Bioenergetic mapping of 'healthy microbiomes' via compound processing potential imprinted in gut and soil metagenomes

Craig Liddicoat[a,*], Robert A. Edwards[a], Michael Roach[a], Jake M. Robinson[a], Kiri Joy Wallace[b], Andrew D. Barnes[b], Joel Brame[a], Anna Heintz-Buschart[c], Timothy R. Cavagnaro[a], Elizabeth A. Dinsdale[a], Michael P. Doane[a], Nico Eisenhauer[d,e], Grace Mitchell[b,f], Bibishan Rai[b], Sunita Ramesh[a], Martin F. Breed[a]

[a]College of Science and Engineering, Flinders University, Bedford Park, South Australia, Australia
[b]Environmental Research Institute, University of Waikato, Hamilton, Aotearoa New Zealand
[c]Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands
[d]German Centre for Integrative Biodiversity Research (iDiv), 04103 Leipzig, Germany
[e]Institute of Biology, Leipzig University, 04103 Leipzig, Germany
[f]Manaaki Whenua – Landcare Research, Hamilton, Aotearoa New Zealand

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

ORCIDs: CL: 0000-0002-4812-7524, RAE: 0000-0001-8383-8949, MR: 0000-0003-1488-5148, JMR: 0000-0001-8108-3271, KJW: 0000-0003-1236-6846, ADB: 0000-0002-6499-381X, JB: 0000-0002-1852-9987, AHB: 0000-0002-9780-1933, TRC: 0000-0002-9922-5677, EAD: 0000-0002-2177-203X, MPD: 0000-0001-9820-2193, NE: 0000-0002-0371-6720, GM: 0000-0002-2839-7981, SR: 0000-0003-2230-4737, MFB: 0000-0001-7810-9696

Author Contributions: CL, RAE, MFB designed research; CL, MR, JMR, KJW, ADB, AHB, MPD, NE, GM, BR performed research; CL, RAE, JMR, KJW, ADB, JB, AHB, TRC, EAD, MPD, NE, SR, MFB analyzed data; CL drafted the manuscript; all authors reviewed and edited the final version.

Preprint Server: bioRxiv.org (ID: BIORXIV/2023/565728) [CC BY-NC-ND]

Competing Interest Statement: The authors declare no competing interest.

Keywords: Metagenomics, compound processing potential, healthy microbiome, human health, bioenergetic mapping, van Krevelen diagram

ABSTRACT

Despite mounting evidence of their importance in human health and ecosystem functioning, the definition and measurement of 'healthy microbiomes' remain unclear. More advanced knowledge exists on health associations for compounds used or produced by microbes. Environmental microbiome exposures (especially via soils) also help shape, and may supplement, the functional capacity of human microbiomes. Given the synchronous interaction between microbes, their feedstocks, and micro-environments, with functional genes facilitating chemical transformations, there exists an intriguing opportunity to examine microbiomes in terms of their capacity to process compounds relevant to human health. Here we integrate functional genomics and biochemistry frameworks to derive new quantitative measures of the potential for human gut and environmental soil metagenomes to process major compound classes (e.g., lipids, carbohydrates) and selected biomolecules (e.g., vitamins, short-chain fatty acids) linked to human health. Metagenome functional potential profile data were translated into a universal compound mapping 'landscape' based on bioenergetic van Krevelen mapping of function-level meta-compounds and corresponding functional relative abundances, reflecting imprinted genetic capacity of microbiomes to metabolize an array of different compounds. We show that measures of 'compound processing potential' associated with human health and disease (examining atherosclerotic cardiovascular disease, colorectal cancer, type 2 diabetes and anxious-depressive behavior case studies), and displayed seemingly predictable shifts along gradients of ecological disturbance in plant-soil ecosystems (three case studies). Ecosystem quality explained 60–92% of variation in soil metagenome compound processing potential measures in a post-mining restoration case study dataset. With growing knowledge of the varying proficiency of environmental microbiota to process human health associated compounds, we might design environmental interventions or nature prescriptions to modulate our exposures, thereby advancing microbiota-oriented approaches to human health. Compound processing potential offers a simplified, integrative approach for applying metagenomics in ongoing efforts to understand and quantify the role and linkages of microbiota in environmental- and human-health.

1. Introduction

Microbial communities (microbiota), their feedstocks (substrates, nutrients) and environmental conditions (e.g., pH, redox potential, temperature, moisture, salinity) work in concert to drive microbially-mediated reactions essential to fueling life on Earth (Averill et al., 2022). Microbiomes (microbiota, genetic material and metabolites) are critical to the health and functioning of humans (Banerjee and van der Heijden, 2022; Gilbert et al., 2018; Round and Mazmanian, 2009) and ecosystems (Delgado-Baquerizo et al., 2020; Sokol et al., 2022). However, clear definitions of what constitutes a 'healthy microbiome' remain elusive (Eisenstein, 2020). More advanced knowledge exists on health associations for the compounds used or produced by microbes. Because microbes, their feedstocks and micro-environments interact synchronously—using functional genes to facilitate chemical transformations—this presents an intriguing opportunity to examine microbiomes through their potential to process compounds associated with human health.

Human microbiomes are supplemented and partly shaped by environmental microbial exposures (Banerjee and van der Heijden, 2022; Gilbert et al., 2018). Therefore, the involvement of environmental microbiomes in processing human health-associated compounds is also of interest. Transfer of environmental microbiota to humans may help supplement important functional capacity, protective microbiota, and immune-signaling agents, particularly in infants, but also in adults who have depleted microbiota due to antibiotic use, poor diet, lifestyle or other health incidents (Banerjee and van der Heijden, 2022; Brame et al., 2021; Flandroy et al., 2018; Roslund et al., 2022). We know that microbiota are shaped by the resources they utilize and the environments they inhabit (Fierer, 2017; Gilbert et al., 2018). If the functional composition of microbiota varies predictably along environmental gradients, then through design, management, and behavior we should be able to modulate our exposure to health-promoting versus disease-associated microbes. Soils, in particular, can represent a rich source of microbial diversity with potential to support human health (Banerjee and van der Heijden, 2022; Li et al.; Liddicoat et al., 2018; Liddicoat et al., 2020; Ottman et al., 2018; Sun et al., 2023). Microbiota in plant-soil systems are shaped by

84 macro-scale factors including climate, soil characteristics, vegetation composition, diversity, land use and
85 management (Delgado-Baquerizo et al., 2018; Fierer, 2017). With the prospect of cost-effectively
86 encouraging health-promoting microbes, it is frequently asked, "What type of environment is best?" Yet,
87 the attributes of health-promoting environmental microbiomes, including potential functional overlaps with
88 human microbiomes, remain understudied.

89 Microbes typically operate as a community (Barton and Northup, 2011) where many taxa may be
90 highly efficient at performing certain functions, but they lack functional capacity for stand-alone survival
91 (Zengler and Zaramela, 2018). Complex cross-feeding and resource sharing in the extracellular space
92 (Daisley et al., 2021), and involvement of multiple taxa in functional pathways, suggest that community-
93 scale functional profiles (rather than specific microbial taxa) underpin the health-supporting capacity of
94 microbiota. Indeed, it is this community-scale complexity that has hindered progress towards clear
95 definitions of a healthy microbiome. Nevertheless, researchers want to better understand the assembly and
96 structure of health-promoting microbiomes to improve the course of microbiome-associated diseases.
97 Disease-associated microbiota are often characterized by a loss of diversity and overgrowth by
98 opportunistic pathogens (Belizario and Napolitano, 2015; Round and Mazmanian, 2009), but it may be
99 unclear whether they are the facilitators or followers of disease.

100 Many microbiome-associated diseases are linked to bioenergetic mechanisms (Daisley et al., 2021),
101 with oxidation-reduction (redox) potential recognized as a key factor shaping microbial communities. The
102 healthy anaerobic gut favors obligate anaerobes, whereas dysbiosis is often accompanied by increased
103 oxygenation of the colonic epithelium and expanded abundance of oxygen-tolerant facultative anaerobic
104 bacteria (Litvak et al., 2018; Rivera-Chávez et al., 2017). Oxygen is a highly electronegative element
105 important in shaping electrochemical gradients, biochemical reactions, and gene expression (Wilson et al.,
106 2020). Oxygen content varies in different types of organic matter (i.e., microbial feedstocks). Yet, the
107 interrelationship between bioenergetic (electrochemical) status of compounds, microbial environments,
108 microbiota development and human health appears to receive insufficient attention. In soils, redox potential

109 varies with weather, vegetation, land use, management, drainage, organic-content, vicinity to roots, soil
110 characteristics, and microbial activity (Hinsinger et al., 2009; Husson, 2013). At the molecular level, redox
111 potential shapes what kind of molecules can be made and how energy is stored. Therefore, a compound-
112 oriented examination of healthy microbiomes might capitalize on available knowledge linking compounds
113 with human health, while also considering deterministic influences of bioenergetic (electrochemical)
114 context.

115 Here, we examine the functional potential of gut and soil microbiota from a compound processing
116 potential (CPP) viewpoint, to assess patterns in human health and disease, and with gradients of ecosystem
117 maturity (or quality). Such an approach might discern health- versus disease-promoting microbiotas from
118 the types of compounds they are attuned to consuming or producing (reflecting microbial feedstocks and
119 metabolites respectively). Previous metabolome prediction frameworks (e.g., Diener et al., 2020; Garza et
120 al., 2018; Mallick et al., 2019; Neveu et al., 2023) rely variously on supplementary metabolome training
121 datasets, microorganism-specific genome-scale metabolic models, taxonomic abundance estimates, and
122 assumed environments (e.g., human gut), to model short-term reaction rates. Here, we wanted to exploit
123 community-scale compound-oriented information that might be embedded within metagenomes, regardless
124 of taxa present. As functional potential profiling from whole genome sequencing (or shotgun
125 metagenomics) does not directly measure functions performed, we characterized microbiota-linked CPP
126 from DNA sequencing, without direct measurement of compounds. Our conceptual interest was to
127 characterize the distribution of available community-scale genetic capacity of microbiomes to process an
128 array of compound classes and specific biomolecules (compounds produced by living organisms), which
129 may be present over a period of time and with variation in feedstocks and environmental conditions,
130 consistent with dynamic influences that have shaped existing microbiota structure. We premised that the
131 ease of transformation between health-associated compounds and other compounds that closely resemble
132 them will depend on stoichiometric and energetic similarities, microbiota functional diversity, and
133 environmental conditions.

134 We utilized a framework that integrates information about compounds, bioenergetics, and
135 environmental conditions. The complexity-reducing van Krevelen (vK) coordinate space offers a simplified
136 and intuitive bioenergetic framework for approximate mapping of compounds based on their carbon (C),
137 oxygen (O) and hydrogen (H) content, while also reflecting energy density and principal axes that explain
138 microbial distribution and chemical speciation (Fig. 1 and Fig. S1) (D'Andrilli et al., 2015; Wu et al., 2018).
139 Compounds are mapped into vK space using their O:C and H:C molar ratios (x- and y-axis respectively).
140 We surmised this framework could offer an exhaustive and intuitive mapping space to summarize the nature
141 of microbiota-mediated functional reactions in a way that reflects the collective mean or dominant
142 compound properties, reaction stoichiometries, and potential overlaps between dietary or environmental
143 substrates, and key health-associated compounds.
144 Specifically, we combined SUPER-FOCUS functional profiling annotations (Silva et al., 2015), the
145 comprehensive ModelSEED (Seaver et al., 2020) functional-biochemistry database system and vK
146 coordinate mapping to assign functional potential relative abundances from human gut and soil sample
147 metagenomes to overall mean function-level vK coordinates. This approach effectively mapped every
148 SUPER-FOCUS function (where feasible via available corresponding database information), based on one
149 or more associated chemical reactions, to a function-level 'meta-compound' represented in the two-
150 dimensional vK space (Fig. S2, Methods). We aimed to: (1) investigate measures of microbiota CPP
151 imprinted in human gut and soil metagenomes; (2) test for differences in human health and disease, and in
152 disturbed, restored and natural ecosystems; and (3) identify compounds linked to human health with
153 differing potential metabolism (i.e., CPP) by environmental microbiota due to differences in ecosystem
154 quality. We hypothesized this universal bioenergetic compound mapping approach might identify CPP
155 profiles, and possible overlaps in human and environmental metagenomes, that could inform the definition
156 and future shaping of 'healthy microbiomes'. We were also keen to explore whether CPP measures might
157 enhance the interpretability and accessibility of metagenomics data to aid hypothesis building and
158 prioritizing future research.

