bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments

Sylvie Estrela[1,*,#], Alicia Sanchez-Gorostiaga[1,2,*], Jean C.C. Vila[1,*], Alvaro Sanchez[1,#]

[1]*Department of Ecology & Evolutionary Biology and Microbial Sciences Institute, Yale University, CT, USA.* [2]*Department of Microbial Biotechnology, Centro Nacional de Biotecnología, CSIC, Cantoblanco, Madrid, Spain.*

* These authors contributed equally.

[#]Corresponding authors: sylvie.estrela@yale.edu (SE) ; alvaro.sanchez@yale.edu (AS)

Abstract

A major open question in microbiome research is whether we can predict how the components of a diet collectively determine the taxonomic composition of microbial communities. Motivated by this challenge, here we ask whether communities assembled in a mixed nutrient environment can be predicted from those assembled in every single nutrient in isolation. To that end, we first formulate a quantitative null model of community assembly in a mixture of nutrients that recruit species independently. We then test the model predictions by assembling replicate communities in synthetic environments that contain either a pair of nutrients, or each nutrient in isolation. We find that the null, naturally additive model generally predicts well the family-level community composition. However, we also identify systematic deviations from the additive predictions that reflect generic patterns of nutrient dominance at the family-level of taxonomy. Pairs of more-similar nutrients (e.g. two sugars) are on average more additive than pairs of more dissimilar nutrients (e.g. one sugar and one organic acid). A simple dominance rule emerges, where we find that sugars generally dominate organic acids. This simple dominance rule extends to most families and most sugar-organic acid pairs in our experiment. Our results suggest that regularities in the ways nutrients interact may help us predict how microbial communities respond to changes in nutrient composition.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Understanding how the components of a complex biological system combine together to produce the system's properties and functions is a fundamental question in biology. Answering this question is central to solving many fundamental and applied problems, such as how multiple genes combine to give rise to complex traits [1,2], how multiple drugs affect the evolution of resistance in bacteria and cancer cells [3,4], how multiple environmental stressors affect bacterial physiology [5], or how multiple species affect the function of a microbial consortium [6–8].

In microbial population biology, a major related open question is whether we can predict how the components of a diet collectively determine the taxonomic and functional composition of microbial communities. Faith and co-workers tackled this question using a defined gut microbial community and a host diet with varying combinations of four macronutrients [9]. This study found that community composition in the combinatorial diets could be predicted from the communities assembled in each of the separate nutrients using an additive linear model [9]. Given the presence of a host and its own possible interactions with the nutrients and resident species, it is not immediately clear whether such additivity is directly mediated by interactions between the community members and the supplied nutrients, or whether it is mediated by the host, for instance by producing additional nutrients, or through potential interactions between its immune system and the community members. In a different study, Enke et al found evidence that marine enrichment communities assembled in mixes of two different polysaccharides could be explained as a linear combination of the communities assembled in each polysaccharide in isolation [10].

Despite the important insights provided by both of these studies, we are far from a general and quantitative understanding of how specific nutrients combine together to shape the composition of self-assembled communities [11]. Motivated by this challenge, here we set out to systematically investigate whether the assembly of enrichment microbial communities in defined nutrient mixes could be predicted from the communities that assembled in each of the single nutrients in isolation.

Results

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A null expectation for community assembly in mixed nutrient environments. To address this question, we must first develop a quantitative null model that predicts community composition in a mixed nutrient environment in the case where each nutrient recruits species independently. Any deviation between the null model prediction and the observed (measured) composition reveals that nutrients are not acting independently, but rather "interact" to shape community composition. This definition of an interaction as a deviation from a null model that assumes independent effects is commonplace in systems-level biology [12].

In order to formulate the null expectation for independently acting nutrients, let us consider a simple environment consisting of two unconnected demes where two bacterial species, A and B, can grow together. The first deme contains a single growth limiting nutrient (nutrient 1), while the second deme contains a different single limiting nutrient (nutrient 2) (**Fig. 1A**). In this scenario, each nutrient influences the abundance of species A and B independently: the microbes growing on nutrient 1 do not have access to nutrient 2 and vice versa. Let's denote the abundance of species A in deme 1 and 2 by $n_{A,1}$ and $n_{A,2}$, and the abundance of species B as $n_{B,1}$ and $n_{B,2}$, respectively. If we now consider the two-deme environment as a whole, the abundance of species A is the sum of its abundance in each deme $n_{A,12} = n_{A,1} + n_{A,2}$ (likewise, for species B $n_{B,12} = n_{B,1} + n_{B,2}$). This example illustrates that in the scenario when two limiting nutrients act independently, each of them recruits species just as if the other nutrients were not there. In such case, the abundance of each species in a nutrient mix is the sum of what we would find in the single nutrient habitats. Note that the lack of nutrient interactions does not mean that species do not interact with each other, but rather that whatever ecological or metabolic interaction they may have (e.g., competition for nutrients, cross-feeding, growth inhibition by toxins), such interaction is not affected by mixing nutrients.

