Overlapping soil and gut microbiome compound processing potential in a gradient of ecosystem quality and subjects with type 2 diabetes

Craig Liddicoat[a,*] and Martin F. Breed[a]

[a]College of Science and Engineering, Flinders University, Bedford Park, South Australia, Australia

***Corresponding author:** Craig Liddicoat; craig.liddicoat@flinders.edu.au

ORCIDs: CL: 0000-0002-4812-7524, MFB: 0000-0001-7810-9696

Competing Interest Statement: The authors declare no competing interest.

Keywords: Healthy microbiome, type 2 diabetes, metagenomics, compound processing potential, soil ecology, human health

Abstract

Background: Type 2 diabetes (T2D) is a complex metabolic disease with uncertain pathogenesis, including recognized influences from gut microbiome-host interactions and environmental factors. Meanwhile, soil microbiomes are known to shift in their composition and functional capacities with ecosystem condition, and represent an important environmental source for priming and resupplying the human gut microbiome.

Objectives: We sought to examine overlapping functional capacities in model case study microbial metagenomes from a gradient of soil ecosystem quality (disturbed to mature/natural sites under forest ecosystem restoration) and gut microbiomes in T2D versus normal health, to investigate plausible conceptual links (via ambient environmental exposures) between ecosystem degradation, soil microbiome disturbance and metabolic anomalies encoded in T2D gut metagenomes.

Methods: We used a novel method to translate metagenome functional capacities (i.e., function potential relative abundances %) to compound processing potential (CPP, %), reflecting potential metabolism at the scale of individual compounds. Soil ecosystem and T2D case study datasets were compared in terms of their potential metabolism for key metabolic compounds, via compound-wide trend analysis and community-wide function and CPP composition analyses.

Results: Potential metabolism of several important groups of compounds in soils and T2D (e.g., sugars, starches, proteins involved in fatty acid biosynthesis, aromatics involved in energy regulation) correlated strongly with soil disturbance and with T2D. Based on our findings, we derived a preliminary gut microbiome index for T2D status informed by consistent CPP trends found in both disturbed soils and T2D gut samples that was capable of discriminating T2D from normal healthy subjects (Wilcox-Mann-Whitney $W = 1001$, $P = <0.001$).

Discussion: Our findings support the profound new hypothesis that soil microbiome exposures from disturbed ecosystems, together with starch-rich diets, may bias the functional capacity profile of the

human gut microbiome in T2D. Our hypothesis, that ecosystem degradation and exposure to associated disturbed soils contribute to T2D gut microbiomes which have (1) excessive potential to metabolise compounds involved in energy harvesting, and (2) deficient potential to metabolise compounds involved in energy management (e.g., storage, regulation), warrants further investigation.

Introduction

Environment and lifestyle exposures under the megatrends of rising urbanization and reducing nature contact are associated with an array of non-communicable diseases including many with suspected microbiome-mediated links (Flandroy et al. 2018; Flies et al. 2019; Frumkin et al. 2017; Sonnenburg and Sonnenburg 2019). Among these, type 2 diabetes (T2D) is a major chronic multifaceted metabolic disease with uncertain pathogenesis, characterized by insulin resistance, loss of insulin secretion, high blood sugar levels and wider metabolic dysregulation (Beulens et al. 2022; Gonzalez et al. 2018; Lu et al. 2024; Pearson 2019; Taylor 2024). A growing evidence base supports the influence of gut microbiomes in human metabolic health and dysfunction, however there is a pressing need to better understand the contribution of microbiomes in T2D (Cunningham et al. 2021; Gurung et al. 2020) due to their potential mediating influence via diet, lifestyle and environmental factors. In particular, the contribution of environmental microbiomes in a period of rapid global change (IPBES 2018), to the maintenance of human metabolic health remains understudied (Flandroy et al. 2018). T2D is a major contributor to the global non-communicable disease pandemic (Cheema et al. 2014). In 2021, T2D affected ~96% of the total 529 million people (6.1% prevalence) with all forms of diabetes globally (Ong et al. 2023). By 2050, more than 1.31 billion people are projected to have diabetes, with prevalence rates expected to exceed 10% in 89 countries (Ong et al. 2023). The global cost of diabetes was estimated at US$ 966 billion in 2021, and is projected to rise to US $1.05 trillion by 2045 (Sun et al. 2022).

The human gut microbiome is directly involved in our metabolism, shaping the biochemical profile of diet via an extensive metabolic repertoire that is distinct yet complementary to mammalian digestive enzymes (Rowland et al. 2018). The gut microbiome and metabolites derived from digesting host diet help to shape our physiological responses and energy regulation through modulation of appetite, energy harvest, modulation of gene expression, gut motility, and forms of dietary energy

72 conversion and storage (Krajmalnik-Brown et al. 2012). Specifically in T2D, shifts in gut microbiome
73 community composition and function have been identified compared to normal healthy subjects, which
74 are in part alleviated by treatment with the antidiabetic drug metformin (Forslund et al. 2015). Gut
75 microbiome structures shift profoundly with long-term changes in diet (Gilbert et al. 2018). Such
76 remodelling of our metabolic machinery also occurs through diet and microbiome changes that modify
77 the expression and operation of intestinal enteroendocrine cells (EECs) responsible for sensing and
78 responding to nutrients (e.g., shifting from serotonin to peptide producing EECs, or microbiome-
79 dependent silencing of EECs; Aliluev et al. 2021; Ye et al. 2019). These are two-way interactions, as
80 loss of EECs and their secreted hormones also changes microbiome structures (Blot et al. 2023).
81 In contrast to the small scale activity of gut microbes and intestinal EECs, numerous
82 assessments attribute the majority of T2D burden to modifiable and environmental risk factors
83 (Beulens et al. 2022). The external exposome of possible contributors to T2D risk extends well beyond
84 lifestyle and dietary behaviors to include the built, natural, social, and physicochemical environments
85 (Beulens et al. 2022). Among this exposome, strong evidence links T2D to urbanization. In a country-
86 scale global ecological analysis, Gassasse et al. (2017) found urban concentration to be strongly
87 positively associated with T2D prevalence, while country-level sugar consumption was not directly
88 identified as a determinant. In rapidly modernizing China, Attard et al. (2012) found twofold higher
89 T2D prevalence in urban versus rural areas. In southern Asia, T2D prevalence in urban areas was more
90 than double the rate of rural areas in over 55 year olds (Cheema et al. 2014). In Peru, urban dwellers
91 had 4.3-fold higher risk for developing T2D compared to rural dwellers, while rural-to-urban migrants
92 with at least 30 years of urban exposure also had 2.7-fold higher risk (Ruiz-Alejos et al. 2018).
93 Microbiomes represent a key potential linkage mechanism between environmental exposures
94 consistent with T2D risk and metabolic processes occurring in the human gut. Gut microbiomes are
95 predominantly shaped by environments, not host genetics (Rothschild et al. 2018). In particular, soils

96 represent a rich source of environmental microbial exposures, providing home to ~59% of species on
97 Earth (Anthony et al. 2023). The gut is dominated by spore-forming bacteria capable of surviving and
98 dispersing through aerobic environments before re-establishing in suitable oxygen-deprived habitats
99 (Browne et al. 2016). Exposure to naturally biodiverse soil microbiomes shape and supplement gut
100 microbiome composition and functional capacity particularly from a young age (Liu et al. 2021;
101 Roslund et al. 2021; Zhou et al. 2024). Airborne microbial communities are also shaped by
102 environments (e.g., land use, green space, soils) (Mhuireach et al. 2016; Robinson et al. 2021) and can
103 at least partially transfer to the gut (Liddicoat et al. 2020). Consequently, soil microbes may be
104 introduced to the gut environment through direct soil contact, food-borne and airborne (dust)
105 microbiomes. Moreover, growing knowledge suggests that soil microbiome composition and functional
106 capacity may shift in generalizable ways with ecosystem quality, with potential to impact human health
107 (Delgado-Baquerizo et al. 2021; Liddicoat et al. 2022; Liddicoat et al. 2024). Therefore, as human
108 populations increasingly urbanize, unchecked processes of soil microbiome degradation in cities and
109 peri-urban landscapes will have increasing potential to negatively impact human health (Sun et al.
110 2023). Exploring overlaps in the functional capacities of (1) naturally biodiverse versus degraded soil
111 microbiomes (linked to aboveground disturbed versus intact ecosystem condition) and (2) normal
112 healthy versus T2D gut microbiomes may reveal previously unappreciated microbiome linkages
113 between environments and T2D.

114 Despite their importance, the measurement and definition of "healthy microbiomes" remain
115 uncertain (Van Hul et al. 2024). Community-scale microbial metabolism, where bioenergetic
116 machinery and resources (including metabolites and electron transfer) are shared in the extracellular
117 space (Daisley et al. 2021) adds complexity and means that multiple taxa may often combine to
118 perform functional roles of interest. The need for improved insight into the contributions of
119 microbiomes in health and disease, and spanning ecosystems to humans, has motivated efforts to

analyze soil and gut metagenomes through the lens of their *compound processing potential* (CPP) (Liddicoat et al. 2024; Figure S1). A compound-oriented approach to the analysis of functional roles of microbiomes carries promise because we typically possess richer knowledge of bioenergetic and health associations with chemical compounds (e.g., microbial substrates, feedstocks and metabolites such as carbohydrates, lipids, proteins, amino-sugars, vitamins, phytochemicals and pharmaceuticals) compared to tens of thousands of poorly characterized taxa. New insights into differential compound processing by microbiomes may also point to mechanistic hypotheses and/or possible therapeutic targets. Microbial metagenomes (entire collective DNA) are used to examine functions, however existing approaches require specialized knowledge of complex biochemical pathways and do not provide resolution at the deeper fundamental level of the chemical compounds involved. Existing metabolite prediction frameworks are also unsuited to highly biodiverse and varying environments such as soils (Liddicoat et al. 2024).

