PNAS
www.pnas.org

Supplementary Information for

Higher-order effects, continuous species interactions, and trait evolution shape microbial spatial dynamics

Anshuman Swain*[#1], Levi Fussell*[#2], and William F Fagan[1]
[1]Department of Biology, University of Maryland, College Park, MD, USA
[2]Institute of Perception, Action and Behaviour, University of Edinburgh, Edinburgh, UK
[#]*contributed equally*
*Corresponding author: Anshuman Swain
Email: answain@terpmail.umd.edu

This PDF file includes:

Figures S1 to S6
Supplementary methods

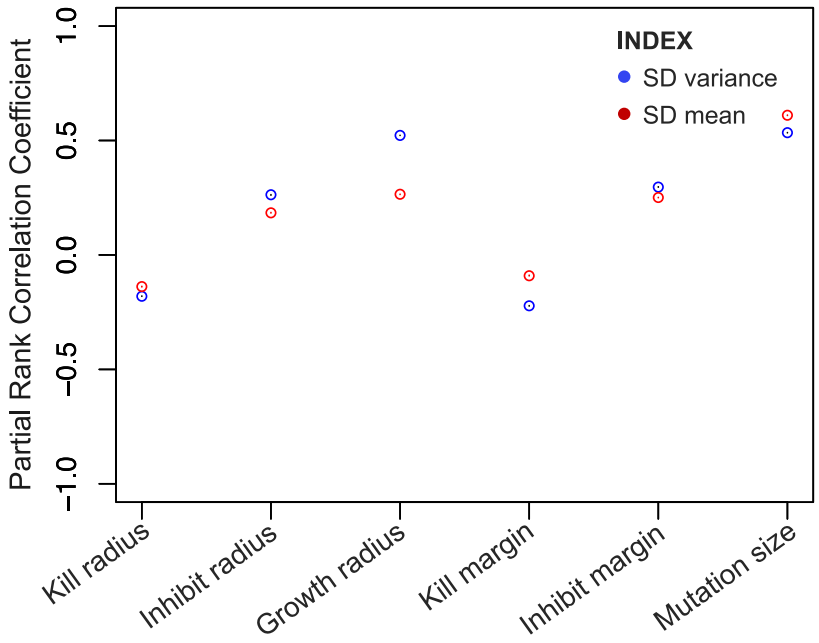

Figure S1: Partial rank correlation coefficients (PRCC) of model parameters in relation to the Shannon Diversity (SD) mean and variance. Please note that different parameters have similar values of PRCC for both the diversity metrics, and the greatest difference in the trends of the two metrics was for *growth radius*.