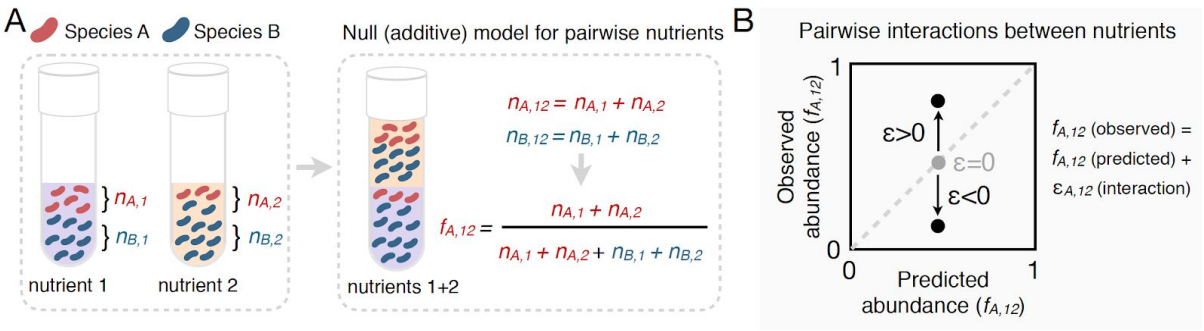

Fig. 1. Predicting community composition in mixed nutrient environments. A. Community composition in a single nutrient (nutrient 1 or nutrient 2) *vs* a mixture of nutrients (nutrient 1+nutrient 2). Assuming that nutrients act independently, the null model predicts that the abundance of each species in

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the mixture is the sum of its abundance in the single nutrients (i.e. additive). **B.** Plotting the experimentally measured (observed) relative abundance in the mixed carbon sources against its predicted (from null model) relative abundance reveals the presence or absence of interactions. Any deviation from the identity line (predicted=observed) is the interaction effect (ε). When $\varepsilon = 0$, there is no interaction between nutrients. When ε is non-zero, community composition is affected by nutrient interactions. If $\varepsilon>0$, the null model underestimates the abundance. If $\varepsilon<0$, the model overestimates the abundance.

Under the null model, the relative abundance of species i in a mix of nutrients 1 and 2 can be written as $f_{i,12}$ (null) = $w_1 f_{i,1} + w_2 f_{i,2}$ where $f_{i,1}$ and $f_{i,2}$ are the relative abundances of i in nutrient 1 and 2, respectively, and w_1 and w_2 are the relative number of cells in nutrients 1 and 2 (Methods). Any quantitative difference between the null model prediction and the observed composition quantifies an "interaction" between nutrients. Accounting for the presence of such interactions, the model can be re-written as $f_{i,12} = f_{i,12}$ (null) + $\varepsilon_{i,12}$ where $\varepsilon_{i,12}$ represents the interaction between nutrients 1 and 2 (**Fig. 1B**).

Experimental system. Equipped with this null model, we can now ask to what extent the nutrients recruit species independently in mixed environments. To address this question, we followed a similar enrichment community approach to the one we have used in previous work for studying the self-assembly of replicate microbial communities in a single carbon source [13,14] (Methods, **Fig. 2A**). Briefly, habitats were initially inoculated from two different soil inocula. Communities were then grown in synthetic (M9) minimal media supplemented with either a single carbon source or a mixture of two carbon sources, and they were serially passaged with transfers to fresh medium every 48h for a total of 10 transfers (dilution factor = 125×) (**Fig. 2A**). The two-carbon source cultures consisted of a focal carbon source mixed 1:1 with one of eight additional carbon sources. We previously found that stable multi-species communities routinely assemble in a single carbon source (which is limiting under our conditions), and they converge at the family level in a manner that is largely governed by the carbon source supplied, while the genus or lower level composition is highly variable [13]. We chose glucose as the focal carbon source because we have previously carried out multiple assembly experiments in this nutrient [13,14]. As the additional carbon sources, we chose cellobiose, fructose, ribose, and glycerol (i.e. a pentose, a hexose, a disaccharide and a sugar alcohol) and fumarate, benzoate, glutamine and glycine (two aminoacids and two organic acids). All carbon sources were also used in single carbon source cultures.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Communities assembled in single sugars contained 5 to 24 ESVs, mainly belonging to the fermenter Enterobacteriaceae family (~0.98±0.03) (**Fig. S1**). In contrast, communities assembled in organic acids exhibited a higher richness (12-36 ESVs), and unlike in sugars, Enterobacteriaceae were generally rare (~0.06±0.06). Instead, communities were dominated by respirators mainly belonging to the Pseudomonadaceae (~0.51±0.25), Moraxellaceae (~0.18±0.2), and Rhizobiaceae (0.11±0.13) families (**Fig. S1**). Because of the observed family-level convergence across carbon sources, which is consistent with previous studies [13–15], we focus our analysis below on family-level abundance.

