Journal of Genetics and Genomics 48 (2021) 32–39

Contents lists available at ScienceDirect

Journal of Genetics and Genomics

Journal homepage: www.journals.elsevier.com/journal-of-genetics-and-genomics/

Original research

The oral microbiome of pregnant women facilitates gestational diabetes discrimination

Xiaoqing Li [a, 1], Jiayong Zheng [a, 1], Xiuling Ma [b, c], Bing Zhang [b, c], Jinyang Zhang [b, c], Wenhuan Wang [a], Congcong Sun [a], Yeping Wang [a], Jianqiong Zheng [a], Haiying Chen [a], Jiejing Tao [a], Hai Wang [a], Fengyi Zhang [b, c], Jinfeng Wang [b, c, *], Hongping Zhang [a, *]

[a] Wenzhou People's Hospital/Wenzhou Maternal and Child Health Care Hospital/The Third Clinical Institute Affiliated to Wenzhou Medical University, Wenzhou 325000, China
[b] Computational Genomics Lab, Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China
[c] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 19 November 2020
Received in revised form
25 November 2020
Accepted 26 November 2020
Available online 20 January 2021

Keywords:
Gestational diabetes mellitus
Oral microbiome
Saliva
Dental plaque
Classification model

ABSTRACT

The oral microbiota plays an important role in the development of various diseases, whereas its association with gestational diabetes mellitus (GDM) remains largely unclear. The aim of this study is to identify biomarkers from the oral microbiota of GDM patients by analyzing the microbiome of the saliva and dental plaque samples of 111 pregnant women. We find that the microbiota of both types of oral samples in GDM patients exhibits differences and significantly varies from that of patients with periodontitis or dental caries. Using bacterial biomarkers from the oral microbiota, GDM classification models based on support vector machine and random forest algorithms are constructed. The area under curve (AUC) value of the classification model constructed by combination of *Lautropia* and *Neisseria* in dental plaque and *Streptococcus* in saliva reaches 0.83, and the value achieves a maximum value of 0.89 by adding clinical features. These findings suggest that certain bacteria in either saliva or dental plaque can effectively distinguish women with GDM from healthy pregnant women, which provides evidence of oral microbiome as an informative source for developing noninvasive biomarkers of GDM.

Introduction

Gestational diabetes mellitus (GDM) is defined as varying degrees of glucose intolerance first found or occurring during pregnancy (Weinert, 2010). GDM is one of the most common maternal complications in middle and late pregnancy. Approximately 5.2–8.8% of pregnant women worldwide suffer from GDM each year (Cheung and Byth, 2003). In some countries and regions, the incidence of this disease is greater than 20%, and the incidence is increasing annually (Damm and Mathiesen, 2015). GDM increases the risk of long-term complications, including obesity, impaired glucose metabolism, and cardiovascular disease, in both mothers and infants (Buchanan et al., 2012; Damm et al., 2016).

The symbiotic microbiota is an important part of the human body (Kinross et al., 2011; Lloyd-Price et al., 2016). Although some studies have found that GDM does not obviously change the microbial community (Hasan et al., 2018), more studies have shown that the microbiota is significantly different between GDM patients and normal pregnant women (Acuna et al., 2011; Crusell et al., 2018). In the placental microbiome of GDM patients, the proportion of Proteobacteria is increased, while Bacteroidetes and Firmicutes are decreased (Zheng et al., 2017). In a recent study, there were also significant differences in the oral microbiota between GDM patients and normal pregnant women in the third trimester of pregnancy and even nine months after delivery (Crusell et al., 2020).

Using microbes as biomarkers for disease prediction has become a promising strategy (Wang et al., 2019). Several studies have found that using bacteria for disease diagnosis has great potential (Martinez et al., 2017). For example, using intestinal bacteria to develop biomarkers of systemic diseases, 16 kinds of bacteria were screened from the intestinal microbiota of colorectal cancer patients that could

* Corresponding authors.
E-mail addresses: wangjf@biols.ac.cn (J. Wang), zjzhp@126.com (H. Zhang).
[1] These authors contribute equally to this work.

https://doi.org/10.1016/j.jgg.2020.11.006

accurately distinguish colorectal cancer patients from normal people with an accuracy of 84% (Thomas et al., 2019). Coincidentally, the accuracy of 15 microbial-related gene markers in the diagnosis of liver cancer can reach 83.6% (Qin et al., 2014). In the exploration of using oral microbes as markers, a dental caries prediction model was established according to the dynamic changes in oral microbes during the occurrence and development of diseases (Teng et al., 2015). Another study using oral microbes to predict periodontitis found that oral microbial prediction could distinguish 95% of healthy people from patients, showing a good ability to predict and diagnose diseases (Huang et al., 2014). Additionally, some studies analyzed the relationship between the oral microbiota and pancreatic diseases and constructed a prediction model for pancreatic cancer using specific microbial species (Farrell et al., 2012).

In our previous study, by comparing the microbial composition of the vagina, intestinal tract, and oral cavity between normal pregnant women and GDM patients, we found that the microbiota in GDM patients was distinct (Wang et al., 2018). In particular, the maximum change occurred in the saliva, which reflects the feasibility of selecting oral microbes as biomarkers for GDM detection. However, many studies have shown that there is some relationship between GDM and periodontitis (Belstrom et al., 2016; Graziani et al., 2018). It has been found that the incidence of GDM is increased in patients with periodontitis (Belstrom et al., 2016). In addition, GDM and periodontitis have the characteristics of accumulative dental plaque, decreased red blood cell count, and increased inflammation (Seraphim et al., 2016). Periodontal infection may increase the risk of GDM by affecting endocrine metabolism and blood glucose control (Gumus et al., 2015), but whether the links between these two diseases are related to microorganisms is unknown. It is also not clear whether the microbial changes caused by oral diseases or other factors (e.g., diet or use of antibiotics) will affect the accuracy of the disease categorization model.

This study intends to analyze the oral microbiome data of saliva and dental plaque from GDM patients and normal pregnant women and explore the possible relationships between GDM and two major oral diseases, dental caries and chronic periodontitis. On this basis, the study will identify suitable microbial markers from the oral microbiota to construct GDM classification models, in addition to developing a simple and noninvasive technique for auxiliary diagnosis and daily follow-up of GDM.

