Relative reduction of biological and phylogenetic diversity of the oral microbiota of diabetes and pre-diabetes patients

Amr T.M. Saeb,∗ Khalid A. Al-Rubeaan, Khalid Aldosary, G.K. Udaya Raja, Balavenkatesh Manic, Mohamed Abouelhoda, Hamsa T. Tayeb

Genetics and Biotechnology Department, Strategic Center for Diabetes Research, College of Medicine, King Saud University, Saudi Arabia
College of Dentistry, King Saud University, Saudi Arabia
Integrated Gulf Biosystems, Riyadh, Saudi Arabia
Saudi Human Genome Project, King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia
Genetics Department, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

ARTICLE INFO
Keywords:
Oral microbiota
Diabetes
Impaired glucose tolerance
16S rRNA
Next-generation sequencing
Ion torrent
Operational taxonomic unit
Biological diversity
Phylogenetic diversity

ABSTRACT
Background: A reciprocal relationship between oral health and systemic disease, such as type 2 diabetes, has been suggested, whereby a systemic disease is a predisposing factor for oral infection. If the infection occurs, it in turn aggravates the progression of the systemic disease. According to several studies, certain constituents of the oral microbiota are linked to diabetes, metabolic syndrome, and obesity. In the current study, we aimed to compare the microbial diversity and population structure of the oral microbiota of normoglycemic, impaired glucose tolerance (IGT), and diabetes patients.

Methodology: The study followed a case-control design, with 15 type 2 diabetes patients, 10 IGT subjects, and 19 control subjects. All subjects underwent assessment of periodontitis and oral health. Saliva samples were collected, and DNA was isolated from these samples. Hypervariable regions of the 16S rRNA gene were amplified and sequenced, and the generated sequences underwent bioinformatics analysis. Statistical analysis and diversity index calculations were made using the statistical software R, vegan R-package, and Past3.20 software.

Results: Overall, 551 operational taxonomic units (OTUs) were identified. Based on OTU analysis, a clear reduction of the number of species was observed in both IGT (412) and diabetes groups (372) compared with that in the normoglycemic group (502). This was associated with a similar pattern of reduction of biological diversity among the three groups. The phylogenetic diversity (PD-SBL) value in the normoglycemic group was higher than that in the diabetes group. The diabetes group exhibited the highest evenness value and the highest microbiota bacterial pathogenic content.

Conclusion: A clear reduction of the biological and phylogenetic diversity was apparent in the diabetes and pre-diabetes oral microbiota in comparison with that in the normoglycemic oral microbiota. However, this was associated with an increase in the pathogenic content of the hyperglycemic microbiota. The results of this study may aid to better understanding of the directionality of the mysterious reciprocal relationship.

1. Introduction
Diabetes is the leading cause of morbidity and mortality globally. The prevalence of type 2 diabetes worldwide reached 415 million people in 2015 and is projected to reach 642 million by 2040 [1]. Diabetes is a significant risk factor for rigorous and progressive periodontitis, i.e., infection or lesions that lead to the obliteration of tissues and the supporting bone for tooth attachment. The oral cavity is a permanent source of infectious agents and its state often mirrors the progression of systemic pathologies. Diabetes and periodontitis are thought to have common pathogenesis, which involves triggering the inflammatory responses that can be perceived at both local and systemic levels. The reciprocal relationship between diabetes and oral health has been supported by multiple lines of evidence [2]. The relationship is an example of systemic disease as a predisposing factor for oral infection. In turn, when that infection occurs, it aggravates the progression of the systemic disease. In addition, several studies suggested that controlling one arm in this reciprocal relationship can lead to improvement of the status of the other arm [2–4]. It has been proposed that increased levels of advanced glycation end-products (AGEs)
associated with a chronic hyperglycemic state, accompanied by infection and an embellished host response, might account for the clinical outcomes observed in diabetes patients with periodontal disease. Furthermore, bacterial cellular components and products, such as lipopolysaccharide (LPS) and endotoxin, also play a role in aggravating the inflammatory response of the host cell, acting through Toll-like receptors (TLRs) and inducing the inflammatory cascade [5,6]. Consequently, oral infection could be an indicator for the incidence and progression of diabetes. The inflammatory response may essentially be viewed as a chronic effect of hyperglycemia, specifically, the presence of biological reactive glycated proteins and lipids. In turn, enhanced innate immune responses and destruction of the periodontal tissue are related to an altered inflammatory response. Since periodontitis can be more than just a localized oral infection, its effects are thought to be systemic [2,5].

Although little studied, the oral microbiota may be important for the progression of chronic diseases, such as diabetes, by directly metabolizing chemical carcinogens and by exerting systemic inflammatory effects [7–9]. Recently, it has been suggested that Interleukin 17 (IL-17) production enhanced by diabetes alters the oral microbiota and renders it more pathogenic [10]. The oral microbiota can affect systemic health by inhibiting potential pathogens, regulating immune responses, and nutrient absorption and metabolism [11]. In addition, bacteria from the oral microbiota and/or their by-products can disseminate in the body, leading to systemic diseases, such as the cardiovascular disease, Alzheimer’s disease, and cancer [9,12–15]. Several studies suggested that some oral microbiota constituents are linked to diabetes, metabolic syndrome, and obesity. For example, the presence of Porphyromonas gingivalis, especially clones with type II fimbriae, in the periodontal pockets affect the glycemic index in diabetes [16]. Moreover, poor glycemic control is associated with increased cell numbers of the red complex bacteria P. gingivalis, T. forsythia, and Treponema denticola in subgingival biofilm [17]. In addition, although Actinobacillus actinomycetemcomitans, Campylobacter rectus, Capnocytophaga spp., Eikenella corrodens, Fusobacterium nucleatum, P. gingivalis, and Prevotella intermedia were observed in both diabetes patients and non-diabetic individuals, the levels of P. gingivalis were significantly elevated in the diabetes patients [18]. Furthermore, the levels of oral bacteria in diabetes patients are slightly higher than those in non-diabetic individuals. In fact, levels of Bifidobacteria in the mouth and stool of subjects with HbA1c ≥ 6.5% are relatively low [19]. In addition, the numbers of total streptococci and lactobacilli in the supragingival plaque from diabetes patients are significantly higher than those in non-diabetic subjects. Furthermore, the numbers of lactobacilli in diabetes patients with active caries are elevated, although saliva eubacterial DNA profiles are not significantly altered in these subjects [20]. Most recently, Long et al. [21] showed that the phylum Actinobacteria is significantly less abundant in patients with diabetes than in the controls. They also showed that multiple bacterial taxa from the phylum Actinobacteria are associated with the risk of type 2 diabetes. However, a major limitation of that study was the classification of oral bacteria only to the genus level in relation to diabetes risk [21]. In the current study, we aimed to compare the microbial diversity and population structure of oral microbes to the species level in normoglycemic, glucose intolerant, and diabetes patients using deep sequencing. In addition, we also aimed to identify a bacterial species combination that could be used as an early biomarker of diabetes.

