



Child type 1 diabetes associated with mother vaginal bacteriome and mycobiome

A. L. Ruotsalainen¹ · M. V. Tejesvi^{1,3} · P. Vänni³ · M. Suokas^{1,5} · P. Tossavainen⁴ · A. M. Pirttilä¹ · A. Talvensaari-Mattila² · R. Nissi²

Received: 10 March 2022 / Accepted: 18 May 2022 / Published online: 14 June 2022
© The Author(s) 2022

Abstract

Mother vaginal microbes contribute to microbiome of vaginally delivered neonates. Child microbiome can be associated with autoimmune diseases, such as type 1 diabetes (T1D). We collected vaginal DNA samples from 25 mothers with a vaginally delivered child diagnosed with T1D and samples from 24 control mothers who had vaginally delivered a healthy child and analyzed bacteriome and mycobiome of the samples. The total DNA of the samples was extracted, and ribosomal DNA regions (16S for bacteria, ITS2 for fungi) were amplified, followed by next-generation sequencing and machine learning. We found that alpha-diversity of bacteriome was increased ($P < 0.002$), whereas alpha-diversity of mycobiome was decreased ($P < 0.001$) in mothers with a diabetic child compared to the control mothers. Beta-diversity analysis suggested differences in mycobiomes between the mother groups ($P = 0.001$). Random forest models were able to effectively predict diabetes and control status of unknown samples (bacteria: 0.86 AUC, fungi: 0.96 AUC). Our data indicate several fungal genera and bacterial metabolic pathways of mother vaginal microbiome to be associated with child T1D. We suggest that early onset of T1D in a child has a relationship with altered mother vaginal microbiome and that both bacteriome and mycobiome contribute to this shift.

Keywords Machine learning · Microbial diversity · Next generation sequencing · Vaginal microbiome

Introduction

Mother vaginal microbiome contributes to the microbial community (i.e., microbiome, including bacteria and fungi) of gut, oral cavities and skin of vaginally delivered children [1–3]. The child gut microbiome, on the other hand, is

associated with autoimmune diseases, such as type 1 diabetes (T1D), during childhood or later in life [4]. Restoration of gut microbiota with *Bifidobacterium infantis* in early life could protect a child from development of T1D [5]. Children born by cesarean section (CS) have altered microbial communities in the gut compared to vaginally delivered children [1, 6, 7] and CS has been considered as a risk factor for the early onset of T1D [8]. Regardless of such associations, the connection between the vaginal microbiome and child microbiome in the development of T1D may be complex [9, 10]. In a case–control-type pilot study, Tejesvi et al. [11] found that the mother vaginal microbiome and child T1D may be associated. The results by Tejesvi et al. suggested that mothers who gave birth to a child with T1D had a more diverse vaginal microbiome than control mothers.

The full microbiome, which has concerned only bacterial species in the large majority of studies, but includes also human fungal communities, i.e., mycobiomes that were long neglected but have lately caught attention [12]. Specifically, the role of human mycobiome and association of the dysbiosis between bacteriome and mycobiome with autoimmune

Edited by Volkhard A.J. Kempf.

✉ A. L. Ruotsalainen
annu.ruotsalainen@oulu.fi

¹ Department of Ecology and Genetics, University of Oulu, POB 3000, 90014 Oulu, Finland

² Department of Obstetrics and Gynecology, University of Oulu, PL 23, FI90029 Oulu, Finland

³ Genobiomics LLC, Oulu, Finland

⁴ Department of Pediatrics, PEDEGO Research Unit and Medical Research Center, University of Oulu and Oulu University Hospital, PO Box 23, 90029 OYS Oulu, Finland

⁵ Biocenter Oulu Sequencing Center, University of Oulu, POB 8000, 90014 Oulu, Finland

diseases has been under investigation [4, 13, 14]. Fungal and bacterial dysbiosis and intestinal inflammation of neonates was recently associated with beta-cell autoimmunity, as children who developed T1D had high amounts of *Saccharomyces* and *Candida* yeasts in their gut [4]. Vaginal mycobiome has been relatively little studied to date. However, studies by Drell et al. [15] and Bradford and Ravel [16] suggest that the vaginal mycobiome is more diverse than previously assumed. Drell et al. noted that the low quality of reference libraries and poor knowledge of fungal taxonomy may complicate the analyses. Nevertheless, interactions between bacterial and fungal communities, especially between *Lactobacillus* spp. and *Candida* spp., have been considered fundamental for future research [16, 17].

We compared bacteriome and mycobiome samples of mothers who had delivered at least one child diagnosed with T1D by the age of 11 ($N=25$, group hereafter named as Diabetes group) with similar samples from control mothers with children without diabetes ($N=24$, group hereafter named as Control group) using next-generation sequencing and machine learning approach. The aim of this study was (1) to find out whether association of mother vaginal microbiome and child T1D found by Tejesvi et al. [11] can be corroborated to a larger sampling from the same research population. We also aimed (2) to find out whether further differences in vaginal mycobiomes exist with respect to the mother groups (Diabetes group vs. Control group). Finally, (3) we analyzed the predicted metabolic pathways to find out whether child T1D-specific signatures were detectable in this study population.

Methods

Description of samples and sampling protocol

Material for this study was collected in the Northern Ostrobothnia Hospital District 2018–2020 covering population in Oulu area, Northern Finland. Institutional Ethics Committee approved the study and the patient consent (Statement of the regional Ethics committee 21.6.2017). All research was performed in accordance with relevant regulations, and informed consent was obtained from all participants. The material for the study was collected from 25 mothers with at least one child with T1D by vaginal delivery (child age less than 11 years at the time of sampling) and 24 control mothers with at least one vaginal delivery and no diagnosed child/children with diabetes. The age of the mothers was set to be between 22 and 40 years at the time of sampling. Mothers included in this study did not have hormonal birth control or constant medication. No other background data on mothers were collected in this study. The samples were collected in a normal protocol of vaginal surface sampling.