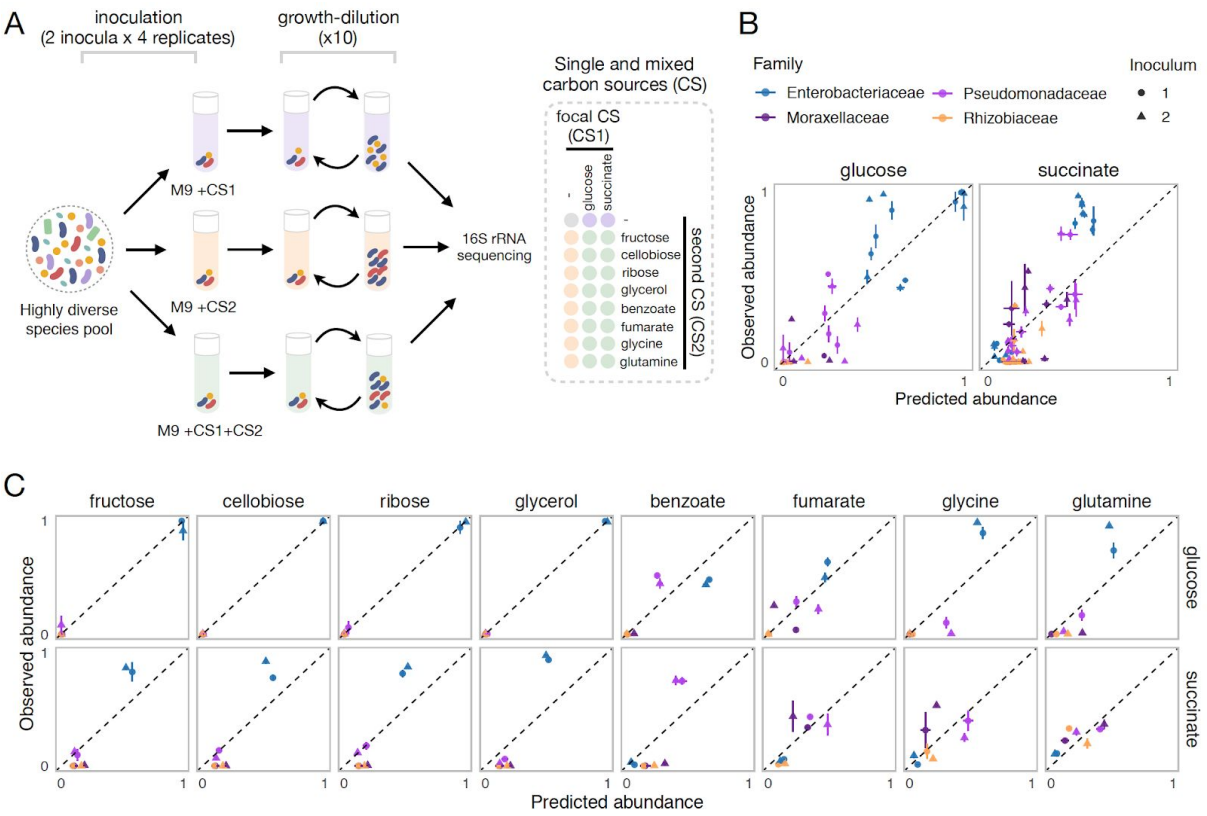

Fig. 2. Systematic deviations from the null prediction reveals that some nutrients interact to shape community assembly. A. Schematic of experimental design. Two different soil samples were inoculated in minimal M9 medium supplemented with either a single carbon source (CS1 or CS2) or a mixture of two carbon sources (CS1 + CS2) (3-4 replicates each). Communities were propagated into fresh media every 48h for 10 transfers, and then sequenced to assess community composition. Carbon source mixtures consisted of a focal carbon source (CS1; glucose or succinate) mixed with a second carbon source (CS2). **B, C.** For each pair of carbon sources, we show the experimentally observed and predicted (by the null additive model) relative abundance of each family in the mixture. Any deviation from the identity line (predicted=observed) reveals an interaction effect. Only the four most abundant families are shown. Error bars represent mean ± SE.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The null model of independently acting nutrients explains a high fraction of the variation observed. To investigate the predictive power of the null (additive) model, we compare the predicted and observed relative abundances of each family for each carbon source pair across all experiments. Our results show that the null model predicts well the family-level abundances on average (Pearson's R=0.95 and p<0.001; RMSE=0.073, N=223) (**Fig. 2B, Fig. S2**). To confirm that the strong predictive power of the null model is not an idiosyncrasy of using glucose as the focal carbon source in the pairs, we repeated the same experiment with succinate (an organic acid) as the focal carbon source. Although the correlation between observed and predicted abundance is lower than when glucose was the focal carbon source, the null additive model remains strongly predictive (Pearson's R=0.87 and p<0.001; RMSE=0.094; N=257) (**Fig. 2B**).

