Title

Predictions of rhizosphere microbiome dynamics with a genome-informed and trait-based energy budget model.

Authors

Zhalnina, Kateryna

Cho, Heejung

Le, Beatrice

et al.

Peer reviewed

Predictions of rhizosphere microbiome dynamics with a genome-informed and trait-based energy budget model

Check for updates

Gianna L. Marschmann [1], Jinyun Tang [1], Kateryna Zhalnina [1,2], Ulas Karaoz [1], Heejung Cho[1,3], Beatrice Le[3], Jennifer Pett-Ridge [4,5] & Eoin L. Brodie [1,6] ✉

Soil microbiomes are highly diverse, and to improve their representation in biogeochemical models, microbial genome data can be leveraged to infer key functional traits. By integrating genome-inferred traits into a theory-based hierarchical framework, emergent behaviour arising from interactions of individual traits can be predicted. Here we combine theory-driven predictions of substrate uptake kinetics with a genome-informed trait-based dynamic energy budget model to predict emergent life-history traits and trade-offs in soil bacteria. When applied to a plant microbiome system, the model accurately predicted distinct substrate-acquisition strategies that aligned with observations, uncovering resource-dependent trade-offs between microbial growth rate and efficiency. For instance, inherently slower-growing microorganisms, favoured by organic acid exudation at later plant growth stages, exhibited enhanced carbon use efficiency (yield) without sacrificing growth rate (power). This insight has implications for retaining plant root-derived carbon in soils and highlights the power of data-driven, trait-based approaches for improving microbial representation in biogeochemical models.

Microbes are major drivers of carbon (C) and nutrient fluxes in Earth's terrestrial ecosystems; however, Earth system models designed to inform climate change adaptation and mitigation strategies have typically not included explicit representation of soil microorganisms, despite mounting evidence that the explicit parameterization of microbial processes improves model prediction and reduces uncertainty in terrestrial systems[1,2]. This lack of representation is rooted in the failure of common organizing principles derived from plant and animal ecology, such as the descendants of r-K selection theory[3] or Grime's competition–stress tolerance–ruderal (C–S–R) framework[4,5],

to fully capture the complexity of microbial systems[6–9]. To accurately predict the sensitivities of soil organic C stocks and plant productivity in response to changing climate conditions, it is crucial to develop a more comprehensive understanding of the role of microbes and their traits. However, the majority of organisms comprising Earth's microbiomes have yet to be cultivated and may never be[10]. This means that our window into the ecology of microorganisms, such as those in soil, remains primarily through the lens of genomic information, which is accumulating far more rapidly than phenotypic information from laboratory isolates[11,12].

[1]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [2]Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [3]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA. [4]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA. [5]Life and Environmental Sciences Department, University of California Merced, Merced, CA, USA. [6]Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, CA, USA. ✉e-mail: elbrodie@lbl.gov

Trait-based models have the potential to represent trait variation by aggregating data from hundreds or thousands of genomes, providing a data-driven approach to organize the complexity of microbial communities with less emphasis on traditional ecological theory and phylogenetic origin[13]. Among the various modelling approaches available for studying microbiomes, trait-based models are an attractive intermediate complexity approach to exploring how the hierarchy of traits interacts to influence the fitness of microorganisms within a community[14]. Yet, despite this promise and potential to scale the representation of microorganisms using a trait-based representation in biogeochemical models, a major challenge remains in their parameterization[15]. Our recent work provides a computational pipeline and a toolset ('microTrait'[16]) to infer microbial traits from genomic data and establish links between each genome-derived trait and ecological strategy at different levels of trait granularity. The resulting information can be used to initialize and parameterize mechanistic trait-based models spanning a hierarchy of structural complexity to explore the drivers of variation in the distribution and co-occurrence of microbial traits[17]. Moreover, incorporating emerging concepts and theory outside of traditional microbiome science, such as thermodynamic and biophysical theory, proves valuable for understanding traits of microorganisms[18]. This integration facilitates the search for generalizable principles or 'rules' that are applicable across diverse microbiome systems[19], furthering theory development, and the elaboration of large-scale predictive models[15].

Here we present a genome-informed, trait-based dynamic energy budget model (DEBmicroTrait; Fig. 1 and Supplementary Table 1). This model integrates genome-predicted traits and their interactions within a dynamic environment, allowing life-history strategies and niches of soil bacteria to emerge from fundamental thermodynamic, biophysical and metabolic principles that constrain trait variation, trait linkages (defined as co-occurrence of traits in the same organism) and ultimately organism fitness. We have initially focused on integrating quantitative genomic traits that distinguish bacteria at a critical soil interface, the rhizosphere (that is, the area surrounding growing plant roots). While quantitative life-history and biophysical traits (such as genome size, maximum growth rate, cell size and rRNA operon (*rrn*) copy number) do not encompass the entirety of potential interactions and traits in bacterial ecology, they contribute substantially to explaining variations in resource utilization and provide ecological insights into species competitiveness at different stages of succession or resource depletion[20,21]. These life-history traits interact with bacterial preferences for diverse types of substrates, which can be predicted from genome sequences[22,23]. The growing development of large collections of sequenced rhizosphere bacterial genomes[22,24], coupled with high-throughput metabolomics methods[25], enables integration of knowledge connecting root exudate dynamics and microbial metabolism into predictive computational models of plant–microbe interactions[26].

The rhizosphere is chemically diverse and a critical hotspot for biogeochemical transformation with high potential for C stabilization through microbial C assimilation and subsequent mineral-surface stabilization[27]. To better understand interactions between life-history traits, biophysical traits and bacterial substrate preferences, we simulated the growth of 39 soil bacteria on 82 plant exudate metabolites. We developed a theory-based approach to estimate microbial substrate uptake parameters using only genome- and substrate-derived traits. We simulated population-level estimates of key traits, including realized growth rate and C assimilation rate, across different substrate classes that are known rhizosphere exudates, and benchmarked simulations quantitatively and qualitatively against observations. We identified microbial growth strategies that arise from the multivariate trait combinations in the absence of soil matrix effects in this case. These strategies emerge from generalizable rules describing how bacterial traits interact with plant exudation traits and are defined by the cellular trade-offs

associated with acquiring and assimilating diverse substrates. Most importantly, these strategies can be explained by trade-offs between growth rate (power) and growth efficiency (yield) that impose strong constraints on microbial community composition in natural habitats[28]. Furthermore, these trade-offs have implications for the formation of soil organic matter via the microbial route[29].

Results

Growing roots dramatically alter the chemical and physical habitat for microorganisms and exude photosynthesis-derived C that can be broadly classified into sugars, amino acids, organic acids, fatty acids, nucleotides and auxins[30]. In fact, root exudation might be a generalizable trait with rates and composition that can be predicted from plant functional traits, with consistent temporal patterns across plant development stages[31–33]. The chemical composition of root exudates interacts with microbial substrate preferences, which are predictable from genomic traits (Fig. 2). Gene annotations assess functional potential—the capacity for organisms to perform. These soil microorganisms differ in genomic traits related to resource acquisition, particularly with regard to their metabolic potential to utilize organic acids (Fig. 2b) and their potential for plant polymer degradation using glycoside hydrolase enzymes (Fig. 2c). Contrary to expectations, we previously found that both fast-growing as well as slow-growing strains are enriched in the rhizosphere (both observed and predicted on the basis of genomic signatures[22,34,35]; Fig. 2a). The presence of potentially distinct life-history strategies challenges the long-standing assumption that living roots select for fast-growing r-strategists that compete for a small set of labile C substrates[36]. With the caveat that it is a simplification of the complex resource requirements of microorganisms, we assumed that bacteria in the rhizosphere are primarily limited by organic C inputs and energy availability. Furthermore, we hypothesized that substrate preference and substrate utilization efficiency interact to confer a growth efficiency-based fitness advantage for many bacteria in the rhizosphere. However, growth efficiency can vary widely with limiting resource concentration and the free energy content of chemical compounds that are released by plants[28]. These patterns are overlaid with physiological variations in resource use between bacteria[37–39] that probably affect phenotypic trade-offs such as C use efficiency (CUE), suggesting that these factors combine to influence rhizosphere microbial community dynamics.