Sampling was carried out by two specialist doctors with one mainly taking samples from Diabetes group and one taking samples from Control group, by following exactly same protocol in sampling. After sampling, the tip of the swab was immediately placed into a sterile Eppendorf tube and preserved at $-20\text{ }^{\circ}\text{C}$. Data is deposited in GenBank <https://www.ncbi.nlm.nih.gov/genbank/> under the bioproject numbers PRJNA751475 for bacteria <https://dataview.ncbi.nlm.nih.gov/object/PRJNA751475?reviewer=jia44284elai976f8b5otp063r> and PRJNA751714 for fungi <https://dataview.ncbi.nlm.nih.gov/object/PRJNA751714?reviewer=h1oaitfif938522cmfv3r5frq>

PCR amplification, DNA extraction, and sequencing

DNeasy Power Soil® Pro DNA isolation kit (Qiagen) and Qiacube robotic workstation (Qiagen) were used to extract genomic DNA from frozen cotton swabs to identify bacterial and fungal communities. Before PCR amplification, genomic DNA was diluted to a concentration of 10 ng/ μl and analyzed using a Nanodrop spectrophotometer.

The Primers 519F (5'-CAGCMGCCCGCGGTAATWC-3') and 926R (5'-CCGTCAATTCCTTTRAGTTT-3') were used to amplify a portion of the bacterial 16S small ribosomal unit gene in bacterial Polymerase chain reactions (PCR). The 519F primer had a 30-bp long adapter sequence A, a 9-bp specific barcode sequence for each sample, and a single nucleotide linker A at the beginning of the Ion Torrent sequencing method. The Ion Torrent adapter series trP1 was present at the beginning of the 926R primer. 1 × Phusion Flash High-Fidelity Master Mix (ThermoFisher Scientific), 0,5 M forward and reverse primers, and 10 ng template DNA were used in duplicate polymerase chain reactions in a 15 μl volume. After a 3-min denaturation period at 98 $^{\circ}\text{C}$, the following conditions were used for 22 cycles: 98 $^{\circ}\text{C}$, 10 s; 64 $^{\circ}\text{C}$, 10 s; 72 $^{\circ}\text{C}$, 30 s. The final extension was done for 5 min at 72 $^{\circ}\text{C}$.

With primers ITS4 (5'-TCCTCCGCTTATTGATATGC-3') and fITS7 (5'-GTGARTCATCGAATCTTTG-3'), the ITS2 region of the ribosomal RNA gene was amplified for fungal community analysis. The ITS4 primer had a 30-bp long adapter sequence A and a 10-bp unique multiplex identifier sequence (MID) at the beginning, while the fITS7B primer had a 30-bp long adapter sequence A and a 10-bp unique multiplex identifier sequence (MID) at the beginning. PCR amplification was performed in the same manner as for 16S rRNA products, with an initial denaturation at 98 $^{\circ}\text{C}$ for 2 min, followed by 32 cycles of 10 s at 98 $^{\circ}\text{C}$, 20 s at 54 $^{\circ}\text{C}$, and 30 s at 72 $^{\circ}\text{C}$, and a final extension for 7 min at 72 $^{\circ}\text{C}$.

Both PCR reactions were done in triplicate, and the PCR products were analyzed on an agarose gel. Following that, the triplicate reactions were mixed, cleaned with a Beckman Coulter Agencourt AMPure XP PCR purification system,

and quantified with an Agilent Bioanalyzer using the DNA-1000 analysis package (Agilent). Individual 16S and ITS samples were then pooled at equimolar ratios for sequencing, and the pools were purified with Ampure XP, tested for purity with a bioanalyzer, and concentration was measured with a picogreen assay. Sequencing was performed using Ion Torrent PGM sequencer, 316 v2 chip, Ion PGM Hi-Q View template kit (400 bp templating program) and Ion PGM Hi-Q View Sequencing kit (850 cycles).

Bioinformatics: 16S sequence and ITS sequence preprocessing

Multiplexed 16S sequences were imported into Qiime2 (version 2019.10) [18]. Barcode sequences were removed using the q2-cutadapt-plugin [19]. Primer sequences (f-primer: CAGCMGCCGCGGTAATWC, r-primer: CCGTCAATTCCTTTRAGTTT) were removed using the q2-cutadapt-plugin. Sequences were denoised to ASV's using the q2-dada2-plugin [20] with truncation length parameter set to 391 base pairs. Naive-Bayes taxonomic classifier was trained using the SILVA (v138) database [21] trimmed to the forward and reverse primers used in sequencing and truncated to 391 bp length. Chimeric sequences were detected and removed using the q2-vsearch-plugin [22]. Features found only in one sample and those with less than 10 frequency across all samples were removed. Taxonomy was assigned using the naive-bayes classifier. Non-bacterial ASV's, mitochondria, and chloroplast sequences were removed using the q2-taxa-plugin. At this point, samples that had a total feature frequency lower than 1000 were removed. Additionally, the ASV-table was then collapsed into a taxonomic level of genera and the metabolic pathway composition was predicted using the q2-picrust2-plugin [23]. The q2-picrust-plugin outputted MetaCyc [24] metabolic pathways. *ITS sequences* were preprocessed similarly to 16S sequences, except 311 was chosen as the q2-dada2 truncation length, and the R-package Decontam [25] was used to identify and remove features identified as contaminants. Taxonomy was assigned to ITS features using the UNITE (v8.2) database [26].

Diversity analyses and differential abundance

Alpha and beta diversity analyses were done using the q2-diversity-plugin and visualized with Matplotlib python package. Shannon index and Bray–Curtis dissimilarity were chosen as diversity metrics. ASV-tables were rarefied to the sampling depth of 1000, while metabolic pathway data were rarefied to a depth of 10,000. Principal coordinates analysis (PCoA) was performed using the q2-diversity-plugin with bacteria, predicted pathways and fungi data independently. Statistical differences between diabetes and control samples were tested with Kruskal–Wallis H-test

with Scipy and PERMANOVA with q2-diversity-plugin in alpha and beta diversity, respectively. Statistical differences in comparison of individual taxonomic groups were carried out using Kruskal–Wallis H-test in R environment (version 4.1.0, [27]). Differentially abundant genera and predicted pathways were investigated using the q2-aldex2-plugin [28].