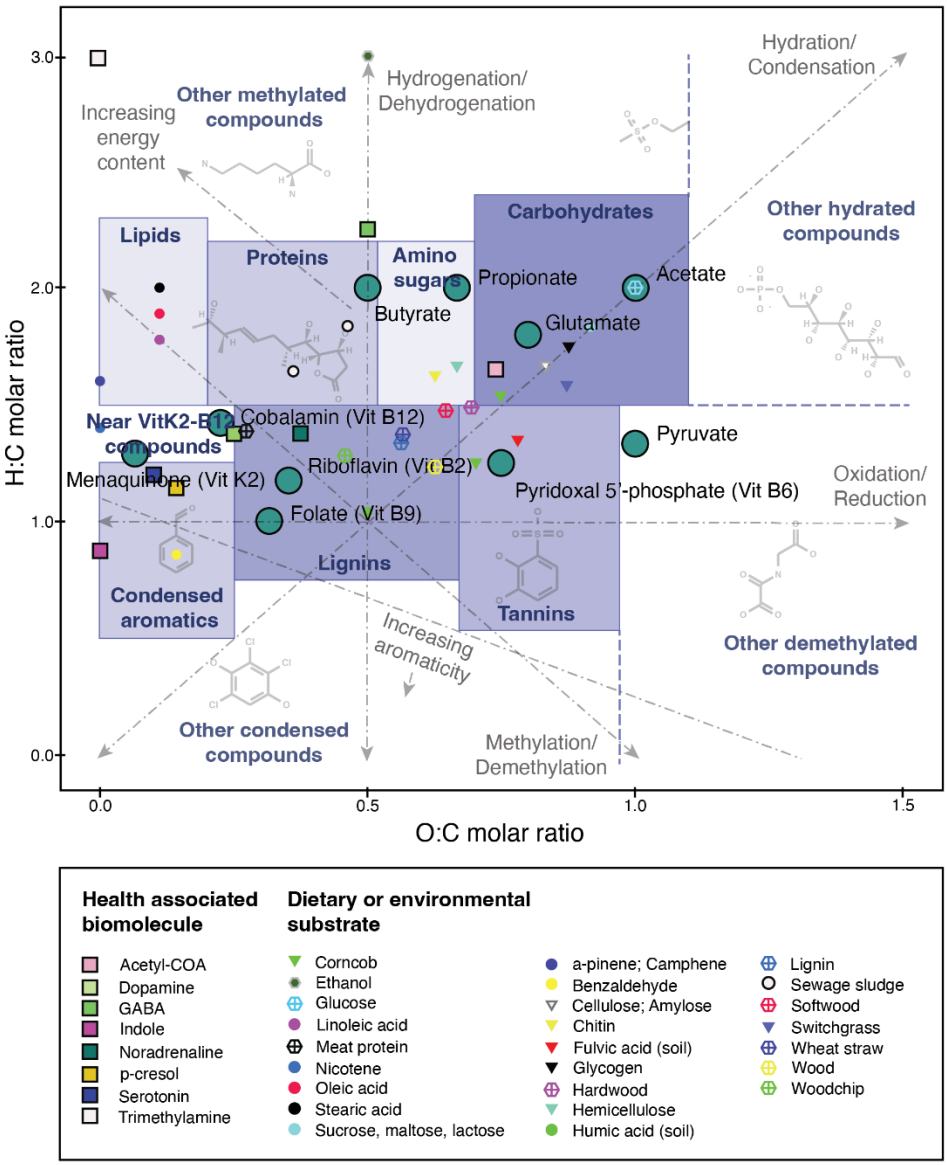

Fig. 1. Van Krevelen (vK) coordinate space (adapted from D'Andrilli et al., 2015), displaying major compound classes (purple zones and text), key gradients (grey axes and text), focus biomolecules examined in this study (large dots), and additional example health-associated biomolecules, dietary or environmental substrates (legend) (see Tables S1, S2). vK zones were adapted from (Wu et al., 2018). Key gradients include oxidation-reduction (x-axis), hydrogenation-dehydrogenation (y-axis), hydration-condensation (top-right to bottom-left), methylation-demethylation (top-left to bottom-right) and increasing energy content (towards top-left).

2. Materials and Methods

2.1. Case study datasets

We assessed metagenomics samples from four human health and disease datasets, comprising atherosclerotic cardiovascular disease (ACVD) (Jie et al., 2017), colorectal cancer (Zeller et al., 2014), type 2 diabetes (T2D) (Forslund et al., 2015) and problem (anxious-depressive) behaviors in children (Flannery et al., 2020); and three environmental soil datasets from ecological restoration and disturbed versus natural plant-soil systems (Barnes et al., 2020; Bissett et al., 2016; Sun and Badgley, 2019) (Table 1; further detail in Appendix A Supporting Information).

Table 1. Description of case study metagenome datasets.

Case study focus	**Main comparison variable or diagnosis groups (and sample/subject numbers[a])**	**Metagenome functional profile characteristics (sample mean ± s.d.)[b]**	**Source data reference; Country of origin**
Human gut			
Atherosclerotic cardiovascular disease (ACVD)	ACVD (n_F = 53, n_M = 157) or normal healthy (n_F 101, n_M = 69). Total n = 380.	Functions^0 n = 21,117 (9301 ± 3523) Functions^{vK} n = 10,829 (5051 ± 1843) Total fxn rel abun^{vK} = 63.2 ± 3.5 % vK coordinates n = 2535 (1509 ± 502)	SRA accession PRJEB21528 (Jie et al., 2017); China
Colorectal cancer	Colorectal cancer (n_F = 24, n_M = 29) or normal healthy (n_F = 33, n_M = 27). Total n = 113.	Functions^0 n = 12,896 (2809 ± 1924) Functions^{vK} n = 6932 (1750 ± 1117) Total fxn rel abun^{vK} = 66.7 ± 2.7 % vK coordinates n = 1954 (711 ± 362)	SRA accession PRJEB6070 (Zeller et al., 2014); France
Type 2 diabetes (T2D) and impaired glucose tolerance (IGT)	T2D with no Metformin treatment (T2D Met-, n = 33), T2D with Metformin (T2D Met+, n = 20), IGT (n = 49), or normal healthy (n = 43). Subjects are females only. Total n = 145.	Functions^0 n = 19,099 (9916 ± 2393) Functions^{vK} n = 9916 (4490 ± 874) Total fxn rel abun^{vK} = 64.4 ± 2.2 % vK coordinates n = 2393 (1431 ± 202)	SRA accession PRJEB1786 (Forslund et al., 2015); Sweden
Problem behaviors in children	First principal component (PC1) of anxious-depressive problem behaviors, examined as either numeric scores (n_F =	Functions^0 n = 20,599 (10282 ± 2481) Functions^{vK} n = 10,322 (5507 ± 1224) Total fxn rel abun^{vK} = 54.9 ± 1.3 %	SRA accession PRJNA496479 (Flannery et al.,

	20, n_M = 17); or high/low PC1 groups[c] (high PC1 n_F = 10, n_M = 8; Low PC1 n_F = 10, n_M = 9). Total n = 37.	vK coordinates n = 2459 (1645 ± 281)	2020); United States
Soils			
People Cities and Nature (PCaN) urban forest ecosystem restoration	Soil samples spanned young to old revegetation age, and remnant sites (treated as ordinal variables). Data were separated into pH-based groups: strongly acidic (pH < 4.5, 10-40 yr old, remnant, n = 8); and acidic-neutral soils (4.5 < pH < 7, 11-48 yr old, n = 10). Total n = 18.	$Functions^0$ n = 36,324 (27,969 ± 1011) $Functions^{vK}$ n = 18,197 (14,690 ± 396) Total fxn rel $abun^{vK}$ = 61.7 ± 0.3 % vK coordinates n = 3302 (2965 ± 49)	Aotearoa Genomic Data Repository project AGDR00045 (Barnes et al., 2020); Aotearoa New Zealand
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.	$Functions^0$ n = 30,125 (20,328 ± 767) $Functions^{vK}$ n = 15,576 (11,051 ± 334) Total fxn rel $abun^{vK}$ = 61.9 ± 0.3 % vK coordinates n = 3076 (2537 ± 54)	MG-RAST project mgp16379 (Sun and Badgley, 2019); United States
Australian Microbiome Initiative (AMI) disturbed versus natural	Disturbed (n = 29) or natural (n = 55) soils, comprising temperate climate zone, surface (0-10cm) with 7.5–45% clay content (i.e., avoiding very sandy and very clayey soils). Total n = 84.	$Functions^0$ n = 37,335 (22,959 ± 1172) $Functions^{vK}$ n = 18,551 (12,212 ± 553) Total fxn rel $abun^{vK}$ = 60.8 ± 1.0 % vK coordinates n = 3326 (2700 ± 69)	AMI Data Portal [Data accessed Sep 2022] (Bissett et al., 2016) [d]; Australia

Notes: [a]F = females, M = males. [b]Number of $functions^0$ from initial SUPER-FOCUS profiling, versus $functions^{vK}$ (fxn) with available compound information for mapping to vK coordinates. [c]PC1 of problem behaviors were analyzed as numeric values for visualizing and assessing CPP_{class} and $CPP_{density}$ data; and high vs. low PC1 groups for CPP_{ASALR} data. High PC1 values represent more problematic behavior. [d]AMI Data portal URL: https://data.bioplatforms.com/organization/australian-microbiome.

183 ***2.2. Compound processing potential (CPP) mapping approach***

184 CPP measures were derived via the following steps (Fig. 2; further detail in Appendix A Supporting
185 Information):

186 a. Shotgun metagenomics raw sequences were accessed, and bioinformatic steps were run on Flinders
187 University DeepThought high-performance computing facility (Flinders_University, 2021).

188 b. Raw sequence data were inspected using FastQC (v0.11.9; Andrews, 2018) and quality control
189 trimming performed using Fastp (v0.23.2; Chen et al., 2018).

190 c. Functional potential profiles were derived from good quality read 1 sequences using SUPER-FOCUS
191 (Silva et al., 2015) functional annotation software, linked to the Diamond sequence aligner (v0.9.19;
192 Buchfink et al., 2021) and version 2 100% identity-clustered reference database (100_v2;
193 https://github.com/metageni/SUPER-FOCUS/issues/66). Where subjects/samples were represented by
194 multiple sequence files, the combined SUPER-FOCUS outputs were normalized so that the total
195 functional relative abundances summed to 100% in each subject/sample.

196 d. Every SUPER-FOCUS function (output row) was translated to one or more corresponding chemical
197 reaction(s) using a purpose-built R-script algorithm based on ModelSEED database lookup tables (from
198 https://github.com/ModelSEED/ModelSEEDDatabase; accessed 10 Aug 2022). The algorithm sought
199 matches based on either: full matching of functional hierarchies (using subsystem-class, -subclass, -
200 name and -role); detection of EC number; or matching of SUPER-FOCUS function name within
201 ModelSEED lookup tables for reactions (reaction name or alias), subsystems (role), or reaction-
202 pathways (external reaction name).

203 e. Every chemical reaction was converted to mean vK coordinates (O:C and H:C molar ratios),
204 considering all C-containing reaction input and product compounds and weighted according to reaction
205 stoichiometry (see example in Fig. 2.). Compounds not containing C were ignored due to undefined
206 O:C and H:C ratios. Data for compounds were based on Hill system chemical formulae in protonated
207 form.

f. Overall mean vK coordinates were calculated for each SUPER-FOCUS function via averaging one or more linked chemical reactions. The majority of SUPER-FOCUS functional relative abundances were successfully mapped to vK coordinates (sample ranges 52-84%, means 55-67%, Table 1, Table S3).

g. In vK coordinate space, we analyzed the spatial assignment of functional relative abundances to derive the following CPP data types:

1) CPP_{class}: summed functional relative abundances mapping to major compound classes (based on pre-defined zones from Wu et al., 2018) (see Fig. 1, Table S1).

2) CPP_{ASALR}: to address large variation in CPP_{class} values across case studies, we implemented a first-pass normalization aiming to account for microbial activity levels, here termed amino sugar adjusted log ratio (ASALR) data. Amino sugars have previously been used as a biomarker of microbial residue turnover (Hu et al., 2018) as they are major components of bacterial and fungal cell walls (peptidoglycan and chitin). CPP_{class} values were divided by the CPP_{class} value for amino sugars, followed by a variance-stabilizing log10-transformation.