Substrate preference in the rhizosphere

Microbial strains vary in their substrate uptake affinities, which in turn influences the composition of the root microbiome through ecological processes such as niche differentiation and competitive exclusion[25]. Recent theoretical advancements based on an analytical approximation of how diffusive substrates are intercepted by microbial cells in soil have allowed the derivation of testable relationships between substrate uptake kinetic parameters, including maximum specific reaction rates (V_{max}) and binding half-saturation constants (also known as substrate affinity parameters, K), while also taking into account the biophysical, metabolic and life-history traits that influence substrate demands of soil bacteria[40]. To synthesize these relationships, we considered genomic traits that provide constraints on the interactions between cell size, cellular C density, cell surface area-to-volume ratio and growth rate potential. We used the substrate uptake potential required to support a given genome-inferred maximum specific growth rate as an objective function and investigated the variation and allometric scaling of both kinetic parameters in the equilibrium chemistry approximation (ECA) for substrate uptake[41]. The estimated substrate binding site densities (that is, transporters) were benchmarked against existing data on nutrient uptake (Supplementary Fig. 1a and Table 4) and subsequently distributed across substrate classes on the basis of the relative gene frequencies of specific transporter genes (Fig. 2b). This approach allows for the consideration of the evolutionary history of microbial substrate
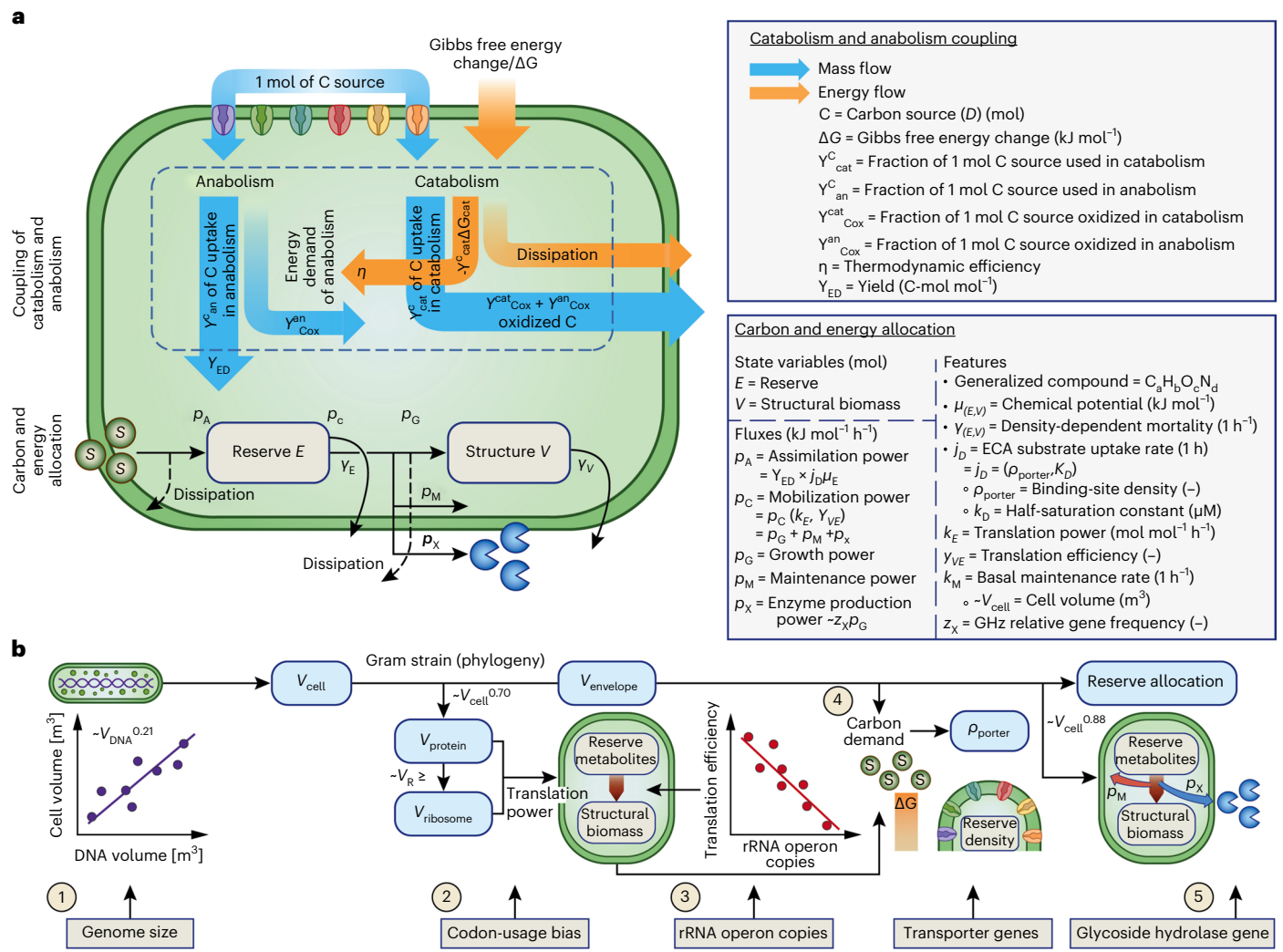
Fig. 1 | Overview of DEBmicroTrait. a, Schematic of the DEBmicroTrait model showing dynamic energy budget (DEB) allocation for a single-reserve (E), single-structure (V) heterotrophic microorganism[67] feeding on different substrates (S). Diffusion-limited substrate uptake occurs through specific substrate binding sites (coloured according to substrate chemical class). Substrate uptake kinetics are described by the ECA[40]. Reserve and structural biomass are conceptualized as generalized chemical compounds characterized by macromolecular composition ($C_aH_bO_cN_d$) and chemical potential ($\mu_{(E,V)}$). Top: the coupling of catabolism and anabolism, that is, the catabolic and anabolic reactions through which energy is obtained and utilized for a metabolism in which the C source is also used as the electron donor[50]. The coupling between catabolism and anabolism exists both in reserve assimilation and structural (and extracellular enzyme) synthesis. Bottom: the sequential assimilation (p_A), partitioning and dissipation of substrate and reserve compounds, with maintenance (p_M) taking priority over growth (p_G) and extracellular enzyme production (p_X). The turnover of reserve and structure ($\gamma_{(E,V)}$) is density-dependent[75]. Essential fluxes are labelled and defined (see also Supplementary Table 1). **b**, Workflow combining biophysical theory and genome inference to constrain DEBmicroTrait model parameters: (1) Cell size covaries with genome size[76]. (2) Codon-usage bias sets an upper bound on protein translation power (k_E[72]). (3) The number of ribosomal RNA operons predicts translation efficiency (y_{VE}[77]). (4) The cellular composition influences C supply and demand, which in turn determines the substrate binding site density required to enable substrate uptake at a rate commensurate with the maximum specific growth rate (ρ_{porter}[84]). Binding sites can be allocated according to relative gene frequencies of transporter genes in the genome (z_p). (5) Basal maintenance rate is proportional to cell volume (k_M[73]). Glycoside hydrolase gene frequencies scale the constitutive extracellular enzyme production rate (z_X[71]).

preference. Within each substrate class, the genome-inferred values fall within the range of variation observed in published experimental data (Supplementary Fig. 1b and Data 1). For amino acids and sugars, the differences between the values reported for cultured organisms in the literature and our predictions were statistically indistinguishable ($P > 0.05$). In contrast, for organic acids, we observed statistically significant deviations from the full data set ($P = 0.04$).

The predicted uptake kinetic parameters vary widely due to factors such as substrate diffusivity, cell size and biomass-specific accessible substrate binding sites in the rhizosphere (Fig. 3). Soil bacteria previously defined as responding negatively to root growth[22] may actually achieve significantly higher maximum specific uptake rates to match

their higher genome-predicted maximum specific growth rates (Fig. 3a). However, their reduced ability to uptake organic acids and auxins that can be prominent in root exudates (Fig. 2b) outweighs their higher maximum growth potential (Fig. 2a). Overall, the optimization of V_{max} and K modulates the competitiveness of rhizosphere organisms according to the placement of their uptake strategies across a concave trade-off curve (Extended Data Fig. 1a). This concave shape allows for the maintenance of optimal substrate uptake phenotypes, enabling organisms to thrive under varying external substrate concentrations[42]. Amino acids and sugars are generally considered to be the most abundant classes in terms of the total amount exuded[43]. Our analysis predicts that the affinity constants of these rhizosphere bacteria for amino acids are
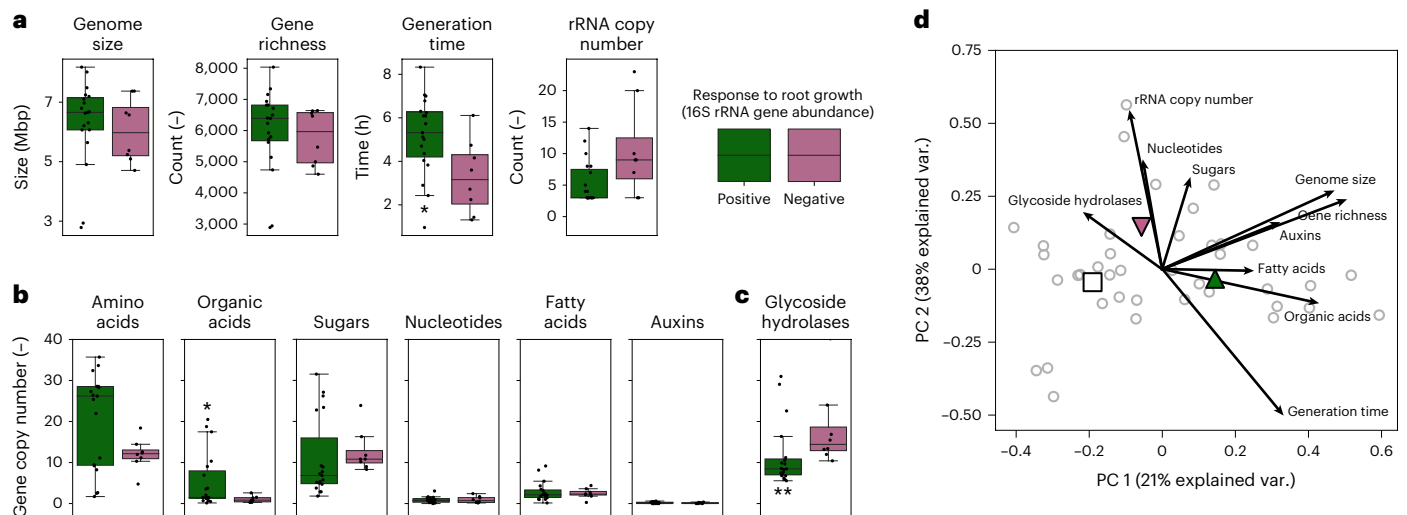
Fig. 2 | Genomic trait distributions of soil bacterial isolates classified on the basis of the response to *Avena barbata* root growth. Modified from ref. 22.
a, Genome size, gene richness, minimum generation time and rRNA operon copy number predicted from genome sequences of soil isolates.
b, Monomer transporters. **c**, Glycoside hydrolase enzymes. Gene copy numbers for monomer transporters and glycoside hydrolases were normalized by genome size. Differences in genomic trait distributions between bacteria that responded positively and negatively to root growth were evaluated using the Kruskal–Wallis one-way analysis of variance (ANOVA) (Supplementary Data 1), and traits with significant differences are annotated according to the R project standard

convention[85] used throughout the paper (***$P \le 0.001$, **$P \le 0.01$, *$P \le 0.05$). Isolate response groups were classified on the basis of changes in 16S rRNA gene abundance over the plant (*A. barbata*) developmental stages. In each boxplot, a point denotes a single isolate. The top and bottom of each box represent the 25th and 75th percentiles, the horizontal line inside each box represents the median and the whiskers represent the range of the points excluding outliers. Positive responders $n = 19$, negative responders $n = 8$. **d**, Principal component analysis illustrating covariations among genomic traits shown in **a**–**c**. Highlighted symbols represent average coordinates of positive, negative and undefined (open square) ($n = 12$) isolate response groups.

nearly optimal at micromolar concentrations (Extended Data Fig. 1b). In addition, we predict that increasing substrate affinity by increasing the substrate binding site density beyond ~0.1% of the cell surface area provides little benefit to soil microorganisms (Supplementary Fig. 2b). This finding contrasts with well-mixed systems at equilibrium, such as oligotrophic marine systems, where the activity of loaded binding proteins monotonically increases affinity and specific uptake rate at low substrate concentrations[44]. It highlights that the ecological trade-offs that define fitness in soil may differ from those that define life-history dichotomies in other habitats[45]. We observe that the substrate affinity constant that maximizes the specific uptake rate at low concentrations is closest to the average value for organic acids (0.26 μM; Extended Data Fig. 1b), suggesting that certain soil bacteria have evolved specialized affinity uptake systems to colonize specific metabolic niches in soil.