Machine learning

Random forest [29] and logistic regression models were trained to predict the diabetes status of the samples using scikit-learn package [30]. Nested cross-validation scheme with tenfolds in each layer was used. In tenfold cross-validation, the whole data are first partitioned to ten different validation and training folds, where each sample is once in the validation fold. In nested cross-validation, a second tenfold cross-validation split is done on the training fold to tune optimal parameters. This way, the validation fold is unseen to the training process of the models. Default parameters were used for random forests, while parameter “C” was tuned using the training folds for logistic regression models. Feature importance values were gathered during model training, where random forest models outputted the normalized gini importance and logistic regression the feature coefficients. Area under the curve (AUC) of receiver operating characteristic (ROC) were chosen as the performance metric as it performs well with class unbalanced data. Nested cross-validation process was repeated 40 times and the model performances, feature importance's and coefficient values from each cross-validation iteration were pooled together, averaged, and finally plotted using Matplotlib.

Results

The total number of raw bacterial sequences was 1 526 318 and fungal sequences 1 095 814. The total frequency, after all quality filtering steps, of bacterial sequences steps was 435 993 (collapsed into 34 genera), 23 860 244 (290 pathways) predicted metabolic pathways, and 144 424 (collapsed into 25 genera) fungal sequences. The bacterial and predicted pathway data had 49 samples in the final analyses, while 44 samples remained in fungal data. The relative abundances of taxonomically assigned sequence reads for bacteria and fungi are presented in Fig. 1 and Supplementary Tables 1 and 2, respectively. The raw sequences were deposited to GenBank under the bioproject numbers PRJNA751475 for bacteria and PRJNA751714 for fungi.

Alpha-diversity (Shannon, diversity within samples) of the vaginal bacteriome of Diabetes group mothers was higher compared to the Control group mothers (Fig. 2A). In contrast, the alpha-diversity of vaginal mycobiome was lower in the Diabetes group than in the Control group

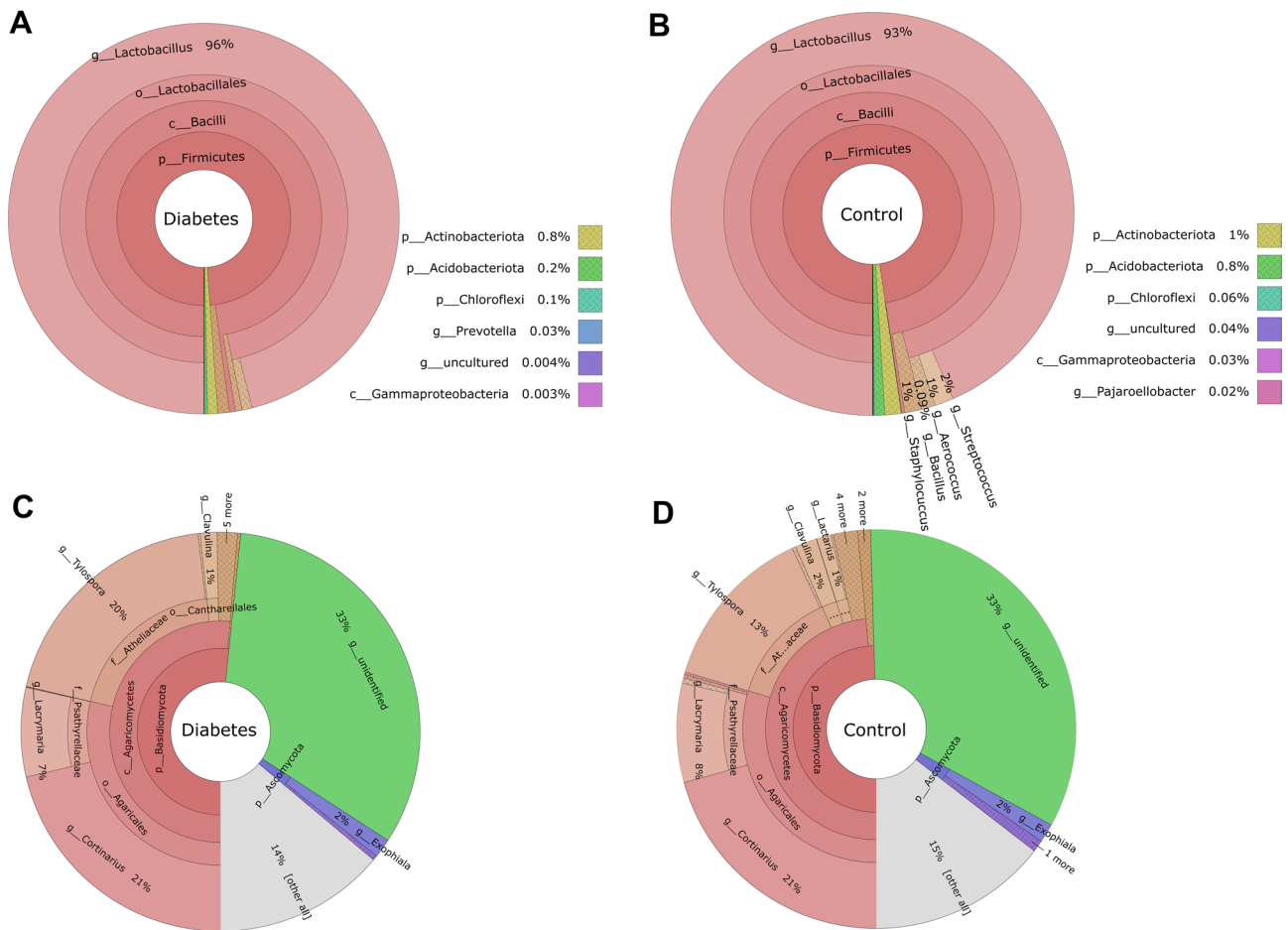


Fig. 1 Averaged relative abundances of bacterial (**A, B**) and fungal (**C, D**) genera in Diabetes and Control groups

(Fig. 2B). Alpha-diversity of bacterial predicted pathways also tended to increase, but the result was not statistically significant ($P=0.08$, Supplementary Fig. 1). We found a difference in beta-diversity (Bray–Curtis dissimilarity, diversity between samples) in the mycobiomes between Diabetes and Control groups (Fig. 2C). In the case of bacteriomes and predicted pathways, there were no statistically significant differences in beta-diversity (Supplementary Fig. 2).