Results

Changes in the oral microbiome in patients with GDM

We enrolled 111 pregnant women with good oral health, including two groups of 44 pregnant women with GDM (GDM+) and 67 without GDM (GDM−). The general information, medications, disease history, dietary habits, and biochemical indexes are shown in supplementary material (Fig. S1). The clinical backgrounds of the two groups were roughly similar, except for blood glucose levels. In total, 105 saliva and 51 dental plaque samples were acquired, of which 45 saliva and dental plaque samples were paired, with each pair collected from the same individual (Fig. S2A). For each sample, the V3–V4 regions of the 16S rRNA gene were sequenced. Gene sequencing totally yielded ~16 million PE reads (2 × 250 bp), with an average of ~108,305 reads per sample (Fig. S2B). Each pair of reads was merged into one sequence by overlaps. Most of the sequences were 400–450 bp in length (Fig. S2C). According to the rarefaction curve (Fig. S3A) and Good's coverage (Fig. S3B), the number of sequences can well represent the microbial diversity of each community.

To investigate whether hyperglycemia that develops during pregnancy is accompanied by extensive changes in the oral microbiota, we explored the microbial communities of saliva and dental plaque of pregnant women who were diagnosed with GDM. We found that both saliva and dental plaque samples of the GDM+ group were divided into different clusters from those of the GDM− group (Fig. 1A), although there was no significant difference in α-diversity (Fig. S4A–S4D). In saliva, the Bray-Curtis (BC) intragroup distances of the GDM+ group were significantly smaller than both the intragroup GDM− and intergroup GDM+ *vs.* GDM− distances ($P < 0.001$, Mann-Whitney test). In dental plaque, the differences in BC distances were not as obvious as those in saliva (Fig. 1B). These results suggest that pregnant women with GDM have a distinct oral microbial community that is different from that of healthy women, and microbial variations in the oral cavity in GDM+ women showed obvious sample-type specificity.

Difference in the oral microbiome between patients with GDM and major oral diseases

To investigate the possible relationships between GDM and periodontitis, we compared the GDM-associated microbiome identified in our study with a downloaded dataset of the periodontal microbiome dataset. Another major oral disease, dental caries, served as a control.

The BC distances of the oral microbiome between patients with GDM and subjects with periodontal health (PH) or disease (PD) were first calculated. The community distances between subjects with PH and GDM+ or GDM− were significantly smaller than those of PD ($P < 0.0001$, Mann-Whitney test), regardless of the use of saliva or dental plaque (Fig. 2A and 2B). However, when bacterial taxa in the oral microbiome were compared among PH, PD, GDM+ and GDM− groups, we did not find that patients with PD shared more bacterial genera with pregnant women than individuals with PH, regardless of whether the women had GDM (Fig. 2C and 2D). There was also no significant difference in the number of shared bacteria in the oral cavity when pregnant women with GDM or without GDM were compared with subjects with no caries (NC), mild caries (LC), moderate caries (MC), and severe caries (HC) (Fig. S5A–S5D). The saliva and dental plaque microbiota of both the GDM+ and GDM− groups showed larger BC distances to the dental caries group than to the NC group (Fig. S5E). These results indicate that the oral microbiome of GDM patients was more similar to that of individuals with healthy periodontal status but different from patients with periodontitis in regard to the community structure. The microbial variations in the oral cavities of pregnant women with GDM may not be equivalent to those of patients with periodontitis, and there should be little relationship in the oral microbial shifts between the GDM and dental caries groups.

To further explore the relationships between GDM and oral diseases as well as to evaluate whether bacteria-based GDM detection would be affected by the microbial changes caused by major oral diseases, we performed odds ratio analysis and identified the differential genera from pregnant women with and without GDM. Significant differences in genera such as *Lautropia*, *Neisseria*, *Streptococcus*, and *Veillonella* were found between the GDM+ and GDM− groups in both saliva and dental plaque samples (Fig. S6A and S6B). Notably, *Streptococcus* and *Veillonella* were also depleted in patients with periodontitis (Fig. S6C), indicating that the possible relationship between GDM and periodontitis may be related to the decreased abundance of these two genera. There was no significant variation in these four bacteria in individuals with dental caries (Fig. S6D–S6F), indicating that there was little relationship between GDM and dental caries in regard to the change in the microbial community.

Classification model of GDM based on the SVM algorithm

To identify specific microbial biomarkers that can be used to discriminate GDM, the differential genera between pregnant women

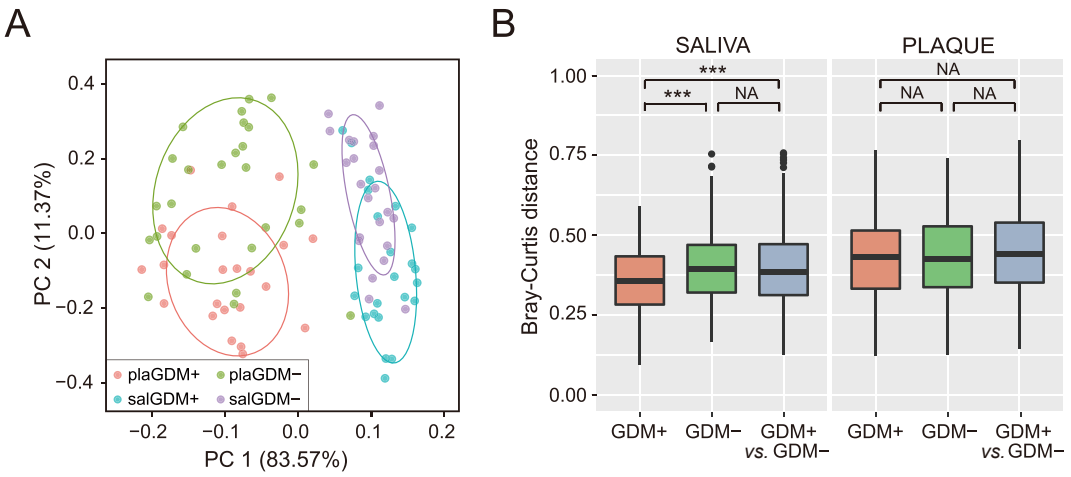

Fig. 1. Oral microbial variations in pregnant women with GDM. **A**: PCoA analysis of saliva (sal_) and dental plaque (pla_) in the GDM+ and GDM− groups. **B**: Bacterial community dissimilarities between the saliva and plaque samples of the GDM+ and GDM− groups. Bray-Curtis distances were independently calculated for intragroup GDM+ and GDM− and intergroup GDM + *vs.* GDM−. Statistical significance was determined by the Mann-Whitney test. ***, $P < 0.001$.