3) $CPP_{density}$: captured the density of functional relative abundances within radial buffers of varying proximity (radii of 0.05, 0.1, 0.15, 0.2, 0.25 vK units) to selected focus biomolecules (Fig. 1, Table S2, Appendix A Supporting Information). These comprised short-chain fatty acids: acetate, propionate, and butyrate; vitamins: riboflavin (B2), cobalamin (B12), pyridoxal 5'-phosphate (B6), folate (B9), menaquinone (K2); and glutamate and pyruvate. Functional relative abundances within radial buffers were summed then divided by the respective area in vK units^2.

4) Compound-associated vK coordinates: Each SUPER-FOCUS functional row was translated to corresponding vK coordinates (as used for the above spatial assignments). Further supplementary analyses were undertaken at the vK coordinate level, including network analyses (in the T2D case study), differential abundance analysis (in the problem behavior case study), and weighted-mean vK coordinate analysis within major compound zones.

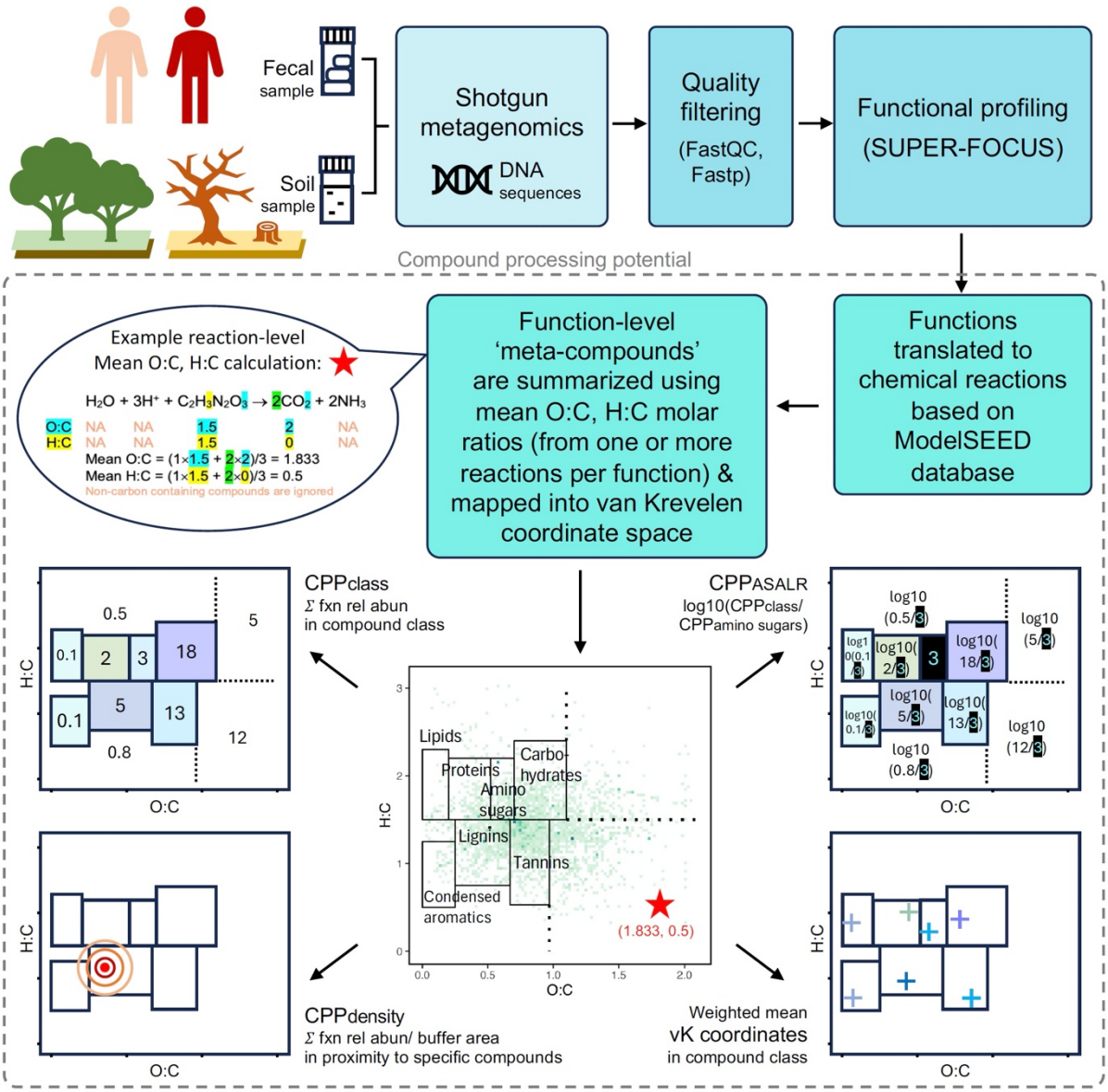

Fig. 2. Flow chart of compound processing potential (CPP) analyses. The example reaction shows calculation of mean O:C and H:C molar ratios (e.g., 1.833, 0.5) based on stoichiometry of all carbon-containing compounds. Mean O:C and H:C values are then averaged across one or more chemical reactions per function, to derive function-level vK coordinates. Each function-level mapping into vK space represents a conceptual 'meta-compound' based on the collective elements and stoichiometry of all C-containing compounds involved in all reactions corresponding to a single microbiota-mediated function.

2.3. Data visualization and statistical analyses

Details of visualization and statistical testing using standard approaches are provided in Appendix A Supporting Information.

3. Results

3.1. Human health and disease

Gut metagenome CPP_{class} (Figs. S4, S8, S12, S18, Table S3) and $CPP_{density}$ (Figs. S5-S6, S9-S10, S13-S14, S19-S20, Table S4) data produced strong associations in ACVD and colorectal cancer compared to normal subjects (detailed below). Many patterns observed in CPP_{class} data were reinforced in the putative activity-normalized CPP_{ASALR} measures (Fig. 3, Table S5), and this transformed data format showed stabilized variance across case study datasets. Interestingly, across all CPP_{class}, $CPP_{density}$, and CPP_{ASALR} measurements, when associations were found in both sexes they were always in the same direction (Tables S3-S5). In the T2D (female only) and problem behavior case studies, we observed far fewer relationships in the coarse CPP_{class}, CPP_{ASALR} or biomolecule-focused $CPP_{density}$ data. Therefore, we pursued network analyses and differential abundance analyses respectively in these case studies, as illustrative examples of more detailed supplementary analyses. Weighted mean vK-coordinate analyses produced striking associations in ACVD (Fig. S7 Tables S6-S7), but weaker effects in other case studies (Figs. S11, S15, S21, Tables S8-S12). The statistical test results described below are detailed in Appendix A.

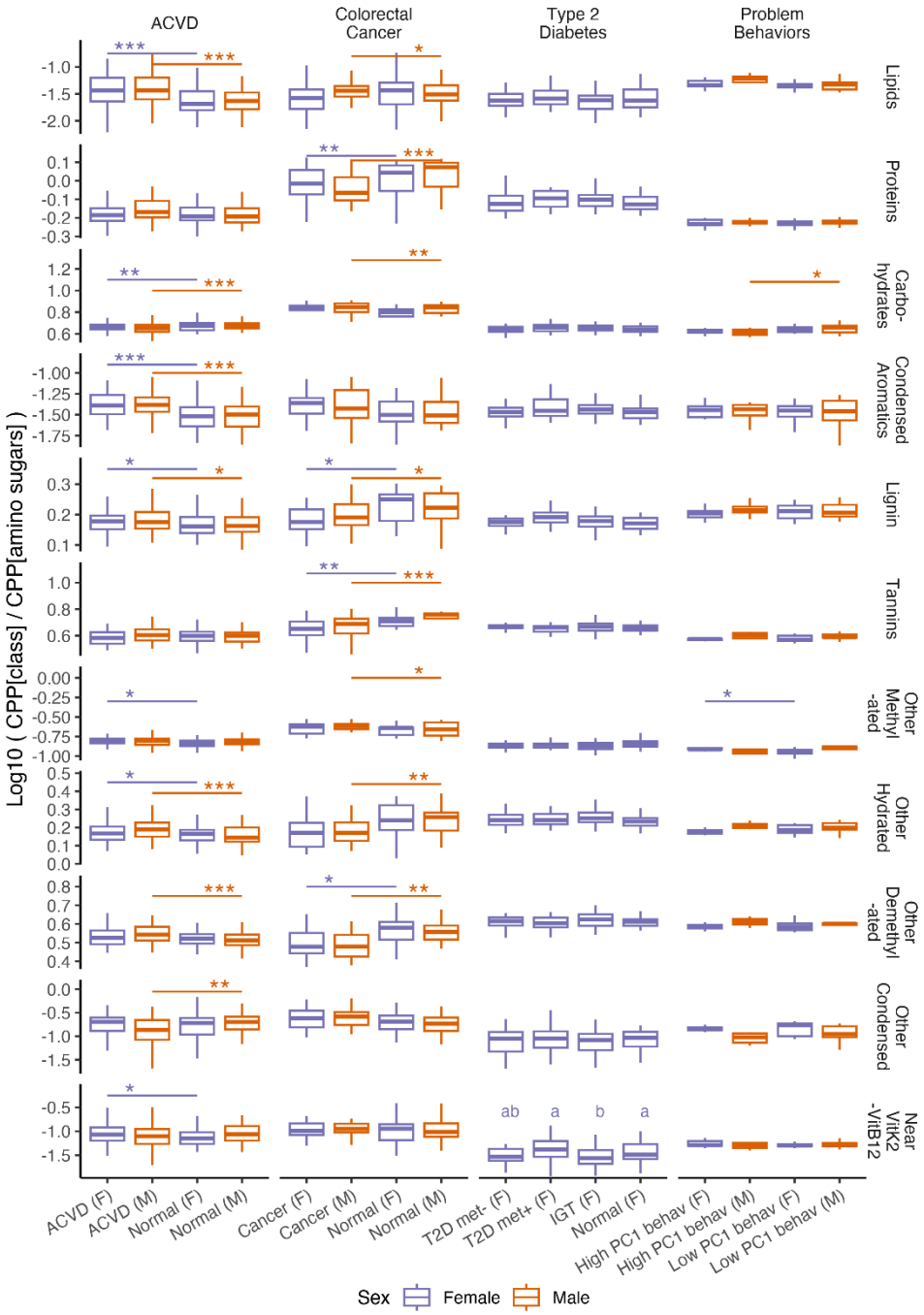

Fig. 3. Amino sugar-adjusted log ratio compound processing potential (CPP_{ASALR}), representing putative microbiota activity-normalized values, in normal healthy and diseased female (F) and male (M) subjects for atherosclerotic cardiovascular disease (ACVD), colorectal cancer, type 2 diabetes (T2D) with and without Metformin treatment (met +/-), impaired glucose tolerance (IGT), and high and low first principal component (PC1) problem behavior values. For visualization purposes outlying values are not shown. However, statistical tests were based on all data (Table S5). Sample sizes are detailed in Table 1. Tests for differences are performed within a single sex. In T2D data, groups not sharing a letter are different.