Machine learning models differentiated well between Diabetes and Control group, where random forest (RF) models achieved high area under the curve (AUC) when using both bacteria (AUC = 0.86, SD = 0.03) and fungi (AUC = 0.93, SD = 0.02) (Fig. 3). Logistic regression (LOG) models had lower AUC values for all types of data, except in predicted metabolic pathway data derived from 16S sequences (0.86 AUC in both bacteria and fungi) (Fig. 3). Machine learning indicated several genera to be characteristic for Diabetes group both in separate (Fig. 4) and combined analyses (Supplementary Fig. 4) of bacterial and fungal genera. When bacteria and fungi data were combined for machine learning analyses, both

RF (AUC = 0.96, SD = 0.02) and LOG (AUC = 0.93, SD = 0.04) models could predict with high accuracy between Diabetes and Control group test samples (Supplementary Fig. 3). Abundance data indicated several fungal genera and bacterial metabolic pathways to be associated with child T1D (Tables 1 and 2).

An amplified sequence variant (ASV) classified within the fungal genus *Tylospora* had the highest feature importances (MDI value) used in the random forest models to predict the differences between the mother groups (Fig. 4). The abundance of *Tylospora* sp. was also significantly higher in the Diabetes group samples ($P=0.001$, Table 2, Supplementary Fig. 6, Supplementary Table 2). The bacterial genus *Aerococcus* was present at a higher frequency in Diabetes group samples than in Control group samples (96% vs 33% of samples, respectively) although an abundance of *Aerococcus* did not differ between groups (data not shown). Considering predicted metabolic pathways of bacteria, PWY-6630 (superpathway of L-tyrosine biosynthesis), PWY-6608 (guanosine nucleotides degradation III) and PWY-6628 (superpathway of L-phenylalanine biosynthesis) were the

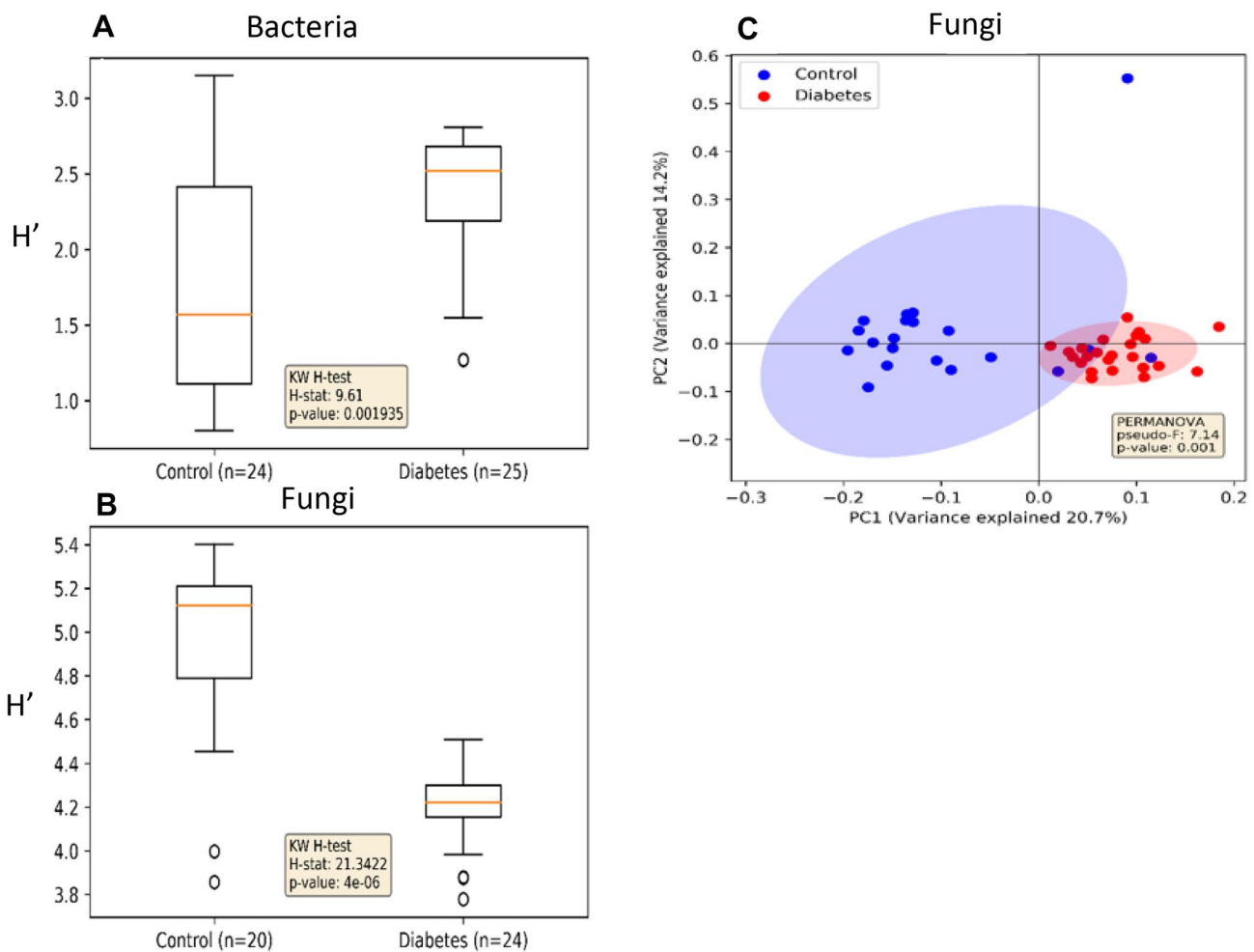


Fig. 2 Alpha and beta diversity of bacteria and fungi in vaginal samples. Alpha diversity boxplots of Shannon's diversity indices for Diabetes and Control group samples of **A** bacteria and **B** fungi. **C** Beta-diversity of fungal community in Diabetes and Control group samples

most important ones to predict the difference between Diabetes and Control groups (Fig. 4; Table 1).