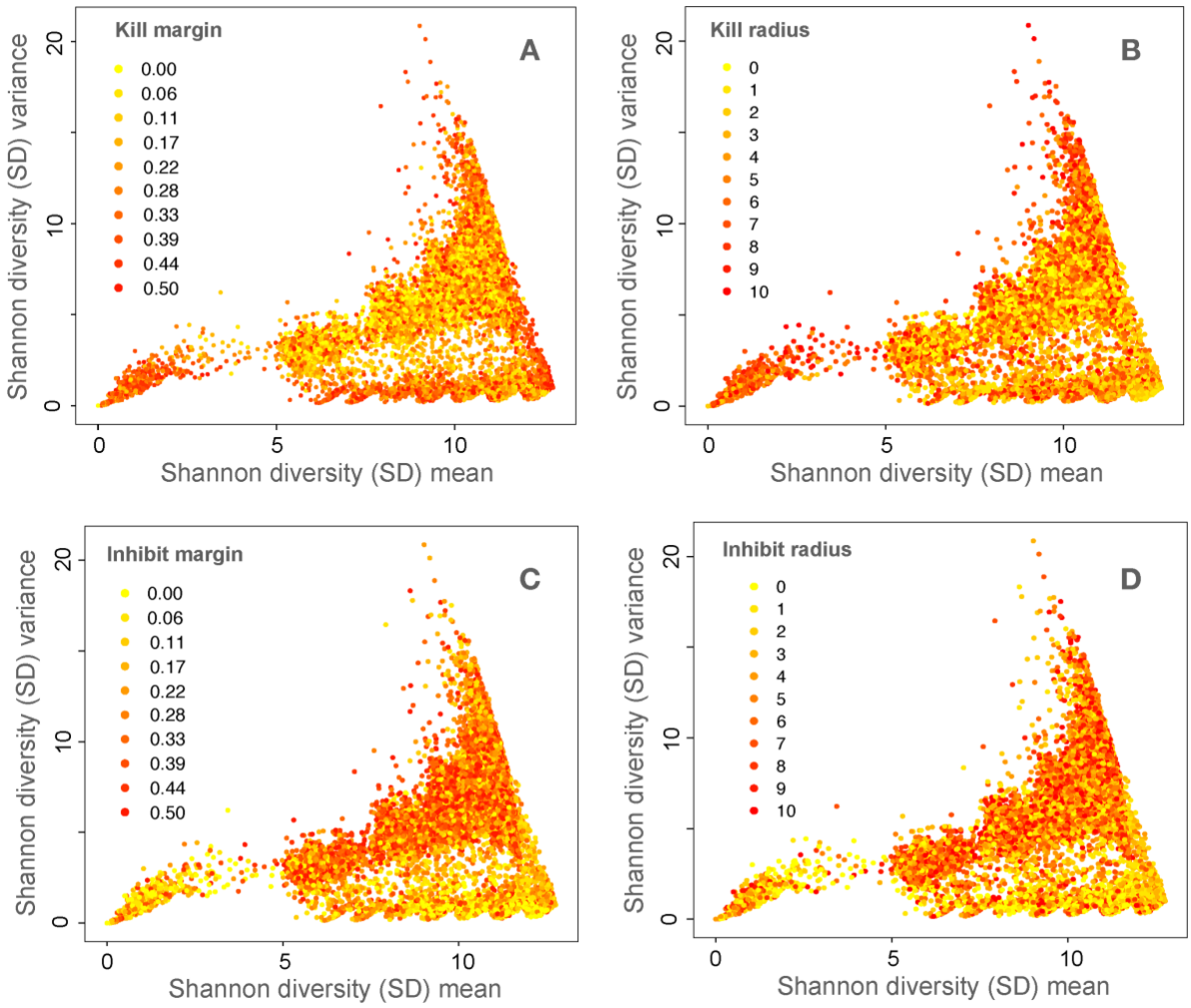

Figure S2: Overlaying values of parameters, (A) *Kill margin*, (B) *Kill radius*, (C) *Inhibit margin*, and (D) *Inhibit radius*, over the respective values on Shannon Diversity (SD) mean and SD variance space. No obvious patterns are discernable in how these parameters affect the diversity space as compared to the ones in Figure 3 D-F, where distinct patterns in clustering of the diversity space can be observed.

