



The dynamics of the gut microbiota in prediabetes during a four-year follow-up among European patients-an IMI-DIRECT prospective study

Lyu, L.; Fan, Y.; Vogt, J.K.; Clos-Garcia, M.; Bonnefond, A.; Pedersen, H.K.; ... ; Pedersen, O.

Citation

Lyu, L., Fan, Y., Vogt, J. K., Clos-Garcia, M., Bonnefond, A., Pedersen, H. K., ... Pedersen, O. (2025). The dynamics of the gut microbiota in prediabetes during a four-year follow-up among European patients-an IMI-DIRECT prospective study. *Genome Medicine*, 17(1). doi:10.1186/s13073-025-01508-7

Version: Publisher's Version

License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/4290266>

Note: To cite this publication please use the final published version (if applicable).

RESEARCH

Open Access



The dynamics of the gut microbiota in prediabetes during a four-year follow-up among European patients—an IMI-DIRECT prospective study

Liwei Lyu^{1,2†}, Yong Fan^{1†}, Josef Korbinian Vogt^{1,3†}, Marc Clos-Garcia^{1†}, Amelie Bonnefond^{4†}, Helle Krogh Pedersen^{3†}, Avirup Dutta¹, Robert Koivula^{5,6}, Sapna Sharma^{7,8}, Kristine Højgaard Allin^{1,9}, Caroline Brorsson^{10,11}, Henna Cederberg¹², Elizaveta Chabanova¹³, Federico De Masi^{10,11}, Emmanouil Dermitzakis¹⁴, Petra J. Elders¹⁵, Marieke T. Blom¹⁵, Monika Hollander¹⁵, Rebeca Eriksen¹⁶, Ian Forgie¹⁷, Gary Frost¹⁶, Giuseppe N. Giordano¹⁸, Harald Grallert^{8,19}, Mark Haid²⁰, Tue Haldor Hansen^{1,21,22}, Bernd Jablonka²³, Tarja Kokkola¹², Anubha Mahajan²⁴, Andrea Mari²⁵, Timothy J. McDonald^{26,27}, Petra B. Musholt²⁸, Imre Pavo²⁹, Cornelia Prehn²⁰, Martin Ridderstråle³⁰, Hartmut Ruetten³¹, Leen M.t Hart³², Jochen M. Schwenk³³, Evelina Stankevič¹, Henrik S. Thomsen³⁴, Jagadish Vangipurapu¹², Henrik Vestergaard^{1,35}, Ana Viñuela¹⁷, Mark Walker³⁶, Torben Hansen¹, Allan Linneberg⁹, Henrik Bjørn Nielsen³, Søren Brunak^{10,37}, Mark I. McCarthy^{5,24,38}, Philippe Froguel^{5,39}, Jerzy Adamski^{40,41,42}, Paul W. Franks¹⁸, Marku Laakso¹², Joline W. J. Beulens³², Ewan Pearson¹⁷ and Oluf Pedersen^{1,2*}

Abstract

Background Previous case–control studies have reported aberrations of the gut microbiota in individuals with prediabetes. The primary objective of the present study was to explore the dynamics of the gut microbiota of individuals with prediabetes over 4 years with a secondary aim of relating microbiota dynamics to temporal changes of metabolic phenotypes.

Methods The study included 486 European patients with prediabetes. Gut microbiota profiling was conducted using shotgun metagenomic sequencing and the same bioinformatics pipelines at study baseline and after 4 years. The same phenotyping protocols and core laboratory analyses were applied at the two timepoints. Phenotyping included anthropometrics and measurement of fasting plasma glucose and insulin levels, mean plasma glucose and insulin under an oral glucose tolerance test (OGTT), 2-h plasma glucose after an OGTT, oral glucose insulin sensitivity index, Matsuda insulin sensitivity index, body mass index, waist circumference, and systolic and diastolic blood pressure. Measures of the dynamics of bacterial microbiota were related to concomitant changes in markers of host metabolism.

[†]Liwei Lyu, Yong Fan, Josef Korbinian Vogt, Marc Clos-Garcia, Amelie Bonnefond, and Helle Krogh Pedersen are equally contributing first authors.

*Correspondence:
Oluf Pedersen
oluf@sund.ku.dk

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

Results Over 4 years, significant declines in richness were observed in gut bacterial and viral species and microbial pathways accompanied by significant changes in the relative abundance and the genetic composition of multiple bacterial species. Additionally, bacterial-viral interactions diminished over time. Despite the overall reduction in bacterial richness and microbial pathway richness, 80 dominant core bacterial species and 78 core microbial pathways were identified at both timepoints in 99% of the individuals, representing a resilient component of the gut microbiota. Over the same period, individuals with prediabetes exhibited a significant increase in glycemia and insulinemia alongside a significant decline in insulin sensitivity. Estimates of the gut bacterial microbiota dynamics were significantly correlated with temporal impairments in host metabolic health.

Conclusions In this 4-year prospective study of European patients with prediabetes, the gut microbiota exhibited major changes in taxonomic composition, bacterial species genetics, and microbial functional potentials, many of which paralleled an aggravation of host metabolism. Whether the temporal gut microbiota changes represent an adaptation to the progression of metabolic abnormalities or actively contribute to these in prediabetes cases remains unsettled.

Trial registration The Diabetes Research on Patient Stratification (DIRECT) study, an exploratory observational study initiated on October 15, 2012, was registered on ClinicalTrials.gov under the number NCT03814915.

Keywords Long-term dynamics, Gut bacterial microbiota, Gut viral microbiota, Microbial functional pathways, Gut bacterial genetics, Prediabetes, Metabolism, Insulin sensitivity

Background

Prediabetes is a condition characterized by elevated blood glucose levels that fall below the threshold for a diabetes mellitus diagnosis [1]. In 2021, the prevalence of prediabetes was estimated at 9.1% for impaired glucose tolerance, affecting approximately 464 million individuals, and 5.8% for impaired fasting glucose, affecting 298 million people worldwide [2]. Even though prediabetes is not classified as a disease, it is often associated with obesity, hypertension and dyslipidemia, and with elevated plasma concentrations of triglycerides and/or low plasma concentrations of HDL cholesterol. Prediabetes is an important risk factor for the development of type 2 diabetes mellitus (T2D) and ischemic heart disease [1].

Prediabetes may, however, be a reversible metabolic condition. Some progress to incident T2D, while others either remain prediabetic or regress to normal glucose metabolism [3]. Therefore, understanding the factors that influence the trajectory of prediabetes is essential for the attempts to develop effective prevention strategies.

One factor that may influence the trajectory of prediabetes is the gut microbiota that in several independent cross-sectional studies have been reported to show gut dysbiosis in prediabetes cases when compared with matched healthy individuals [4–6]. Still, longitudinal studies are needed to capture gut microbial features that are stable or variable over time and to assess their associations with host metabolic fluctuations.

To date, studies of the dynamics of the gut microbiota have primarily been conducted in healthy populations [7–9]. For instance, a 1-year prospective study of 75

healthy Swedish individuals found that intra-individual variation accounted for 23% of gut microbiota variance, with lower variability linked to higher abundances of gut bacteria related to metabolic health such as *Faecalibacterium prausnitzii* and *Bifidobacterium* species [8]. In another gut microbiota study of 338 individuals from the Netherlands followed for 4 years, a microbial fingerprint was identified. This approach achieved up to 85% accuracy in classifying microbiota samples taken 4 years apart [7]. These studies offer important insights into gut microbiota dynamics in healthy individuals or focus primarily on broad population-level trends, rather than examining specific metabolic health associations in at-risk groups.

Here, we conducted a 4-year longitudinal study involving 486 European adults with prediabetes as part of the Innovative Medicines Initiative Diabetes Research on Patient Stratification (IMI-DIRECT) Project [10]. Gut microbiota profiling of stool samples collected at two timepoints with 4 years apart was conducted using the same metagenomic sequencing protocol and identical pipelines for data processing, ensuring consistency across all samples. Bioclinical assessments following the same protocols were performed at study baseline and after 4 years.

The primary aim of our study was to characterize the long-term dynamics of the gut microbiota in prediabetic individuals and the secondary aim was to associate the dynamics of the gut microbiota with temporal changes of host metabolic variables. An overview of the study design and main outcomes is given in Fig. 1.

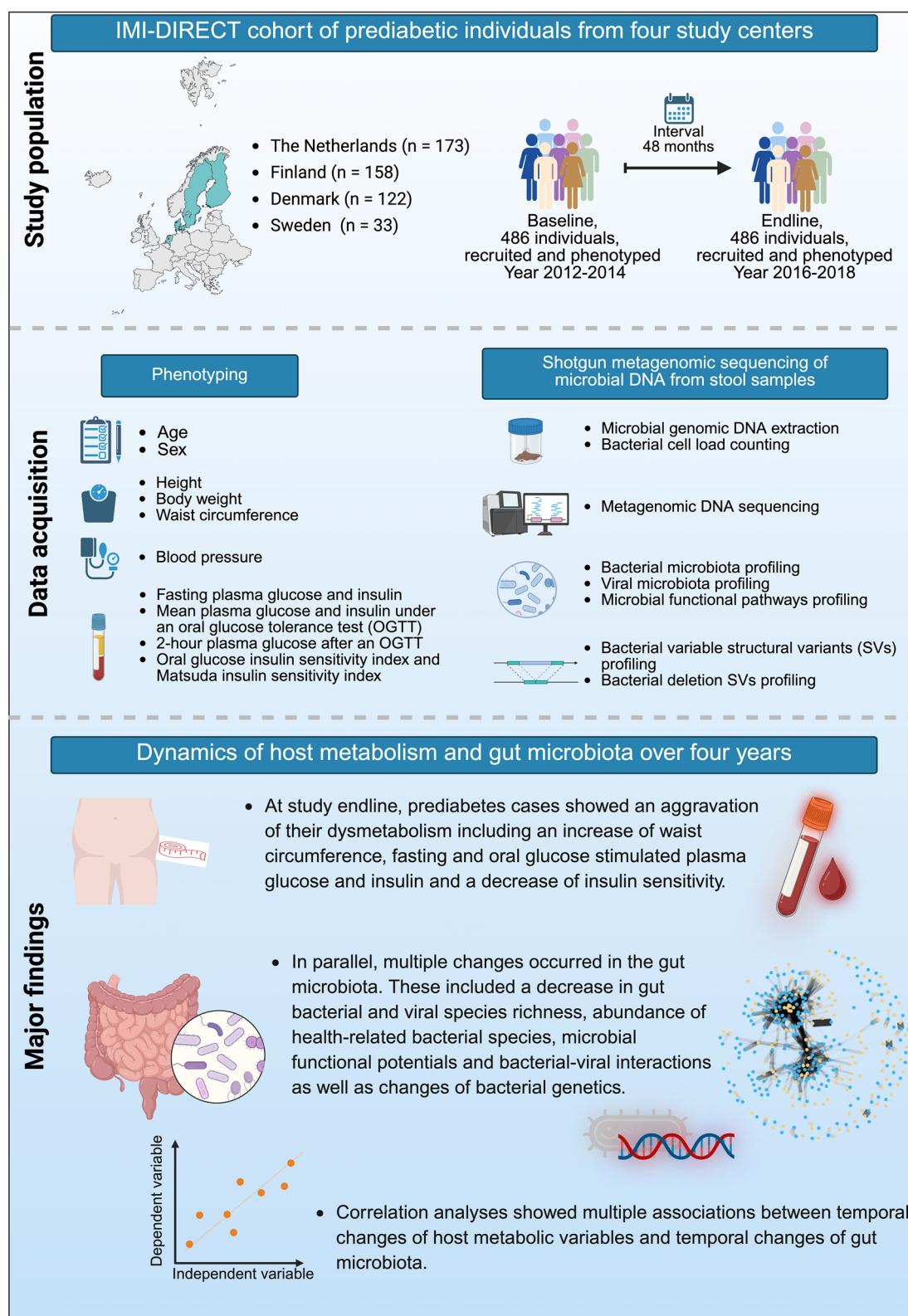


Fig. 1 Overview of study materials, methods, and major study outcomes. The figure is created with biorender.com

Table 1 Anthropometric and bioclinical characteristics of 486 individuals in the IMI-DIRECT study. Values are presented as mean \pm standard deviation or count and percentage. Statistical significance between baseline and endline measurements was assessed using Wilcoxon paired *t*-test for samples where both baseline and endline data were available. Adjusted *p* values were derived using the Benjamini–Hochberg method to correct for multiple comparisons

	Baseline <i>n</i> = 486	Endline <i>n</i> = 486 [#]	Adjusted <i>p</i> value
Study center			
The Netherlands	173 (36%)	173 (36%)	
Finland	158 (33%)	158 (33%)	
Denmark	122 (25%)	122 (25%)	
Sweden	33 (7%)	33 (7%)	
Age (years)	62 \pm 6	66 \pm 6	
Sex			
Female	169 (35%)	169 (35%)	
Male	317 (65%)	317 (65%)	
Waist circumference (cm)	99.7 \pm 10.8	100.7 \pm 11.0	9.9e-05
Body mass index (kg/m ²)	28.2 \pm 3.8	28.3 \pm 4.0	1.1e-01
Fasting plasma glucose (mmol/L)	5.9 \pm 0.5	6.2 \pm 0.7	1.5e-14
Mean plasma glucose (mmol/L)	8.0 \pm 1.5	8.6 \pm 1.7	4.4e-15
Fasting plasma insulin (pmol/L)	93 \pm 67	96 \pm 70	4.5e-02
Mean plasma insulin (pmol/L)	466 \pm 313	529 \pm 365	8.6e-08
Oral glucose insulin sensitivity index	358 \pm 58	336 \pm 61	2.6e-15
Matsuda insulin sensitivity index	3.3 \pm 1.9	3.1 \pm 2.0	4.4e-05
2-h plasma glucose after an OGTT (mmol/L)	6.2 \pm 1.7	6.9 \pm 2.1	2.4e-11
Systolic blood pressure (mmHg)	130 \pm 15	131 \pm 16	5.6e-02
Diastolic blood pressure (mmHg)	80 \pm 8	79 \pm 9	6.8e-02

[#] For all metadata, data were available for over 89% of individuals at both time points

Methods

Recruitment and phenotyping of study participants

In this protocol, we recruited participants who met the inclusion criteria: White European ethnicity, aged between 35 and 75 years, and meeting at least one of the three 2011 American Diabetes Association (ADA) criteria for prediabetes were recruited and phenotyped in the baseline study between November 2012 and August 2014 from four European study centers: Finland [11], the Netherlands [12, 13] Denmark [14–16], and Sweden [17], as part of the IMI-DIRECT consortium [18]. The ADA prediabetes criteria include fasting plasma glucose (100 mg/dL (5.6 mmol/L) to 125 mg/dL (6.9 mmol/L)); 2-h plasma glucose after an OGTT (140 mg/dL (7.8 mmol/L) to 199 mg/dL (11.0 mmol/L)); or HbA1c (5.7–6.4%, 39–47 mmol/mol) [19]. A total of 813 participants with impaired glucose regulation were recruited and phenotyped. However, 327 individuals were excluded due to failure to meet specific criteria, including antibiotics use within the past 3 months, insufficient stool sample quantity or quality, or lack of available metagenomic sequencing data after 4 years (Additional file 2: Fig. S1). The final prediabetes cohort comprised 486 adults with

prediabetes (169 females and 317 males), with 173 participants from the Netherlands, 158 from Finland, 122 from Denmark, and 33 from Sweden (Fig. 1). At study entry, participants had a mean age of 62 \pm 6 years (Table 1). Study participants were phenotyped in the baseline study between November 2012 and August 2014, and in the endline study between November 2016 and August 2018.

