Novel artificial selection method improves function of simulated microbial communities

Björn Vessman[1,*], **Pablo Guridi-Fernández**[1,2,*], **Flor Inés Arias-Sánchez**[3], **and Sara Mitri**[1,4,✉]

[1]Département de Microbiologie Fondamentale, Université de Lausanne, 1015 Lausanne, Switzerland
[2]Department of Microbiology & Genomics, Instituto de Biomedicina y Biotecnología de Cantabria
[3]BIH Center for Regenerative Therapies (BCRT), Charité - Universitätsmedizin Berlin, Berlin, Germany
[4]Swiss Institute of Bioinformatics
[*]These authors contributed equally

Abstract. There is increasing interest in artificially selecting or breeding microbial communities, but experiments have reported modest success. Here, we develop computational models to simulate two previously known selection methods and compare them to a new "disassembly" method. We evaluate all three methods in their ability to find a community that could efficiently degrade toxins, whereby investment into degradation results in slower growth. Our disassembly method relies on repeatedly competing different communities of known species combinations against one another, while regularly shuffling around their species combinations. This approach allows many species combinations to be explored, thereby maintaining enough between-community diversity for selection to act on, and resulting in communities with high performance. Nevertheless, selection at the community level in our simulations did not counteract selection at the individual level, nor the communities' ecological dynamics. Species in our model evolved to invest less into community function and more into growth, but increased growth compensated for reduced investment, such that overall community performance was barely affected by within-species evolution. Within-community ecological dynamics were more of a challenge, as we could control them during the selection process, but community composition and function dropped in the longer term. Our work shows that the strength of disassembly lies mainly in its ability to explore different species combinations, and helps to propose alternative designs for community selection experiments.

Community breeding; artificial community selection; directed evolution; microbial ecology; community function; pollutant degradation; individual-based model; levels of selection

Correspondence: ***sara.mitri@unil.ch***

Introduction

Humans have been breeding plants and animals for centuries by allowing individuals with the most desirable traits to selectively produce offspring. Also known as "artificial selection" or "directed evolution", breeding has altered traits such as the size of fruits or the enzymatic activity of proteins used in biotechnology (1). More recently, we have started to appreciate that microbes — often multi-species communities of microbes — play an important role for health and the environment. One way to improve or optimize the functions and services that these microbes provide is to select for their traits in the same way as traditional breeding.

However, breeding microbial communities is less straightforward than individual organisms (2–6), mainly because the breeder selects whole groups of organisms rather than individual plants, animals or proteins. According to evolutionary theory, group-level selection suffers from reduced heritability, one of the main requirements for evolution by natural selection (7). The problem arises because a single community-level "generation", which we call a "round of selection" to avoid confusion, can comprise several generations of cells, each belonging to different genotypes (i.e. species or strains). Since the genotypes all reproduce at varying rates, their relative abundances can change during one round of community growth and over subsequent rounds. Because community traits depend on the traits of all of its constituent genotypes and their proportions, an "offspring" community may not resemble its "parent" (7–9). Another issue with community selection is that within- and between-species selection continue to operate within a round. If there are trade-offs between growth and contribution to the community trait, cheaters that contribute less can emerge and sweep to fixation (2, 10). A third challenge is to find a good constellation of different community members and their proportions that can best achieve the desired function. Generating different constellations of member species at each round of selection is also important to have enough variability for selection to act on (7). The major challenges for community-level selection then, are (i) ensuring that community functions are heritable, (ii) that within-community selection does not dominate over between-community selection, and (iii) ensuring variability, that communities differ in phenotype.

Since the earliest community breeding experiments by Swenson *et al.* where microbial communities were selected to in-

crease plant yields and to control pH (11), many attempts have been made, aiming to optimize different microbial community traits (12–21). Some of these studies have managed to significantly improve the *average* community function over several rounds of selection, but sometimes only as an effect of time without any significant differences between selection treatments (11, 18, 19). Overall, community breeding experiments have shown mixed success (3, 4, 22), with computer simulations providing some clues on how to improve them (6, 8, 9, 23–27).

Most previous experiments have followed one of two methods to propagate the communities with the highest scores to the next round: in the "propagule" selection method (PS), a fraction of the cells in the highest-scoring communities are selected and transferred by dilution, while in "migrant pool" selection (MS), all populations of the selected communities are mixed in a *pool* before they are diluted in equal proportions to the new tubes. These methods suffer from a rapid decrease in between-community variability (28), such that selection has little to act on. This suggests that we need novel selection methods that can better explore the search space of species combinations (25).

In this manuscript, we explore a selection method that we call "disassembly selection" (DS) that is designed to maintain between-community variability. After each round, we disassemble the selected communities by isolating the constituent species before recombining them into new communities for the next round of growth. Species are combined in equal proportions at every round to ensure that community traits are heritable. We use an individual-based model of evolving microbes in a well-mixed liquid culture to systematically compare our approach to the classical propagule selection (PS) and migrant pool selection (MS) methods. To inform the design of experiments to select for communities that degrade industrial pollutants (29, 30), microbial communities in our simulations are selected for their ability to degrade toxins.

Our results confirm our intuition that propagule and migrant pool selection do not maintain enough variability to explore many different species combinations. In contrast, our disassembly approach maintains variability between communities, allowing it to find some of the best possible species combinations. Nevertheless, disassembly selection still suffers from an important problem: once our scaffold of restarting communities with equal abundances of each species is removed, communities lose their high function. Our work thereby shows that this new method can find species combinations whose community function is high, but leaves controlling the ecological dynamics of these successful communities as an open problem.