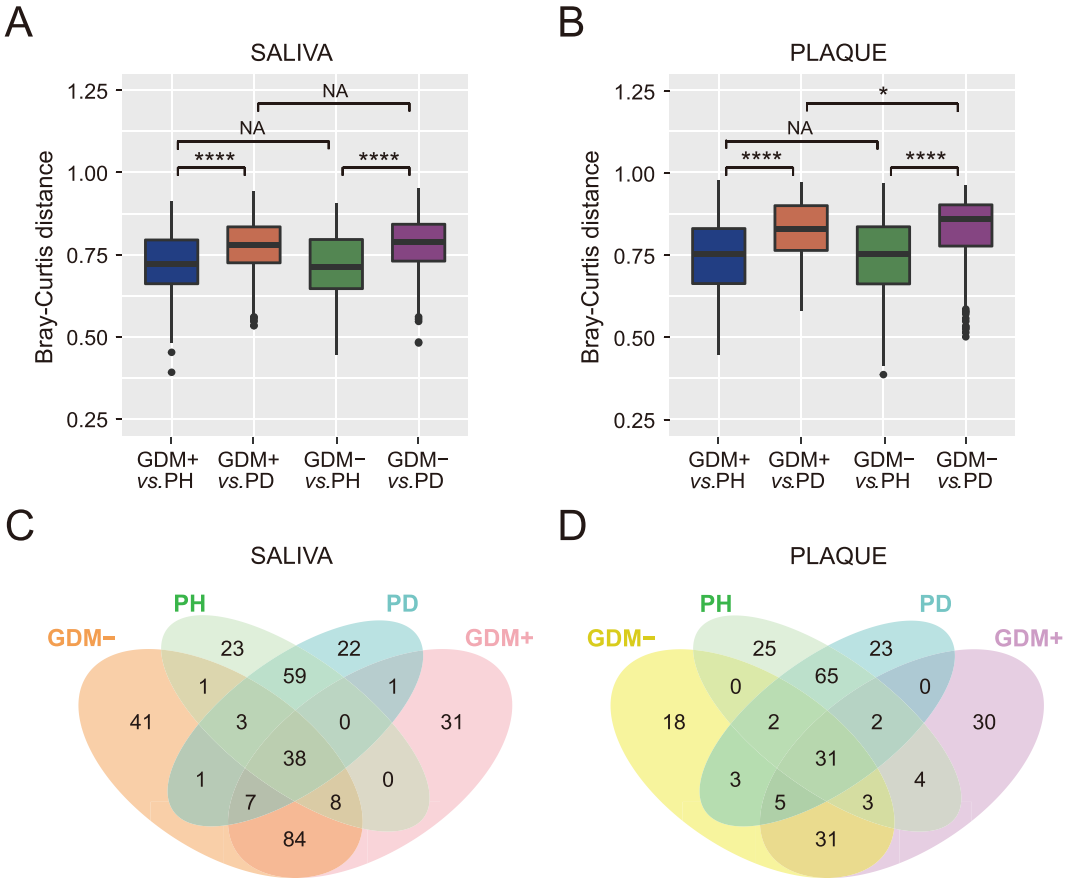

Fig. 2. Comparison of the oral microbiota between periodontitis and GDM. **A** and **B**: Bray-Curtis distances of the oral microbiota between the periodontitis, periodontal health, GDM+ and GDM− groups in saliva (**A**) or in plaque (**B**). Statistical significance is determined by the Mann-Whitney test. *, $P < 0.05$; ****, $P < 0.0001$. **C** and **D**: The number of shared bacterial genera between the periodontitis, periodontal health, GDM+ and GDM− groups in saliva (**C**) or in dental plaque (**D**).

with GDM and those without GDM were further investigated. We compared the two groups by linear discriminant analysis effect size (LEfSe) analysis, with the threshold value of LDA 3.0 (Fig. S7A and S7B). For saliva, Leptotrichiaceae, *Lautropia*, *Neisseria*, Neisseriales, and four other bacterial taxa were significantly enriched in the GDM+ group, while *Selenomonas*, *Leptotrichia*, F16, and three other taxa were depleted (Fig. 3A). Regarding plaque, significant enrichment was shown in the abundance of *Lautropia*, *Neisseria*, and Neisseriales, while the microbiota was depleted of bacteria such as *Streptococcus* and *Veillonella* in the GDM+ group (Fig. 3B). *Lautropia* and *Neisseria* were the common characteristic bacteria in both saliva and dental plaque.

According to the results of both odds ratios and LEfSe (Figs. S6 and S7), we found that the abundances of the genera *Lautropia*, *Neisseria*, *Streptococcus*, and *Veillonella* were significantly different between the two sample types in the GDM+ and GDM− groups, so these genera were used to construct classification models based on the support vector machine (SVM) algorithm. First, to find the optimal combination of microbial biomarkers in addition to optimizing the efficiency of GDM identification, we performed orthogonal experiments using paired samples of saliva and dental plaque (Fig. 4A). Using *Lautropia* and *Neisseria* of dental plaque and *Streptococcus* of saliva microbiota as bacterial features, the model achieved the optimal area under curve (AUC) value of 0.84 (95% confidence interval [CI]: 0.81–0.87). In another two combinations, the AUC value using *Streptococcus* and *Veillonella* for the two sample types was 0.78 (95% CI: 0.75–0.81), while the value using only *Streptococcus* for the two sample types also reached 0.75 (95% CI: 0.71–0.78). Subsequently, by performing 1,000 iterations (Fig. 4B), the AUC value of the combination of *Lautropia* and *Neisseria* and saliva *Streptococcus* from dental plaque was as high as 0.83 (95% CI: 0.82–0.84). Even if only *Streptococcus* was used in the two sample types, an AUC value of 0.74 (95% CI: 0.73–0.75) was obtained.