2. Materials and methods

This study was conducted in agreement with institutional ethics review board of the King Saud University, Collage of Medicine Riyadh, Kingdom of Saudi Arabia and the subjects provided written informed consent for participating in this study. The current study was designed as a cross-sectional case-control hospital-based study involving three cohorts (normoglycemic, pre-diabetes, and type 2 diabetes subjects).

The participants were recruited from May 2013 to March 2015 at the University Diabetes Center, King Saud University. After exclusion of individuals who did not meet the inclusion criteria (see Results section), 44 subjects were recruited: 19 normoglycemic subjects (the control cohort), 10 IGT subjects (the pre-diabetes cohort), and 15 type 2 diabetes patients. The subjects represented both genders, were aged between 40 and 55 years, and Saudi nationals. Any subjects used an antibiotic for the last 6 months or known to have severe periodontal disease was excluded. Immunocompromised patients, e.g., post organ transplantation, with end-stage renal disease (ESRD), with severely debilitating diseases (cancer), or receiving chemotherapy were excluded. Patients with autoimmune diseases and pregnant women were also excluded. For the diabetes cohort, any patient with type 1, gestational, and other types of diabetes were excluded. Normoglycemic subjects and pre-diabetes subjects were recruited from among the patients’ relatives attending the clinics, after a 100-g oral glucose tolerance (OGT) test after 10-h overnight fast. The criteria of the American Diabetes Association (ADA) were employed, as follows: normoglycemic: FPG below 100 mg/dl or 2-h post-load glucose below 140 mg/dl; pre-diabetic: FPG of 100–125 mg/dl or 2-h post-load glucose between 140 and 199 mg/dl. The type 2 diabetes patients were recruited from among the patients attending the University Diabetes Center clinics and diagnosed as per ADA criteria (FPG ≥126 mg/dl or 2-h plasma glucose ≥200 mg/dl), and managed with oral hypoglycemic agents only.

Each subject was interviewed individually by a research physician and provided written informed consent. General demographic data, including the age, gender, weight, height, and history of any dental problems, in addition to drug history and mouth hygiene practice were obtained. For the diabetes patients, additional information was collected, including diabetes duration, the presence of chronic complications or associated disease. The details of metabolic control, including HbA1c, FPG, and 2-h post-prandial glucose levels, were obtained from the patients’ files.

2.1. Periodontal assessment

All subjects (44) were referred to a specialized dental clinic for periodontal assessment. A full mouth clinical examination was conducted by three examiners who had been previously exposed to the National Health and Nutrition Examination Survey (NHANES) reference examiner (Dr. Bruce Dye). The Periodontal disease was assessed by determining the probing pocket depth (PPD) and clinical attachment loss (CAL) at six sites, including the distobuccal, mid-buccal, mesiobuccal, distolingual, mid-lingual, and mesiobuccal buccal surfaces, for all teeth, excluding the third molar. The determinations were done using a periodontal probe (no. PCP2; Hu-Friedy, Chicago, IL, USA) and rounded off upwards to the nearest millimeter. Periodontitis was defined according to the Centers for Disease Control and Prevention/American Academy of Periodontology (CDC/AAP) [22]. Severe periodontitis was defined as having at least two interproximal sites with CAL ≥ 6 mm (not on the same tooth) and at least one interproximal site with PPD ≥ 5 mm. Moderate periodontitis was defined as having at least two interproximal sites with CAL ≥ 4 mm (not on the same tooth) or at least two interproximal sites with PPD ≥ 5 mm (not on the same tooth). Mild periodontitis was defined as having at least two interproximal sites with CAL ≥ 3 mm and at least two interproximal sites with PPD ≥ 4 mm (not on the same tooth) or one site with PPD ≥ 5 mm. During the determination of PPD, the periodontal probe was inserted at the base of the sulcus or pocket with a maximum force of 20 g; bleeding on probing (BOP) was considered positive if the probe site bled approximately 20 s after probing the lingual and buccal surfaces of each tooth. The BOP was classified as high if 30% or more teeth were BOP positive, and as low otherwise [23]. The Silness-Löe plaque index, a measure of oral hygiene status, was determined by visual assessment of the presence of bacterial plaque after passing a periodontal probe around the tooth surface of six pre-selected Ramfjord teeth. The plaque
index was coded as 0, if no plaque was present; 1, if the dental plaque was present after passing the periodontal probe around the tooth; 2, if the plaque was visible along the gingival margin; and 3, if the tooth surface was covered with abundant plaque.

2.2. Molecular analysis

All study participants were asked to provide saliva in the Omingene saliva collection tubes of kit OM501 for microbial DNA analysis (Ontario, Canada) following the instructions of the manufacturer, after rubbing the tongue for approximately 60s. Prior to sample collection, the subjects were instructed to abstain from eating, drinking, smoking, and mouth washing for at least 60 min. The samples were placed immediately in an icebox and transferred to a laboratory of the Strategic Center for Diabetes Research, where they were stored at −20 °C until DNA isolation.

2.3. Isolation, purification, quantification, and enrichment of bacterial DNA

Bacterial DNA was isolated from the collected saliva samples using the Promega Maxwell 16 automated DNA isolation machine, according to the instructions of the manufacturer. The concentration of isolated DNA was determined using the NanoDrop 2000c UV–Vis spectrophotometer. Agilent 2100 Bioanalyzer system was used for the sizing, quantitation, and quality control assessment for the isolated DNA. The NEBNEXT® Microbiome DNA kit was used to eliminate human DNA and to facilitate enrichment of microbial DNA in the samples. This kit selectively removes the CpG-methylated host DNA to maintain microbial DNA only after enrichment. All isolated DNA samples were then stored at −20 °C.

