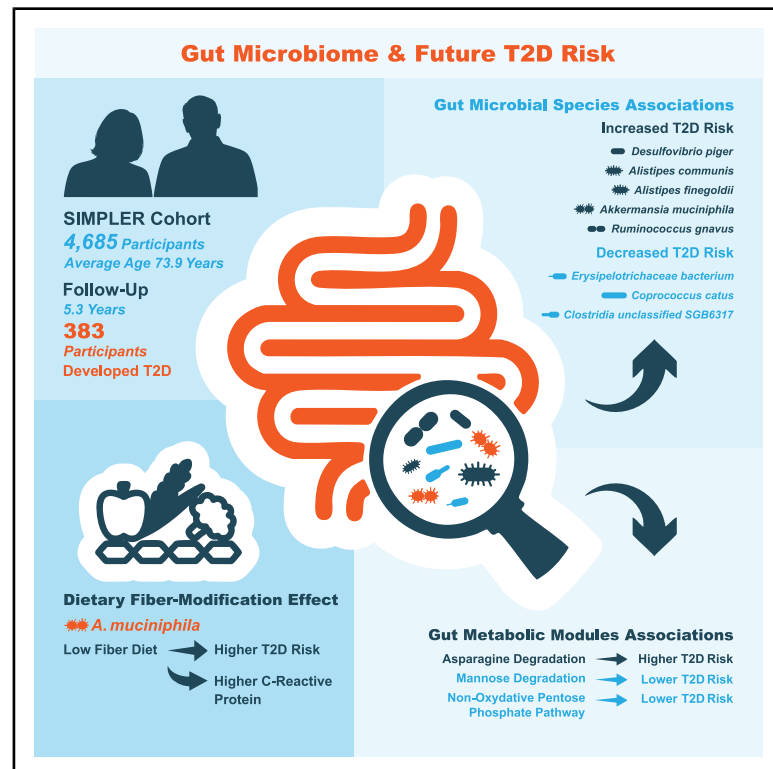


Gut microbiome composition and functional potential associate with incident type 2 diabetes in 4,685 adults from a Swedish prospective cohort

Graphical abstract



Authors

Gaël Toubon, Fredrik Boulund, Cecilia Martinez Escobedo, ..., Alicja Wolk, Clemens Wittenbecher, Rikard Landberg

Correspondence

gael.toubon@chalmers.se (G.T.), rikard.landberg@chalmers.se (R.L.)

In brief

Toubon et al. show that the gut microbiome measured years before diagnosis predicts future type 2 diabetes risk. Using a large prospective Swedish cohort of 4,685 adults, they identify specific microbial species and metabolic pathways associated with diabetes incidence, advancing understanding of the microbiome's role in diabetes development and prevention.

Highlights

- Large-scale prospective metagenomics ($n = 4,685$) links gut microbiome to T2D risk
- Findings are derived from a T2D-free and antidiabetic medication-naïve population
- Nine species and three gut metabolic modules associate with future risk of T2D
- Dietary fiber modifies the association between *A. muciniphila* and T2D risk



Article

Gut microbiome composition and functional potential associate with incident type 2 diabetes in 4,685 adults from a Swedish prospective cohort

Gaël Toubon,^{1,2,*} Fredrik Boulund,² Cecilia Martinez Escobedo,¹ Carl Brunius,¹ Lars Engstrand,² Susanna C. Larsson,^{3,4} Elise Nordin,¹ Ina Schuppe-Koistinen,² Alicja Wolk,³ Clemens Wittenbecher,¹ and Rikard Landberg^{1,5,*}

¹Division of Food and Nutrition Science, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden

²Centre for Translational Microbiome Research, Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

³Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

⁴Medical Epidemiology, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

⁵Lead contact

*Correspondence: gael.toubon@chalmers.se (G.T.), rikard.landberg@chalmers.se (R.L.)

<https://doi.org/10.1016/j.xcrm.2026.102835>

SUMMARY

Cross-sectional studies link gut microbiome alterations to type 2 diabetes (T2D), but prospective evidence remains limited. We aim to identify taxonomic and functional features associated with future T2D risk. We analyze shotgun metagenomic data from 4,685 participants (mean age, 73.9 years; 49.0% women) in the Swedish SIMPLER cohort, followed for a median 5.3 years, during which 383 developed T2D. Six species are associated with increased T2D risk: *Desulfovibrio piger*, *Alistipes communis*, *Alistipes finegoldii*, *Akkermansia muciniphila*, *Ruminococcus gnavus*, and GGB3614_SGB4886 (Lachnospiraceae), while three are protective: *Erysipelotrichaceae* bacterium, *Coprococcus catus*, and *Clostridia* unclassified SGB6317. We observe context-specific associations, including a dietary fiber-modified effect for *A. muciniphila* indicative of diet-dependent patterns. Three gut metabolic modules are associated with incident T2D: asparagine degradation (higher risk), mannose degradation, and the non-oxidative pentose phosphate pathway (lower risk). These prospective findings offer insights into T2D etiology and may support microbiome-informed strategies for risk prediction and prevention.

INTRODUCTION

Type 2 diabetes (T2D) is a major global public health burden, and its prevalence is projected to increase by 61.2% until 2050, affecting more than 1.27 billion people worldwide.¹ There is a growing recognition of the role of the gut microbiome in the pathophysiology of T2D.^{2–6} Currently, most of the evidence linking the gut microbiome to T2D stems from cross-sectional studies, where reduced gut microbial diversity and a lower abundance of butyrate-producing bacteria in individuals with T2D are among the most recurrent findings.^{5,7} However, due to their cross-sectional nature, these studies are limited in their ability to infer causality. Moreover, increasing evidence suggests that some of the observed associations may be confounded by the use of antidiabetic drugs that can impact the gut microbiome composition, such as metformin.^{8,9}

Findings from experimental models support the role of the gut microbiome in T2D development through several interconnected pathways, including altered glucose and lipid metabolism, reduced insulin sensitivity, increased gut permeability, and immune modulation.³ A recurring feature across these pathways is the emergence of chronic low-grade inflammation, which con-

tributes to insulin resistance and disease progression. Changes in gut microbiome composition, shaped by dietary and other environmental factors, appear to influence these processes.¹⁰ Among dietary factors, fiber intake plays a central role in modulating the gut microbiome composition and activity. Its protective effect against T2D¹¹ is thought to be partly mediated by its prebiotic action on fiber-fermenting bacteria that produce short-chain fatty acids (SCFAs), such as butyrate.¹² In contrast, low fiber intake has been linked to reduced SCFA production and impaired intestinal mucus barrier,^{13,14} which may permit microbial translocation that triggers metabolic endotoxemia ultimately promoting systemic low-grade inflammation and increasing T2D risk.

Despite mounting evidence, prospective studies investigating the gut microbiome in relation to incident T2D remain scarce. Most of the few available studies rely on 16S rRNA gene sequencing¹⁵ and are limited by small sample sizes,^{16,17} with inconsistent findings across studies. To date, only one large-scale prospective study using shotgun metagenomics has been published.¹⁸ This study identified four microbial species (including *Ruminococcus gnavus*) associated with incident T2D in a Finnish population. Besides, while these studies have



Table 1. Baseline characteristics of the population

	Uppsala	Västmanland	
	SMCC-U (N = 925)	SMCC-V (N = 1,370)	COSMC-V (N = 2,390)
Women	925 (100%)	1,370 (100%)	0 (0%)
Baseline age (years)	78.8 (6.64)	73.7 (3.18)	72.1 (4.90)
Education			
<10 years	191 (20.6%)	341 (24.9%)	543 (22.7%)
10–12 years	323 (34.9%)	670 (48.9%)	1,278 (53.5%)
>12 years	411 (44.4%)	359 (26.2%)	569 (23.8%)
Height (cm)	162 (6.67)	163 (6.14)	177 (6.46)
Waist (cm)	87.7 (11.6)	89.0 (11.6)	97.4 (10.1)
Smoking status			
Non-smokers	572 (61.8%)	805 (58.8%)	1,160 (48.5%)
Current smokers	46 (5.0%)	84 (6.1%)	147 (6.2%)
Former smokers	307 (33.2%)	481 (35.1%)	1,083 (45.3%)
Total energy intake (kcal/day)	1,790 (497)	1,850 (526)	2,450 (683)
Coffee intake (cups/day)	2.51 (1.26)	2.52 (1.56)	2.80 (2.67)
Alcohol intake (g/day)	5.12 (5.57)	4.95 (5.12)	10.5 (9.29)
Walking/biking this past month			
Never	88 (9.5%)	199 (14.5%)	382 (16.0%)
<20 min	431 (46.6%)	559 (40.8%)	898 (37.6%)
20–40 min	194 (21.0%)	350 (25.5%)	558 (23.3%)
40–60 min	182 (19.7%)	202 (14.7%)	430 (18.0%)
>60 min	30 (3.2%)	60 (4.4%)	122 (5.1%)
Exercise this past month			
Almost never	352 (38.1%)	517 (37.7%)	1,184 (49.5%)
<1 h/week	87 (9.4%)	132 (9.6%)	277 (11.6%)
1 h/week	207 (22.4%)	318 (23.2%)	365 (15.3%)
2–3 h/week	233 (25.2%)	348 (25.4%)	447 (18.7%)
≥4 h/week	46 (5.0%)	55 (4.0%)	117 (4.9%)
Whole grains (g/day)	129 (81.2)	132 (84.6)	174 (105)
Yogurt (g/day)	196 (229)	202 (215)	246 (293)
Red/processed meat (g/day)	33.7 (21.4)	36.8 (23.3)	62.0 (33.3)
Sugary food/sweetened beverages (g/day)	135 (122)	128 (116)	200 (199)
Fasting plasma glucose (mmol/L)	5.55 (0.797)	5.40 (0.807)	5.67 (0.861)
Statin medication	272 (29.4%)	515 (37.6%)	949 (39.7%)

Categorical variables are presented as number of events (%), and continuous variables are presented as mean (SD).

investigated gut microbial composition in relation to incident T2D, only a few have comprehensively examined gut metabolites or other functional readouts in relation to T2D in a prospective setting.^{19–21} This limited evidence emphasizes the need for further research using large-scale datasets and more comprehensive analytical approaches.

To address these limitations, we aimed to investigate both taxonomic and functional gut microbiome features associated with T2D incidence in a T2D medication-naïve population. We leveraged data from the Swedish Infrastructure for Medical Population-based Life-course and Environmental Research (SIMPLER), a well-phenotyped cohort of men and women. We used whole-genome shotgun sequencing to profile the gut microbiome in 4,685 individuals, providing an opportunity to

examine microbial features in relation to incident T2D using a large-scale, prospective design.

RESULTS

Over a median follow-up of 5.3 years (interquartile range, 3.3–6.9 years), a total of 383 individuals (8.2%) with incident T2D were identified among 4,685 participants. This included 925 from the Swedish Mammography Cohort clinical subcohort Uppsala (SMCC-U) (mean [SD] age, 78.8 [6.6] years; 34 [3.7%] with incident T2D), 1,370 from the Swedish Mammography Cohort clinical subcohort Västmanland (SMCC-V) (mean [SD] age, 73.7 [3.2] years; 96 [6.8%] with incident T2D), and 2,390 from the Cohort of Swedish Men clinical subcohort Västmanland

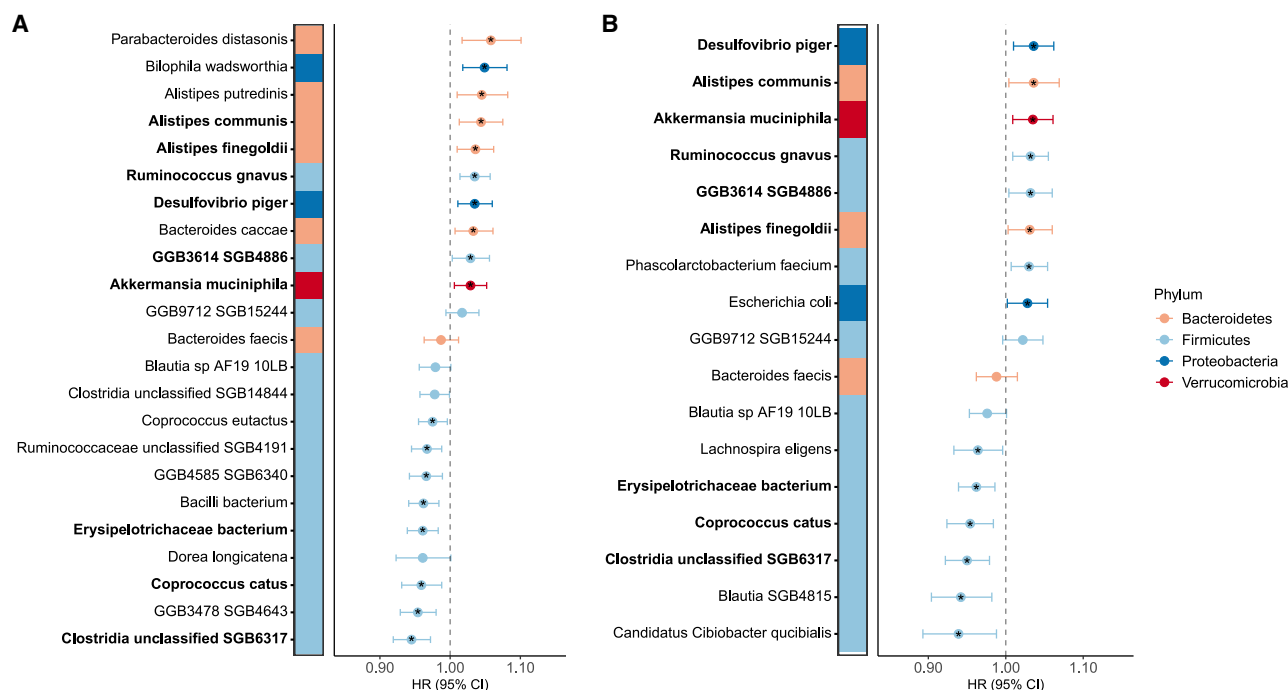


Figure 1. Gut microbial species associated with incident T2D

Hazard ratios (HRs) of Cox regression models between (A) the 23 selected species and incident T2D after Elastic Net feature selection using the full analysis set ([FAS], $n_{\text{total}} = 4,685$ with 383 T2D incident cases) and (B) the 17 selected species and incident T2D after Elastic Net feature selection, using the lag time analysis set ([LTAS], $n_{\text{total}} = 4,633$ with 331 T2D incident cases). All models were adjusted for sex, baseline age, level of education, height, waist circumference, smoking status, walking/cycling, exercise, coffee consumption, daily intakes of total energy, alcohol, whole grains, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth. Species are colored according to their phylum, and species in bold correspond to overlapping species selected by Elastic Net and associated with incident T2D across both analysis sets. * q value < 0.05.

(COSMC-V) (mean [SD] age, 72.1 [4.9] years; 256 [10.7%] with incident T2D). The characteristics of the study participants are reported in Table 1.

Gut microbiome α - and β -diversity and T2D incidence

In the full analysis set (FAS) of 4,685 participants, we observed weak and statistically non-significant inverse associations between α -diversity and incident T2D (species richness: hazard ratio [HR] = 0.999, 95% confidence interval [CI] = 0.997–1.001, $q = 0.567$; Shannon index: HR = 0.847, 95% CI = 0.659–1.09, $q = 0.263$). For β -diversity, two principal components (PCs) were statistically significantly associated with higher T2D risk (PC5: HR = 1.016, 95% CI = 1.007–1.025, $q = 0.007$ and PC6: HR = 1.013, 95% CI = 1.004–1.023, $q = 0.028$) (Table S2). However, these associations were driven to null in the lag time analysis set (LTAS) of 4,633 participants (q value > 0.05) (Table S2).

Gut microbiome species associations with T2D incidence

For species-level analyses in the FAS, 23 species were consistently selected by Elastic Net as strong predictors of future T2D development (Figures S2 and S3), suggesting that they may form a core microbiome set of particular biological relevance to T2D development. Predictive models including these 23 selected species demonstrated moderate ability to differentiate risk (mean C-index, 0.78 $\pm$ 0.02) (Figure S4). Among the

23 species selected as T2D predictors, 18 were significantly associated with incident T2D in subsequent multivariable Cox regression models (q value < 0.05) (Figure 1A). Ten species showed a positive association and eight showed a negative association with T2D risk. Half of the high T2D risk-associated species belonged to the Bacteroidetes (Bacteroidota) phylum, while all inversely associated species belonged to the Firmicutes (Bacillota) phylum.

In the LTAS, 17 species emerged from the Elastic Net selection procedure as strong predictor of incident T2D (Figure S2). Model performance was slightly higher and showed a mean C-index of 0.81 $\pm$ 0.02, indicating good predictive performance (Figure S4). Of the 17 selected species, 12 overlapped with those identified in the FAS (Figure S2), and 9 of these were associated with incident T2D in subsequent multivariable Cox regression models, supporting their robust association with incident T2D (Figure 1B).