Then, we combined bacterial and several clinical features, such as weight gain during pregnancy, to construct models. Although there was a slight improvement (0.76) when using the single bacteria *Streptococcus*, the maximum AUC values (0.82) exhibited no obvious increase (Fig. 4C). Considering that saliva sampling is more convenient, rapid, safe, and noninvasive, we employed 105 saliva samples to develop classification models (Fig. S2A). In the SVM classification model of GDM, the AUC of *Streptococcus* and *Veillonella* was 0.76 (95% CI: 0.75–0.76) (Fig. S8A and S8B).

GDM classification using the RF algorithm or only salivary bacteria

To provide more choices, a classifier based on the random forest (RF) algorithm was constructed to distinguish GDM, first using paired dental plaque and saliva samples. The recursive feature elimination method was used to rank the importance of all the features, and the top 20 features and their abundance information are shown (Figs. 5A and S9). We then selected different features to calculate the AUC of the model (Fig. 5B). When using seven or three genera to build the model, the model had the best performance, and the AUC values were 0.82 (95% CI: 0.74–0.90) and 0.81 (95% CI: 0.71–0.91), respectively (Fig. 5B and 5C). And, we found that the clinical feature of weight gain during pregnancy had the highest importance as the features of GDM classification combined with bacteria (Fig. 5B), the AUC reached 0.89 (95% CI: 0.83–0.95) (Fig. 5D). Furthermore, we found that using *Atopobium*, *Veillonella*, *Bulleidia*, *Streptococcus*, *Kingella*, and *Lautropia* to build the model resulted in the best performance (Fig. S8C), and the AUC was 0.79 (95% CI: 0.74–0.84) (Fig. S8D).

Discussion

In this study, the microbiome of saliva and dental plaque was used to analyze the oral microbiota associated with GDM and to screen the microbiological markers that could potentially distinguish GDM patients from healthy pregnant women. This is an attempt to construct a classification model for GDM discrimination using microbes from saliva and dental plaque as biomarkers. This is also the first study to reveal the relationship between GDM and major oral diseases, including periodontitis and dental caries, by comparing the microbial shift and to evaluate the potential impact of oral diseases on using oral microbes as diagnostic markers.

The four bacteria with the most variation in the oral microbiota of pregnant women with GDM were identified. Among these bacteria, *Streptococcus* was positively correlated with *Actinomyces* (Kolenbrander et al., 2002), while the latter participates in the Embden-Meyerhof-Parnas pathway in which glucose is degraded into pyruvate and further degraded to lactate, formate, and acetate (Takahashi and Yamada, 1999). *Veillonella* could use lactic acid as a carbon source and energy source (Ng and Hamilton, 1971), as well as to regulate pH to promote the proliferation of *Streptococcus* (Dzunkova et al., 2018; Kim et al., 2018). In addition, the other two bacteria, *Lautropia* and *Neisseria*, may be related to the synthesis of bacterial motion proteins, linoleic acid metabolism, and flavonoids (Dzunkova et al., 2018; Kim et al., 2018). *Streptococcus*, *Veillonella*, and *Neisseria* in dental plaque and saliva are positively correlated with glycolysis, fructose metabolism, and alanine metabolism but negatively correlated with arginine metabolism (Koopman et al., 2015). This knowledge suggests that these four bacteria have complex interactions and are closely related to glucose metabolism, so they may be used to indicate the occurrence and development of GDM as a metabolic disease.

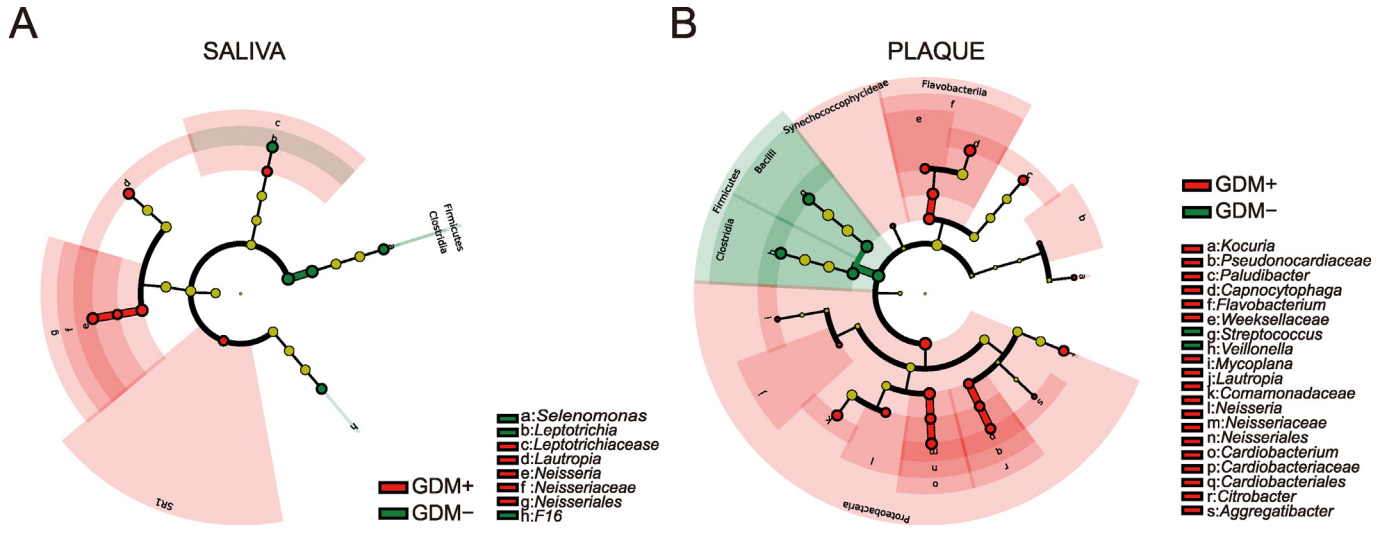

Fig. 3. Bacterial features most likely explain differences between GDM+ and GDM−. Cladogram of bacterial biomarkers down to the genus level identified by LDA using LEfSe in saliva (**A**) and dental plaque (**B**). Color indicates the group in which a differentially abundant taxon is enriched (red: GDM+, green: GDM−).