However, resource acquisition traits are only one facet of ecological strategy variation, and interactions with cellular resource allocation strategies are key. To address this, we analysed biomass production (BP) and respiration (BR) rates of 39 soil bacteria growing on 82 root metabolites across combined 3,198 batch simulations (Fig. 4) to explore relationships between realized growth rate and yield (that is, CUE), calculated as BP/(BP+BR).

Power–yield signatures in bacterial rhizosphere succession
Microbial access to soil C, as well as species-specific differences in the energetic demands required to use different substrate classes, select for distinct growth regimes in the rhizosphere (Fig. 4a). While growth rates in the low growth rate regime (0.0044–0.039 h^{-1}) are reflective of growth rates of autochthonous bacteria occurring in unamended soils[46], the growth rates in the high growth rate regime span typical values observed in pure cultures under laboratory conditions (0.039–0.46 h^{-1}). Maximum realized growth rates were partially confirmed with measured growth rates at C concentrations representative of the original growth media used to cultivate the bacterial isolates ($r^2 = 0.85$; Supplementary Fig. 3 and Table 5). Potential CUE values range

from 0.07 to 0.74 with a median of 0.49, suggesting that on average approximately half of the consumed C is typically lost via respiration (Fig. 4b). Bacterial traits related to the assimilation of C had the largest influence on model-simulated CUE (Supplementary Fig. 4). A similar amount of variation in CUE was explained by phylogeny at class level (20%) relative to substrate class (15%; Supplementary Table 2), indicating that interactions between substrate traits (for example, mean differences in molecular size and nominal oxidation state of C among substrate classes) and microbial traits (for example, transporter gene frequencies, protein synthesis efficiency, relative maintenance costs) provide a strong foundation for this aspect of soil bacterial ecology[47,48].

Bacterial growth rate and CUE trade off during growth on substrates of high bioavailability (defined by solubility and hydrophobicity[49]; Supplementary Data 1) exuded early during plant growth (sugars, amino acids), highlighting an early successional growth strategy where power is optimized over yield ($F_{1,3} = 13.83$, $r^2 = 0.82$, $P = 0.03$; Fig. 4c). Maximum CUE is achieved for those bacteria that responded positively to root growth when simulated with glucose as the substrate, while maximum growth rates occur at suboptimal CUE for organisms with the fastest growth rates. The rate–yield trade-off emerges from thermodynamic constraints reflecting energy generation and transfer from catabolism to anabolism, resulting in decreasing structural biomass yield to accommodate increasing protein synthesis rates (Supplementary Fig. 5), as well as increased maintenance costs at high growth rates. Energy dissipation increases with the rate of C uptake[50], resulting in lower thermodynamic efficiency for growth on amino acids as compared with sugars, and an overall disproportional scaling between biomass production and respiration rates (with slope that is significantly different from one, $\beta \in (0.56, 0.63)$, $t_{1,437} = -24$, $P < 1 \times 10^{-5}$; Fig. 4d). Fast growth is constrained by internal substrate limitation caused by the accelerated dilution of storage compounds due to (volume) growth, indicating that these bacteria grow faster than they can assimilate new substrate during early rhizosphere successional growth stages. The corresponding realized growth rates are strongly correlated
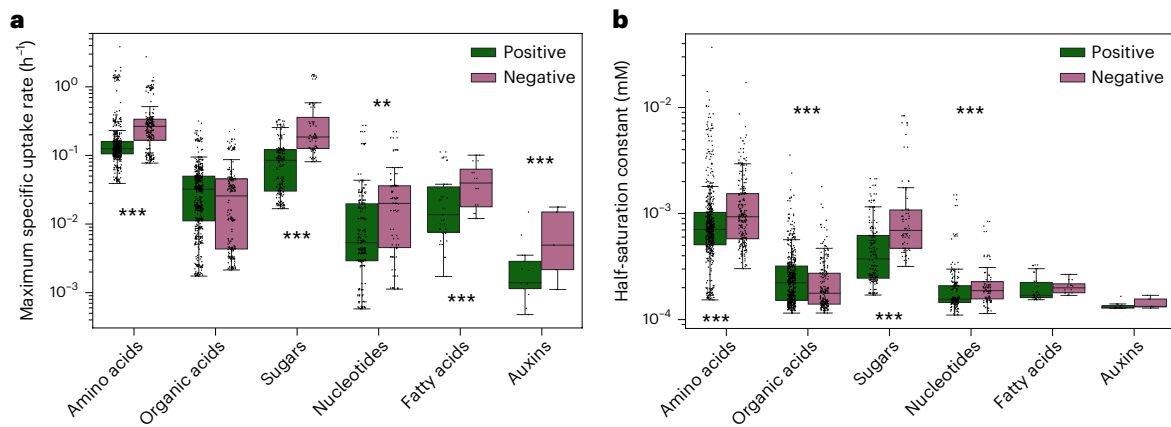
Fig. 3 | Theory-predicted distributions of root exudate metabolite uptake parameters of soil bacterial isolates. a, Maximum specific uptake rate. **b**, Half-saturation constant. Differences in uptake trait distributions between bacteria that responded positively or negatively to root growth were evaluated using the Kruskal–Wallis one-way ANOVA (Supplementary Data 1). In each boxplot, a point denotes a single substrate–consumer relation in the ECA[40]. The top and bottom of each box represent the 25th and 75th percentiles, the horizontal line inside each box represents the median and the whiskers represent the range of the points excluding outliers. Substrates $n = 82$, consumers $n = 27$.

with the number of *rrn* copies in the genome (Supplementary Table 3), where this single genomic trait explains about a third of the variation ($r^2 = 0.30$). This supports the hypothesis that genomic traits related to maximum growth potential are good predictors of growth responses to initial resource pulses in the rhizosphere[46].

The contribution of bacteria to soil C cycling in the rhizosphere is primarily determined by anabolic processes[51] that control the allocation of C and energy for maintenance, the synthesis of storage compounds, as well as extracellular enzyme production and assimilation. These processes represent nearly independent (orthogonal) axes of variation (Fig. 4e), providing compelling evidence for the ecological classification of rhizosphere bacteria into the yield–acquisition–stress (Y–A–S) framework of life-history strategies[5]. Within this framework, bacteria that responded positively to root growth could be considered Y-strategists, actively exploiting available low-molecular-weight C sources through direct capture and assimilation. However, these bacteria also showed greater latency in their growth response (Kruskall–Wallis test, $P < 1 \times 10^{-16}$), suggesting that they may be assimilating additional C in the rhizosphere that has been transformed by previous microbial processing. Interestingly, the absence of a trade-off between niche breadth and yield (Extended Data Fig. 2) indicates a high degree of resource specialization among these organisms. Niche differentiation is further facilitated by resource acquisition strategies of negative responders (A-strategists), which appear to be better adapted to the breakdown of root polymeric carbohydrates via the constitutive expression of carbohydrate-active enzymes[52]. Both Y- and A-strategies align with functional differences encoded in genomes along the copiotroph–oligotroph continuum. Copiotrophs frequently possess a greater prevalence of gene families involved in transcription, transport and metabolism of carbohydrates and amino acids, as well as carbohydrate-active enzymes such as glycoside hydrolases and polysaccharide lyases. This overall strategy enables copiotrophs to rapidly acquire nutrients and produce proteins[35].

Across time, we find that yield and resource acquisition strategies are tightly linked along gradients of resource availability, as bacteria grow more slowly on root exudates that are released during later plant developmental stages (organic acids, fatty acids, nucleosides, auxins). For slower-growing bacteria, only 4% of the variance in growth rates is explained by *rrn* copy number and genome size (Supplementary Table 3), consistent with findings that there is essentially no selective advantage to optimizing translational power via codon usage at low growth rates[35]. The observed proportional scaling between biomass production and respiration (with slope $\beta \in (0.95, 1.05)$, $t_{1,332} = 0.06$, $P = 0.95$; Fig. 4d) is consistent with an overall oligotrophic strategy of energy production, conversion and cell maintenance. Indeed, basal maintenance requirements and extracellular enzyme production explain most of the variability in the growth strategies of isolates at low growth rates (Fig. 4f). The ecological strategies of bacteria that responded negatively to root growth or showed no response to root growth coalesced along both principal components, implying that succession in the rhizosphere is accompanied by a significant decrease in functional diversity, which is consistent with previous observations[53]. As the plant matures, the soil surrounding roots harbours more conservative, slower-growing organisms that do not exhibit a trade-off between growth rate and efficiency since turnover is slower for these organisms ($F_{1,11} = 7.74$, $r^2 = 0.41$, $P = 0.02$; Fig. 4c). Taken together with CUE values (0.07–0.74) that span almost the whole range of values typically observed in soil (Fig. 4a)[54], we hypothesized that selection for efficiency via substrate preference is a primary driver of rhizosphere community composition during the later stages of plant development.