The goal of this study was to explore and compare model case study soil and gut metagenome datasets from a gradient of ecosystem quality and in subjects with T2D versus normal health through quantifying their genomic potential to metabolize an array of chemical compounds. Specifically, the aims were to improve functional and metabolic insights to the role of microbiomes in ecosystems and T2D, and their potential overlaps, through identifying: (1) compounds with differential potential metabolism (i.e., with trending CPP) based on analysis of soil metagenomes across an ecosystem disturbance-quality gradient; (2) compounds with differential potential metabolism (based on CPP values) from gut metagenomes in health versus untreated T2D; and (3) overlapping compounds and biochemical/functional pathways between (1) and (2) that may reveal new insights to potential environmental microbiome contributions to T2D.

Methods

144 ***Case study data***

145 We examined two model case study datasets (Table 1) comprising previously published (1) soil
146 metagenomes from degraded post-mining to reforested and unmined natural reference sites (Sun and
147 Badgley 2019), and (2) gut metagenomes from subjects with T2D and normal healthy controls
148 (Forslund et al. 2015).

149

150 **Table 1.** Description of case study metagenome datasets.

Case study focus	**Main comparison groups (and sample/subject numbers)**	**Metagenome functional profile characteristics (sample mean ± s.d.)**	**Source data reference**	**Study nation of origin**
Soil				
Post-mining forest ecosystem restoration	Soil samples spanned revegetation ages of 6, 12, 22, 31 years, and unmined (UM) samples (treated as ordinal variables). Comprising five age-based groups, each with three replicates. Total n = 15	Functions from SUPER-FOCUS profiling: n = 30,125 (20,328 ± 767). Functional % all sum to 100%. Compounds from CPP profiling: n = 8,370 (7,777 ± 92). CPP (%): (67 ± 0.4 %)	MG-RAST project mgp16379 (Sun and Badgley 2019)	United States
Human gut				
Type 2 diabetes (T2D) vs. Normal	T2D with no metformin treatment (T2D met-, n = 33), vs. normal healthy (n = 43). Subjects are females only. Ages: 69-72 years	Functions from SUPER-FOCUS profiling: n = 17,962 (7861 ± 1699). Functional % all sum to 100%. Compounds from CPP profiling: n = 7,031 (5,292 ± 584). CPP (%): (70.5 ± 2.9 %)	SRA accession PRJEB1786 (Forslund et al. 2015; Karlsson et al. 2013)[a]	Sweden

151 Note: [a]Forslund et al. (2015) applied metformin treatment data in a reanalysis of metagenome data
152 from Karlsson et al. (2013). A subset of the Swedish cohort is analyzed here, excluding subjects that
153 were treated with metformin or diagnosed with impaired glucose tolerance (prediabetes).

154

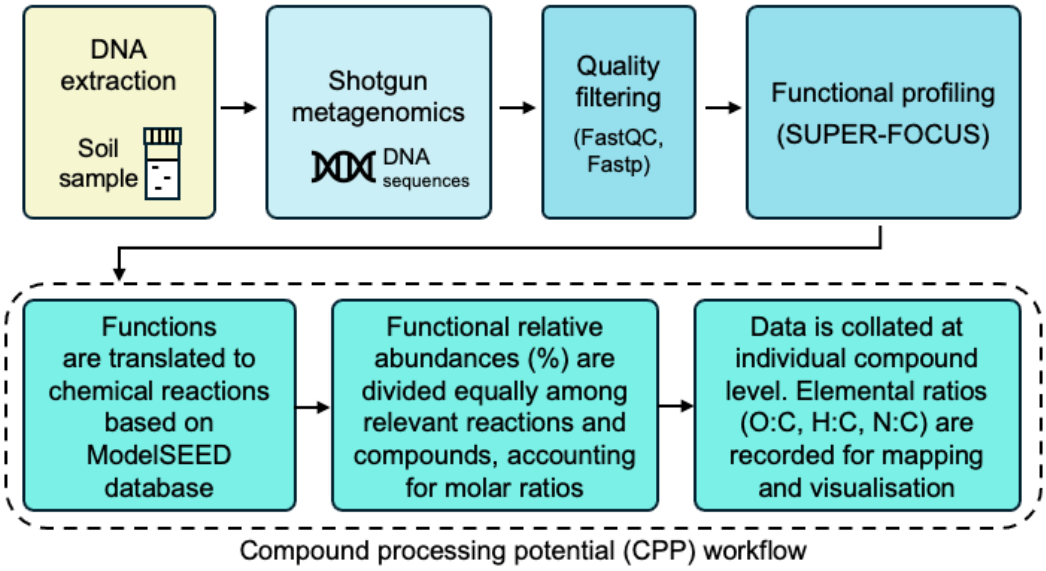
155 ***Metagenome functional profiling via compound processing potential (CPP)***

156 CPP measures were derived following steps outlined in Figure 1. Initially, following Liddicoat et al.
157 (2024), raw metagenomic sequences were accessed, trimmed for quality control, then classified to
158 SEED subsystem (Overbeek et al. 2013) functional annotations and relative abundances using SUPER-
159 FOCUS software (Silva et al. 2015). Every SUPER-FOCUS function (output row) was translated to
160 one or more corresponding chemical reaction(s) using a purpose-built R- script algorithm based on
161 ModelSEED database lookup tables (from https://github.com/ModelSEED/ModelSEEDDatabase;
162 accessed 10 Aug 2022). Linking of functions to chemical reactions was based on either: full matching
163 of functional hierarchies (using subsystem-class, -subclass, -name and -role); detection of EC number;
164 or matching of SUPER- FOCUS function name within ModelSEED lookup tables for reactions
165 (reaction name or alias), subsystems (role), or reaction-pathways (external reaction name). Reactions
166 were then linked to corresponding compounds using ModelSEED database tables ('reactions.tsv',
167 'compounds.tsv'). Advancing from the previous method used in Liddicoat et al. (2024), here we
168 derived CPP measures at the resolution of individual compounds (all software parameters and
169 supporting R code for CPP calculations are available at: https://github.com/liddic/cpp3d):

170 a. Function level relative abundances (%) from SUPER-FOCUS were allocated to compounds by
171 dividing values equally across all sub-functions and reactions, and dividing again across all
172 corresponding compounds, with weightings to account for stoichiometric coefficients for each
173 compound within each reaction, with equal consideration to reactant and product compounds.

174 b. Relative abundances (%) for respective unique compounds (which may arise across multiple
175 reactions and functions/pathways) were then collated and summed to provide a summary CPP
176 relative abundance (%) value at the individual compound level.

177 c. To assist further mapping visualization of CPP patterns for carbon-containing compounds,
178 elemental ratios of O:C, H:C, and N:C were calculated. Such chemical mapping frameworks
179 have been used previously to display and interpret groupings of energetically and chemically
180 similar compounds (Liddicoat et al. 2024; Rivas-Ubach et al. 2018).
181



182

183 **Figure 1.** Overview of the CPP analysis workflow used to translate microbial metagenome functional
184 potential relative abundances to quantify potential metabolism at the level of individual compounds.
185

186 ***Statistical Analysis***

187 Three strategies were used to analyze microbiome CPP profiles across the soil ecosystem and T2D gut
188 samples (further detail is in Supplemental Material and at https://github.com/liddic/cpp3d):

189 a. ***Selected compounds*** were chosen due to their relevance to soil and/or gut health (e.g., Daisley
190 et al. 2021; Jardon et al. 2022; Six et al. 2006; compounds and reference information are listed
191 in Table S1). CPP trends with soil disturbance were evaluated using ordinal (Kendall tau) rank
192 correlation tests. CPP trends (i.e., differences) with T2D versus normal health were evaluated
193 using Wilcoxon-Mann-Whitney tests.

194 b. ***Compound-wide association tests*** were also performed spanning all compounds represented in
195 the respective CPP datasets, using the same trend evaluation methods as in a. In soils,
196 compounds that did not have relative abundance data for at least a quarter of the total sample
197 number were excluded due to low data (while permitting detection of effects at either extreme
198 of the most disturbed or natural/unmined states). In soils, strong trends were discovered and
199 only significant results based on adjusted *P*-values were considered further (after applying
200 Benjamini and Hochberg false discovery rate correction). In the separate compound-wide
201 analysis of T2D gut samples, compounds were only included if at least one or other of the T2D
202 or normal groups contained at least 50% non-zero CPP values. The T2D Wilcoxon-Mann-
203 Whitney test results yielded many *P*-values < 0.05, however none were strong enough to
204 withstand Benjamini and Hochberg *P*-adjustment for multiple testing, therefore *P*-adjustment
205 was not performed in the compound-wide T2D gut data to retain the most interesting results.
206 Next, compounds with trends in soils and T2D gut microbiomes, and overlaps between the two,
207 were 'mapped' using their chemical formula elemental ratios of O:C, H:C, and N:C. From this
208 mapping space, prominent clusters of consistently trending compounds were identified and
209 further analysed to interpret their possible metabolic importance.
210 c. ***Beta diversity ordinations*** were performed to examine community-scale compositional patterns
211 and differences between sample groups and analysis approaches – comparing conventional
212 functional potential (%) data (termed 'functions') versus CPP (%) data. This analysis used
213 principal coordinate analysis (PCoA) ordinations based on Bray-Curtis distances, and testing for
214 differences between groups via permutational analysis of variance (PERMANOVA) and beta
215 dispersion testing for homogeneity of groupings using the vegan R package (Oksanen et al.
216 2020).
217