270 ACVD associated with CPP_{class} values, compared to normal subjects, in the form of increased
271 potential metabolism of lipids (in females – F, and in males – M), proteins (M), condensed aromatics (F, M)
272 lignin (F, M), tannins (M), other hydrated compounds (F, M), other demethylated compounds (M), and near
273 vitamin K2-vitamin B12 compounds (F); contrasting with decreased potential metabolism of carbohydrates
274 (F, M) and other condensed compounds (M). Despite all samples initially summing to 100% total functional
275 relative abundances, there was less complete conversion to identifiable reactions and total sum CPP_{class} data
276 in females only for ACVD compared to normal cases. Many of these patterns were reinforced in the
277 CPP_{ASALR} data, with ACVD notably associated in both sexes with increased potential metabolism of lipids,
278 condensed aromatics, lignin, and other hydrated compounds; but decreased potential metabolism of
279 carbohydrates. For $CPP_{density}$ measures within 0.05 vK unit radii, ACVD associated with increased potential
280 metabolism of propionate (F), vitamin B12 (F), vitamin B6 (M), vitamin B9 (F, M), vitamin K2 (F, M) and
281 pyruvate (M); but decreased potential metabolism of acetate (M) and glutamate (F, M).
282 Colorectal cancer also associated prominently with CPP_{class} values, specifically increased potential
283 metabolism of lipids (M), amino sugars (F, M), condensed aromatics (M), lignin (M), other condensed
284 compounds (F, M) and near vitamin K2-vitamin B12 compounds (M); but decreased potential metabolism
285 of proteins (F, M) and tannins (F, M). Total sum CPP_{class} data were reduced in male colorectal cancer cases
286 compared to normal. CPP_{ASALR} values in colorectal cancer subjects of both sexes were associated with
287 decreased potential metabolism of proteins, lignin, tannins, and other demethylated compounds. For
288 $CPP_{density}$ measures within 0.05 vK unit radii, colorectal cancer associated with increased potential
289 metabolism of vitamin B2 (M), vitamin B12 (F, M), vitamin B9 (F, M), and pyruvate (M). Weighted mean
290 vK-coordinates varied between colorectal cancer and normal subjects with different centroids in condensed
291 aromatics (F), lignin (M), and tannins (M); and different beta-dispersion or spread in proteins (F, M) and
292 condensed aromatics (F, M).
293 In the female-only T2D case study results, we found CPP_{class} values decreased for other methylated
294 compounds in IGT compared to normal subjects and decreased for near vitamin K2-vitamin B12

295 compounds in IGT compared to normal, with a reduced total sum CPP_{class} in normal compared to other
296 subjects. No associations were found with $CPP_{density}$ values. CPP_{ASALR} values were decreased for near
297 vitamin K2-vitamin B12 compounds in IGT compared to normal and T2D Met+ subjects. From analysis of
298 weighted mean vK-coordinates among diagnosis groups, a difference in centroids was found only in the
299 compound class of proteins. Lacking interpretable results at the compound class or focus biomolecule
300 levels, further investigation was performed via network analysis (Appendix A Supporting Information) on
301 20 subjects in each diagnosis group based on inferred correlations between functional relative abundances
302 mapped to unique vK-coordinates. Network analyses were based on commonly observed vK coordinates
303 (present in at least 60% of samples and minimum 2% sum functional relative abundance across samples).
304 Network diagrams (Fig. 4) and structure dendograms (Fig. S16) display a transition in their complexity,
305 number of nodes and fraction of negative edges (interactions) from simplest in normal and IGT subjects to
306 most complex in T2D Met+ and T2D Met- subjects. Comparing network characteristics for the four groups
307 (normal, IGT, T2D Met-, T2D Met+; n = groups of 20) to a bootstrapped (B = 1000) density distribution of
308 randomly resampled networks (n = 20, drawn from the same pool of 80 subjects) (Fig. S17, Table S17) we
309 found normal healthy subjects had the lowest fraction of negative edges and the highest degree
310 centralization. Untreated disease, T2D Met-, had the lowest closeness centralization (graph-level inverse of
311 average geodesic distance between nodes); and borderline significant results for the highest fraction of
312 negative edges (negative correlations between vK-coordinates), lowest betweenness centralization (graph-
313 level centrality based on broker positions connecting others), and lowest mean distance (average path length
314 between nodes). In short, normal subjects appear to have far less correlations between vK-coordinates
315 (fewer nodes / vertices), and for the nodes and links that are present they are largely positively correlated
316 and highly interlinked. Whereas T2D Met- (untreated disease) is characterized by a much larger number of
317 negatively correlated vK-coordinate nodes, which on average have shorter links, and are less well
318 connected-up across the whole network.

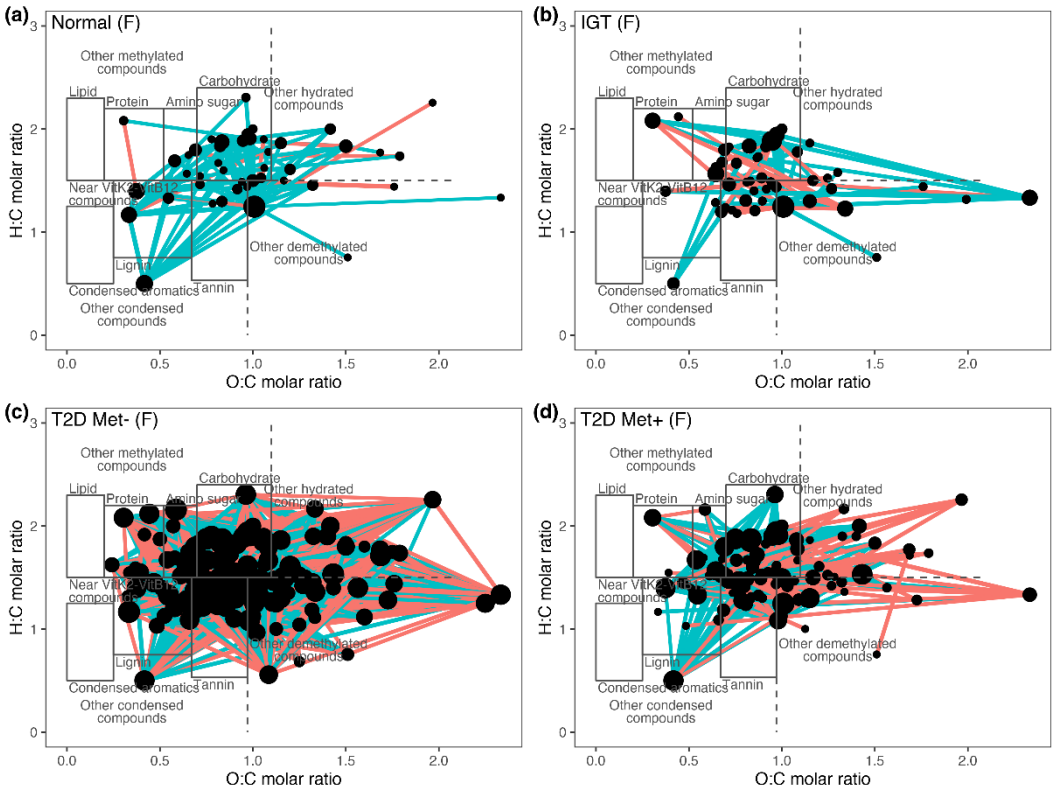

Fig. 4. Network diagrams based on commonly observed vK coordinates for female subjects (n = groups of 20) with diagnoses: (a) normal, (b) impaired glucose tolerance (IGT), (c) type 2 diabetes without Metformin (T2D Met-), and (d) type 2 diabetes with Metformin (T2D Met+). Nodes are located according to unique compound-associated vK coordinates, with size reflecting node degree (number of linked significant correlations). Links between nodes display positive (aqua color) and negative (red color) correlations ($p \leq 0.05$).

327 Problem behaviors displayed few significant associations in CPP_{class} values: increasing PC1 of
328 problem behaviors associated with increased potential metabolism of lipids (M), amino sugars (F), other
329 demethylated compounds (M), and near vitamin K2-vitamin B12 compounds (F); but decreased potential
330 metabolism of carbohydrates (M). For $CPP_{density}$ values, patterns were found in males only: increasing PC1
331 of problem behaviors associated with increased potential metabolism of vitamin B6 (M) and vitamin B9
332 (M); but decreased potential metabolism of acetate (M), vitamin B12 (M) and glutamate (M). For CPP_{ASALR}
333 values, the comparison was made between groups of high PC1 versus low PC1 of problem behaviors (for
334 consistent display with other case studies in Fig. 3). High PC1 values associated with increased potential
335 metabolism of other methylated compounds (F); and decreased potential metabolism of carbohydrates (M).
336 Weighted mean vK-coordinates between high PC1 and low PC1 of problem behaviors showed a difference
337 in beta-dispersion within condensed aromatics for females only. To explore this dataset in more detail, we
338 identified differentially abundant functions and vK-coordinates (i.e., aggregated functional relative
339 abundances via bioenergetic mapping) in high PC1 versus low PC1 subjects, separately within each sex. In
340 females, 22 differentially abundant functions, compared to only 2-vK coordinates (with 3 corresponding
341 functions), were identified (Figs. S22, S24, Table S18, S20). In males, 6 functions compared to 2 vK-
342 coordinates (with 8 corresponding functions) were identified (Figs. S23, S25, Table S19, S21). Not all
343 functions could be mapped into vK coordinate space. Interestingly, there was no overlap in functions
344 identified directly versus indirectly (from aggregation into vK-coordinates). This means that differential
345 abundance analysis using vK-coordinates can provide entirely different focal areas for investigation
346 compared to the standard function-level analysis. From the vK-coordinate level analysis, high PC1
347 (compared to low PC1) females exhibited increased potential metabolism of Uridine phosphorylase (EC
348 2.4.2.3) (fxn_14491) involved in pyrimidine conversions; and decreased aldehyde lyases dihydroneopterin
349 phosphate phosphatase and dihydroneopterin aldolase (EC 4.1.2.25) (fxn_12938; fxn_12942). High PC1
350 males exhibited increased phosphoenolpyruvate carboxykinase (GTP) (EC 4.1.1.32) (fxn_2926) associated
351 with pyruvate metabolism; and also increases in multiple functions (fxn_821; fxn_2958; fxn_2973;

352 fxn_12703; fxn_12705; fxn_12786; fxn_12788) all involving alcohol dehydrogenase (EC 1.1.1.1) and
353 acetaldehyde dehydrogenase (EC 1.2.1.10), with or without pyruvate-formate-lyase deactivase – involved in
354 degradation of aromatic compounds, biphenyl, tryptophan, and pyruvate metabolism. Results and
355 visualizations for the standard function-level analysis are included for comparison but are not discussed
356 further.
357
358 ***3.2. Plant-soil systems***
359 We found remarkable consistency in many observed patterns across the environmental soil case studies.
360 Results reported here are for relative trends with increasing ecosystem maturity (i.e., older revegetation and
361 natural samples) in CPP_{class} (Figs. S26, S30, S33, Table S3), $CPP_{density}$ (Figs. S27-S28, S31-S32, S34-35,
362 Table S4), and CPP_{ASALR} (Figure 5, Table S5). Below, we highlight trends found in at least two of three
363 case studies.
364 In CPP_{class} data we observed: increased potential metabolism of lipids (post-mining, AMI) and
365 condensed aromatics (post-mining, AMI); but decreased potential metabolism of proteins (post-mining,
366 AMI), carbohydrates (post-mining, AMI, with marginal indications in both PCaN soil groups), lignin (post-
367 mining, AMI), other methylated compounds (post-mining, AMI), and other hydrated compounds (AMI,
368 PCaN acidic-neutral soils). Mixed or isolated results included potential metabolism: either decreased (AMI)
369 or increased (PCaN strongly acidic soils) for tannins; decreased for other demethylated compounds (AMI);
370 increased for other condensed compounds (AMI); and decreased (post-mining, PCaN acidic-neutral soils)
371 or increased (PCaN strongly acidic soils) for near vitamin K2-vitamin B12 compounds. Total CPP_{class}
372 compounds appeared to be less well characterized and mapped to functional reactions in AMI natural
373 compared to disturbed samples, but more well characterized in older revegetation (compared to younger
374 revegetation) within the PCaN strongly acidic soils.
375 CPP_{ASALR} results reinforced many patterns observed in the CPP_{class} data: we observed increased
376 potential metabolism of lipids (post-mining, AMI, marginal in PCaN strongly acidic soils) and condensed

aromatics (post-mining, AMI); but decreased potential metabolism of proteins (AMI, marginal in post-mining), carbohydrates (post-mining, AMI, PCaN acidic-neutral soils), and other methylated compounds (post-mining, AMI). Isolated results included, potential metabolism: increased for other condensed compounds (AMI); but decreased for lignin (AMI), tannins (AMI), other hydrated compounds (AMI), and other demethylated compounds (AMI). Differing from CPP_{class} results, for CPP_{ASALR} potential metabolism of near vitamin K2-vitamin B12 compounds increased (post-mining).

$CPP_{density}$ results were also quite consistent across case studies: we observed increased potential metabolism of vitamin B9 (AMI, marginal in post-mining and PCaN strongly acidic soils) and vitamin K2 (post-mining, AMI, marginal in PCaN strongly acidic); but decreased potential metabolism of acetate (post-mining, AMI, PCaN acidic-neutral soils), propionate (AMI), vitamin B2 (post-mining, marginal in PCaN strongly acidic soils), vitamin B12 (post-mining, AMI, PCaN strongly acidic soils), vitamin B6 (post-mining, AMI), glutamate (post-mining, AMI, PCaN acidic-neutral soils), and pyruvate (AMI). While butyrate $CPP_{density}$ values were low and indistinguishable across all samples using 0.05 vK unit radial buffers, for an example comparison we tested near butyrate $CPP_{density}$ using a larger 0.1 vK unit buffer in the AMI soils and found increased levels in natural compared to disturbed soils (Fig. S36).