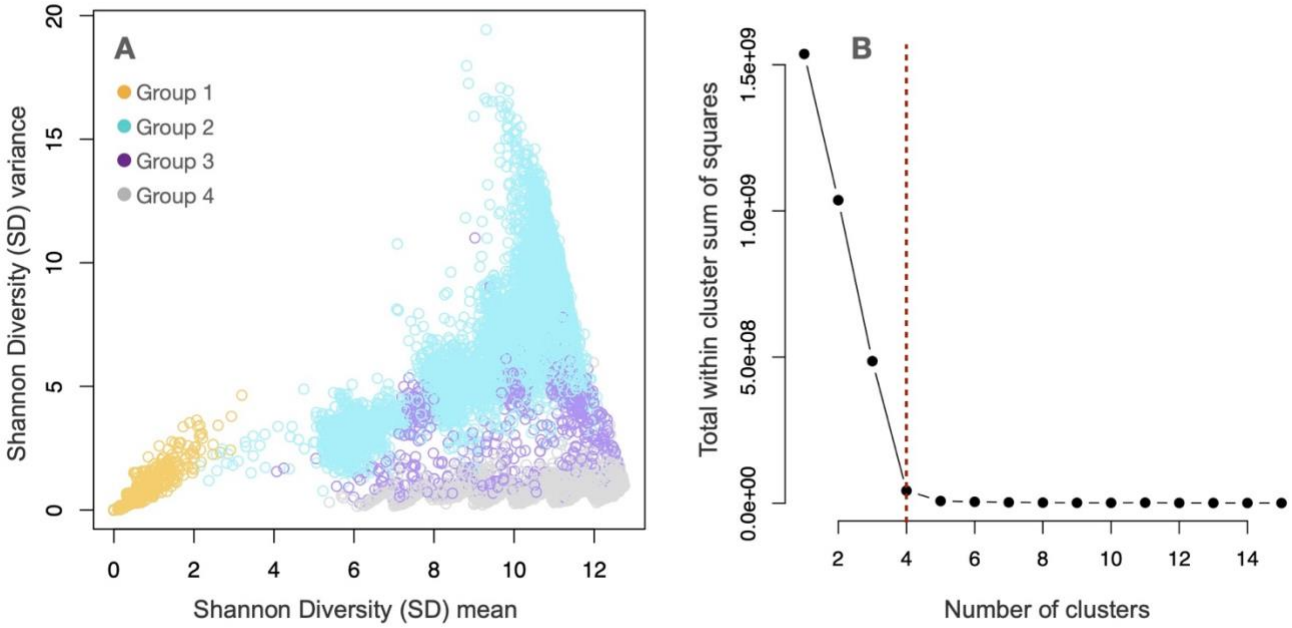

Figure S3: (A) k-means clustering of the Shannon Diversity (SD) mean-variance space by the community formation time (CFT) value. (B) shows the optimal number of clusters using the elbow method. We obtained similar results when optimizing clusters using the D index and the Hubert indices.