2.4. Amplification of hypervariable regions of the 16S rRNA gene

16S rRNA gene was amplified using Ion 16S™ metagenomics kit (Thermo Fisher Scientific, Massachusetts, USA) following the manufacturer’s instructions. To increase the resolving power of 16S rRNA profiling, the primers were designed to amplify the variable regions 2, 4, and 8 in a single tube, yielding amplicon fragments of approximately 250 bp, 288 bp, and 295 bp, respectively. In a second tube, a multiplex PCR reaction targeted the variable regions 3, 6–7, and 9, yielding amplicon fragments of approximately 215 bp, 260 bp, and 209 bp, respectively. The primer pools in the kit are designed to target > 80% of sequences found in the Greengenes database with 100% identity for a primer pair amplifying at least one variable region. The PCR amplifications were performed using Applied Biosystems Veriti™ 96-Well thermal cycler, with the following cycling conditions: 95 °C for 10 min; followed by 18–25 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 20 s; a final extension step at 72 °C for 7 min; and a hold at 4 °C overnight. The amplicons were purified using an Agencourt AMPure XP kit (Beckman Coulter, Brea, CA, USA) according to the manufacturer's instructions. The PCR-amplified products were used for library preparation using Ion Plus fragment library kit with sample indexing using IonXpress barcode adapters 1–16 kit, following the manufacturer's instructions. The library concentration was determined using the Ion universal library quantification kit and diluted to 13 pM. The diluted libraries were pooled together in equimolar quantities and used as a template for emulsion PCR. Emulsion PCR and enrichment of template-positive particles were performed using an Ion PGM HI-Q OT2 kit (Thermo Fisher Scientific, Massachusetts, USA) and the Ion OneTouch 2 system (Life Technologies), according to the manufacturers’ instructions. The enriched template-positive ion sphere particles were loaded onto an Ion 318 chip v2 (Thermo Fisher Scientific) and sequenced on an Ion PGM (Thermo Fisher Scientific) using an Ion PGM HI-Q sequencing kit (Thermo Fisher Scientific), according to the manufacturers’ instructions, to ensure high degree of sensitivity required and considering the high complexity of samples.

2.5. DNA sequence analysis

Data sequence analysis was performed using automated streamlined software Torrent Suite software v5.0 (TSS) and IonReporter software v5.0. The Torrent Suite Software was used for sequencing base calling from the PGML-generated data, and the IonReporter was used for annotation and taxonomical assignments. The standard analysis parameters in the Torrent Suite software for primary base calling included trimming low-quality 3’-ends of reads; removing duplicated reads; filtering out entire reads with average quality scores below Q20 (base-call quality); and removing adapter sequences, 3’-ends with low-quality scores, short reads, adapter dimers, polynucleotide reads, etc. Q20 is a Phred score based on the probability of error (−10 × log10). It corresponds to a predicted error rate of 1%. AQ20 is a read length at which the error rate is 1% or less. Ion Torrent PGM-generated sequences were saved in a binary alignment map (ubam) file format; the data consisted of base-call sequences, flow (flowgram), and qual (quality score). IonReporter Uploader plugin was used to transfer all high-quality data from the Torrent Suite software to the Cloud Ion Reporter software.

Analysis of the three groups of samples (normal glycemic, IGT, and diabetes) was launched in the IonReporter software within the 16S Metagenomics analysis workflow. The Metagenomics workflow contains curated Greengenes and premium-curated MicroSEQ ID 16S rRNA reference databases. The Greengenes database was manually curated from the available existing public library because of the relatively large size of the resource, incorporating over 1.2 million references. Since the Greengenes mainly contains taxonomy information at family level and above, taxonomy information at genus and species levels was not available for almost 1 million references. A custom utility was used to curate Greengenes for genus and species-level information by querying the NCBI database. The final library contained 215,967 references directly from Greengenes, and 146,822 references with updated genus and species names from NCBI, totaling 362,789 references. Data in both databases were converted into a BLAST-compatible format for the analysis. The analysis pipeline included annotation, taxonomic assignment, and classification of reads by alignment with either the curated MicroSEQ ID or curated Greengenes database.

The IonReporter identifies chimeric sequences based on a built-in algorithm and removes them before subsequent downstream analysis. Samples from each group were analyzed together, and the reads were classified at species level using the 16S Metagenomics analysis workflow. The analysis steps included primer trimming and read length check. The processed reads were placed in a “hash table.” The hash table contained all the unique reads, and their abundance or copy number information. When a new read is added to the hash table, the read will have a copy number of 1, if it does not exist in the table already; if it does exits, the copy number will be increased by 1. After adding all reads, the hash table contained all the unique reads, each with copy number information. All unique reads were then converted to the FASTA file format. Each read had a unique ID for identification throughout the process. Multistage BLAST search of reads against information in the databases was then initiated. This algorithm generates input files and starts one BLAST process for each sample. One BLAST parameter was E-value equal to 0.01; results with E-value over 0.01 were not included in the output. The final dataset of Ion Reporter-generated results consisted of species-level OTU identification, with primer and classification information, percent ID, and mapping information, etc.

Some OTUs were identified to species level by the IonReporter. Such sequences were retrieved and BLAST analysis was performed in HOMD. For each unsolved sequence, the top-four hits were used for further species-level assignment, based on the following criteria: 100% identity match resulted in direct species assignment; if none of the top-four hits
resulted in 100% identity match, the selected maximum percentage identities with identified species. In HOMD, some sequences are not assigned at the species level; these unsolved cases are designed by a respective clone or hot strain species name in BLAST output. All sequence data were submitted to NCBI under an SRA accession number SRP159425, and BioSample accessions numbers SAMN09929596–SAMN09929639 in the BioProject PRJNA488297.