Six species were positively associated as follows: *Desulfovibrio piger* (FAS: HR = 1.035, 95% CI = 1.011–1.06, $q = 0.011$; LTAS: HR = 1.036, 95% CI = 1.01–1.062, $q = 0.016$), *Alistipes communis* (FAS: HR = 1.044, 95% CI = 1.013–1.075, $q = 0.011$; LTAS: HR = 1.036, 95% CI = 1.004–1.069, $q = 0.039$), *Akkermansia muciniphila* (FAS: HR = 1.029, 95% CI = 1.006–1.052, $q = 0.021$; LTAS: HR = 1.035, 95% CI = 1.009–1.061, $q = 0.016$), *R. gnavus* (FAS: HR = 1.035, 95% CI = 1.014–1.057, $q < 0.001$; LTAS: HR = 1.032, 95% CI = 1.009–1.055, $q = 0.016$), GGB3614 SGB4886 (corresponding to CAG-194 sp000432915

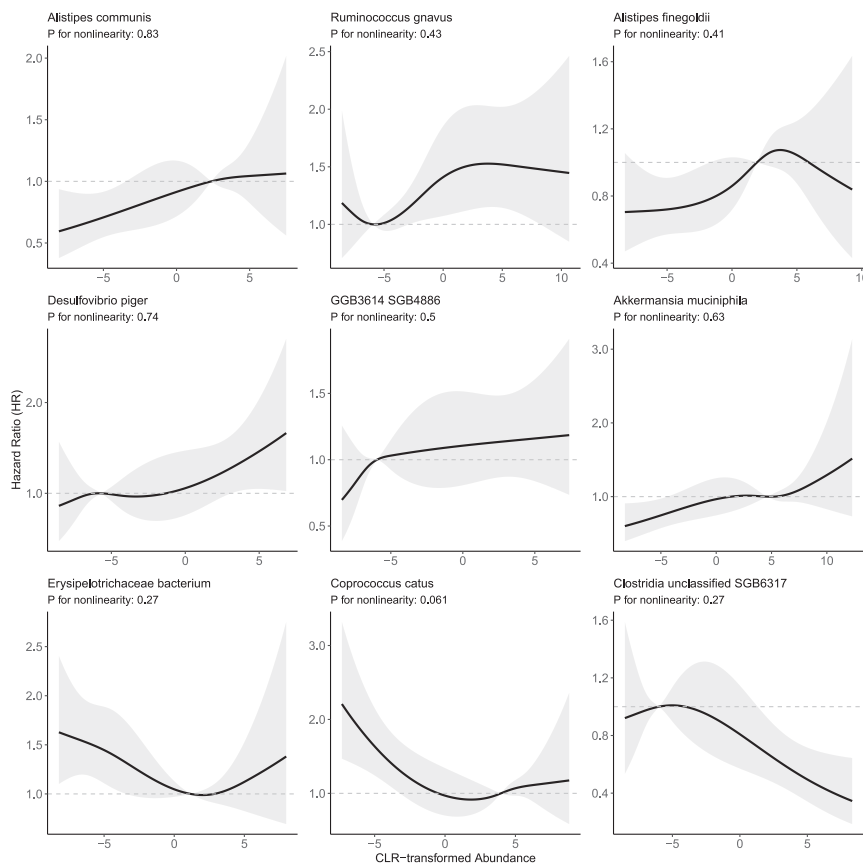


Figure 2. Nonlinear association between gut microbial species abundance and incident T2D

Restricted cubic spline (RCS) curves from Cox proportional hazards models showing adjusted hazard ratios (HRs) for incident T2D across Centered Log-Ratio (CLR) abundance of the 9 robust species identified in both analysis sets. The solid line corresponds to the estimated HR, and the shaded band corresponds to the 95% CI. The horizontal dashed line indicates HR = 1, corresponding to the reference point set at the median abundance of each species. The p value for the nonlinear term (p -nonlinearity) was estimated from a likelihood ratio test comparing the RCS term with a linear term. Models are adjusted for statin medication use, sex, baseline age, education, height, waist circumference, smoking, total energy intake, coffee, alcohol, walking/biking, exercise, whole grains, yogurt, red/processed meat, sweet foods/beverages, aliquoting plate, and sequencing depth. Knots are placed at the 5th, 35th, 65th, and 95th percentiles. The RCS is shown using data from the full analysis set ($n_{\text{total}} = 4,685$ with 383 T2D incident cases).

CI's crossing 1). This pattern suggests a low-abundance penalty with little evidence of further benefit across typical and higher abundances. Overall, within the observed distributions, associations between species abundance and T2D risk were approximately linear for most species (Figure 2). When analyses were restricted to the LTAS, similar patterns were observed (Figure S5).

in the Genome Taxonomy Database [GTDB],²² FAS: HR = 1.029, 95% CI = 1.003–1.056, $q = 0.036$; LTAS: HR = 1.032, 95% CI = 1.004–1.06, $q = 0.039$), and *Alistipes finegoldii* (FAS: HR = 1.036, 95% CI = 1.01–1.062, $q = 0.011$; LTAS: HR = 1.031, 95% CI = 1.003–1.06, $q = 0.039$). Interestingly, these overlapping species showed the strongest associations among the positively associated species in the LTAS. Three species were consistently found inversely associated with incident T2D. These species were *Erysipelotrichaceae bacterium* (FAS: HR = 0.961, 95% CI = 0.939–0.983, $q = 0.004$; LTAS: HR = 0.962, 95% CI = 0.939–0.986, $q = 0.016$), *Coprococcus catus* (FAS: HR = 0.959, 95% CI = 0.931–0.988, $q = 0.011$; LTAS: HR = 0.954, 95% CI = 0.924–0.984, $q = 0.016$), and *Clostridia unclassified SGB6317* (FAS: HR = 0.945, 95% CI = 0.919–0.972, $q = 0.002$; LTAS: HR = 0.95, 95% CI = 0.922–0.979, $q = 0.014$).

We further investigated species-level associations using restricted cubic splines to assess potential deviations from linearity between species abundance and T2D risk. Across the 9 robustly associated species, there was little evidence of nonlinearity (p for nonlinearity ≥ 0.10), except for *C. catus*, which showed borderline evidence (p for nonlinearity = 0.06). For *C. catus*, the decile contrasts referenced to the median showed that elevated risk was confined to the very low tail (10th percentile HR = 1.374, 95% CI = 1.059–1.784), whereas HRs from the 20th through the 90th percentiles were near unity (0.924–1.105,

As *A. muciniphila* is known to be linked to dietary fiber intake, we further investigated the potential modifying effect of dietary fiber intake by stratifying the *A. muciniphila* model by dietary fiber intake quartiles without adjusting for whole grains. Interestingly, we observed a stronger significant effect of *A. muciniphila* on T2D risk in participants within the lowest quartile of dietary fiber intake, which corresponded to an intake of ≤ 20.7 g/day (FAS: Q1, HR = 1.098, 95% CI = 1.039–1.161, $p = 0.017$ vs. Q4, HR = 1.040, 95% CI = 0.988–1.096, $p = 0.205$; LTAS: Q1, HR = 1.111, 95% CI = 1.046–1.180, $p < 0.001$ vs. Q4, HR = 1.036, 95% CI = 0.980–1.095, $p = 0.217$) (Figure 3A). Among participants who developed T2D, we observed a slightly lower *A. muciniphila* abundance along higher dietary fiber intakes, but no significant differences were observed except in the LTAS, where lower abundance of *A. muciniphila* was observed at higher levels of dietary fiber intake (Q1 vs. Q4) (Figure 3B). Formal interaction tests did not indicate any statistically significant effect modification (FAS, $p_{\text{interaction}} = 0.302$; LTAS, $p_{\text{interaction}} = 0.321$), likely due to low statistical power given the low variability of *A. muciniphila* abundance across dietary fiber intake levels (Table S3). Among participants who developed T2D with C-reactive protein (CRP) data available ($n = 381$), we observed a significant interaction between *A. muciniphila* abundance and dietary fiber intake in relation to inflammation ($p = 0.019$). Opposite patterns across fiber intake levels were observed with higher *A. muciniphila* associated with

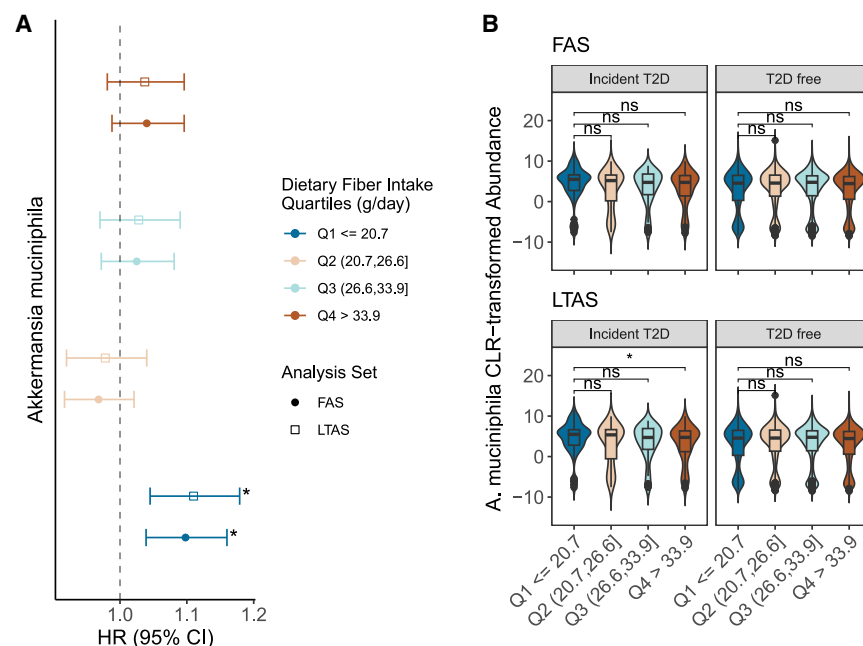


Figure 3. Association between *Akkermansia muciniphila*, dietary fiber intake, and incident T2D

(A) HRs of Cox regression models between the *A. muciniphila* and incident T2D using the full analysis set ([FAS], $n_{\text{total}} = 4,685$ with 383 T2D incident cases) and the lag time analysis set ([LTAS], $n_{\text{total}} = 4,633$ with 331 T2D incident cases) stratified by levels of dietary fiber intake. Quartiles of dietary fiber intake are expressed as g/day. The model is adjusted for sex, baseline age, level of education, height, waist circumference, smoking status, walking/cycling, physical exercise, coffee consumption, daily intakes of total energy, alcohol, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth. * p value < 0.05. (B) Boxplot indicating the distribution of *A. muciniphila* across the four quartiles of dietary fiber intake for both the FAS and LTAS. ns, non-significant, * p < 0.05 (Student's t test).

increased odds of elevated CRP in a low-fiber intake context, whereas in a high-fiber intake context, it was associated with lower odds of inflammation (Figure S6), suggesting a dietary fiber intake-dependent dual role of *A. muciniphila* in relation to inflammation.

Next, we used anpan's phylogenetic generalized linear mixed model to assess whether within-species phylogeny of *A. muciniphila* explained inter-individual heterogeneity in T2D risk. Strain-level analysis did not reveal evidence of lineage-specific associations with T2D risk. Individuals who developed T2D were distributed across the phylogenetic tree without clustering within specific clades (Figure S7), and no phylogenetic signal was detected ($\text{elpd_diff} = -0.9 \pm 0.3$). These findings suggest that the observed positive association between *A. muciniphila* abundance and incident T2D was not driven by specific phylogenetic lineages in our cohort.

Gut microbiome functional capability associations with incident T2D

Among the 103 gut metabolic modules (GMMs) investigated, 3 GMMs were consistently associated with incident T2D across both analysis sets (Figure 4A). Asparagine degradation showed the strongest risk association (MF0042, FAS: HR = 1.122, 95% CI = 1.053–1.196, $q = 0.014$; LTAS: HR = 1.121, 95% CI = 1.046–1.201, $q = 0.024$). Two GMMs showed protective associations: mannose degradation (MF0018; FAS: HR = 0.435, 95% CI = 0.292–0.648, $q = 0.004$; LTAS: HR = 0.415, 95% CI = 0.271–0.637, $q = 0.006$) and non-oxidative pentose phosphate pathway (PPP) (MF0071; FAS: HR = 0.245, 95% CI = 0.101–0.595, $q = 0.048$; LTAS: HR = 0.173, 95% CI = 0.066–0.454, $q = 0.019$). Restricted cubic splines indicated a tendency toward nonlinearity for MF0042 asparagine degradation and MF0018 mannose degradation without

reaching significance (p for nonlinearity < 0.1). For asparagine degradation, the HR increased steeply from low to mid abundance and appeared to plateau, while for mannose degradation, the HR declined with increasing abundance, with a more pronounced drop at higher levels, especially from the third quartile onward (Figure 4B). Overall, similar patterns were observed for the LTAS (Figure S8).

Sensitivity analysis

After additional adjustment for fasting plasma glucose, several microbial features remained significantly associated with incident T2D in the FAS and the LTAS. In the FAS, *A. communis*, *R. gnavus*, Erysipelotrichaceae bacterium, Clostridia unclassified SGB6317, and the GMMs MF0018 (mannose degradation) and MF0042 (asparagine degradation) remained significantly associated with T2D risk ($q < 0.05$). In the LTAS, *A. communis*, *D. piger*, *A. muciniphila*, Clostridia unclassified SGB6317, and the same two metabolic modules (MF0018 and MF0042) retained significant associations with T2D risk (Table S4). The competing risk analysis, accounting for death as a competing event, yielded similar results with effect sizes comparable to those from the primary Cox regression analyses (Table S5). We replicated species-level associations in the FAS using the CHAMP taxonomic profiler. Of the nine species initially identified, eight were again selected by Elastic Net as strong predictors of T2D risk. Effect estimates were consistent in direction and magnitude across profiling methods, including for *Merdibacter merdipullorum* corresponding to Erysipelotrichaceae bacterium in MetaPhlAn (not selected by Elastic Net). Associations remained statistically significant after multiple testing correction ($q < 0.05$) for all species except *Merdicola* sp001915925 (CHAMP; corresponding to Clostridia unclassified SGB6317 in MetaPhlAn; $q = 0.051$), CAG-194 sp000432915 (CHAMP; corresponding to GGB3614 SGB4886 in MetaPhlAn; $q = 0.056$), and *Merdibacter merdipullorum* ($q = 0.065$).

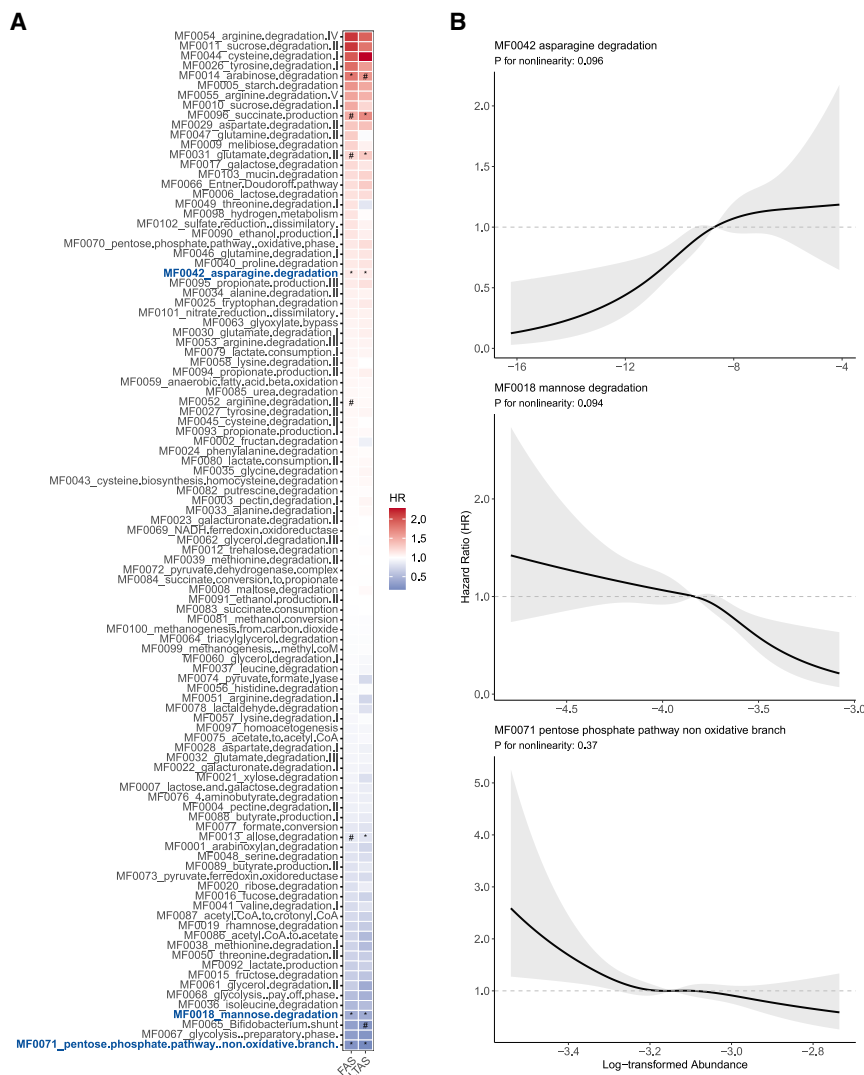


Figure 4. Gut metabolic modules associated with incident T2D

(A) Heatmap of HRs from Cox models between GMMs and incident T2D. Results are presented for both analysis sets (FAS [full analysis set], $n_{\text{total}} = 4,685$ with 383 T2D incident cases and LTAS [lag time analysis set], $n_{\text{total}} = 4,633$ with 331 T2D incident cases). All models were adjusted for sex, age at baseline, level of education, height, waist circumference, smoking status, walking/cycling, exercise, coffee consumption, daily intakes of alcohol, whole grains, yogurt, red/processed meat, sugary food/sweetened beverages and total energy intake, statin medication use, aliquoting plate, and sequencing depth. * q value < 0.05 , # $0.05 \leq q$ value < 0.1 . Robust GMMs associated with incident T2D in both analysis sets are colored in blue. (B) Restricted cubic spline (RCS) curves from Cox proportional hazards models showing adjusted hazard ratios (HRs) for incident T2D across log-transformed abundance of the three GMMs consistently associated with incident T2D. The solid line corresponds to the estimated HR, and the shaded band corresponds to the 95% CI. The horizontal dashed line indicates HR = 1, corresponding to the reference point set at the median abundance of each GMM. The p value for the nonlinear term (p -nonlinear) was estimated from a likelihood-ratio test comparing the RCS term with a linear term. Models are adjusted for the same set of covariates as the main models. Knots are placed at the 5th, 35th, 65th, and 95th percentiles. The RCS is shown using data from the FAS.