This result seems to indicate that, at the family level, a simple model that assumes that nutrients act independently can predict community composition in a pair of nutrients (for an analysis of this point at the genus and ESV level, see **Fig. S3**). However, when we looked at this more closely and broke down our results by carbon source and family, we found consistent and systematic deviations from the null model (**Fig. 2C**). For example, across all succinate-sugar pairs, Enterobacteriaceae are significantly more abundant than predicted by the null model (ε = 0.347±0.107, Mean±SD; p < 0.001, one-sample Student's t-test, N=32) while both Rhizobiaceae and Moraxellaceae are less abundant than predicted (ε = -0.136±0.0339 and ε=-0.152±0.0415; p< 0.001, one-sample Student's t-test, N=32) (**Fig. 2C**). The null 'interaction-free' model also predicts species abundance better in certain carbon source combinations (e.g. glucose + ribose) than in others (e.g. glucose + glutamine) (**Fig. 2C**). The existence of systematic deviations from the null prediction reveals that some nutrient pairs do not act independently, but instead interact with each other to affect the abundance of specific families.

A simple dominance rule in mixed nutrient environments: sugars generally dominate organic acids. To map the regularities we have observed in nutrient interactions, we next sought to characterize the nature of these interactions for each carbon source pair and every family. One helpful way of visualizing nutrient interactions is to draw the pairwise abundance landscape for each species and carbon source pair (**Fig. 3A**). For instance, a species could be either more abundant in a pair of nutrients than it is in any of them independently (synergy). Or it could be less abundant than it is in any of the two (antagonism). Dominance is a less extreme interaction which can be visualized by the pushing of a species abundance towards one single nutrient and

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

away from the average that is the predicted value, thus overriding the effect of the paired nutrient (**Fig. 3A**).