Results

Selected compounds with shared trends for potential metabolism in disturbed soils and T2D

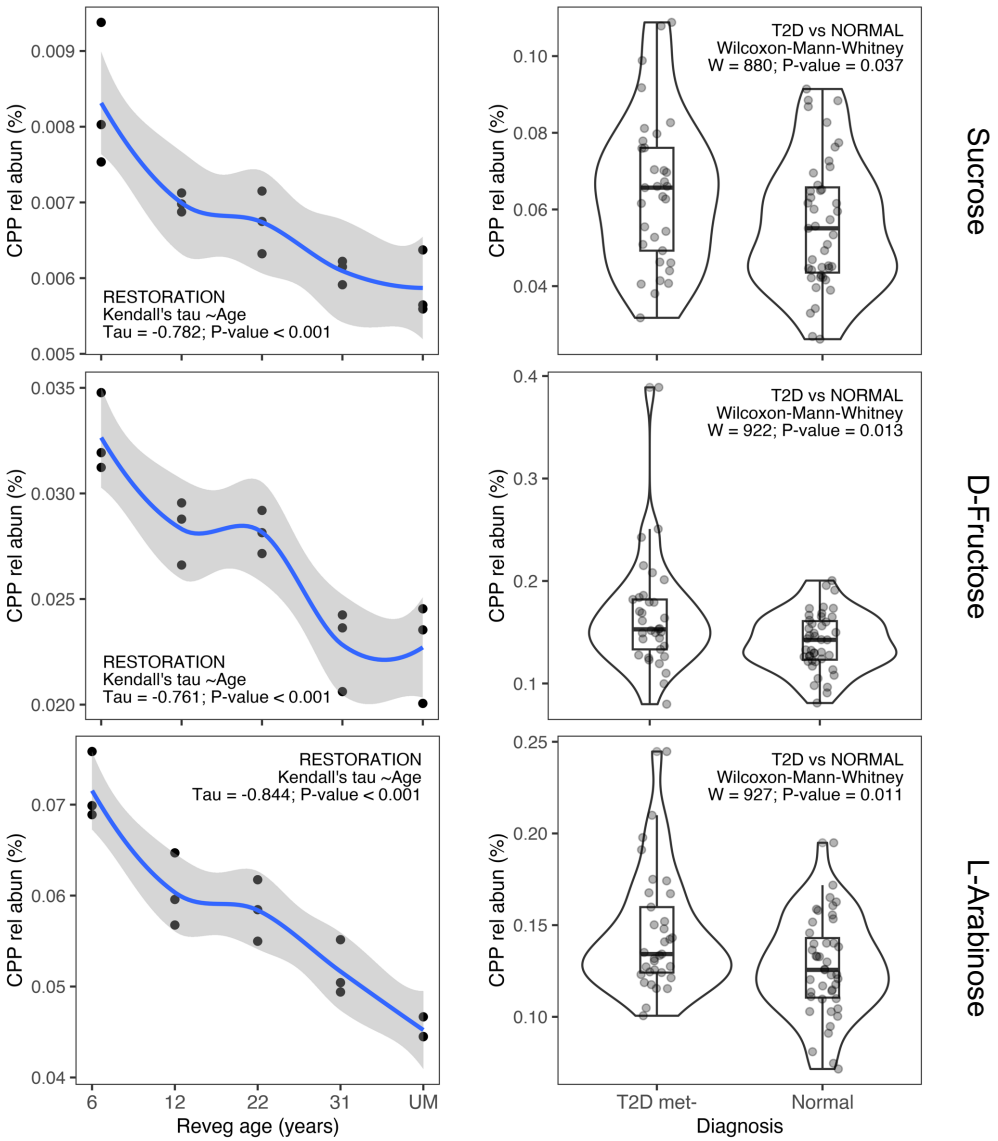
Analysis of CPP trends for selected compounds (Table 1; Figures S2-S34) found that potential metabolism for the sugars: fructose, sucrose, and arabinose increased with both soil disturbance (all $P < 0.001$) and in T2D (Figure 2). CPP values for ammonia also increased in T2D and showed marginal increase ($P < 0.1$) in disturbed soils (Figure S5). Conversely, CPP values indicating potential metabolism of lignin and trimethylamine N-oxide (TMAO) decreased with both soil disturbance and in T2D (Figure 3). Trimethylamine (TMA) potential metabolism also decreased in disturbed soils and showed marginal decrease ($P < 0.1$) in T2D (Figure S26).

Selected compounds with increasing CPP trends in soils only

Glucose, galactose, xylose, mannose (sugars), glucosamine (amino sugar), chitin (polysaccharide), arabinoxylan (hemicellulose) and hydrogen sulfide – all displayed increased potential metabolism with soil disturbance (Figures S6-S13). The potential metabolism trends for glucose, xylose, and chitin in disturbed soils were all strong ($P < 0.001$).

Selected compounds with decreasing CPP trends in soils only

Glycogen (polysaccharide), acetate, propionate, butyrate (short chain fatty acids), methane and carbon dioxide – all displayed decreased potential metabolism with soil disturbance (Figures S18-S23). The strongest trends were shown in carbon dioxide (Kendall tau = 0.84, $P < 0.001$) and acetate (Kendall tau = 0.72, $P < 0.001$).



241

242 **Figure 2.** Potential metabolism of sucrose, D-fructose and L-arabinose compounds displayed increases

243 with soil disturbance (younger revegetation age) and in T2D gut samples.

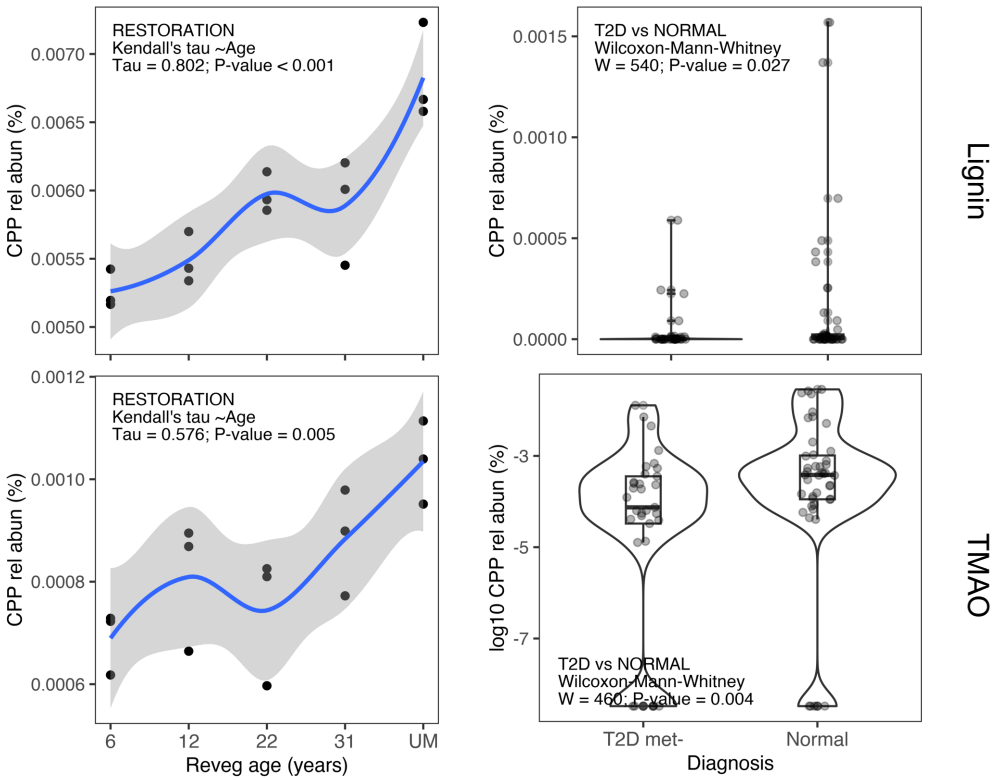


Figure 3. Potential metabolism of lignin and trimethylamine N-oxide (TMAO) decreased in more disturbed soils (younger revegetation age) and in T2D.

Selected compounds with increasing CPP trends in T2D gut only

Xyloglucan (hemicellulose), starch, and starch components amylopectin and amylose displayed increased potential metabolism in T2D gut samples (Figures S14-S17)

Selected compounds with decreasing CPP trends in T2D gut only

Potential metabolism for pectin (polysaccharide) and muramic acid (amino sugar acid) were reduced in T2D gut samples (Figures S27-S28).

Compound-wide CPP patterns and overlaps with soil disturbance and T2D gut microbiomes

Ordinal correlation trends were assessed across 8370 compounds in the soil ecosystem restoration dataset. After false discovery rate correction for multiple testing, 2122 compounds displayed significant trends (n = 757 increasing and n = 1365 decreasing with restoration age). From these, the carbon-containing compounds (total n = 1975, n = 720 increasing and n = 1255 decreasing with restoration age) were isolated and mapped using their O:C, H:C and N:C elemental ratios to enhance visualization and interpretation of energetically and chemically similar compounds that shift together (Figure 4). Example CPP % data for a single sample plotted in this manner are shown in Figure S35.

Wilcoxon-Mann-Whitney testing of differences in CPP values in T2D versus normal subjects revealed 1243 compounds with $P < 0.05$ (n = 328 increasing and n = 915 decreasing with T2D). From these, the carbon-containing compounds were isolated and visualized based on their O:C, H:C and N:C elemental ratios (Figure 5; total n = 1158, n = 294 increasing and n = 864 decreasing with T2D).