Rhizosphere carbon stabilization via resource specialization

To test whether bacterial preference for specific substrates could interact with CUE to confer a selective advantage in the rhizosphere, we analysed simulations of growth on mixed media (a combination of 82 exudate substrates) for differences in substrate uptake across the positive and negative rhizosphere responder groups. We found that 39 of the 82 exudate substrates show substantial differences in substrate uptake between bacteria that responded positively or negatively to root growth (Fig. 5). For 16 of the 39 root exudates that had previously been identified experimentally, the largest cumulative differences in substrate uptake by bacteria that responded positively to root growth were for plant hormones (indole-3 acetic acid, abscisic acid), followed by a cluster of aromatic organic acids (caffeic, shikimic, 3-dehydroshikimic, *trans*-cinnamic, salicylic, nicotinic). Nucleosides, on the other hand, were more preferentially consumed by the bacteria that responded negatively to root growth. Differences in uptake of 16 out of the 39 substrates agree qualitatively with uptake from the growth medium measured by liquid chromatography–mass spectrometry (LC–MS), without accounting for interactions with the soil matrix[22]. Furthermore, bacteria that responded positively to root growth had a 39% higher growth efficiency on average for organic acids with aromatic rings than those isolates that responded negatively to growing roots (Fig. 5 inset).
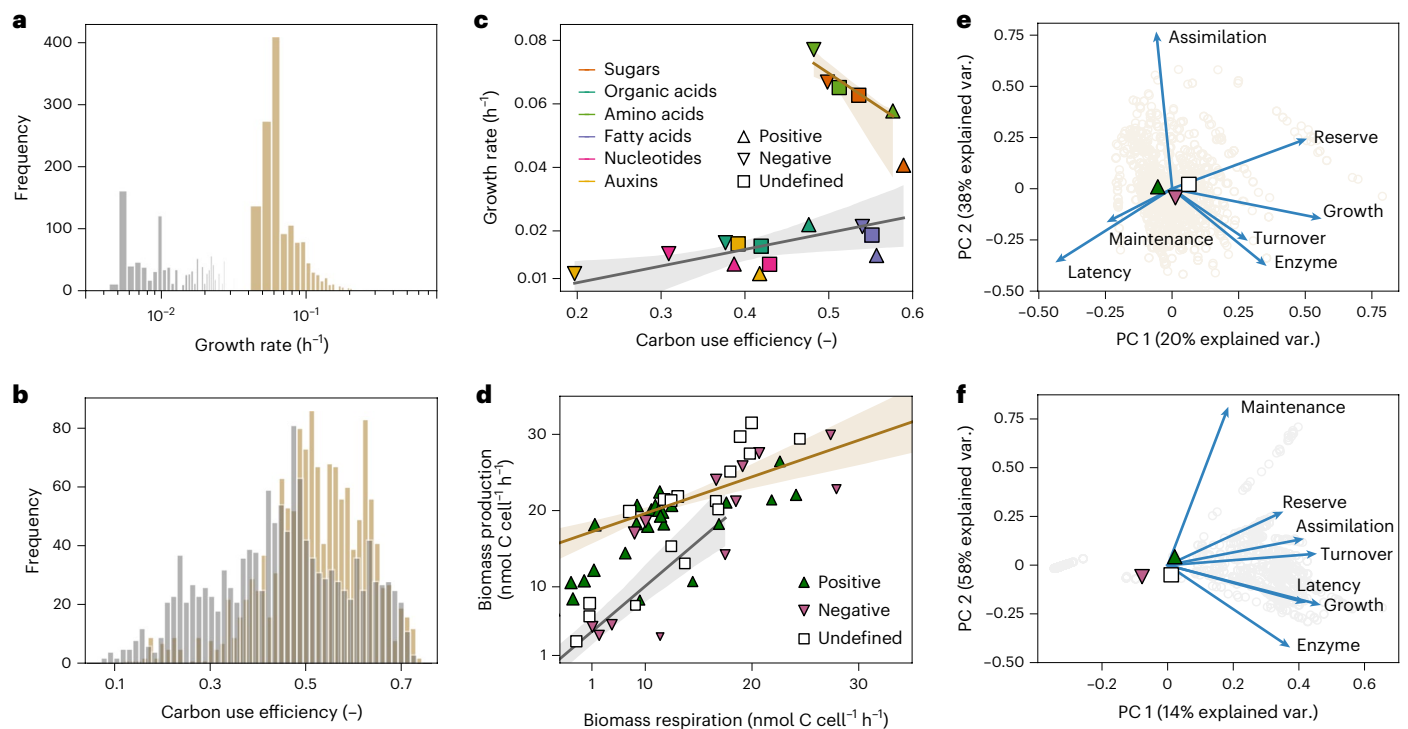
Fig. 4 | Phenotypic traits and trade-offs during batch growth on root exudate metabolites. a,b, Histogram of predicted growth rate (**a**) and predicted CUE (**b**) across isolates and root exudate metabolites in DEBmicroTrait batch simulations. Throughout plots **a**–**f**, two distinct growth regimes are distinguished by brown (high growth regime) and grey (low growth regime) colours on the basis of the bimodal growth rate distribution shown in **a. c,** Relationships between realized growth rate and carbon use efficiency. Median trait values are plotted using different colours and shape depending on substrate class (sugars, organic acids, amino acids, fatty acids, nucleotides, auxins) and isolate response to plant root growth (positive, negative, undefined). **d,** Relationships between biomass production and respiration rates. Symbol size is scaled by carbon use efficiency. In **c** and **d**, solid lines indicate the observed regression lines, while shaded areas indicate the 95% confidence bands. **e,f,** Principal component analysis illustrating covariations among modelled fluxes delineating growth strategies of isolates at high (**e**) and low (**f**) realized growth rates. Highlighted symbols represent average coordinates of positive, negative and undefined isolate response groups. The significance threshold for two-sided *P* values was set at 0.05. The principal components of the different response groups are significantly different ($H_{PC1}(2) = 329$, $H_{PC2}(2) = 105$, $P = 2.2 \times 10^{-16}$). Dunn's test with Benjamini–Hochberg correction confirmed the significant differences between positive and negative responders along PC1 (PC2) ($P = 2 \times 10^{-16}$ (9.3×10^{-4})), and negative and undefined responders ($P = 1.6 \times 10^{-5}$ (2×10^{-16})) in the high growth regime. In the low growth regime, the ecological strategies of negative and undefined responders coalesce along both principal components ($P = 0.43$ (0.09)). They differ significantly from positive responders ($P = 2.5 \times 10^{-5}$). Substrates $n = 82$, consumers $n = 39$, simulations $n = 3198$.

Discussion

Root–microbial interactions have significant effects on the soil C cycle, altering the amount and types of organic matter that become associated with mineral surfaces. Focusing on recent studies that have identified direct predictive links between plant exudate composition and rhizosphere community assembly, and a theory-based approach to predict microbial substrate uptake kinetic traits directly from genome sequences, we synthesized a suite of genome-derived traits into model-based predictions of life-history strategies for a set of soil bacteria. We found that interacting microbial traits (maximum specific growth rate, substrate uptake kinetics, ribosome biosynthesis potential and extracellular enzyme synthesis) have additional interactions with the dynamics of root exudate chemistry, creating emergent patterns of bacterial C use efficiency. These combinations of traits manifest as life-history strategies and have consequences for the path that small molecules take on the way to becoming stabilized soil organic matter.

Root exudates can follow distinct pathways before integrating into the soil matrix: direct sorption to mineral surfaces or microbial trans-formation[55]. While root-derived compounds may rapidly exchange with mineral surfaces[56], it is now widely recognized that the direct microbial transformation of labile photosynthate C into stabilized forms allows microorganisms to contribute disproportionately to persistent C in the soil[29,57,58]. On the basis of our results, we propose

that multidimensional trait interactions influence the pathways of mineral-associated soil organic matter (SOM) formation throughout the plant growing season (Fig. 6). As microbial composition and abundance change over the growing season, the initial exudation of sugars and amino acids with weak sorption to mineral surfaces coincides with power-optimized growth strategies (resulting in more C overall into microbes). C inputs from sugars and amino acids stimulate the growth of resource acquisition (A-) strategists and may accelerate the decomposition of surrounding organic residues, mostly of plant origin, as previously observed in this study system[52,59]. The growth response of individuals is highly dynamic, with fast growth correlating with high density-dependent turnover of microbial biomass (Fig. 4e). This rapid initial turnover of biomass during the vegetative phase implies that secondary processing, fuelled by the products of microbial anabolism, comprises an important component of the C that cycles through rhizosphere communities[60]. While the metabolic products of microbes subsisting on glucose or amino acids earlier in the growing season may contribute to direct mineral-surface stabilization[61], the subsequent emergence of more yield-optimized guilds (more C per specific microbe) may selectively enhance the mineral stabilization of compounds. This enhancement occurs through the deposition of senesced microbial biomass containing greater proportions of lipids and proteins and fewer aromatics[62]. Therefore, these pathways may represent two distinct routes to mineral stabilization of C—which of these
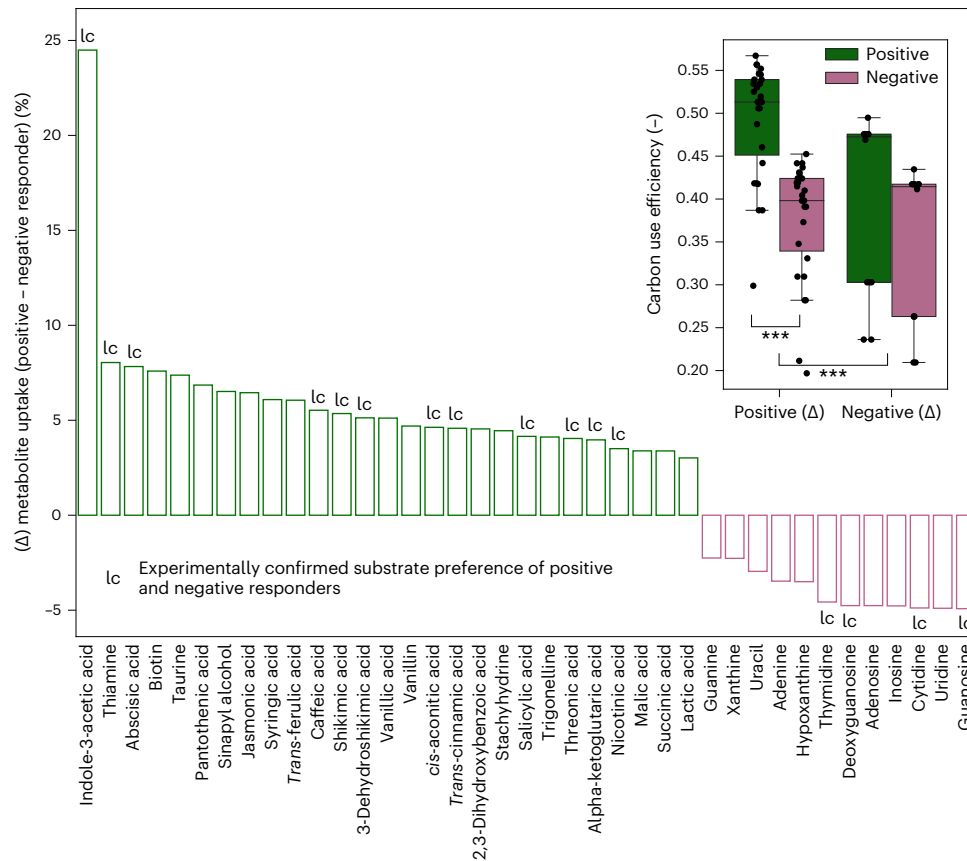
Fig. 5 | Substrate preference and carbon use efficiency in mixed root exudate medium. Substrates with the largest differences in uptake are shown on the *x* axis (*n* = 39). Each bar corresponds to differences in substrate uptake quantified as a percentage of the difference in substrate depletion from the medium. Substrate uptake preferences that were confirmed experimentally[22] are denoted by the letters 'lc'. Inset: differences in carbon use efficiency for substrates that were preferentially consumed by bacteria that responded positively (positive (Δ), *n* = 27) and negatively (negative (Δ), *n* = 12) to root growth. Differences in carbon use efficiency within (χ^2 = 9.6, *P* = 0.002) and between (χ^2 = 23.6, *P* = 1 × 10^{-6}) substrate preference groups were evaluated using the Kruskal–Wallis one-way ANOVA. In each boxplot, a point denotes a single substrate–isolate pair. The top and bottom of each box represent the 25th and 75th percentiles, the horizontal line inside each box represents the median and the whiskers represent the range of the points excluding outliers. Substrates *n* = 82, consumers *n* = 27, simulations *n* = 27.

pathways is quantitatively most effective in creating mineral-associated SOM remains to be determined.