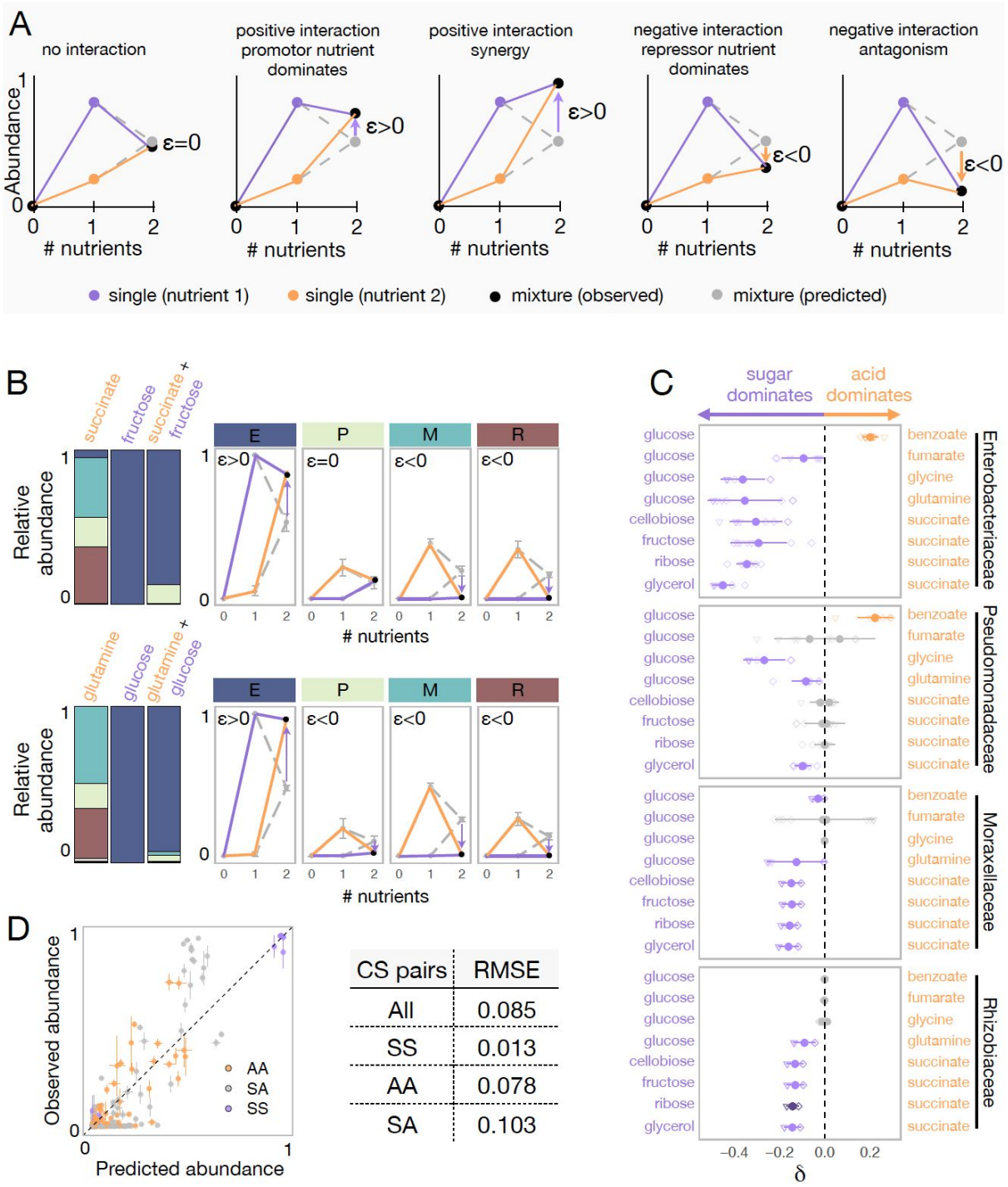

Fig. 3. Sugars generally dominate over organic acids. A. Detecting interactions and hierarchies of dominance between nutrients on microbial community composition. Drawing the single and pairwise abundance landscapes for each species allows us to visualise interactions between nutrients. Multiple types of interactions are possible, including dominance, synergy and antagonism. Interactions occur when ε is significantly different from 0 (one-sample Student's t-test, $p < 0.05$). Synergy (antagonism) occurs when the abundance in the mixture is greater (lower) than the abundances in any of the single nutrients

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

independently (Methods). Dominance occurs when the abundance in the mixture is closer or similar to the abundance in one of the singles. The landscape also allows us to identify which carbon source has a dominating effect within the pair. When ε>0, the growth-promoting nutrient dominates and has an overriding effect in the community composition. In contrast, when ε<0, the growth-repressing nutrient dominates. **B.** Two examples of nutrient interactions (succinate+fructose and glucose+glutamine) exhibiting sugar dominance. Barplots show a representative replicate from one of the inocula (**Fig. S1, S2**). For instance, the landscape for succinate-fructose shows that fructose overrides the effect of succinate by promoting Enterobacteriaceae (E), and repressing Moraxellaceae (M) and Rhizobiaceae (R) (purple arrows), whereas no interaction effect is observed for Pseudomonadaceae (P). Error bars represent mean ± SD of the four replicates. **C.** Panel shows dominance index for the eight sugar-acid pairs and the four dominant families. Filled circles show the mean±SD of the 2 inocula x 4 replicates for each pair of nutrients, and open symbols show all 8 independent replicates (different shapes for different inocula). Purple indicates that the sugar dominates while orange indicates that the acid dominates. Lighter purple and orange indicate dominance while darker purple and orange indicate super-dominance (synergy or antagonism). No interaction is shown in gray, and occurs when $\varepsilon=0$ (one-sample Student's t-test, $p < 0.05$, N=8) or when dominance is undefined because the two inocula exhibit opposite dominant nutrient (in which case δ is shown as both $-\delta$ *and* $+\delta$). **D**. Predicted vs observed family-level abundance. For each pair of carbon sources (CS), shown is the experimentally observed and predicted (by the null model) relative abundance of each family in the mixed carbon sources. Any deviation from the identity line (predicted=observed) is the interaction effect. The colours show whether the carbon source pairs are sugar-sugar (SS), acid-acid (AA), or sugar-acid (SA). Error bars represent mean ± SE. Table shows RMSE for each carbon source pair type.

When the interaction is positive ($\varepsilon>0$), the dominant nutrient is the one where the family grew to a higher abundance. When the interaction is negative ($\varepsilon<0$), the dominant nutrient is the one where the species grew less well. Mathematically, dominance occurs when $|\varepsilon|>0$ and $min(f_{i,1}, f_{i,2}) \leq f_{i,12} \leq max(f_{i,1}, f_{i,2})$, while synergy and antagonism (forms of super-dominance) occur when $|\varepsilon|>0$ and $f_{i,12} > max(f_{i,1}, f_{i,2})$ and $f_{i,12} < min(f_{i,1}, f_{i,2})$ respectively (Methods). **Fig. 3B** shows representative examples of dominant carbon source interactions. For instance, Moraxellaceae and Rhizobiaceae grow strongly on succinate, but they are not present in fructose. When fructose is mixed with succinate, both families drop dramatically in abundance, despite their high fitness in succinate alone. Interestingly, however, the dominance of fructose over succinate is not observed for all families: those two nutrients do not interact on Pseudomonas, whose abundance is well predicted by the null model. Using this framework, we then systematically quantified the prevalence of dominance, antagonism and synergy between nutrients for each family **(Fig. S4A)**. While 59% of the nutrient pair combinations exhibited no significant interaction, dominance was by far the most common interaction amongst those that interacted (73%, **Fig. S4A**). It occurred predominantly in the sugar-acid pairs, and to a lesser extent in the acid-acid pairs, and only rarely in the sugar-sugar pairs (**Fig. S4B**). This result strongly suggests that nutrient interactions are not random but do have a specific structure that is conserved at the family-level (**Fig. S4C**).