2.6. Statistical and phylogenetic analysis

Statistical analysis was performed as described in Lourenco et al., 2018 [24]. Data were analyzed by using Statistical Package for Social Studies (SPSS 22; IBM Corp., New York, NY, USA) and Past 3.21 software. The demographic data, frequency and means were computed for each patient and group. The clinical parameters were averaged for each patient and the three glycemic groups. Comparisons amongst groups were evaluated by Chi-square (for categorical data), Mann-Whitney (for pairs of groups). The continuous variables were expressed as mean ± standard deviation. The t-test was used for continuous variables which have normal distribution and Mann-Whitney test was used test for non-normal distribution. Shapiro-Wilk test was used to assess the normality of the data. A p-value < 0.05 was considered statistically significant. All diversity indices statistical software R, vegan R-package, and Past321 software [25]. Cluster analysis (Unweighted pair-group average UPGMA) using Bary-Curtis, Manhattan, Elucidation and correlation similarity indices with bootstrap number = 1000 were performed using Past321 software. The principal component analysis using variance-covariance and correlation matrix was also performed using Past321 software. MEGA 7 software was used to calculate phylogenetic diversity values [26]. The 16S rRNA based phylogenetic diversity (Maximum likelihood total branch length) was calculated as described Saeb at al., 2017 [27].

3. Results

3.1. Recruitment of the study subjects

Overall, 800 subjects were screened from a group of patients followed at the University Diabetic Center at King Saud University and their non-diabetic family members. Of these, 180 patients fulfilled the inclusion criteria, i.e., type 2 diabetes, aged 40–55 years, and Saudi nationals. The patients were interviewed by research physicians. Only 44 subjects fulfilled the study protocol, after excluding any subjects who had taken antibiotics in the preceding 6 months or had a severe periodontal disease. Patients administered folic acid or refusing dental appointment or to provide a saliva sample were also excluded. The final study cohort included 19 normoglycemic subjects, 10 individuals with impaired glucose tolerance (IGT), and 15 type 2 patients with diabetes. The age ranges for the selected study cohort were 41–56 years, 33–53 years, and 45–53 years for the normoglycemic, IGT, and diabetes groups, respectively. Males were more predominant than females in the normoglycemic and diabetes groups, but their number was equal to that of females in the IGT group. Mild periodontitis was observed in one normoglycemic and diabetes groups; and 8 for the IGT and diabetes groups. The number of species per sample was close to a plateau over 10,000 to 15,000 sequence reads (Fig. S1). The number of OTUs shared by the three studied groups was 311; it was 78 for the normoglycemic and IGT groups; 21 for the normoglycemic and diabetes groups; and 8 for the IGT and diabetes groups. The number of species identified only in the normoglycemic group was 86; 15 in the IGT group; and 26 in the diabetes group (Table S2).

3.3. The core microbiota: class level

In the sampled population (44 individuals), 11 classes, all from the domain Bacteria, represented the core taxa (95% of reads). These were Actinobacteria, Bacilli, Bacteroidia, Betaproteobacteria, Clostridia, Epsilonproteobacteria, Erysipelotrichia, Flavobacteria, Gammaproteobacteria, Fusobacteria, and Negativicutes. The class Bacteroidia represented 28% of the core microbiota, followed by the class Bacilli (19%). The least abundant component was the class Flavobacteria (0.2% of the core microbiota). For the normoglycemic group, 11 classes similarly represented the core taxa (95% of reads). The class Bacteroidia represented 31% of the core microbiota, followed by the class Bacilli (21.1%), with the least abundant class of Flavobacteria (0.25%). For the IGT group, the same number and identities of classes represented the core taxa (95% of reads). The class Bacteroidia represented the highest percentage (26%) of the core microbiota. However, Gammaproteobacteria was the second dominant taxon (21%), followed by Actinobacteria (17%), and Bacilli (16.4%). The class Erysipelotrichia was the least abundant component of the IGT study was determined by 16S rRNA gene amplicon analysis using an Ion Personal Genome Machine (PGM). After applying the read quality and length filters, 24,568,466 raw reads were obtained. The number of reads per sample was 293,442–4,016,984, with a mean of 558,374 reads per sample. The IonReporter software automatically removed the chimeric reads; further, UCHIME, Chimera Slayer, and Decipher programs identified 0.66% chimeric reads, which were later excluded. The average read length was 175 base pair.

The sequences were assigned to 551 operational taxonomic units (OTUs), using a cutoff distance of 0.05, by sequence comparison to the updated Human Oral Microbial Database (HOMD) and trusted-HOMDext (Fig. 1). Most of the reads (99% OTUs) enabled identification to the species level. The total number of species identified was 502 in the normoglycemic group; 412 in the IGT group; and 372 in the diabetes group. The rarefaction curve for the number of observed species per sample was close to a plateau over 10,000 to 15,000 sequence reads (Fig. S1). The number of OTUs shared by the three studied groups was 311; it was 78 for the normoglycemic and IGT groups; 21 for the normoglycemic and diabetes groups; and 8 for the IGT and diabetes groups. The number of species identified only in the normoglycemic group was 86; 15 in the IGT group; and 26 in the diabetes group (Table S2).
core microbiota (0.22%). By contrast, in the diabetes group, albeit 11 classes represented the core taxa (95% of reads), these were Actinobacteria, Bacilli, Bacteroidia, Betaproteobacteria, Clostridia, Epsilonproteobacteria, Erysipelotrichia, Fusobacteriia, Gammaproteobacteria, Negativicutes, and Spirochaetia. In this group the class Spirochaetia contributed in the core microbiota but not Flavobacteriia. Again, class Bacteroidia were the most abundant component (28.5%) of the core microbiota, followed by Bacilli (19.4%). The class Spirochaetia was the least abundant component of the core microbiota in diabetes (0.4%) (Fig. 2).

3.4. The core microbiota: family level

In the sampled population, 22 families represented the core taxa (95% of reads). The family Prevotellaceae was the most abundant component of the core microbiota (25.26%), while the family Clostridiales XI Incertae Sedis was the least abundant component (0.17%). In the normoglycemic group, 24 families represented the core taxa (95% of reads), with the family Prevotellaceae being the most abundant component (29.41%) and the family Clostridiales XI Incertae Sedis the least abundant component (0.14%) of the core microbiota. In the IGT group, 23 families represented the core taxa (95% of reads). The family Prevotellaceae was the most abundant component of the core microbiota (23.41%), followed by the genus *Haemophilus* (17.75%), while the genus *Filifactor* was the least abundant component (0.16%). In the diabetes group, 29 genera represented the core taxa (95% of reads). The genus *Prevotella* was the most abundant component of the core microbiota (27%), while the genus *Peptostreptococcus* was the least abundant component (0.23%) (Fig. 3).