Discussion

We found differences in vaginal bacteriome and mycobiome between mothers who had a child with T1D (Diabetes group) and mothers with non-diabetic children (Control group). Alpha diversity of mother vaginal bacteriome was higher among the Diabetes group compared to the Control group. Increased alpha diversity, i.e., higher within-sample variation of the vaginal microbiome (bacteriome) of mother, related to the child T1D, has been previously reported by Tejesvi et al. [11]. In addition, observations of increased diversity in child's bacterial gut microbiota in relation to maternal gestational diabetes have been observed by Wang et al. [12]. Accordingly, Vatanen et al. [31] reported increased diversity of child's bacterial gut microbiota toward

T1D onset in childhood. The relationship between increased diversity in mother's vaginal bacteriome and child's T1D suggests that there exist unidentified vaginal bacterial groups that play a role in development of T1D.

Tejesvi et al. [11] reported that within-genus beta-diversity of *Lactobacillus* was altered in the group of mothers with T1D children, and the bacterial genus *Prevotella* was characteristic to this group. High amounts of *Prevotella* have been found in vaginal microbiomes of women who have other than *Lactobacillus*-dominated microbiome, but *Prevotella* has also been linked to bacterial vaginosis [32]. In this study, *Prevotella* was absent from Control group samples and present in Diabetes group samples (at an average of 3% abundance), but this difference was not statistically significant. Drell et al. [15] found that increased microbiome diversity was linked to an increase in vaginal pH. In this study, we found a bacterial ASV classified as *Aerococcus* (Firmicutes, Bacilli) characterizing the difference between the Diabetes and Control group mothers. *Aerococcus*, which

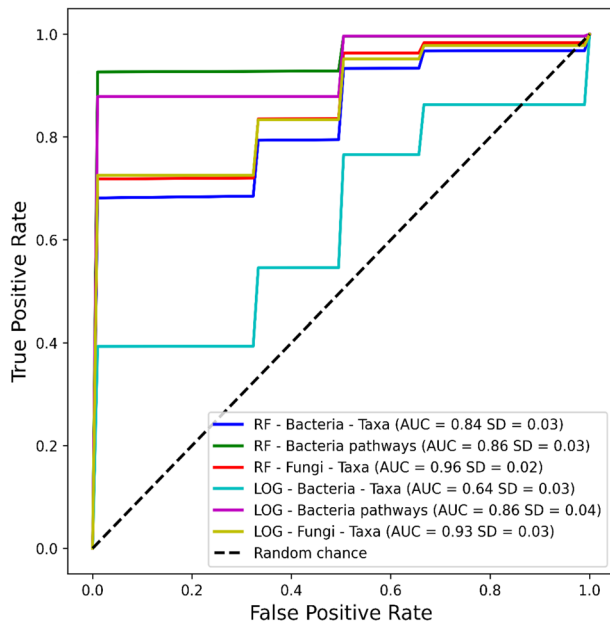


Fig. 3 Cross-validated machine learning model performance when differentiating Diabetes and Control samples in the test samples. Dotted black line represents the performance of a model that is completely random (0.5 area under the curve, AUC), while a model that is always correct would have an AUC of 1.0

is a common member of air and vegetation microbial communities, is often found in the vaginal bacteriome, but it is also associated with bacterial vaginosis and urinary tract infections [33]. Unhealthy vaginal bacteriome of mother has been found to be linked to offspring health in a mouse model [34]. Whether changes in certain bacterial groups, such as in *Lactobacillus*, *Prevotella* or *Aerococcus* in mother vaginal microbiome are associated with child T1D, warrants further research.

We also predicted specific bacterial metabolic pathways in vaginal microbiomes that were associated with child T1D. Specifically, the predicted bacterial metabolic pathways characterizing differences between Diabetes and Control groups were related to amino acid biosynthesis and nucleotide degradation. The combination of these pathways may imply of direction of nitrogen usage within the bacterial communities, as nitrogen is specifically required for both amino acid and nucleotide synthesis. Because amino acids are building blocks for proteins, required for metabolic activity, and nucleotides are needed for replication of DNA and proliferation [35], the result suggests that bacteria were directing their nitrogen use toward survival and not cell division.

