



Dissimilarity–Overlap analysis of replicate enrichment communities

Jean C. C. Vila ^{1,2} · Yang-Yu Liu ^{3,4} · Alvaro Sanchez ^{1,2}

Received: 26 February 2020 / Revised: 29 May 2020 / Accepted: 9 June 2020 / Published online: 18 June 2020

© The Author(s), under exclusive licence to International Society for Microbial Ecology 2020

Abstract

The taxonomic composition of microbial communities can vary substantially across habitats and within the same habitat over time. Efforts to build quantitative and predictive models of microbial population dynamics are underway, but fundamental questions remain. How different are population dynamics in different environments? Do communities that share the same taxa also exhibit identical dynamics? In vitro communities can help establish baseline expectations that are critical towards resolving these questions in natural communities. Here, we applied a recently developed tool, Dissimilarity–Overlap Analysis (DOA), to a set of experimental in vitro communities that differed in nutrient composition. The Dissimilarity and Overlap of these communities are negatively correlated in replicate habitats, as one would expect if microbial population dynamics were on average strongly convergent (or “universal”) across these replicate habitats. However, the existence of such a negative correlation does not necessarily imply that population dynamics are always universal in all communities. Even in replicate, identical habitats, two different communities may contain the same set of taxa at different abundances in equilibrium. The formation of alternative states in community assembly is strongly associated with the presence of specific taxa in the communities. Our results benchmark DOA, providing support for some of its core assumptions, and suggest that communities sharing the same taxa and external abiotic factors generally (but not necessarily) have a negative correlation between Dissimilarity and Overlap.

Introduction

Microorganisms grow and thrive in all habitats throughout the biosphere [1–4]. This includes the human body, where

they form rich ecological communities made of large numbers of interacting species [5–8]. The taxonomic composition of these communities can vary substantially between body sites, reflecting their different ecological, physical, and biochemical conditions [9]. Even for the same body site, community composition may vary widely between individuals, as well as within the same individual over time [10, 11]. In order to understand how microbiomes change longitudinally and over the lifespan of an organism, and to design effective strategies that enable us to manipulate microbiomes towards desirable states, it is critical to develop predictive quantitative models of microbial population dynamics [12].

Models of dynamic ecosystems vary in their level of description, which is typically chosen to capture the specific phenomena under study. A detailed population dynamics model of microbial communities would have to include mechanistic microbial interactions (due to cross-feeding [13, 14], direct secretion of substances such as bacteriocins, antibiotics, or extracellular enzymes [15, 16], or competition for the same nutrient), spatial structure of the particular habitat [17], and environment–microbiome or host–microbiome interactions. Building such detailed models can be daunting due to (i) a huge number of model

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-0702-7>) contains supplementary material, which is available to authorized users.

- ✉ Jean C. C. Vila
jean.vila@yale.edu
- ✉ Yang-Yu Liu
yyl@channing.harvard.edu
- ✉ Alvaro Sanchez
alvaro.sanchez@yale.edu

- ¹ Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06511, USA
- ² Microbial Sciences Institute, Yale University West Campus, West Haven, CT 06516, USA
- ³ Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA 02115, USA
- ⁴ Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

parameters which need to be inferred from experimental data; and (ii) many environmental variables (such as the concentrations of bacteriocins and nutrients) which are hard to measure in real time.

To avoid those difficulties, an alternative modeling framework focuses on exploring the impact that any given microbial species has on the abundance of other microbial species [18]. In this phenomenological modeling framework, one only needs to consider a simple population dynamics model written as a set of ordinary differential equations: $dX(t)/dt = f(X(t), \Theta)$. Here, f is a nonlinear function characterizing the population dynamics of the microbial community, $X(t) = (x_1(t), \dots, x_i(t), \dots, x_N(t))$ is an N -dimensional vector with $x_i(t)$ denoting the abundance of the i -th microbial species at time t , and Θ captures all the ecological parameters (such as intrinsic growth rates, intra- and inter-species interaction strengths, etc.). Note that those ecological parameters depend on environment- or host-independent factors, such as biochemical processes and microbial metabolic pathways; as well as environment- or host-specific ones, such as pH, temperature, nutrient intake, host genetic make-up, etc. Hence, environmental or host factors are not explicitly considered in this modeling framework but are absorbed in the ecological parameters [18].

Generally, the ecological parameters estimated from a given habitat with certain characteristic environmental conditions do not necessarily map to other habitats with different environmental conditions. One can ask, however, whether those parameters (Θ) are strongly similar (“universal”) for microbiomes that assemble in similar habitats. Addressing this fundamental question has important consequences for the applicability and predictive power of quantitative models of microbial community dynamics. If the interaction parameters were highly similar across habitats of a certain type, such as the guts of different human subjects, this will facilitate the development of generic microbiome-based therapeutics. By contrast, if the ecological parameters and microbial dynamics are strongly host-specific, we must design truly personalized interventions, which need to consider not only the highly personalized microbial composition of each individual but also the unique dynamics of the underlying microbial ecosystem.