3.5. The core microbiota: genus level

In the sampled population, 29 genera represented the core taxa (95% of reads). The genus *Prevotella* was the most abundant component of the core microbiota (26.25%), while the genus *Selenomonas* was the least abundant component (0.14%). In the normoglycemic group, 34 genera represented the core taxa (95% of reads). The genus *Prevotella* was the most abundant component of the core microbiota (28.29%), followed by the genus *Streptococcus* (23.35%), whereas the genus *Peptostreptococcus* was the least abundant component (0.12%). In the IGT group, 27 genera represented the core taxa (95% of reads). The genus *Prevotella* was the most abundant component of the core microbiota (23.41%), followed by the genus *Haemophilus* (17.75%), while the genus *Filifactor* was the least abundant component (0.16%). In the diabetes group, 29 genera represented the core taxa (95% of reads). The genus *Prevotella* was the most abundant component of the core microbiota (25.83%), followed by the genus *Rothia* (11.26%), whereas the genus *Actinomyces* was least abundant (0.11%) (Fig. 4).

3.6. The core microbiota: species level

In the sampled population, 77 species represented the core taxa (95% of reads). *Prevotella melaninogenica* was the most abundant component of the core microbiota (13.19%), followed by *Rothia mucilaginosa* (10.34%), and *Haemophilus parainfluenzae* (9.42%), while *Actinomyces denitrificans* was least abundant (0.11%). In the IGT group, 71 species represented the core taxa
Fig. 3. The core microbiota (family level) of the three groups and family-level comparison of three glycemic groups (a) group illustration (b) taxa illustration.
Fig. 4. The core microbiota (genus level) of the three groups and genus-level comparison of three glycemic groups (a) group illustration (b) taxa illustration.
H. parainfluenzae was the most abundant component of the core microbiota (15.83%), followed by R. mucilaginosa (13.04%), and P. melaninogenica (10.87%), while Streptococcus infantis was the least abundant component (0.15%). In the diabetes group, 76 species represented the core taxa (95% of reads). R. mucilaginosa was the most abundant component of the core microbiota (10.12%), followed by P. melaninogenica (9.61%), H. parainfluenzae (6.12%), and Streptococcus salivarius (5.35%). Prevotella auricularis was the least abundant component (0.12%). An overview of the taxonomy of the common core microbiota in the study population and in the three subject groups, at the taxonomic levels mentioned above, is provided in Supplemental Fig. 2. A quantitative comparison of 88 species of the core microbiota in three glycemic groups is shown in Fig. 5 as a heatmap and as a comparative area demonstration in Fig. 6.

3.7. Biological diversity among different glycemic groups

An overview of biological alfa-diversity and diversity indices calculated for the three investigated glycemic control groups is given in Table 1. As stated above, the number of identified OTUs (species) was
the highest in the normoglycemic group, followed by the IGT group, and the diabetes group. The most well-known and accepted diversity indices, i.e., Chao-1, Margalef, and Fisher alpha, indicated that the biological diversity decreased with progression of the normoglycemic status to diabetes. The Chao-1 index was 502 for the normoglycemic group; 412 for the IGT group; and 372 for the diabetes group. The Margalef index was 35.2 for the normoglycemic group; 29.27 for the IGT group; and 27.49 for the diabetes group. The Fisher alpha index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.

The following beta-diversity values for the three groups were calculated: Whittaker index of 0.28538; Harrison index of 0.14269; Cody index of 141; Routledge index of 0.074622; Wilson-Shmida index of 0.32893; Mourelle index of 0.16446; Harrison 2 index of 0.048805; and Williams index of 0.088929. The pairwise Cody index was 68 for the normoglycemic and IGT group comparison; 372 for the IGT and diabetes group comparison. The pairwise diversity t-tests were also performed, giving the Shannon index for the three tested groups. For the normoglycemic and IGT groups, the Shannon H index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.

The following beta-diversity values for the three groups were calculated: Whittaker index of 0.28538; Harrison index of 0.14269; Cody index of 141; Routledge index of 0.074622; Wilson-Shmida index of 0.32893; Mourelle index of 0.16446; Harrison 2 index of 0.048805; and Williams index of 0.088929. The pairwise Cody index was 68 for the normoglycemic and IGT group comparison; 372 for the IGT and diabetes group comparison. The pairwise diversity t-tests were also performed, giving the Shannon index for the three tested groups. For the normoglycemic and IGT groups, the Shannon H index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.

The following beta-diversity values for the three groups were calculated: Whittaker index of 0.28538; Harrison index of 0.14269; Cody index of 141; Routledge index of 0.074622; Wilson-Shmida index of 0.32893; Mourelle index of 0.16446; Harrison 2 index of 0.048805; and Williams index of 0.088929. The pairwise Cody index was 68 for the normoglycemic and IGT group comparison; 372 for the IGT and diabetes group comparison. The pairwise diversity t-tests were also performed, giving the Shannon index for the three tested groups. For the normoglycemic and IGT groups, the Shannon H index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.

The following beta-diversity values for the three groups were calculated: Whittaker index of 0.28538; Harrison index of 0.14269; Cody index of 141; Routledge index of 0.074622; Wilson-Shmida index of 0.32893; Mourelle index of 0.16446; Harrison 2 index of 0.048805; and Williams index of 0.088929. The pairwise Cody index was 68 for the normoglycemic and IGT group comparison; 372 for the IGT and diabetes group comparison. The pairwise diversity t-tests were also performed, giving the Shannon index for the three tested groups. For the normoglycemic and IGT groups, the Shannon H index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.

The following beta-diversity values for the three groups were calculated: Whittaker index of 0.28538; Harrison index of 0.14269; Cody index of 141; Routledge index of 0.074622; Wilson-Shmida index of 0.32893; Mourelle index of 0.16446; Harrison 2 index of 0.048805; and Williams index of 0.088929. The pairwise Cody index was 68 for the normoglycemic and IGT group comparison; 372 for the IGT and diabetes group comparison. The pairwise diversity t-tests were also performed, giving the Shannon index for the three tested groups. For the normoglycemic and IGT groups, the Shannon H index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.