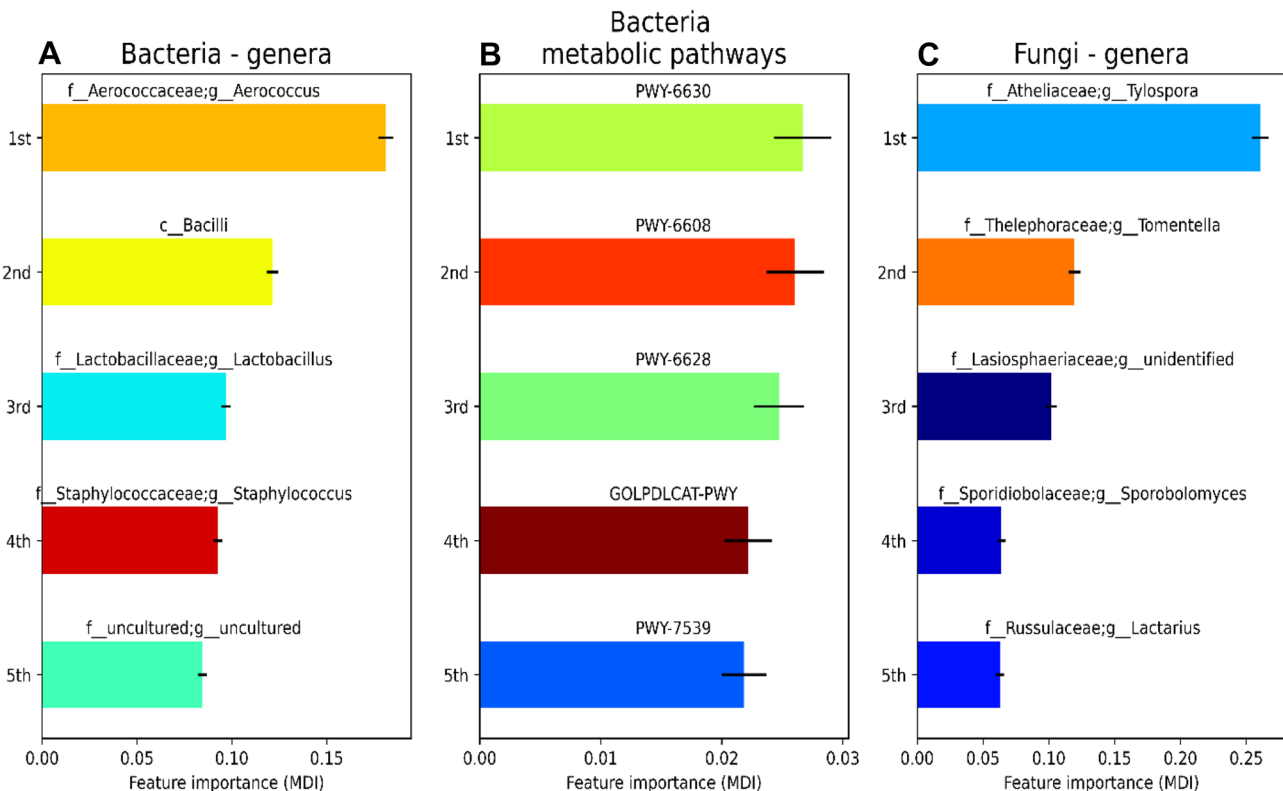


Fig. 4 Importances of features used by random forest models (separate analyses for bacteria and fungi). Mean Decrease in Impurity (MDI) was used as the feature importance metric. Models were

trained to predict unknown samples of Diabetes and Control groups in **A** bacteria, **B** pathways, and **C** fungi

Table 1 Predicted metabolic pathway features in Diabetes and Control groups

| Feature name | Diabetes vs Control Adjusted <i>p</i> -value | Effect Direction |
|----------------|--|------------------|
| ASPASN-PWY | 0.045 | 0.705 |
| CENTFERM-PWY | 0.049 | 0.577 |
| HSERMETANA-PWY | 0.031 | 0.498 |
| PWY-5695 | 0.011 | 0.943 |
| PWY-6507 | 0.033 | 0.635 |
| PWY-6590 | 0.035 | 0.591 |
| PWY-6608 | 0.004 | 1.058 |
| PWY-6628 | 0.001 | 1.106 |
| PWY-6630 | 0.001 | 1.097 |
| PWY-6891 | 0.04 | 0.655 |
| PWY-6892 | 0.04 | 0.685 |
| PWY-6895 | 0.02 | 0.697 |
| PWY-6901 | 0.024 | 0.612 |
| RHAMCAT-PWY | 0.036 | 0.592 |
| THISYN-PWY | 0.016 | 0.749 |

Effect: direction of change. Positive value: feature is more abundant in Diabetes group

Table 2 Aldex2 results for Fungi genera feature table

| Feature name | Diabetes vs Control Adjusted <i>p</i> -value | Effect Direction |
|-------------------|--|------------------|
| K_Fungi;_:_:_:_:_ | 0.02 | 0.757 |
| G_Exophiala | 0.04 | 0.623 |
| G_Cortinarius | 0.006 | 0.889 |
| G_Lacrymaria | 0.026 | 0.593 |
| G_Tylospora | 0.001 | 1.194 |
| G_Tomentella | 0.044 | -0.69 |
| G_unidentified | 0.022 | 0.745 |

Effect=direction of change. Positive value=feature, i.e., fungal group is more abundant in Diabetes group

In contrast to the bacteriome, alpha-diversity of mycobiome was lower in Diabetes group mothers compared to the Control group mothers. Beta diversity of the mycobiomes also differed between the groups. Decreased diversity of vaginal mycobiome and altered community structure (beta-diversity) suggest that the vaginal mycobiome may play a more significant role in the full microbiome functions and, consequently, in the child’s gut microbiome, than has previously been thought. In general, recent findings have indicated that vaginal mycobiome is more diverse than previously thought [15, 16]. According to Hall & Noverr [36], the majority of vaginal fungi are opportunists, which suggests that they are sensitive to the living conditions of the vagina. Fungal communities of vaginally delivered children are also

affected by environmental conditions, such as fungal dispersal via air, living environment, or caretakers [37]. In our study, the decreased diversity was related to several fungal genera/groups that were associated with vaginal mycobiome of mothers in the Diabetes group. In particular, *Tylospora* sp. was highly specific to the Diabetes group. Because the human mycobiomes are relatively little studied to date [15, 16], the importance of this finding remains to be discovered. This is due to the low capacity for taxonomic identification in the reference libraries, which can be expected to increase as more fungal reference samples of humans are identified [15]. However, regardless of poor classification of the taxa, identification of specific fungal groups in the mother vaginal mycobiome and changes in fungal diversity indices clearly indicate that fungal community changes may play a role in early onset of T1D. Dysbiosis, i.e., a change in the balance between gut bacteriome and mycobiome is reported to play a role in the onset of T1D in child gut microbiome [4]. We suggest that a similar kind of interaction between bacterial and fungal communities in mother’s vagina (as suggested also by Bradford & Ravel for vaginal communities in general [16]) may take place and that the altered microbiome is associated with the onset of T1D in a child [17].

Machine learning studies have shown that type 1 and type 2 diabetes can be predicted based on gut microbiome using both 16S and whole genome sequencing data [38–41]. In these studies, the machine learning models achieved moderate prediction performance in the range of 0.7–0.8 AUC, except for a deep learning-based model that achieved 0.9 AUC in a cohort of European women with type 2 diabetes [39]. In our study, the best models predicted T1D with a high performance of 0.86–0.96 AUC, indicating that the vaginal microbiome and mycobiome of the mother is a reliable predictor for T1D in children.