Using batch simulations to infer microbial life-history strategies and niches entails a trade-off. This is because the behaviour of organisms in culture may not always align with their performance in natural settings[46]. However, ground truthing of genome-inferred microbial traits is needed to enhance the credibility and confidence in genome-informed model predictions. Furthermore, using a consistent methodology to measure CUE across a broad range of microbial taxa is necessary to determine how physiological variation in substrate use between rhizosphere bacteria impacts CUE[37,38]. The same holds true for substrate uptake measurements, the accuracy of which is significantly influenced by experimental design elements, such as incubation duration and the preceding physiological state of the organisms. Consequently, deviations from model predictions might potentially stem from estimation inaccuracies or a lack of correlation between the half-saturation concentration for uptake and the half-saturation concentration for growth over extended incubation periods[63]. The simulation results presented here represent model-based hypotheses to confirm with appropriately designed experiments. These include considering a range of abiotic and biotic factors such as diurnal exudation dynamics[64], competition with roots and mineral surfaces for essential nutrients[65], and interactions with other organisms[66]. By integrating across a hierarchy of traits, genome-informed trait-based modelling

facilitates the generation and testing of hypotheses and can provide a robust foundation for the data-driven representation of microorganisms in many complex systems.

Methods

DEBmicroTrait model description

DEBmicroTrait assumes that microbial metabolism can be described by the standard dynamic energy budget (DEB) theory[67]. DEB theory partitions the total biomass into generalized chemical compounds (reserve and structural biomass) with specific functions: (1) reserve, denoted *E*, comprising the cellular growth machinery, including ribosomal proteins and RNA involved in the biosynthesis of proteins and/or additional ribosomes, as well as localized storage compounds (for example, glycogen, polyhydroxybutyrate) that buffer metabolism against external fluctuations in resource supply and (2) structural biomass, denoted *V*, comprising essential cellular proteins, DNA and other macromolecules that make up the cell wall and membrane[68]. Structural biomass is subjected to maintenance while reserves are subjected to continuous external supply and utilization (Fig. 1). The utilization of reserves follows first-order dynamics when expressed as a reserve density, $m_E = E/V$. The first-order turnover rate is given by
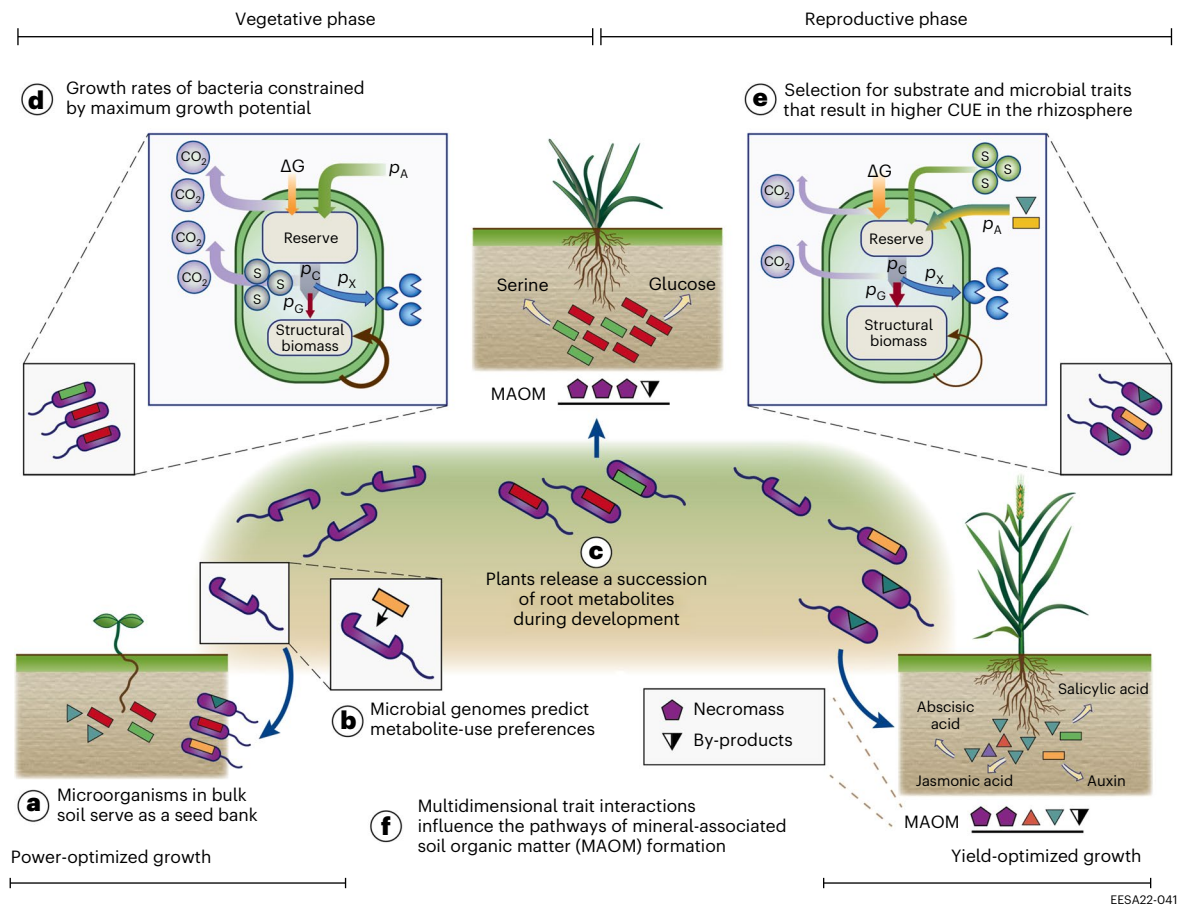
$$k_E = v/L_c = p_{Am}/m_E^*,\qquad(1)$$

Fig. 6 | Bacterial and root exudate metabolite traits interact to determine growth strategies in the rhizosphere. a,b, The microorganisms in bulk soil function as a seed bank and vary in their genetic potential to import and utilize distinct metabolites as substrates. **c**, Plants are genetically programmed to release a succession of root exudate metabolites throughout developmental stages (shown here are the vegetative and reproductive phases, with distinct metabolites coloured according to the six chemical classes used throughout the paper). **d**, During the vegetative growth phase, bacteria grow faster than they can assimilate new substrate. Rapid growth on sugars and amino acids is constrained by maximum growth potential and correlated with high biomass turnover. **e**, Bacteria grow more slowly on root exudates that are released during the reproductive growth phase (organic acids, fatty acids, auxins). Substrate

limitation occurs from environmental supply, except for select root metabolites that are preferentially consumed in the rhizosphere. These compounds select for organisms with traits (higher relative transporter gene frequencies, lower relative maintenance investment, higher protein synthesis efficiency, lower biomass turnover) that result in higher carbon use efficiency than typically observed experimentally when phenolic or polyvalent organic acids are added to bulk soil[55]. These results suggest that microbial community assembly patterns can be predicted from models coupling substrate-use preferences and chemical succession in the rhizosphere[22,86]. **f**, Understanding these patterns can guide strategies for engineering plant phenotypes and microbial communities, enhancing carbon stabilization through microbial carbon assimilation and subsequent stabilization on mineral surfaces.

where ν is an energy conductance parameter, $L_c = V_c^{1/3}$ is volumetric length, p_{Am} is the maximum assimilation flux and m_E^* represents the ratio of assimilation and mobilization fluxes, that is, the reserve capacity of the organism that is reached at long exposure to high substrate concentrations. The reserve pool buffers between environmental substrate uptake and microbial cell metabolism[69].

If the net synthesis rates of reserve and structure are balanced, weak homoeostasis implies that substrate limitation may occur inside the cell, capturing the dilution of reserve compounds due to (volume-) growth at a rate proportional to the reserve density[70]. It follows that the reserve mobilization power of the organism is given by the difference between the first-order turnover rate of reserve and dilution by growth,

$$p_C = \mu_E(k_E m_E - m_E r), \qquad (2)$$

where μ_E, measured in Gibbs energy per mol or C-mol, is the chemical potential of reserve and $r \equiv \frac{1}{V}\frac{dV}{dt}$ denotes the specific growth rate of the organism.

The reserve mobilization power, p_C, is partitioned between growth, p_G, maintenance, p_M, and extracellular enzyme production, p_X, power, that is,

$$p_C = p_G + p_M + p_X, \qquad (3)$$

where, in the absence of regulation, extracellular enzymes are produced constitutively at a rate proportional to the specific growth rate[71]. Dissipative loss in the growth and protein synthesis machinery is linked to the energetic cost of functional protein increasing with translation speed[72]. The basal maintenance rate is taken as proportional to cell volume—a valid assumption if the cumulative cost of protein turnover and replacement plays an important role in the overall energy budget and the amount of protein is proportional to cell volume[73]. Since the maintenance costs of protein turnover are preferentially on the expense of reserve materials, selection for high translation power (leading to reduced translational yield, y_{VE}) may also trade-off with the tolerance for starvation of the organism. However, somatic maintenance costs, k_M, can be supplemented with structural biomass when starvation leads

to the depletion of the reserve density, albeit at higher overhead costs in the transformation[74].

An explicit equation for the specific growth rate is obtained by combining equations (1) and (2) as

$$r = \frac{p_G}{\mu_E y_{EV}} = \frac{k_E m_E - (p_M + p_X)/\mu_E}{m_E + y_{EV}} \quad (4)$$

where $y_{EV} = 1/y_{VE}$ denotes the inverse of the structural growth efficiency. The inefficiencies associated with assimilation, growth, maintenance and enzyme production are captured as respiratory CO_2 losses (BR) and contribute to the overall CUE of the organism. CUE was calculated as BP/(BP+BR), where BP denotes total reserve and structural biomass production taking into account density-dependent biomass turnover[75].

A schematic for the coupling of anabolism and catabolism, as well as C and energy allocation in DEBmicroTrait is shown in Fig. 1. DEBmicroTrait trait integration and auxiliary equations to translate traits into model quantities are described below. An overview of the trait integration framework is provided in Supplementary Table 1.