The following beta-diversity values for the three groups were calculated: Whittaker index of 0.28538; Harrison index of 0.14269; Cody index of 141; Routledge index of 0.074622; Wilson-Shmida index of 0.32893; Mourelle index of 0.16446; Harrison 2 index of 0.048805; and Williams index of 0.088929. The pairwise Cody index was 68 for the normoglycemic and IGT group comparison; 372 for the IGT and diabetes group comparison. The pairwise diversity t-tests were also performed, giving the Shannon index for the three tested groups. For the normoglycemic and IGT groups, the Shannon H index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.

The following beta-diversity values for the three groups were calculated: Whittaker index of 0.28538; Harrison index of 0.14269; Cody index of 141; Routledge index of 0.074622; Wilson-Shmida index of 0.32893; Mourelle index of 0.16446; Harrison 2 index of 0.048805; and Williams index of 0.088929. The pairwise Cody index was 68 for the normoglycemic and IGT group comparison; 372 for the IGT and diabetes group comparison. The pairwise diversity t-tests were also performed, giving the Shannon index for the three tested groups. For the normoglycemic and IGT groups, the Shannon H index was 48.49 for the normoglycemic group; 39.78 for the IGT group; and 37.71 for the diabetes group.
3.9. Phylogenetic diversity analysis

The phylogenetic diversity information (PD-SBL) for the three investigated glycemic control groups is provided in Table 1. The normoglycemic group exhibited the highest PD-SBL value (72.93) (Supplemental Fig. 3), followed by the diabetes group (67.36) (Supplemental Fig. 4), and the IGT group (61.43) (Supplemental Fig. 5). The phylogenetic diversity information for the total microbiota of individuals from the normoglycemic group is presented in Table 2. The highest PD-SBL value (62.89) was observed for individual 14 (Supplemental Fig. 6), while the lowest PD-SBL value (46.91) was observed for individual 6 with a mean value of 54.58842 and a median of 54.14 (Supplemental Fig. 7). The phylogenetic diversity information for the total microbiota of individuals from the IGT group is presented in Table 3. The highest PD-SBL value (57.58) was observed for subject 2 (Supplemental Fig. 8) and the lowest PD-SBL value (43.98) was observed for subject 7 with a mean value of 52.191 and a median of 53.53 (Supplemental Fig. 9). The phylogenetic diversity information for the total microbiota of individuals from the diabetes group is presented in Table 4. The highest PD-SBL value (57.45) was observed for subject 14 (Supplemental Fig. 10), while the lowest PD-SBL value (41.66) was observed for subject 4 with a mean value of 50.86133 and a median of 51.15 (Supplemental Fig. 11). The phylogenetic diversity comparison amongst the three glycemic groups is presented as violin plots in Fig. 7 (c).

3.10. Cluster analysis and principal component analysis (PCA)

Both cluster analysis and Principal component analysis (PCA) (Supplemental Fig. 12 and Fig. 8) showed the presence of clustering pattern, namely, Diabetic cluster, IGT cluster and two Normoglycemic clusters. The Diabetic cluster all the 15 diabetic microbiomes, the IGT cluster contained all IGT microbiomes in addition to the normoglycemic microbiome 57. The Normoglycemic cluster I contained 10 microbiomes while the Normoglycemic cluster II contained 7 microbiomes. There was a normoglycemic microbiome located between clusters, namely, Nor 71. Both cluster analysis and Principal component analysis confirmed the reduction of microbiome diversity toward the diabetic status.

4. Discussion

In the current study, we aimed to investigate the oral microbiota in both normal and hyperglycemic subjects. We selected diabetes and IGT patients to represent conditions of abnormal glucose metabolism. We employed very strict criteria to select individuals to participate in the study. The normoglycemic control subjects were selected from the immediate family members of the patients, to account for the host genetic background and environmental effects. The genetic background and environment are important factors shaping the host-associated microbial communities. Namely, the similarity of the oral microbiota increases with the shared host genotype, irrespective of the dental state. In addition, many taxa are inherited from the parent to the offspring. On the other hand, the occurrence of potential cariogenic bacterial taxa is likely not controlled by genetic factors [28]. In the current study, we selected subjects and diabetes patients who were either not affected by periodontitis or had only mild periodontitis. This was meant to diminish the effect of oral health on the oral microbiota structure, in an attempt to link the observed changes in the oral flora and the glycemic status with minimal impact from other factors.

Clear changes in the oral flora of diabetes patients with both
aggressive or chronic periodontitis have been documented [11,29,30].
In the current study, the number of species (OTUs) identified in the IGT
and in the diabetes group was clearly reduced in comparison with that
in the normoglycemic subjects. This could be explained by three dif-
f erent mechanisms. First, elevated glucose levels in the saliva of subjects
diagnosed with diabetes and pre-diabetes could have impacted the oral
environment, enhancing the growth of certain bacterial species at the
expense of others [31,32]. Second, mouth dehydration, usually asso-
ciated with diabetes, could result in the observed reduction of microbial
diversity [33]. Third, hyperglycemia could have led to the acidification
of the oral environment, perturbing the oral microbiota [34]. Similarly,
it was found that oral microbial diversity decreased in diabetics and
increased with progression of periodontal disease compared with perio-
dontally healthy controls [35]. Moreover, in the microbiota of preg-
nant women with gestational diabetes mellitus (GDM). The microbiota
was remarkably altered with a sturdy correlation between bacteria and
the oral glucose tolerance test [36].

In general, the three clinical groups shared more than 50% of all
identified species; 20% were shared by any two groups; and the
remaining 30% were group-specific (Fig. 1 and Supplementary Table 2).
Indeed, only 86 OTUs were normoglycemic group-specific, 15 OTUs
were IGT group-specific, and 26 OTUs were diabetes group-specific.
In addition, 78 OTUs were shared by the normoglycemic and IGT groups,
and 27 were shared by the normoglycemic and diabetes groups, and 8
shared by the IGT and diabetes groups. The shared OTUs are considered
transitional taxa. We propose that these differential and transitional
species associated with each clinical state might have a clinical value
and require closer screening.