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To systematically characterize and quantify nutrient dominance, we developed a dominance index (δ) (Methods). For visualization purposes, the dominance index for the sugar-acid pairs (we will discuss the acid-acid pairs later) is written as $\delta_i = -|\varepsilon_{12}|$ when the sugar dominates and as $\delta_i = |\varepsilon_{12}|$ when the acid dominates. If $\varepsilon_{12} = 0$, then $\delta_i = 0$. That is, in the absence of interaction between nutrients, there is no dominance. By plotting the dominance index for each pair of nutrients and each family, we observe a generic pattern of dominance of sugars over acids (**Fig. 3C**). The families Moraxellaceae or Rhizobiaceae are recruited to the community by most organic acids in isolation, but they are not found in most sugar communities. When sugars and organic acids are mixed together, the sugar dominates and both families are at much lower abundances (by ~6-fold in the case of Moraxellaceae and ~114-fold in Rhizobiaceae) than the predicted average, even though the organic acid where they thrived is present in the environment. Consistent with this pattern, we found that pairs of more similar nutrients (a pair of sugars or a pair of organic acids) were significantly better predicted by the null model than mixed organic acid-sugar pairs (**Fig. 3D**). No generic pattern of dominance was observed in the acid-acid mixtures (**Fig. S5**). Together, these results indicate that interactions between nutrients are not universal, but rather they are conserved at the family-level.

Discussion

Understanding how the available nutrients affect the composition of microbial communities is a fundamental question in microbiome biology. Here, we have shown that a simple additive model that assumes that nutrients act independently is predictive of community composition at the family (and to a lesser extent also at the genus or ESV) level of taxonomic composition. Our results add to the growing evidence that nutrients may combine linearly to determine taxonomic abundance [9,10], and suggest that neither host action nor biochemically complex dietary sources are necessary for this additivity. Our results also highlight the existence of systematic and predictable deviations that are conserved at the family-level.

In particular, we find that there exist generic patterns of dominance between nutrients for specific families. In our experiments, sugars generally dominate over organic acids, a nutrient interaction rule that is conserved at the family level. When we examine interactions and dominance at the genus-level, we find that sugars do not exhibit the same dominance for all genera within the same

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

family (**Fig. S6 and S7**). This result is consistent with the convergence of community structure at the family level (despite substantial variation at lower levels of taxonomy) which we have reported for communities assembled in a single nutrient [13,14]. The predictable nutrient-interaction rules we have found thus represent another case of emergent simplicity in microbial community assembly. Together, our results highlight the importance of considering the taxonomic rank, and cautions against simply focusing on environmental complexity (i.e. the number of different nutrients) to understand how community properties, including taxonomic composition and function, are affected by the nutritional environment. Instead, our work suggests that not all nutrients are equal, and that taking into account the nature of the different nutrients that are combined is crucial for predicting how communities respond to different diets.

Our findings leave open many important questions about the mechanisms behind the emergent nutrient interaction rules observed. For instance, why are pairs of more similar nutrients better predicted than pairs of more dissimilar nutrients? Why do sugars dominate organic acids? Are the observed dominance rules followed in more complex environments? While addressing these questions is beyond the scope of this paper, we hope that posing them will stimulate future work. Answering these questions could help guide engineering diets to modulate the composition and function of microbial communities in desired directions, including promoting the growth of beneficial species and preventing the growth of undesired species, including the spread of pathogens.

Methods

Null model for relative abundance. Let's consider a simple scenario with two species (A and B) growing in two separate nutrients (1 and 2). This is similar to cocultures of A and B growing in two separate demes/tubes (one per nutrient). The fractions of A and B in nutrient 1 are $f_{A,1} = n_{A,1} /(n_{A,1} + n_{B,1})$ and $f_{B,1} = n_{B,1} /(n_{A,1} + n_{B,1})$ respectively, and similarly, the fractions of A and B in nutrient 2 are $f_{A,2} = n_{A,2} /(n_{A,2} + n_{B,2})$ and $f_{B,2} = n_{B,2} /(n_{A,2} + n_{B,2})$ (where n is the total number of cells of species A or B). If we mix nutrients 1 and 2 together (i.e. mix the two tubes), the fractions of A and B in the mixture are given by:

$f_{A,12} = (n_{A,1} + n_{A,2}) /(n_{A,1} + n_{B,1} + n_{A,2} + n_{B,2})$ and

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$f_{B,12} = (n_{B,1} + n_{B,2}) / (n_{A,1} + n_{B,1} + n_{A,2} + n_{B,2})$.