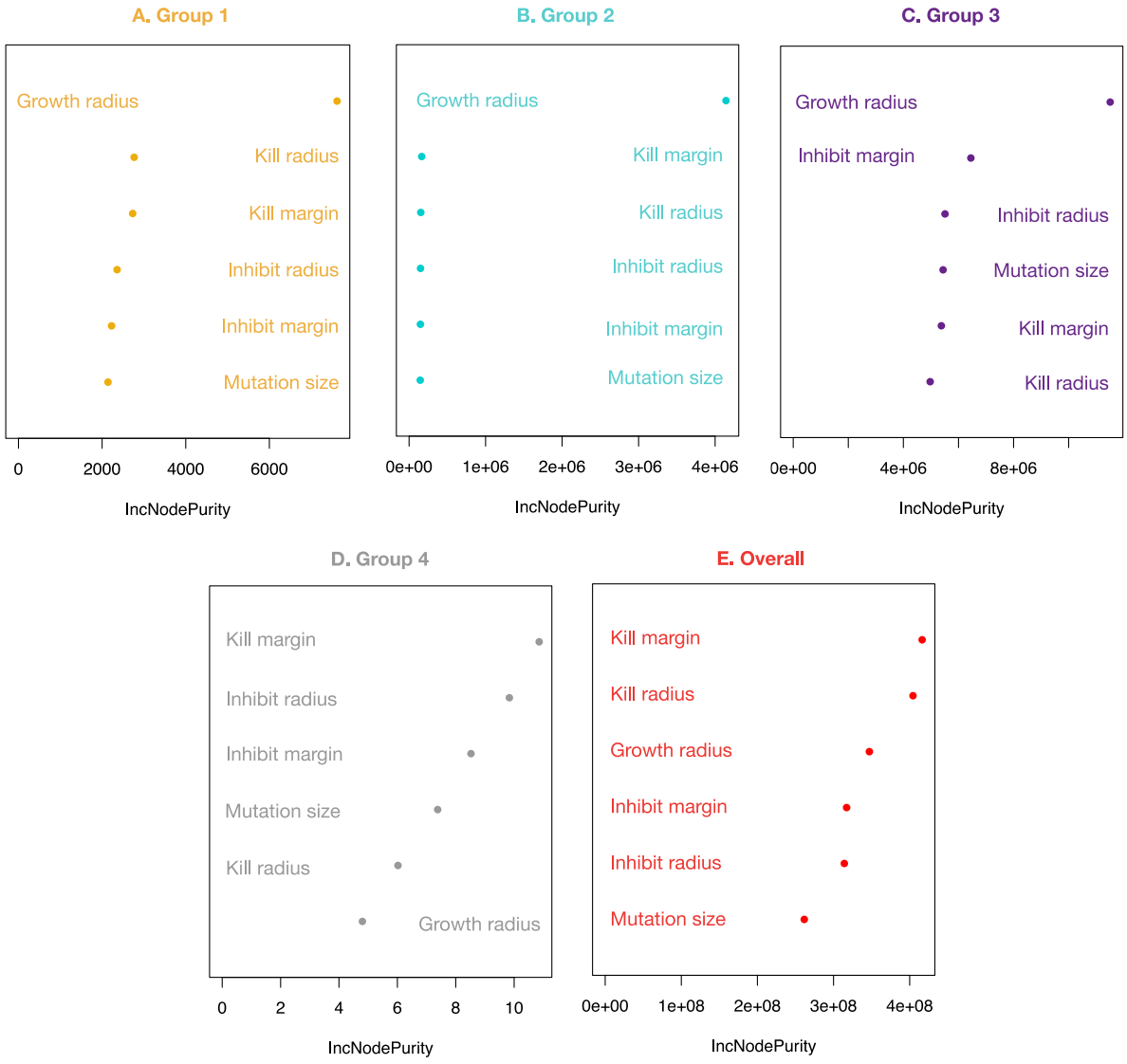

Figure S4: Random forest regressions (10,000 trees each with optimization for splits using out-of-bag (OOB) error) for community formation time (CFT) for the four diversity groups clustered using k-means. Variance explained for- Group 1: 30.2%, Group 2: 90.6%, Group 3: 27.3%, Group 4: 6.0%, Overall: 22.3%. Note: IncNodePurity refers to the total decrease in node impurities from splitting on a given parameter, averaged over all trees.