Consequently, we investigated the number and types of bacterial
species that were non-pathogenic, pathogenic, or beneficial/probiotic.
The non-pathogenic taxa are bacteria that are not associated with the
oral tissue infection or caries. Pathogenic species are associated with
any oral infection, such as periodontitis or related to dental caries.
Probiotic microorganisms are those with oral or general health benefits.
Accordingly, of the 86 OTUs exclusive to the normoglycemic group, 63
(73.3%) were not oral pathogens, 15 (17.4%) were oral/dental patho-
gens, and 8 (9.3%) were beneficial probiotic bacteria (Supplementary
Table 3). Examples of recovered dental pathogens include Actinomycys
radiiidentis, previously isolated in pure culture from infected root canals
of teeth [37], and Mogibacterium neglectum, previously recovered from
tongue plaque and necrotic dental pulp [38]. On the other hand,
Lactobacillus paracasei, which is known to suppress the growth of many
pathogens, could result in the observed reduction of microbial
diversity [33]. Third, hyperglycemia could have led to the acidification
of the oral environment, enhancing the growth of certain bacterial species at the
expense of others [31,32]. Second, mouth dehydration, usually asso-
ciated with diabetes, could result in the observed reduction of microbial
diversity [33]. Third, hyperglycemia could have led to the acidification
of the oral environment, perturbing the oral microbiota [34]. Similarly,
it was found that oral microbial diversity decreased in diabetics and
increased with progression of periodontal disease compared with perio-
dontally healthy controls [35]. Moreover, in the microbiota of preg-
nant women with gestational diabetes mellitus (GDM). The microbiota
was remarkably altered with a sturdy correlation between bacteria and
the oral glucose tolerance test [36].

Further, of the 15 OTUs exclusive to the IGT group, 7 (46.7%) were
not oral pathogens, 5 (33.3%) were oral/dental pathogens, and 3 (20%)
were beneficial or probiotic bacteria. The oral/dental pathogens ex-
clusive to the IGT group were Enterococcus gallinarum, Streptococcus
genomos, Capnocytophaga sp. strain F0488, Treponema pectinovorum,
and Moraxella caprae (Supplementary Table 4). E. gallinarum has been
implicated in spontaneous bacterial peritonitis [42,43]; S. genomos and
Capnocytophaga sp. strain F0488 are involved in caries and opportun-
istic oral infection, respectively [44,45]; and T. pectinovorum is asso-
ciated with endodontic treatment failure and periapical dental lesions
[46]. Two bacterial species with anti-inflammatory properties were also
identified in this group, namely, Weisella paramesteroides, which pro-
duces bacteriocins against numerous pathogens [47,48], and Bifi-
dobacterium breve.

Of the 26 OTUs exclusive to the diabetes group, 16 (61.54%) were
not oral pathogens and 10 (38.46%) were oral/dental pathogens. No
beneficial probiotic bacteria were identified (Supplementary Table 5).
Among the observed oral pathogens in this group was *Treponema pallidum*, which is highly associated with oral cavity infections in individuals with immunodeficiency virus-infected individuals and is also the pathogen that causes oral syphilis [49,50]. Another bacterium, *Treponema* sp. clone AF026, has been isolated from human periodontal pockets of several subjects exhibiting a range of periodontal conditions [51]. Another identified pathogen, *Lactobacillus ultunensis*, is implicated in advanced dental caries [52,53]. Interestingly, we also identified the non-oral pathogen *Peptoniphilus asaccharolyticus* in this group. This bacterium was reportedly associated with septic arthritis and osteomyelitis in a woman with osteoarthritis and diabetes mellitus, and has been also isolated from diabetic foot infections [54,55].

The normoglycemic and IGT groups shared 78 bacterial species. These transitional species included 10 confirmed oral pathogens (12.82%), 12 potential oral pathogens (15.38%), and 56 non-oral pathogens (71.79%). No beneficial probiotic bacteria were identified (Supplementary Table 6). *Actinomyces israelii* was one of the oral pathogens identified among the normoglycemic/IGT transitional species. It is the causal agent of actinomycosis, oral periapical lesion, and oral-cervical faucal (“lumpy jaw”) infection [56,57]. *Arcanobacterium haemolyticum*, the causal agent of Lemierre syndrome, including oropharyngeal infection [58], and *Brevundimonas diminuta*, the causal agent of refractory periodontitis and advanced Noma lesions [59,60], were also identified.

The IGT and diabetes groups shared eight bacterial species. These transitional species included four confirmed oral pathogens (50%), one potential oral pathogen (12.5%), and three non-oral pathogens (37.5%). As above, no beneficial probiotic bacteria were identified (Supplementary Table 7). *Staphylococcus warneri*, which plays a role in apical periodontitis lesions of obturated tooth and persistent root canal infections [61,62], was one of the identified oral pathogens among the IGT/diabetes transitional species. In addition, *Lepothrix* sp. clone AV011a, the causal agent of advanced Noma lesions and dentine caries [60,63], was also observed, as was the potential oral pathogen *Streptococcus downei*, associated with dental plaque in human [64]. Even though three non-oral pathogens were identified in this transitional set, these bacteria can potentially cause infections at other sites. For example, *Haemophilus paraphrophemolyticus* is the causal agent of liver abscess [65].

The marked increase in the pathogenic content of the hyperglycemic microbiota could be explained by reduced immunity associated with this condition and/or the acidification of saliva (on account of the presence of glucose), since those two factors are known to affect the growth of microorganisms in the oral cavity [66]. Among type 2 diabetes patients, the microbiota distribution was totally different from that in the normoglycemic subjects or IGT patients. This group of patients had the highest pathogenic microbiota content (38.5%), and more non-pathogenic bacteria than IGT patients but less than the control subjects. No probiotic microorganisms were isolated from the oral cavity of these patients. The increase in pathogenic oral microbiota content observed in this group may be explained by the same factors as those affecting the microbiota in the IGT group, although of greater magnitude, namely: reduced immunity and elevated acidity of the oral cavity. The absence of probiotic microorganisms could be associated with the increased prevalence of the pathogenic microbiota, which might be associated with the presence of toxic compounds that would inhibit the probiotic bacteria. Other factors that are related to hyperglycemia may have had contributed to this effect, which requires