(Figure S9). Evidence of nonlinearity was observed for *C. catus* (p for nonlinearity = 0.029), consistent with the primary analysis (Figure S10).

Finally, complete case analyses indicated that our results were unaffected by confounders' imputation as overall, similar strength, and direction of associations were observed in complete case analyses (Table S6).

DISCUSSION

This prospective cohort study of 4,685 individuals identified six gut microbial species and one GMM associated with a higher risk of T2D, whereas three species and two GMMs were associated with a lower risk of T2D. Gut microbiome diversity was not associated with incident T2D in our study. We comprehensively adjusted our analyses for an extensive set of potential confounding factors spanning sociodemographic, lifestyle, and dietary factors and focused on gut microbiome features that were

our results extend previous evidence, although key discrepancies with earlier work warrant careful interpretation.

A. muciniphila, widely recognized as beneficial for cardiometabolic health,^{23–26} showed a counterintuitive, albeit modest, positive association with T2D risk in our study. This finding contrasts with most cross-sectional studies but aligns with two prospective studies: one Finnish cohort reporting neutral associations¹⁸ and a European prediabetic cohort showing that an increase of *A. muciniphila* was linked to a deterioration in metabolic health.²⁷ The discordance may reflect cohort-specific factors such as strain-level heterogeneity between study populations, since the above-mentioned studies were also conducted mostly in Nordic European countries, as well as the advanced age of our population and the potential effect modification of *A. muciniphila* abundance by low dietary fiber intake (≤ 20.7 g/day), which may contribute to context-dependent outcomes.

Although *A. muciniphila* exhibits substantial phylogenetic and functional heterogeneity across populations,²⁸ with

strain-specific differences in gut barrier integrity and metabolic regulation in experimental models,^{29,30} we found no evidence that strain-level variation contributed to the observed association with T2D risk in our cohort. This is consistent with prior multi-ethnic analyses showing that strain-specific associations with T2D are not uniformly observed across species; notably, some species associated with T2D at the species level do not exhibit lineage-specific effects, while strain-level signals may emerge in species without species-level associations.³¹ In the aforementioned study, most strain-level signals were heterogeneous across populations, with specific strain clusters associated with T2D enriched in particular geographic or ethnic groups, suggesting that such associations may depend on the broader microbial genetic diversity present at the population level. Our findings indicate that the relationship between *A. muciniphila* and T2D risk in our population was driven by overall species abundance rather than lineage-specific effects. Nevertheless, the potentially limited phylogenetic diversity captured within our cohort may have constrained the ability to detect strain-level signals, and this possibility cannot be excluded.

Furthermore, experimental evidence suggests that the role of *A. muciniphila* as a mucin-degrading species may also explain our findings. Murine studies have shown that dietary fiber deprivation leads to an increase in *A. muciniphila*, accompanied by greater mucous layer erosion and subsequent gut barrier dysfunction, resulting in increased susceptibility to pathogen infection.^{32,33} In our stratification analysis, the *A. muciniphila*-T2D risk association was strongest in the lowest fiber quartile (HR = 1.10 vs. 1.04 in the highest fiber quartile), supporting diet as a potential modifier of gut microbiome-T2D relationships, although formal tests for interaction did not reach statistical significance likely due to limited statistical power arising from low variability in *A. muciniphila* abundance across fiber strata in our cohort. However, among participants who developed T2D, we observed a significant interaction between *A. muciniphila* abundance and dietary fiber intake in relation to inflammation. Higher *A. muciniphila* was associated with higher odds of elevated CRP under low reported fiber intake but with lower odds under high reported fiber intake, suggesting that dietary fiber may modulate the inflammatory effects of *A. muciniphila* in line with previous evidence that dietary fiber deprivation influences its ecological niche and metabolic activity.³² In a low-fiber context, *A. muciniphila* compromise gut barrier integrity by thinning the mucin layer, thereby increasing intestinal permeability and promoting inflammation both locally and systemically. This is supported by human and animal studies linking microbial encroachment to chronic low-grade inflammation, insulin resistance, and dysglycemia, all hallmarks of T2D.^{34,35}

Moreover, during aging, the protective function of the mucus barrier is reduced partly because of the natural decline of mucus thickness.³⁶ In our elderly cohort, it is conceivable that the age-related impairment of the mucus may have contributed to the observed detrimental effect of *A. muciniphila* on T2D risk. Although evidence suggests a protective effect of *A. muciniphila* in the context of cardiometabolic diseases partially by maintaining intestinal homeostasis in the host, this protective effect may be context dependent. Under certain conditions, such as aging or low dietary fiber intake, increased *A. muciniphila* may exacerbate

intestinal epithelial damage, suggesting a context-dependent, double-edged role for this bacterium.^{37–40} These context-dependent effects likely arise from interactions between microbial, dietary, and host factors that shape its ecology and function. In addition to dietary fiber, modulators such as age-related physiological changes, inflammation, and dietary components such as polyphenols can influence *A. muciniphila* abundance and activity.^{28,41} Therefore, these factors may contribute to variability across studies and highlight the importance of considering host context when interpreting its links to metabolic health.

R. gnavus, characterized as an inflammatory species, has been consistently linked to metabolic disorders across study designs.^{18,42,43} In our study, *R. gnavus* replicated its prospective association with T2D risk, aligning with Ruuskanen et al.'s Finnish cohort findings,¹⁸ strengthening evidence for its role in T2D pathogenesis. In the context of T2D, it has been implicated in the production of tryptamine and phenethylamine, which contribute to gut dysbiosis-induced insulin resistance.⁴⁴ *R. gnavus* has been associated with elevated levels of imidazole propionate, a histidine-derived microbial metabolite that impairs glucose metabolism and is associated with systemic inflammation and T2D.⁴⁵ *R. gnavus* is also linked to the production of trimethylamine N-oxide (TMAO), a gut microbiome-dependent metabolite⁴⁶ notably associated with red meat intake⁴⁷ that is consistently linked to increased T2D risk.^{48,49}

D. piger has been previously associated, in cross-sectional studies, with several metabolic diseases including T2D,⁵⁰ metabolic dysfunction-associated steatotic liver disease (MASLD),⁵¹ and obesity.⁵² *D. piger* is a sulfate-reducing bacterium that produces hydrogen sulfide (H₂S) which promotes both toxic and pro-inflammatory effects for intestinal epithelial cells via activation of T helper 17 cells.⁵³ Interestingly, a recent murine model study showed that *Desulfovibrio*-derived H₂S compromises glucagon-like peptide-1 (GLP-1) production.⁵⁴ This hormone is responsible for a variety of glucoregulatory effects, including glucose-dependent secretion of insulin and inhibition of glucagon release, which are impaired in people with T2D.⁵⁵ Moreover, H₂S can further facilitate the degradation of mucin by acting on the disulfide bonds,⁵⁶ therefore, acting directly on the gut barrier integrity, which is linked to low-grade inflammation and metabolic dysregulation, as discussed above.

Two *Alistipes* species, *A. communis* (previously named *A. obes*)⁵⁷ and *A. finegoldii*, were robustly associated with T2D risk in our study. Although recent human and experimental research suggests glucose-lowering or metabolically beneficial effects for some *Alistipes* strains,⁵⁸ the genus shows mixed health associations,⁵⁹ and cross-sectional studies in westernized middle-aged cohorts report inconsistent links with T2D.^{60,61} However, Mendelian randomization⁶² and animal studies⁶³ support roles in insulin resistance and metabolic disruption, consistent with the pro-inflammatory properties described for *A. finegoldii*.⁵⁹

In our study, three species were inversely associated with incident T2D: an unclassified Erysipelotrichaceae bacterium species, *C. catus*, and an unclassified Clostridia species (SGB6317).

Members of the Erysipelotrichaceae family have been linked to metabolic disorders, with decreased abundance reported in individuals with metabolic syndrome,⁶⁴ although higher

resolution analyses down to the species-level were lacking. A metagenomic study demonstrated that fecal microbiome transplants from donors with different T2D severities (mild and severe T2D) induced corresponding metabolic phenotypes in mice. Among the key species, *Erysipelotrichaceae* bacterium I46, which was enriched in recipients of mild T2D gut microbiome, was negatively associated with glucose tolerance parameters, suggesting a potential protective role against metabolic disturbances in the context of T2D.⁶⁵ While species-level associations remain rare within this phylogenetic group, our findings raise the possibility that other related, yet uncharacterized, members may also contribute to metabolic health.

C. catus has been previously negatively associated with insulin resistance⁶⁶ in line with its role as a butyrate-producing bacterium, whose reduced abundance is a well-established feature of T2D.^{5,6,10} SCFAs, including butyrate, improve glucose homeostasis by stimulating GLP-1 and PYY secretion, which promote satiety and improve insulin sensitivity,⁵ and by supporting gut barrier integrity, as butyrate serves as the primary energy source for colonocytes.⁶⁷ Interestingly, a previously mentioned study²⁷ showed that loss of hydrogenase-encoding bacterial genes in *C. catus* was linked to lower plasma insulin levels.²⁷ Bacterial hydrogenases may play a critical role in butyrate production,⁶⁸ which may potentially enhance host insulin sensitivity. Additionally, *C. catus* may contribute to the production of 3,4-dihydroxyphenylacetic acid (DOPAC),⁶⁹ a phenolic compound shown to alleviate hyperglycemia and insulin resistance in T2D mouse models.⁷⁰ For this species, we observed a potential nonlinear relationship suggesting that this species abundance range may mediate metabolic outcomes. Two additional species associated with T2D risk, one positively (GGB3614_SGB4886, *Lachnospiraceae* family) and one negatively (*Clostridia* species SGB6317, *Clostridia* class), have not been previously implicated in metabolic disorders. These taxa belong to species-level genome bins (SGBs), clusters of metagenome-assembled genomes grouped based on sequence similarity, often lacking formal taxonomic annotation.⁷¹ While MetaPhlAn4 assigns all species to SGBs, some remain labeled by their SGB identifier due to a lack of known reference genomes or taxonomic names. Their associations with T2D risk suggest that they may play previously unrecognized roles in disease pathogenesis, highlighting both the strength of shotgun metagenomics in uncovering such associations and the ongoing challenge of interpreting unannotated microbial taxa.

The asparagine degradation GMM (asparagine to aspartate) showed the strongest microbial association with T2D risk. This aligns with previous cross-sectional⁷² and prospective studies^{73,74} reporting negative associations between circulating asparagine, insulin resistance, and T2D risk, respectively, suggesting that microbial asparagine degradation may contribute to hyperglycemia and subsequent T2D. However, evidence remains conflicting, as one study linked a higher asparagine-to-aspartate ratio to increased T2D risk⁷⁵ and its role in diabetes pathogenesis remains uncertain and warrants further investigation.

We also identified inverse associations of two functional modules with T2D risk: first, GMM MF0018, which is involved in mannose degradation (mannose-6-phosphate to fructose-6-

phosphate). Mannose is considered an insulin-regulated metabolite, reflecting systemic insulin sensitivity,⁷⁶ and elevated plasma mannose levels have been consistently linked to higher T2D and cardiovascular disease risk in prospective cohorts, independent of glycemic status.^{76–78} Previous studies have shown enrichment of mannose degradation genes in the gut microbiome following metformin treatment, supporting a potential protective role of microbial mannose degradation in T2D management.⁷⁹ Our findings strengthen the hypothesis that an increased microbial mannose degradation capacity may contribute to lower circulating mannose levels, improved insulin sensitivity, and reduced risk of T2D.

Second is GMM MF0071, which represents the non-oxidative branch of PPP. The inverse association of GMM MF0071 with T2D risk may be explained by its link to SCFA production. The non-oxidative PPP contributes to bacterial carbohydrate fermentation by generating intermediates, such as pyruvate, that fuel SCFA synthesis.^{80,81} SCFAs, particularly butyrate, are known to enhance glucose homeostasis, improve insulin sensitivity, and reduce inflammation through modulation of cytokines like TNF- α and IL-6.⁸² Although individual modules involved in butyrate production showed negative but non-significant associations, the increased microbial PPP capacity observed may either better capture total SCFA synthesis, or relate to other potentially unknown mechanisms, which could support metabolic health and lower T2D risk.

Our study has several strengths, including the leverage of shotgun metagenomics that provides high taxonomic resolution in one of the largest prospective population-based cohorts to date, employing a highly standardized protocol from fecal sample collection to DNA sequencing. Incident T2D was comprehensively defined using inpatient and outpatient diagnoses alongside prescription data, offering a more objective and accurate measure than self-reports. We sought to reduce any potential reverse causation by excluding participants who developed T2D within the first year. This 1-year window removed those most likely to have marked metabolic disturbances at baseline while preserving statistical power. Although this strategy may not eliminate all potential reverse causation, it helps identify the most robust and biologically meaningful microbial associations. After further adjustment for fasting glucose, some associations weakened, suggesting early metabolic dysregulation, whereas most remained robust ($q < 0.1$), indicating long-term microbial contributors to T2D risk independent of baseline glycemia. Furthermore, our results were unaffected by antibiotic use and antidiabetic treatments, including metformin, and were not confounded by statin medication. Additionally, for species-level analyses, we employed a robust Elastic Net resampling and cross-validation framework ensuring robust feature selection and reliable model performance evaluation. We validated the species-level associations using an independent taxonomic profiling pipeline, demonstrating robustness to bioinformatic methodology and strengthening confidence in the observed findings. Additionally, we had access to comprehensive information on potential confounders and adjusted our models for a wide range of demographic, lifestyle, and dietary factors, as well as sequencing-related technical variables.

In conclusion, our study advances gut microbiome research in T2D by identifying several gut microbial species associated with incident T2D, some of which appeared to be context and diet dependent, such as the suggested fiber-modified effect of *A. muciniphila*. By establishing temporal relationships, we extend prior cross-sectional evidence linking the gut microbiome to T2D and provide prospective validation of *R. gnavus* as a conserved microbial risk factor across populations. Moreover, we provide early evidence of putative metabolic connections between microbiome functional potential and subsequent T2D risk, years before clinical onset. If replicated, these findings could enhance risk stratification and inform personalized interventions that target diet-microbiome interactions.

Limitations of the study

Our study also has limitations that warrant consideration. Although we adjusted for several potential cofounders throughout the modeling process, residual confounding may remain. In addition, the cohort represents an older Swedish population (mean age, 73.9 years), which may limit generalizability. Nevertheless, several associations identified in this study have also been reported in cross-sectional studies across diverse age groups and ethnic backgrounds, supporting their broader generalizability. For example, while *R. gnavus* has been linked to metabolic disorders in Norwegian adults (age range, 20–94 years),⁴³ similar associations were observed in a Chinese cohort (mean age, 58 years) showing enrichment in individuals with insulin resistance and dyslipidemia.⁸³ Likewise, *D. piger* has been associated with T2D in a Nigerian cohort (mean age, 57 years),⁵⁰ and *C. catus*, inversely with insulin resistance in a Kazakh cohort (age range, 30–59 years).⁶⁶ However, only one microbial association (*R. gnavus*) was replicated against the only large-scale reported prospective analysis to date.¹⁸ Discrepancies may be attributed to differences in follow-up duration and age-driven microbial shifts, as well as other differences (e.g., ethnic, cultural and dietary). Furthermore, recent evidence suggests that variations in circadian rhythm influence microbial community structure, including taxa related to glucose metabolism and T2D risk.⁸⁴ Since stool samples in our study were collected at a single time point without standardized timing, some circadian-related variability may remain, although systematic bias is unlikely given the large, population-based design. In addition, while we propose diet-microbe interactions (e.g., fiber-*A. muciniphila*), causal mediation analyses were beyond our scope and mechanisms remain uncertain.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Rikard Landberg (rikard.landberg@chalmers.se).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The dataset supporting the conclusions of this article was provided by the national research infrastructure SIMPLER access board and cannot be shared publicly because of the sensitive nature of the data and the GDPR legislation.

Data are available from the national research infrastructure SIMPLER for researchers who meet the criteria for access to confidential data. Details of obtaining data from the national research infrastructure SIMPLER can be obtained at the website www.simpler4health.se. The analysis-specific programs are publicly available under a permanent DOI in Zenodo: <https://doi.org/10.5281/zenodo.19454951>. Any additional information required to re-analyze the data reported in this work paper is available from the [lead contact](#) upon request.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the project HealthFerm, which is co-funded by the European Union under the Horizon Europe grant agreement no. 101060247, and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract no. 22.00210. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the REA can be held responsible for them. The authors acknowledge the SIMPLER board for enabling the current study. The computations were performed on resources provided by the National Academic Infrastructure for Super-computing in Sweden (NAISS) support for sensitive data NAISS-SENS through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project simp2023012. The Swedish Research Council financially supports NAISS. We acknowledge support from the Swedish Research Council (grant 2022-00924), which covered part of R.L.'s salary.

AUTHOR CONTRIBUTIONS

G.T., C.W., and R.L. planned and designed the study. A.W. initiated the clinical subcohorts and designed feces collection. F.B. coordinated the microbiome profiling in the SIMPLER cohort. G.T. carried out statistical analyses. G.T. wrote the original draft of the manuscript with support from R.L. F.B., C.B., L.E., S.C.L., C.M.E., E.N., I.S.-K., A.W., C.W., and R.L. contributed with critical interpretation of the results for important intellectual content and reviewing and editing the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Assessment of type 2 diabetes and follow-up
 - Microbiome data
 - Covariates
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2026.102835>.