Consortium-wise standard operation procedures (SOPs) for phenotyping were applied in each of the four study centers. Clinical examinations and blood sampling for biochemistry were conducted in the morning after a 10-h overnight fast. Height was measured without shoes using calibrated wall-mounted stadiometers, while weight was recorded without shoes and with participants wearing light clothing using calibrated scales. Body mass index (BMI) was calculated by dividing weight in kilograms by the square of height in meters. Waist circumference was measured in the standing position with non-stretchable measuring tapes placed midway between the lower rib margin and the iliac crest. Blood pressure was measured after 10 min of rest using calibrated manual or automatic sphygmomanometers with appropriately sized arm cuffs. On the day of the examination,

three seated blood pressure readings were recorded and averaged for each study participant.

A standardized OGTT was done where the individual over 5 min drank 75-g glucose dissolved in 325 mL water. Blood samples were collected at fasting and at 30, 60, 90, and 120 min to measure plasma glucose and insulin levels. Mean plasma glucose or insulin represents the average concentration of glucose or insulin measured during the OGTT. This value is calculated using the trapezoidal rule for numerical integration, which divides the total area under the concentration–time curve (AUC) by the duration time of the test. This variable provided a standardized estimate of the mean glucose or insulin levels over the test period. Oral glucose insulin sensitivity (OGIS) index was determined using a validated mathematical model that incorporates plasma glucose and insulin levels at fasting, 90, and 120 min [20]. Additionally, the Matsuda insulin sensitivity index, which assesses whole-body insulin sensitivity, was calculated using plasma glucose and insulin levels at fasting, 30, 60, 90, and 120 min based on the defined formula [21]. Plasma glucose and insulin concentrations were measured centrally in batches at the University of Eastern Finland, Kuopio. Plasma glucose was quantified using the enzymatic glucose hexokinase method with photometric detection on a KoneLab 20 XT Clinical Chemistry Analyzer (Thermo Fisher ScienPfifc, Vantaa, Finland). Plasma insulin was measured through electrochemiluminescence on Roche E170 Analyzers (Hoffmann-La Roche). All plasma samples were stored at -80°C prior to analysis. To control for inter-assay variability, reference samples were included in all assays. This core laboratory regularly engaged in international external quality assessments and to control for inter-assay variability, reference samples were included in all assays.

Stool sample collection, bacterial cell counting, and gut microbial DNA extraction

Participants collected stool samples at home, adhering to SOPs that included immediate freezing of the samples at -18°C in their home freezers. The samples were then transported to the laboratory in an insulating cooler bag or styrofoam box containing cooling elements or dry ice. Upon arrival at the laboratory, the samples were stored at 80°C until DNA extraction.

Bacterial cells in stool samples were counted using staining and flow cytometry [22]. Data on bacterial cell counts was used for quantitative microbial profiling, as described [22].

Microbial DNA was extracted and purified from the frozen fecal samples using the NucleoSpin Soil DNA extraction kit (Machery-Nagel, catalog No. 740780.50) following the manufacturer's protocol.

Metagenomic sequencing and data processing

Library preparation and next-generation sequencing were conducted at the University of Lille-CNRS, France. Shotgun sequencing was performed on the Illumina HiSeq 4000 system, utilizing a paired-end 2×150 base pair (bp) protocol with one pool per lane. The resulting reads underwent quality filtering with KneadData (http://huttenhower.sph.harvard.edu/knead_data) to remove low-quality bases. Reads aligning to the human genome were excluded by mapping the quality-filtered reads to the human genome (GRCh38 release) using Bowtie2 [23] (version 0.2.3.2). Polymerase chain reaction (PCR) and optical duplicates were removed from the data using samtools [24] (version 1.6).

High-quality metagenomic sequencing reads were processed using Phanta [25] (version 1.1.0) to generate species count data, including species of bacteria, virus, and fungi, with default settings. The default Phanta database was used for annotations. For bacteria and archaea, the HumGut collection, comprising 30,691 dereplicated genomes from Unified Human Gastrointestinal Genome (UHGG) [26] and RefSeq [27], was used. For viruses, the Metagenomic Gut Virus catalog [28] and RefSeq were employed, while for gut eukaryotes, annotation was based on RefSeq. We rarefied this dataset to a minimum of 3.9 million total reads among all the sequencing samples. The compositional data was then categorized into gut bacterial species, gut viral species, and gut fungi, based on kingdom classification. Within kingdom, species abundances were filtered stringently based on both read number and prevalence. Gut fungi data is not included in the present communication due to insufficient sequencing coverage.

To reduce bias from very low-abundance taxa, we included only species with more than ten reads that were present in at least 20% of the samples. The calculation of Bray–Curtis distance was made on relative abundance following rarefaction. The abundance data for gut bacterial and viral species were analyzed independently to describe the stability of gut microbiota across bacterial and viral kingdoms.

HUMAnN3 [29] was used to construct MetaCyc [30] pathways from gene family data with default settings, providing insights into the functional potentials of the microbial community. MetaCyc pathways with relative abundance $> e - 05$ in $> 20\%$ of samples were included.

Structural variant (SV) profiles were extracted based on high-quality metagenomic sequencing reads. SGV-Finder [31] was used to identify both deletion SVs and variable SVs using the default parameters [7, 31, 32]. The SV data with $> 20\%$ non-missing values at both study baseline and endline were included to minimize bias from sequencing errors.

Statistical analysis—general approaches

Statistical analyses of differences between study baseline and endline were performed for microbial richness, Shannon and Simpson diversity, inter-individual distances, the first and second principal coordinates (PCo1 and PCo2) from principal coordinate analysis (PCoA) [33], and microbial relative abundance using the Wilcoxon signed-rank test. In the association analyses, continuous variables were scaled using the scale function in R to ensure comparability and mitigate differences in measurement scales. Unless otherwise stated, temporal association analyses were conducted using linear regression models, adjusted for baseline age, sex, study center, and bacterial cell load changes. All adjusted *p* values were calculated using the Benjamini–Hochberg method to account for multiple comparisons.

Analysis of the temporal dynamics at the microbial community level

Alpha diversity (microbiota composition) of the gut microbial species was assessed as richness, and Shannon's and Simpson's diversity indices [34], following data rarefaction but before any filtering. Specifically, compositional richness, which is the number of different microbial species present within one sample, was calculated using the *specnumber* function from the “vegan” package (version 2.6–6.1) in R (version 4.3.2). Species diversity and dominance were analyzed using the Shannon and Simpson indices [34], respectively, from the *diversity* function. The Shannon index measures species diversity by accounting for both the richness (number of species) and the evenness (distribution of individuals among species) within a community. Higher values indicate greater diversity, with a balance in the abundance of species. The Simpson index, on the other hand, focuses on species dominance by quantifying the probability that two randomly selected individuals from a sample belong to the same species. Lower values in the Simpson index reflect higher diversity, while higher values suggest dominance by one or a few species. Similarly, MetaCyc pathway richness was calculated applying the *specnumber* function.

The overall compositional variation and beta diversity of the gut microbiota were assessed using the Bray–Curtis distance from microbial species-level abundance profiles. This variation was further partitioned into intra-individual and inter-individual distances. Intra-individual distance quantified the dissimilarity between paired samples from the same individual across two different timepoints, illustrating temporal shifts in overall bacterial species abundance. The “inter-individual distance at study baseline” and “inter-individual distance at study endline” were calculated to measure the median value of

dissimilarities between a sample of one individual against those of other individuals at the same timepoints.

Additionally, permutational multivariate analysis of variance (PERMANOVA) using the *adonis2* function was conducted on Bray–Curtis distance matrix to evaluate the impact of various host or environmental factors on the compositional variation. This analysis included categorical factors such as timepoint, study center, and sex, alongside scaled continuous variables including age, body mass index, waist circumference, fasting plasma glucose and insulin, mean plasma glucose and insulin under an OGTT, 2-h plasma glucose after an OGTT, OGIS index, Matsuda insulin sensitivity index, systolic and diastolic blood pressure, and bacterial cell load. In our PERMANOVA analysis, 999 permutations were involved, with the proportion of variance explained and adjusted *p* values reported. Principal coordinate analysis (PCoA) was executed through the *cmdscale* function in R. The first two principal coordinates were used for visualization. The variance explained by these coordinates was derived from their eigenvalues and expressed as a percentage of the total variance.

To specifically assess host factors influencing the dynamics of bacterial species richness and the association between baseline bacterial species richness and intra-individual distance, Spearman's rank correlation analysis was performed in two settings: (1) between changes in bacterial species richness and changes in host metabolic variables and (2) between study baseline richness and intra-individual distance.

Analysis of microbiota dynamics at the level of the single microbial species or pathway

The variation of specific taxa and pathways were explored with higher resolution. Differentially abundant microbial species and pathways between study baseline and endline samples were identified using the Wilcoxon signed-rank test, with significance set at an adjusted *p* value < 0.05 .

In the assessment of temporal variance (intra-individual variance) or individual variance (inter-individual variance), we applied linear mixed-effects models with subject identity (subject ID) as the random effect. We included all variables used in PERMANOVA as fixed effects. Consequently, the calculations of intra- and inter-individual variances were adjusted to account for systematic influences from these variables. For each microbial feature (bacterial species, viral species, or microbial pathways), we calculated variance components from the results of linear mixed-effects models. Inter-individual variance was extracted from the first component of the random effect variance–covariance matrix (subject ID), and intra-individual variance from the residual variance [8, 9]. The intraclass correlation coefficient (ICC) [35]

was calculated as inter-individual variance/(inter-individual variance+intra-individual variance) to quantify the proportion of total variance attributable to variance between individuals.

To explore variance patterns across all gut microbial taxa or pathways, we conducted association analyses between total variances and relative abundance of the pertinent microbial feature using Spearman's rank correlation analysis.

Analysis of the variations at bacterial genetics level

To evaluate the dynamics of bacterial genetics within bacterial species over time and across individuals, we calculated intra-individual and inter-individual distance for each species from their SV profiles.

For all SVs derived from the same bacterial species, intra-individual and inter-individual distances were calculated using the *Canberra* distance metric for variable SVs and the *Jaccard* distance metric for deletion SVs, implemented via the *vegdist* function from the R package *vegan* (version 2.6–8).

The degree of microbial individuality (DMI) [9] was calculated as the mathematical difference between inter-individual distance and intra-individual distance. Given that distance is mathematically defined as $distance = 1 - similarity$, the DMI effectively captures the mathematical difference between intra-individual similarity ($1 - intra-individual\ distance$) and inter-individual similarity ($1 - inter-individual\ distance$). This index highlights how much more similar SV profiles are within individuals compared to between individuals, serving as a measure of microbial individuality.

Network analyses of gut microbiota

To analyze the interaction network among bacterial and viral species, we applied the *SparCC* (sparse correlations for compositional data) algorithm [36] to identify significant correlations both within and between microbial kingdoms. We focused on correlations with an absolute value of the correlation coefficient >0.3 , which were considered strong enough to construct the interaction network [37]. The network was visualized using the Fruchterman–Reingold layout, which was implemented through the R *igraph* package (version 2.1.1). To assess the significance of each microbial node in the network's transition between study baseline and endline groups, we calculated the NetMoss [38] score for each node. This score was compared to a null distribution generated by a permutation test with 100 iterations. In each iteration, sample labels were randomly shuffled, and the network was reconstructed to create a distribution of expected node scores under random conditions. For visualization and comparison of bacterial-viral interactions between

study baseline and endline groups, we excluded bacterial-bacterial and viral-viral interactions from the network to focus solely on the cross-kingdom interactions.

Association analyses between temporal changes in the relative abundance of the individual bacterial species and temporal changes in gut bacterial community indices, or host metabolic variables.

The host metabolic variables included in the association analyses were waist circumference, body mass index, fasting plasma glucose and insulin, mean plasma glucose and insulin under an OGTT, 2-h plasma glucose after an OGTT, OGIS index, Matsuda insulin sensitivity index, and systolic and diastolic blood pressure. For the analysis, linear regression models were used where changes in the relative abundance of the bacterial species were regressed against richness change and intra-individual distance or host metabolic variables alterations, controlling for baseline age, sex, study center, and bacterial cell load changes from study baseline to endline.