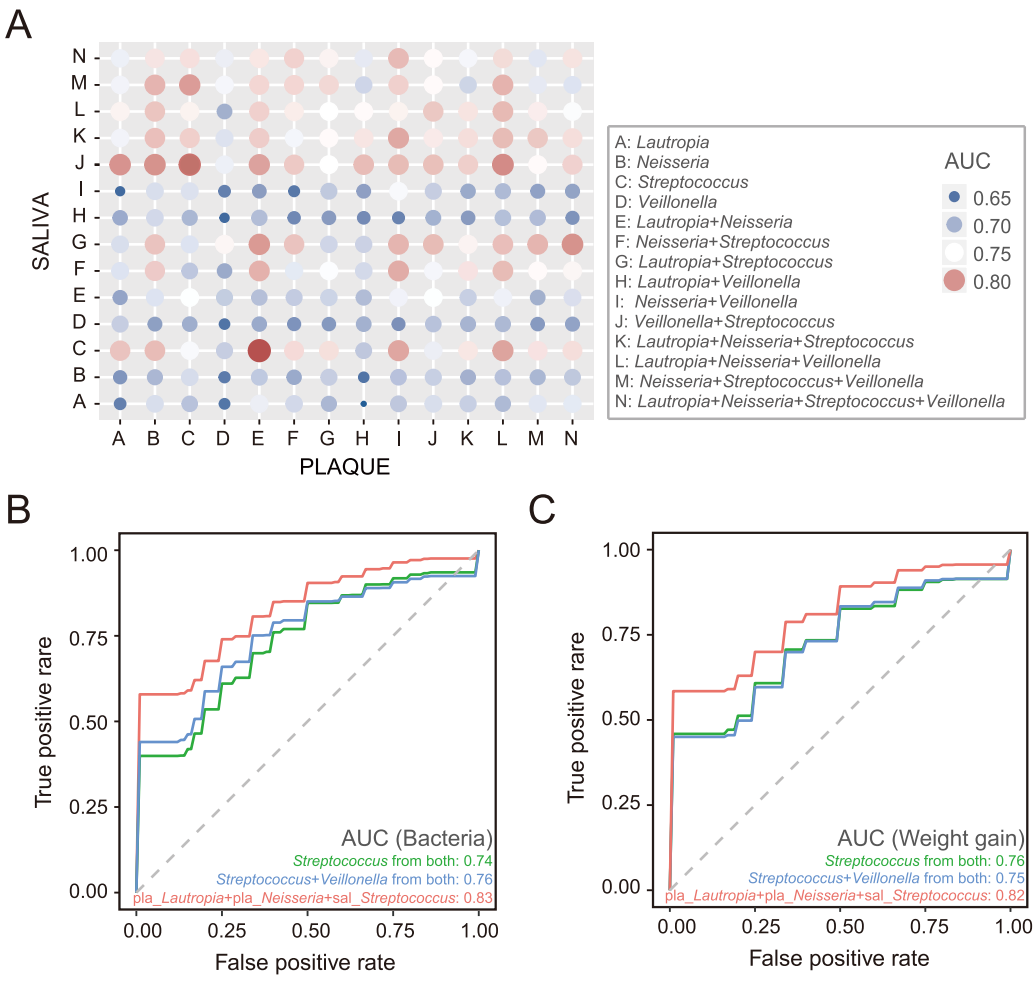

Fig. 4. SVM classification model of GDM based on oral microbes. **A**: The orthogonal results of four genera in saliva and plaque ($n = 45$). **B**: The ROC curve of SVM classification models using the genus abundance in paired saliva and dental plaque samples ($n = 45$). **C**: The ROC curve of SVM classification models using oral bacteria and weight gain during pregnancy. pla_ represents bacteria from dental plaque, and sal_ represents bacteria from saliva.

Previous studies have shown that the symbiotic microbiota in the human body is a global hot topic in the contexts of promoting the progress of disease diagnosis, assisting disease treatment, and developing personalized medicines (Farrell et al., 2012; Teng et al., 2015; Thomas et al., 2019). A model for predicting periodontitis was constructed using oral bacteria such as *Lautropia*, *Streptococcus*, *Selenomonas*, *Peptostreptococcus*, *Oribacterium*, and Veillonellaceae (Huang et al., 2014). In a predictive model for caries, using only eight marker *Prevotella* species could predict caries with an AUC of up to 0.74, while 20 bacteria, including *Streptococcus*, *Veillonella* and *Prevotella,* predict caries with an AUC of up to 0.77 (Teng et al., 2015). Obviously, the markers mentioned in the contexts of the above-mentioned oral diseases were not exactly identical to the microbes used in this study for GDM classification, which ensures the specificity of our model. However, it should be noted that many studies have shown a link between GDM and periodontitis (Belstrom et al., 2016; Graziani et al., 2018), and partial overlaps in their microbial markers (viz. *Streptococcus* and *Veillonella*) were also found in this study. This reminds us to pay special attention to the oral status of the subjects and to choose a bacterial combination that is not disturbed by periodontitis as much as possible when developing a similar method for GDM testing in the future.

Scholars around the world have been trying to develop various noninvasive methods for GDM detection (Nassr et al., 2018). A recent study used a machine learning approach to predict GDM using retrospective data from 588,622 pregnancies in Israel (Artzi et al., 2020). These researchers devised a model based on nine questions that a patient could answer, which may identify low-risk women and avoid glucose tolerance tests. Similarly, we used bacterial features from the oral microbiota to build GDM prediction models based on both SVM and RF algorithms. The performance of RF and SVM was similar when using bacterial features only, but the RF model performed better when clinical features were added. Perhaps, this method currently costs more than the oral glucose tolerance test (OGTT), but it is easier for pregnant women to provide saliva and plaque samples than to participate in an OGTT. Oral samples can be collected by doctors, nurses, or even pregnant women themselves under guidance. It is very encouraging to see that a few bacteria could achieve a good discrimination effect in our study, which contributed to the design of specific bacterial probes for GDM detection. Pregnant women can even test by themselves at home if it