Examination of overlapping compounds with trends in both T2D and in soils under gradients of ecosystem condition, from the analyses above, revealed a total of n = 276 compounds (n = 70 were 'decreasing in T2D (reduced exposure in disturbed ecosystems)', n = 58 were 'increasing in T2D (increased exposure in disturbed ecosystems)', n = 98 were 'decreasing in T2D (reduced exposure in quality ecosystems)' , n= 50 were 'increasing in T2D (increased exposure in quality ecosystems)'). From these, the carbon-containing compounds were isolated and visualized based on their O:C, H:C and N:C elemental ratios (Figure 6; total n = 260, n = 66 were 'decreasing in T2D (reduced exposure in disturbed ecosystems)', n = 52 were 'increasing in T2D (increased exposure in disturbed ecosystems)', n = 93 were 'decreasing in T2D (reduced exposure in quality ecosystems)' , n= 49 were 'increasing in T2D (increased exposure in quality ecosystems)').

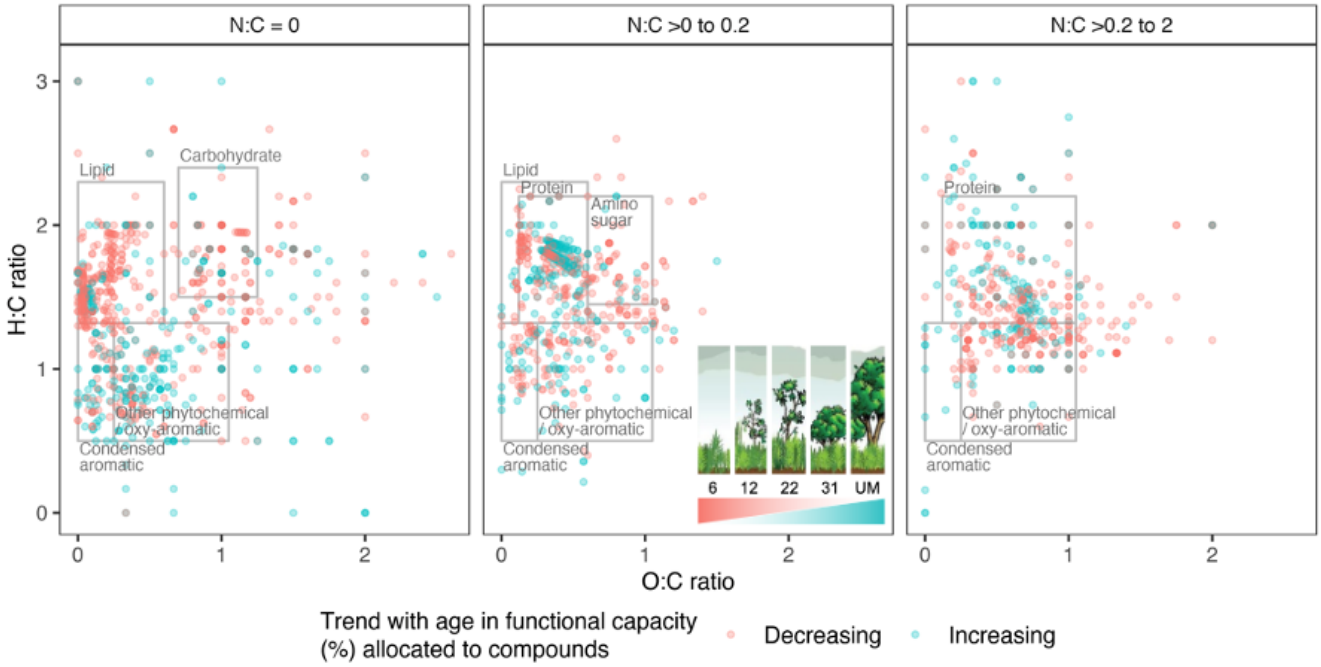


Figure 4. Compound-wide trend analysis results derived from soil metagenomes (described in Table 1) showing carbon-containing compounds with increasing (aqua, n = 720) or decreasing (red, n = 1255) compound processing potential (CPP, i.e., functional capacity % or potential metabolism at individual compound scale) with revegetation age. Each point represents a carbon-containing compound, mapped via its chemical formula O:C, H:C and N:C elemental ratios. Only compounds that have trending potential metabolism (i.e., shifting CPP % values) from disturbed to mature/natural ecosystem quality are displayed. Clusters of points indicate energetically and chemically similar compounds. The O:C by H:C panels are zoomed in from their full extent to improve visualization of more data-rich regions. N:C ratios are calculated as continuous numeric values but displayed as three depth-slices of varying thickness to represent the third dimension. Rectangular zones denoting compound classes are provided as a visual guide only (adapted from Rivas-Ubach et al. 2018; Wu et al. 2018).

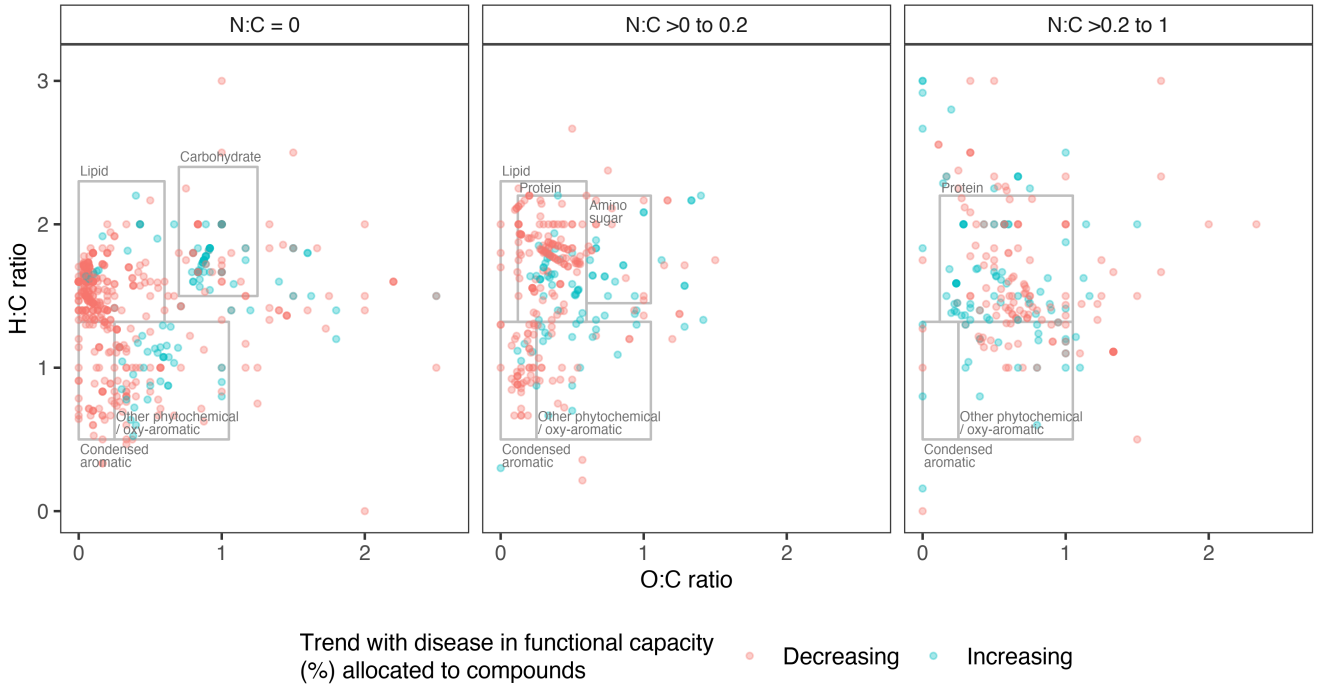


Figure 5. Compound-wide trend analysis results derived from T2D and normal gut metagenomes (described in Table 1) showing carbon-containing compounds with increasing (aqua, n = 328) or decreasing (red, n = 864) compound processing potential (CPP) in T2D. Further description of the plot format is provided in Figure 4.