Mediation analysis

To explore whether changes in gut bacterial richness influenced host metabolic variables via specific bacterial species, we conducted mediation analyses using the R package *mediation* (v4.5.0). Candidate triplets (Δ gut bacterial richness \rightarrow Δ bacterial abundance \rightarrow Δ metabolic variable) were selected based on significant pairwise associations (adjusted p value <0.1). For each triplet, two linear models were fitted, adjusting for baseline age, sex, study center, and changes in total bacterial load, including (1) mediator model: Δ bacterial abundance $\sim \Delta$ gut bacterial richness + covariates and (2) outcome model: Δ metabolic variable $\sim \Delta$ gut bacterial richness + Δ bacterial abundance + covariates. We estimated the average causal mediation effect (ACME), average direct effect (ADE), and the proportion mediated (ACME/[ACME + ADE]). Mediation triplet with adjusted p value <0.1 of ACME were considered significant.

Analysis of relationships between variation in bacterial genetics and host metabolic variables

SVs in bacterial species were tested for correlations with host metabolic variables. The first set of association studies was performed between the delta values of host metabolic variables and the intra-individual distance derived from multiple SVs within each bacterial species. The analytical approach followed that used in delta association analyses between host metabolic changes and bacterial species abundance changes but applied at the bacterial genetics level.

The second set of association analyses was conducted using a linear mixed-effects model to explore the relationships between host metabolic variables and the

prevalence of deletion SVs or the coverage of variable SVs. The models were constructed with the formula: host phenotypic markers ~ coverage of variable SVs or prevalence of deletion SVs + sex + age + study center + bacterial cell load + (1|SubjectID).

Results

Deterioration of metabolism in prediabetes cases over 4 years

For all measurements, data were available for >89% of individuals at both timepoints. After 4 years, we observed significant deteriorations in host metabolism across multiple markers. These were evidenced by a 4.4% increase in fasting plasma glucose (adjusted p value = 1.5e - 14), a 7.4% increase in mean plasma glucose under an OGTT (adjusted p value = 4.4e - 15), a 13.8% increase in 2-h plasma glucose after an OGTT (adjusted p value = 2.4e - 11), a 14.6% increase of fasting plasma insulin (adjusted p value = 4.5e - 02), a 19.3% increase in mean plasma insulin under an OGTT (adjusted p value = 8.6e - 08), and a 1.2% increase of waist circumference (adjusted p value = 9.9e - 05), but a 4.9% decrease in OGIS index (adjusted p value = 2.6e - 15) and a 3.8% decrease in Matsuda insulin sensitivity index (adjusted p value = 4.4e - 05) (Table 1). In line with these changes, we observed heterogeneous glycemic outcomes among individuals with impaired glycemic regulation at baseline: 16% reverted to normal glycemic regulation, 73% remained impaired, and 11% progressed to type 2 diabetes over 4 years (Additional file 1: Table S1).

Identification of a core bacterial microbiota and core microbial pathways that are present both at study baseline and endline

To investigate the temporal dynamics of the human gut microbiota, we first analyzed the shotgun metagenomic sequencing reads. After rarefying the sequencing data to a minimum total matched reads of 3.9 million (Additional file 1: Table S2) and filtering out low-abundance species across all 972 metagenomes, we identified 571 bacterial species and 183 viral species (Additional file 1: Tables S3 and S4).

For the gut bacterial species, the median cumulative relative abundance across all 571 bacterial species was 98% (interquartile range, IQR: 96–99%) (Additional file 1: Table S3). Within this bacterial community, we identified 80 core bacterial species that were consistently present (read counts > 10) in > 99% of the 486 individuals at both timepoints, representing a stable and ubiquitous component of the gut bacterial microbiota. The median cumulative relative abundance of these 80 dominant core species was 59% (IQR: 53–66%) despite representing only 14% (80 out of 571) of the total number of distinct bacterial

species. Within the core bacterial microbiota, the ten most abundant genera included *Clostridium*, *Bacteroides*, *Ruminococcus*, *Eubacterium*, *Alistipes*, *Collinsella*, *Roseburia*, *Blautia*, *Streptococcus*, and *Prevotella*. Among these, *Faecalibacterium prausnitzii* exhibited the highest relative abundance of 4.6% (Additional file 1: Table S3).

In the analysis of the gut viral microbiota, we identified 183 prevalent viral species across the 972 samples with a median cumulative relative abundance of 36% (IQR: 27–44%, Additional file 1: Table S4). Most of the identified viral species belonged to *Brigitvirus*, *Mushuvirus*, *Svunavirus*, *Taranisvirus*, *Toutatisvirus*, *Oengusvirus*, and *Lilyvirus* genera. Notably, unlike the core bacterial microbiota, no core viral gut microbiota at species level was observed in our study, potentially implying the high sensitivity of viral species to environmental factors or host health condition.

After filtering out unannotated and unintegrated pathways, we identified 531 microbial pathways, representing a relative abundance range of 3–6%, illustrating the fact that only a small portion of the gut microbial pathways are known. After filtering of low abundant pathways, we defined 278 prevalent microbial pathways. Among these, 78 core pathways were consistently present (relative abundance > 0.01% in > 99% of individuals), representing a stable and likely essential configuration of the gut microbial functional potentials (Additional file 1: Table S5). The cumulative relative abundance of the 78 core pathways represents a substantial proportion, amounting to 73% of total abundance of 531 annotated pathways (IQR: 70–75%). Among the core pathways, sucrose biosynthesis II pathway showing the highest relative abundance accounted for 1.4% of total abundance of all annotated pathways in the present analysis.

Temporal declines of bacterial and viral microbiota richness that correlate with an aggravation of host metabolism

Over 4 years, the median value of compositional richness of gut bacterial species was decreased by 9.1% at the study endline (adjusted p value = 1.1e - 31, Fig. 2A). Despite the decrease in bacterial species richness over time, Shannon and Simpson diversity indices (Additional file 2: Fig. S2A and B) remained unchanged between the two timepoints, suggesting a preserved ecological balance over time. In addition, the number of distinct microbial pathways decreased by 2.4% (p value = 9.5e - 04, Fig. 2D) over the 4 years. The compositional richness of gut viruses showed a 14.3% decline at the study endline (p value = 2.1e - 15, Fig. 2G), with virulent viruses showing the larger decline (16.3% and p value = 1.9e - 17, Fig. 2J–L). Yet, Shannon and Simpson diversity indices

for the viral microbiota remained unchanged (Additional file 2: Fig. S2E and F).

Importantly, the decrease in bacterial species richness was significantly associated with adverse metabolic changes over the 4-year period, such as increases in body mass index (BMI), waist circumference, mean plasma insulin, and reductions in insulin sensitivity (Fig. 3 and Additional file 1: Table S7). These findings suggest that the lowering of gut bacterial richness and the loss of key bacterial species are closely linked to a worsening of the metabolic health in individuals with prediabetes. Stratified analyses further revealed that the temporal associations between declining bacterial richness and worsening metabolic traits were more pronounced in males than in females (Additional file 2: Fig. S3), which may reflect underlying sex-specific host–microbiota interactions as well as reduced statistical power in the female subgroup due to sample size imbalance ($n=317$ males vs. $n=169$ females, Table 1).

Prediabetes individuals with higher initial microbial species richness show smaller intra-individual distance calculated from overall microbial species relative abundance

Bray–Curtis dissimilarity analysis based on the relative abundance of gut bacterial species showed smaller compositional differences within the same individual than between individuals (Fig. 2B). This was the case both at study baseline (adjusted p value = $6.7e-116$) and at study endline (adjusted p value = $3.4e-101$, Fig. 2B). Compared to the inter-individual distance observed at study baseline, the inter-individual distance at study endline was smaller, indicating a greater resemblance in bacterial microbiota composition among the cohort's endline samples (adjusted p value = $4.3e-20$, Fig. 2B).

Permutational multivariate analysis of variance (PERMANOVA) based on the Bray–Curtis distance matrix revealed that both host and environmental factors significantly influenced the variation of the bacterial composition at the species level (Additional file 2: Fig. S2C).

Notably, the study center confounder accounted for 1.6% of the explained variance, implying that the demographics of each of the four study centers have impacts on the observed microbial community structures (Additional file 2: Fig. S2C). In addition, host factors like age, sex, waist circumference, body mass index, mean plasma glucose level under an OGTT, and Matsuda insulin sensitivity index exerted modest, yet significant influence (Additional file 2: Fig. S2C). The timepoint factor alone, though contributing a small fraction, significantly impacted the bacterial composition variance (PERMANOVA $R^2=4.9e-03$, adjusted p value = $3.2e-03$). Similarly, principal coordinate analysis (PCoA) based on the Bray–Curtis distance metric of bacterial species-level profiles revealed significant shifts in bacterial composition over time, as indicated by a change in PCo1 values between samples obtained at study baseline and endline (p value = $2.7e-04$, Fig. 2C).

Next, we found an inverse link between bacterial species richness at study baseline and intra-individual distance (Spearman's correlation coefficient = -0.13 , $p=2.2e-18$, Additional file 2: Fig. S2D), indicating that individuals with higher initial bacterial species richness experienced smaller intra-individual changes in gut bacterial abundance over time.

Similarly, for microbial pathways and gut viral species, Bray–Curtis distance index revealed greater inter-individual differences than intra-individual differences, with reduced variability at study endline (Fig. 2E and H). PERMANOVA showed that both host metabolism markers and demographics significantly influenced viral composition (Additional file 2: Fig. S2G). PCoA analysis confirmed the significant temporal shifts in microbial functional pathways and viral composition (Fig. 2F and I). Additionally, baseline viral species richness was inversely associated with intra-individual distance (Spearman's correlation coefficient = -0.26 , $p=1.6e-18$, Additional file 2: Fig. S2H). Interestingly, our observations also showed that microbial pathways exhibited a higher degree of

(See figure on next page.)

Fig. 2 Temporal changes of gut microbial features between study baseline and endline. **A, D**, and **G** show the compositional richness of gut bacterial species, microbial pathways, and viral species, respectively, at study baseline and study endline. **B, E**, and **H** display the intra- and inter-individual Bray–Curtis distances of gut bacterial species, microbial pathways, and viral species, respectively. The dots in brown, yellow, or pink colors show the Bray–Curtis distances derived from the relative abundance of gut bacterial species, microbial pathways, or viral species, respectively, illustrating variability within and between individuals at study baseline and study endline. In **C, F**, and **I**, the principal coordinate analysis (PCoA) of overall composition of gut bacterial species, microbial pathways, and viral species, respectively, based on Bray–Curtis dissimilarity matrix, are shown. Yellow and pink dots represent the mean PCo1 and PCo2 coordinates for all study baseline and study endline samples with error bars indicating the standard error of the mean (SEM). **J–L** Changes in the compositional richness of temperate and virulent viruses, as well as the ratio of virulent to temperate viral species, respectively, from study baseline to endline. All adjusted p values were derived from p values corrected for multiple comparisons using the Benjamini–Hochberg method. The compositional richness and Bray–Curtis distance are expressed in arbitrary units

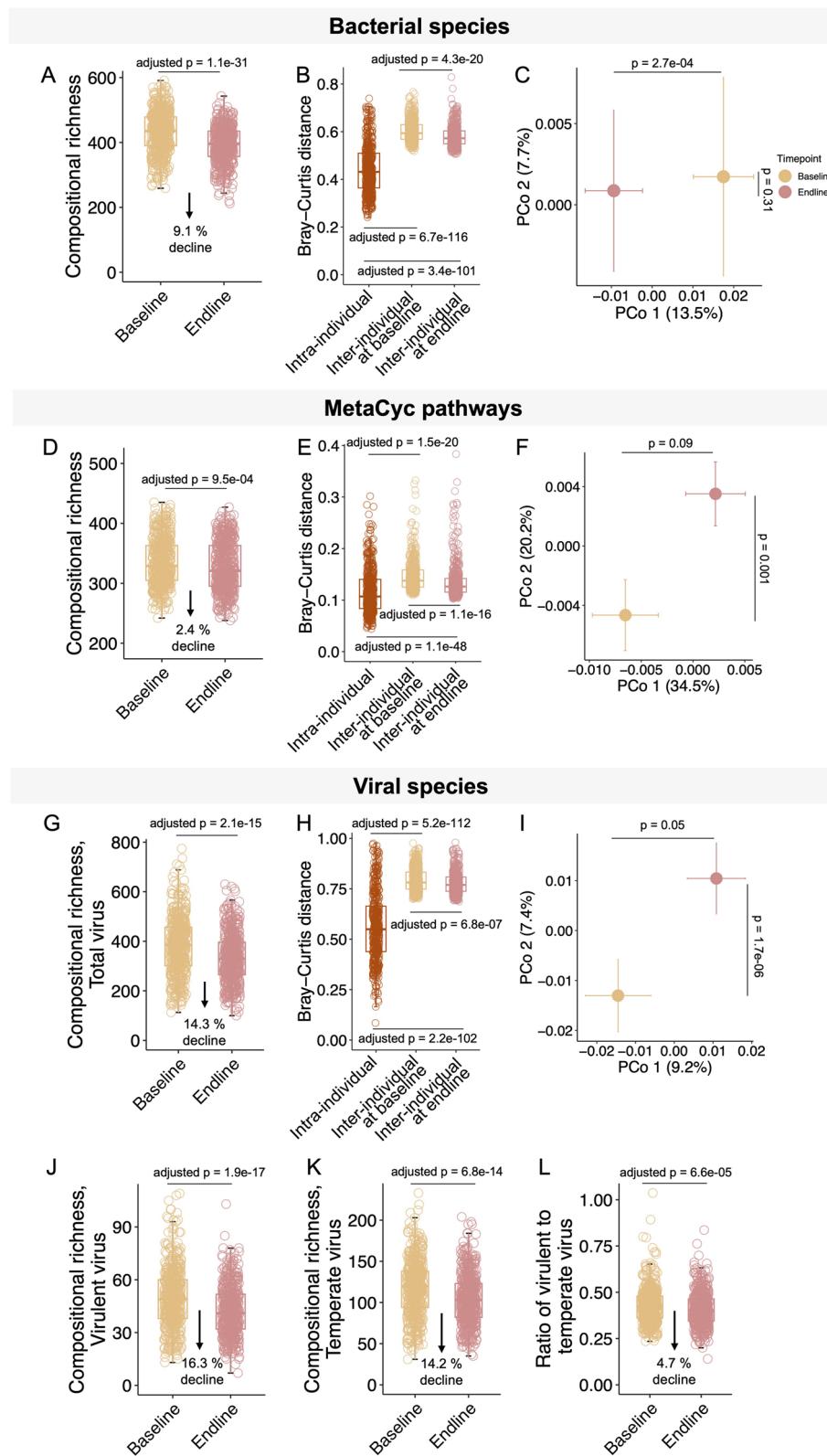


Fig. 2 (See legend on previous page.)