Directly addressing the above question would require one to infer all of the ecological parameters and fit the population dynamics $f(X(t), \Theta)$ from the microbiome data of each local community or host. Doing this for a large collection of communities is both logistically and computationally challenging. Recently, an indirect method called Dissimilarity–Overlap Analysis (DOA) was proposed [18]. DOA relies on two mathematically independent measures between any two local communities: Overlap (O), which is defined as half of the sum of relative abundances of the shared species; and Dissimilarity (D), which is defined as

the divergence between the renormalized abundance profiles of the shared species (Methods) [18]. DOA is based on the following two assumptions. First, the abundance profiles of the microbiome samples represent the steady states X^* of the microbial ecosystem and hence the fixed points of the underlying population dynamics that satisfy $f(X^*, \Theta) = 0$. Second, if any two local communities that have the same species collection also have the same abundance profile (steady state), i.e., $O = 1$ and $D = 0$, then the two communities should share universal microbial dynamics $f(X, \Theta)$ characterized by the same set of ecological parameters Θ . Mathematically, this means that if X^* satisfies both $f(X^*, \Theta^{(1)}) = 0$ and $f(X^*, \Theta^{(2)}) = 0$, given the large number of species and all the other levels of complexity in their interactions (encoded in the highly nonlinear function f), we conclude that generically $\Theta^{(1)} = \Theta^{(2)}$. In general, since D is mathematically not constrained by any value of $O > 0$, any constraints of D by O observed from real data deserve ecological interpretations. In particular, even if we do not have any steady state pair satisfying $O = 1$ and $D = 0$ (which is the typical case for host-associated microbial communities, such as the human gut microbiome, due to highly personalized microbial compositions), as long as steady state pairs with higher O tend to have lower D , i.e., there is a negative slope in the high-Overlap region of the Dissimilarity–Overlap Curve (DOC). This particular statistical constraint of D by O is consistent with the hypothesis of universal dynamics across all habitats in the sample, and it is a foundation of DOA [18]. It is also consistent with alternative hypotheses, such as communities assembling in environmental gradients, or situations when only a small fraction of the habitats have highly similar interaction parameters [19]. The former is a particularly important scenario, and was recognized in the original study by Bashan et al. In many instances enough is known about the habitats to exclude from the analysis factors that can lead to environmental heterogeneity [18].

A negative slope in the high-Overlap region of the DOC has been found in the gut microbiome samples collected from different healthy individuals [18]. Yet, the complete set of selective pressures experienced by microorganisms in the same habitat (e.g., the same body site of different individuals), and their variation across a host group cannot be known exactly. Hence, one cannot account for all the potential factors that may conceivably influence the microbial communities assembled in the same habitat, and so cannot provide an entirely conclusive answer regarding the universality of the underlying microbial dynamics. In other words, we cannot unambiguously attribute the negative slope of the DOC to universal dynamics and completely rule out the alternative explanation of environment or host factors. A more definitive demonstration would require a comparison between experimental communities assembled

in well-controlled replicate habitats to those assembled in nonidentical habitats.

Benchmarking DOA against well-controlled in vitro communities, ideally assembled in multiple replicates of habitats that are either identical to each other, or different from each other in well-understood ways, would be necessary to understand the limitations and potential of DOA for its application to natural communities [20]. To address this need, here we perform DOA for a large set of in vitro communities that meet these requirements: close to 300 independent enrichment communities assembled in multiple, replicate synthetic habitats on three different limiting nutrient conditions, and assembled to equilibrium under periodic serial dilution cycles [13].

Results

Data Set

In a recent study [13], we reported the assembly of a total of 276 enrichment communities in three different synthetic environments: M9 minimal medium with either glucose, citrate, or leucine as the only carbon source. These enrichment communities were assembled from twelve different environmental sources (including various soil samples and plant matter collected near Yale University in New Haven, CT). Seven or eight biological replicates of each inoculum were propagated in each of the three nutrient environments, under serial dilution with transfers every 48 h with a dilution factor of 125 \times . A diagrammatic summary of the experiment is presented in Fig. 1a. As reported in [13], communities were initially very diverse ($N = 110$ – 1290 unique Exact Sequence Variants, or ESVs). They typically converged to an approximately stable composition (containing $N = 2$ – 22 ESVs) after 50–60 generations, suggesting that communities were close to a steady state. Metabolic cross-feeding was found to be widespread and critical for the coexistence of multiple species on a single limiting resource [13].

Communities assembled in identical environments exhibit a negative correlation between Dissimilarity and Overlap

We first addressed the question of whether communities assembled in identical environments do indeed give rise to a negatively sloped DOC at high Overlap. To test this prediction, we took all pairs of communities that had been assembled in the same nutrient-limited habitats, by pooling together every possible pair of glucose communities, as well as every pair of citrate (and of leucine) assembled communities. We then measured the Dissimilarity and Overlap for each pair. Applying the same type of statistical analysis

used in the original study by Bashan et al. (Robust LOW-ESS regression; see “Methods” for details [18]), we find that the DOC (but not the controls; Fig. S1) does indeed exhibit a negative slope at high values of Overlap (Fig. 1b), and this is also confirmed by a standard linear regression ($D = D_0 + mO$; $m = -0.56$, $p < 0.002$ by Bootstrapping; Methods) applied to the points with higher than median O (Fig. 1b, inset; Methods). The same is true when we analyze each of the three nutrient environments separately (Fig. 1c), and it also holds when we separate those communities assembled from either the same or different inoculum (Fig. 1d, e). In contrast, a statistically significant negative correlation between Dissimilarity and Overlap is not observed for community pairs that are assembled in different environments (e.g., one in citrate medium, one in glucose medium) (Fig. S2).