We can define $n_{t,1} = n_{A,1} + n_{B,1}$ and $n_{t,2} = n_{A,2} + n_{B,2}$ as the total number of cells in the nutrient demes 1 and 2, respectively. We can thus write $f_{A,12} = (n_{A,1} + n_{A,2}) / (n_{t,1} + n_{t,2})$. Defining $w_1 = n_{t,1} / (n_{t,1} + n_{t,2})$ and $w_2 = n_{t,2} / (n_{t,1} + n_{t,2})$, it is straightforward to show that: $f_{A,12} = w_1 f_{A,1} + w_2 f_{A,2}$. By the same reasoning, we find that $f_{B,12} = w_1 f_{B,1} + w_2 f_{B,2}$.

Sample collection. Soil samples were collected from two different natural sites in West Haven (CT, USA), with sterilized equipment, and placed into sterile bottles. Once in the lab, five grams of each soil sample were then transferred to 250mL flasks and soaked into 50mL of sterile 1x PBS (phosphate buffer saline) supplemented with 200 µg/mL cycloheximide (Sigma, C7698) to inhibit eukaryotic growth. The soil suspension was well mixed and allowed to sit for 48 hrs at room temperature. After 48hrs, samples of the supernatant solution containing the 'source' soil microbiome were used as inocula for the experiment or stored at −80ºC after mixing with the same volume of 80% glycerol.

Preparation of media plates. Stock solutions of carbon sources (CS, for a full list see **Table S1**) were prepared at 0.7 C-mol/L (10x) in 50 mL of double distilled sterile water and sterilized through 0.22µm filters (Millipore). CS were aliquoted into 96 deep-well plates (VWR) as single CS or mixed in pairs at 1:1 (vol:vol) and stored at -20ºC. To keep the total amount of carbon constant across all treatments at 0.07 C-mol/L, pairs contained half the amount of each carbon source when compared to their respective single CS. Synthetic minimal growth media was prepared from concentrated stocks of M9 salts, $MgSO_4$, $CaCl_2$, and 0.07C-mol/L (final concentration) of single or pairs of CS. The pH of all growth media (i.e. for each carbon source in M9) were determined and shown in **Table S1**.

Community assembly experiment. Starting inocula were obtained directly from the initial 'source' soil microbiome solution by inoculating 40µL into 500 µL culture media prepared as indicated above. For each sample and CS, 4 µL of the culture medium was dispensed into fresh media plates containing the different single or pairs of CS in quadruplicate. Bacterial cultures were allowed to grow for 48 hrs at 30 °C in static broth in 96 deep-well plates (VWR). After 48 hrs each culture was homogenized by pipetting up and down 10 times before transferring 4 µL to 500µL of fresh media, and cells were allowed to grow again. Cultures were passaged 10 times

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(~70 generations). Optical density (OD620) was used to measure biomass in cultures after the 48-hour growth cycle. Samples were frozen at -80ºC after mixing with 400 µL of 80% glycerol.

DNA extraction, library preparation, and sequencing. Samples were centrifuged for 40mins at 3500rpm, and the pellet was stored at -80ºC until DNA extraction. DNA extraction was performed with the DNeasy 96 Blood & Tissue kit for animal tissues (QIAGEN), as described in the kit protocol, including the pre-treatment step for Gram-positive bacteria. DNA concentration was quantified using the Quan-iTPicoGreen dsDNA Assay kit (Molecular Probes, Inc) and the samples were normalized to 5ng/uL before sequencing. The 16S rRNA gene amplicon library preparation and sequencing were performed by Microbiome Insights, Vancouver, Canada (www.microbiomeinsights.com). For the library preparation, PCR was done with dual-barcoded primers [16] targeting the 16S V4 region and the PCR reactions were cleaned up and normalized using the high-throughput SequalPrep 96-well Plate Kit. Samples were sequenced on the Illumina MiSeq using the 300-bp paired end kit v3.chemistry.

Taxonomy assignment. The taxonomy assignment was performed as described in previous work [14]. Following sequencing, the raw sequencing reads were processed, including demultiplexing and removing the barcodes, indexes and primers, using QIIME (version1.9, [17]), generating fastq files with the forward and reverse reads. DADA2 (version 1.6.0) was then used to infer exact sequence variants (ESVs) [18]. Briefly, the forward and reverse reads were trimmed at position 240 and 160, respectively, and then merged with a minimum overlap of 100bp. All other parameters were set to the DADA2 default values. Chimeras were removed using the "consensus" method in DADA2. The taxonomy of each 16S exact sequence variant (ESV) was then assigned using the naïve Bayesian classifier method [19] and the Silva reference database version [20] as described in DADA2. The analysis was performed on samples rarefied to 10779 reads.