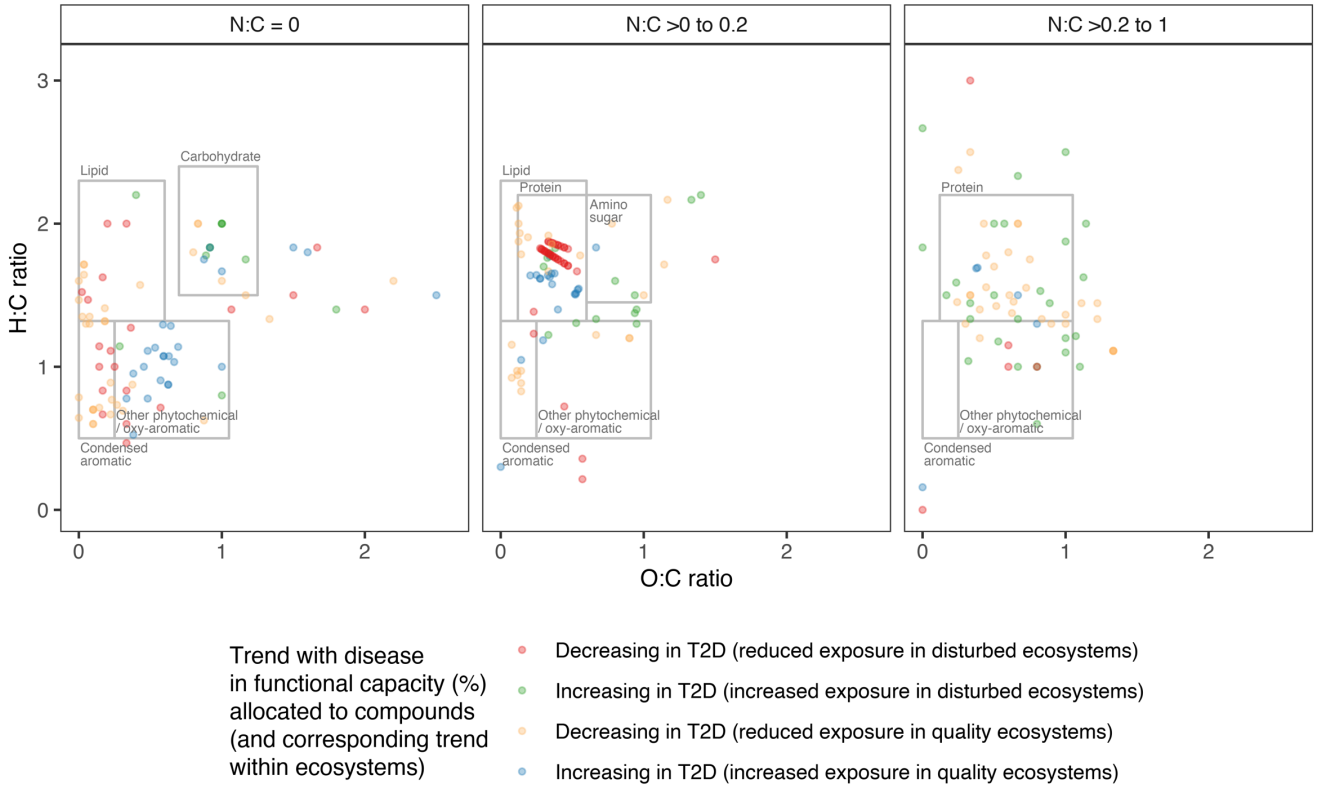


Figure 6. Overlap analysis showing carbon-containing compounds with trending potential metabolism (i.e., shifting CPP or functional capacity % at individual compound scale) in both T2D and along a gradient of soil-ecosystem condition, derived from overlapping compounds in Figures 4-5. Clusters of each color represent energetically and chemically similar compounds occurring in each of the four trend groups listed (numbers of compounds are provided in the text). Further description of the plot format is provided in Figure 4.

Exploring compounds with consistent trends in T2D and disturbed soils

To explore overlaps in compounds that share trending CPP values in T2D and soil disturbance, we examined three example groupings of compounds from Figure 6. In this analysis we emphasized trends associated with more disturbed soil ecosystems (i.e., towards younger revegetation age) for the purpose of exploring the influence of ecosystem degradation on microbiome functional capacity that may also be relevant to T2D. With regards to our emphasis on disturbance, it should be noted that the sign of the

correlation statistics reported here will be opposite, as soil CPP trend statistics were calculated based on increasing revegetation age. The first group ('group A', Table S2, Figure S36-37; these are the red points arranged in two diagonal lines in the middle panel of Figure 6) comprise 35 unique acyl-carrier proteins involved in fatty acid synthesis, which all have reduced CPP in T2D and reduced CPP in disturbed ecosystems. When analyzed collectively by summing CPP % values for all group A proteins we found that they collectively exhibit a strong decrease (Kendall tau = 0.741, P = <0.001) with soil disturbance and also show substantial reduction in T2D (Wilcox-Mann-Whitney W = 513, $P = 0.02$). Group A decreasing proteins all occur on two straight lines in the O:C, H:C, N:C mapping space (line 1: H:C = -0.625 × O:C + 2 and line 2: H:C = -0.375 × O:C + 2, where N:C is in the range >0 to 0.2).

Example group B comprise 6 carbohydrate compounds that all display increasing CPP in both T2D and with soil disturbance (Table S2, Figures S38-39, compounds are L-arabinose, melibiose, manninotriose, D-lyxulose, sucrose, D-fructose). In the O:C, H:C, N:C mapping space, these carbohydrates all occur on the line H:C = 2*O:C, with N:C = 0 (Figure S38). Analysis of summed CPP % values for group B sugars showed that they collectively exhibit a strong increase (Kendall tau = -0.844, P = <0.001) with soil disturbance and substantial increase in T2D (Wilcox-Mann-Whitney W = 984, $P = 0.002$).

Example group C comprise 10 aromatic compounds which all have decreased CPP in T2D and decreased CPP in disturbed ecosystems (Table S2, Figures S40-41, compound names are phenyl acetate (PACT), 3-chlorocatechol, 3-chlorobenzyl alcohol, phenolic-donors, phenoxyl-rad-of-phenolic-donors, sinapyl alcohol, p-coumaryl alcohol, baicalein, 6,7-dehydrobaicalein, phenylcarbinol). Collectively, these aromatics exhibit a strong decrease (Kendall tau = 0.802, P = <0.001) with soil disturbance and in T2D (Wilcox-Mann-Whitney W = 474, $P = 0.007$).

Preliminary index of T2D status informed by soil and gut microbiome CPP trends

Based on our findings, we proposed a preliminary index of T2D status informed by consistent CPP trends in both disturbed soils and T2D gut samples, and calculated using gut microbiome CPP values only. The index used CPP values rescaled to the range 1-100 from selected groups of compounds representing potential metabolism associated with energy harvest (/availability) and energy management (equation 1). The compounds used in this index calculation are detailed in the Supplemental Material and their rationale for consideration are discussed below.

T2D Index = log10(Energy harvest / Energy management) (equation 1)

Where: Energy harvest = $({}^{\#}\Sigma\ CPP_{sugars} \times {}^{\#}\Sigma\ CPP_{starches} \times {}^{\#}CPP_{ammonia})$

Energy management = $({}^{\#}\Sigma\ CPP_{\text{acyl carrier proteins}} \times {}^{\#}\Sigma\ CPP_{aromatics} \times {}^{\#}\Sigma\ CPP_{fiber} \times {}^{\#}CPP_{TMAO})$

${}^{\#}$CPP (%) values were rescaled to 1-100 before the terms were multiplied.

All CPP values used in the T2D index calculation are from gut microbiome samples.

Both energy harvest and energy management sub-indices held significant explanatory value in their own right (log10(Energy harvest): Wilcox-Mann-Whitney W = 990, $P = 0.001$; log10(1/Energy manangement): Wilcox-Mann-Whitney W = 970, $P = 0.003$), however the ability to discriminate between T2D and normal subjects improved further when these were combined (log10(Energy harvest/Energy manangement): Wilcox-Mann-Whitney W = 1001, $P = <0.001$; Figure 7).

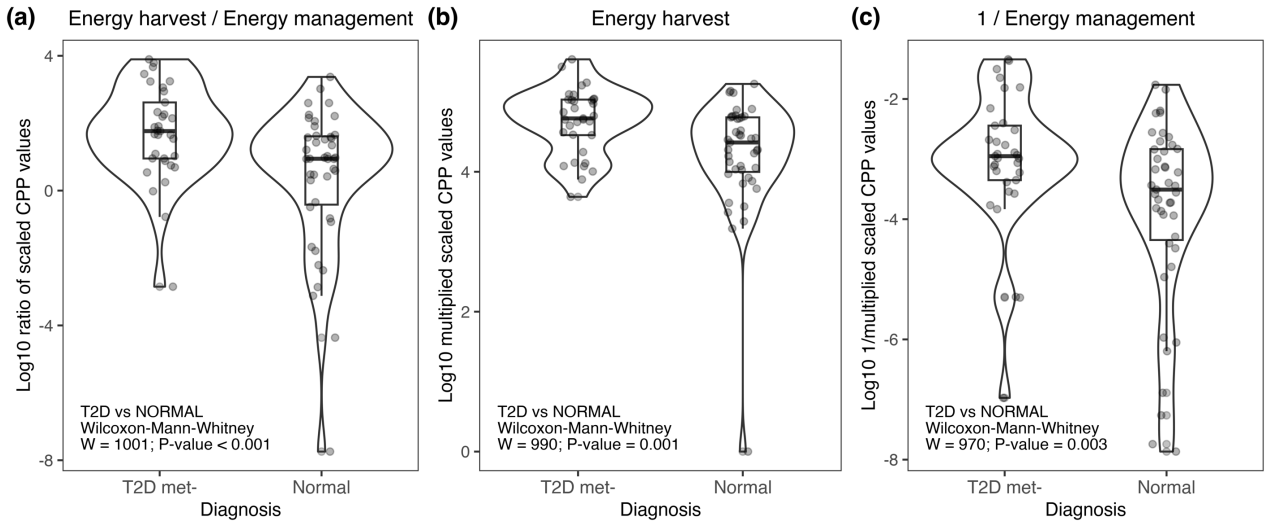


Figure 7. Preliminary T2D index values in subjects with untreated T2D and normal health, based on potential metabolism (CPP values) associated with (a) energy harvest and energy management, (b) energy harvest alone, and (c) energy management alone.

Community-scale profiling of metagenomic capacity using functions vs. CPP

Beta diversity ordinations (Figure 8) show that soil samples from different revegetation ages readily separated into groups based on the community scale function and CPP compositions. PCoA ordination axes provided greater explanation of overall compositional variation following conversion from functions to CPP measures (PCo1 + PCo2 in functions: 49.6 + 18.8 = 68.4% vs. CPP: 61.5 + 17.9 = 79.4%) and improved the explanatory value of revegetation age as a key variable affecting compositional shifts ($\text{PERMANOVA}_{\text{soil functions}}$: Df = 4, R^2 = 0.67, F = 5.08, P = 0.002; Beta $\text{dispersion}_{\text{soil functions}}$: F = 0.34, P = 0.838; $\text{PERMANOVA}_{\text{soil CPP}}$: Df = 4, R^2 = 0.75, F = 7.87, P = 0.001; Beta $\text{dispersion}_{\text{soil CPP}}$: F = 0.36, P = 0.842; Figure 1a-b). These results show that revegetation age explained 75% of the compositional variation when using CPP measures spanning 8,370 compounds, compared to 67% using 30,125 functions.