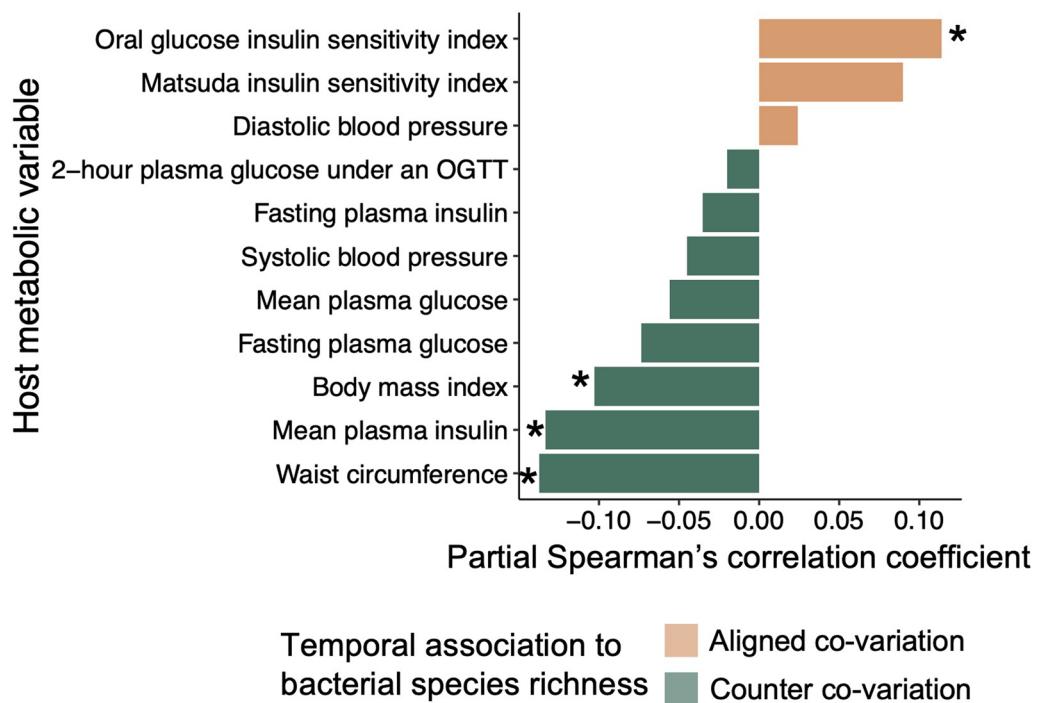


Fig. 3 Relationships between temporal changes of bacterial species richness and temporal changes of host metabolic variables. This bar plot illustrates the delta associations, referring to the relationship between changes in bacterial species richness and host metabolic variables over time, as determined by partial Spearman's correlation analyses adjusting for baseline age, sex, study centers, and bacterial cell load. The x-axis shows the strength of the associations by partial Spearman's correlation coefficient values, with orange bars indicating positive correlation coefficients (aligned co-variation) and green bars indicating negative correlation coefficients (counter co-variation). The y-axis lists the host metabolic variables. OGTT means oral glucose tolerance test. Significance levels are denoted as * for adjusted p values < 0.1 derived from p values corrected for multiple comparisons using the Benjamini–Hochberg method

DMI than microbial species (Additional file 2: Fig. S4), implying a greater intra-individual stability of the former.

Temporal changes of the prevalence of gut bacterial species

Following the observed decline in gut bacterial microbiota richness, we examined the prevalence changes of bacterial species that underwent the most significant shifts. Our analysis revealed that the reduction in bacterial richness was accompanied by a $> 30\%$ decrease in the prevalence of 14 highly prevalent bacterial species (prevalence $> 50\%$ at study baseline). These species include *Coprococcus eutactus*, *Bacteroides eggerthii*, *Alistipes inops*, *Phocaeicola massiliensis*, and *Evtepia gabavorous* (Additional file 2: Fig. S5A and Additional file 1: Table S6). For gut microbial pathways and gut viral microbiota, by performing the same filtering criteria, we did not find prevalent pathways or viral microbiota that showed any major decrease (data not shown).

Temporal shifts of the abundance of gut microbial species and pathways

Building upon the community-level analysis of alpha and beta diversity, we then focused on species-specific and pathway-specific features to explore their abundance dynamics from study baseline to endline. Significant changes were observed in the relative abundance of species and microbial pathways. Specifically, 295 bacterial species, 51 viral species, and 64 microbial pathways showed differential abundance (adjusted p value < 0.05), as shown in Additional file 2: Fig. S5B, C and Additional file 1: Tables S3–S5. Notably, multiple bacterial species from *Bifidobacterium* genus, such as *Bifidobacterium adolescentis*, *Bifidobacterium pseudocatenulatum*, and *Bifidobacterium catenulatum*, showed a decreased relative abundance over time, while the relative abundance of *Ruthenibacterium lactatiformans*, which may have pro-inflammatory effects, was higher at study endline (Additional file 1: Table S3). Additionally, microbial pathways involved in glucose degradation, as well as in energy production via the TCA cycle II, also declined over time (Additional file 1: Table S5).

Variance of microbial species and pathways observed in this longitudinal study

Furthermore, the longitudinal study design allowed us to investigate the variance observed among individuals. Hence, after adjusting for all the potential covariates, we profiled intra-individual variance, inter-individual variance, ICC, and total variance for all microbial species and pathways (Additional file 1: Tables S3–S5).

For total variance, bacterial and viral species exhibited similar values with bacterial species showing slightly lower medians. MetaCyc pathways had the lowest median value of total variance, indicating a higher temporal stability (Additional file 2: Fig. S6A). Interestingly, for gut bacterial species and microbial pathways, total variance was inversely correlated with the relative abundance of these microbiota components, suggesting that more abundant features tend to have lower variance both across two timepoints and between individuals (Spearman's correlation coefficient = -0.37 for bacteria and -0.93 for pathways, with both p value $<2.2e-16$, Additional file 2: Fig. S6B and C).

Dynamics of structural variants of gut bacteria

Applying the filtering criterion of $\geq 20\%$ non-missing values at both study baseline and endline, we found 3831 deletion SVs and 1820 variable SVs across the genomes of 39 bacterial species (Additional file 2: Fig. S7) with *Dorea formicigenerans* and *Dorea longicatena* having the highest number of SVs (350 and 269 SVs, respectively).

To evaluate the genetic stability of the gut bacterial species using the DMI metric, inter-individual and intra-individual dissimilarities were calculated separately for both deletion SVs and variable SVs. For both types of bacterial SVs, we found a wide range of DMI values across bacterial species (Fig. 4A). For deletion SVs, *Prevotella copri* had the highest DMI of 0.52, followed by *Akkermansia muciniphila* (DMI = 0.50) and *Ruminococcus bicirculans* (DMI = 0.49), indicating a high degree of genetic stability over time. *Prevotella copri* also stood out with a highest DMI value of 0.32 for variable SVs (Fig. 4B). *Roseburia hominis*, *Roseburia intestinalis*, and *Faecalibacterium prausnitzii* exhibited low DMI values based on both types of SV profiles. Particularly, *Roseburia hominis* showed similar intra- and inter-individual distance values for its deletion SVs (DMI = 0.06) suggesting a very low temporal genetic stability.

Fewer bacterial-viral interactions in the gut microbiota at study endline

To explore how overall gut microbial interactions evolve over time, we first analyzed the interactions within kingdom and between kingdoms at the study baseline and endline. Network analyses of bacterial and viral species

revealed a diminished microbial interactome at study endline compared to study baseline (Fisher's exact test p value = $3.0e-02$, Additional file 2: Fig. S8A–C). To identify the taxa driving the interactome shifts, we calculated the NetMoss score for each microbial node (see Methods) and visualized the top 50 features significantly contributing to the network alterations from study baseline to endline. Notably, the interactions centered at *Coprococcus comes*, a butyrate producer, that contributed most substantially to the network changes (Additional file 2: Fig. S8D and Additional file 1: Table S8).

Next, we specifically examined bacterial-viral interactions by excluding bacterial-bacterial and viral-viral interactions from the overall network to assess the influence of viral dynamics on bacterial populations. The network at study baseline was denser and more interconnected, whereas the endline network became more dispersed and sparser (Fig. 5A and B). Compared to study baseline, there were significantly fewer bacterial-viral interactions after 4 years (Fisher's exact test p value = $3.0e-02$, Fig. 5C). Both positive and negative associations decreased, with positive mutualistic relationships declining more markedly (615 versus 359 and 248 versus 186, respectively, for study baseline versus endline in positive and negative correlations, Fig. 5C). This reduction in bacterial-viral interactions aligns with the observed temporal decrease in species richness among both bacteria and viruses (Fig. 2A and G). Additionally, the substantial loss of virulent viruses may indicate a reduced viral predation pressure (Fig. 2J), potentially leading to fewer trans-kingdom interactions.

Temporal changes of the abundance of specific gut bacterial species are associated with overall bacterial community dynamics

By performing delta values correlation analysis on bacterial species abundance, we found that temporal abundance changes of *Alistipes putredinis*, *Oxalobacter formigenes*, and *Coprobacter secundus* co-varied directly with total bacterial species richness (beta coefficient = 0.14 to 0.19, adjusted p value = $2.8e-03$ to $2.5e-02$), suggesting that the increased abundance of these bacteria may drive a higher overall bacterial species richness (Fig. 6A and Additional file 1: Table S9). Conversely, inverse co-occurrence was seen between compositional richness and abundance changes of *Anaerostipes hadrus*, *Blautia wexlerae*, *Evtepia gabavorous*, and *Shigella flexneri* (beta coefficient = -0.17 to -0.25 , adjusted p value = $2.2e-05$ to $1.7e-03$), indicating that the increased abundance of these bacteria may contribute to reduce overall bacterial species richness.

Further association analyses showing counter- or aligned co-variation to the intra-individual distance

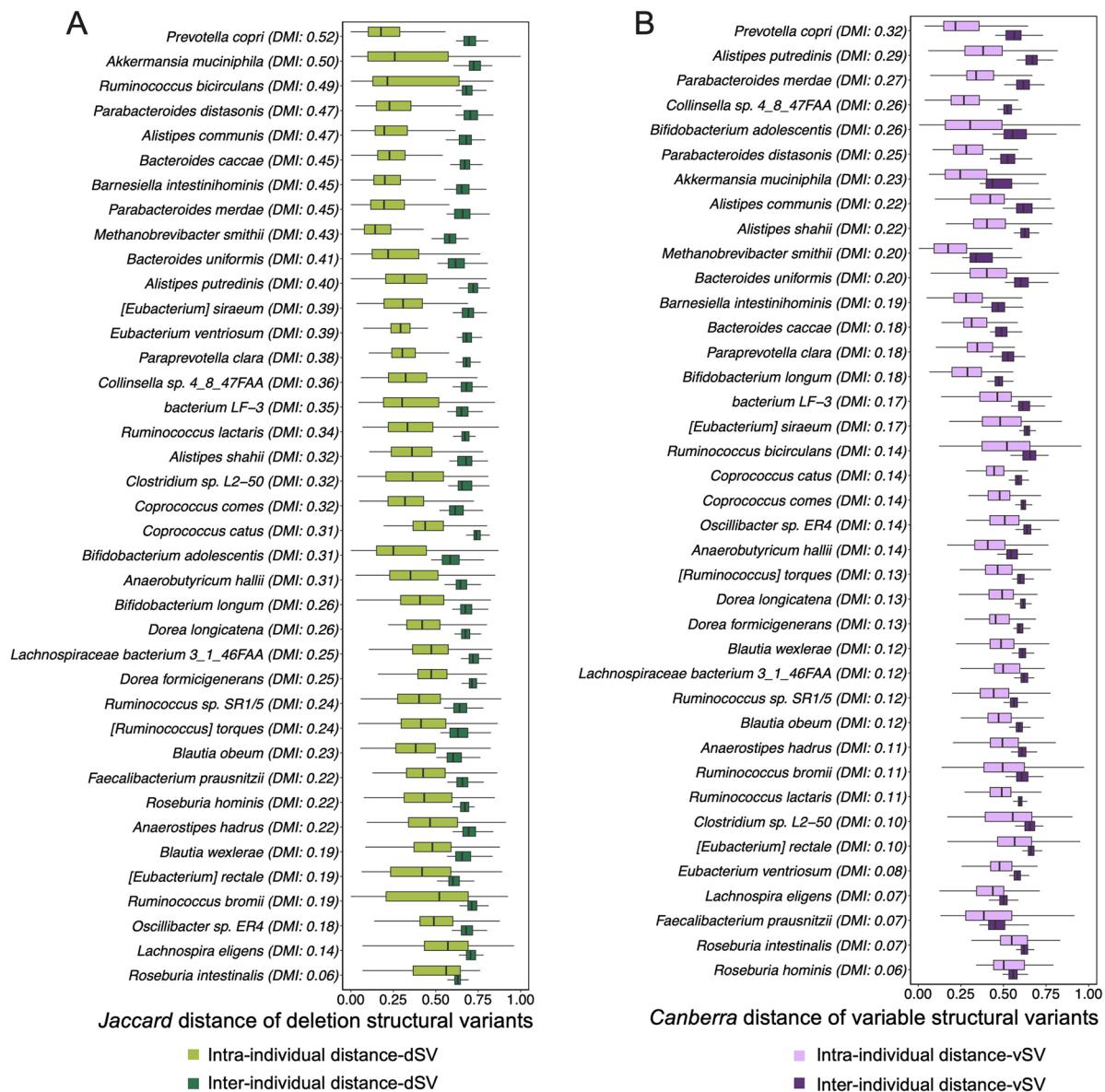


Fig. 4 Gut bacterial genetics stability within the individual and between individuals. To evaluate the dynamics of bacterial genetics within bacterial species over time and between individuals, we calculated intra-individual and inter-individual distance for each species from their structural variant (SV) profiles. **A** Profiles of Jaccard distance of deletion structural variants (dSVs) and **B** profiles of Canberra distance of variable structural variants (vSVs) of 39 bacterial species. Degree of microbial individuality (DMI) was labeled. Each box plot represents the distance calculated from the genetic structural variant profiles within one bacterial species (see Methods), with light-colored boxes indicating intra-individual distances, while dark-colored boxes are showing inter-individual distances. Bacterial species are listed along the y-axis in descending order based on their DMI values. Distances are displayed on the x-axis

(See figure on next page.)