Quantification of total abundances, interactions, and dominance. We used OD620 of the cultures after the 48-hour growth cycle as a proxy for total population size (community biomass). The predicted relative abundance of species i in a mix of nutrients 1 and 2 was then calculated as $f_{i,12}(\text{null}) = w_1 f_{i,1} + w_2 f_{i,2}$ where $f_{i,1}$ and $f_{i,2}$ are the relative abundances of i in nutrients 1 and 2, respectively, and $w_1 = (\text{OD620}_1 / (\text{OD620}_1 + \text{OD620}_2)$ and $w_2 = (\text{OD620}_2 / (\text{OD620}_1 + \text{OD620}_2)$. For each carbon source pair and inoculum, $f_{i,12}(\text{null})$ is calculated as the mean of the two single carbon source-replicate pairwise combinations (N=16). In order to quantify interactions, we first

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

determine whether an interaction between nutrients exists for each nutrient pair (nutrient 1 and nutrient 2) and family *i*. An interaction exists when $\varepsilon = f_{i,12} - f_{i,12}$(null) is significantly different from 0 (one-sample Student's t-test, $p<0.05$), that is when there is a deviation from the null prediction. Under such condition (i.e. $|\varepsilon|>0$), synergy and antagonism (which are forms of super-dominance) occur when $f_{i,12} > max(f_{i,1}, f_{i,2})$ and $f_{i,12} < min(f_{i,1}, f_{i,2})$ respectively, while dominance occurs when $min(f_{i,1}, f_{i,2}) <= f_{i,12} <= max(f_{i,1}, f_{i,2})$ (Welch two sample t-test, $p<0.05$). When $\varepsilon>0$, the nutrient with greater abundance dominates; when $\varepsilon<0$, the nutrient with lower abundance dominates. For visualization purposes, we developed a dominance index (δ). The dominance index for the sugar-acid pairs is written as $\delta_i = -|\varepsilon_{12}|$ when the sugar dominates and as $\delta_i = |\varepsilon_{12}|$ when the acid dominates. The dominance index for the sugar-sugar and acid-acid pairs is written as $\delta_i = -|\varepsilon_{12}|$ when the focal carbon source (glucose or succinate) dominates and as $\delta_i = |\varepsilon_{12}|$ when the additional carbon source dominates.

Statistical analysis. All data analysis was performed in R. Pearson's R was calculated using the R function 'cor.test' from the 'stats' package and the RMSE was calculated using the 'rmse' function from the 'Metrics' package.

Acknowledgments

We want to thank members of the Sanchez lab for helpful discussions. This work was supported by the National Institutes of Health through grant 1R35 GM133467-01, and by a Packard Foundation Fellowship to AS.

References

1. Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9: 855–867.

2. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet. 2014;15: 22–33.

3. Michel J-B, Yeh PJ, Chait R, Moellering RC Jr, Kishony R. Drug interactions modulate the potential for evolution of resistance. Proc Natl Acad Sci U S A. 2008;105: 14918–14923.

4. Wood KB, Wood KC, Nishida S, Cluzel P. Uncovering scaling laws to infer multidrug

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

response of resistant microbes and cancer cells. Cell Rep. 2014;6: 1073–1084.

5. Cruz-Loya M, Kang TM, Lozano NA, Watanabe R, Tekin E, Damoiseaux R, et al. Stressor interaction networks suggest antibiotic resistance co-opted from stress responses to temperature. ISME J. 2019;13: 12–23.

6. Sanchez-Gorostiaga A, Bajić D, Osborne ML, Poyatos JF, Sanchez A. High-order interactions distort the functional landscape of microbial consortia. PLoS Biol. 2019;17: e3000550.

7. Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, Korasidis N, et al. Microbiome interactions shape host fitness. Proc Natl Acad Sci U S A. 2018;115: E11951–E11960.

8. Guo X, Boedicker JQ. The Contribution of High-Order Metabolic Interactions to the Global Activity of a Four-Species Microbial Community. PLoS Comput Biol. 2016;12: e1005079.

9. Faith JJ, McNulty NP, Rey FE, Gordon JI. Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science. 2011;333: 101–104.

10. Enke TN, Datta MS, Schwartzman J, Cermak N, Schmitz D, Barrere J, et al. Modular Assembly of Polysaccharide-Degrading Marine Microbial Communities. Curr Biol. 2019;29: 1528–1535.e6.

11. Pacheco AR, Segrè D. The effects of environmental complexity on microbial community yield and structure. bioRxiv. 2020. p. 2020.07.09.195404. doi:10.1101/2020.07.09.195404

12. Sanchez A. Defining Higher-Order Interactions in Synthetic Ecology: Lessons from Physics and Quantitative Genetics. cels. 2019;9: 519–520.

13. Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. Science. 2018;361: 469–474.

14. Estrela S, Vila JCC, Lu N, Bajic D, Rebolleda-Gomez M, Chang C-Y, et al. Metabolic rules of microbial community assembly. bioRxiv. 2020. p. 2020.03.09.984278. doi:10.1101/2020.03.09.984278

15. Lu N, Sanchez-Gorostiaga A, Tikhonov M, Sanchez A. Cohesiveness in microbial community coalescence. bioRxiv. 2018. p. 282723. doi:10.1101/282723

16. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol. 2013;79: 5112–5120.

17. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7: 335–336.

18. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13: 581–583.

19. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.06.239897; this version posted August 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5261–5267.

20. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41: D590–6.