### Table 3

<table>
<thead>
<tr>
<th>Number of OTUs</th>
<th>IGT1</th>
<th>IGT2</th>
<th>IGT3</th>
<th>IGT4</th>
<th>IGT5</th>
<th>IGT6</th>
<th>IGT7</th>
<th>IGT8</th>
<th>IGT9</th>
<th>IGT10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individuals</td>
<td>267</td>
<td>251</td>
<td>268</td>
<td>261</td>
<td>288</td>
<td>264</td>
<td>267</td>
<td>249</td>
<td>273</td>
<td>242</td>
</tr>
<tr>
<td>Dominance_D</td>
<td>0.0441</td>
<td>0.07338</td>
<td>0.08687</td>
<td>0.07445</td>
<td>0.06784</td>
<td>0.07375</td>
<td>0.05932</td>
<td>0.08096</td>
<td>0.04439</td>
<td>0.03642</td>
</tr>
<tr>
<td>Evenness_e^H/S</td>
<td>0.1514</td>
<td>0.1204</td>
<td>0.1093</td>
<td>0.1132</td>
<td>0.1332</td>
<td>0.1132</td>
<td>0.1345</td>
<td>0.1726</td>
<td>0.1211</td>
<td>0.1425</td>
</tr>
<tr>
<td>Brillouin</td>
<td>1.115</td>
<td>1.144</td>
<td>1.211</td>
<td>1.008</td>
<td>1.315</td>
<td>1.186</td>
<td>1.359</td>
<td>1.198</td>
<td>1.214</td>
<td>1.134</td>
</tr>
<tr>
<td>Margalef</td>
<td>226</td>
<td>235</td>
<td>267</td>
<td>245</td>
<td>276</td>
<td>248</td>
<td>267</td>
<td>249</td>
<td>273</td>
<td>242</td>
</tr>
<tr>
<td>Equitability_J</td>
<td>0.5693</td>
<td>0.6172</td>
<td>0.5642</td>
<td>0.5883</td>
<td>0.5957</td>
<td>0.6085</td>
<td>0.6461</td>
<td>0.6093</td>
<td>0.6409</td>
<td>0.6816</td>
</tr>
<tr>
<td>Fisher_alpha</td>
<td>35.28</td>
<td>34.82</td>
<td>37.21</td>
<td>31.94</td>
<td>37.39</td>
<td>36.26</td>
<td>42.36</td>
<td>36.73</td>
<td>37.24</td>
<td>34.35</td>
</tr>
<tr>
<td>Chao-1</td>
<td>49.42</td>
<td>50.63</td>
<td>48.53</td>
<td>41.53</td>
<td>57.16</td>
<td>51.16</td>
<td>54.05</td>
<td>51.17</td>
<td>51.55</td>
<td>48.93</td>
</tr>
<tr>
<td>PD-SBL*</td>
<td>49.42</td>
<td>50.63</td>
<td>48.53</td>
<td>41.53</td>
<td>57.16</td>
<td>51.16</td>
<td>54.05</td>
<td>51.17</td>
<td>51.55</td>
<td>48.93</td>
</tr>
</tbody>
</table>

* Phylogenetic Diversity-Total branch length.
further study. In addition, a future pathogenomics study will be required to confirm our observed results on a metagenomic level.

5. Conclusions

In the present study, we observed a clear reduction of the biological and phylogenetic diversity in the diabetes and pre-diabetes oral microbiota in comparison with that in the normoglycemic oral microbiota. However, this reduction was associated with an increase in the pathogenic components of the hyperglycemic microbiota. A future pathogenomics study, i.e., one focusing on the pathogenicity and virulence factor determinants, will be essential for confirming the findings on a metagenomic level.

Ethics approval and consent to participate

This study was approved by the institutional review board of the King Saud University, Collage of Medicine Riyadh, Kingdom of Saudi Arabia. The subjects provided written informed consent for participating in this study.

Consent to publish

All authors have consented for publication of this manuscript.

Availability of data and materials

All sequence data were submitted to NCBI under an SRA accession number SRP159425, and BioSample accessions numbers SAMN09929596–SAMN09929639 in the BioProject PRJNA488297.

Authors’ contributions

ATMS: Involved in study conception and design, and data analysis and interpretation. Involved in drafting of the manuscript and revising it critically for important intellectual content. Prepared the final version to be published. KA: Involved in study conception and design. Prepared the final version to be published. KD: Performed the dental screening and prepared the relevant dental screening report. BM: Performed DNA sequencing and was involved in drafting of the manuscript. URGK: Involved in the acquisition of data, and analysis and interpretation of data. Involved in drafting of the manuscript. MA: Involved in study design. Involved in the acquisition of data, and analysis and interpretation of data. Involved in preparation and drafting of the manuscript. HTT: Involved in drafting of the manuscript and revising it critically for important intellectual content.

Conflicts of interest

The authors declare that they have no competing interests.

Acknowledgement

The authors want to thank the members of the University Diabetes Center at King Saud University for their help throughout the study. Special thanks to Dr. Dhekra Alnakeeb and Mrs. Katherine Anne Uy Saeb for their great effort in coordinating the clinical part of the study. In addition, we want to acknowledge the support of the Saudi Human Genome Program (SHGP) at King Abdulaziz City for Science and Technology and King Faisal Specialist Hospital & Research Center for NGS experiments and analysis.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.micpath.2019.01.009.

Funding

The authors received internal research fund from the King Faisal Specialist Hospital and Research Center to support the publication.

References


228


List of abbreviations

NGS: next-generation sequencing
16S rRNA: 16S ribosomal RNA gene
Mi: mega base pairs
BLAST: Basic Local Alignment Search Tool
AGE: advanced glycation end-product
LPS: lipopolysaccharide
TLR: Toll-like receptor
ESRD: end-stage renal disease
OGT: oral glucose tolerance
ADA: The American Diabetes Association
FBG: fasting blood glucose
NIHANES: National Health and Nutrition Examination Survey
PPD: probing pocket depth
CAL: clinical attachment loss
BOP: bleeding on probing
OTU: operational taxonomic unit
HOMD: Human Oral Microbiome Database
PGM: Ion Personal Genome Machine