Post-mining and AMI samples displayed significant and mostly consistent directional shifts in weighted mean vK-coordinate centroids across all compound classes (vK mapping zones) considered, while PCaN acidic to neutral soils also displayed shifts within carbohydrates and condensed aromatics (Fig. 6, Fig. S29, Tables S13-S16). Ecosystem maturity (/quality) explained 60–92% of the variation in weighted mean vK-coordinates in post-mining soils (Fig. 6a, Table S15). The following general patterns emerged with increasing ecosystem maturity: lipids became more reduced (lower oxygen content), proteins became more hydrated, amino sugars became more dehydrogenated or condensed, carbohydrates became more condensed (post-mining, PCaN acidic-neutral) or reduced (AMI), condensed aromatics became more condensed, lignin showed mixed trends (dehydrogenation in post-mining, reduction in AMI), and tannins

became more demethylated. Except for carbohydrates and amino sugars, these trends largely represented an outward extension of sample profile mapping into vK coordinate space with older ecosystems.

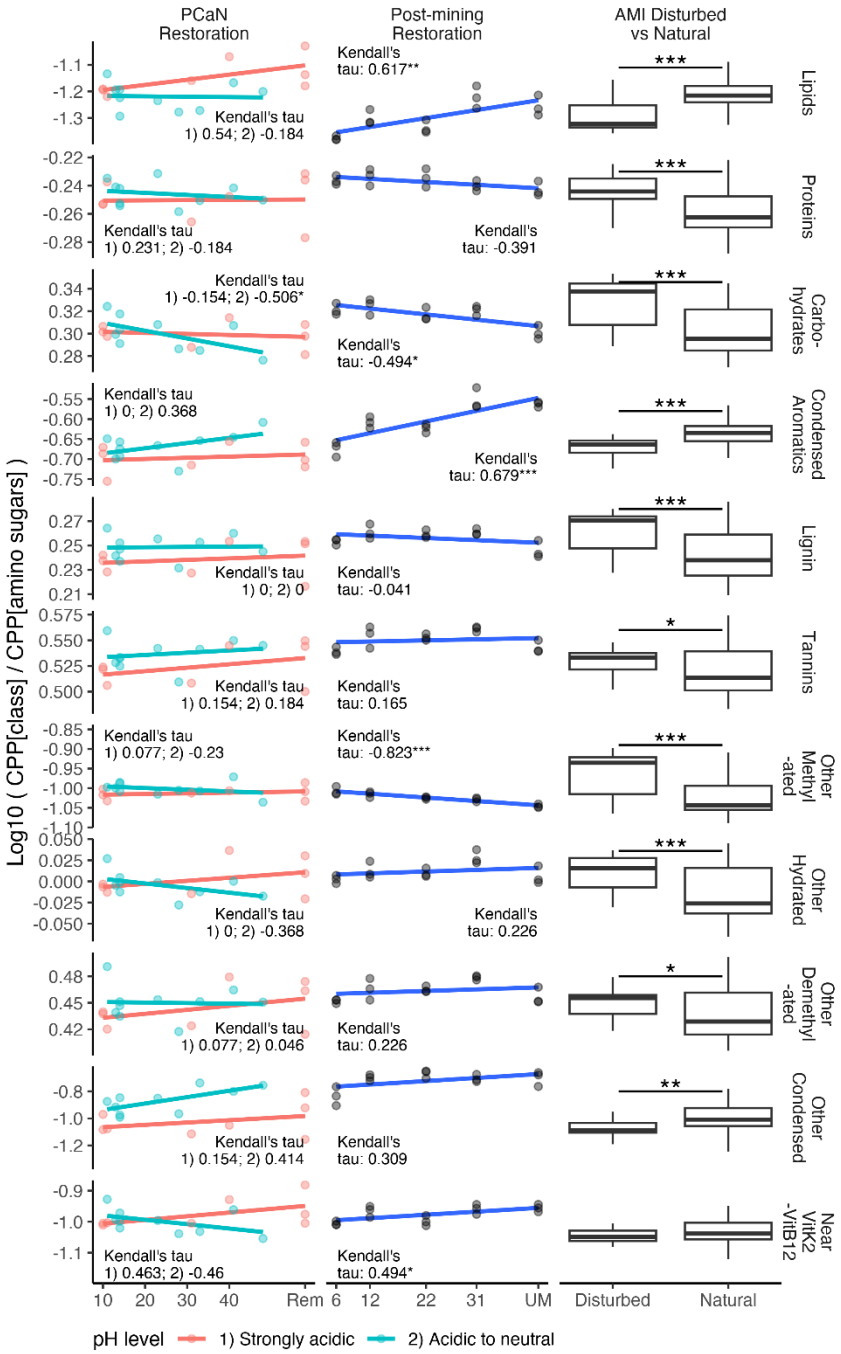

Fig. 5. Amino sugar-adjusted log ratio compound processing potential (CPP_{ASALR}), representing putative microbiota activity-normalized values, from People Cities and Nature (PCaN) urban restoration, post-mining restoration soils, and Australian Microbiome Initiative (AMI) disturbed versus natural soils. Linear

trends were used for visualization purposes. However, Kendall's tau correlation tests (suited to ordinal data) were applied. Sample sizes are detailed in Table 1.

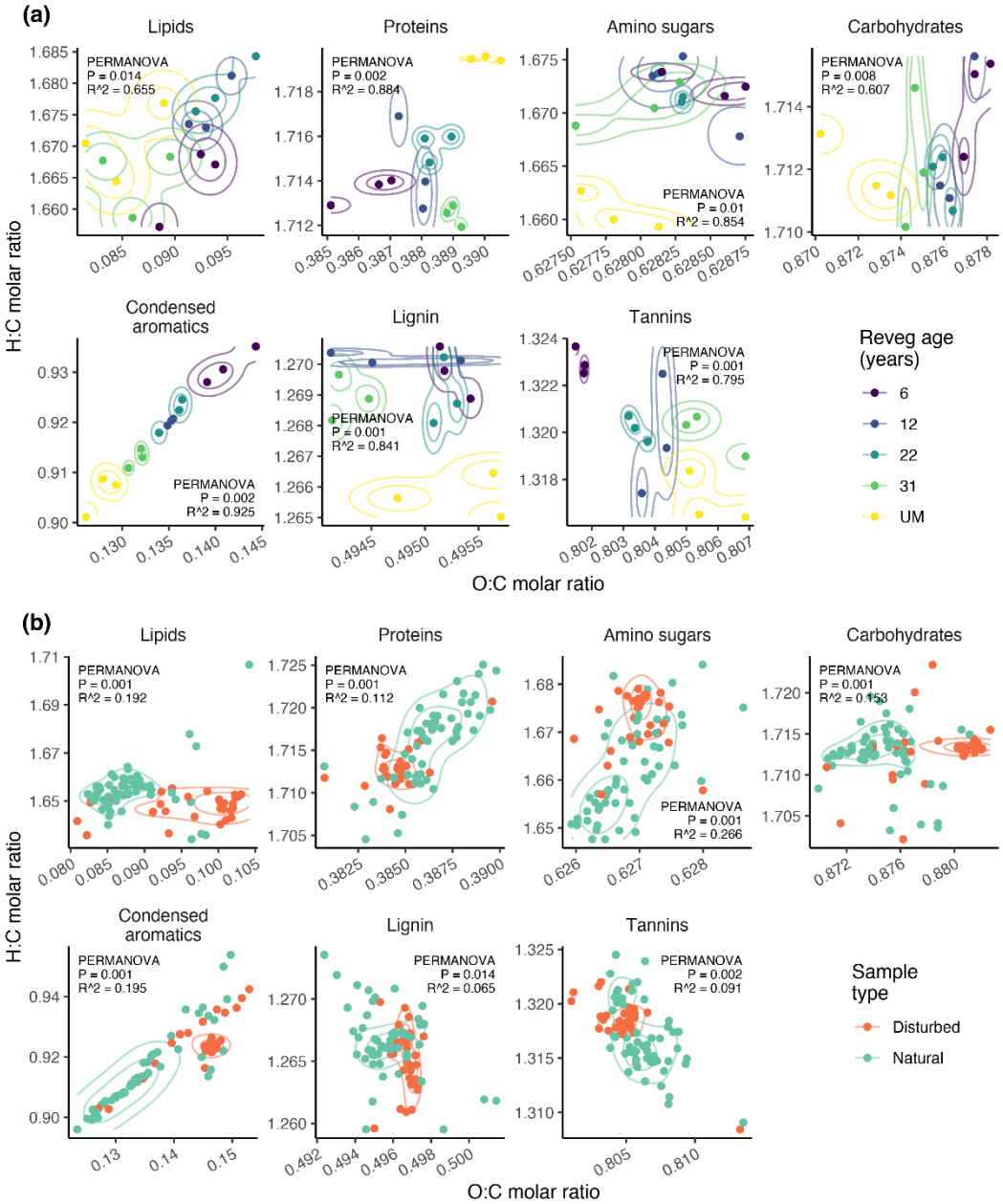

Fig. 6. Weighted-mean van Krevelen coordinates within select compound classes display significant shifts with maturity of plant-soil ecosystems. Patterns are from (a) post-mining forest ecosystem restoration soil samples (n = age-based groups of 3), and (b) AMI disturbed (n = 29) vs natural (n = 55) soil samples (b). PERMANOVA and beta-dispersion results are in Table S15-S16.

417 ***3.3. Consistency in bioenergetic 'topography' mapping***

418 $CPP_{density}$ plots visualizing local polynomial regression fitting ('loess') smoothed vK mapping profiles for
419 increasing radii buffer areas displayed striking consistency of form within each compound class, with
420 alternative expressions in either gut (Figs. S5, S9, S13, S19) or soils (Figs. S27, S31, S34).

421

422 **4. Discussion**

423 We show a meaningful bioenergetic basis to the development and differentiation of community-scale
424 microbiomes. Our CPP metrics quantified putative shifts in a microbial community's relative proficiency to
425 process different types of compounds across gradients of human- and environmental-health. We found
426 significant patterns of CPP association at varying resolutions: within major classes of compounds, near
427 focus biomolecules, and for unique vK coordinates. In the environmental soil metagenomes, we observed
428 surprising consistency in CPP profiles, which suggests they may link to plant-soil system conditions in
429 coherent and predictable ways. Our findings align with the notion that microbiota are shaped by the
430 bioenergetic status of prevailing substrates and micro-environments, and this information is simultaneously
431 recorded in their metagenomes. From a methodological perspective, our compound-focused bioenergetic
432 mapping approach demonstrates new pathways for assessing and interpreting microbial systems, capable of
433 supporting ongoing efforts to define healthy microbiomes. For example, aggregating functions via vK-
434 coordinates can provide entirely different focal areas for investigation compared to standard function-level
435 analyses.

436

437 ***4.1. Patterns in human health***

438 *4.1.1. Atherosclerotic cardiovascular disease (ACVD)*

439 We found strong links in both sexes between ACVD and increased potential gut microbial metabolism of
440 lipids, condensed aromatics, lignin, other hydrated compounds, vitamin B9 and vitamin K2; and decreased
441 potential metabolism of carbohydrates and glutamate. Globally, ACVD is the leading cause of death, with

442 established treatment involving lipid-lowering statins, while human gene editing approaches to disrupt lipid
443 processing are being trialed (Li and Wu, 2022; Vaduganathan et al., 2022).
444 Links between high-fat diets and ACVD are well established (Law, 2000). Gut microbiota can
445 contribute to ACVD by metabolizing the dietary lipid phosphatidylcholine, with subsequent production of
446 harmful trimethylamine oxide (Albenberg and Wu, 2014). Polycyclic aromatic hydrocarbons are known
447 risk factors in ACVD (Burstyn et al., 2005; Mallah et al., 2021). Lignin (different to lignan) is a complex
448 and ubiquitous structural plant polymer, considered predominantly insoluble fiber, with content ranging
449 from 1-2 g/100g in vegetables, fruits and cereals, up to 30-40 g/100g in nut shells and stone fruit kernels
450 (Tao et al., 2020). Lignin inhibits the enzymatic activity of α-glucosidase, delaying carbohydrate digestion
451 and absorption, with potential for low post-meal blood sugar levels (Chen et al., 2015). Acute hypoglycemia
452 (low blood sugar) can trigger cardiac events (Frier et al., 2011), with greater adverse risks in subjects with
453 significant comorbidities (e.g., T2D, ACVD). Our finding for decreased carbohydrates in ACVD might
454 align with these impacts on blood sugar. Alternatively, we speculate that result might be symptomatic of a
455 more Westernized diet (high in animal protein, sugar, starch, and fat – and lower in carbohydrate content
456 than a plant-rich diet; Albenberg and Wu, 2014).
457 Excessive vitamin B9 (folate) is associated with ACVD risk via a non-linear u-shaped dose-
458 response relationship (Xu et al., 2023). Non-linear u-shaped dose-response relationships are common in
459 biological systems (i.e., hormesis, or deficiency-sufficiency-toxicity) (Calabrese et al., 2007). Similarly, our
460 results linking increased vitamin K2 with ACVD appear contrary to recent opinion (Essa et al., 2021),
461 although non-linear u-shaped dose-responses have also been observed (Bellinge et al., 2021). Vitamin K2 is
462 commonly found in fermented foods which are less common in Western diets (McFarlin et al., 2017). Here,
463 the ACVD case study was based on Chinese subjects whose diets potentially contained higher quantities of
464 fermented foods including vitamin K2. Possibly, these subjects were more susceptible to adverse effects if
465 excessive levels of vitamin K2 were reached.