Fig. 5 Gut bacterial-viral interactions at study baseline and endline. The interaction network of bacterial and viral microbiota at the species level at study baseline (**A**) and study endline (**B**) was constructed using SparCC (sparse correlations for compositional data). A more dispersed and sparser network was observed at study endline (**B**) compared to study baseline (**A**). Interactions with absolute value of correlation coefficient > 0.3 are shown in the network. Each edge in the network represents an interaction between a pair of taxa with edge thickness reflecting the absolute value of the correlation coefficient. Bacterial species and viral species are shown as blue or yellow nodes, respectively. **C** Bar plot showing the counts of positive or negative bacterial-viral interactions that decline at study endline compared to that at study baseline, with a more pronounced reduction observed in positive mutualistic relationships

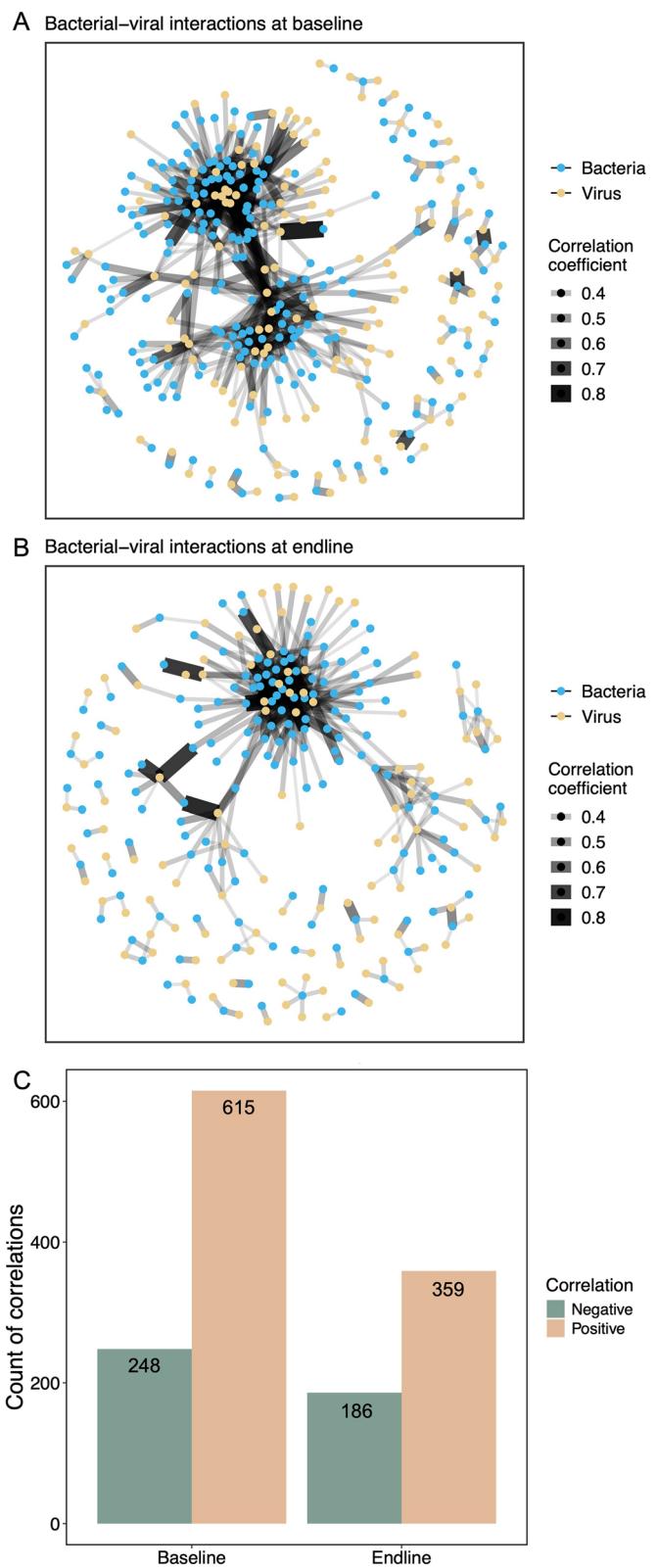


Fig. 5 (See legend on previous page.)

metric of bacterial species abundance (Fig. 6A and Additional file 1: Table S9) suggested the existence of both stabilizing bacterial species and destabilizing species. Thus, the abundance changes of *Dorea* sp. AF36-15AT or *Prevotella copri* exhibited inverse correlations to the intra-individual distance calculated from the relative abundances of overall bacterial species (beta coefficient = -0.17 and -0.16 , adjusted p values = $5.6e-02$), suggesting their potential role in stabilizing the overall gut bacterial community (lowering intra-individual distance). In contrast, the abundance changes in *Intestinimonas butyriciproducens* showed an inverse correlation to the intra-individual distance calculated from the relative abundances of overall bacterial species (beta coefficient = 0.19 , adjusted p value = $1.1e-02$), implying that the abundance of *Intestinimonas butyriciproducens* may contribute to, or serve as an indicator of the overall gut bacterial microbiota variation.

Temporal changes of the abundance of specific gut bacterial species are associated with changes in host metabolism

Association analysis revealed that changes in the relative abundance of specific bacterial species correlate with alterations in host metabolism, especially insulin sensitivity over time. Notably, increases in *Oxalobacter formigenes* and an uncultured *Clostridiales bacterium* positively correlate with changes in the Matsuda insulin sensitivity index and OGIS index (beta coefficients = 0.13 to 0.18 , adjusted p values = $6.0e-03$ to $3.0e-02$, Fig. 6B and Additional file 1: Table S9). Furthermore, additional associations were noted between changes in the Matsuda insulin sensitivity index and in the abundance change of *Proteobacteria bacterium CAG:495*, *Ruminococcaceae bacterium D5*, *Ruminococcus sp. CAG:382*, and *Lachnospiraceae bacterium OM04-12BH* (beta coefficient = 0.16 to 0.22 , adjusted p value = $2.0e-02$ to $3.0e-02$) as well as *Parabacteroides goldsteinii* (beta coefficient = -0.20 ,

adjusted p value = $2.0e-02$). Changes in mean plasma insulin levels were shown to be positively correlated with changes in the relative abundance of *Anaeromassilibacillus* sp. *An250* and *Clostridium sp. CAG:451* (beta coefficient = 0.19 to 0.22 , adjusted p value = $1.0e-02$ and $9.0e-02$). Conversely, 2-h plasma glucose levels after an OGTT were negatively associated with *Clostridium sp. CAG:253* (beta coefficient = -0.18 , adjusted p value = $8.0e-02$). To further explore potential causal links beyond associations, mediation analysis revealed that part of the association between gut bacterial richness and host metabolic traits—most notably BMI—may be indirectly mediated through specific bacterial taxa, such as *Clostridia bacterium DTU025* (Additional file 2: Fig. S9).

Structural variants in gut bacterial genomes are associated with markers of host metabolism

In the analysis of SVs within bacterial genomes (Fig. 7A and Additional file 1: Table S10), intra-individual distances of SV profiles in *Ruminococcus sp. SR1/5* were negatively associated with changes of fasting plasma glucose (beta coefficient = -0.24 , adjusted p value = $1.0e-03$, Fig. 7A and B). Positive temporal associations were identified between intra-individual distances calculated from SVs in *Methanobrevibacter smithii* and changes in BMI, diastolic blood pressure, and fasting plasma glucose (beta coefficients = 0.22 to 0.27 , adjusted p values = $2.0e-02$ to $8.0e-02$, Fig. 7A and C), while inversely associated with changes in OGIS index (beta coefficient = -0.27 , adjusted p value = $2.3e-02$, Fig. 7A). Similarly, intra-individual distances calculated from SVs in *Alistipes shahii* and *Alistipes putredinis* demonstrated positive associations with changes in mean plasma insulin during an OGTT and Matsuda insulin sensitivity index (beta coefficients = 0.17 to 0.29 and adjusted p values = $7.0e-02$ to $8.0e-02$, Fig. 7A). These findings suggest that genetic convergence, divergence, or sub-species shifts

(See figure on next page.)

Fig. 6 Associations of temporal changes of the relative abundance of bacterial species and temporal changes in gut bacterial community indices or host metabolic variables. **A** Correlations between temporal changes of the relative abundance of bacterial species and temporal shifts in overall gut bacterial species richness and intra-individual distance of bacterial species abundance profiles, highlighting the bacterial species that are driving the community changes. The y-axis lists bacterial species, and the x-axis shows the beta coefficient and standard error values calculated from linear regression models. The figure includes the top 10 bacterial species, ranked in ascending order by adjusted p value. The dots are colored yellow and blue for positive and negative coefficient values, respectively. Dots size indicates the $-\log_{10}$ (adjusted p value) of the correlation, with larger dots showing smaller adjusted p value. **B** Correlations between temporal changes in relative abundance of bacterial species and temporal changes of host metabolic variables, highlighting changes in relative abundance of bacterial species with parallel changes in host metabolism. Y-axis lists bacterial species, while x-axis lists host metabolic variable. The dots are colored yellow for positive coefficient values and blue for negative coefficient values with color intensity indicating the effect size. Dots size indicates the $-\log_{10}$ (adjusted p value) of the correlation, with larger dots showing smaller adjusted p value. In **A** and **B**, beta coefficients and adjusted p values were calculated from linear regression models after adjusting for co-variates of individual's age at baseline, sex, study centers, and delta value of bacterial cell load. All correlations shown are statistically significant after adjustment for multiple comparisons using the Benjamin-Hochberg procedure, with adjusted p value <0.1 . OGTT means oral glucose tolerance test

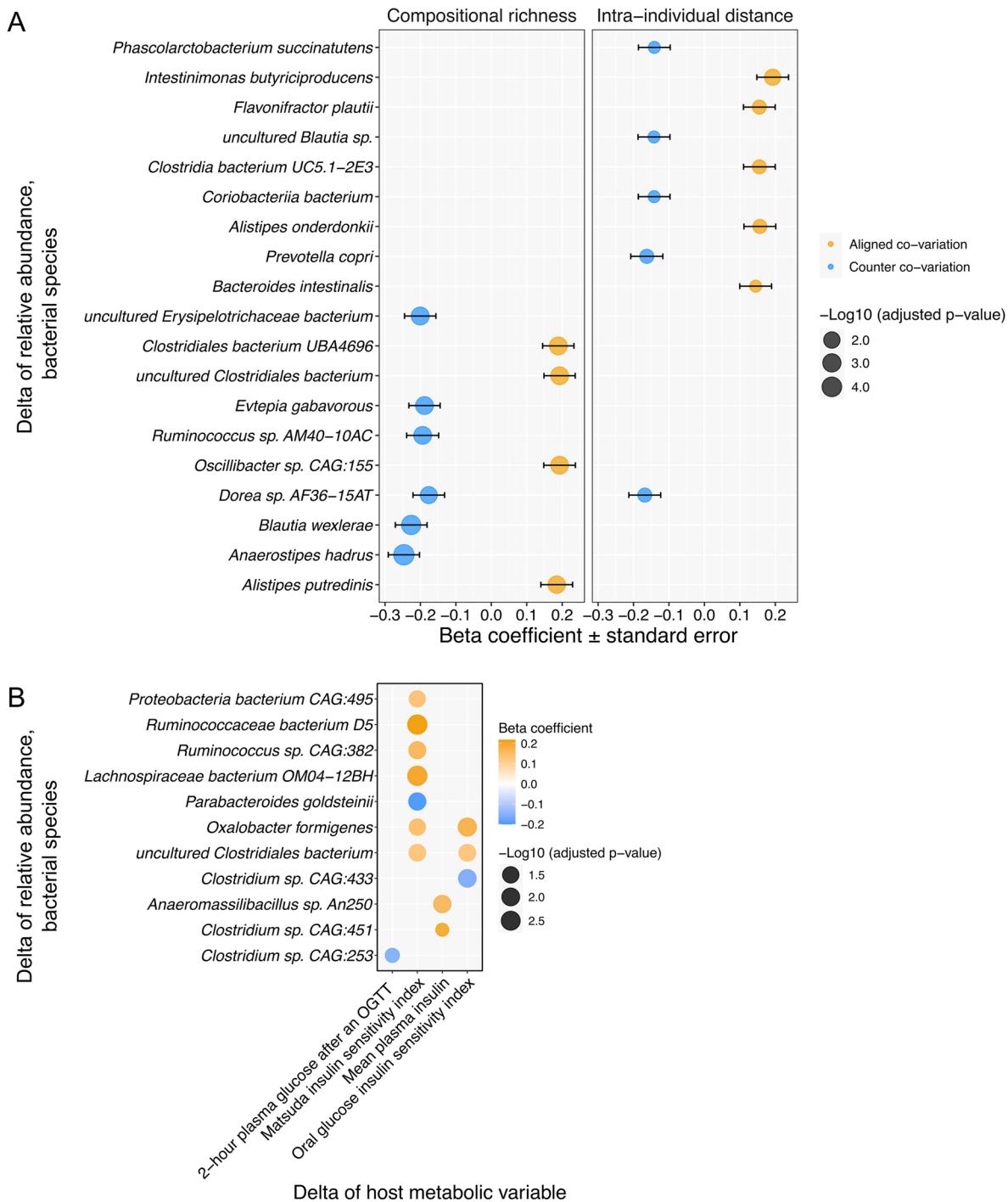


Fig. 6 (See legend on previous page.)

within bacterial species may influence or respond (adapt or contribute) to alterations in host metabolism.