Model assumptions for ribosome requirements. In our choice of parameterization, we assumed that natural selection has fine-tuned the translational power of ribosomes for optimal activity in a given environment[68]. Shifts in translational power across different bacteria follow allometric trends in macromolecular composition according to

$$k_E = \frac{r_{max} V_P}{V_R}, \quad (5)$$

where r_{max} is the genome-inferred maximum specific growth rate, V_P is the total protein volume and V_R is the ribosomal volume of the cell. V_p scales sublinearly with total cellular volume and impacts the ribosomal volume by considering how many ribosomes are required to replicate all ribosomes and proteins within a division cycle while also replacing proteins and ribosomes that have been degraded[76]. This places a lower bound on overall ribosomal volume of the cell according to the following inequality,

$$V_R \geq \frac{\bar{l}_P v_P N_P (\phi/r + 1)}{v_R (\bar{r}_R/r - \bar{l}_R(\eta/r + 1))}, \quad (6)$$

where r is growth rate, $\bar{l}_R$ is the average length of a ribosome in base pairs, $\bar{r}_R$ is the maximum base pair processing rate of the ribosome, $\bar{l}_P$ is the average protein length found to be invariant across bacteria, η and ϕ are specific degradation rates for ribosomes and proteins, N_P is the total number of proteins, and $\bar{v}_P$ and $\bar{v}_R$ represent the volume of an average protein or ribosome, respectively. While conceptually simple, the model explains more variance in translation phenotypes across species ($r^2 = 0.78$, $n = 11$; Supplementary Fig. 6a) than genomic signatures based on rrn copy number alone ($r^2 = 0.49$; Supplementary Fig. 6b). For simplicity, we assumed that rrn copy number is a good predictor of translational yield[77]. Consequently, we found that the fraction of reserve that is mobilized for growth, that is, $y_V = r/(k_E m_E)$, is negatively correlated with k_E at high k_E (Supplementary Fig. 5). At low k_E, the structural biomass yield remains relatively constant because reserves can be mobilized efficiently. As a result, maximum growth rates can occur at suboptimal yield[78].

Membrane requirements for substrate uptake. ECA kinetics provide a robust mathematical representation of complex substrate–consumer interactions in soil environments[41]. When applied to diffusive substrate uptake, the substrate affinity K_{ij} (mM) for the binding of substrate D_i ($i \in [1, ..., I]$) to free cellular binding sites B_j ($j \in [1, ..., J]$) in chemical equilibrium can be approximated as

$$K_{ij} = K_{0,ij} \left(1 + \frac{k_{ij}^+ B_{j,T}}{4\pi \bar{D}_i r_{c,j} n_j} \right), \quad (7)$$

where $K_{0,ij} \sim k_{2,ij}/k_{ij}^+$ is the ratio of forward reaction coefficients, $B_{j,T}$ denotes total (free and occupied) binding sites, $\bar{D}_i$ is the substrate diffusivity (m² s⁻¹), $r_{c,j}$ is the spherical cell radius (m) and n_j is the cell number density (m⁻³). The maximum substrate processing rate $k_{2,ij}$ (s⁻¹) defines the maximum specific uptake rate according to $V_{max,ij} = k_{2,ij} B_{j,T}$. Hence, if a cell increases its volumetric binding site density ($B_{j,T}/n_j$), it decreases its substrate affinity. At the same time, decreasing the volumetric size while keeping the same area-specific binding site density, $B_{j,T} = n_j \rho_{B,j} 4\pi r_{c,j}$, can increase substrate affinity.

We considered binding site densities in the context of cellular substrate supply and demand under conditions of balanced growth. Then, the ratios of extensive properties (that is, intensive properties, such as the reserve density) remain constant[79]. It follows that the reserve density is given by the ratio of assimilation power to reserve turnover rate (equation 1). Furthermore, cells have evolved optimal protein densities in cellular compartments (for example, the cell membrane) that maximize reaction rates[80]. Cellular demand depends on biochemical processes associated with substrate assimilation into generalized reserve compounds. Accordingly, cells can differ in their assimilation yield as given by

$$y_{ED} = \frac{1}{Y_{cat}^c + \lambda Y_{an}^c}, \lambda = \frac{\Delta G_{an} + \Delta G_{diss}}{-\Delta G_{cat}}, \quad (8)$$

where λ couples the stoichiometric vectors for catabolism (Y_{cat}^c) and anabolism (Y_{an}^c) by determining how many times the catabolic reaction (the energy production through substrate degradation, ΔG_{cat}) needs to run to provide the Gibbs energy for anabolism (ΔG_{an}), with the remaining energy dissipated into the environment (ΔG_{diss}). Determination of λ requires the calculation of the Gibbs free energy changes ΔG_{cat} and ΔG_{an}, the latter of which is equivalent to the energy conversion to generalized reserve compounds, plus the synthesis costs of membrane-bound binding site proteins, that is, $\Delta G_{an} = \Delta G_{block} + \Delta G_{\rho_{porter}}$ in our formulation. For simplicity, the elemental composition of reserve was treated as similar to structural cell components and follows allometric trends in cellular biomolecule composition with volume[76]. The Gibbs energy of dissipation denotes the Gibbs energy change of the conversion of the biomass building block to the different biomass components, that is, $\Delta G_{diss} = \Delta G_{syn}/v$, where v is a constant fraction of the energy dissipated in all enzymatic steps of the process. We assumed that ΔG_{syn} is constant[81], but organisms can differ in their protein synthesis efficiency as described above.

The cellular membrane area that needs to be covered with binding sites (ρ_{porter}) to reach transport rates commensurate with maximum specific growth rates was obtained by solving equations (1), (4) and (8). The estimated binding site densities of isolates matched for specific substrate molecules in the rhizosphere ranged from 0.0005% to 0.19%, with a median covering ~0.1% of the cell surface. The average estimated binding site density for cumulative uptake of plant metabolites amounted to ~0.08, or 8% of the total membrane area. These estimates were then scaled using normalized relative gene frequencies of specific transporter genes in the genome of isolates. To normalize the results per unit C microbial biomass[40], we calculated the number of cells in one mol C equivalent biomass (λ_B) using the allometric scaling of cell dry mass components (M_{dry}) and the assumption that 47% of dry biomass is C[76].

To represent substrate competition in a network of single-substrate to product reactions, the consumption of substrate D_i by a consumer B_j is then given by

$$j_{D,ij} = \frac{dD_{i,j}}{dt} = -k_{2,ij}^+ N_{ij} B_j \left(\frac{F_{c,j}^{(i)}}{1 + F_{r,i} + F_{c,j}} \right), \quad i \in [1, ..., I], j \in [1, ..., J] \quad (9)$$

where

$$F_{c,j} = \sum_{l=1}^{l=I} (D_l/K_{lj}) = \sum_{l=1}^{l=I} F_{c,j}^{(l)} \tag{10}$$

is the normalized substrate flux that describes the influence of all competing substrate fluxes towards consumer B_j, with conjugate flux

$$F_{r,i} = \sum_{l=1}^{l=J} (N_{il}B_l/K_{il}) = \sum_{l=1}^{l=J} F_{r,i}^{(l)} \tag{11}$$

describing all competing consumers' demands for a given substrate S_i, $k_{2,ij}^+$ is the maximal substrate conversion rate, $N_{ij} = \frac{M_C 4\pi r_{c,j}^2 \rho_{porter,ij}}{0.47 M_{dry} \pi r_p^2 N_A}$ (mol C^{-1}) is the matrix of biomass-specific binding sites (with 'radius' r_p), M_C is the molar mass of C, N_A is Avogrado's constant and K_{ij} (mM) denotes the binding half-saturation constants (equation 7).

Simulations

We simulated laboratory batch culture conditions for 39 bacterial isolates growing on 82 plant exudate metabolites. The concentration of exudates was selected to match the original growth medium concentration (125 mg C l^{-1}), assuming that nitrogen and other essential nutrients are non-limiting for the synthesis of biomass[22]. Exudate concentrations corresponded to concentrations of dissolved organic C detected in soil excavated from the 'Little Buck' pasture at the University of California Hopland Research and Extension Center (38° 59′ 34.5768″ N, 123° 4′ 3.7704″ W), which is the traditional and ancestral territory of the Shóqowa and Hopland People, where these bacteria were originally isolated. Inocula corresponding to 10^3 cells per gram soil were split into 90% reserve and 10% structural biomass[74] and simulations were run for 500 h or until substrate was exhausted. The reported CUE and realized growth rates corresponded to median values averaged over the simulated growth curves. The latency of C assimilation was determined through the timing of peak growth mineralization in each simulation.

To compare maximum specific growth rates of isolates to measured growth rates in a defined growth medium (Supplementary Fig. 3 and Table 5), we defined the in silico molecular input for commercial compound mixtures (yeast extract, proteose peptone) according to manufacturer instructions, following guidelines developed for genome-scale metabolic modelling[82].

Simulations were then extended to represent a mixed growth medium by evenly distributing the original batch exudate concentration across the different metabolites. Mixed medium simulations were run for 1,500 h or until the first metabolite was depleted, and the differences in relative uptake between rhizosphere bacterial response groups were calculated from the concentration differences at the start and end of the simulation. Using these data, resource niche breadth was calculated using Levins index[83].

Statistical analysis

Statistical analysis was performed using publicly available R (v.3.6.2) and Python (v.3.7.1) packages as described below.

To assess the relative proportion of variance explained by taxonomy and resource type, we used restricted maximum log-likelihood and variance partitioning analyses using the lmer function in the 'lm4' (v.1.1.27.1) package with default settings, followed by the r.squaredGLMM function in the 'MuMIn' (v.1.43.17) package. A typical analysis was coded as

```
fm.trait <- lmer(trait ~ as.factor(main) + (1|nested/ontology), data)

r.squaredGLMM(fm.trait)
```

and used to assess the relative importance of each variable when grouped together in a nested framework (Supplementary Table 2).

In addition, to assess the relative importance of genomic traits on isolate growth rate and C use efficiency, we used multiple regression models including rRNA operon copy number (*rrn*) and genome size (G), and the best model was selected on the basis of the smallest Akaike information criterion value. The regression analysis was performed using the lm function in R with the generic form

dependent variable = copy number * slope rrn + genome size * slope G

+copy number * genome size * slope rrn : G + intercept

and results are summarized in Supplementary Table 3.

A linear regression model with growth regime (high vs low) as categorical variable was used to test for the relationship between C use efficiency and growth rate, as well as changes in BP rate due to BR rate. BP and BR were log$_{10}$-transformed to match model assumptions (Fig. 4d). We then tested for the scaling relationship between BP and BR using a one-sample t-test to determine whether the slope was different from unity.