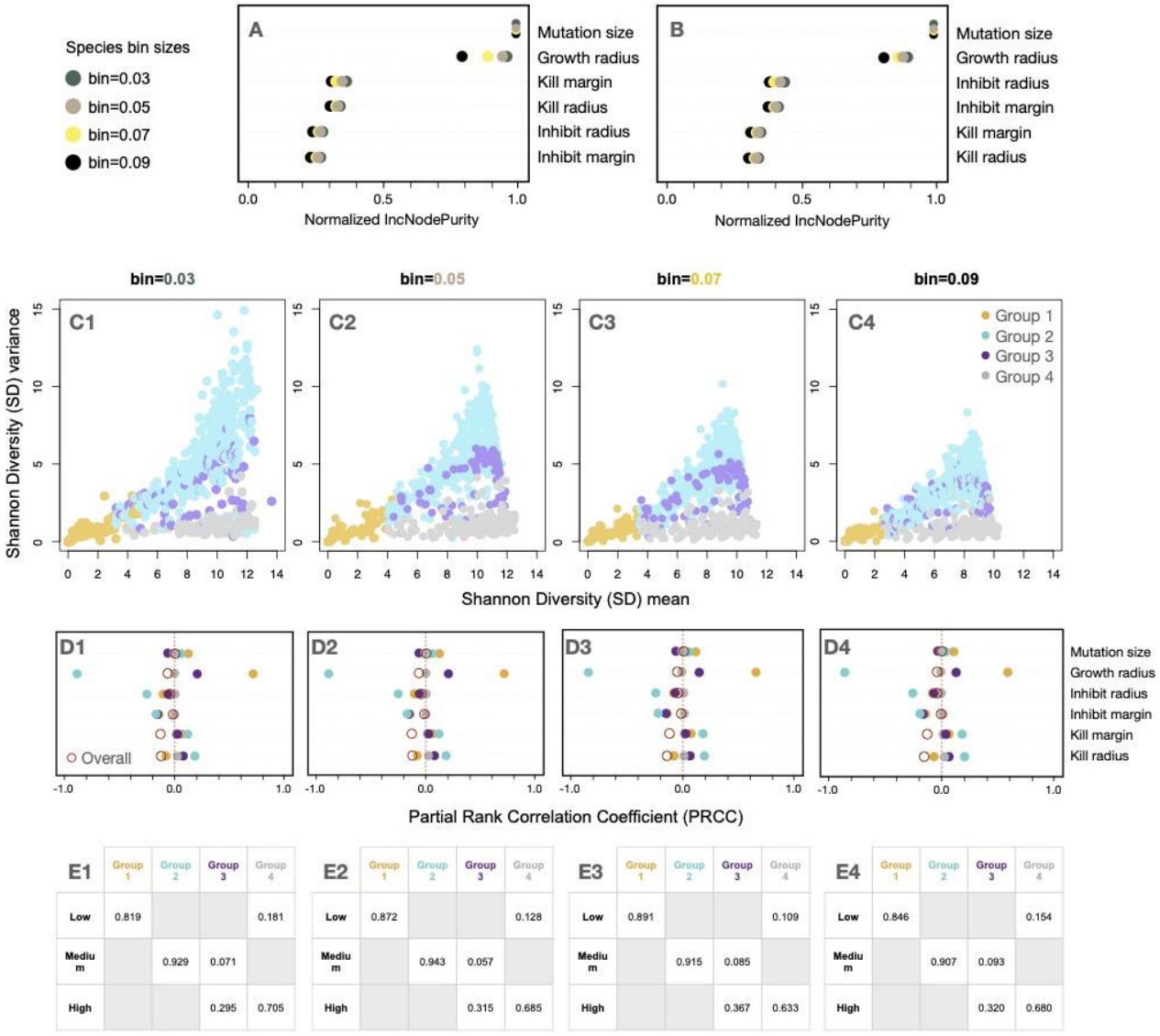

Figure S5: Dependence of the results on species bin sizes. (A) and (B) depict the normalized variable importance scores for a random forest regression of Shannon diversity (SD) mean and variance, respectively, at the four bin sizes. Percent variance explained remained between 68-75% for all bin sizes, and all other results were similarly consistent. (The only thing to note is that the dependence on growth radius (a mobility-related metric) weakens with increasing bin size, which may occur because increasing bin size results in greater discretization for a given length scale). (C1-4) depict the k-means clustering of the space of SD mean and variance for the four bin sizes. Increasing bin size collapses the range of SD mean and SD variance of the resultant communities, but otherwise no changes are obvious. (D1-4) depict the Partial Rank Correlation Coefficients (PRCC) of variables on Community Formation Time (CFT), which are consistent across bin sizes. (E1-4) depict how the spatial categorizations correspond to the diversity groups at four bin sizes, and again the results are consistent.

Supplemental Methods. Species interactions can be generalized by keeping the distance-based threshold function but changing the geometric representation of the trait space by defining some function $f(x)$: $\boldsymbol{R^3} \rightarrow \boldsymbol{R^3}$ which maps the uniform trait space to a *transformed trait space*. The threshold function (margin) is then applied to this transformed trait space to determine the interactions between species. In other words, if we imagine the interactions in our model to be represented by a graph (as in the case of Kelsic et al. 2015), the original trait space defines the identity of the species (i.e., nodes in the graph) and the transformed trait space (after being acted upon by the threshold function) defines the interactions (i.e., the edges in the graph).

Our current model simply uses the identity mapping $f(x) = x$, and therefore the transformed trait space and the original trait space are equivalent. The identity map is the simplest functional assumption. Instead of the uniform toroidal space in our current model, other trait space geometries are possible which could represent very intricate non-linear relationships between how individual species, and their traits, interact and evolve (See Figure S6 for simplified examples). The nature of these interactions is not explicitly known from current experimental literature and there is much debate around it, i.e., there is a dearth of empirical information on the structure and dependencies of these traits (e.g., whether they are independent, or constrained; see Schoustra et al., 2012; Pérez-Gutiérrez et al., 2013; Gonzalez et al., 2019; Granato et al., 2019; Peterson et al., 2020; Gorter et al., 2021). In the absence of any guiding information, we assumed an independent, uniform linear space of microbial traits for simplicity. This leaves room to develop the model to more complex definitions of trait interactions based on future experimental findings.