Our study has been carried out in one local population at maximum 11 years after the child delivery. Our findings therefore need to be corroborated by a longitudinal study, where the microbial sampling would be carried out at the time of birth and linked with a follow-up of child cohorts, similar to large-scale child diabetes studies (e.g., [42]), preferably covering several populations (countries and continents). In addition, our results rely on current microbial reference libraries, which in the case of humans, are changing fast (<https://www.arb-silva.de>, [43]). It is known that hormonal contraceptives may change the vaginal microbiota [44], and for this reason, the use of hormonal contraceptives was excluded in our study. None of the mothers had constant medication, or hormonal contraceptives, but we suggest that factors such as frequency of intercourse, glycemic index of food and the quality of hygiene products need to be taken into account in future studies. Vaginal microbiomes during childhood, reproductive-age and menopause are different also suggesting

that within reproductive age vaginal microbiome is relatively stable [45, 46]. However, antibiotic use, gestation and mother hormonal status affect vaginal microbiome in short term [47, 48]. During the last trimester of pregnancy microbial diversity has been found to increase toward delivery [3].

Taken together, we have detected systematic variation in vaginal microbial communities between mothers with or without a child with diabetes in a one-population study. Although the current research has indicated that mother's microbial communities may play a smaller role in determining the general microbiome of a neonate/child than have been previously assumed [3, 37], our results suggest that mother's vaginal microbiome, even collected years after childbirth, may be linked with the development of T1D. This yet unknown link may act via vaginal hormonal status [49]. As our study was carried out in a relatively small geographical area, the results may be different in a wider population, and our observations should therefore be confirmed in a larger, cross-population study, where vaginal swab samples are preserved at the time of delivery.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00430-022-00741-w>.

Acknowledgements We thank Päivikki and Sakari Sohlberg foundation for funding this study and we thank cordially all the volunteers providing samples for this study.

Author contributions R.N., A.T.-M. and P.T. coordinated sampling, M.T. and M.S. carried out laboratory and sequencing analyses, P.V. carried out bioinformatics and machine learning, A.L.R. produced the first version and coordinated manuscript writing with the help of all authors. All authors approved the final version of the manuscript.

Funding Open Access funding provided by University of Oulu including Oulu University Hospital. Päivikki and Sakari Sohlbergin säätiö (foundation) has funded this study (for R.N.).

Data availability statement Data is deposited in GenBank <https://www.ncbi.nlm.nih.gov/genbank/> under the bioproject numbers PRJNA751475 for bacteria <https://dataview.ncbi.nlm.nih.gov/object/PRJNA751475?reviewer=jia44284elai976f8b5otp063r> and PRJNA751714 for fungi <https://dataview.ncbi.nlm.nih.gov/object/PRJNA751714?reviewer=h1oaitfif938522cmfv3r5frq>

Declarations

Conflict of interest The authors declare no conflict of financial or non-financial interest.

Ethical standards This study was performed in line with the principles of the Declaration of Helsinki. Institutional Ethics Committee of Northern Ostrobothnia Hospital District approved the study and the patient consent (Statement of the regional Ethics committee 21.6.2017).

Informed consent Informed consent was obtained from all individual participants included in the study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Dominquez-Bello MG, Dominguez-Bello EK, Costello M, Contreras MM, Glida H, Fierer N et al (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *PNAS* 107:11971–11975. <https://doi.org/10.1073/pnas.1002601107>
- Stewart CJ, Ajami NJ, Jacqueline L, O'Brien JL, Diane S, Hutchinson DS, Smith DP, Wong MC et al (2018) Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562:583–588. <https://doi.org/10.1038/s41586-018-0617-x>
- Rasmussen MA, Thorsen J, Dominguez-Bello MG, Blaser MS, Mortensen AD, Breyndrod SA et al (2020) Ecological succession in the vaginal microbiota during pregnancy and birth. *ISME J* 14:2325–2335. <https://doi.org/10.1038/s41396-020-0686-3>
- Honkanen J, Vuorela A, Muthas D, Orivuori L, Luopajarvi K, Tejesvi MVG et al (2020) Fungal dysbiosis and intestinal inflammation in children with beta-cell autoimmunity. *Front Immunol.* <https://doi.org/10.3389/fimmu.2020.00468>
- Insel R, Knip M (2018) Prospects for primary prevention of type 1 diabetes by restoring a disappearing microbe. *Pediatr Diabetes* 19:1400–1406. <https://doi.org/10.1111/pedi.12756>
- Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C (2010) Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev* 86(suppl):13–15. <https://doi.org/10.1016/j.earlhumdev.2010.01.004>
- Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS et al (2013) Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ* 185:385–394. <https://doi.org/10.1503/cmaj.121189>
- Cardwell CR, Stene LC, Joner G, Cinek O, Svensson J, Goldacre MJ et al (2008) Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia* 51:726–735. <https://doi.org/10.1007/s00125-008-0941-z>
- Stinson LF, Payne MS, Keelan JA (2018) A critical review of the bacterial baptism hypothesis and the impact of cesarean delivery on the infant microbiome. *Front Med.* <https://doi.org/10.3389/fmed.2018.00135>
- Tanoey J, Gulati A, Patterson C, Becher H (2019) Risk of type 1 diabetes in the offspring born through elective or non-elective caesarean section in comparison to vaginal delivery: a meta-analysis of observational studies. *Curr Diab Rep.* <https://doi.org/10.1007/s11892-019-1253-z>
- Tejesvi MV, Nissi R, Saravesi K, Pirttilä AM, Markkola A, Talvensaaari-Mattila A et al (2019) Association of prevalent vaginal microbiome of mother with occurrence of type I diabetes in child. *Sci Rep.* <https://doi.org/10.1038/s41598-018-37467-w>
- Wang J, Zheng J, Shi W, Du N, Xu X, Zhang Y et al (2018) Dysbiosis of maternal and neonatal microbiota associated with