Received: September 4, 2025

Revised: March 20, 2026

Accepted: April 29, 2026

Published: May 27, 2026

REFERENCES

- GBD 2021 Diabetes Collaborators (2023). Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* 402, 203–234.
- Sharma, S., and Tripathi, P. (2019). Gut microbiome and type 2 diabetes: where we are and where to go? *J. Nutr. Biochem.* 63, 101–108.
- Gurung, M., Li, Z., You, H., Rodrigues, R., Jump, D.B., Morgun, A., and Shulzhenko, N. (2020). Role of gut microbiota in type 2 diabetes pathophysiology. *EBioMedicine* 51, 102590.
- Yang, G., Wei, J., Liu, P., Zhang, Q., Tian, Y., Hou, G., Meng, L., Xin, Y., and Jiang, X. (2021). Role of the gut microbiota in type 2 diabetes and related diseases. *Metabolism* 117, 154712.
- Byndloss, M., Devkota, S., Duca, F., Niess, J.H., Nieuwdorp, M., Orholm-Melander, M., Sanz, Y., Tremaroli, V., and Zhao, L. (2024). The Gut Microbiota and Diabetes: Research, Translation, and Clinical Applications—2023 Diabetes, *Diabetes Care*, and *Diabetologia* Expert Forum. *Diabetes* 73, 1391–1410.
- Barlow, G.M., and Mathur, R. (2022). Type 2 Diabetes and the Microbiome. *J. Endocr. Soc.* 7, bvac184.
- Cunningham, A.L., Stephens, J.W., and Harris, D.A. (2021). Gut microbiota influence in type 2 diabetes mellitus (T2DM). *Gut Pathog.* 13, 50.
- Wu, H., Esteve, E., Tremaroli, V., Khan, M.T., Caesar, R., Mannerås-Holm, L., Ståhlman, M., Olsson, L.M., Serino, M., Planas-Félix, M., et al. (2017). Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat. Med.* 23, 850–858.
- Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., Le Chatelier, E., Sunagawa, S., Prifti, E., Vieira-Silva, S., Gudmundsdóttir, V., Pedersen, H.K., et al. (2015). Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* 528, 262–266.
- Scheithauer, T.P.M., Rampanelli, E., Nieuwdorp, M., Vallance, B.A., Verchere, C.B., van Raalte, D.H., and Herrema, H. (2020). Gut Microbiota as a Trigger for Metabolic Inflammation in Obesity and Type 2 Diabetes. *Front. Immunol.* 11, 571731.
- Reynolds, A.N., Akerman, A.P., and Mann, J. (2020). Dietary fibre and whole grains in diabetes management: Systematic review and meta-analyses. *PLoS Med.* 17, e1003053.
- Salamone, D., Rivellese, A.A., and Vetrani, C. (2021). The relationship between gut microbiota, short-chain fatty acids and type 2 diabetes mellitus: the possible role of dietary fibre. *Acta Diabetol.* 58, 1131–1138.
- Makki, K., Deehan, E.C., Walter, J., and Bäckhed, F. (2018). The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* 23, 705–715.
- Zhang, F., Fan, D., Huang, J.L., and Zuo, T. (2022). The gut microbiome: linking dietary fiber to inflammatory diseases. *Med. Microecol.* 14, 100070.
- Wang, H., Gou, W., Su, C., Du, W., Zhang, J., Miao, Z., Xiao, C., Jiang, Z., Wang, Z., Fu, Y., et al. (2022). Association of gut microbiota with glycaemic traits and incident type 2 diabetes, and modulation by habitual diet: a population-based longitudinal cohort study in Chinese adults. *Diabetologia* 65, 1145–1156.
- Aasmets, O., Lüll, K., Lang, J.M., Pan, C., Kuusisto, J., Fischer, K., Laakso, M., Lusi, A.J., and Org, E. (2021). Machine Learning Reveals Time-Varying Microbial Predictors with Complex Effects on Glucose Regulation. *mSystems* 6, e01191-20.
- Vals-Delgado, C., Alcalá-Díaz, J.F., Molina-Abril, H., Roncero-Ramos, I., Caspers, M.P.M., Schuren, F.H.J., Van den Broek, T.J., Luque, R., Perez-Martinez, P., Katsiki, N., et al. (2022). An altered microbiota pattern precedes Type 2 diabetes mellitus development: From the CORDIOPREV study. *J. Adv. Res.* 35, 99–108.
- Ruuskanen, M.O., Erawijantari, P.P., Havulinna, A.S., Liu, Y., Méric, G., Tuomilehto, J., Inouye, M., Jousilahti, P., Salomaa, V., Jain, M., et al. (2022). Gut Microbiome Composition Is Predictive of Incident Type 2 Diabetes in a Population Cohort of 5,572 Finnish Adults. *Diabetes Care* 45, 811–818.
- Zhou, W., Sailani, M.R., Contrepois, K., Zhou, Y., Ahadi, S., Leopold, S.R., Zhang, M.J., Rao, V., Avina, M., Mishra, T., et al. (2019). Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* 569, 663–671.
- Schüssler-Florenza Rose, S.M., Contrepois, K., Moneghetti, K.J., Zhou, W., Mishra, T., Mataraso, S., Dagan-Rosenfeld, O., Ganz, A.B., Dunn, J., Hornburg, D., et al. (2019). A longitudinal big data approach for precision health. *Nat. Med.* 25, 792–804.
- Wu, H., Lv, B., Zhi, L., Shao, Y., Liu, X., Mitteregger, M., Chakaroun, R., Tremaroli, V., Hazen, S.L., Wang, R., et al. (2025). Microbiome–metabolome dynamics associated with impaired glucose control and responses to lifestyle changes. *Nat. Med.* 31, 2222–2231.
- Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.A., and Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* 50, D785–D794.
- Abbasi, A., Bazzaz, S., Da Cruz, A.G., Khorshidian, N., Saadat, Y.R., Sabahi, S., Ozma, M.A., Lahouty, M., Aslani, R., and Mortazavian, A.M. (2024). A Critical Review on Akkermansia muciniphila: Functional Mechanisms, Technological Challenges, and Safety Issues. *Probiotics Antimicrob. Proteins* 16, 1376–1398.
- Yan, J., Sheng, L., and Li, H. (2021). *Akkermansia muciniphila*: is it the Holy Grail for ameliorating metabolic diseases? *Gut Microbes* 13, 1984104.
- Zhao, Q., Yu, J., Hao, Y., Zhou, H., Hu, Y., Zhang, C., Zheng, H., Wang, X., Zeng, F., Hu, J., et al. (2023). *Akkermansia muciniphila* plays critical roles in host health. *Crit. Rev. Microbiol.* 49, 82–100.
- Cani, P.D., Depommier, C., Derrien, M., Everard, A., and De Vos, W.M. (2022). Akkermansia muciniphila: paradigm for next-generation beneficial microorganisms. *Nat. Rev. Gastroenterol. Hepatol.* 19, 625–637.
- Lyu, L., Fan, Y., Vogt, J.K., Clos-Garcia, M., Bonnefond, A., Pedersen, H.K., Dutta, A., Koivula, R., Sharma, S., Allin, K.H., et al. (2025). The dynamics of the gut microbiota in prediabetes during a four-year follow-up among European patients—an IMI-DIRECT prospective study. *Genome Med.* 17, 78.
- Ioannou, A., Berkhout, M.D., Geerlings, S.Y., and Belzer, C. (2025). Akkermansia muciniphila: biology, microbial ecology, host interactions and therapeutic potential. *Nat. Rev. Microbiol.* 23, 162–177.
- Liu, Q., Lu, W., Tian, F., Zhao, J., Zhang, H., Hong, K., and Yu, L. (2021). Akkermansia muciniphila Exerts Strain-Specific Effects on DSS-Induced Ulcerative Colitis in Mice. *Front. Cell. Infect. Microbiol.* 11, 698914.
- Kumar, R., Kane, H., Wang, Q., Hibberd, A., Jensen, H.M., Kim, H.S., Bak, S.Y., Auzanneau, I., Bry, S., Christensen, N., et al. (2022). Identification and Characterization of a Novel Species of Genus Akkermansia with Metabolic Health Effects in a Diet-Induced Obesity Mouse Model. *Cells* 11, 2084.
- Mei, Z., Wang, F., Bhosle, A., Dong, D., Mehta, R., Ghazi, A., Zhang, Y., Liu, Y., Rinott, E., Ma, S., et al. (2024). Strain-specific gut microbial signatures in type 2 diabetes identified in a cross-cohort analysis of 8,117 metagenomes. *Nat. Med.* 30, 2265–2276.
- Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., Pudlo, N.A., Kitamoto, S., Terrapon, N., Muller, A., et al. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* 167, 1339–1353.e21.
- Wolter, M., Grant, E.T., Boudaud, M., Pudlo, N.A., Pereira, G.V., Eaton, K.A., Martens, E.C., and Desai, M.S. (2024). Diet-driven differential response of Akkermansia muciniphila modulates pathogen susceptibility. *Mol. Syst. Biol.* 20, 596–625.
- Chassaing, B., Raja, S.M., Lewis, J.D., Srinivasan, S., and Gewirtz, A.T. (2017). Colonic Microbiota Encroachment Correlates With Dysglycemia in Humans. *Cell. Mol. Gastroenterol. Hepatol.* 4, 205–221.

35. Kordahi, M.C., Daniel, N., Gewirtz, A.T., and Chassaing, B. (2025). Mucus-penetrating microbiota drive chronic low-grade intestinal inflammation and metabolic dysregulation. *Gut Microbes* 17, 2455790.
36. Sovran, B., Hugenholtz, F., Elderman, M., Van Beek, A.A., Graversen, K., Huijskes, M., Boekschoten, M.V., Savelkoul, H.F.J., De Vos, P., Dekker, J., and Wells, J.M. (2019). Age-associated Impairment of the Mucus Barrier Function is Associated with Profound Changes in Microbiota and Immunity. *Sci. Rep.* 9, 1437.
37. Rajpurohit, Y.S. (2024). Dual-edged health benefit of *Akkermansia muciniphila*: impact on metformin and insulin resistance in type 2 diabetes – a perspective. *Curr. Top. Diabetes* 4, 118–127. <https://doi.org/10.5114/ctd/194184>.
38. Luo, Y., Lan, C., Li, H., Ouyang, Q., Kong, F., Wu, A., Ren, Z., Tian, G., Cai, J., Yu, B., et al. (2022). Rational consideration of *Akkermansia muciniphila* targeting intestinal health: advantages and challenges. *Npj Biofilms Microbiomes* 8, 81.
39. Qu, S., Zheng, Y., Huang, Y., Feng, Y., Xu, K., Zhang, W., Wang, Y., Nie, K., and Qin, M. (2023). Excessive consumption of mucin by over-colonized *Akkermansia muciniphila* promotes intestinal barrier damage during malignant intestinal environment. *Front. Microbiol.* 14, 1111911.
40. Grant, E.T., Monzel, E., and Desai, M.S. (2026). Navigating the duality of *Akkermansia muciniphila*. *Nat. Microbiol.* 11, 20–30.
41. Aja, E., Zeng, A., Gray, W., Connelley, K., Chaganti, A., and Jacobs, J.P. (2025). Health Effects and Therapeutic Potential of the Gut Microbe *Akkermansia muciniphila*. *Nutrients* 17, 562.
42. Crost, E.H., Coletto, E., Bell, A., and Juge, N. (2023). *Ruminococcus gnavus*: friend or foe for human health. *FEMS Microbiol. Rev.* 47, fuad014.
43. Grahnmø, L., Nethander, M., Coward, E., Gabrielsen, M.E., Sree, S., Billo, J.M., Engstrand, L., Abrahamsson, S., Langhammer, A., Hveem, K., and Ohlsson, C. (2022). Cross-sectional associations between the gut microbe *Ruminococcus gnavus* and features of the metabolic syndrome: the HUNT study. *Lancet Diabetes Endocrinol.* 10, 481–483.
44. Zhai, L., Xiao, H., Lin, C., Wong, H.L.X., Lam, Y.Y., Gong, M., Wu, G., Ning, Z., Huang, C., Zhang, Y., et al. (2023). Gut microbiota-derived tryptamine and phenethylamine impair insulin sensitivity in metabolic syndrome and irritable bowel syndrome. *Nat. Commun.* 14, 4986.
45. Molinaro, A., Bel Lassen, P., Henriksen, M., Wu, H., Adriouch, S., Belda, E., Chakaroun, R., Nielsen, T., Bergh, P.O., Rouault, C., et al. (2020). Imidazole propionate is increased in diabetes and associated with dietary patterns and altered microbial ecology. *Nat. Commun.* 11, 5881.
46. Coletto, E., Latousakis, D., Pontifex, M.G., Crost, E.H., Vaux, L., Perez Santamarina, E., Goldson, A., Brion, A., Hajhosseini, M.K., Vauzour, D., et al. (2022). The role of the mucin-glycan foraging *Ruminococcus gnavus* in the communication between the gut and the brain. *Gut Microbes* 14, 2073784.
47. Li, J., Li, Y., Ivey, K.L., Wang, D.D., Wilkinson, J.E., Franke, A., Lee, K.H., Chan, A., Huttenhower, C., Hu, F.B., et al. (2022). Interplay between diet and gut microbiome, and circulating concentrations of trimethylamine N-oxide: findings from a longitudinal cohort of US men. *Gut* 71, 724–733.
48. Li, C., Bishop, T.R.P., Imamura, F., Sharp, S.J., Pearce, M., Brage, S., Ong, K.K., Ahsan, H., Bes-Rastrollo, M., Beulens, J.W.J., et al. (2024). Meat consumption and incident type 2 diabetes: an individual-participant federated meta-analysis of 1.97 million adults with 100 000 incident cases from 31 cohorts in 20 countries. *Lancet Diabetes Endocrinol.* 12, 619–630.
49. Gu, X., Drouin-Chartier, J.P., Sacks, F.M., Hu, F.B., Rosner, B., and Willett, W.C. (2023). Red meat intake and risk of type 2 diabetes in a prospective cohort study of United States females and males. *Am. J. Clin. Nutr.* 118, 1153–1163.
50. Doumatey, A.P., Adeyemo, A., Zhou, J., Lei, L., Adebamowo, S.N., Adebamowo, C., and Rotimi, C.N. (2020). Gut Microbiome Profiles Are Associated With Type 2 Diabetes in Urban Africans. *Front. Cell. Infect. Microbiol.* 10, 63.
51. Lin, Y.-C., Lin, H.-F., Wu, C.-C., Chen, C.-L., and Ni, Y.-H. (2022). Pathogenic effects of *Desulfovibrio* in the gut on fatty liver in diet-induced obese mice and children with obesity. *J. Gastroenterol.* 57, 913–925.
52. Palmas, V., Pisanu, S., Madau, V., Casula, E., Deledda, A., Cusano, R., Uva, P., Vascellari, S., Loviselli, A., Manzin, A., and Velluzzi, F. (2021). Gut microbiota markers associated with obesity and overweight in Italian adults. *Sci. Rep.* 11, 5532.
53. Singh, S.B., Carroll-Portillo, A., and Lin, H.C. (2023). *Desulfovibrio* in the Gut: The Enemy within? *Microorganisms* 11, 1772.
54. Qi, Q., Zhang, H., Jin, Z., Wang, C., Xia, M., Chen, B., Lv, B., Peres Diaz, L., Li, X., Feng, R., et al. (2024). Hydrogen sulfide produced by the gut microbiota impairs host metabolism via reducing GLP-1 levels in male mice. *Nat. Metab.* 6, 1601–1615.
55. Brunton, S.A., and Wysham, C.H. (2020). GLP-1 receptor agonists in the treatment of type 2 diabetes: role and clinical experience to date. *Postgrad. Med. J.* 132, 3–14.
56. Ijssennagger, N., Belzer, C., Hooiveld, G.J., Dekker, J., van Mil, S.W.C., Müller, M., Kleerebezem, M., and van der Meer, R. (2015). Gut microbiota facilitates dietary heme-induced epithelial hyperproliferation by opening the mucus barrier in colon. *Proc. Natl. Acad. Sci.* 112, 10038–10043.
57. Sakamoto, M., Ikeyama, N., Ogata, Y., Suda, W., Iino, T., Hattori, M., and Ohkuma, M. (2020). *Alistipes communis* sp. nov., *Alistipes dispar* sp. nov. and *Alistipes onderdonkii* subsp. *vulgaris* subsp. nov., isolated from human faeces, and creation of *Alistipes onderdonkii* subsp. *onderdonkii* subsp. nov. *Int. J. Syst. Evol. Microbiol.* 70, 473–480.
58. Takeuchi, T., Kubota, T., Nakanishi, Y., Tsugawa, H., Suda, W., Kwon, A.T.J., Yazaki, J., Ikeda, K., Nemoto, S., Mochizuki, Y., et al. (2023). Gut microbial carbohydrate metabolism contributes to insulin resistance. *Nature* 621, 389–395.
59. Parker, B.J., Wearsch, P.A., Veloo, A.C.M., and Rodriguez-Palacios, A. (2020). The Genus *Alistipes*: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. *Front. Immunol.* 11, 906.
60. Wu, H., Tremaroli, V., Schmidt, C., Lundqvist, A., Olsson, L.M., Krämer, M., Gummesson, A., Perkins, R., Bergström, G., and Bäckhed, F. (2020). The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study. *Cell Metab.* 32, 379–390.e3.
61. Cui, J., Ramesh, G., Wu, M., Jensen, E.T., Crago, O., Bertoni, A.G., Gao, C., Hoffman, K.L., Sheridan, P.A., Wong, K.E., et al. (2022). Butyrate-Producing Bacteria and Insulin Homeostasis: The Microbiome and Insulin Longitudinal Evaluation Study (MILES). *Diabetes* 71, 2438–2446.
62. Zhang, H., Ma, L., Peng, W., Wang, B., and Sun, Y. (2024). Association between gut microbiota and onset of type 2 diabetes mellitus: a two-sample Mendelian randomization study. *Front. Cell. Infect. Microbiol.* 14, 1327032.
63. Zhou, Y., Tang, J., Du, W., Zhang, Y., and Ye, B.-C. (2024). Screening potential biomarkers associated with insulin resistance in high-fat diet-fed mice by integrating metagenomics and untargeted metabolomics. *Microbiol. Spectr.* 12, e0409423.
64. Chávez-Carbajal, A., Nirmalkar, K., Pérez-Lizaur, A., Hernández-Quiroz, F., Ramírez-Del-Alto, S., García-Mena, J., and Hernández-Guerrero, C. (2019). Gut Microbiota and Predicted Metabolic Pathways in a Sample of Mexican Women Affected by Obesity and Obesity Plus Metabolic Syndrome. *Int. J. Mol. Sci.* 20, 438.
65. Debédat, J., Le Roy, T., Volland, L., Belda, E., Allili, R., Adriouch, S., Bel Lassen, P., Kasahara, K., Hutchison, E., Genser, L., et al. (2022). The human gut microbiota contributes to type-2 diabetes non-resolution 5-years after Roux-en-Y gastric bypass. *Gut Microbes* 14, 2050635.
66. Abidinova, G.Z., Benberin, V.V., Vochshenkova, T.A., Mussin, N.M., Afshar, A., and Tamadon, A. (2025). Distinct gut microbial species, but not phylum-to-genus composition, associate with insulin resistance: a unique perspective from the Kazakh population. *Front. Microbiol.* 16, 1683885.