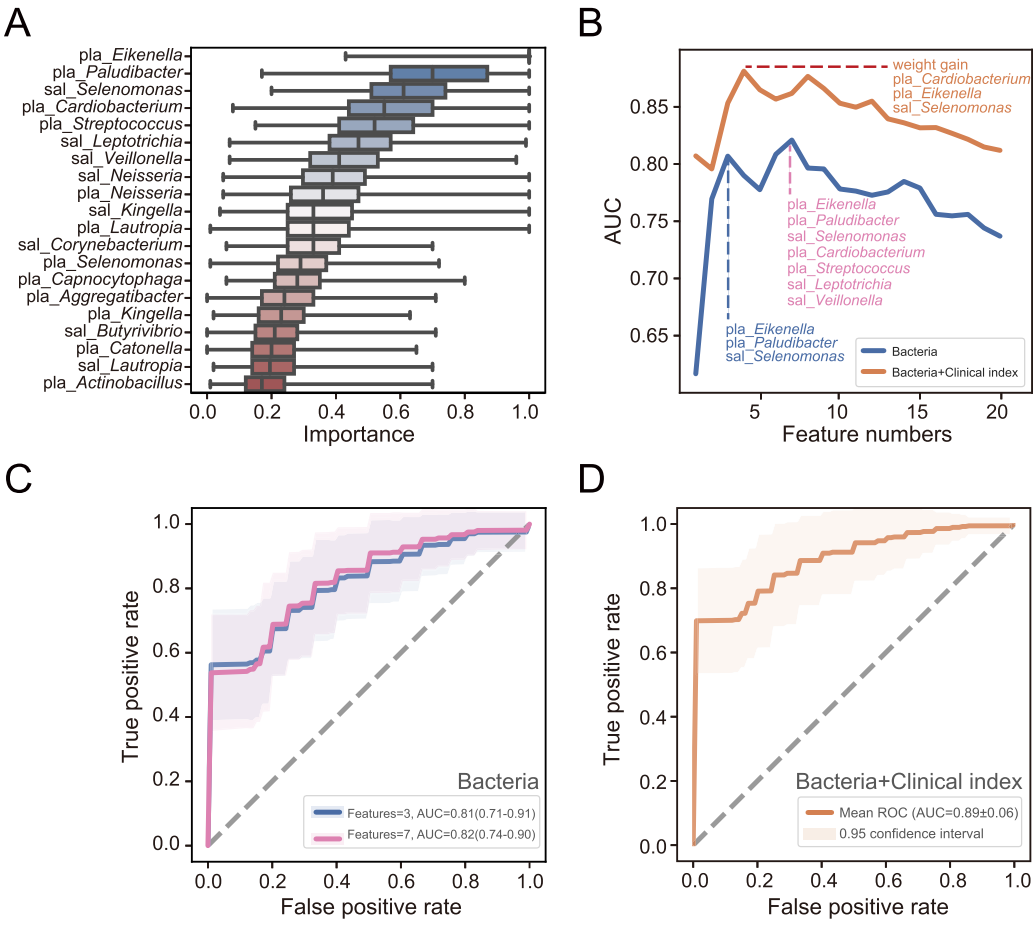

Fig. 5. RF classification model for GDM based on oral microbes. **A**: Top 20 bacteria with the highest importance for GDM classification ($n = 45$). **B**: The AUC of different combinations of bacterial and clinical features. **C**: The RF classification model using bacteria from saliva and dental plaque samples ($n = 45$). **D**: The RF classification model using bacteria and weight gain during pregnancy. pla_ represents bacteria from dental plaque, and sal_ represents bacteria from saliva.

becomes a simpler solution or a rapid-screening kit becomes available.

Materials and methods

Subject recruitment

The study was approved by the Ethics Committee of Wenzhou People's Hospital. Pregnant women were recruited at Wenzhou People's Hospital. Informed consent was obtained from all participants. All the pregnant women were of Han ethnicity and permanent residents of Wenzhou. All pregnant women were non-vegetarian and had no history of smoking, alcohol consumption, or any other systemic, metabolic or oral diseases, especially periodontitis and dental caries (Fig. S1).

Based on the diagnostic criteria recommended by the International Association of the Diabetes and Pregnancy Study Groups (IADPSG) in 2011, pregnant women with GDM (GDM+) were diagnosed by specialized doctors according to the results of the OGTT and were recruited as a case group. For women without GDM (GDM−), plasma glucose concentrations were lower than the threshold values of 5.1, 10.0 and 8.5 mmol/L before and 1 h and 2 h after drinking a 75 g glucose solution, respectively. Any woman with a blood glucose level equal to or greater than the threshold values was diagnosed as GDM+.

Sample collection

Saliva and dental plaque samples were collected from third trimester pregnant women according to the sampling methods described in our previous study (Wang et al., 2013). The whole mouths of all subjects were examined, and samples were collected by professional doctors. Partial saliva and dental plaque were paired samples that were collected from the same individual. Prior to sample collection, all subjects were instructed to avoid eating and brushing their teeth for 2 h. Briefly, ~2 mL saliva was collected from each pregnant woman with a sterile tube and stored at −80°C. Dental plaque was scraped from the tooth surface, resuspended into a centrifugal tube, and stored at −80°C until total DNA extraction for later sequencing.

DNA extraction

In a strictly controlled, separate and sterile workplace, approximately 0.2 mL saliva and 50 μL PBS containing the plaque sample were mixed with Qiagen's AL buffer by pulse vortexing for 30 s (Qiagen, Valencia, CA). Total DNA was extracted from the suspension of each sample using a QIAamp DNA Mini Kit (Qiagen, Valencia, CA). Isolated DNA was eluted in 50 μL distilled water. The DNA quality and concentration of all samples were measured

by agarose gel electrophoresis and a Qubit 3.0 fluorometer (Life Technologies, Waltham, MA) before downstream processing.

High-throughput sequencing

For each sample, variable regions 3 and 4 (V3–V4) of the 16S rRNA gene were amplified using 341F and 805R primers. Purified positive amplicons with different index sequences were pooled in equimolar amounts. Amplicon length and integrity of the libraries were assessed by a Fragment Analyzer (Advanced Analytical Technologies) before paired-end (PE) sequencing (2 × 125 bp) on a HiSeq2500 platform (Illumina, San Diego, CA). The sequencing data were deposited under the accession number CRA002189 in the Genome Sequence Archive (https://bigd.big.ac.cn/gsa/).