Focusing on single SVs within gut bacterial genomes, we identified 25 associations between host metabolic markers and SVs across 15 bacterial species (Fig. 7D and Additional file 1: Table S11). *Coprococcus catus* had

the highest number of SVs that correlated with host metabolism. As noteworthy examples, in the genome of *Prevotella copri*, deletion SVs involved in genes encoding TonB-dependent receptor were positively associated with Matsuda insulin sensitivity index of the host (beta coefficient = 0.55, adjusted *p* value = 4.0e - 02, Fig. 7E,

Additional file 1: Table S11). In *Coprococcus catus*, deletion SVs at 1101–1102 kbp and 3414–3415 kbp within genes encoding hydrogenase were associated with a higher mean plasma insulin level (beta coefficient=0.92, adjusted p value=9.9e – 02, see Fig. 7F and Additional file 1: Table S11).

Discussion

In the present 4-year prospective study of 486 European prediabetic patients, we explored the temporal dynamics of the gut bacterial and viral microbiota at different resolution levels (Fig. 8) and related these gut microbial changes to the concomitant deterioration of host metabolism. The gut microbiota underwent a temporal community-level shift characterized by reduced bacterial and viral richness, fewer trans-kingdom interactions, and increased microbial convergence within the population. These microbiota shifts were correlated with changes in host metabolic markers that included an increase in glycemia and a decline of insulin sensitivity. When compared to the European LifeLines-DEEP cohort [7], which included non-prediabetic individuals, the prediabetic individuals in the present European cohort exhibited a more substantial increase in fasting plasma glucose during a 4-year follow-up.

At study endline, the gut bacterial microbiota showed a depletion of several bacterial species that are known to relate to metabolic health such as *Bifidobacterium adolescentis*, *Bifidobacterium pseudocatenulatum* [39], *Bifidobacterium catenulatum*, and *Coprococcus eutactus* [40], but an enrichment of pro-inflammatory bacteria, like *Ruthenibacterium lactatiformans* [41]. In parallel, the abundance of pathways involved in glucose degradation and energy production

decreased, while transitioning to less efficient, non-oxidative metabolic processes potentially associated with inflammation and metabolic stress occurred. Some of these findings correspond with discoveries in the gut microbiota of individuals at risk for type 1 diabetes (T1D) [42–47]. The overlapped bacteria and the potential mechanisms for T1D and T2D are shown in Additional file 1: Tables S12. Besides these partially overlapped gut microbiota features, individuals at risk of T1D and T2D exhibit disease-specific gut microbiota signatures that may reflect pathogenesis differences and age at diabetes onset [48].

Despite the metabolic deterioration observed in prediabetic individuals at study endline, we noted an increase in the relative abundance of *Akkermansia muciniphila*. This finding may seem paradoxical given the beneficial role of various *A. muciniphila* strains in metabolism [49–53]. However, the changes in *A. muciniphila* species abundance in our study despite an aggravation of metabolism may reflect broader ecological shifts in the gut microbiota. It might also represent a compensatory response to metabolic stress [54], with *A. muciniphila* attempting to restore gut ecological balance and mitigate further metabolic damage. Aligning with findings from a multi-cohort study that reported an upward trend in *Flavonifractor plautii* abundance across normoglycemic, prediabetic, and T2D individuals [55], we observed an increase in *F. plautii* abundance at study endline. Again, an unexpected finding that may exhibit some of the knowledge gaps in the current understanding of intestinal microbial ecology, since this bacterium has the genetic potential to produce monophenolic acid that has been suggested to counteract liver steatosis [56].

(See figure on next page.)

Fig. 7 Correlations between gut bacterial structural variants and host metabolic variables. **A** Temporal associations between changes in profiles of bacterial genetics (shown by intra-individual distance within bacterial species) and changes in host metabolic variables. The y-axis lists bacterial species, and the x-axis displays delta values of host metabolic variables. Dots are shaped by types of structural variants (circle for deletion structural variants (dSVs) and square for variable structural variants (vSVs)), colored based on beta coefficient values, and sized according to $-\log_{10}$ (adjusted p value). Beta coefficients and adjusted p values were calculated from linear regression models after adjusting for co-variates of individual's age at baseline, sex, study centers, and delta value of bacterial cell load. **B** and **C** Scatter plots showing selected examples of the temporal association results in **A**. **B** Temporal associations between the delta values of diastolic blood pressure and the intra-individual *Canberra* distance calculated from the profile of 19 vSVs within the *Methanobrevibacter smithii* genome, and **C** temporal associations between the delta values of fasting plasma glucose and the intra-individual Jaccard distance calculated from the profiles of 81 dSVs within the *Ruminococcus* sp. *SR1/5* genome. Each dot represents an individual participant. **D** A circular chord diagram illustrating the associations between specific gut bacterial species (right side) and host phenotypes (left side) categorized by SV types, either dSVs or vSVs, summarizing the links between single SVs and host phenotypes. Each bacterial species is labeled with its name and the count of associated SVs in parentheses, with the color of each link and species label showing the variant type involved (green for dSVs; purple for vSVs). The width of each link reflects the count of associated SVs. Beta coefficients and adjusted p values were calculated from linear mixed-effects models after adjusting for co-variates of individual's age, sex, study center, and bacterial cell load. **E** and **F** Bar plots showing selected examples of association results given in **D**. **E** Comparison of Matsuda insulin sensitivity index values in samples retaining ($n=105$) or deleting ($n=111$) gene fragments of 50–52 kbp in *Prevotella copri*, predicted to encode a TonB-dependent receptor protein. **F** Comparison of fasting plasma insulin in samples where gene fragments 1101–1102 and 3414–3415 kbp in *Coprococcus catus* are retained ($n=486$) or deleted ($n=13$). The encoding protein is predicted as hydrogenases. Statistical significance from linear mixed-effects model was labeled. All correlations shown are statistically significant after adjustment for multiple comparisons using the Benjamin-Hochberg procedure, with adjusted p value < 0.1. OGTT means oral glucose tolerance test

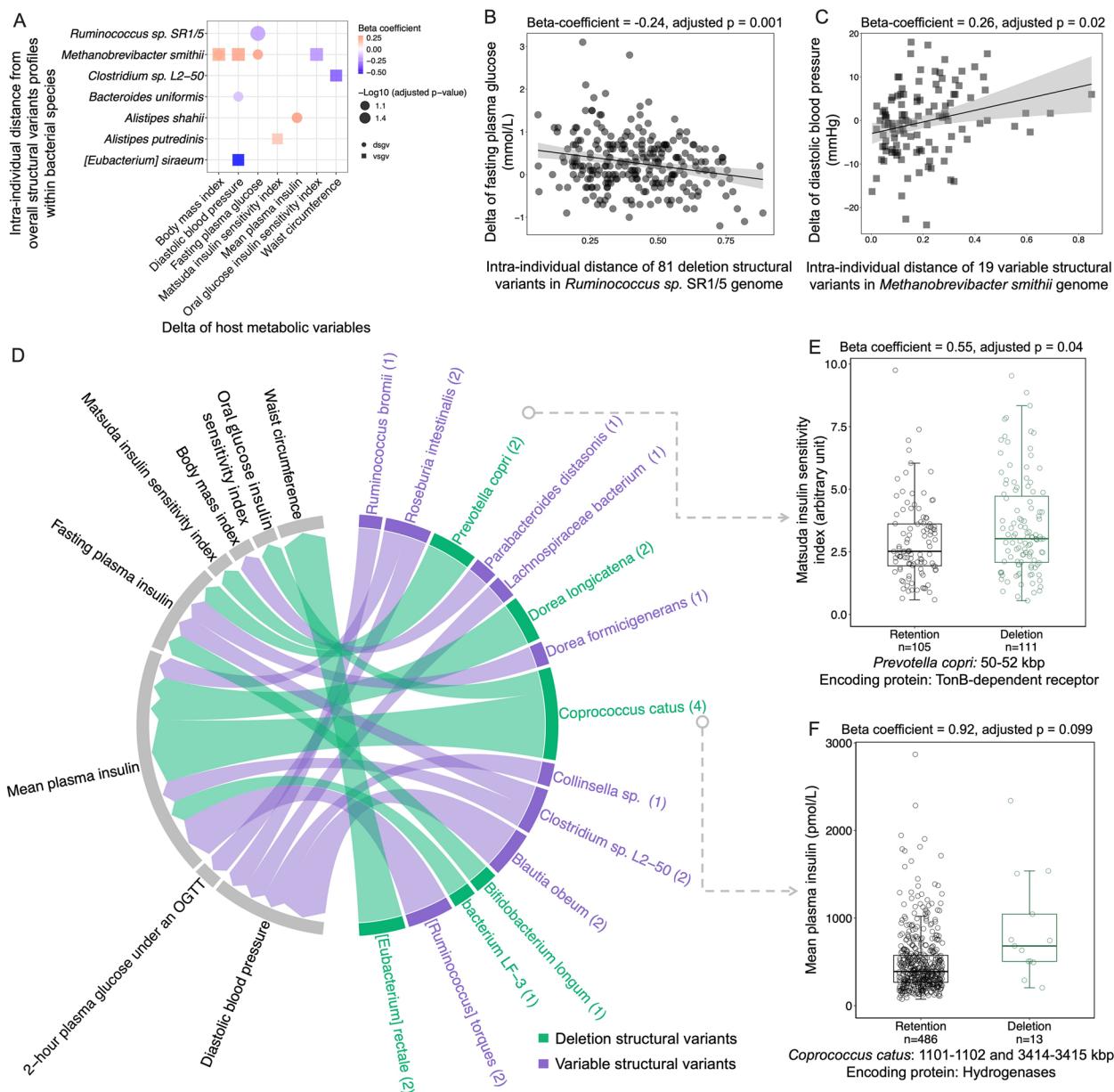


Fig. 7 (See legend on previous page.)

Despite the reduction in the richness of bacteria, viruses, and microbial pathways over time, we mapped the presence of 80 dominant core bacterial species and 78 dominant core microbial pathways. Moreover, microbial pathways showed only a small decrease in richness from study baseline to endline, with major functional potentials being largely preserved. Collectively, these findings suggest a robust microbial ecology over 4 years in prediabetes individuals that maintains essential microbial functions, likely through functional redundancy and ecological compensation [8]. The presence of a core

bacterial microbiota and stable microbial pathways in prediabetes individuals across diverse European demographics may suggest that these features play an essential role in gut microbial ecology that strive to maintain host metabolic health. Future lifestyle and drug interventions are warranted to explore whether targeting these core microbiota components can add to prevention of T2D in individuals at risk.

Notably, no core viral microbiota at species levels was observed in our study, potentially indicating a higher sensitivity of viral species to environmental factors

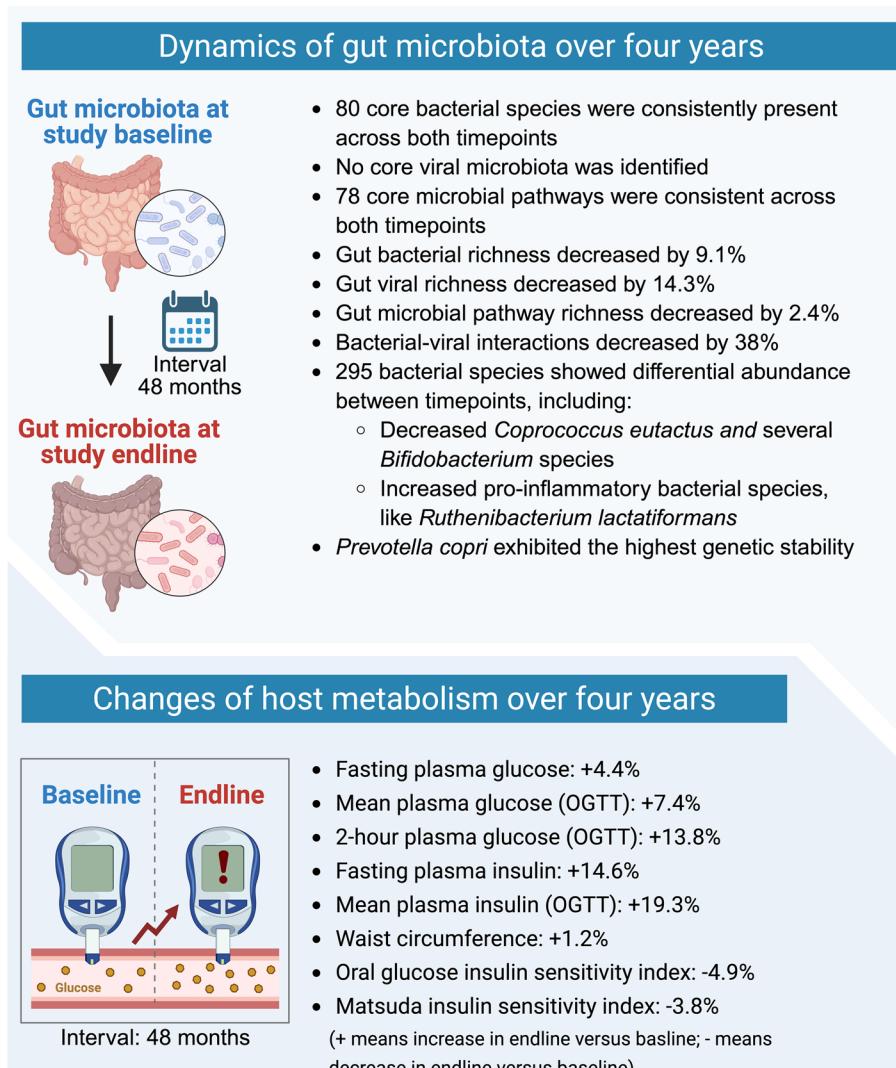


Fig. 8 Summary of temporal changes in host metabolism and gut microbiota dynamics in prediabetes over 4-year follow-up. Created with biorender.com

or host health conditions [57]. Additionally, the relatively shallow sequencing depth (with a rarefaction to 3.9 million reads) may have further limited the detection of low-abundance viral species, contributing to the absence of a distinct core gut viral microbiota in the present study.