67. Nogal, A., Valdes, A.M., and Menni, C. (2021). The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. *Gut Microbes* 13, 1.
68. Campbell, A., Gdanetz, K., Schmidt, A.W., and Schmidt, T.M. (2023). H₂ generated by fermentation in the human gut microbiome influences metabolism and competitive fitness of gut butyrate producers. *Microbiome* 11, 133.
69. Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., et al. (2019). The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat. Microbiol.* 4, 623–632.
70. Liu, M., Wang, L., Huang, B., Lu, Q., and Liu, R. (2022). 3,4-Dihydroxyphenylacetic acid ameliorates gut barrier dysfunction via regulation of MAPK-MLCK pathway in type 2 diabetes mice. *Life Sci.* 305, 120742.
71. Manghi, P., Blanco-Míguez, A., Manara, S., NabiNejad, A., Cumbo, F., Beghini, F., Armanini, F., Golzato, D., Huang, K.D., Thomas, A.M., et al. (2023). MetaPhlAn 4 profiling of unknown species-level genome bins improves the characterization of diet-associated microbiome changes in mice. *Cell Rep.* 42, 112464.
72. Cheng, S., Rhee, E.P., Larson, M.G., Lewis, G.D., McCabe, E.L., Shen, D., Palma, M.J., Roberts, L.D., Dejam, A., Souza, A.L., et al. (2012). Metabolite Profiling Identifies Pathways Associated With Metabolic Risk in Humans. *Circulation* 125, 2222–2231.
73. Rebholz, C.M., Yu, B., Zheng, Z., Chang, P., Tin, A., Köttgen, A., Wagenknecht, L.E., Coresh, J., Boerwinkle, E., and Selvin, E. (2018). Serum metabolomic profile of incident diabetes. *Diabetologia* 61, 1046–1054.
74. Ottosson, F., Smith, E., Melander, O., and Fernandez, C. (2018). Altered Asparagine and Glutamate Homeostasis Precede Coronary Artery Disease and Type 2 Diabetes. *J. Clin. Endocrinol. Metab.* 103, 3060–3069.
75. Luo, H.-H., Feng, X.-F., Yang, X.-L., Hou, R.-Q., and Fang, Z.-Z. (2020). Interactive effects of asparagine and aspartate homeostasis with sex and age for the risk of type 2 diabetes risk. *Biol. Sex Differ.* 11, 58.
76. Ferrannini, E., Bokarewa, M., Brembeck, P., Baboota, R., Hedjazifar, S., Andersson, K., Baldi, S., Campi, B., Muscelli, E., Saba, A., et al. (2020). Mannose is an insulin-regulated metabolite reflecting whole-body insulin sensitivity in man. *Metabolism* 102, 153974.
77. Mardinoglu, A., Stančáková, A., Lotta, L.A., Kuusisto, J., Boren, J., Blüher, M., Wareham, N.J., Ferrannini, E., Groop, P.H., Laakso, M., et al. (2017). Plasma Mannose Levels Are Associated with Incident Type 2 Diabetes and Cardiovascular Disease. *Cell Metab.* 26, 281–283.
78. Fortin, E., Campi, B., Ferrannini, E., Mari, A., Mellbin, L.G., Norhammar, A., Näsman, P., Rydén, L., Saba, A., and Ferrannini, G. (2024). High Mannose Correlates With Surrogate Indexes of Insulin Resistance and Is Associated With an Increased Risk of Cardiovascular Events Independently of Glycemic Status and Traditional Risk Factors. *Diabetes Care* 47, 246–251.
79. Ezzamouri, B., Rosario, D., Bidkhor, G., Lee, S., Uhlen, M., and Shoaie, S. (2023). Metabolic modelling of the human gut microbiome in type 2 diabetes patients in response to metformin treatment. *npj Syst. Biol. Appl.* 9, 2.
80. Macfarlane, S., and Macfarlane, G.T. (2003). Regulation of short-chain fatty acid production. *Proc. Nutr. Soc.* 62, 67–72.
81. Deleu, S., Machiels, K., Raes, J., Verbeke, K., and Vermeire, S. (2021). Short chain fatty acids and its producing organisms: An overlooked therapy for IBD? *EBioMedicine* 66, 103293.
82. Zhu, T., and Goodarzi, M.O. (2020). Metabolites Linking the Gut Microbiome with Risk for Type 2 Diabetes. *Curr. Nutr. Rep.* 9, 83–93.
83. Xu, T., Wang, X., Chen, Y., Li, H., Zhao, L., Ding, X., and Zhang, C. (2023). Microbiome Features Differentiating Unsupervised-Stratification-Based Clusters of Patients with Abnormal Glycometabolism. *mBio* 14, e0348722.
84. Reitmeier, S., Kiessling, S., Clavel, T., List, M., Almeida, E.L., Ghosh, T.S., Neuhaus, K., Grallert, H., Linseisen, J., Skurk, T., et al. (2020). Arrhythmic Gut Microbiome Signatures Predict Risk of Type 2 Diabetes. *Cell Host Microbe* 28, 258–272.e6.
85. Boulund, F., and Arzoomand, A. (2022). ctmrbio/stag-mwc: StaG v0.5.1. Zenodo. <https://doi.org/10.5281/zenodo.7405309>.
86. Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* 27, 626–638.
87. Pita, S., Myers, P.N., Johansen, J., Russel, J., Nielsen, M.C., Eklund, A.C., and Nielsen, H.B. (2024). CHAMP delivers accurate taxonomic profiles of the prokaryotes, eukaryotes, and bacteriophages in the human microbiome. *Front. Microbiol.* 15, 1425489.
88. Darzi, Y., Falony, G., Vieira-Silva, S., and Raes, J. (2016). Towards biome-specific analysis of meta-omics data. *ISME J.* 10, 1025–1028.
89. Ghazi, A.R., Thompson, K.N., Bhosle, A., Mei, Z., Yan, Y., Wang, F., Wang, K., Franzosa, E.A., and Huttenhower, C. (2025). Quantifying Metagenomic Strain Associations from Microbiomes with Anpan. *bioRxiv*. <https://doi.org/10.1101/2025.01.06.631550>.
90. Harris, H., Håkansson, N., Olofsson, C., Julin, B., Åkesson, A., and Wolk, A. (2013). The Swedish mammography cohort and the cohort of Swedish men: Study design and characteristics of 2 population-based longitudinal cohorts. *OA Epidemiol.* 1.
91. Dekkers, K.F., Pertiwi, K., Baldanzi, G., Lundmark, P., Hammar, U., Moksnes, M.R., Coward, E., Nethander, M., Salih, G.A., Miari, M., et al. (2026). Genome-wide association analyses highlight the role of the intestinal molecular environment in human gut microbiota variation. *Nat. Genet.* 58, 540–549.
92. Vieira-Silva, S., Falony, G., Darzi, Y., Lima-Mendez, G., Garcia Yunta, R., Okuda, S., Vandeputte, D., Valles-Colomer, M., Hildebrand, F., Chaffron, S., and Raes, J. (2016). Species–function relationships shape ecological properties of the human gut microbiome. *Nat. Microbiol.* 1, 16088.
93. Larsson, S.C., Bergkvist, L., and Wolk, A. (2009). Long-term dietary calcium intake and breast cancer risk in a prospective cohort of women. *Am. J. Clin. Nutr.* 89, 277–282.
94. Andreu-Sánchez, S., Blanco-Míguez, A., Wang, D., Golzato, D., Manghi, P., Heidrich, V., Fackelmann, G., Zhernakova, D.V., Kurilshikov, A., Valles-Colomer, M., et al. (2025). Global genetic diversity of human gut microbiome species is related to geographic location and host health. *Cell* 188, 3942–3959.e9.
95. Austin, P.C., Lee, D.S., and Fine, J.P. (2016). Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* 133, 601–609.
96. Berry, S.D., Ngo, L., Samelson, E.J., and Kiel, D.P. (2010). Competing Risk of Death: An Important Consideration in Studies of Older Adults. *J. Am. Geriatr. Soc.* 58, 783–787.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human fecal sample	This Paper	N/A
Deposited data		
The analysis-specific code of this study	This paper	https://doi.org/10.5281/zenodo.19454951
Software and algorithms		
StaG-mwc (v0.5.1)	Boulund et al. ⁸⁵	https://stag-mwc.readthedocs.io/en/latest/index.html
StrainPhlAn (v4.0.6)	Truong et al. ⁸⁶	http://segatalab.cibio.unitn.it/tools/strainphlan/
CHAMP	Pita et al. ⁸⁷	https://cmbio.io/omics-sequencing-services/microbiome-profiling
R software (v4.3.1)	R Foundation	https://www.r-project.org
Omixer-RPM (v0.3.2)	Darzi et al. ⁸⁸	https://github.com/raeslab/omixer-rpm?tab=readme-ov-file
vegan (v2.6–4)	CRAN	https://cran.r-project.org/web/packages/vegan/index.html
glmnet (v4.1-8)	CRAN	https://cran.r-project.org/web/packages/glmnet/index.html
anpan (v0.3.0)	Ghazi et al. ⁸⁹	https://github.com/biobakery/anpan
missForest (v1.5.0)	CRAN	https://cran.r-project.org/web/packages/missForest/index.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The study used data from the **Swedish Infrastructure for Medical Population-Based Life-Course and Environmental Research (SIMPLER; www.simpler4health.se)** which consists of two prospective cohort studies: the Swedish Mammography Cohort (SMC) and the Cohort of Swedish Men (COSM). SMC was established in 1987–1990 and included women born between 1914 and 1948 ($n = 90,303$) and residing in the Swedish counties Uppsala or Västmanland at time of study enrollment. COSM was established in 1997, including men born between 1918 and 1952 ($n = 100,303$) residing in the counties Örebro or Västmanland at time of study enrollment.⁹⁰ Data on diet, lifestyle, and health status from all SIMPLER participants were collected with questionnaires in 10-year intervals. In addition, data from SIMPLER participants can be linked to individual's information from the Swedish population registries through unique personal identity numbers, providing almost complete continuously updated information on disease diagnoses and medication prescriptions for the entire study population.

A random selection of participants from each cohort were invited to take part in an extended collection of health-related information, referred to as the clinical subcohorts. For a subset of these participants, biological samples, including fecal samples, were also collected. Data collection was conducted via clinical visits in several subgroups according to the regions of residency of the cohort participants. The SMC clinical subcohort (SMCC) consists of two subgroups: SMCC Uppsala (SMCC-U, $n = 1,539$), conducted between 2012 and 2021; and SMCC Västmanland (SMCC-V, $n = 2,630$), conducted between 2010 and 2019. All COSM clinical subcohort were conducted in Västmanland from 2010 to 2019 (COSMC-V, $n = 4,750$). In total, 8,919 invited participants (60%) were included in the clinical subcohorts comprising fecal sample collection. At clinic study centers, all subcohort participants underwent extensive standardized physical examination, physiological and cognitive tests, and information on dietary habits, health status, and other health-related lifestyle information was collected with validated questionnaires.

Metagenomic analyses were conducted before sample collection was completed. At the time of analysis, of the 8,919 participants in the clinical subcohorts, 7,249 had donated fecal samples (SMCC-U, $n = 1,464$; SMCC-V, $n = 2,245$; COSMC-V, $n = 3,540$) and gut microbiome metagenomic data were available for 6,150 of them. In the present study, we excluded participants meeting one or more exclusion criteria: colorectal cancer or inflammatory bowel disease ($n = 270$), antibiotic use in the past 6 months before the clinical visit ($n = 704$, defined as a dispensed prescription, Anatomical Therapeutic Chemical code J01), and prevalent diabetes (including

those identified based on T2D medication; $n = 491$). After exclusions, 4,685 participants were included in the current study.

The current study was approved by the Regional Ethical Review Board in Stockholm, Sweden under the registration number Dnr 2023-02665-02 and amendment number Dnr 2024-00303-01. All participants provided written informed consent prior to participation.

METHOD DETAILS

Assessment of type 2 diabetes and follow-up

For all study participants, complete information on the diagnosis of T2D before or during the study period was obtained by linkage with the Swedish National Patient Register (NPR), which contains inpatient and outpatient specific diagnostic data from the Swedish healthcare system. T2D was defined according to ICD-10 code E11, ICD-9 code 250, and ICD-8 code 250. We further identified prevalent and incident cases using the Swedish Prescribed Drug Register targeting prescriptions of drugs used in diabetes (ATC A10) which includes metformin, reported to have an effect on the gut microbiome.^{8,9} All participants with a T2D identification date before the individual fecal sample collection date were defined as prevalent T2D case in this study, and all participants with a T2D identification date between fecal sample collection and study censoring date were defined as incident T2D cases. For prevalent cases only, register data were supplemented with measured fasting plasma glucose concentrations of ≥ 7.0 mmol/L according to WHO guidelines.

Dates of death were obtained from the Swedish Death Register. In prospective analysis with T2D incidence as outcome, for participant under risk (free of T2D at baseline), follow-up time accrued from fecal sampling date until the date of T2D diagnosis, death, or the censoring date of the study (31 December 2021), whichever came first.

Microbiome data

Participants received a pre-packaged fecal sample collection kit including instructions on how to collect the sample at home. The participants were asked to store the samples at -20°C in the home freezer at least 24h before the clinical visit. Once received at the clinical visit, samples were aliquoted with DNA/RNA shield (Zymo Research) and stored at -80°C until shipment on dry ice to the Center for Translational Microbiome Research (CTMR) at the Karolinska Institutet (Stockholm, Sweden) for DNA extraction.

DNA extraction was performed using MagPure Stool DNA LQ kit from Magen Biotechnology Co, Ltd. Briefly, 800 μL of the fecal samples were transferred to cryo tubes. A negative control (800 μL DNA/RNA shield) and a positive control (75 μL Zymo mock) were included. The samples were centrifuged at 14 000 g for 25 min and the supernatant was removed. To each tube, 600 μL ATP/PVP, 600 μL PCI and MagPure bead were added. The samples were bead beaten in a FastPrep 96 at 1600 rpm for 1 min. The samples were incubated at 65°C for 20 min and thereafter centrifuged for 3 min at 14 000 $\times$ g. Then, 340 μL of the upper phase of each sample was transferred to a deep-well plate and placed in an SP960 (MGI, Shenzhen). The following reagent plates were prepared and placed in an SP960, RNaseA 10 μL /well (15 mg/mL), Reagent mix 640 μL /well (MagPure Particles N 30 μL , Proteinase K (20 mg/mL) 20 μL and Buffer MLE 590 μL), GW1 (650 μL /well), 75% Ethanol (1.1 mL/well) and EB buffer (100 μL /well). The DNA was stored at -20°C . The DNA was then shipped on dry ice to MGI facilities in Riga, Latvia (Latvia MGI Tech).

The genomic DNA was subjected to library preparation using MGI's FS library prep set as per the manufacturer's protocol with 50 ng of input DNA. Quality assessment of the prepared libraries was performed using the TapeStation D1000 kit from Agilent, USA, while the quantity was determined by QuantIT HighSensitivity dsDNA Assay on a Tecan Spark (Tecan, Switzerland). The equimolarly pooled libraries were circularized using MGI Easy Circularization kit (MGI Tech) and subjected to DNBseq $2 \times 100\text{bp}$ paired-end sequencing on the DNBSEQ G400 or T7 sequencing instrument (MGI) following the manufacturer's instructions. The sequencing generated a median read count of 51.8 million (IQR = 34.8–61.2 million) read-pairs.

Sequencing data from the samples were processed with StaG-mwc, a reproducible Snakemake workflow for metagenomic analysis, version 0.5.1.⁸⁵ Tools used in the workflow included fastp v0.23.0 for quality control and filtering, Kraken2 v2.1.2 executed with “–confidence 0.1” and “–quick” for host removal using human genome version GRCh38, and MetaPhlAn v4.0.3 was used for taxonomic profiling, yielding a species-level relative abundance table comprising 2,470 bacterial species. All tools were run with their default settings.