To determine model variable importance for C use efficiency and growth rate, we analysed the mean decrease in model accuracy as a measurement of the change in the accuracy of boosted random forest predictions when the variable in question was randomly permuted using the 'gbm' (v.2.1.8) library and within it the gbm function, as

```
fm.boost <- gbm (trait ~ ., data = train, distribution = "gaussian",

n.trees = 10000, interaction.depth = 8, shrinkage = 0.001 ).
```

The number of trees, tree depth and the shrinkage parameter were selected on the basis of cross validation against the test root mean square error, using the observations for isolates originally classified as an 'undefined' response group as 994 out-of-bag samples (Supplementary Fig. 4).

Covariation patterns in standardized energy budget flux distributions were illustrated via a principal component analysis (PCA) using the PCA function in the 'scikit-learn' (v.1.02) Python package (Fig. 4e,f).

To test for significant differences in trait distributions among rhizosphere response groups, we used a Kruskal−Wallis test followed by Dunn's test with Benjamini−Hochberg correction as implemented in the 'FSA' (v.0.9.2) R package.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw experimental data can be accessed from public repositories provided in ref. 22. Data to reproduce the work in this paper are provided in the Supplementary Data File and registered in Zenodo at https://doi.org/10.5281/zenodo.7879221.

Code availability

The DEBmicroTrait source code is available from the standalone modular Julia (v.1.6.7) library DEBmicroTrait.jl (https://github.com/giannamars/DEBmicroTrait.jl). All scripts to reproduce the work in this paper are registered in Zenodo at https://doi.org/10.5281/zenodo.7879221.

References

1. Bradford, M. A. et al. Managing uncertainty in soil carbon feedbacks to climate change. *Nat. Clim. Change* **6**, 751–758 (2016).
2. Berardi, D. et al. 21st-century biogeochemical modeling: challenges for century-based models and where do we go from here? *Glob. Change Biol. Bioenergy* **12**, 774–788 (2020).
3. Fierer, N., Bradford, M. A. & Jackson, R. B. Toward an ecological classification of soil bacteria. *Ecology* **88**, 1354–1364 (2007).

4. Fierer, N. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* **15**, 579–590 (2017).

5. Malik, A. A. et al. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J.* **14**, 1–9 (2020).

6. Beier, S., Werner, J., Bouvier, T., Mouquet, N. & Violle, C. Trait–trait relationships and tradeoffs vary with genome size in prokaryotes. *Front. Microbiol.* **13**, 985216 (2022).

7. Stone, B. W. et al. Life history strategies among soil bacteria—dichotomy for few, continuum for many. *ISME J.* **17**, 611–619 (2023).

8. Treseder, K. K. Ecological strategies of microbes: thinking outside the triangle. *J. Ecol.* **111**, 1832–1843 (2023).

9. Piton, G. et al. Life history strategies of soil bacterial communities across global terrestrial biomes. *Nat. Microbiol.* **8**, 2093–2102 (2023).

10. Steen, A. D. et al. High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J.* **13**, 3126–3130 (2019).

11. Sharon, I. & Banfield, J. F. Genomes from metagenomics. *Science* **342**, 1057–1058 (2013).

12. Madin, J. S. et al. A synthesis of bacterial and archaeal phenotypic trait data. *Sci. Data* **7**, 170 (2020).

13. Anantharaman, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).

14. Kumar, M., Ji, B., Zengler, K. & Nielsen, J. Modelling approaches for studying the microbiome. *Nat. Microbiol.* **4**, 1253–1267 (2019).

15. Schimel, J. Modeling ecosystem-scale carbon dynamics in soil: the microbial dimension. *Soil Biol. Biochem.* **178**, 108948 (2023).

16. Karaoz, U. & Brodie, E. L. microtrait: a toolset for a trait-based representation of microbial genomes. *Front. Bioinform.* **2**, 918853 (2022).

17. Sokol, N. W. et al. Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nat. Rev. Microbiol.* **20**, 415-430 (2022).

18. Milligan-McClellan, K. C. et al. Deciphering the microbiome: integrating theory, new technologies, and inclusive science. *mSystems* **7**, e00583-22 (2022).

19. Lajoie, G. & Kembel, S. W. Making the most of trait-based approaches for microbial ecology. *Trends Microbiol.* **27**, 814–823 (2019).

20. Hall, E. K. et al. Understanding how microbiomes influence the systems they inhabit. *Nat. Microbiol.* **3**, 977–982 (2018).

21. Westoby, M. et al. Strategic traits of bacteria and archaea vary widely within substrate-use groups. *FEMS Microbiol. Ecol.* **97**, fiab142 (2021).

22. Zhalnina, K. et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* **3**, 470–480 (2018).

23. Gralka, M., Pollak, S. & Cordero, O. X. Genome content predicts the carbon catabolic preferences of heterotrophic bacteria. *Nat. Microbiol.* **8**, 1799–1808 (2023).

24. Levy, A. et al. Genomic features of bacterial adaptation to plants. *Nat. Genet.* **50**, 138–150 (2018).

25. Jacoby, R. P. & Kopriva, S. Metabolic niches in the rhizosphere microbiome: new tools and approaches to analyse metabolic mechanisms of plant–microbe nutrient exchange. *J. Exp. Bot.* **70**, 1087–1094 (2019).

26. Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. Plant–microbiome interactions: from community assembly to plant health. *Nat. Rev. Microbiol.* **18**, 607–621 (2020).

27. Sokol, N. W. & Bradford, M. A. Microbial formation of stable soil carbon is more efficient from belowground than aboveground input. *Nat. Geosci.* **12**, 46–53 (2019).

28. Roller, B. R. & Schmidt, T. M. The physiology and ecological implications of efficient growth. *ISME J.* **9**, 1481–1487 (2015).

29. Tao, F. et al. Microbial carbon use efficiency promotes global soil carbon storage. *Nature* **618**, 981–985 (2023).

30. Sasse, J., Martinoia, E. & Northen, T. Feed your friends: do plant exudates shape the root microbiome? *Trends Plant Sci.* **23**, 25–41 (2018).

31. Chaparro, J. M., Badri, D. V. & Vivanco, J. M. Rhizosphere microbiome assemblage is affected by plant development. *ISME J.* **8**, 790–803 (2014).

32. Williams, A. et al. Root functional traits explain root exudation rate and composition across a range of grassland species. *J. Ecol.* **110**, 21–33 (2022).

33. McLaughlin, S., Zhalnina, K., Kosina, S., Northen, T. R. & Sasse, J. The core metabolome and root exudation dynamics of three phylogenetically distinct plant species. *Nat. Commun.* **14**, 1649 (2023).

34. Vieira-Silva, S. & Rocha, E. P. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet.* **6**, e1000808 (2010).

35. Weissman, J. L., Hou, S. & Fuhrman, J. A. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl Acad. Sci. USA* **118**, e2016810118 (2021).

36. Whipps, J. & Lynch, J. Substrate flow and utilization in the rhizosphere of cereals. *New Phytol.* **95**, 605–623 (1983).

37. Saifuddin, M., Bhatnagar, J. M., Segrè, D. & Finzi, A. C. Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nat. Commun.* **10**, 3568 (2019).

38. Muscarella, M. E., Howey, X. M. & Lennon, J. T. Trait-based approach to bacterial growth efficiency. *Environ. Microbiol.* **22**, 3494–3504 (2020).

39. Pold, G. et al. Carbon use efficiency and its temperature sensitivity covary in soil bacteria. *mBio* **11**, e02293-19 (2020).

40. Tang, J. & Riley, W. J. Competitor and substrate sizes and diffusion together define enzymatic depolymerization and microbial substrate uptake rates. *Soil Biol. Biochem.* **139**, 107624 (2019).

41. Tang, J. & Riley, W. A total quasi-steady-state formulation of substrate uptake kinetics in complex networks and an example application to microbial litter decomposition. *Biogeosciences* **10**, 8329–8351 (2013).

42. Maharjan, R. et al. The form of a trade-off determines the response to competition. *Ecol. Lett.* **16**, 1267–1276 (2013).

43. Jaeger, C. H. 3rd, Lindow, S. E., Miller, W., Clark, E. & Firestone, M. Mapping of sugar and amino acid availability in soil around roots with bacterial sensors of sucrose and tryptophan. *Appl. Environ. Microbiol.* **65**, 2685–2690 (1999).

44. Bosdriesz, E. et al. Low affinity uniporter carrier proteins can increase net substrate uptake rate by reducing efflux. *Sci. Rep.* **8**, 5576 (2018).

45. Norris, N., Levine, N. M., Fernandez, V. I. & Stocker, R. Mechanistic model of nutrient uptake explains dichotomy between marine oligotrophic and copiotrophic bacteria. *PLoS Comput. Biol.* **17**, e1009023 (2021).

46. Li, J. et al. Predictive genomic traits for bacterial growth in culture versus actual growth in soil. *ISME J.* **13**, 2162–2172 (2019).

47. Barnett, S. E., Youngblut, N. D., Koechli, C. N. & Buckley, D. H. Multisubstrate DNA stable isotope probing reveals guild structure of bacteria that mediate soil carbon cycling. *Proc. Natl Acad. Sci. USA* **118**, e2115292118 (2021).

48. Dang, C. et al. Phylogenetic organization in the assimilation of chemically distinct substrates by soil bacteria. *Environ. Microbiol.* **24**, 357–369 (2022).

49. Cheng, T. et al. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* **47**, 2140–2148 (2007).

50. Calabrese, S., Chakrawal, A., Manzoni, S. & Van Cappellen, P. Energetic scaling in microbial growth. *Proc. Natl Acad. Sci. USA* **118**, e2107668118 (2021).

51. Liang, C., Schimel, J. P. & Jastrow, J. D. The importance of anabolism in microbial control over soil carbon storage. *Nat. Microbiol.* **2**, 17105 (2017).

52. Nuccio, E. E. et al. Niche differentiation is spatially and temporally regulated in the rhizosphere. *ISME J.* **14**, 999–1014 (2020).

53. Ling, N., Wang, T. & Kuzyakov, Y. Rhizosphere bacteriome structure and functions. *Nat. Commun.* **13**, 836 (2022).

54. Manzoni, S., Taylor, P., Richter, A., Porporato, A. & Ågren, G. I. Environmental and stoichiometric controls on microbial carbon-use efficiency in soils. *New Phytol.* **196**, 79–91 (2012).

55. Sokol, N. W., Sanderman, J. & Bradford, M. A. Pathways of mineral-associated soil organic matter formation: integrating the role of plant carbon source, chemistry, and point of entry. *Glob. Change Biol.* **25**, 12–24 (2019).

56. Neurath, R. A. et al. Root carbon interaction with soil minerals is dynamic, leaving a legacy of microbially derived residues. *Environ. Sci. Technol.* **55**, 13345–13355 (2021).