An expectation of DOA is that communities that contain the same taxa in identical habitats should have them at highly similar equilibrium abundances, as the underlying population dynamics would be strongly similar. By contrast, communities assembled in different nutrient habitats are not expected to have similar species abundances even when they happen to share a large proportion of common species (high Overlap), as we do not necessarily expect their dynamical equations to be similar. Consistent with the first hypothesis, we find that, in identical nutrient conditions, the majority of our high-Overlap ($O > 0.98$) communities have low Dissimilarity (Fig. 1f, see S1 for comparison to null) (Mean = 0.188, Median = 0.132, IQR = 0.23). To test the second hypothesis, we considered “mixed” pairs of communities, where each community in the pair was assembled in a different environment. For instance, we find that glucose–citrate pairs (which exhibit no correlation between D and O (Figs. S2 and 3)) have a similar Overlap distribution to glucose–glucose and citrate–citrate pairs (Fig. 1g). Yet, the distribution of Dissimilarities for these high-Overlap pairs ($O > 0.98$) is shifted up compared to glucose–glucose pairs ($t = -12.79$, p value < 0.002 by Bootstrapping; Methods) and citrate–citrate pairs ($t = -11.965$, p value < 0.002 , by Bootstrapping; Methods) (Fig. 1h). This shift persists even if we only consider communities assembled from the same inoculum (Fig. S4) and is robust to the Overlap threshold chosen (Fig. S5).

This last finding is consistent with the idea that population dynamics and equilibria are strongly convergent when the environments are identical, but not necessarily when the environments are different. This lends support to the null assumption that species interactions with the environment and with each other are different in different environments, but strongly convergent in identical environments. Our results support the prediction that identical environments will generate a negative statistical correlation between D and O , whereas different environments will not.