Strain-level profiling of *Akkermansia muciniphila* (SGB9226), was performed using StrainPhlAn v4.0.6 and the mpa_vJan21_CHO-COPhAnSGB_202103 database. Among 171 clade-specific marker genes, 169 passed filtering and were retained. Marker sequences were trimmed by 50 bp at both ends, and samples and markers were filtered by requiring $\geq 80\%$ marker presence across samples and $\geq 80\%$ marker coverage per sample based on StrainPhlAn default parameters. After filtering, 3,156 of the 6,150 initial metagenomic samples were retained for phylogenetic analysis, of which 2,425 were included in the present study. StrainPhlAn uses the marker genes defined in the database to identify polymorphisms per sample, and alignments were subsequently used to infer a maximum-likelihood phylogenetic tree using the PhyloPhlAn pipeline in fast mode with RAXML.

The SIMPLER raw sequencing data were also processed at Cmbio using the CHAMP profiler⁸⁷ as previously described elsewhere,⁹¹ with taxonomic annotation based on the Genome Taxonomy Database (GTDB) release 214. Species with matching taxonomic labels were directly aligned, while SGBs without species names were mapped to GTDB taxonomy using the bioBakery SGB2GTDB translation file, enabling consistent species-level annotation across methods. The functional potential of the gut microbiome was characterized using gut metabolic modules (GMMs). Catalog genes were mapped to the curated GMMs v1.07 database⁹²

using EggNOG-mapper-software v2.0.1. To calculate the modules abundances, Omixer-RPM v0.3.2 was used with a minimum module coverage of 66.6%.

Covariates

Covariate information was obtained from either the questionnaire at the clinical examination visit, clinical measurements or register data. Assessment of anthropometric measurements was performed by the nurse during the clinic visit and included height (cm) and waist circumference (cm). Sociodemographic and lifestyle information retrieved from the questionnaire at the clinical visit included age at baseline (years), level of education, smoking status, walking/cycling and exercise during the past month. Dietary assessment was also retrieved from the questionnaire at the clinical visit and included consumption of wholegrains (g/day), yogurt (g/day), red/processed meat (g/day), sugary food/sweetened beverages (g/day), coffee (cups/day), alcohol (g/day), and total daily energy intake (kcal/day). Total energy intake was calculated by multiplying the age-specific portions by consumption frequency of each food item with the nutrient content obtained from the Swedish Food Database.⁹³ Daily intake of alcohol, wholegrains, yogurt, red/processed meat, and sugary food and sweetened beverages were estimated from a validated food frequency questionnaire included in the clinical examination visit questionnaire. Wholegrains, yogurt, red/processed meat, and sugary food and sweetened beverages intakes were calculated using different items in the FFQ (Table S1). Information on blood lipid modifying drug prescriptions (statin medication) was retrieved from the Swedish Prescribed Drug Register using Anatomical Therapeutic Chemical (ATC) code C10.

QUANTIFICATION AND STATISTICAL ANALYSIS

The microbiome α -diversity was estimated using the Shannon diversity index, and richness (number of different species). β -Diversity was estimated using Aitchison distance (vegan v2.6–4). Principal Component Analysis (PCA) was performed and the PCs explaining more than 1% of total variation were retained for downstream analyses leading to the inclusion of the first 6 PCs explaining together 18.9% of the total variation in the gut microbiome (Figure S1). Diversity metrics were calculated on the whole dataset without prior filtering. For downstream analyses at species level, we filtered out species that were not detected at a minimum relative abundance of 0.1% in at least 10% of samples yielding 193 species that met the criteria. Species relative abundances were then Centered Log-Ratio (CLR) transformed. For the functional potential of the gut microbiome, GMM abundances were log transformed prior to analysis.

The associations between gut microbiome features and the risk of developing T2D were estimated as hazard ratios (HRs) with 95% confidence intervals (CIs) using multivariable-adjusted Cox proportional hazards regression models with age as the underlying time-scale. We constructed models for alpha diversity (species richness and Shannon index), beta diversity (6 first PC axes), species CLR transformed abundances and GMMs log transformed abundances. Models were all adjusted for the covariables mentioned above and two metagenome sequencing-related technical variables, i.e., aliquoting plate and sequencing depth.

To balance the control of false positive findings while maintaining sensitivity for detecting biologically relevant species associations, we employed a two-stage approach: first using penalized Cox regression approach with Elastic net regularization and cross-validation to identify a subset of species with collective predictive value for T2D risk, followed by individual hypothesis testing on this reduced set with appropriate multiple testing correction. To ensure robustness, we performed 10 iterations of stratified re-sampling (70% training, 30% testing), preserving the initial case:control ratio in each split. For each resample, penalized Cox regression models were fitted using the Elastic Net framework with a 10-fold cross-validation to optimize the combination of alpha (mixing parameter) and lambda (regularization parameter) (glmnet v4.1–8). Model performance was evaluated on the test set using the Harrell's C-index, and species selected across resamples were summarized with their selection frequency. Species selected in at least 6 out of the 10 resamplings were considered as strong predictors of T2D risk and used for downstream analyses.

The study outcome T2D was primarily identified through register data, which raises the possibility that some participants may have had subclinical dysglycemia or undiagnosed T2D at the time of fecal sampling which may have influenced their gut microbiome.⁶⁰ Hence, we replicated all analyses including for the species-level association, the feature selection and the Cox regression models as lag time analysis, excluding 52 participants who developed T2D within the first year of follow-up. Therefore, our analyses were performed in two analysis sets in the current study: the “Full Analysis Set” (FAS, $n_{\text{total}} = 4,685$ with 383 T2D incident cases) and the “Lag Time Analysis Set” (LTAS, $n_{\text{total}} = 4,633$ with 331 T2D incident cases) and only gut microbiome features consistently associated with incident T2D across the two analysis sets were considered as robust gut microbiome features associated with T2D risk. For these robust features, we further modeled the associations with restricted cubic splines curves (RCS) to examine whether there was a non-linear relationship between the feature and T2D risk. For all RCS terms, we pre-specified four knots placed at the 5th, 35th, 65th, and 95th percentiles of the feature's distribution. We evaluated departures from linearity using a likelihood-ratio test (LRT) comparing the spline specification with a model containing only the linear term for the feature.

To test for statistical interaction between dietary fiber intake and *A. muciniphila* in relation to T2D risk, we compared Cox regression models with the same covariate adjustment as the main models except for wholegrain intake with and without the interaction term, using LRTs. The LRT p -value was used to assess whether inclusion of the interaction term significantly improved model fit, indicating evidence of interaction. To investigate strain-level variation in *A. muciniphila* associated with T2D, we applied phylogenetic generalized linear mixed models (PGLMM) using anpan⁸⁹ to identify a phylogenetic signal related to specific health outcomes as previously reported.^{31,94} In this framework, a phylogenetic term is incorporated as a sample-specific random effect whose covariance structure

is derived from the phylogenetic tree, allowing closely related strains to have correlated effects. PGLMMs are fitted in a Bayesian framework using either binomial or Gaussian likelihoods depending on the outcome variable. Here, we modeled incident T2D as a binary variable (incident T2D vs. non-T2D during follow-up), and the phylogeny as the independent variable with sex, baseline age, waist circumference, dietary fiber intake, and statin medication use included as covariates. Two models were fitted: one including the within-species phylogenetic random effect and a null model excluding the phylogenetic term but retaining all other covariates. Model performance was evaluated using leave-one-out cross-validation with Pareto-smoothed importance sampling to estimate the expected log pointwise predictive density (ELPD). The difference in ELPD between the phylogenetic and null models (elpd_diff), together with its standard error (se), was used to assess whether incorporating phylogenetic structure improved prediction of the T2D outcome. A significant phylogenetic signal was defined as an absolute difference in ELPD ($|\text{elpd_diff}| \geq 4$) with confidence interval of the elpd_diff not overlapping zero ($\text{elpd_diff} \pm 2 \text{ SE}$), indicating improved predictive performance with inclusion of phylogeny as previously reported.^{31,94}

For sensitivity analyses, we further adjusted the Cox regression models for fasting plasma glucose to assess whether the associations between the robust features and incident T2D were influenced by baseline glycemic status. Furthermore, Cox regression models are not designed to account for the competing risk of death, and thus, they can overestimate risk of disease in elderly individuals with high mortality rate.^{95,96} Given the relatively advanced age of the study population and that a substantial part of the participants died during the follow-up ($n = 387$), we additionally performed a competing risk analysis using Fine–Gray subdistribution hazard models, treating death as a competing event and reporting subdistribution hazard ratio (SHRs). We further replicated the species-level associations using the CHAMP profiled data. Finally, the overall proportion of missing data in the cohort was 2.3%, and missing values for covariates were imputed using a random forest-based algorithm, missForest (v1.5.0). To assess the potential impact of this imputation, we conducted a complete case analysis including only participants without missing data for covariates.

All p -values from high dimensional analyses were adjusted for multiple comparisons using the Benjamini–Hochberg procedure with a target rate of 0.05 for q -values. All statistical analyses were conducted using the R software version 4.3.1 (R Foundation).

Supplemental information

**Gut microbiome composition and functional
potential associate with incident type 2 diabetes
in 4,685 adults from a Swedish prospective cohort**

Gaël Toubon, Fredrik Boulund, Cecilia Martinez Escobedo, Carl Brunius, Lars Engstrand, Susanna C. Larsson, Elise Nordin, Ina Schuppe-Koistinen, Alicja Wolk, Clemens Wittenbecher, and Rikard Landberg

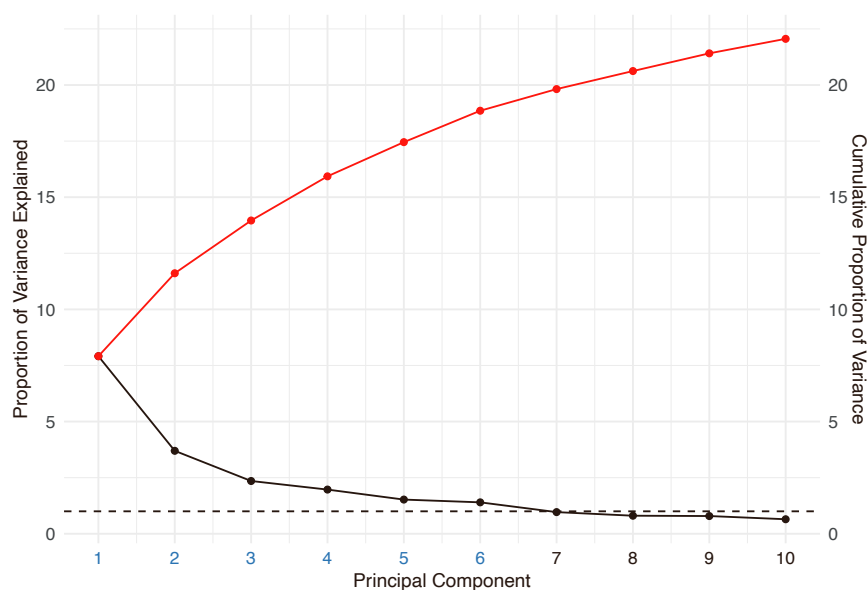


Figure S1. Percentage of variance explained across the 10 first Principal Components (PCs) after PCA analysis based on Aitchison distance using the Full Analysis set. Related to STAR Methods. PCs marked in blue explained each more than 1% in total variance. The black line corresponds to the proportion of Variance explained by each PCs while the red line corresponds to the cumulative variance explained across the 10 first PCs. The dashed line corresponds to the 1% variance explained threshold.

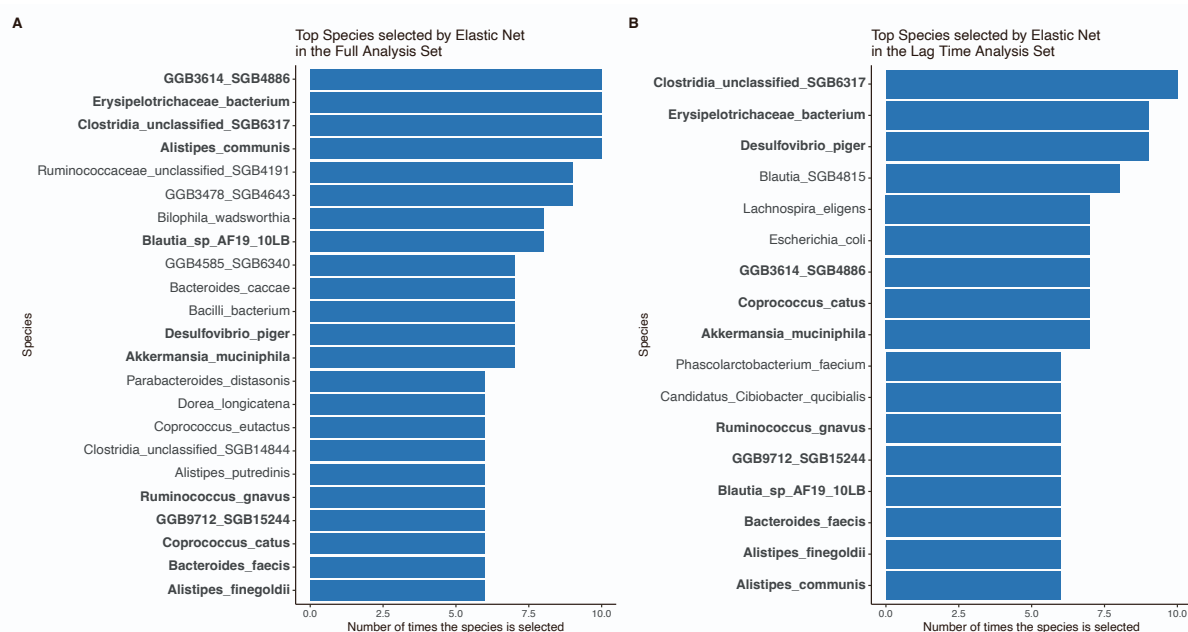


Figure S2. Species selected at least 6 times across 10 random resampling 10-fold cross validation Elastic Net. Related to Figure 1. Overlapping selected species between (A) the Full Analysis Set ($n_{\text{total}}=4,685$ with 383 T2D incident cases) and (B) the Lag Time Analysis Set ($n_{\text{total}}=4,633$ with 331 T2D incident cases) are marked in bold.

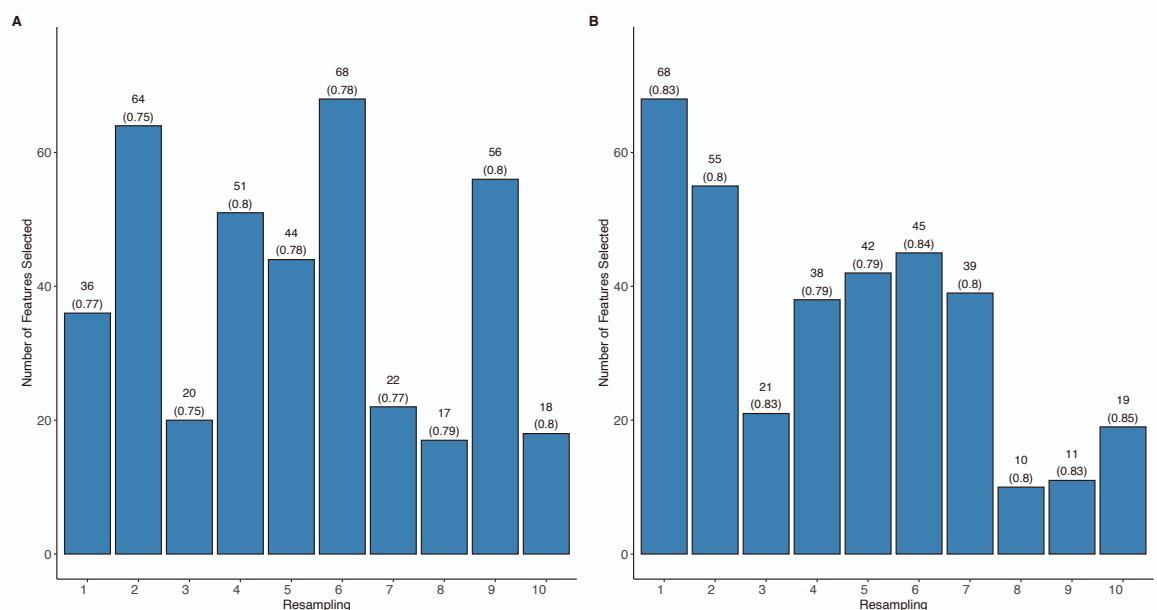


Figure S3. Number of species selected across each resample in (A) the Full Analysis Set and (B) the Lag Time Analysis Set. Related to Figure 1. Number on top of the bars corresponds to the number of species selected while number in parentheses corresponds to the C-index evaluated on the test set for each resample.