- gestational diabetes mellitus. *Gut* 67:1614–1625. <https://doi.org/10.1136/gutjnl-2018-315988>
13. Hoarau G, Mukherjee PK, Gower-Rousseau C, Hager C, Chandra J, Retuerto MA et al (2016) Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. *MBio*. <https://doi.org/10.1128/mBio.01250-16>
 14. Kowalewska B, Zorena K, Szmigiero-Kawko M, Wąz P, Myśliwiec M (2016) Higher diversity in fungal species discriminates children with type 1 diabetes mellitus from healthy control. *Patient Prefer Adherence* 10:591–599. <https://doi.org/10.2147/PPA.S97852>
 15. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspõllu A, Väin E et al (2013) Characterization of the vaginal micro- and mycobiome in asymptomatic reproductive-age Estonian women. *PLoS One*. <https://doi.org/10.1371/journal.pone.0054379>
 16. Bradford LL, Ravel J (2017) The vaginal mycobiome: A contemporary perspective on fungi in women's health and diseases. *Virulence* 8:342–351. <https://doi.org/10.1080/21505594.2016.1237332>
 17. Siljander H, Honkanen J, Knip M (2019) Microbiome and type 1 diabetes. *Lancet* 46:512–521. <https://doi.org/10.1016/j.ebiom.2019.06.031>
 18. Bolyen E, Rideout JR, Dillon MR, Bokulich NB, Abnet CC, Al-Ghalith GA et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 27:852–857. <https://doi.org/10.1038/s41587-019-0209-9>
 19. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>
 20. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>
 21. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>
 22. Rognes T, Flouri T, Nichols B, Quince C, Mahe F (2016) VSEARCH: A versatile open source tool for metagenomics. *PeerJ*. <https://doi.org/10.7717/peerj.2584>
 23. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM et al (2020) PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol* 38:685–688. <https://doi.org/10.1038/s41587-020-0548-6>
 24. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM et al (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 42:D459–D471. <https://doi.org/10.1093/nar/gkv1164>
 25. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ (2018) Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*. <https://doi.org/10.1186/s40168-018-0605-2>
 26. Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D et al (2019) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res* 47:D259–D264. <https://doi.org/10.1093/nar/gky1022>
 27. R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
 28. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB (2013) ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-seq. *PLoS One*. <https://doi.org/10.1371/journal.pone.0067019>
 29. Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
 30. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B (2011) Scikit-learn: Machine learning in python. *J Mach Learn Res* 12:2825–2830
 31. Vatanen T, Franzosa EA, Schwager R, Tripathi S, Arthur TD, Vehik K et al (2018) The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature* 562:589–594. <https://doi.org/10.1038/s41586-018-0620-2>
 32. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL (2011) Vaginal microbiome of reproductive-age women. *PNAS* 108(Supplement 1):4680–4687. <https://doi.org/10.1073/pnas.1002611107>
 33. Lewis AL, Gilbert NM (2020) Roles of the vagina and the vaginal microbiota in urinary tract infection: evidence from clinical correlations and experimental models. *GMS Infect Dis*. <https://doi.org/10.3205/id000046>
 34. Madigan MT, Bender KS, Buckley DH, Sattley WM, Stahl DA (2018) Brock Biology of Microorganisms. 15th Global Edition. Benjamin Cummings, Boston, US
 35. Jašarević E, Hill EM, Kane PJ, Rutt L, Gyles T, Folts L et al (2021) The composition of human vaginal microbiota transferred at birth affects offspring health in a mouse model. *Nat Commun*. <https://doi.org/10.1038/s41467-021-26634-9>
 36. Hall RA, Noverr MC (2017) Fungal interactions with the human host: exploring the spectrum of symbiosis. *Curr Opin Microbiol* 40:58–64. <https://doi.org/10.1016/j.mib.2017.10.020>
 37. Ward TL, Dominguez-Bello MG, Heisel T, Al-Ghalith G, Knights D, Gale C (2018) Development of the human mycobiome over the first month of life and across body sites. *mSystems*. <https://doi.org/10.1128/mSystems.00140-17>
 38. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G et al (2021) Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. *Genome Biol*. <https://doi.org/10.1186/s13059-021-02306-1>
 39. Oh M, Zhang L (2020) DeepMicro: deep representation learning for disease prediction based on microbiome data. *Sci Rep*. <https://doi.org/10.1038/s41598-020-63159-5>
 40. Gou W, Ling C-W, He Y, Jiang Z, Fu Y, Xu F et al (2021) Interpretable machine learning framework reveals robust gut microbiome features associated with type 2 diabetes. *Diabetes Care* 44:358–366. <https://doi.org/10.2337/dc20-1536>
 41. Pasolli E, Truong DT, Malik F, Waldron L, Segata N (2016) Machine learning meta-analysis of large metagenomic datasets: tools and biological insights. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1004977>
 42. Knip M (2021) Type 1 diabetes in Finland: past, present, and future. *Lancet Diabetes Endocrinol* 9:259–260. [https://doi.org/10.1016/S2213-8587\(21\)00074-7](https://doi.org/10.1016/S2213-8587(21)00074-7)
 43. Sierra MA, Li Q, Pushalkar S, Paul B, Sandoval TA, Kamer AR et al (2020) The Influences of Bioinformatics Tools and Reference Databases in Analyzing the Human Oral Microbial Community. *Genes*. <https://doi.org/10.3390/genes11080878>
 44. Achilles SL, Austin MN, Meyn LA, Mhlanga F, Chirenje ZM, Hillier SL (2018) Impact of contraceptive initiation on vaginal microbiota. *Am J Obstet Gynecol* 218:622.e1–622.e10
 45. Huang B, Fettweis JM, Brooks JP, Jefferson KK, Buck GA (2014) The changing landscape of the vaginal microbiome. *Clin Lab Med* 34:747–761. <https://doi.org/10.1016/j.cll.2014.08.006>
 46. Auriemma RS, Sciarati R, Del Vecchio G, Liccardi A, Verde N, Pirchio R et al (2021) The vaginal microbiome: a long urogenital colonization throughout woman life. *Front Cell Infect Microbiol* 11:686167. <https://doi.org/10.3389/fcimb.2021.686167>
 47. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Lorraine N et al (2014) The composition and stability of the vaginal microbiota of normal pregnant women is different from that of

- non-pregnant women. *Microbiome* 2:4. <https://doi.org/10.1186/2049-2618>
48. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A et al (2015) Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A* 112:11060–11065. <https://doi.org/10.1073/pnas.1502875112>
49. Farage MA, Miller KW, Sobel JD (2010) Dynamics of the vaginal ecosystem—hormonal influences. *Infect Dis* 3:1–15. <https://doi.org/10.4137/IDRT.S3903>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.