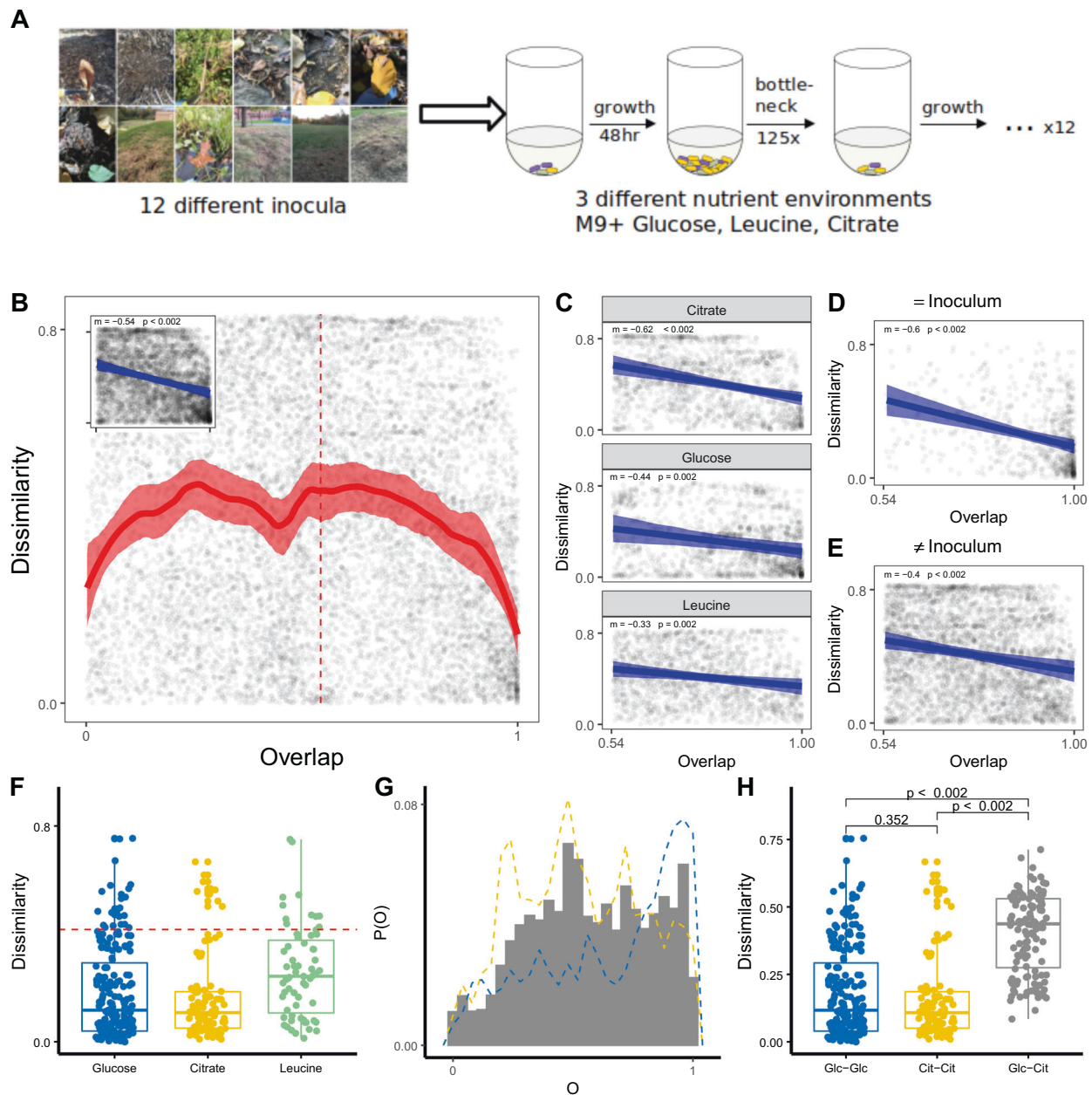


Fig. 1 Communities assembled in the same environment show a negative correlation between Dissimilarity and Overlap. **a** Schematic description of the experiments in ref. [13]. **b** DOC of all microbial community pairs that have been assembled in the same environment ($n = 276$ Samples). Shaded regions indicate the 95% confidence interval (Methods). The vertical dotted red line represents the median Overlap (0.543). The inset shows a linear regression for communities above the median Overlap. We repeat this regression over the same region, subsetting the data to consider **(c)** each nutrient environment separately; **(d)** subsets of pairs that have been assembled from the same inoculum **(e)**; subsets of the pairs that have been assembled different inoculum. For each regression, we report m (slope of the linear regression) and a p value calculated as the fraction of bootstrap realization in which this slope is negative (see Methods).

f Distributions for community pairs assembled in the same environment (both in glucose, both in citrate, or both in leucine) with high Overlap ($O > 0.98$). The dotted red line is at half the maximum possible dissimilarity $\frac{\sqrt{\log(2)}}{2}$. **g** Histogram showing distributions of Overlaps for community pairs where one has been assembled on glucose and the other on citrate. The dotted lines give the frequency polygon for glucose–glucose community pairs and citrate–citrate community pairs (blue and yellow, respectively). We use the same binwidth (0.04) for both histograms and frequency polygons so the two are comparable. **h** Glucose–citrate communities with high-Overlap ($O > 0.98$) have significantly higher mean dissimilarity than glucose–glucose communities or citrate–citrate communities in the same Overlap range ($O > 0.98$). Displayed p values are computed by bootstrapping (see Methods).

Table 1 Specific taxa are strongly associated with high Dissimilarity in replicate habitats.

ESV	<i>N</i>	Difference in mean	<i>T</i> -statistic	<i>P</i> value
<i>Citrobacter</i>	25	0.319974235	9.632238	<0.002
<i>Yersinia</i>	19	0.289481747	7.3186745	0.00209205
<i>Enterobacteriaceae</i>	38	0.248081124	6.6420427	0.005988024
<i>Pseudomonas</i>	57	0.175009032	6.5060298	0.003992016
<i>Pseudomonas.2</i>	41	0.131673553	3.716131	0.093812375
<i>Raoultella</i>	45	0.103170657	3.2164266	0.081836327
<i>Pseudomonas.3</i>	18	−0.002248539	−0.0671654	0.55228806
<i>Pseudomonas.1</i>	24	−0.031053732	−0.969162	0.638722555
<i>Pseudomonas.4</i>	27	−0.044317786	−1.6107989	0.756487026
<i>Klebsiella</i>	65	−0.281428827	−7.4501771	0.98997996

For each of the ten most commonly observed ESVs on Glucose we performed one-tailed *t*-tests to determine whether pairs of highly overlapping communities ($O > 0.98$) both containing that ESV had higher Dissimilarity than pairs where at least one community did not contain the ESV. In this table, we report the identity of the ESV, the number of communities in which that ESV is found (*N*) the difference in mean Dissimilarity, the *T*-statistic and the *p* value (obtained by bootstrapping, see Methods). See Tables S1 and S2 for a similar analysis applied to leucine and citrate communities.

Specific taxa can be strongly associated with high Dissimilarity in replicate habitats

As can be visually appreciated in Fig. 1, when environments are identical and Overlap is high the Dissimilarity in our experimental communities is generally small. However there are numerous deviations from this rule, and we find multiple community pairs with high Overlap that still show high levels of Dissimilarity. Considering only communities with an Overlap > 0.98 , we find that 12% of glucose pairs, 9% of citrate pairs, and 17% of leucine pairs have Dissimilarity $> \frac{\sqrt{\log(2)}}{2}$ (which is half of the maximum possible Dissimilarity, calculated through the root Jensen–Shannon divergence) (Fig. 1f). Similar values are also obtained if we only consider communities assembled from the same inoculum, (14% of glucose pairs, 9% citrate pairs, 8% of leucine pairs) (Fig. S4A). In sum, we find that communities can be dominated by the same set of ESVs in identical habitats, yet these ESVs, may exist at very different abundances.