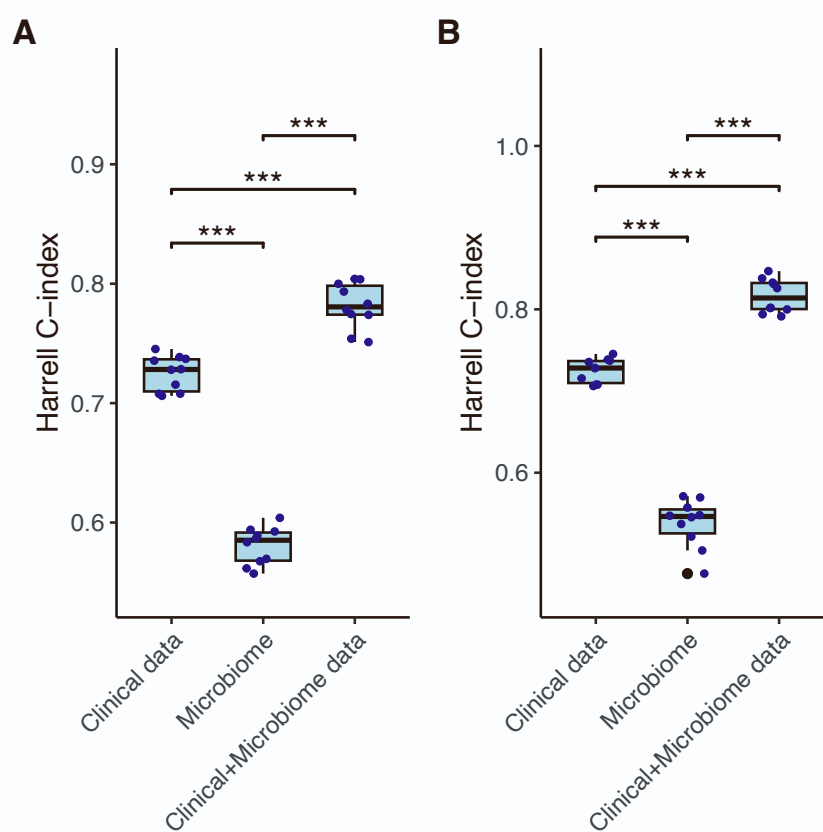


Figure S4. Feature selection model performance based on Harell c-index across the 10 resamples from penalized Cox regression with Elastic Net regularization models using either clinical data (covariates used for model adjustment, see covariates paragraph in the methods section) alone, microbiome data alone (species CLR transformed abundances) or a combination of clinical and microbiome data in (A) the Full Analysis Set and (B) the Lag Time Analysis Set. Related to Figure 1. For each resample, model performance was evaluated on the test set. *** $p < 0.001$

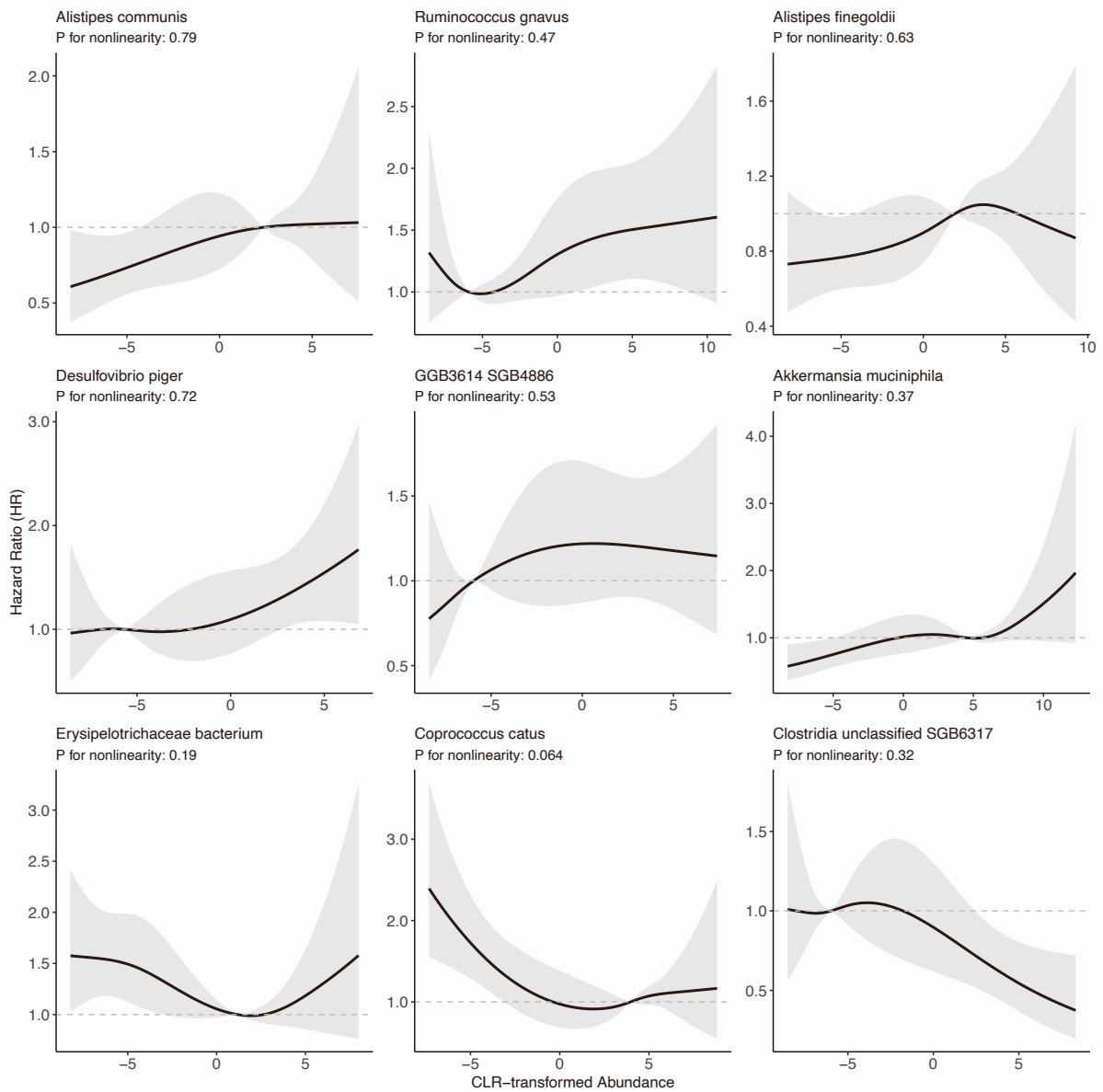


Figure S5. Restricted cubic spline (RCS) curves from Cox proportional hazards models from the LTAS showing adjusted hazard ratios (HR) for incident T2D across centered log-ratio (CLR) abundance of the 9 robust species identified in both analysis sets. Related to Figure 2. The solid line corresponds to the estimated HR and the shaded band the 95% CI. The horizontal dashed line indicates HR = 1, corresponding to the reference point set at the median abundance of each species. The p-value for the nonlinear term (P-nonlinear) was estimated from a likelihood-ratio test comparing the RCS term with a linear term. Models are adjusted for statin medication use, sex, baseline age, education, height, waist circumference, smoking, total energy intake, coffee, alcohol, walking/biking, exercise, whole grains, yogurt, red/processed meat, and sweet foods/beverages, aliquoting plate, and sequencing depth. Knots are placed at the 5th, 35th, 65th, and 95th percentiles. The RCS are shown using data from the Lag Time Analysis Set ($n_{\text{total}}=4,633$ with 331 T2D incident cases).

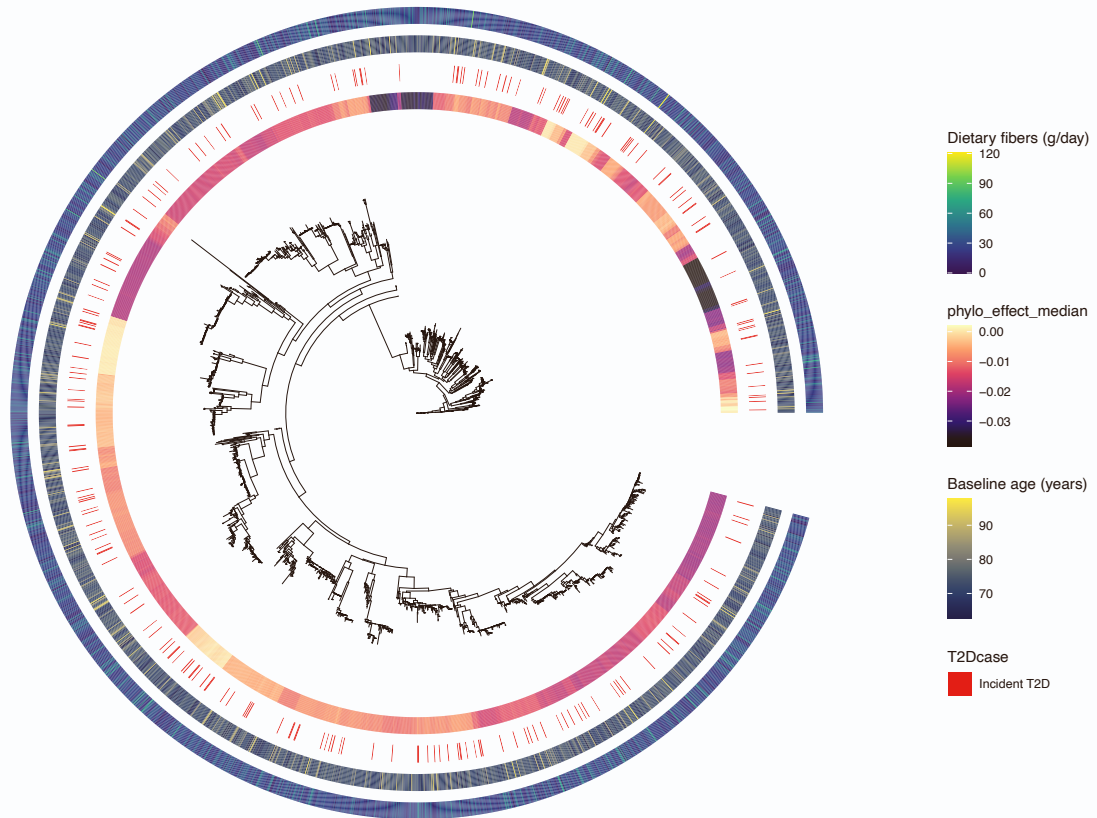


Figure S7: Phylogenetic tree of *Akkermansia muciniphila* in relation to T2D outcome. Related to Figures 1 and 3. Gut metagenomic samples from 2,425 individuals, of which, 213 developed T2D during the follow-up (red annotation) showing no strong phylogenetic signal for the T2D outcome. The phylogenetic effect median (phylo_effect_median) indicating the strength of the phylogenetic signal on T2D prediction per sample adjusted for sex, baseline age, waist circumference, dietary fibers, and statin medication. Only, baseline age, and dietary fibers are annotated in the outer rings.

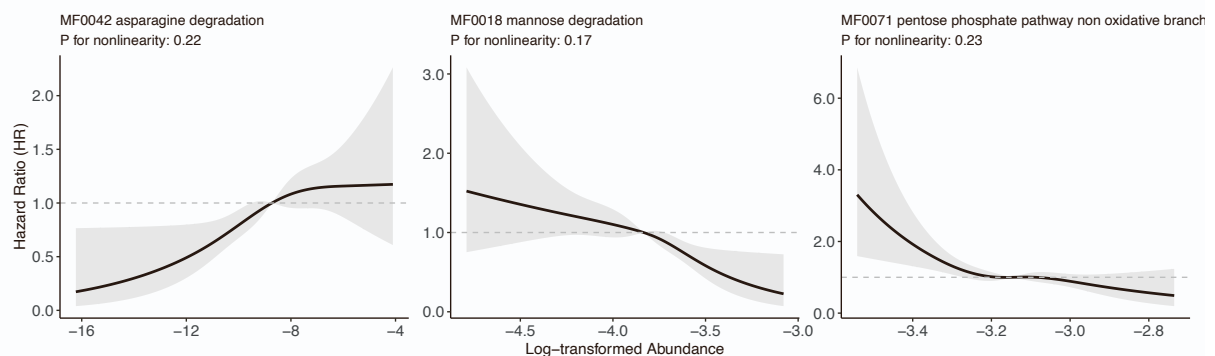


Figure S8. Restricted cubic spline (RCS) curves from Cox proportional hazards models from the LTAS showing adjusted hazard ratios (HR) for incident T2D across log-transformed abundance of the three GMMs consistently associated with incident T2D. Related to Figure 4. The solid line corresponds to the estimated HR and the shaded band the 95% CI. The horizontal dashed line indicates HR = 1, corresponding to the reference point set at the median abundance of each GMM. The p-value for the nonlinear term (P-nonlinear) was estimated from a likelihood-ratio test comparing the RCS term with a linear term. Models are adjusted for statin medication use, sex, baseline age, education, height, waist circumference, smoking, total energy intake, coffee, alcohol, walking/biking, exercise, whole grains, yogurt, red/processed meat, and sweet foods/beverages, aliquoting plate, and sequencing depth. Knots are placed at the 5th, 35th, 65th, and 95th percentiles.

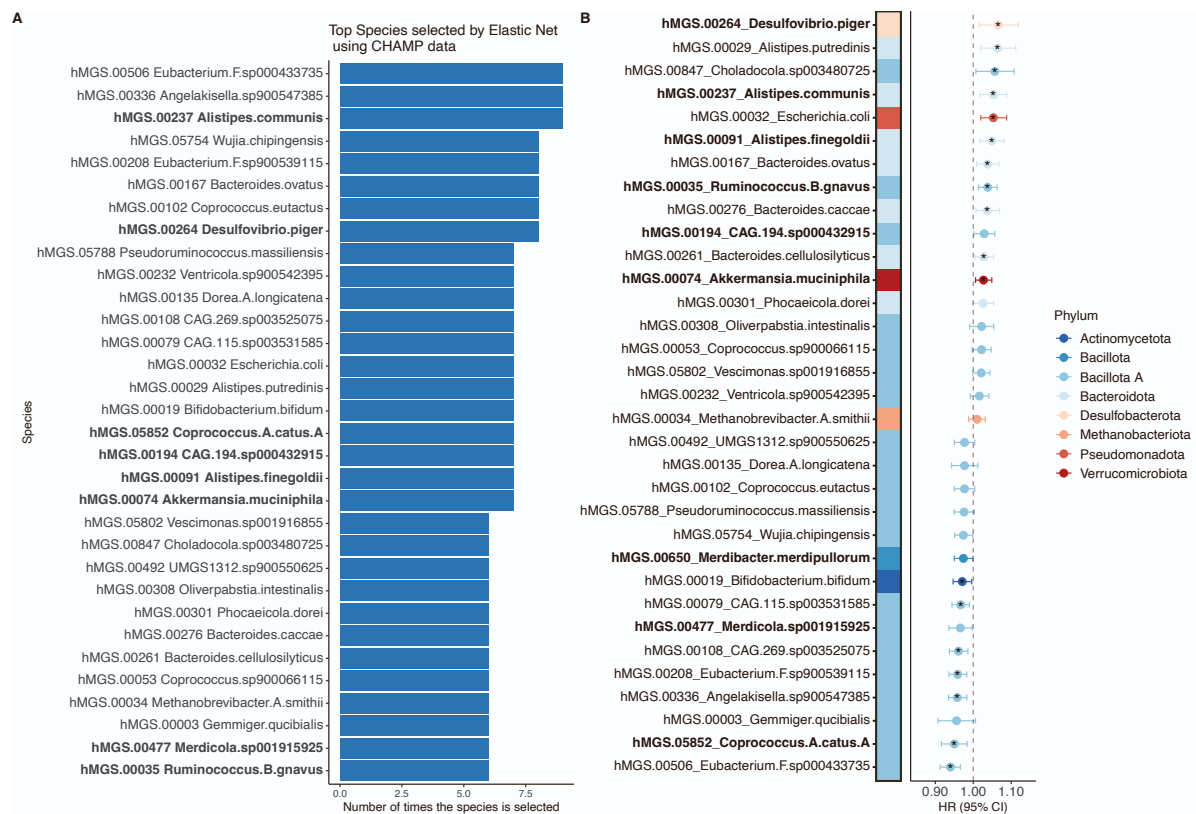


Figure S9. Replication of the species-level association using CHAMP taxonomic profiler. Related to Figure 1. (A) Species selected at least 6 times across 10 random resampling 10-fold cross validation Elastic Net. **(B)** HRs of Cox regression models between the 32 selected species by Elastic Net + *Merdibacter merdipullorum* and incident T2D. The analysis was done using the Full Analysis Set ($n_{\text{total}}=4,685$ with 383 T2D incident cases) and all models were adjusted for sex, baseline age, level of education, height and waist circumference, smoking status, walking/cycling, exercise, coffee consumption, daily intakes of total energy, alcohol, wholegrains, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth. Species are colored according to their phylum, and species in bold correspond to the CHAMP-labeled strong species predictors identified in the primary analysis using MetaPhlAn taxonomic profiling data. * q-value < 0.05.