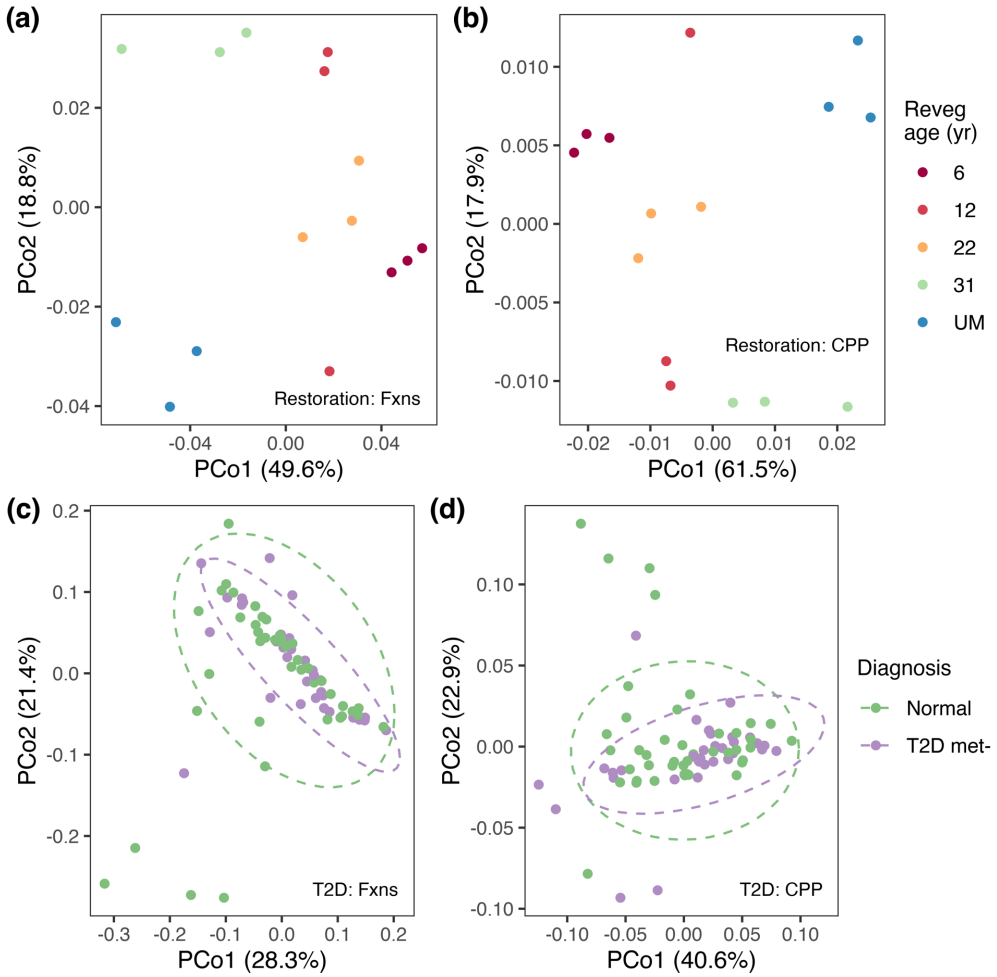


Figure 8. Visualisation of functional and CPP compositional differences in soils under ecosystem restoration (a-b) and gut microbiomes in T2D versus normal (c-d), as measured using function (/biochemical pathway) potential (a, c) and compound processing potential (b, d). Statistical information is provided in the text and Table 1. T2D met- denotes subjects have not received metformin treatment.

When the T2D samples were evaluated at the entire community scale, neither functions nor CPP data were able to discriminate between T2D and normal healthy subjects ($PERMANOVA_{T2D\ functions}$: Df = 1, R^2 = 0.02, F = 1.55, P = 0.136; Beta dispersion: F = 0.16, P = 0.712; $PERMANOVA_{T2D}$

383 $_{CPP}$: Df = 1, R^2 = 0.02, F = 1.49, *P* = 0.174; Beta dispersion: F = 9 x 10^{-4}, *P* = 0.979). PCoA ordination
384 axes using CPP data increased the scope to visualize compositional variation compared to functions in
385 the T2D samples (PCo1 + PCo2 in functions: 28.3 + 21.4 = 49.7 % using 17,962 functions vs. CPP:
386 40.6 + 22.9 = 63.5 % using 7031 compounds; Figure 1c-d). While these entire community PCoA
387 analyses cannot discriminate between T2D and normal subjects, the increased explanatory value
388 demonstrated in soils provides support for the translation from functions to using CPP values,
389 consistent with the notion that compound-level data represent a more fundamental basis for analyses.
390

391 **Discussion**

392 In this study we found several meaningful overlapping trends in the potential metabolism of soil
393 microbiomes within disturbed ecosystems and gut microbiomes in T2D. These patterns may have
394 profound implications for potential linkages between the interlinked, widespread and increasing
395 processes of global change, biodiversity loss, ecosystem degradation and soil microbiome disturbance –
396 because soil microbiomes provide a key reservoir and source for priming and resupplying the human
397 gut microbiome. We found that these microbiomes in the soil and human gut, and their associated
398 functional / metabolic capacities shifted strongly with ecosystem quality and T2D versus health, when
399 measured in terms of the potential metabolism of key compounds linked to energy harvesting/
400 availability and energy management (discussed below).
401 Using a novel method we translated metagenome functional capacities (i.e., function potential
402 relative abundances %) to compound processing potential (%), reflecting potential metabolism at the
403 scale of individual compounds. This compound-oriented approach allowed us to access a wealth of
404 published literature that associates chemical compounds with health, biological processes and
405 biochemical pathways in soils and gut microbiome-mediated human health. T2D is in essence a
406 metabolic disease that appears to be driven by a breakdown in complex bioenergetic processes (Daisley

407 et al. 2021). Biochemical processes relevant in soil microbiomes and gut microbiome-host interactions,
408 are also expected to be shaped by bioenergetic gradients and (bio)chemical relationships, and our
409 mapping of compounds into the bioenergetically informed O:C, H:C and N:C coordinate space helped
410 to reveal apparently meaningful patterns of simultaneously trending compounds with soil ecosystem
411 quality and T2D versus health.
412
413 ***Compounds linked to energy harvesting and availability in soils and T2D***
414 We found a number of compounds linked to energy harvesting and availability that increased in T2D,
415 where some also increased in disturbed soils, supporting the hypothesis that environmental microbial
416 exposures from disturbed ecosystems may excessively supplement and reinforce gut microbiome
417 functional capacities and biochemical pathways that exacerbate metabolic imbalances underlying T2D.
418 Potential metabolism of several sugars (see group B in Table 2) and ammonia increased in both
419 disturbed soils and T2D. Sugars and ammonia represent sources of energy. Venous ammonia levels
420 have been associated with diabetes through potential mechanisms including autonomic dysfunction,
421 extended gastrointestinal transit times, bacterial overgrowth and increased protein catabolism (Balcar et
422 al. 2023). CPP values for the group B sugars and ammonia identified through earlier trend analyses
423 were included into the 'energy harvest' sub-index calculation described earlier. The remaining
424 component used in the energy harvest sub-index comprised the sum of CPP % values for starch,
425 amylopectin and amylose. These compounds did not trend with soil ecosystem condition, however it is
426 reasonable to consider that functional capacity to process these compounds is likely to be amplified
427 through regular dietary choices over a sufficient period of time, perhaps even including dietary choices
428 within prior generations that create a degree of heritability in gut microbiomes (Sonnenburg et al.
429 2016).
430

431 ***Compounds linked to energy management in soils and T2D***

432 We found potential metabolism for several different types of compounds with links to energy
433 management (e.g., energy storage via fatty acids and lipids, gut hormone secretion, appetite regulation)
434 that decreased in both disturbed soils and in T2D. This finding supports a second hypothesis that
435 exposures to disturbed soil ecosystems that are deficient in potential metabolism (CPP) of compounds
436 involved in energy management may supplement and shape gut microbiomes that are also deficient in
437 this way, thereby contributing to further metabolic imbalances found in T2D. Specifically, the
438 compounds of interest for energy management include acyl carrier proteins involved in fatty acid
439 biosynthesis (group A in Table S2), aromatics (group C in Table S2), fiber and Trimethylamine N-
440 oxide (TMAO), as discussed below.

441 The 35 acyl carrier proteins (see Supplemental Material), for which potential metabolism was
442 deficient in disturbed soils and T2D, were linked to three enzymes involved in fatty acid biosynthesis
443 (EC 5.3.3.14, EC 1.3.1.9, EC 1.3.1.10). One enzyme in particular (EC 5.3.3.14) is involved in energy
444 storage via forming unsaturated fatty acids such as palmitoleic acid and vaccenic acid
445 (https://www.genome.jp/entry/5.3.3.14). Alternatively, energy storage as the saturated fatty acid,
446 palmitic acid, has been linked to increasing risk of T2D through decreasing β-cell function (Taylor
447 2024). Higher ratios of palmitic acid to palmitoleic acid are also associated with a higher risk of
448 diabetes (Qureshi et al. 2019), and this ratio would likely increase if capacity to produce unsaturated
449 palmitoleic acid is decreased.