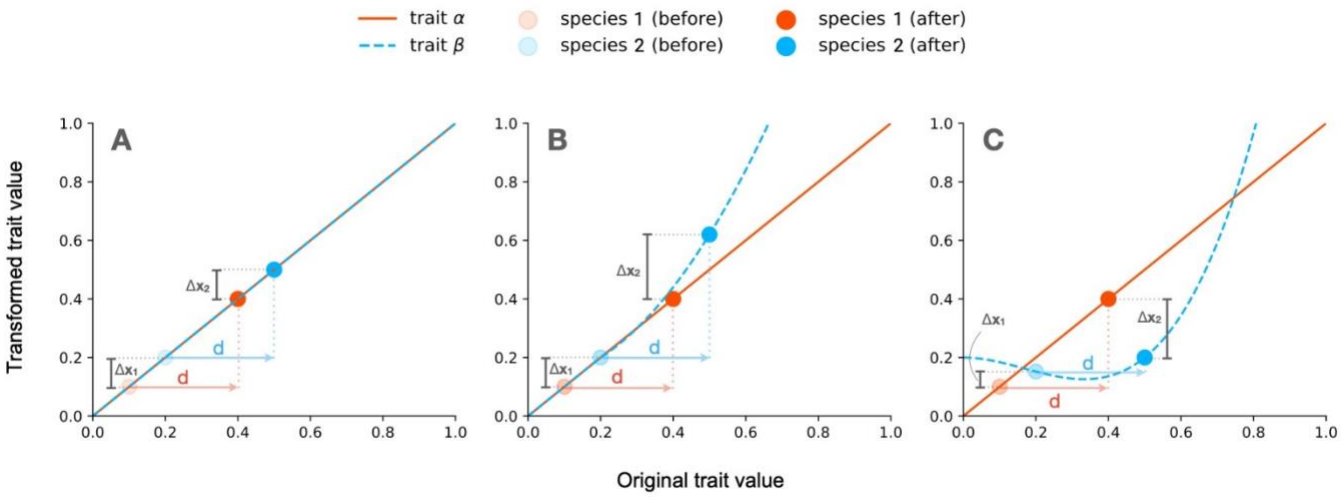

Figure S6: Simple example of two-trait space: (A) depicts the distance between the original trait α for species 1 and trait β for species 2 before and after both traits are mutated by a value '*d*'. The transforming function (i.e., the function that transforms the trait value for use in interaction mapping) is linear (i.e., identity function, as is the case with our model). The trait distance between both species is the same before and after mutation ($\Delta x_1 = \Delta x_2$). These principles of linearity and equidistance underpin the simplicity of our approach to trait evolution. **(B)** depicts the same species and traits, but trait β has a different geometry (here the transforming function is exponential) such that mutating both traits by a value '*d*' does not preserve the trait distance between the two species ($\Delta x_1 \neq \Delta x_2$). **(C)** is similar to **(B)** but with a more non-linear transforming function. Both **(B)** and **(C)** represent more complex (and more specific) assumptions about trait interrelationships than the scenario in **(A)** which we adopted. If we imagine the interactions in our model to be represented by a graph (as in the case of Kelsic et al. 2015), the original trait space (X-axis) defines the identity of the species (i.e., nodes in the graph) and the transformed trait space (Y-axis) (after being acted upon by the threshold function) defines the interactions (i.e., the edges in the graph).

Supplementary References:

1. E. D. Kelsic, J. Zhao, K. Vetsigian, R. Kishony, Counteraction of antibiotic production and degradation stabilizes microbial communities. Nature, 521(7553), 516-519 (2015).
2. S. E. Schoustra, J. Dench, R. Dali, S. D. Aaron, R. Kassen, Antagonistic interactions peak at intermediate genetic distance in clinical and laboratory strains of Pseudomonas aeruginosa. BMC microbiology, 12(1), 1-9 (2012).
3. R. A. Pérez-Gutiérrez et al., Antagonism influences assembly of a Bacillus guild in a local community and is depicted as a food-chain network. The ISME journal 7.3, 487-497 (2013).
4. D. Gonzalez, D. A. Mavridou, Making the best of aggression: the many dimensions of bacterial toxin regulation. Trends in microbiology, 27(11), 897-905 (2019).
5. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, The evolution and ecology of bacterial warfare. Current biology, 29(11), R521-R537 (2019).
6. S. B. Peterson, S. K. Bertolli, J. D. Mougous, The Central Role of Interbacterial Antagonism in Bacterial Life. Current Biology, 30(19), R1203-R1214 (2020).
7. F. A. Gorter, C. Tabares-Mafla, R. Kassen, S. E. Schoustra, Experimental Evolution of Interference Competition. Frontiers in microbiology, 12, 602 (2021).