57. Liang, C., Amelung, W., Lehmann, J. & Kästner, M. Quantitative assessment of microbial necromass contribution to soil organic matter. *Glob. Change Biol.* **25**, 3578–3590 (2019).

58. Villarino, S. H., Pinto, P., Jackson, R. B. & Piñeiro, G. Plant rhizodeposition: a key factor for soil organic matter formation in stable fractions. *Sci. Adv.* **7**, eabd3176 (2021).

59. Bird, J. A., Herman, D. J. & Firestone, M. K. Rhizosphere priming of soil organic matter by bacterial groups in a grassland soil. *Soil Biol. Biochem.* **43**, 718–725 (2011).

60. Daly, R. A. et al. Viruses control dominant bacteria colonizing the terrestrial deep biosphere after hydraulic fracturing. *Nat. Microbiol.* **4**, 352–361 (2019).

61. Wilhelm, R. C. et al. Tracing carbon metabolism with stable isotope metabolomics reveals the legacy of diverse carbon sources in soil. *Appl. Environ. Microbiol.* **88**, e00839-22 (2022).

62. Kögel-Knabner, I. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biol. Biochem.* **34**, 139–162 (2002).

63. Fink, J. W., Held, N. A. & Manhart, M. Microbial population dynamics decouple growth response from environmental nutrient concentration. *Proc. Natl Acad. Sci. USA* **120**, e2207295120 (2023).

64. Hubbard, C. J. et al. The plant circadian clock influences rhizosphere community structure and function. *ISME J.* **12**, 400–410 (2018).

65. Whitman, T. et al. Microbial community assembly differs across minerals in a rhizosphere microcosm. *Environ. Microbiol.* **20**, 4444–4460 (2018).

66. Hungate, B. A. et al. The functional significance of bacterial predators. *mBio* **12**, e00466-21 (2021).

67. Kooijman, S., Muller, E. & Stouthamer, A. Microbial growth dynamics on the basis of individual budgets. *Antonie van Leeuwenhoek* **60**, 159–174 (1991).

68. Franklin, O., Hall, E. K., Kaiser, C., Battin, T. J. & Richter, A. Optimization of biomass composition explains microbial growth–stoichiometry relationships. *Am. Nat.* **177**, E29–E42 (2011).

69. Mason-Jones, K., Breidenbach, A., Dyckmans, J., Banfield, C. C. & Dippold, M. A. Intracellular carbon storage by microorganisms is an overlooked pathway of biomass growth. *Nat. Commun.* **14**, 2240 (2023).

70. de Jong, H. et al. Mathematical modelling of microbes: metabolism, gene expression and growth. *J. R. Soc. Interface* **14**, 20170502 (2017).

71. Traving, S. J., Thygesen, U. H., Riemann, L. & Stedmon, C. A. A model of extracellular enzymes in free-living microbes: which strategy pays off? *Appl. Environ. Microbiol.* **81**, 7385–7393 (2015).

72. Dethlefsen, L. & Schmidt, T. M. Performance of the translational apparatus varies with the ecological strategies of bacteria. *J. Bacteriol.* **189**, 3237–3245 (2007).

73. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl Acad. Sci. USA* **112**, 15690–15695 (2015).

74. Tolla, C., Kooijman, S. A. & Poggiale, J.-C. A kinetic inhibition mechanism for maintenance. *J. Theor. Biol.* **244**, 576–587 (2007).

75. Georgiou, K., Abramoff, R. Z., Harte, J., Riley, W. J. & Torn, M. S. Microbial community-level regulation explains soil carbon responses to long-term litter manipulations. *Nat. Commun.* **8**, 1223 (2017).

76. Kempes, C. P., Wang, L., Amend, J. P., Doyle, J. & Hoehler, T. Evolutionary tradeoffs in cellular composition across diverse bacteria. *ISME J.* **10**, 2145–2157 (2016).

77. Roller, B. R., Stoddard, S. F. & Schmidt, T. M. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* **1**, 16160 (2016).

78. Wong, W. W., Tran, L. M. & Liao, J. C. A hidden square-root boundary between growth rate and biomass yield. *Biotechnol. Bioeng.* **102**, 73–80 (2009).

79. Bruggeman, F. J., Planqué, R., Molenaar, D. & Teusink, B. Searching for principles of microbial physiology. *FEMS Microbiol. Rev.* **44**, 821–844 (2020).

80. de Groot, D. H., van Boxtel, C., Planqué, R., Bruggeman, F. J. & Teusink, B. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLoS Comput. Biol.* **15**, e1006858 (2019).

81. Sousa, T., Mota, R., Domingos, T. & Kooijman, S. M. Thermodynamics of organisms in the context of dynamic energy budget theory. *Phys. Rev. E* **74**, 051901 (2006).

82. Marinos, G., Kaleta, C. & Waschina, S. Defining the nutritional input for genome-scale metabolic models: a roadmap. *PLoS ONE* **15**, e0236890 (2020).

83. Colwell, R. K. & Futuyma, D. J. On the measurement of niche breadth and overlap. *Ecology* **52**, 567–576 (1971).

84. Flynn, K. J., Skibinski, D. O. & Lindemann, C. Effects of growth rate, cell size, motion, and elemental stoichiometry on nutrient transport kinetics. *PLoS Comput. Biol.* **14**, e1006118 (2018).

85. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2022); https://www.R-project.org/

86. Beattie, G. A. Metabolic coupling on roots. *Nat. Microbiol.* **3**, 396–397 (2018).

Acknowledgements

Extended Data Fig. 1 | Trade-offs between substrate uptake parameters and optimal affinity parameter values at different external substrate concentrations. A Trade-off curve between predicted maximum specific uptake rate (V_{max}) and half-saturation constant (also known as substrate affinity parameter, K) in the equilibrium chemistry approximation (ECA[40]) for substrate uptake across substrates (n = 82) and isolates (n = 39). Isolates are colored by their response to root growth (green: positive, magenta: negative, white: undefined). **B** Normalized specific substrate uptake rate (J) as a function of the affinity parameter (K) for three different external substrate concentrations (D). The dashed lines indicate optimal affinity parameter values (K^{opt}). In **A** and **B**, the solid lines correspond to locally weighted linear regression models across substrates and isolates (n = 3198).
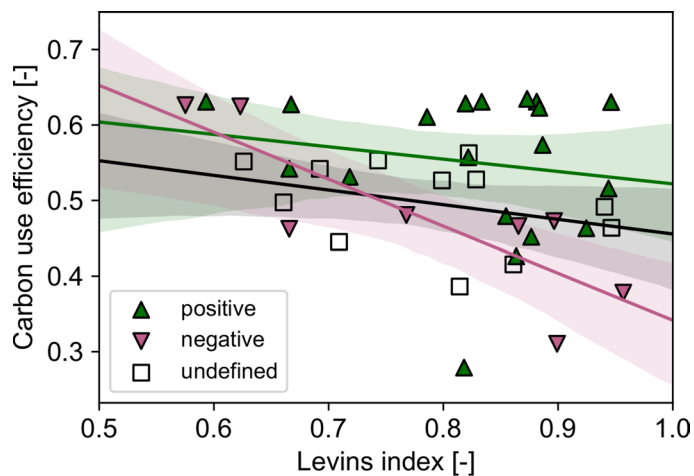
Extended Data Fig. 2 | Niche breadth of isolates for growth on mixed root exudate metabolites. Relationship between carbon use efficiency and resource niche breadth (Levins index[83]) of isolates for growth in a mixed medium composed of 82 root exudate metabolites. Isolates are colored by their response to root growth (green: positive, magenta: negative, white: undefined). The solid lines indicate the respective regression line (positive: $F_{1,17} = 0.44$, $r^2 = 0.03$, $p = 0.52$, negative: $F_{1,6} = 10.99$, $r^2 = 0.65$, $p = 0.01$, undefined: $F_{1,10} = 1.31$, $r^2 = 0.12$, $p = 0.28$), while shaded areas indicate the 95% confidence bands. Niche breadth was unimodally distributed across isolates based on Hartigan's dip test (all: $D = 0.04$, $p = 0.76$, positive: $D = 0.07$, $p = 0.68$, negative: $D = 0.11$, $p = 0.52$, undefined: $D = 0.08$, $p = 0.81$).

Corresponding author(s): Eoin Brodie

Last updated by author(s): October 21, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Custom computer code used to generate results was typed in the Julia programming language (Julia (v1.6.7)). The source code is available from the standalone, modular library DEBmicroTrait.jl (https://github.com/giannamars/DEBmicroTrait.jl). All scripts to reproduce the work in this manuscript are registered under digital object identifier 10.5281/zenodo.7879221.
Data analysis	For statistical analysis, we used R (v3.6.2) with packages lm4 (v1.1.27.1), MuMIn (v1.43.17), gbm (v2.1.8), FSA (v0.9.2), Python (v3.7.1) with scikit-learn (v1.02)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The raw experimental data can be accessed from public repositories provided in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.
Data exclusions	Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.
Replication	Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.
Randomization	Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.
Blinding	Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).
Research sample	State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.
Sampling strategy	Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.
Data collection	Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.

Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Research sample	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Sampling strategy	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Data collection	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Timing and spatial scale	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Data exclusions	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Reproducibility	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Randomization	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.
Blinding	Details are published in Zhalnina, Kateryna, et al. "Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly." Nature microbiology 3.4 (2018): 470-480.

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern
☒ ☐	Plants

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Antibodies

Antibodies used	Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.
Validation	Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.
Authentication	Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.
Mycoplasma contamination	Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Palaeontology and Archaeology

Specimen provenance	Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.
Specimen deposition	Indicate where the specimens have been deposited to permit free access by other researchers.
Dating methods	If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.
Wild animals	Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.
Reporting on sex	Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall

	numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No | Yes
- ☒ ☐ Public health
- ☒ ☐ National security
- ☒ ☐ Crops and/or livestock
- ☒ ☐ Ecosystems
- ☒ ☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No | Yes
- ☒ ☐ Demonstrate how to render a vaccine ineffective
- ☒ ☐ Confer resistance to therapeutically useful antibiotics or antiviral agents
- ☒ ☐ Enhance the virulence of a pathogen or render a nonpathogen virulent
- ☒ ☐ Increase transmissibility of a pathogen
- ☒ ☐ Alter the host range of a pathogen
- ☒ ☐ Enable evasion of diagnostic/detection modalities
- ☒ ☐ Enable the weaponization of a biological agent or toxin
- ☒ ☐ Any other potentially harmful combination of experiments and agents

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor*

	was applied.
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| Data access links
May remain private before publication.	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session
(e.g. [UCSC](#))	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See [Eklund et al. 2016](#))	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph,*

Graph analysis

subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.