We set out to investigate whether communities that deviate from the average trend captured by the DOC could be associated with the presence of specific taxa. To that end, we first selected pairs of glucose communities with high Overlap (higher than the median of 0.98) that were assembled from the same initial pool of species (for a similar analysis in citrate or leucine, see Tables S1 and S2). We then tested whether any of the ten most commonly observed ESVs (corresponding to ESVs found in at least 18 of the 92 glucose communities), were statistically associated with high Dissimilarity. Four of these ESVs had higher Dissimilarity than expected by chance ($p < 0.05$, by Bootstrapping, Methods). Of these, an ESV of the genus *Citrobacter* had the largest effect (Table 1).

To further investigate this point, we split all pairs of communities by whether both contain this *Citrobacter* ESV (group I), only one contains it (group II), or none does (group III) (Fig. 2a). The mean Dissimilarity is higher for group I than group II (0.427 vs 0.175, $p < 0.003$, by Bootstrapping; Methods), and for group II than group III (0.175 vs 0.088, $p = 0.06$ by Bootstrapping; Methods) Fig. 2b. As shown in Fig. S6, our results are robust to our choice of a “High-Overlap” threshold of 0.98. If *Citrobacter* ESV was indeed associated with alternative dynamical states (either through multistability, or through their contribution to alternative dynamical equations when they are part of the community), we would also expect the DOC to flatten for glucose–glucose communities that contain it, relative to those that do not. That is indeed the case, as shown in Fig. 2c, d, see Fig. S7 for controls).

Similar results were found for citrate communities and leucine communities, where we found a *Raoultella* ESV and a *Pseudomonas* ESV associated with higher Dissimilarity respectively (Supp Tables S1, S2, and Figs. S8, S9). Our results thus reveal that the presence of a single ESV in a community may be strongly associated with alternative states in community assembly, even amongst communities that are assembled in the same environment from the same regional pool and contain highly overlapping sets of taxa.

Discussion

The first part of this paper tests two fundamental predictions of DOA. First, using publicly available data from a recent experiment we show that community assembly in identical environments does lead to a negative correlation between Dissimilarity and Overlap and a negatively sloped DOC.