A particularly interesting observation was the decrease of bacterial-viral interactions over the four years, coherently with the observed reduction in both bacterial and viral richness. The decline in virulent viruses may imply a reduced viral predation pressure, which could lead to less regulation of bacterial populations [58]. Viruses regulate bacterial communities by preventing overgrowth, thereby maintaining balance within the microbiota [59]. A reduction in these

interactions might allow opportunistic or pathogenic bacterial species to dominate, potentially exacerbating metabolic dysfunction [60].

Regarding the interplay of gut microbiota and host metabolism, we observed several correlations between temporal changes in microbial abundance and host metabolic markers. For example, we found inverse co-variations between the abundance of *Clostridium sp. CAG:253* and 2-hour plasma glucose after an OGTT, as well as positive co-variations between *Ruminococcaceae bacterium D5* abundance and insulin sensitivity. The findings that align with previous reports [55, 61, 62] may suggest a beneficial role for these gut bacteria in enhancing glucose metabolism and insulin sensitivity of their host.

At the bacterial genetics level, we observed significant associations between bacterial genome variation and host metabolic changes over time. Notably, the retention of hydrogenase-encoding bacterial genes in *Coprococcus catus* was linked to low plasma insulin level. Bacterial hydrogenases may play a critical role in butyrate production [63] and in neutralization of reactive oxygen species [64], both of which may potentially enhance host insulin sensitivity.

In addition, the temporal association analysis identified key bacterial species influencing the overall dynamics of the gut bacterial community. Thus, changes in the abundance of *Alistipes putredinis*, *Oxalobacter formigenes*, and *Coprobacter secundus* were positively linked to increases in bacterial species richness. It has been reported that *A. putredinis* supports butyrate production in a bacterial community [65], *O. formigenes* aids in oxalate regulation and hyperoxaluria prevention [66–68], and *C. secundus* contributes to butyrate synthesis [69, 70]. These versatile bacteria may potentially foster cross-feeding to other bacteria, enhancing microbial diversity. Conversely, temporal changes of the abundance of *Prevotella copri*, with its anti-inflammatory and short-chain fatty acid (SCFA)-producing functions [71], showed an inverse association with the measure of intra-individual distance, underscoring its potential role in stabilizing gut microbiota resilience and temporal stability.

In accordance with what has been reported in a previous study of the gut bacterial microbiota in healthy individuals [8], we observed that the dynamics of gut microbiota tend to follow specific patterns. For example, high-abundance bacterial species in our study were more stable both within individuals and between individuals, while low-abundance bacterial species often exhibited multi-modal distributions in abundance as indicated by a high variance. These findings may imply that high-abundance bacterial species which often are evolutionarily dominant within microbial communities [72] are more resilient to perturbations. Another pattern we observed was the inverse relationship between bacterial species richness at study baseline and the temporal stability as measured by the intra-individual distance metric, indicating that a diverse microbial community may confer resilience against environmental or physiological perturbations [7].

Our study has limitations. The present prospective study focused on the dynamics of the gut microbiota in prediabetes cases and was not designed to explore any role of the gut microbiota in prediction of metabolic outcomes over the four years. Achieving such a goal would require a much larger study population monitored over a much longer period, along with detailed information on additional potential confounders, such as medication,

lifestyle factors and interfering comorbidities. Although height has been linked to 2-hour glucose levels [73], we did not adjust for it as its effect is largely captured by sex, and including both would risk multicollinearity, but it may be relevant in future stratified or mediation analyses. To explore if the aberrant gut microbiota in prediabetes is a therapeutic target, drug or lifestyle interventions are needed. Another shortcoming includes the DNA extraction and sequencing protocols that were optimized for bacterial DNA recovery and not for viral recovery. Therefore, the actual genetic material analyzed may only represents a fraction of the total viral DNA in stool samples.

Conclusions

In conclusion, in this 4-year longitudinal study of European prediabetic patients, we observed a decline in gut bacterial and viral richness, microbial functional potentials and bacterial-viral interactions as well as changes of bacterial relative abundance, and bacterial genetics, with the former being directly linked to a deterioration of metabolic health. At present, it is unsettled whether the observed dynamics of the gut microbiota in individuals with prediabetes reflect adaptations or contributions to the observed metabolic deterioration.

Abbreviations

OGTT	Oral glucose tolerance test
T2D	Type 2 diabetes mellitus
IMI DIRECT	Innovative Medicines Initiative Diabetes Research on Patient Stratification
SOPs	Standard operation procedures
BMI	Body mass index
AUC	Area under the curve
OGIS	Oral glucose insulin sensitivity
PCR	Polymerase chain reaction
SV	Structural variants
PCoA	Principal coordinate analysis
PERMANOVA	Permutational multivariate analysis of variance
SparCC	Sparse correlations for compositional data
IQR	Interquartile range
DMI	Degree of microbial individuality
ICC	Intraclass correlation coefficient

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-025-01508-7>.

Additional file 1.

Additional file 2.

Acknowledgements

We extend our heartfelt thanks to the participants of the IMI-DIRECT study for volunteering in phenotyping. We also appreciate the dedication of the clinical and technical staff across the four European study centers who contributed to the recruitment and phenotypic assessment of participants.

Author contributions

In the IMI-DIRECT consortium projects, JWB, ML, OP, PWF, TH, HV, TH, JV, MCG, KHA, JV, RK, MW, Lth, HC, EC, HST, PJE, and EW either supervised or conducted the recruitment and phenotyping of study participants. Bacterial cell counting

was conducted by LL and ES. Library preparation and metagenomic sequencing were carried out by AB and PF. LL, YF, JV, MCG, HKP, AD and OP were responsible for microbiota and phenotype data quality assurance and integrative analyses of the gut microbiota and host phenotypes. The manuscript was drafted by LL with contributions by YF and OP. All authors involved in these studies contributed to various aspects such as participant recruitment, phenotyping, study coordination, data management, or data interpretation. OP conceived, designed, and supervised the IMI-DIRECT microbiota project. All authors read and approved the final manuscript.

Funding

Open access funding provided by Copenhagen University. The work leading to this publication has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement 115317 (DIRECT), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and European Federation of Pharmaceutical Industries and Associations companies' in-kind contribution. Information on the project can be found at <https://www.direct-diabetes.org/>. European Union's Seventh Framework Programme, FP7/2007–2013, FP7/2007–2013, FP7/2007–2013, FP7/2007–2013, FP7/2007–2013, FP7/2007–2013, FP7/2007–2013, FP7/2007–2013

Data Availability

Owing to the informed consent obtained from study participants, the necessary national ethical approvals for this research, and adherence to the European General Data Protection Regulation (GDPR), individual-level phenotypic and microbiota data are not available for public access. However, data access can be granted upon request. Requests should be directed towards the corresponding author (oluf@sund.ku.dk) or the management team of the IMI-DIRECT Study (DIRECTdataaccess@Dundee.ac.uk), which oversees a formal data access protocol. Requests are typically reviewed within one month, and if approved, the data will be shared in accordance with GDPR requirements following the execution of a Data Transfer Agreement (DTA). The codes used in this study are publicly available on GitHub [74] (<https://github.com/Liwei-Lyu/IMI-DIRECT-dynamics>).

Declarations

Ethics approval and consent to participate.

The DIRECT study (Diabetes Research on Patient Stratification) was registered under ClinicalTrials.gov ID NCT03814915. Approval of the study protocol was obtained from each of the regional research ethics review boards separately: Lund, Sweden: 2013031210545997; Copenhagen, Denmark: H-1–2012-166 and H-1–2012-100; Amsterdam, the Netherlands: NL40099.029.12; Newcastle, Dundee, and Exeter, UK: 12/NE/0132. All participants gave written informed consent to participate in the study at enrollment and the research conformed to the ethical principles for medical research involving human participants outlined in the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

RK has received consulting fees from Novo Nordisk; he was also funded by a STAR Award Novo Nordisk co-financed PhD fellowship and a Novo Nordisk Foundation postdoctoral fellowship (NNF18OC0031650). PWF has received research funding from Boehringer Ingelheim, Eli Lilly, Janssen, Novo Nordisk A/S, Sanofi Aventis and Servier, received consulting fees from Eli Lilly, Novo Nordisk and Zoe Global Ltd and has stock options in Zoe Global Ltd. HR is an employee of Boehringer Ingelheim and a shareholder of Sanofi Aventis. MMcC declares that the views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health; he has served on advisory panels for Pfizer, Novo Nordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda; as of June 2019, he is an employee of Genentech, and a holder of Roche stock. BJ and PBM are employees of Sanofi Deutschland GmbH. IP is employed by Eli Lilly Regional Operations GmbH. HR is an employee of Boehringer Ingelheim International GmbH. MR is employed by Novo Nordisk A/S. OP and YF are

co-founders of GutCRINE. The remaining authors declare that they have no competing interests.

Author details

- ¹Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
- ²Center for Clinical Metabolic Research, Herlev-Gentofte University Hospital, Hellerup, Denmark.
- ³Cmbio (Denmark), Copenhagen, Denmark.
- ⁴INSERM UMR 1283, CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille University Hospital, 59000 Lille, France.
- ⁵Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK.
- ⁶Department of Clinical Sciences, Lund University, Malmö, Sweden.
- ⁷Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Bavaria, Germany.
- ⁸German Center for Diabetes Research (DZD), Ingolstädter Landstr. 1, Neuherberg 85764, Germany.
- ⁹Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, The Capital Region, Copenhagen, Denmark.
- ¹⁰Section for Bioinformatics, Department of Health Technology, Technical University of Denmark, Kgs. Lyngby, Denmark.
- ¹¹Disease Systems Biology Program, Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark.
- ¹²Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland.
- ¹³Department of Diagnostic Radiology, Copenhagen University Hospital Herlev Gentofte, Herlev, Denmark.
- ¹⁴Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland.
- ¹⁵Department of General Practice, Amsterdam UMC- location VUmc, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands.
- ¹⁶Section for Nutrition Research, Faculty of Medicine, Imperial College London, London, UK.
- ¹⁷Population Health & Genomics, School of Medicine, University of Dundee, Dundee, UK.
- ¹⁸Genetic and Molecular Epidemiology Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Lund University and Helsingborg Hospital, Helsingborg, Sweden.
- ¹⁹Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany.
- ²⁰Research Unit Molecular Endocrinology and Metabolism, Genome Analysis Center, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, Neuherberg 85764, Germany.
- ²¹Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
- ²²Medical Department, Zealand University Hospital, Køge, Denmark.
- ²³Strategy & Innovation, Sanofi-Aventis Deutschland GmbH, Frankfurt, Germany.
- ²⁴Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.
- ²⁵CNR Institute of Neuroscience, Padova, Italy.
- ²⁶Institute of Clinical and Biological Sciences, University of Exeter Medical School, Exeter, UK.
- ²⁷Blood Sciences, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK.
- ²⁸R&D Global Development, Translational Medicine & Clinical Pharmacology (TMCP), Sanofi-Aventis Deutschland GmbH, Frankfurt, Germany.
- ²⁹Eli Lilly, Regional Operations GmbH, Vienna, Austria.
- ³⁰Clinical Obesity, Department of Clinical Sciences, Lund University Diabetes Centre, CRC, Lund University, Skåne University Hospital, Malmö, Sweden.
- ³¹Boehringer Ingelheim, International GmbH, Ingelheim am Rhein, Germany.
- ³²Amsterdam UMC, Location VUmc, Department of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands.
- ³³Affinity Proteomics, Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH - Royal Institute of Technology, Solna, Sweden.
- ³⁴Faculty of Medical and Health Sciences, University of Copenhagen, Copenhagen, Denmark.
- ³⁵Steno Diabetes Center Copenhagen, Herlev, Denmark.
- ³⁶Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle, UK.
- ³⁷Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
- ³⁸Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, UK.
- ³⁹Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK.
- ⁴⁰Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstädter Landstraße 1, Neuherberg, Germany.
- ⁴¹Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore, Singapore.
- ⁴²Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Vrazov Trg 2, Ljubljana 1000, Slovenia.