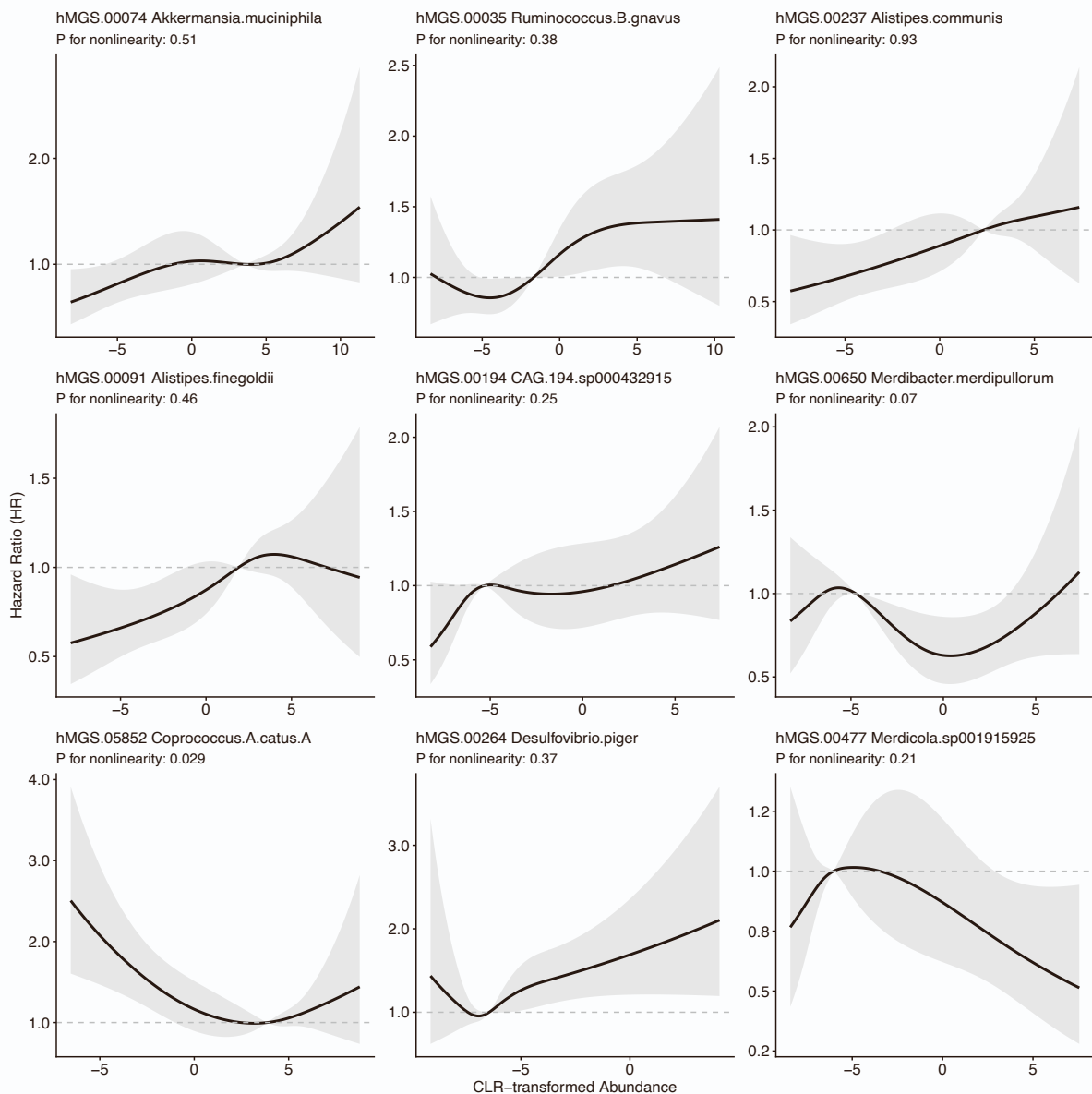


Figure S10. Restricted cubic spline (RCS) curves from Cox proportional hazards models showing adjusted hazard ratios (HR) for incident T2D across centered log-ratio (CLR) abundance of the 9 robust species identified in primary analysis using CHAMP taxonomic profiling data. Related to Figure 2. The solid line corresponds to the estimated HR and the shaded band the 95% CI. The horizontal dashed line indicates HR = 1, corresponding to the reference point set at the median abundance of each species. The p-value for the nonlinear term (P-nonlinear) was estimated from a likelihood-ratio test comparing the RCS term with a linear term. Models are adjusted for statin medication use, sex, baseline age, education, height, waist circumference, smoking, total energy intake, coffee, alcohol, walking/biking, exercise, whole grains, yogurt, red/processed meat, and sweet foods/beverages, aliquoting plate, and sequencing depth. Knots are placed at the 5th, 35th, 65th, and 95th percentiles. The RCS are shown using data from the Full Analysis Set ($n_{\text{total}}=4,685$ with 383 T2D incident cases).

Table S1. FFQ items used for the construction of dietary variables. Related to STAR Methods.

FFQ Sugary food/Sweetened beverages variables	FFQ Red/Processed variables	FFQ Wholegrains variables	FFQ Yogurt variables
Sugary food	Red meat	Wholegrain pasta/spaghetti/macaroni Wholegrain rice (råris)	Sour milk/Yogurt (≥ 3% fat) Low-fat Sour milk/Yogurt (0.5% fat) Fruit flavoured Yogurt/Sour milk
Sugar	Pork (steak/casserole)		
Honey	Beef/veal (steak/casserole)	Whole grain/meal bread	
Buns/cookies	Minced meat dishes (Meatballs, hamburger, minced meat sauce)	Crisp bread/Hard bread	
Biscuits/wafers	Processed meat	Oatmeal porridge	
Cakes/pastries	Falukorv (sausage)	Muesli	
Fruit soup/kissel	Liver paté		
Lingonberry jam	Liver paté (low-fat)		
Other jam	Cold cut meats (e.g. ham/turkey)		
Chocolate	Cold cut sausages (e.g. salami)		
Candy (not chocolate)			
Ice cream			
Sweetened beverages			
Orange/grapefruit juice			
Coca Cola/Pepsi			
Other Soda/Squash (carbonated/non-carbonated)			

Each variable was derived as g/day.

Table S2. Alpha and Beta diversity association with incident T2D in the Full Analysis Set and the Lag Time Analysis Set. Related to Results.

Diversity metrics	Full Analysis Set			Lag Time Analysis Set		
	HR 95% IC	p-value	q-value	HR 95% IC	p-value	q-value
PC1	1.004 (95% CI, 1-1.008)	0.072	0.193	1.004 (95% CI, 1-1.009)	0.065	0.129
PC2	1.001 (95% CI, 0.995-1.007)	0.771	0.771	1 (95% CI, 0.994-1.007)	0.982	0.982
PC3	0.994 (95% CI, 0.987-1.002)	0.164	0.263	0.996 (95% CI, 0.987-1.004)	0.311	0.355
PC4	0.993 (95% CI, 0.985-1.001)	0.105	0.210	0.991 (95% CI, 0.982-1)	0.051	0.129
PC5	1.016 (95% CI, 1.007-1.025)	0.001	0.007	1.012 (95% CI, 1.002-1.022)	0.018	0.129
PC6	1.013 (95% CI, 1.004-1.023)	0.007	0.028	1.011 (95% CI, 1.001-1.022)	0.033	0.129
Shannon	0.847 (95% CI, 0.659-1.09)	0.197	0.263	0.813 (95% CI, 0.619-1.068)	0.137	0.218
Species richness	0.999 (95% CI, 0.997-1.001)	0.496	0.567	0.999 (95% CI, 0.997-1.001)	0.282	0.355

Cox regression models adjusted for sex, baseline age, level of education, height and waist circumference, smoking status, walking/cycling, physical exercise, coffee consumption, daily intakes of total energy, alcohol, wholegrains, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth. p-values and q-values < 0.05 are marked in bold.

Table S3. Interaction analysis of *Akkermansia muciniphila* and dietary fiber intake on in Cox proportional hazards models. Related to Figure 3.

Analysis set	exposure	interaction_var	lrt_p_value	aic_no_int	aic_with_int	int_var_type	int_var_category	int_term_1	int_p_value
FAS	<i>A. muciniphila</i>	dietary_fibers	0.538	5248.107	5249.729	continuous	NA	<i>A. muciniphila</i> :dietary_fibers	0.535
	<i>A. muciniphila</i>	dietary_fibers_Q	0.302	5249.310	5251.665	categorical	Q1, Q2, Q3, Q4	<i>A. muciniphila</i> :dietary_fibers_Q2	0.058
								<i>A. muciniphila</i> :dietary_fibers_Q3	0.401
								<i>A. muciniphila</i> :dietary_fibers_Q4	0.476
LTAS	<i>A. muciniphila</i>	dietary_fibers	0.340	4487.136	4488.227	continuous	NA	<i>A. muciniphila</i> :dietary_fibers	0.335
	<i>A. muciniphila</i>	dietary_fibers_Q	0.321	4487.596	4490.102	categorical	Q1, Q2, Q3, Q4	<i>A. muciniphila</i> :dietary_fibers_Q2	0.069
								<i>A. muciniphila</i> :dietary_fibers_Q3	0.282
								<i>A. muciniphila</i> :dietary_fibers_Q4	0.207

Results of the Cox proportional hazards models evaluating the interaction between *Akkermansia muciniphila* (exposure) and dietary fiber intake (interaction variable) on the risk of T2D. The analysis compares models with and without the interaction term.

FAS: Full Analysis Set (n_{total}=4685 with 383 T2D incident cases), LTAS: Lag Time Analysis Set (n_{total}=4,633 with 331 T2D incident cases).

lrt p-value: P-value from the likelihood ratio test comparing the models with and without the interaction term.

aic_no_int: Akaike Information Criterion (AIC) for the model without the interaction term,

aic_with_in: AIC for the model including the interaction term.

Models were adjusted for all previous covariates i.e, sex, baseline age, level of education, height and waist, smoking status, walking/cycling, physical exercise, coffee consumption, daily intakes of total energy, alcohol, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth, except for wholegrains.

Table S4. Robust features association with incident T2D in the Full Analysis Set and the Lag Time Analysis Set adjusting for fasting plasma glucose. Related to Figures 1 and 4.

Species features	Full Analysis Set			Lag Time Analysis Set		
	HR 95% IC	p-value	q-value	HR	p-value	q-value
Alistipes_communis	1.053 (95% CI, 1.02-1.088)	0.002	0.010	1.048 (95% CI, 1.012-1.084)	0.008	0.031
Ruminococcus_gnavus	1.026 (95% CI, 1.004-1.048)	0.021	0.045	1.023 (95% CI, 0.999-1.047)	<i>0.061</i>	<i>0.081</i>
Desulfovibrio_piger	1.018 (95% CI, 0.993-1.044)	0.160	0.192	1.029 (95% CI, 1.001-1.056)	0.039	<i>0.081</i>
Akkermansia_muciniphila	1.018 (95% CI, 0.994-1.042)	0.149	0.192	1.026 (95% CI, 1-1.053)	0.047	<i>0.081</i>
Alistipes_finegoldii	1.011 (95% CI, 0.985-1.038)	0.407	0.444	1.012 (95% CI, 0.984-1.041)	0.416	0.416
GGB3614_SGB4886	1.01 (95% CI, 0.983-1.038)	0.475	0.475	1.016 (95% CI, 0.987-1.046)	0.283	0.308
Coprococcus_catus	0.976 (95% CI, 0.945-1.008)	0.137	0.192	0.975 (95% CI, 0.942-1.009)	0.153	0.183
Erysipelotrichaceae_bacterium	0.972 (95% CI, 0.949-0.996)	0.023	0.045	0.975 (95% CI, 0.95-1)	<i>0.054</i>	<i>0.081</i>
Clostridia_unclassified_SGB6317	0.953 (95% CI, 0.924-0.982)	0.002	0.010	0.958 (95% CI, 0.928-0.989)	0.008	0.031
Gut Metabolic Modules (GMMs)						
MF0042_asparagine degradation	1.088 (95% CI, 1.016-1.164)	0.015	0.045	1.092 (95% CI, 1.015-1.174)	0.018	<i>0.055</i>
MF0018_mannose degradation	0.589 (95% CI, 0.389-0.893)	0.013	0.045	0.524 (95% CI, 0.337-0.816)	0.004	0.031
MF0071 pentose phosphate pathway (non oxidative branch)	0.486 (95% CI, 0.19-1.243)	0.132	0.192	0.349 (95% CI, 0.126-0.964)	0.042	<i>0.081</i>

Cox regression models adjusted for sex, baseline age, level of education, height and waist circumference, smoking status, walking/cycling, physical exercise, coffee consumption, daily intakes of total energy, alcohol, wholegrains, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth and **fasting plasma glucose**. p-values and q-values < 0.05 are marked in bold and p-values and q-values < 0.1 are marked in italic.

Table S5. Subdistribution hazard ratios for incident type 2 diabetes from Fine and Gray competing risk models, considering death as a competing event. Related to Figures 1 and 4.

Species features	Full Analysis Set			Lag Time Analysis Set		
	SHR 95% IC	p-value	q-value	SHR 95% IC	p-value	q-value
Alistipes_communis	1.045 (95% CI, 1.013-1.078)	0.005	0.008	1.037 (95% CI, 1.004-1.071)	0.028	0.033
Ruminococcus_gnavus	1.035 (95% CI, 1.014-1.057)	0.001	0.003	1.032 (95% CI, 1.008-1.056)	0.0083	0.015
Desulfovibrio_piger	1.033 (95% CI, 1.010-1.058)	0.006	0.008	1.034 (95% CI, 1.008-1.061)	0.009	0.015
Akkermansia_muciniphila	1.027 (95% CI, 1.004-1.051)	0.023	0.025	1.032 (95% CI, 1.007-1.058)	0.013	0.017
Alistipes_finegoldii	1.034 (95% CI, 1.008-1.060)	0.011	0.015	1.027 (95% CI, 0.999-1.056)	0.062	0.068
GGB3614_SGB4886	1.023 (95% CI, 0.997-1.050)	0.087	0.087	1.025 (95% CI, 0.996-1.054)	0.088	0.088
Coprococcus_catus	0.961 (95% CI, 0.932-0.992)	0.015	0.018	0.957 (95% CI, 0.924-0.990)	0.012	0.017
Erysipelotrichaceae_bacterium	0.960 (95% CI, 0.938-0.984)	0.001	0.003	0.962 (95% CI, 0.937-0.987)	0.0028	0.006
Clostridia_unclassified_SGB6317	0.949 (95% CI, 0.923-0.976)	0	0.001	0.954 (95% CI, 0.927-0.982)	0.0016	0.005
Gut Metabolic Modules (GMMs)						
MF0042_asparagine degradation	1.124 (95% CI, 1.056-1.197)	0	0.001	1.119 (95% CI, 1.044-1.199)	0.0015	0.004
MF0018_mannose degradation	0.446 (95% CI, 0.304-0.653)	0	0	0.438 (95% CI, 0.291-0.660)	0	0
MF0071 pentose phosphate pathway (non oxidative branch)	0.256 (95% CI, 0.099-0.662)	0.005	0.008	0.188 (95% CI, 0.067-0.527)	0.0015	0.005

Subdistribution hazard ratios (SHRs) for incident type 2 diabetes estimated using Fine and Gray competing risk models, treating death as a competing event. Models were adjusted for sex, baseline age, level of education, height and waist circumference, smoking status, walking/cycling, physical exercise, coffee consumption, daily intakes of total energy, alcohol, wholegrains, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth.

Table S6. Complete cases analysis including 3114 participants without missing data for the covariates based on the Full Analysis Set. Related to Figures 1 and 4.

Diversity metrics	HR 95% CI	p-value	q-value
PC1	1.005 (95% CI, 1-1.01)	0.054	0.187
PC2	1.003 (95% CI, 0.995-1.011)	0.491	0.491
PC3	0.996 (95% CI, 0.986-1.006)	0.404	0.491
PC4	0.996 (95% CI, 0.986-1.007)	0.491	0.491
PC5	1.011 (95% CI, 0.999-1.022)	0.07	0.187
PC6	1.012 (95% CI, 0.999-1.024)	0.063	0.187
Shannon	0.785 (95% CI, 0.572-1.077)	0.134	0.214
Species richness	0.998 (95% CI, 0.996-1)	0.095	0.19
Species features			
Bilophila_wadsworthia	1.048 (95% CI, 1.009-1.089)	0.015	0.051
Akkermansia_muciniphila	1.046 (95% CI, 1.015-1.077)	0.003	0.014
Alistipes_communis	1.04 (95% CI, 1.002-1.079)	0.038	0.096
Desulfovibrio_piger	1.039 (95% CI, 1.009-1.069)	0.010	0.039
Ruminococcus_gnavus	1.029 (95% CI, 1.003-1.056)	0.028	0.081
Alistipes_finegoldii	1.029 (95% CI, 0.997-1.062)	0.077	0.127
Parabacteroides_distasonis	1.028 (95% CI, 0.979-1.078)	0.266	0.340
Alistipes_putredinis	1.026 (95% CI, 0.983-1.071)	0.233	0.320
GGB3614_SGB4886	1.02 (95% CI, 0.987-1.053)	0.237	0.320
GGB9712_SGB15244	1.007 (95% CI, 0.977-1.036)	0.663	0.727
Bacteroides_caccae	1.003 (95% CI, 0.972-1.034)	0.857	0.857
Dorea_longicatena	0.99 (95% CI, 0.936-1.048)	0.733	0.767
GGB4585_SGB6340	0.987 (95% CI, 0.959-1.016)	0.376	0.432
Coprococcus_eutactus	0.986 (95% CI, 0.96-1.014)	0.329	0.399
Bacteroides_faecis	0.981 (95% CI, 0.952-1.011)	0.214	0.320
Clostridia_unclassified_SGB14844	0.975 (95% CI, 0.949-1.003)	0.077	0.127
Blautia_sp_AF19_10LB	0.975 (95% CI, 0.947-1.003)	0.077	0.127
Ruminococcaceae_unclassified_SGB4191	0.972 (95% CI, 0.945-0.999)	0.042	0.096
Coprococcus_catus	0.965 (95% CI, 0.93-1.001)	0.054	0.112
Erysipelotrichaceae_bacterium	0.955 (95% CI, 0.928-0.982)	0.001	0.008
Bacilli_bacterium	0.953 (95% CI, 0.927-0.981)	0.001	0.007
Clostridia_unclassified_SGB6317	0.94 (95% CI, 0.908-0.973)	0.001	0.006
GGB3478_SGB4643	0.933 (95% CI, 0.902-0.965)	0	0.001
Gut Metabolic Modules (GMMs)			
MF0042_asparagine degradation	1.116 (95% CI, 1.029-1.21)	0.008	0.149
MF0018_mannose degradation	0.363 (95% CI, 0.215-0.612)	0	0.015
MF0071 pentose phosphate pathway (non oxidative branch)	0.239 (95% CI, 0.078-0.726)	0.012	0.149

Cox regression models adjusted for sex, baseline age, level of education, height and waist circumference, smoking status, walking/cycling, physical exercise, coffee consumption, daily intakes of total energy, alcohol, wholegrains, yogurt, red/processed meat, sugary food/sweetened beverages, statin medication use, aliquoting plate, and sequencing depth. p-values and q-values < 0.05 are marked in bold.