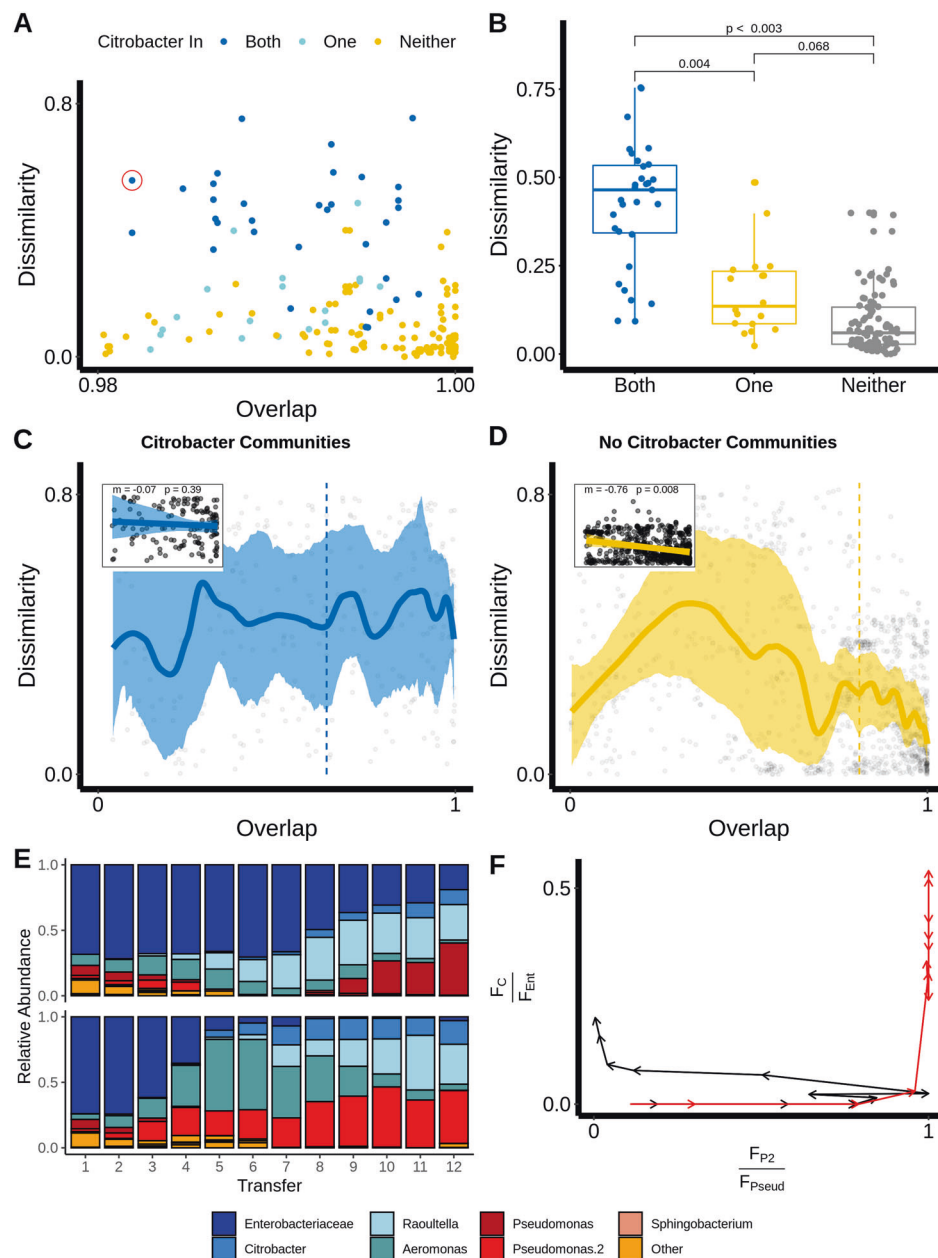


Fig. 2 A *Citrobacter* ESV is associated with dynamical dissimilarity in communities assembled in replicate environments. **a** Dissimilarity and Overlap of microbial community pairs assembled from the same regional pool on M9 + glucose with above-median Overlap ($O > 0.98$). **b** Dissimilarity of the same set of communities. For **(a)** and **(b)** we label communities by whether *Citrobacter* ESV is found in both communities (dark blue), only in one community (light blue) or in neither community (yellow). **c** DOC of all pairs of microbial communities assembled on glucose that contain *Citrobacter* ESV ($n = 25$). **d** DOC of all pairs of microbial communities assembled on glucose that do not contain *Citrobacter* ($n = 67$). **e** Population dynamics for one pair of glucose communities with high Overlap ($O = 0.98$) and

high Dissimilarity ($D = 0.56$) (highlighted in **(a)** with the red circle). Structure of the two communities at the ESV level at every transfer. **f** Phase portraits illustrate the dynamics of the most abundant Enterobacteriaceae and Pseudomonadaceae ESV within those two communities. That black line corresponds to the top community in **(e)** and the red line corresponds to the bottom community in **(e)**. F_C and F_{P2} represent the fraction of the *Citrobacter* and Pseudomonas.2 ESVs in the population, whereas F_{Ent} and F_{Pseud} represent the fractions of the Enterobacteriaceae and Pseudomonadaceae families. Dynamics are highly convergent until the third transfer, after which the communities diverge to alternative states.

Second, we show that, as expected, communities assembled in identical environments that also contain highly overlapping sets of taxa have them at strongly convergent

abundances. This is consistent with what one would expect if their dynamics were describable by the same equations. Likewise, we would not expect identical dynamical

equations in different nutrient habitats. Consistent with this expectation, we found that communities assembled in different nutrient habitats do not have convergent species abundances, even when they happen to share a high number of taxa. This is reflected in the data by a shift in the Dissimilarity to higher values for those communities with very large Overlap.

Our results also indicate the existence of specific taxa that are associated with high Dissimilarity in replicate habitats. In glucose communities, we find an ESV of the genus *Citrobacter*, which not only predicts high Dissimilarity between community pairs when it is present in at least one of the communities, but it also flattens the DOC. Intriguingly, these results are in line with one of the main findings of the Bashan study, which was that microbiomes disrupted by *C. difficile* infection did not exhibit a negative correlation between Dissimilarity and Overlap, and Dissimilarity remained constant even as Overlap increased [18]. The negative slope was recovered after a fecal microbiota transplantation, which restored a healthy microbiota and cured the disease. Although the reasons for this finding may be very different from the similar result found in our communities, our results indicate that a tight association between specific taxa and the flattening of the DOC may be seen even in the absence of an immune system or a complex host.

The publicly available data of Goldford et al. [13] includes the measurement of the complete population dynamics for one glucose pair with high-Overlap and high Dissimilarity (highlighted in Fig. 2a). Both communities in the pair contain the *Citrobacter* ESV. For this pair, we find that the population dynamics are initially strongly convergent between the two communities, but bifurcate after ~20 generations and subsequently diverge into alternative compositions (Fig. 2e, f), suggesting the potential presence of true multistability (i.e., multiple stable steady states are associated with the same set of species).

Of course, other possibilities exist. For example, the population dynamics may just not be identical even in replicate habitats due to the violation of the first assumption of DOA (that the communities are at equilibrium steady state). Communities may either have not reached equilibrium after 12 transfers, or they may be at a non-equilibrium steady state, undergoing oscillatory or cyclical dynamics. Neutral population dynamics and stochastic population dynamics can also lead to variation in community composition that may lead to increased Dissimilarity [19, 21]. These dynamics would be observed if communities are moving on a shallow attractor where selection is weak, or if non-accounted environmental fluctuations shift the position of the fixed points in the community. The high overlap of the two communities at the ESV level might also reflect their differences at the strain level due to rapid evolution. *Citrobacter* may also have more sensitive

interactions with the rarer members of the community. It is important to consider that environments do not just passively select for species and determine their interactions, but rather they are dynamically shaped by the taxa growing in them [22–25]. Therefore, it is possible that although the supplied nutrients are the same in two communities, the environment experienced by the members of our communities is actually different, through the different effects that species have on it [22].