466 Our results linking reduced glutamate with ACVD appear contrary to findings from large US cohort
467 studies which found higher glutamate levels, and lower glutamine:glutamate ratios, correlated with
468 increased ACVD risk (Ma et al., 2017; Zheng et al., 2016). Dietary proteins are a major source of glutamate
469 (Loï and Cynober, 2022). However, some ethnic populations may have inadequate protein in their diets (Liu
470 et al., 2013). Interestingly, in female adults from rural western China with inadequate (and largely plant-
471 derived) protein intake, increasing animal protein associated with reduced risk of hypertension (Liu et al.,
472 2013). Together, these observations suggest that a u-shaped dose-response may also operate for animal-
473 based proteins, glutamate, and ACVD risk.

474 *4.1.2. Colorectal cancer*

475 In the colorectal cancer subjects (from France) we found consistent associations across CPP metrics in both
476 sexes with increased potential metabolism of amino sugars, vitamin B12 and B9; and decreased potential
477 metabolism of proteins and tannins. Colon cancer has been associated with low dietary fiber, low fruit and
478 vegetable consumption, and high red meat consumption (Law, 2000). Our results were consistent with
479 reports for anticancer activity, including protective effects against colorectal cancer, from some tannins or
480 polyphenols (e.g., components in green and black tea, resveratrol in red wine and grapes) (Alam et al.,
481 2018). Unfortunately, our data did not distinguish between animal- and plant-based protein. However, red
482 meat is widely consumed in France, with 41% of males and 24% of females consuming above guideline
483 levels (Mota et al., 2021). Amino sugars are sugar molecules with at least one hydroxyl group substituted
484 by an amino group. In biological systems, they are formed by catalytic activity acting on amino acids
485 (glutamate, glutamine – building blocks of protein) to transfer an amino functionality to a sugar phosphate
486 or sugar nucleotide (Skarbek and Milewska, 2016). Therefore, both glucose (sugar) and amino acids
487 contribute to amino sugar formation. Meanwhile, metabolism of both glucose and amino acids plays a key
488 role in colorectal cancer development (Zhang et al., 2023). Possibly, our finding of increased potential
489 metabolism of amino sugars with colorectal cancer may reflect dysregulated activity of glucose and amino
490 acids with the product of their interaction (amino sugars) recorded by the gut microbiome. Consistent with

491 our findings, vitamin B9 (folate or folic acid) and vitamin B12 supplementation have been associated with
492 increased risk of colorectal cancer (Oliai Araghi et al., 2019).

493 *4.1.3. Type 2 diabetes (T2D)*

494 In the T2D case study, our lack of clear findings linked to major compound classes or focus biomolecules
495 seems consistent with reports that T2D is a complex, multifaceted, highly heterogeneous polygenic disease
496 with uncertain etiology (Pearson, 2019). We found untreated (Met-) T2D exhibited an anomalous and
497 complex CPP network, including a high number of negative correlations (indicating negative feedbacks)
498 ranging widely across vK coordinate space (i.e., covering a spectrum of compounds and bioenergetic
499 status). Diagnosis of T2D is based on elevated blood glucose, primarily arising from insulin resistance and
500 inadequate insulin secretion (Forslund et al., 2015). However, a clear diagnostic test for T2D is lacking,
501 except by exclusion of other causes (Pearson, 2019). A range of factors including genetics, dietary habits,
502 sedentary lifestyle, and gut microbiota are involved in disease development (Forslund et al., 2015).
503 Possibly, with more detailed examination, diagnostic relationships (e.g., correlations, ratios) might be
504 uncovered in relative abundance patterns of compound-associated vK-coordinates underpinning the
505 anomalous T2D Met- network.

506 *4.1.4. Problem behaviors*

507 In the problem behavior case study, different compound associations were observed in female and male
508 children. Increased problem behavior (higher PC1) in females associated with increased potential
509 metabolism of amino sugars, other methylated compounds, and near vitamin K2-vitamin B12 compounds.
510 From differential abundance analysis, high PC1 females exhibited decreased dihydroneopterin aldolase—an
511 enzyme involved in converting dihydroneopterin (a molecule involved in folate biosynthesis) into other
512 compounds (Deng et al., 2000). Excessive serum levels of dihydroneopterin have been associated with
513 major depression (Kusunoki et al., 1999). Possibly, in case study subjects, reduced levels of
514 dihydroneopterin-degrading enzyme have promoted accumulation of dihydroneopterin in association with
515 problem behaviors. High PC1 females also exhibited increased uridine phosphorylase, an enzyme involved

516 in pyrimidine metabolism that converts uridine to uracil (Cao and Pizzorno, 2004), therefore possibly
517 degrading uridine levels in those subjects. Uridine is linked to energy metabolism and glutamate-mediated
518 excitatory neurotransmission in the brain, and supplemental uridine treatments have been used to reduce
519 depressive symptoms in adolescents (Douglas G. Kondo, 2011).

520 In males, increased PC1 associated with increased potential metabolism of lipids, vitamins B6 and
521 B9; and decreased potential metabolism of carbohydrates, acetate, vitamin B12, and glutamate. Vitamins
522 B12 (cobalamin) and B9 (folate) are recognized precursors involved in forming key neurotransmitters
523 dopamine, noradrenaline (norepinephrine), and serotonin (Hutto, 1997). These three neurotransmitters
524 occur in the vicinity of vitamin B12 and K2 in vK coordinate space. Vitamin B12 deficiency has been
525 associated with depressive disorders in older subjects (Henning Tiemeier et al., 2002). Glutamate is the
526 major excitatory neurotransmitter of the healthy mammalian brain, and an abundant free amino acid
527 important in multiple metabolic pathways, which requires regulation at optimal levels in extracellular fluids
528 (Zhou and Danbolt, 2014). High PC1 males also exhibited increased phosphoenolpyruvate carboxykinase—
529 an enzyme involved in cataplerosis, or removal of intermediate 4- and 5-carbon compounds from the TCA
530 cycle (Owen et al., 2002; Yang et al., 2009). These intermediates are removed because they cannot be fully
531 oxidized for energy metabolism within the TCA cycle, but are converted elsewhere to glucose, fatty acids
532 or amino acids (Owen et al., 2002). High PC1 males also exhibited increases in alcohol dehydrogenase,
533 acetaldehyde dehydrogenase, and pyruvate-formate-lyase deactivase, variously involved in pyruvate
534 metabolism, degradation of aromatics and biphenyl, and tryptophan catabolism. Key processes of energy
535 metabolism involving glucose, lipids, protein and the TCA cycle (via keystone molecules pyruvate, acetyl-
536 CoA, and glutamate) have been implicated in major depressive disorder, although precise pathways of
537 pathogenesis are still unclear (Gu et al., 2021).
538

4.2. Patterns in plant-soil systems

Our findings point to generalizable patterns with older ecosystems for increasing microbial CPP associated with lipids, condensed aromatics, vitamin B9, and vitamin K2; and decreasing CPP associated with proteins, carbohydrates, lignin, other methylated compounds, other hydrated compounds, acetate, vitamin B12, vitamin B6, and glutamate. Drivers of shifting CPP in soils are expected to include changing: (1) composition of biota and biotic materials including plants, organic debris, and re-assembly of invertebrate and microbial communities, and (2) soil abiotic conditions due to plant-soil feedbacks (e.g., pH, nutrients, organic carbon content, temperature, moisture regime) (van der Putten et al., 2016; Wardle et al., 2004). This includes macro-environmental influences with development of vegetation structure and canopy cover (e.g., shading, rainfall interception, altered drainage). CPP values also likely reflect a dynamic balance between resource availability and use by microbiota.

For example, we might expect greater accumulation of lignin in soils of older ecosystems due to plant inputs such as dead roots, bark, leaf litter, and other structural plant residues. However, we observed reduced CPP for lignin in these sample types. Fungi are major lignin degraders (Janusz et al., 2017) and fungal communities vary with ecosystem disturbance and abiotic conditions (Rodriguez-Ramos et al., 2021). Interestingly, our results were counter to expectations for elevated fungal decomposition of lignin in older ecosystems. Reforestation with native mixed-species can produce higher levels of recalcitrant soil organic matter (Cunningham et al., 2015) (e.g., humic acid which is hard to decompose and maps to lignin in vK space). Our CPP metrics are relative and compositional (based on functional relative abundances summing to a maximum of 100%), so it may be that in relative terms, the metabolic foci of microbiota are shifted to processing other materials. Or possibly, structural plant materials may be more accessible for degradation in disturbed (e.g., agricultural) soil environments, depending on plant residue management, nutrient availability and other factors.

We expect some CPP quantities are driven primarily by plant material inputs. For example, soils from more mature ecosystems in temperate climates, represented in samples from AMI and post-mining (in

564 the Appalachian Plateau, southwestern Virginia USA; Avera et al., 2015; Sun and Badgley, 2019),
565 displayed a positive relationship with CPP for lipids and condensed aromatics. These two compound classes
566 are represented in plant-based essential oils and volatile, aromatic organic compounds. Oils are found in
567 high densities in much of the fire-adapted Australian flora (unlike New Zealand flora) (Bowman et al.,
568 2012). Increased CPP for lipids might also arise due to increased density of energy storage linked to
569 primary production, or more active plant signaling in response to abiotic stress (Hou et al., 2016). High
570 levels of lipids and condensed aromatics in mature ecosystem soils could also be a result of increasing plant
571 investment into defensive mechanisms via antimicrobial essential oils (Hammer et al., 1999), and volatile
572 and aromatic secondary defense compounds induced by herbivory (typically by invertebrates) (Erb and
573 Kliebenstein, 2020).

574 Shifting weighted mean vK coordinates across many compound classes (in AMI and post-mining)
575 suggests broad changes in the composition of microbial substrates with more mature ecosystems. The
576 changing composition of the microbiota itself may contribute to this. Carbohydrates are of interest due to
577 the potential contribution of plant-based material to human diet, and CPP_{class} values for carbohydrates were
578 consistently assigned the largest sum of functional relative abundances in the human gut samples. With
579 more mature ecosystems, CPP for carbohydrates decreased in relative terms, but weighted vK coordinates
580 suggest carbohydrate CPP shifts towards favoring processing materials with reduced oxygen content per
581 unit of carbon. There is likely to be global variation in environmental soil CPP driven by soil abiotic factors
582 and changing biota (vegetation and animals), previously outlined.