16S rRNA sequence analysis

Raw sequencing reads of the 16S rRNA gene sequences were quality filtered and analyzed using QIIME V.1.8.0.12. The operational taxonomic units (OTUs) were classified taxonomically using the Greengenes 16S rRNA gene reference database. Two microbial classification datasets of adults (nonpregnant), including 28 patients with periodontitis (PD), 22 subjects with good periodontal health (PH), 62 patients with high caries (HC), 37 patients with middle caries (MC), 32 patients with low caries (LC), and 29 subjects with no caries (NC) were retrieved for comparison (Camelo-Castillo et al., 2015; Xiao et al., 2016).

Analysis of microbial community composition

The taxonomic composition of microbial communities was visualized using Calypso (Zakrzewski et al., 2017). Community clustering was measured by unweighted UniFrac distance based on the normalized taxonomy table. Bray-Curtis dissimilarity between different sample types was calculated using the R package ecodist. The difference in alpha diversity between groups was statistically analyzed by the Mann-Whitney test ($P < 0.05$). LEfSe and odds ratio analysis were used to identify the characteristic genera in the GDM+ and GDM− groups, and a score of log linear discriminant analysis (LDA) > 3.0 or an odds ratio with $P < 0.05$ was considered to indicate a differential signature that better discriminated between groups.

Biomarker screening and classification model construction

The genera whose abundance was ≥ 0.1% in each sample were retained, and low-abundance bacteria were excluded from the following analysis. The SVM algorithm and ROC calculation were performed by the e1071 and ROCR packages in R, respectively. Based on the genus abundance in saliva-plaque paired samples ($n = 45$) and only saliva samples ($n = 105$), cross-validation with 1,000 random permutations was executed to evaluate the performance of these models. RF models were also trained using bacterial taxonomy profiles of the oral microbiota to differentiate the disease status of GDM. The recursive feature elimination method was used to sort the importance of all bacterial and clinical features and to draw the ROC curve. Then, five-fold cross validation with 1,000 iterations was used to evaluate the performance of these models.

Data availability

The sequencing data were deposited under the accession number CRA002189 in the Genome Sequence Archive (https://bigd.big.ac.cn/gsa/).

CRediT authorship contribution statement

Xiaoqing Li: Writing - Original draft, Sample collection, Experimentation, Data analysis, Data visualization. **Jiayong Zheng**: Sample collection, Experimentation. **Xiuling Ma**: Data analysis, Data visualization. **Bing Zhang**: Data analysis, Data visualization. **Jinyang Zhang**: Data analysis, Data visualization. **Wenhuan Wang**: Sample collection, Experimentation. **Congcong Sun**: Sample collection, Experimentation. **Yeping Wang**: Sample collection, Experimentation. **Jianqiong Zheng**: Sample collection, Experimentation. **Haiying Chen**: Sample collection, Experimentation. **Jiejing Tao**: Sample collection, Experimentation. **Hai Wang**: Sample collection, Experimentation. **Fengyi Zhang**: Data analysis, Data visualization. **Jinfeng Wang**: Conceptualization, Supervision, Data interpretation, Writing - Review & Editing, Data analysis, Data visualization. **Hongping Zhang**: Supervision, Data interpretation, Sample collection, Experimentation.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (31722031, 31670119, 31870107 and 32070122), the Key Research Program of the Chinese Academy of Sciences (KFZD-SW-219), and the Medical Health Science and Technology Project of Zhejiang Provincial (2019ZD048, 2018275012). The authors also thank Xuehan Li, Lina Hou, Ming He, Shuai Chen, Zhanzhan Li, and Jingwei Wang for their assistance in this study.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jgg.2020.11.006.

References

Acuna, J., Cohavy, O., Solt, I., Reeder, J., Kim, M., Lebovics, I., Paster, B., Knight, R., Rotmensch, S., 2011. Preliminary observations on the microbial phylogeny of the oral, vaginal, and rectal microbiome in gestational diabetes and healthy pregnancies. Am. J. Obstet. Gynecol. 204, S109–S110.

Artzi, N.S., Shilo, S., Hadar, E., Rossman, H., Barbash-Hazan, S., Ben-Haroush, A., Balicer, R.D., Feldman, B., Wiznitzer, A., Segal, E., 2020. Prediction of gestational diabetes based on nationwide electronic health records. Nat. Med. 26, 71–76.

Belstrom, D., Paster, B.J., Fiehn, N.E., Bardow, A., Holmstrup, P., 2016. Salivary bacterial fingerprints of established oral disease revealed by the Human Oral Microbe Identification using Next Generation Sequencing (HOMINGS) technique. J. Oral Microbiol. 8, 30170.

Buchanan, T.A., Xiang, A.H., Page, K.A., 2012. Gestational diabetes mellitus: risks and management during and after pregnancy. Nat. Rev. Endocrinol. 8, 639–649.

Camelo-Castillo, A.J., Mira, A., Pico, A., Nibali, L., Henderson, B., Donos, N., Tomas, I., 2015. Subgingival microbiota in health compared to periodontitis and the influence of smoking. Front. Microbiol. 6, 119.

Cheung, N.W., Byth, K., 2003. Population health significance of gestational diabetes. Diabetes Care 26, 2005–2009.

Crusell, M.K.W., Brink, L.R., Nielsen, T., Allin, K.H., Hansen, T., Damm, P., Lauenborg, J., Hansen, T.H., Pedersen, O., 2020. Gestational diabetes and the human salivary microbiota: a longitudinal study during pregnancy and postpartum. BMC Pregnancy Childbirth 20, 69.

Crusell, M.K.W., Hansen, T.H., Nielsen, T., Allin, K.H., Ruhlemann, M.C., Damm, P., Vestergaard, H., Rorbye, C., Jorgensen, N.R., Christiansen, O.B., et al., 2018. Gestational diabetes is associated with change in the gut microbiota composition in third trimester of pregnancy and postpartum. Microbiome 6, 89.

Damm, P., Mathiesen, E.R., 2015. Diabetes: therapy for gestational diabetes mellitus–time for a change? Nat. Rev. Endocrinol. 11, 327–328.

Damm, P., Houshmand-Oeregaard, A., Kelstrup, L., Lauenborg, Jeannet, Mathiesen, E.R., Clausen, T.D., 2016. Gestational diabetes mellitus and long-term consequences for mother and offspring: a view from Denmark. Diabetologia 59, 1396–1399.