One limitation of our study is that the communities investigated are species-poor, and many of our community pairs will only share a few species in common. This can be potentially problematic because it may affect the Dissimilarity measurement [18]. Despite this caveat, the overall good agreement between our findings and the expectations of DOA suggests that species richness is not necessarily an impediment for the application of DOA to taxonomically poor natural communities. Further research would be needed to establish the precise conditions under which this would be true. More generally, others have argued that communities assembled along an environmental gradient may also give rise to negative correlations between Dissimilarity and Overlap [19]. It would be important to test this prediction experimentally (for instance by establishing mixed nutrient habitats with varying concentrations of glucose and citrate in between the two “pure” habitats studied by Goldford et al.) but this falls beyond the scope of this paper.

It is thus important to remark that although sets of communities assembled in identical habitats present a negatively sloped DOC, the reverse statement is not necessarily true: the presence of a negatively sloped DOC does not necessarily mean that the habitats are identical to each other. In fact, grouping together all communities in the experiment, including those assembled in identical and different environments and projecting them all into the same Dissimilarity–Overlap plot, we find a negative correlation between D and O and a negatively sloped DOC (Fig. S10). The reason is that the strong effect of identical habitats overpowers the effect of nonidentical ones. In our case, we know the environmental factor that was critically different among these habitats (the single limiting nutrient), but this is not something that is trivial to identify in any given natural habitat, even if we remove any known factors of variation across habitats.

Notwithstanding these important caveats, our results confirm in a controlled set of experiments that microbial dynamics in replicate habitats are strongly convergent on average and lead to a negatively sloped DOC. In recent years, negatively sloped DOC has been detected for the microbial communities assembled in some (but not all) human body sites [18], as well as for mycorrhizal fungal communities [20]. In order to correctly interpret these results, it is critical to benchmark the technique not only

against simulations, but also against experimental communities whose assembly is well understood. Our results have provided a first empirical benchmark using well-controlled communities. We hope that these findings will contribute to grounding the expectations and intuitions behind DOA, and contribute to its application to microbial communities assembled in natural environments. More generally we hope that these results will encourage other researchers to benchmark novel statistical methods in microbial ecology using well-controlled *in vitro* communities.

Methods

Community assembly experiment

We analyzed publicly available data from a recent set of *in vitro* microcosms experiments [13]. Briefly, diverse microbial communities were isolated from natural ecosystems and used as the inoculum for a batch culture containing M9 + one of three carbon sources (Glucose, Citrate, Leucine). Cultures were passaged every 48 h with a dilution factor of 125× and after each transfer, a sample was taken and stored for 16S community sequencing (Fig. 1a). The experiment was conducted for 12 transfers by which point communities appear to have reached a stable population equilibrium. In total, 276 communities were allowed to self assemble (7–8 replicates per Inoculum and 12 Inoculum per carbon source). Community structure was determined at the end of the 12th growth period for all communities using 16s rRNA amplicon sequencing. A subset of communities was sequenced at each transfer allowing community structure to be tracked through time.

Calculating Dissimilarity and Overlap for community pairs

To account for differences in the sequencing depth of different communities we first normalize all communities so that each community is represented by the same number of sequences. Briefly, for each community we create a sample community of N reads sampled from the original read pool without replacement. Here $N = 4397$ was used as this is the minimum number of reads for all communities analyzed. We then calculated the relative abundance of each ESV in each community. For each pair of communities, we follow Bashan et al. [18] and calculate Overlap and Dissimilarity between relative abundance vectors x and y . The Overlap is given by

$$O(x, y) = \sum_{i \in S} \left(\frac{x_i + y_i}{2} \right),$$

where S is the set of ESVs found in both communities. For the calculation of Dissimilarity only the shared ESVs are

considered, and the relative abundance of shared taxa is renormalized to add up to 1. The Dissimilarity between renormalized vectors X and Y can be calculated as root Jensen–Shannon divergence

$$D(x, y) = \sqrt{\frac{1}{2} \left(\sum_i X_i \log \frac{X_i}{M_i} + \sum_i Y_i \log \frac{Y_i}{M_i} \right)},$$

where $M_i = \frac{X_i + Y_i}{2}$.

We calculated the Overlap and Dissimilarity for every pair of our 276 communities at transfer 12 (a total of 37,950 pairwise comparisons). From this dataset, any pairwise comparison in which the communities shared fewer than two taxa in common is excluded. These were removed as for these pairs $D(x, y)$ is always 0. In total this gave us a dataset consisting of 23592 pairwise comparisons.