583

584 ***4.3. Potential environment-human health links***

585 This work opens new avenues for investigating environment-human health connections as environments
586 vary in their production of human health-associated compounds. Moreover, varying environmental
587 microbiota exposures may supply modulating CPP profiles for colonizing or transient impacts to human
588 microbiomes (e.g., skin, airway, gut), which are intimately linked to our health. We show CPP patterns

589 imprinted in environmental soil metagenomes are linked with the maturity (or quality) of plant-soil systems
590 and abiotic factors such as soil pH. We demonstrate that CPP 'landscapes' (or the range of meta-
591 compounds) involved in functions that can be performed by soil and gut metagenomes show a large degree
592 of overlap with similar distributions of functional relative abundances across vK space (Fig. S2). We also
593 show a similar spectrum of CPP measures that are impacted by ecosystem quality (e.g., increased potential
594 metabolism of lipids, condensed aromatics, vitamin B9, vitamin K2; and decreased potential metabolism of
595 proteins, carbohydrates, lignin, acetate, vitamin B12, vitamin B6, and glutamate) are also variously linked
596 to human health and disease. However, we urge caution in attempting to directly translate CPP trends in
597 plant-soil environments to infer possible implications for human health. We stress that non-linear, u-shaped
598 dose-response relationships, as discussed earlier (Calabrese et al., 2007), are common and relevant in the
599 context of dietary and environmental exposure-human health links. With potential for u-shaped dose-
600 response effects (i.e., deficiency, sufficiency, toxicity), we might expect environmental microbiota CPP to
601 have varying influences in different geographic/cultural populations. Also, the gut represents a more tightly
602 controlled micro-environment (redox, pH, etc.) unsuited to many environmental microbes.
603 Example evidence for potential environment-human transfer of microbial CPP comes from
604 Endomicrobia species found in oral microbiota of indigenous peoples from central Australia (Handsley-
605 Davis et al., 2022). Endomicrobia species provide energetic advantage for cellulose digestion in the guts of
606 termites and wood-eating insects—and transfer to humans has occurred likely through use of termites and
607 termite mounds in traditional food and medicine (Handsley-Davis et al., 2022). Speculatively, our results
608 suggest that if broad supplementation of human microbiota CPP capacity is required (spanning a range of
609 health-supporting biomolecules), this may require exposure to multiple types of environments.
610 Alternatively, to address particular microbiota-associated disease states, specific environment types may
611 provide more targeted microbiota CPP supplementation. As knowledge builds of the varying proficiency of
612 environmental microbiota to process human health associated compounds in different qualities of

613 ecosystems, we might design environmental interventions or nature prescriptions to modulate human-
614 microbial exposures, thereby advancing microbiota-oriented approaches to human health.
615
616 ***4.4. Limitations***
617 There are important limitations in this study. CPP metrics do not measure actual compounds; rather, they
618 quantify conceptual 'meta-compounds' or assemblies of elements based on function-level summary mean
619 O:C and H:C ratios consistent with compounds of interest. Quantification occurred via mapping into vK
620 space and aggregating functional relative abundances into major compound classes, near focus
621 biomolecules, or at unique vK coordinates, to assess CPP structural profiles of metagenomes. Conceptually,
622 mean function-level attributes (O:C and H:C ratios) represent a mid-point of chemical transformation
623 mediated by microbiota (i.e., interpreted as 'X% of functions were involved in processing
624 compound/biomolecule type Y'). Stoichiometry rules determine that reaction inputs and products will have
625 balanced O, H, and C atomic counts. However, different mean O:C and H:C ratios can arise due to
626 uncounted O and H atoms in non-C containing species (O:C, H:C values become undefined). For future
627 work, the CPP mapping algorithm could be readily adjusted to separately target reaction inputs, or products,
628 or individual chemical species.
629 Our coarse compound classes did not distinguish (for example) plant versus animal proteins or high-
630 fiber versus low-fiber carbohydrates. Finer-resolution vK mapping zones would increase the precision of
631 results. We could not discern CPP differences for butyrate between sample types using 0.05 vK radii.
632 Butyrate is often present at low concentrations in the gut compared to other SCFAs, with rapid consumption
633 by colonocytes (den Besten et al., 2013). The volatility of butyrate (Wagner et al., 2017) may make it
634 susceptible to loss from soils. Our butyrate $CPP_{density}$ profiles spanning large to small vK radii may depict
635 source-sink dynamics. Compound mapping using O, H, and C content enable exhaustive and
636 compartmented mapping within vK coordinate space, however this represents a simplified, imperfect
637 approach. Multi-element compound mapping would offer increased precision (Rivas-Ubach et al., 2018)

638 and may be developed to provide exhaustive and compartmented mapping across a range of compound
639 types. CPP measures used here were relative, not absolute. Mapping of SUPER-FOCUS functions was
640 incomplete (sample functional relative abundances ranged from 52-84%, with means 55-67%). This may be
641 improved with future algorithm refinement.

642 Further work is required to refine our first-pass ASALR normalization, to examine CPP-disease
643 links in wider geographic/cultural populations, and to explore other potential explanatory variables not
644 considered in our analyses. Like many microbiome studies, our analyses do not permit causal insight to
645 interpret whether increased or decreased CPP may facilitate or follow disease. For example, excessive CPP
646 may produce metabolites at toxic levels, or degrade substrates leading to deficiency. Reduced CPP
647 measures might correspond to dietary deficiencies or suppression of functional pathways due to
648 dysregulated environmental conditions. Nonetheless, observed CPP trends may assist hypothesis-building
649 and prioritizing mechanistic research. We suggest that future work might address these limitations.
650

651 **5. Conclusions**

652 Our assessment of microbial metagenomes using the compound processing potential framework offers a
653 new approach to examine potential overlapping functional capacities between human and environmental
654 microbiomes. Our findings reveal a meaningful bioenergetic basis to the development and differentiation of
655 microbiomes in human health and disease, and with variation in ecosystem quality. We show that
656 ecosystem quality can shape environmental microbiota in seemingly predictable ways, with varying
657 proficiency to process a range of compounds that are associated with human health and disease—as
658 measured by the functional potential relative abundances imprinted in their collective DNA. This work
659 informs the ongoing development of microbiota-oriented approaches in human health, including the
660 potential for nature-based solutions to help address global healthcare challenges.
661
662

Data availability

All data used in this study are available from repositories listed in Table 1. Code used to support this study is available from: https://github.com/liddic/compound_potential

Acknowledgments

We acknowledge the authors of previously published metagenomics datasets used here which enabled comparison of diverse 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), Marsden Fund Fast-Start grant funding (Royal Society Te Apārangi), funding from the German Research Foundation iDiv ([DFG]–FZT 118, 202548816; Ei 862/29-1; Ei 862/31-1; Ei 862/27-1; HE 8266/4-1), and the Australian Research Council (grant numbers LP190100051, LP190100484). We acknowledge the contribution of the Australian Microbiome consortium in the generation of data used in this publication. The Australian Microbiome Initiative is supported by funding from Bioplatforms Australia and the Integrated Marine Observing System (IMOS) through the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS), Parks Australia through the Bush Blitz program funded by the Australian Government and BHP, and the CSIRO.

682 **References**
683

684 Alam MN, Almoyad M, Huq F. Polyphenols in Colorectal Cancer: Current State of Knowledge including
685 Clinical Trials and Molecular Mechanism of Action. BioMed Research International 2018; 2018:
686 4154185.
687 Albenberg LG, Wu GD. Diet and the Intestinal Microbiome: Associations, Functions, and Implications for
688 Health and Disease. Gastroenterology 2014; 146: 1564-1572.
689 Andrews S. FastQC - a quality control application for high throughput sequence data.
690 http://www.bioinformatics.babraham.ac.uk/projects/fastqc, 2018.
691 Avera BN, Strahm BD, Burger JA, Zipper CE. Development of ecosystem structure and function on
692 reforested surface-mined lands in the Central Appalachian Coal Basin of the United States. New
693 Forests 2015; 46: 683-702.
694 Averill C, Anthony MA, Baldrian P, Finkbeiner F, Hoogen Jvd, Kiers T, et al. Defending Earth's terrestrial
695 microbiome. Nature Microbiology 2022; 7: 1717-1725.
696 Banerjee S, van der Heijden MGA. Soil microbiomes and one health. Nature Reviews Microbiology 2022.
697 Barnes A, Eisenhauer N, Wallace KJ, Rai B, Mitchell G, Heintz-Buschart A. People Cities and Nature
698 (PCaN) pilot metagenomics dataset. In: Waikato Uo, editor, Aotearoa New Zealand, 2020.
699 Barton LL, Northup DE. Microbial Ecology: Wiley-Blackwell, 2011.
700 Belizario JE, Napolitano M. Human microbiomes and their roles in dysbiosis, common diseases, and novel
701 therapeutic approaches. Frontiers in Microbiology 2015; 6.
702 Bellinge JW, Dalgaard F, Murray K, Connolly E, Blekkenhorst LC, Bondonno CP, et al. Vitamin K Intake
703 and Atherosclerotic Cardiovascular Disease in the Danish Diet Cancer and Health Study. Journal of
704 the American Heart Association 2021; 10: e020551.
705 Bissett A, Fitzgerald A, Meintjes T, Mele PM, Reith F, Dennis PG, et al. Introducing BASE: the Biomes of
706 Australian Soil Environments soil microbial diversity database. GigaScience 2016; 5: 21.
707 Bowman DMJS, Murphy BP, Burrows GE, Crisp MD. Fire regimes and the evolution of the Australian
708 biota. CSIRO Publishing, Collingwood, Vic, 2012, pp. 27-47.
709 Brame JE, Liddicoat C, Abbott CA, Breed MF. The potential of outdoor environments to supply beneficial
710 butyrate-producing bacteria to humans. Science of The Total Environment 2021; 777: 146063.
711 Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND.
712 Nature Methods 2021; 18: 366-368.
713 Burstyn I, Kromhout H, Partanen T, Svane O, Langård S, Ahrens W, et al. Polycyclic Aromatic
714 Hydrocarbons and Fatal Ischemic Heart Disease. Epidemiology 2005; 16: 744-750.
715 Calabrese EJ, Bachmann KA, Bailer AJ, Bolger PM, Borak J, Cai L, et al. Biological stress response
716 terminology: Integrating the concepts of adaptive response and preconditioning stress within a
717 hormetic dose–response framework. Toxicology and Applied Pharmacology 2007; 222: 122-128.
718 Cao D, Pizzorno G. Uridine phosophorylase: an important enzyme in pyrimidine metabolism and
719 fluoropyrimidine activation. Drugs of Today 2004; 40: 431-443.
720 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;
721 34: i884-i890.
722 Chen Y, Liu Y, Li X, Zhang J, Li G. Lignin Interacting with α-glucosidase and its Inhibitory Effect on the
723 Enzymatic Activity. Food Biophysics 2015; 10: 264-272.
724 Cunningham SC, Cavagnaro TR, Mac Nally R, Paul KI, Baker PJ, Beringer J, et al. Reforestation with
725 native mixed-species plantings in a temperate continental climate effectively sequesters and
726 stabilizes carbon within decades. Global Change Biology 2015; 21: 1552-1566.
727 D'Andrilli J, Cooper WT, Foreman CM, Marshall AG. An ultrahigh-resolution mass spectrometry index to
728 estimate natural organic matter lability. Rapid Communications in Mass Spectrometry 2015; 29:
729 2385-2401.
730 Daisley BA, Koenig D, Engelbrecht K, Doney L, Hards K, Al KF, et al. Emerging connections between gut
731 microbiome bioenergetics and chronic metabolic diseases. Cell Reports 2021; 37: 110087.

732 Delgado-Baquerizo M, Reich PB, Trivedi C, Eldridge DJ, Abades S, Alfaro FD, et al. Multiple elements of
733 soil biodiversity drive ecosystem functions across biomes. Nature Ecology & Evolution 2020; 4:
734 210-220.
735 Delgado-Baquerizo M, Reith F, Dennis PG, Hamonts K, Powell JR, Young A, et al. Ecological drivers of
736 soil microbial diversity and soil biological networks in the Southern Hemisphere. Ecology 2018; 99:
737 583-596.
738 den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud D-J, Bakker BM. The role of short-chain
739 fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. Journal of
740 Lipid Research 2013; 54: 2325-2340.
741 Deng H, Callender R, Dale GE. A Vibrational Structure of 7,8-Dihydrobiopterin Bound to
742 Dihydroneopterin Aldolase. Journal of Biological Chemistry 2000; 275: 30139-30143.
743 Diener C, Gibbons SM, Resendis-Antonio O. MICOM: Metagenome-Scale Modeling To Infer Metabolic
744 Interactions in the Gut Microbiota. mSystems 2020; 5: 10.1128/msystems.00606-19.
745 Douglas G. Kondo Y-HS, Tracy L. Hellem, Kristen K. Delmastro, Eun-Kee Jeong, Namkug Kim, Xianfeng
746 Shi, and Perry F. Renshaw. Open-Label Uridine for Treatment of Depressed Adolescents with
747 Bipolar Disorder. Journal of Child and Adolescent Psychopharmacology 2011; 21: 171-175.
748 Eisenstein M. The hunt for a healthy microbiome. Nature 2020; 577: S6-S8.
749 Erb M, Kliebenstein DJ. Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites:
750 The Blurred Functional Trichotomy. Plant Physiology 2020; 184: 39-52.
751 Essa H, Nicholas K, Jean-Pierre I, Leon JS, Anas S, Omar A, et al. Vitamin K2—a neglected player in
752 cardiovascular health: a narrative review. Open Heart 2021; 8: e001715.
753 Fierer N. Embracing the unknown: Disentangling the complexities of the soil microbiome. Nat Rev
754 Microbiol 2017; 15: 579-590.
755 Flandroy L, Poutahidis T, Berg G, Clarke G, Dao M-C, Decaestecker E, et al. The impact of human
756 activities and lifestyles on the interlinked microbiota and health of humans and of ecosystems.
757 Science of The Total Environment 2018; 627: 1018-1038.
758 Flannery JE, Stagaman K, Burns AR, Hickey RJ, Roos LE, Giuliano RJ, et al. Gut Feelings Begin in
759 Childhood: the Gut Metagenome Correlates with Early Environment, Caregiving, and Behavior.
760 mBio 2020; 11: e02780-19.
761 Flinders_University. DeepThought (HPC). Retrieved from
762 https://doi.org/10.25957/FLINDERS.HPC.DEEPTHOUGHT, 2021.
763 Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, et al. Disentangling type 2
764 diabetes and metformin treatment signatures in the human gut microbiota. Nature 2015; 528: 262-
765 266.
766 Frier BM, Schernthaner G, Heller SR. Hypoglycemia and Cardiovascular Risks. Diabetes Care 2011; 34:
767 S132-S137.
768 Garza DR, van Verk MC, Huynen MA, Dutilh BE. Towards predicting the environmental metabolome from
769 metagenomics with a mechanistic model. Nature Microbiology 2018; 3: 456-460.
770 Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human
771 microbiome. Nature Medicine 2018; 24: 392.
772 Gu X, Ke S, Wang Q, Zhuang T, Xia C, Xu Y, et al. Energy metabolism in major depressive disorder:
773 Recent advances from omics technologies and imaging. Biomedicine & Pharmacotherapy 2021;
774 141: 111869.
775 Hammer KA, Carson CF, Riley TV. Antimicrobial activity of essential oils and other plant extracts. Journal
776 of Applied Microbiology 1999; 86: 985-990.
777 Handsley-Davis M, Kapellas K, Jamieson LM, Hedges J, Skelly E, Kaidonis J, et al. Heritage-specific oral
778 microbiota in Indigenous Australian dental calculus. Evolution, Medicine, and Public Health 2022;
779 10: 352-362.
780 Henning Tiemeier, M.D. ,, H. Ruud van Tuijl, M.Sc. ,, Albert Hofman, M.D., Ph.D. ,, John Meijer, M.Sc. ,,
781 Amanda J. Kiliaan, Ph.D. , and, Monique M.B. Breteler, M.D., Ph.D. Vitamin B12, Folate, and

