.

[10] Ringner M(2008).What is principal component analysis?*Natural Biotechnology*. 26(3):303–304. DOI: 10.1038/nbt0308-303.

[11] Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374(2065):20150202. https://doi.org/10.1098/rsta.2015.0202.

[12] Huang T, Li J, Zhang W. Application of principal component analysis and logistic regression model in lupus nephritis patients with clinical hypothyroidism. BMC medical research methodology. 2020;20:1-7. https://doi.org/10.1186/s12874-020-00989-x.

[13] Ling Z, Liu X, Luo Y, Wu X, Yuan L, Tong X, et al. Associations between vaginal pathogenic community and bacterial vaginosis in Chinese reproductive-age women. PLoS One. 2013;8(10):e76589.

[14] Oshim IO, Agbakoba NR, Anukam KC. Gut Microbiota Compositions and Modulation of Bacterial Metabolic Functional Genes in Type-2 Diabetes Mellitus Individuals at Nnewi, Anambra State. J Med Lab Sci. 2020;30(3):136-50. http://doi.org/10.5281/zenodo.4050381.

[15] Larsen N, Vogensen FK, Van Den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PloS one. 2010;5(2):e9085.

[16] Hur KY, Lee MS. Gut microbiota and metabolic disorders. Diabetes & metabolism journal. 2015;39(3):198. http://dx.doi.org/10.4093/dmj.2015.39.3.198.

How to Cite this Article: Oshim IO, Agbakoba NR, Oguejiofor OC, Anukam KC. Selective Microbial Biomarkers in Type-2 Diabetes with Principal Component Analysis and Receiver-operating Characteristic Curves. International Journal of Scientific Research in Dental and Medical Sciences, 2021;3(1):23-34. doi:10.30485/IJSRDMS.2021.272435.1110.