Fitting DOCs

In Bashan et al. [18] estimated using the Robust LOWESS a nonparametric scatterplot smoothing method. When analyzing all pairs of communities (Fig. 1b, Fig. S2A, Fig. 2c, d), we implement the same method using the LOESS function from R default stats package with the following parameters (span = 0.2, family = “symmetric”, iterations = 5). To compare slopes across different subsets of pairs (Insets Fig. 1b, Fig. 1c–e, Fig. S2B–D, Insets Fig. 2c, d) we use a simple OLS regression on data points with above-median Overlap as was also done by Bashan et al. [18] and later by Kalyuzhny and Shnerb [19].

Estimating confidence intervals and P value for DOC

We implemented the same bootstrapping algorithm used by Bashan et al. [18]. We repeat this bootstrap algorithm 500 times and repeated all our analysis on every bootstrap realization. Confidence intervals in Fig. 1 and Fig. 2 represent 95% percentiles of the curves fitted to the bootstrapped data. The reported p values for the regression slopes (m) represent the fraction of bootstrap realization for which the OLS slope is positive (main text and Fig. 1, Fig. 2, Fig. S3).

Bootstrapped Welch-tests

To account for the nonindependence of groups of sample pairs, all t -tests were performed on every bootstrap realization. The reported p values represent the fraction of bootstrap realizations in which the t -statistic has a different sign to the one calculated from the original dataset. When calculating this fraction we excluded bootstrap realizations in which some groups of sample pairs were unrepresented and so no t -statistic could be obtained.

Randomized data

Each time a DOC is shown, we repeat the analysis on a randomized dataset in which species assemblage and abundance distribution are kept but the abundance of each taxon is randomized following Bashan et al. [18]. The randomized results are shown in Fig. S1, Fig. S3, and Fig. S7.

Code availability

The R Code and datasets analyzed are available at <https://doi.org/10.5281/zenodo.3817698>.

Acknowledgements We thank members of AS and YYL Laboratories for helpful discussion about the work. We thank Josh Goldford for providing us with the experimental data analyzed in this paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci USA*. 2016;113:5970–5.
- Hunter P. Plant microbiomes and sustainable agriculture: deciphering the plant microbiome and its role in nutrient supply and plant immunity has great potential to reduce the use of fertilizers and biocides in agriculture. *EMBO Rep*. 2016;17:1696–9.
- Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. *Science*. 2016;353:1272–7.
- Blaser MJ, Cardon ZG, Cho MK, Dangl JL, Donohue TJ, Green JL, et al. Toward a predictive understanding of Earth's microbiomes to address 21st century challenges. *Mbio* 2016;7:e00714–e00716.
- Ley RE, Peterson DA, Gordon JI. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 2006;124:837–48.
- Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. *Science*. 2016;352:560–4.
- Shafquat A, Joice R, Simmons SL, Huttenhower C. Functional and phylogenetic assembly of microbial communities in the human microbiome. *Trends Microbiol*. 2014;22:261–6.
- Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017;550:61–66.
- Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486:207–14.
- David LA, Materna AC, Friedman J, Campos-Baptista MI, Blackburn MC, Perrotta A, et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol*. 2014;15:R89.
- David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505:559–63.
- Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol*. 2012;10:538–50.
- Goldford JE, Lu N, Bajić D, Estrela S, Tikhonov M, Sanchez-Gorostiaga A, et al. Emergent simplicity in microbial community assembly. *Science*. 2018;361:469–74.
- Machado D, Maistenko OM, Andrejev S, Kim Y, Bork P, Patil KR. Polarization of microbial communities between competitive and cooperative metabolism. 2020. Preprint at <https://doi.org/10.1101/2020.01.28.922583v1>.
- 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.
- Cornforth DM, Foster KR. Competition sensing: the social side of bacterial stress responses. *Nat Rev Microbiol*. 2013;11:285–93.
- Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, et al. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep*. 2014;7:1104–15.
- Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, et al. Universality of human microbial dynamics. *Nature*. 2016;534:259–62.
- Kalyuzhny M, Shnerb NM. Dissimilarity-overlap analysis of community dynamics: opportunities and pitfalls. *Methods Ecol Evol*. 2017;8:1764–73.
- Verbruggen E, Sheldrake M, Bainard LD, Chen B, Ceulemans T, De Gruyter J, et al. Mycorrhizal fungi show regular community compositions in natural ecosystems. *ISME J*. 2018;12:380–5.
- Hubbell SP. *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology 32. Princeton, USA: Princeton University Press; 2001.
- Bajić D, Vila JCC, Blount ZD, Sánchez A. On the deformability of an empirical fitness landscape by microbial evolution. *Proc Natl Acad Sci USA*. 2018;115:11286–91.
- Callahan BJ, Fukami T, Fisher DS. Rapid evolution of adaptive niche construction in experimental microbial populations. *Evolution*. 2014;68:3307–16.
- Laland KN, Uller T, Feldman MW, Sterelny K, Müller GB, Moczek A, et al. The extended evolutionary synthesis: its structure, assumptions and predictions. *Proc Biol Sci*. 2015;282:20151019.
- Lewontin RC. The organism as the subject and object of evolution. *Scientia*. 1983;77:65.