Homocysteine in Depression: The Rotterdam Study. American Journal of Psychiatry 2002; 159: 2099-2101.
Hinsinger P, Bengough AG, Vetterlein D, Young IM. Rhizosphere: biophysics, biogeochemistry and ecological relevance. Plant and Soil 2009; 321: 117-152.
Hou Q, Ufer G, Bartels D. Lipid signalling in plant responses to abiotic stress. Plant, Cell & Environment 2016; 39: 1029-1048.
Hu Y, Zheng Q, Zhang S, Noll L, Wanek W. Significant release and microbial utilization of amino sugars and d-amino acid enantiomers from microbial cell wall decomposition in soils. Soil Biology and Biochemistry 2018; 123: 115-125.
Husson O. Redox potential (Eh) and pH as drivers of soil/plant/microorganism systems: a transdisciplinary overview pointing to integrative opportunities for agronomy. Plant and Soil 2013; 362: 389-417.
Hutto BR. Folate and cobalamin in psychiatric illness. Comprehensive Psychiatry 1997; 38: 305-314.
Janusz G, Pawlik A, Sulej J, Świderska-Burek U, Jarosz-Wilkołazka A, Paszczyński A. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. FEMS Microbiology Reviews 2017; 41: 941-962.
Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. Nature Communications 2017; 8: 845.
Kusunoki K, Ozaki N, Sawada M, Sato T, Hirano S, Narita T. Serum. Levels of Dihydroneopterin and Soluble Cytokine Receptors in Major Depression. Pteridines 1999; 10: 24-26.
Law M. Dietary fat and adult diseases and the implications for childhood nutrition: an epidemiologic approach. The American Journal of Clinical Nutrition 2000; 72: 1291s-1296s.
Li N, Zhang H, Bai Z, Jiang H, Yang F, Sun X, et al. Soil exposure accelerates recovery of the gut microbiota in antibiotic-treated mice. Environmental Microbiology Reports; n/a.
Li Z-F, Wu N-Q. The Progression of Treatment for Refractory Hypercholesterolemia: Focus on the Prospect of Gene Therapy. Frontiers in Genetics 2022; 13.
Liddicoat C, Bi P, Waycott M, Glover J, Breed M, Weinstein P. Ambient soil cation exchange capacity inversely associates with infectious and parasitic disease risk in regional Australia. Science of The Total Environment 2018; 626: 117-125.
Liddicoat C, Sydnor H, Cando-Dumancela C, Dresken R, Liu J, Gellie NJC, et al. Naturally-diverse airborne environmental microbial exposures modulate the gut microbiome and may provide anxiolytic benefits in mice. Science of The Total Environment 2020; 701: 134684.
Litvak Y, Byndloss MX, Bäumler AJ. Colonocyte metabolism shapes the gut microbiota. Science 2018; 362: eaat9076.
Liu R, Dang S, Yan H, Wang D, Zhao Y, Li Q, et al. Association between dietary protein intake and the risk of hypertension: a cross-sectional study from rural western China. Hypertension Research 2013; 36: 972-979.
Loï C, Cynober L. Glutamate: A Safe Nutrient, Not Just a Simple Additive. Annals of Nutrition and Metabolism 2022; 78: 133-146.
Ma W, Heianza Y, Huang T, Wang T, Sun D, Zheng Y, et al. Dietary glutamine, glutamate and mortality: two large prospective studies in US men and women. International Journal of Epidemiology 2017; 47: 311-320.
Mallah MA, Mallah MA, Liu Y, Xi H, Wang W, Feng F, et al. Relationship Between Polycyclic Aromatic Hydrocarbons and Cardiovascular Diseases: A Systematic Review. Frontiers in Public Health 2021; 9.
Mallick H, Franzosa EA, McLver LJ, Banerjee S, Sirota-Madi A, Kostic AD, et al. Predictive metabolomic profiling of microbial communities using amplicon or metagenomic sequences. Nature Communications 2019; 10: 3136.
McFarlin BK, Henning AL, Venable AS. Oral Consumption of Vitamin K2 for 8 Weeks Associated With Increased Maximal Cardiac Output During Exercise. Altern Ther Health Med 2017; 23: 26-32.

Mota JDO, Guillou S, Pierre F, Membré J-M. Public health risk-benefit assessment of red meat in France: Current consumption and alternative scenarios. Food and Chemical Toxicology 2021; 149: 111994.
Neveu V, Nicolas G, Amara A, Salek RM, Scalbert A. The human microbial exposome: expanding the Exposome-Explorer database with gut microbial metabolites. Scientific Reports 2023; 13: 1946.
Oliai Araghi S, Kiefte-de Jong JC, van Dijk SC, Swart KMA, van Laarhoven HW, van Schoor NM, et al. Folic Acid and Vitamin B12 Supplementation and the Risk of Cancer: Long-term Follow-up of the B Vitamins for the Prevention of Osteoporotic Fractures (B-PROOF) Trial. Cancer Epidemiology, Biomarkers & Prevention 2019; 28: 275-282.
Ottman N, Ruokolainen L, Suomalainen A, Sinkko H, Karisola P, Lehtimäki J, et al. Soil exposure modifies the gut microbiota and supports immune tolerance in a mouse model. Journal of Allergy and Clinical Immunology 2018; 143: 1198-1206.e12.
Owen OE, Kalhan SC, Hanson RW. The Key Role of Anaplerosis and Cataplerosis for Citric Acid Cycle Function. Journal of Biological Chemistry 2002; 277: 30409-30412.
Pearson ER. Type 2 diabetes: a multifaceted disease. Diabetologia 2019; 62: 1107-1112.
Rivas-Ubach A, Liu Y, Bianchi TS, Tolić N, Jansson C, Paša-Tolić L. Moving beyond the van Krevelen Diagram: A New Stoichiometric Approach for Compound Classification in Organisms. Analytical Chemistry 2018; 90: 6152-6160.
Rivera-Chávez F, Lopez CA, Bäumler AJ. Oxygen as a driver of gut dysbiosis. Free Radical Biology and Medicine 2017; 105: 93-101.
Rodriguez-Ramos JC, Cale JA, Cahill Jr JF, Simard SW, Karst J, Erbilgin N. Changes in soil fungal community composition depend on functional group and forest disturbance type. New Phytologist 2021; 229: 1105-1117.
Roslund MI, Parajuli A, Hui N, Puhakka R, Grönroos M, Soininen L, et al. A Placebo-controlled double-blinded test of the biodiversity hypothesis of immune-mediated diseases: Environmental microbial diversity elicits changes in cytokines and increase in T regulatory cells in young children. Ecotoxicology and Environmental Safety 2022; 242: 113900.
Round JL, Mazmanian SK. The gut microbiome shapes intestinal immune responses during health and disease. Nature Reviews. Immunology 2009; 9: 313-323.
Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. Nucleic Acids Research 2020; 49: D575-D588.
Silva GGZ, Green KT, Dutilh BE, Edwards RA. SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. Bioinformatics 2015; 32: 354-361.
Skarbek K, Milewska MJ. Biosynthetic and synthetic access to amino sugars. Carbohydrate Research 2016; 434: 44-71.
Sokol NW, Slessarev E, Marschmann GL, Nicolas A, Blazewicz SJ, Brodie EL, et al. Life and death in the soil microbiome: how ecological processes influence biogeochemistry. Nature Reviews Microbiology 2022.
Sun S, Badgley BD. Changes in microbial functional genes within the soil metagenome during forest ecosystem restoration. Soil Biology and Biochemistry 2019; 135: 163-172.
Sun X, Liddicoat C, Tiunov A, Wang B, Zhang Y, Lu C, et al. Harnessing soil biodiversity to promote human health in cities. npj Urban Sustainability 2023; 3: 5.
Tao J, Li S, Ye F, Zhou Y, Lei L, Zhao G. Lignin – An underutilized, renewable and valuable material for food industry. Critical Reviews in Food Science and Nutrition 2020; 60: 2011-2033.
Vaduganathan M, Mensah GA, Turco JV, Fuster V, Roth GA. The Global Burden of Cardiovascular Diseases and Risk. Journal of the American College of Cardiology 2022; 80: 2361-2371.
van der Putten WH, Bradford MA, Pernilla Brinkman E, van de Voorde TFJ, Veen GF. Where, when and how plant–soil feedback matters in a changing world. Functional Ecology 2016; 30: 1109-1121.

880 Wagner AO, Markt R, Puempel T, Illmer P, Insam H, Ebner C. Sample preparation, preservation, and
881 storage for volatile fatty acid quantification in biogas plants. Engineering in Life Sciences 2017; 17:
882 132-139.
883 Wardle DA, Bardgett RD, Klironomos JN, Heikki S, van der Putten WH, Wall DH. Ecological Linkages
884 Between Aboveground and Belowground Biota. Science 2004; 304: 1629-1633.
885 Wilson JW, Shakir D, Batie M, Frost M, Rocha S. Oxygen-sensing mechanisms in cells. The FEBS Journal
886 2020; 287: 3888-3906.
887 Wu X, Wu L, Liu Y, Zhang P, Li Q, Zhou J, et al. Microbial Interactions With Dissolved Organic Matter
888 Drive Carbon Dynamics and Community Succession. Frontiers in Microbiology 2018; 9.
889 Xu J, Zhu X, Guan G, Zhang Y, Hui R, Xing Y, et al. Non-linear associations of serum and red blood cell
890 folate with risk of cardiovascular and all-cause mortality in hypertensive adults. Hypertension
891 Research 2023; 46: 1504-1515.
892 Yang J, Kalhan SC, Hanson RW. What is the metabolic role of phosphoenolpyruvate carboxykinase? J Biol
893 Chem 2009; 284: 27025-9.
894 Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for
895 early-stage detection of colorectal cancer. Molecular Systems Biology 2014; 10: 766.
896 Zengler K, Zaramela LS. The social network of microorganisms — how auxotrophies shape complex
897 communities. Nature Reviews Microbiology 2018; 16: 383-390.
898 Zhang J, Zou S, Fang L. Metabolic reprogramming in colorectal cancer: regulatory networks and therapy.
899 Cell & Bioscience 2023; 13: 25.
900 Zheng Y, Hu FB, Ruiz-Canela M, Clish CB, Dennis C, Salas-Salvado J, et al. Metabolites of Glutamate
901 Metabolism Are Associated With Incident Cardiovascular Events in the PREDIMED
902 PREvención con DIeta MEDiterránea (PREDIMED) Trial. Journal of the American
903 Heart Association 2016; 5: e003755.
904 Zhou Y, Danbolt NC. Glutamate as a neurotransmitter in the healthy brain. Journal of Neural Transmission
905 2014; 121: 799-817.
906