450 The aromatic compounds for which potential metabolism was deficient in disturbed soils and
451 T2D were involved in many functional pathways including catabolism of aromatic amine compounds,
452 ubiquinone biosynthesis, omega-amidase enzymatic activity, oxidative stress responses, fermentation,
453 and synthesis of acetone, butanol and ethanol. Ubiquinone is a redox active lipid produced across all
454 domains of life that functions in electron transport and oxidative phosphorylation and whose deficiency

455 causes human diseases (Stefely and Pagliarini 2017). Omega-amidase is an enzyme involved in L-
456 glutamine and L-asparagine metabolism which comprise important components in the wider
457 metabolism of mammals (Cooper et al. 2016). In particular, glutamine is believed to play an important
458 role in stimulating regulatory gut hormones such as glucagon-like peptide 1 (GLP1), to increase insulin
459 levels and reduce blood sugar (Jafari-Vayghan et al. 2020). One of the compounds was baicalein, a
460 dietary flavonoid which has recently been found to enhance the insulin-sensitizing effect of metformin
461 to prevent type 2 diabetes via the regulation of lipid metabolism and gut microenvironment (Xing et al.
462 2024).

463 Dietary fiber is known to decrease plasma ghrelin (hunger hormone) and increase other key gut
464 regulatory hormones: cholecystokinin (CCK), glucagon-like-peptide-1 (GLP-1), and peptide YY
465 (PYY) compared to energy matched low fiber control diets (Bulsiewicz 2023). Short chain fatty acids
466 (SCFAs) are produced when certain members of the gut microbiome ferment non-digestible fibers, and
467 these SCFAs bind to receptors in the colon leading to the production of a range of key hormones that
468 regulate food intake, body weight, and energy metabolism (Bulsiewicz 2023). To represent fiber we
469 included both lignin and pectin which had been detected in the datasets and are among the recognized
470 key types of dietary fibers (Blanco-Pérez et al. 2021). Potential metabolism (CPP values) for lignin
471 decreased with soil disturbance and in T2D gut microbiomes. Potential metabolism for pectin also
472 decreased in T2D, but not in disturbed soils. We also searched for oligosaccharides (e.g., fructan,
473 inulin, galactan; Blanco-Pérez et al. 2021) but could not detect these compounds in our datasets.

474 Trimethylamine N-oxide (TMAO) was another compound with deficient potential metabolism
475 in the T2D gut and disturbed soils. TMAO has reported beneficial effects including potential protection
476 from hyperammonemia and glutamate neurotoxicity, alleviation of endoplasmic reticulum stress, and
477 improved glucose homeostasis by stimulating insulin secretion by pancreatic cells (Hoyles et al. 2018).
478 However, with increasing levels TMAO is also positively associated with metabolic diseases such as

479 T2D (Daisley et al. 2021). Such contradictory reports may align with generalizable u-shaped dose-
480 health response relationships (i.e., deficiency-sufficiency-toxicity) seen in many biological systems
481 (Calabrese et al. 2007; Figure S1). Certain key members of the gut microbiome may help to detoxify
482 and thereby manage levels of TMAO, via reducing it in the gut to trimethylamine (TMA) followed by
483 host uptake of TMA from the gut into the bloodstream, then conversion back to TMAO by host
484 enzymes in the liver, through a process termed metabolic retroconversion (Hoyles et al. 2018).
485

486 ***Limitations and future work***

487 Our study examined separate model case study metagenome datasets and did not consider variables
488 associated with microbiome-host interactions. We did not test environmental exposures or microbiome
489 transfer from soils to the human gut (e.g., Roslund et al. 2021). Some compounds of interest were not
490 included in the case study datasets, or sometimes recorded using synonyms or closely chemically
491 related species. Alternative naming for some chemical species may have introduced minor uncertainties
492 in our analyses. We analyzed only a portion of the compounds present, featuring those that we pre-
493 selected (Table S1), and example groupings from the compound-wide trend analysis that featured
494 conspicuously in our mapping into the O:C, H:C, N:C bioenergetic coordinate space. However, a much
495 broader spectrum of ~7,000 to >8,000 compounds are potentially metabolized by the gut and soil
496 microbiomes considered here. Notwithstanding, our analysis provides strong support for the notion that
497 differences in potential metabolism of key compounds by microbiomes do contribute to T2D
498 (complementing previous work by Forslund et al. 2015), and we show several overlapping patterns of
499 this differential potential metabolism (CPP trends) in disturbed soils. Our study provides evidence
500 supporting the hypothetical linkage from ecosystem degradation and disturbed soil microbiomes, via
501 environmental exposures, to influence metabolic anomalies encoded in T2D gut metagenomes. The
502 widespread scale of global land degradation (causing rising loss of biodiversity, soil and ecosystem

503 functions, recently estimated to impact 75% of the Earth's land surface; IPBES 2018) seems to align
504 with the rising pandemic scale of T2D (Ong et al. 2023; Sun et al. 2022). However, further evidence of
505 these hypothetical links are required through controlled experimentation.
506

507 **Conclusions**

508 We found meaningful overlapping trends in the potential metabolism of key compounds involved in
509 energy harvesting and energy management (storage, regulation) within model case study metagenomes
510 from soils under disturbed ecosystems and in T2D gut microbiomes. Our results support existing
511 studies implicating, at least partially, a role for microbiomes in T2D. Additionally, our findings also
512 support the conceptual involvement of soil microbiomes from disturbed environments at least partially
513 contributing to the metabolic (energy harvesting and management) imbalance found in T2D. Our novel
514 method used to examine microbiome potential metabolism at the resolution of individual compounds
515 has derived new insights to potential links between soil and T2D gut health that warrant further
516 investigation.
517
518

Data availability

Code used to support this study is available from: https://github.com/liddic/cpp3d

Acknowledgments

We acknowledge the authors of previously published metagenomics datasets used here which enabled comparison of samples spanning human and environmental health. We acknowledge First Nations Peoples and traditional custodians of the lands from which the case study datasets originate. This work was supported by the People, Cities and Nature research program (Aotearoa New Zealand Ministry of Business, Innovation and Employment, grant UOWX2101).

531 **References**

532

533 Aliluev A, Tritschler S, Sterr M, Oppenländer L, Hinterdobler J, Greisle T, et al. 2021. Diet-induced
534 alteration of intestinal stem cell function underlies obesity and prediabetes in mice. Nature Metabolism
535 3:1202-1216.

536 Anthony MA, Bender SF, van der Heijden MGA. 2023. Enumerating soil biodiversity. Proceedings of
537 the National Academy of Sciences 120:e2304663120.

538 Attard SM, Herring AH, Mayer-Davis EJ, Popkin BM, Meigs JB, Gordon-Larsen P. 2012. Multilevel
539 examination of diabetes in modernising china: What elements of urbanisation are most associated with
540 diabetes? Diabetologia 55:3182-3192.

541 Balcar L, Krawanja J, Scheiner B, Paternostro R, Simbrunner B, Semmler G, et al. 2023. Impact of
542 ammonia levels on outcome in clinically stable outpatients with advanced chronic liver disease. JHEP
543 Reports 5:100682.

544 Beulens JWJ, Pinho MGM, Abreu TC, den Braver NR, Lam TM, Huss A, et al. 2022. Environmental
545 risk factors of type 2 diabetes—an exposome approach. Diabetologia 65:263-274.

546 Blanco-Pérez F, Steigerwald H, Schülke S, Vieths S, Toda M, Scheurer S. 2021. The dietary fiber
547 pectin: Health benefits and potential for the treatment of allergies by modulation of gut microbiota.
548 Current Allergy and Asthma Reports 21:43.

549 Blot F, Marchix J, Ejarque M, Jimenez S, Meunier A, Keime C, et al. 2023. Gut microbiota remodeling
550 and intestinal adaptation to lipid malabsorption after enteroendocrine cell loss in adult mice. Cellular
551 and Molecular Gastroenterology and Hepatology 15:1443-1461.

552 Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. 2016. Culturing of
553 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature 533:543-546.

554 Bulsiewicz WJ. 2023. The importance of dietary fiber for metabolic health. American Journal of
555 Lifestyle Medicine 17:639-648.

556 Calabrese EJ, Bachmann KA, Bailer AJ, Bolger PM, Borak J, Cai L, et al. 2007. Biological stress
557 response terminology: Integrating the concepts of adaptive response and preconditioning stress within a
558 hormetic dose–response framework. Toxicology and Applied Pharmacology 222:122-128.

559 Cheema A, Adeloye D, Sidhu S, Sridhar D, Chan KY. 2014. Urbanization and prevalence of type 2
560 diabetes in southern asia: A systematic analysis. J Glob Health 4:010404.

561 Cooper AJL, Shurubor YI, Dorai T, Pinto JT, Isakova EP, Deryabina YI, et al. 2016. Ω-amidase: An
562 underappreciated, but important enzyme in l-glutamine and l-asparagine metabolism; relevance to
563 sulfur and nitrogen metabolism, tumor biology and hyperammonemic diseases. Amino Acids 48:1-20.

564 Cunningham AL, Stephens JW, Harris DA. 2021. Gut microbiota influence in type 2 diabetes mellitus
565 (t2dm). Gut Pathogens 13:50.

566 Daisley BA, Koenig D, Engelbrecht K, Doney L, Hards K, Al KF, et al. 2021. Emerging connections
567 between gut microbiome bioenergetics and chronic metabolic diseases. Cell Reports 37:110087.