Received: 22 December 2024 Accepted: 30 June 2025
Published online: 15 July 2025

References

1. American Diabetes Association. 2. Classification and diagnosis of diabetes. *Diabetes Care*. 2016;40:S11–24.
2. Rooney MR, Fang M, Ogurtsova K, Ozkan B, Echouffo-Tcheugui JB, Boyko EJ, et al. Global prevalence of prediabetes. *Diabetes Care*. 2023;46:1388–94.
3. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. *Lancet*. 2012;379:2279–90.
4. Allin KH, Tremaroli V, Caesar R, Jensen BAH, Damgaard MTF, Bahl MI, et al. Aberrant intestinal microbiota in individuals with prediabetes. *Diabetologia*. 2018;61:810–20.
5. Gravdal K, Kirste KH, Grzelak K, Kirubakaran GT, Leissner P, Saliou A, et al. Exploring the gut microbiota in patients with pre-diabetes and treatment naïve diabetes type 2 - a pilot study. *BMC Endocr Disord*. 2023;23:179.
6. Wu H, Tremaroli V, Schmidt C, Lundqvist A, Olsson LM, Krämer M, et al. The gut microbiota in prediabetes and diabetes: a population-based cross-sectional study. *Cell Metab*. 2020;32:379–390.e3.
7. Chen L, Wang D, Garmava S, Kurilshikov A, Vila AV, Gacesa R, et al. The long-term genetic stability and individual specificity of the human gut microbiome. *Cell*. 2021;184:2302–2315.e12.
8. Olsson LM, Boulund F, Nilsson S, Khan MT, Gummesson A, Fagerberg L, et al. Dynamics of the normal gut microbiota: a longitudinal one-year population study in Sweden. *Cell Host Microbe*. 2022;30:726–739.e3.
9. Zhou X, Shen X, Johnson JS, Spakowicz DJ, Agnello M, Zhou W, et al. Longitudinal profiling of the microbiome at four body sites reveals core stability and individualized dynamics during health and disease. *Cell Host Microbe*. 2024;32:506–526.e9.
10. Bizzotto R, Jennison C, Jones AG, Kurbasic A, Tura A, Kennedy G, et al. Processes underlying glycemic deterioration in type 2 diabetes: an IMI DIRECT study. *Diabetes Care*. 2020;44:511–8.
11. Stančáková A, Javorský M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes*. 2009;58:1212–21.
12. Rutters F, Nijpels G, Elders P, Stehouwer CDA, van der Heijden AA, Groeneveld L, et al. Cohort profile: the Hoorn studies. *Int J Epidemiol*. 2018;47:396–396j.
13. Hills SA, Balkau B, Coppock SW, Dekker JM, Mari A, Natali A, et al. The EGIR-RISC STUDY (The European group for the study of insulin resistance: relationship between insulin sensitivity and cardiovascular disease risk): I. Method Object Diabetol. 2004;47:566–70.
14. Thuesen BH, Cerqueira C, Aadahl M, Ebstrup JF, Toft U, Thyssen JP, et al. Cohort profile: the health 2006 cohort, research centre for prevention and health. *Int J Epidemiol*. 2014;43:568–75.
15. Christensen AL, Ekholm O, Glümer C, Andreasen AH, Hvidberg MF, Kristensen PL, et al. The Danish National Health Survey 2010. Study design and respondent characteristics. *Scand J Public Health*. 2012;40:391–7.
16. Dantoft TM, Ebstrup JF, Linneberg A, Skovbjerg S, Madsen AL, Mehlsen J, et al. Cohort description: the Danish study of functional disorders. *CLEP*. 2017;9:127–39.
17. Manjer J, Carlsson S, Elmståhl S, Gullberg B, Janzon L, Lindström M, et al. The Malmö diet and cancer study: representativity, cancer incidence and mortality in participants and non-participants. *Eur J Cancer Prev*. 2001;10:489.
18. Koivula RW, Forgie IM, Kurbasic A, Viñuela A, Heggie A, Giordano GN, et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: descriptive characteristics of the epidemiological studies within the IMI DIRECT Consortium. *Diabetologia*. 2019;62:1601–15.
19. American Diabetes Association Professional Practice Committee. 2. Diagnosis and classification of diabetes: standards of care in diabetes—2024. *Diabetes Care*. 2023;47:S20–42.
20. Mari A, Pacini G, Murphy E, Ludvik B, Nolan JJ. A model-based method for assessing insulin sensitivity from the oral glucose tolerance test. *Diabetes Care*. 2001;24:539–48.
21. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care*. 1999;22:1462–70.
22. Fan Y, Støving RK, Berreia Ibrahim S, Hyötyläinen T, Thirion F, Arora T, et al. The gut microbiota contributes to the pathogenesis of anorexia nervosa in humans and mice. *Nat Microbiol*. 2023;8:787–802.
23. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
24. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCtools. *GigaScience*. 2021;10: giab008.
25. Pinto Y, Chakraborty M, Jain N, Bhatt AS. Phage-inclusive profiling of human gut microbiomes with Phanta. *Nat Biotechnol*. 2024;42:651–62.
26. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol*. 2021;39:105–14.
27. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2007;35:D61–65.
28. Nayfach S, Páez-Espino D, Call L, Low SJ, Sberro H, Ivanova NN, et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol*. 2021;6:960–70.
29. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Turnbaugh P, Franco E, Brown CT, editors. eLife*. 2021;10:e65088.
30. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res*. 2020;48:D445–53.
31. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, et al. Structural variation in the gut microbiome associates with host health. *Nature*. 2019;568:43–8.
32. Liu R, Zou Y, Wang W-Q, Chen J-H, Zhang L, Feng J, et al. Gut microbial structural variation associates with immune checkpoint inhibitor response. *Nat Commun*. 2023;14:7421.
33. Gower JC. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*. 1966;53:325–38.
34. Roswell M, Dushoff J, Winfree R. A conceptual guide to measuring species diversity. *Oikos*. 2021;130:321–38.
35. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420–8.
36. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. *PLoS Comput Biol*. 2012;8: e1002687.
37. Squillario M, Bonaretti C, La Valle A, Di Marco E, Piccolo G, Minuto N, et al. Gut-microbiota in children and adolescents with obesity: inferred functional analysis and machine-learning algorithms to classify microorganisms. *Sci Rep*. 2023;13:11294.
38. Xiao L, Zhang F, Zhao F. Large-scale microbiome data integration enables robust biomarker identification. *Nat Comput Sci*. 2022;2:307–16.
39. Sarkar A, Mandal S. Bifidobacteria-insight into clinical outcomes and mechanisms of its probiotic action. *Microbiol Res*. 2016;192:159–71.
40. Yang R, Shan S, Shi J, Li H, An N, Li S, et al. *Coprococcus eutactus*, a potent probiotic, alleviates colitis via acetate-mediated IgA response and microbiota restoration. *J Agric Food Chem*. 2023;71:3273–84.
41. Zhou X, Baumann R, Gao X, Mendoza M, Singh S, Sand IK, et al. Gut microbiome of multiple sclerosis patients and paired household healthy controls reveal associations with disease risk and course. *Cell*. 2022;185:3467–3486.e16.
42. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen A-M, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe*. 2015;17:260–73.
43. de Goffau MC, Luopajarvi K, Knip M, Ilonen J, Ruohola T, Härkönen T, et al. Fecal microbiota composition differs between children with β -cell autoimmunity and those without. *Diabetes*. 2013;62:1238–44.
44. Yue T, Tan H, Wang C, Liu Z, Yang D, Ding Y, et al. High-risk genotypes for type 1 diabetes are associated with the imbalance of gut microbiome and serum metabolites. *Front Immunol*. 2022;13: 103393.
45. Cinek O, Kramna L, Lin J, Oikarinen S, Kolarova K, Ilonen J, et al. Imbalance of bacteriome profiles within the Finnish Diabetes Prediction and Prevention study: parallel use of 16S profiling and virome sequencing in stool

samples from children with islet autoimmunity and matched controls. *Pediatr Diabetes*. 2017;18:588–98.

46. Mokhtari P, Jambal P, Metos JM, Shankar K, Anand Babu PV. Microbial taxonomic and functional shifts in adolescents with type 1 diabetes are associated with clinical and dietary factors. *EBioMedicine*. 2023;93: 104641.
47. van Heck JJP, Gacesa R, Stienstra R, Fu J, Zhernakova A, Harmsen HJM, et al. The gut microbiome composition is altered in long-standing type 1 diabetes and associates with glycemic control and disease-related complications. *Diabetes Care*. 2022;45:2084–94.
48. Stanescu DE, Lord K, Lipman TH. The epidemiology of type 1 diabetes in children. *Endocrinol Metab Clin North Am*. 2012;41:679–94.
49. Ioannou A, Berkholz MD, Geerlings SY, Belzer C. Akkermansia muciniphila: biology, microbial ecology, host interactions and therapeutic potential. *Nat Rev Microbiol*. 2025;23:162–77.
50. Cani PD, Depommier C, Derrien M, Everard A, de Vos WM. Akkermansia muciniphila: paradigm for next-generation beneficial microorganisms. *Nat Rev Gastroenterol Hepatol*. 2022;19:625–37.
51. Zeng Z, Chen M, Liu Y, Zhou Y, Liu H, Wang S, et al. Role of Akkermansia muciniphila in insulin resistance. *J Gastroenterol Hepatol*. 2025;40:19–32.
52. Liu E, Ji X, Zhou K. Akkermansia muciniphila for the prevention of type 2 diabetes and obesity: a meta-analysis of animal studies. *Nutrients*. 2024;16: 3440.
53. Depommier C, Van Hul M, Everard A, Delzenne NM, De Vos WM, Cani PD. Pasteurized Akkermansia muciniphila increases whole-body energy expenditure and fecal energy excretion in diet-induced obese mice. *Gut Microbes*. 2020;11:1231–45.
54. Pellegrino A, Coppola G, Santopaolo F, Gasbarrini A, Ponziani FR. Role of Akkermansia in human diseases: from causation to therapeutic properties. *Nutrients*. 2023;15: 1815.
55. Mei Z, Wang F, Bhosle A, Dong D, Mehta R, Ghazi A, et al. Strain-specific gut microbial signatures in type 2 diabetes identified in a cross-cohort analysis of 8,117 metagenomes. *Nat Med*. 2024;30:2265–76.
56. Osborn LJ, Schultz K, Massey W, DeLucia B, Choucair I, Varadharajan V, et al. A gut microbial metabolite of dietary polyphenols reverses obesity-driven hepatic steatosis. *Proc Natl Acad Sci U S A*. 2022;119: e2202934119.
57. Ritz NL, Draper LA, Bastiaanssen TFS, Turkington CJR, Peterson VL, van de Wouw M, et al. The gut virome is associated with stress-induced changes in behaviour and immune responses in mice. *Nat Microbiol*. 2024;9:359–76.
58. Fernández L, Rodríguez A, García P. Phage or foe: an insight into the impact of viral predation on microbial communities. *ISME J*. 2018;12:1171–9.
59. Mahmud MdR, Tamanna SK, Akter S, Mazumder L, Akter S, Hasan MdR, et al. Role of bacteriophages in shaping gut microbial community. *Gut Microbes*. 16(1):2390720. <https://doi.org/10.1080/19490976.2024/2390720>.
60. Bäumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature*. 2016;535:85–93.
61. Wang J, Qie J, Zhu D, Zhang X, Zhang Q, Xu Y, et al. The landscape in the gut microbiome of long-lived families reveals new insights on longevity and aging - relevant neural and immune function. *Gut Microbes*. 2022;14: 2107288.
62. Sakhaee K. Unraveling the mechanisms of obesity-induced hyperoxaluria. *Kidney Int*. 2018;93:1038–40.
63. Campbell A, Gdanetz K, Schmidt AW, Schmidt TM. H₂ generated by fermentation in the human gut microbiome influences metabolism and competitive fitness of gut butyrate producers. *Microbiome*. 2023;11:133.
64. Smith NW, Shorten PR, Altermann EH, Roy NC, McNabb WC. Hydrogen cross-feeders of the human gastrointestinal tract. *Gut Microbes*. 2019;10(3):270–88. <https://doi.org/10.1080/19490976.2018.1546522>.
65. Ke X, Walker A, Haange S-B, Lagkouvardos I, Liu Y, Schmitt-Kopplin P, et al. Symbiotic-driven improvement of metabolic disturbances is associated with changes in the gut microbiome in diet-induced obese mice. *Mol Metab*. 2019;22:96–109.
66. Daniel SL, Moradi L, Paiste H, Wood KD, Assimos DG, Holmes RP, et al. Forty years of Oxalobacter formigenes, a gutsy oxalate-degrading specialist. *Appl Environ Microbiol*. 2021;87:e00544–621.
67. Liu M, Koh H, Kurtz ZD, Battaglia T, PeBenito A, Li H, et al. Oxalobacter formigenes-associated host features and microbial community structures examined using the American Gut Project. *Microbiome*. 2017;5:108.
68. Jaunet-Lahary T, Shimamura T, Hayashi M, Nomura N, Hirasawa K, Shimizu T, et al. Structure and mechanism of oxalate transporter OxIT in an oxalate-degrading bacterium in the gut microbiota. *Nat Commun*. 2023;14:1730.
69. Bold TD. A new hope; the M. tuberculosis strikes back. *Cell Host and Microbe*. 2023;31:321–2.
70. Luan M, Niu M, Yang P, Han D, Zhang Y, Li W, et al. Metagenomic sequencing reveals altered gut microbial compositions and gene functions in patients with non-segmental vitiligo. *BMC Microbiol*. 2023;23:265.
71. Bedarf JR, Hildebrand F, Coelho LP, Sunagawa S, Bahram M, Goeser F, et al. Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Medicine*. 2017;9:39.
72. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature*. 2012;489:220–30.
73. Palmu S, Rehunen S, Kautiainen H, Eriksson JG, Korhonen PE. Body surface area and glucose tolerance - the smaller the person, the greater the 2-hour plasma glucose. *Diabetes Res Clin Pract*. 2019;157: 107877.
74. Lyu LL. IMI-DIRECT-dynamics. Github; 2025. <https://github.com/Liwei-Lyu/IMI-DIRECT-dynamics>.