Dzunkova, M., Martinez-Martinez, D., Gardlik, R., Behuliak, M., Jansakova, K., Jimenez, N., Vazquez-Castellanos, J.F., Marti, J.M., D'Auria, G., Bandara, Hmhn,

et al., 2018. Oxidative stress in the oral cavity is driven by individual-specific bacterial communities. NPJ Biofilms Microbiomes 4, 29.
Farrell, J.J., Zhang, L., Zhou, H., Chia, D., Elashoff, D., Akin, D., Paster, B.J., Joshipura, K., Wong, D.T., 2012. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. Gut 61, 582–588.
Graziani, F., Gennai, S., Solini, A., Petrini, M., 2018. A systematic review and meta-analysis of epidemiologic observational evidence on the effect of periodontitis on diabetes an update of the EFP-AAP review. J. Clin. Periodontol. 45, 167–187.
Gumus, P., Ozcaka, O., Ceyhan-Ozturk, B., Akcali, A., Lappin, D.F., Buduneli, N., 2015. Evaluation of biochemical parameters and local and systemic levels of osteoactive and B-cell stimulatory factors in gestational diabetes in the presence or absence of gingivitis. J. Periodontol. 86, 387–397.
Hasan, S., Aho, V., Pereira, P., Paulin, L., Koivusalo, S.B., Auvinen, P., Eriksson, J.G., 2018. Gut microbiome in gestational diabetes: a cross-sectional study of mothers and offspring 5 years postpartum. Acta Obstet. Gynecol. Scand. 97, 38–46.
Huang, S., Li, R., Zeng, X., He, T., Zhao, H., Chang, A., Bo, C., Chen, J., Yang, F., Knight, R., et al., 2014. Predictive modeling of gingivitis severity and susceptibility via oral microbiota. ISME J. 8, 1768–1780.
Kim, B.S., Han, D.H., Lee, H., Oh, B., 2018. Association of salivary microbiota with dental caries incidence with dentine involvement after 4 years. J. Microbiol. Biotechnol. 28, 454–464.
Kinross, J.M., Darzi, A.W., Nicholson, J.K., 2011. Gut microbiome-host interactions in health and disease. Genome Med. 3, 14.
Kolenbrander, P.E., Andersen, R.N., Blehert, D.S., Egland, P.G., Foster, J.S., Palmer Jr., R.J., 2002. Communication among oral bacteria. Microbiol. Mol. Biol. Rev. 66, 486–505.
Koopman, J.E., Röling, W.F., Buijs, M.J., Sissons, C.H., ten Cate, J.M., Keijser, B.J., Crielaard, W., Zaura, E., 2015. Stability and resilience of oral microcosms toward acidification and Candida outgrowth by arginine supplementation. Microb. Ecol. 69, 422–433.
Lloyd-Price, J., Abu-Ali, G., Huttenhower, C., 2016. The healthy human microbiome. Genome Med. 8, 51.
Martinez, K.B., Leone, V., Chang, E.B., 2017. Microbial metabolites in health and disease: navigating the unknown in search of function. J. Biol. Chem. 292, 8553–8559.
Nassr, A.A., Shazly, S.A., Trinidad, M.C., El-Nashar, S.A., Marroquin, A.M., Brost, B.C., 2018. Body fat index: a novel alternative to body mass index for prediction of gestational diabetes and hypertensive disorders in pregnancy. Eur. J. Obstet. Gynecol. Reprod. Biol. 228, 243–248.
Ng, S.K., Hamilton, I.R., 1971. Lactate metabolism by Veillonella parvula. J. Bacteriol. 105, 999–1005.
Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., Guo, J., Le Chatelier, E., Yao, J., Wu, L., et al., 2014. Alterations of the human gut microbiome in liver cirrhosis. Nature 513, 59–64.
Seraphim, A.P., Chiba, F.Y., Pereira, R.F., Mattera, M.S., Moimaz, S.A., Sumida, D.H., 2016. Relationship among periodontal disease, insulin resistance, salivary cortisol, and stress levels during pregnancy. Braz. Dent. J. 27, 123–127.
Takahashi, N., Yamada, T., 1999. Glucose and lactate metabolism by Actinomyces naeslundii. Crit. Rev. Oral Biol. Med. 10, 487–503.
Teng, F., Yang, F., Huang, S., Bo, C., Xu, Z.Z., Amir, A., Knight, R., Ling, J., Xu, J., 2015. Prediction of early childhood caries via spatial-temporal variations of oral microbiota. Cell Host Microbe 18, 296–306.
Thomas, A.M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., Pozzi, C., et al., 2019. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat. Med. 25, 667–678.
Wang, J., Jia, Z., Zhang, B., Peng, L., Zhao, F., 2019. Tracing the accumulation of *in vivo* human oral microbiota elucidates microbial community dynamics at the gateway to the GI tract. Gut 69, 1355–1356.
Wang, J., Qi, J., Zhao, H., He, S., Zhang, Y., Wei, S., Zhao, F., 2013. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. Sci. Rep. 3, 1843.
Wang, J., Zheng, J., Shi, W., Du, N., Xu, X., Zhang, Y., Ji, P., Zhang, F., Jia, Z., Wang, Y., et al., 2018. Dysbiosis of maternal and neonatal microbiota associated with gestational diabetes mellitus. Gut 67, 1614–1625.
Weinert, L.S., 2010. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy: comment to the international association of diabetes and pregnancy study groups consensus panel. Diabetes Care 33, e97.
Xiao, C., Ran, S., Huang, Z., Liang, J., 2016. Bacterial diversity and community structure of supragingival plaques in adults with dental health or caries revealed by 16S pyrosequencing. Front. Microbiol. 7, 1145.
Zakrzewski, M., Proietti, C., Ellis, J.J., Hasan, S., Brion, M.J., Berger, B., Krause, L., 2017. Calypso: a user-friendly web-server for mining and visualizing microbiome-environment interactions. Bioinformatics 33, 782–783.
Zheng, J., Xiao, X., Zhang, Q., Mao, L., Yu, M., Xu, J., Wang, T., 2017. The placental microbiota is altered among subjects with gestational diabetes mellitus: a pilot study. Front. Physiol. 8, 675.