568 Delgado-Baquerizo M, Eldridge DJ, Liu Y-R, Sokoya B, Wang J-T, Hu H-W, et al. 2021. Global
569 homogenization of the structure and function in the soil microbiome of urban greenspaces. Science
570 Advances 7:eabg5809.

571 Flandroy L, Poutahidis T, Berg G, Clarke G, Dao M-C, Decaestecker E, et al. 2018. The impact of
572 human activities and lifestyles on the interlinked microbiota and health of humans and of ecosystems.
573 Science of The Total Environment 627:1018-1038.

574 Flies EJ, Mavoa S, Zosky GR, Mantzioris E, Williams C, Eri R, et al. 2019. Urban-associated diseases:
575 Candidate diseases, environmental risk factors, and a path forward. Environment International
576 133:105187.

577 Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, et al. 2015. Disentangling
578 type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528:262-266.

579 Frumkin H, Bratman GN, Breslow SJ, Cochran B, Kahn PH, Jr., Lawler JJ, et al. 2017. Nature contact
580 and human health: A research agenda. Environmental health perspectives 125:075001-075001.

581 Gassasse Z, Smith D, Finer S, Gallo V. 2017. Association between urbanisation and type 2 diabetes:
582 An ecological study. BMJ Global Health 2:e000473.

583 Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. 2018. Current understanding of
584 the human microbiome. Nature Medicine 24:392.

585 Gonzalez LL, Garrie K, Turner MD. 2018. Type 2 diabetes – an autoinflammatory disease driven by
586 metabolic stress. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1864:3805-3823.

587 Gurung M, Li Z, You H, Rodrigues R, Jump DB, Morgun A, et al. 2020. Role of gut microbiota in type
588 2 diabetes pathophysiology. eBioMedicine 51.

589 Hoyles L, Jiménez-Pranteda ML, Chilloux J, Brial F, Myridakis A, Aranias T, et al. 2018. Metabolic
590 retroconversion of trimethylamine n-oxide and the gut microbiota. Microbiome 6:73.

591 IPBES. 2018. The ipbes assessment report on land degradation and restoration. Bonn, Germany.

592 Jafari-Vayghan H, Varshosaz P, Hajizadeh-Sharafabad F, Razmi HR, Amirpour M, Tavakoli-
593 Rouzbehani OM, et al. 2020. A comprehensive insight into the effect of glutamine supplementation on
594 metabolic variables in diabetes mellitus: A systematic review. Nutrition & Metabolism 17:80.

595 Jardon KM, Canfora EE, Goossens GH, Blaak EE. 2022. Dietary macronutrients and the gut
596 microbiome: A precision nutrition approach to improve cardiometabolic health. Gut 71:1214.

597 Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. 2013. Gut
598 metagenome in european women with normal, impaired and diabetic glucose control. Nature 498:99-
599 103.

600 Krajmalnik-Brown R, Ilhan Z-E, Kang D-W, DiBaise JK. 2012. Effects of gut microbes on nutrient
601 absorption and energy regulation. Nutrition in Clinical Practice 27:201-214.

602 Liddicoat C, Sydnor H, Cando-Dumancela C, Dresken R, Liu J, Gellie NJC, et al. 2020. Naturally-
603 diverse airborne environmental microbial exposures modulate the gut microbiome and may provide
604 anxiolytic benefits in mice. Science of The Total Environment 701:134684.

605 Liddicoat C, Krauss SL, Bissett A, Borrett RJ, Ducki LC, Peddle SD, et al. 2022. Next generation
606 restoration metrics: Using soil edna bacterial community data to measure trajectories towards
607 rehabilitation targets. Journal of Environmental Management 310:114748.

608 Liddicoat C, Edwards RA, Roach M, Robinson JM, Wallace KJ, Barnes AD, et al. 2024. Bioenergetic
609 mapping of 'healthy microbiomes' via compound processing potential imprinted in gut and soil
610 metagenomes. Science of The Total Environment 940:173543.

611 Liu W, Sun Z, Ma C, Zhang J, Ma C, Zhao Y, et al. 2021. Exposure to soil environments during earlier
612 life stages is distinguishable in the gut microbiome of adult mice. Gut Microbes 13:1-13.

613 Lu X, Xie Q, Pan X, Zhang R, Zhang X, Peng G, et al. 2024. Type 2 diabetes mellitus in adults:
614 Pathogenesis, prevention and therapy. Signal Transduction and Targeted Therapy 9:262.

615 Mhuireach G, Johnson BR, Altrichter AE, Ladau J, Meadow JF, Pollard KS, et al. 2016. Urban
616 greenness influences airborne bacterial community composition. Science of The Total Environment
617 571:680-687.

618 Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. 2020. Vegan:
619 Community ecology package. R package version 2.5-7.

620 Ong KL, Stafford LK, McLaughlin SA, Boyko EJ, Vollset SE, Smith AE, et al. 2023. Global, regional,
621 and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: A
622 systematic analysis for the global burden of disease study 2021. The Lancet 402:203-234.

623 Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. 2013. The seed and the rapid
624 annotation of microbial genomes using subsystems technology (rast). Nucleic Acids Research
625 42:D206-D214.

626 Pearson ER. 2019. Type 2 diabetes: A multifaceted disease. Diabetologia 62:1107-1112.

627 Qureshi W, Santaren ID, Hanley AJ, Watkins SM, Lorenzo C, Wagenknecht LE. 2019. Risk of
628 diabetes associated with fatty acids in the de novo lipogenesis pathway is independent of insulin
629 sensitivity and response: The insulin resistance atherosclerosis study (iras). BMJ Open Diabetes
630 Research & Care 7:e000691.

631 Rivas-Ubach A, Liu Y, Bianchi TS, Tolić N, Jansson C, Paša-Tolić L. 2018. Moving beyond the van
632 krevelen diagram: A new stoichiometric approach for compound classification in organisms. Analytical
633 Chemistry 90:6152-6160.

634 Robinson JM, Cando-Dumancela C, Antwis RE, Cameron R, Liddicoat C, Poudel R, et al. 2021.
635 Exposure to airborne bacteria depends upon vertical stratification and vegetation complexity. Scientific
636 Reports 11:9516.

637 Roslund MI, Puhakka R, Nurminen N, Oikarinen S, Siter N, Grönroos M, et al. 2021. Long-term
638 biodiversity intervention shapes health-associated commensal microbiota among urban day-care
639 children. Environment International 157:106811.

640 Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. 2018. Environment
641 dominates over host genetics in shaping human gut microbiota. Nature.

642 Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, et al. 2018. Gut microbiota functions:
643 Metabolism of nutrients and other food components. European Journal of Nutrition 57:1-24.

644 Ruiz-Alejos A, Carrillo-Larco RM, Miranda JJ, Anderson CAM, Gilman RH, Smeeth L, et al. 2018.
645 Addressing the impact of urban exposure on the incidence of type 2 diabetes mellitus: The peru migrant
646 study. Scientific Reports 8:5512.

647 Silva GGZ, Green KT, Dutilh BE, Edwards RA. 2015. Super-focus: A tool for agile functional analysis
648 of shotgun metagenomic data. Bioinformatics 32:354-361.

649 Six J, Frey SD, Thiet RK, Batten KM. 2006. Bacterial and fungal contributions to carbon sequestration
650 in agroecosystems. Soil Science Society of America Journal 70:555-569.

651 Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL. 2016. Diet-
652 induced extinctions in the gut microbiota compound over generations. Nature 529:212-215.

653 Sonnenburg ED, Sonnenburg JL. 2019. The ancestral and industrialized gut microbiota and
654 implications for human health. Nature Reviews Microbiology 17:383-390.

655 Stefely JA, Pagliarini DJ. 2017. Biochemistry of mitochondrial coenzyme q biosynthesis. Trends in
656 Biochemical Sciences 42:824-843.

657 Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. 2022. Idf diabetes atlas:
658 Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045.
659 Diabetes Res Clin Pract 183:109119.

660 Sun S, Badgley BD. 2019. Changes in microbial functional genes within the soil metagenome during
661 forest ecosystem restoration. Soil Biology and Biochemistry 135:163-172.

662 Sun X, Liddicoat C, Tiunov A, Wang B, Zhang Y, Lu C, et al. 2023. Harnessing soil biodiversity to
663 promote human health in cities. npj Urban Sustainability 3:5.

664 Taylor R. 2024. Understanding the cause of type 2 diabetes. The Lancet Diabetes & Endocrinology
665 12:664-673.

666 Van Hul M, Cani PD, Petitfils C, De Vos WM, Tilg H, El-Omar EM. 2024. What defines a healthy gut
667 microbiome? Gut 73:1893.

668 Wu X, Wu L, Liu Y, Zhang P, Li Q, Zhou J, et al. 2018. Microbial interactions with dissolved organic
669 matter drive carbon dynamics and community succession. Frontiers in Microbiology 9.

670 Xing Y, Liu H, Zhang D, Jiang H, Xiu Z, Dong Y. 2024. Baicalein, a dietary flavonoid, enhances the
671 insulin-sensitizing effect of metformin to prevent type 2 diabetes via the regulation of lipid metabolism
672 and gut microenvironment. Food Frontiers 5:668-690.

673 Ye L, Mueller O, Bagwell J, Bagnat M, Liddle RA, Rawls JF. 2019. High fat diet induces microbiota-
674 dependent silencing of enteroendocrine cells. eLife 8:e48479.

675 Zhou G-L, Xie Y, Zhang Y-D, Wang L, Xie Y-T, Qiu H-L, et al. 2024. Exposure to greenspaces
676 sourced soils improves mice gut microbiota. Environmental Pollution 362:124931.

677