Results

Simulating community-level selection. To make sure that our results are not biased by the type of model we chose or its exact implementation, we developed an individual-based model (IBM) and a model based on ordinary differential equations (ODE). As both models showed very similar results, we present the results of the IBM in the main text and leave the ODE model as a comparison in the supplement.

As these simulations were designed to inform a real artificial selection experiment, parameters were chosen to resemble the experimental setup (30). Each simulation starts with 21 communities of 4 species each, chosen at random with replacement from a set of 15 initial species. We generated 5 such 15-species sets and repeated the simulations for each to make sure that our results are not specific to a given set of parameters. Each species is described by randomly drawn model parameters: its growth and uptake rates for each of 4 available nutrients, its death and degradation rates for each of 10 toxic compounds, and a set of fractions f representing its investment into the degradation of the 10 toxins. The remaining fraction ($1 - f\cdot$, shorthand for $1 - \sum f$) of consumed nutrients can be invested into growth (see Methods). The microbes in our model thereby face a dilemma: whether to invest consumed nutrients into growth or into degradation of toxic compounds that would otherwise cause cell death.

In each of the 21 simulated "tubes", we place cells of different types, nutrients and toxic compounds (Fig. 1A). At each time-step, with fixed probabilities (Table 1), a cell takes up nutrients that get invested into cell growth (activation and division) and mutation, or into toxin degradation. Mutations that occur only alter the investment fractions f, while all other species' parameters remain unchanged throughout the simulations. After 80 such time-steps, each of the 21 communities is scored based on degradation of the ten toxic compounds. The best 7 communities are then selected and propagated to the next round, depending on the selection method (Fig. 1B).

Propagation methods. In propagule selection (PS), the 7 communities with the highest score are diluted to make offspring communities, with the frequency of each offspring community corresponding to the score of their parent. In migrant pool selection (MS), the 7 best communities are combined and then all cells distributed amongst the 21 tubes of the next round (see Methods, Fig. 1B).

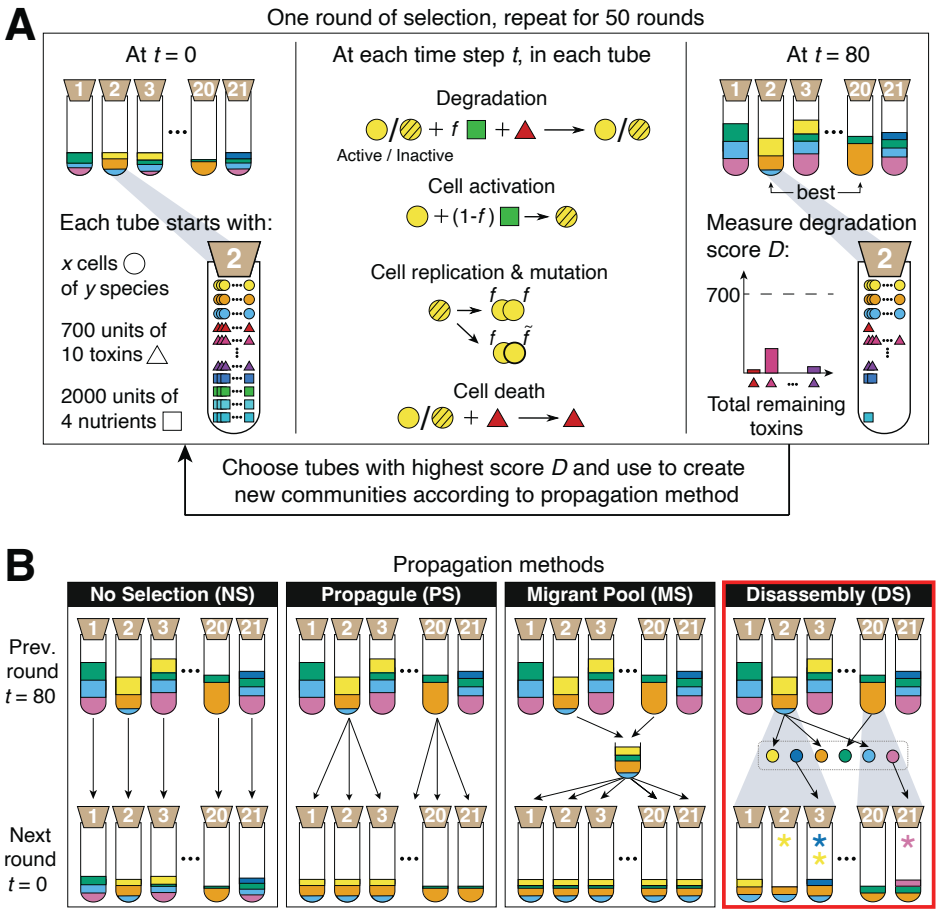

Fig. 1. (A) One round of artificial selection as implemented in the IBM (see Fig. S1 for the ODE model). Communities are illustrated as test tubes with bacterial "species" in different colors. At $t = 0$, each tube is inoculated with cells (circles), toxins (triangles) and nutrients (squares) as indicated. In disassembly selection, $x = 10$ cells of each species are used. We begin all communities with $y = 4$ species in round 0. At each time step t, each cell in each tube is subject to one of four processes (degradation, cell activation, cell division and mutation, and cell death) that occur with different probabilities (Table 1, see Methods for details). After 80 such time-steps, the degradation score D is measured (see Equation 4), and the 7 communities with the highest score (illustrated with 2) are selected for propagation. In the random controls, 7 random communities are propagated. **(B)** Propagation methods. The 7 best communities (represented by tubes 2 and 20 here) are propagated to make the same number of communities for the next round. Propagule selection (PS): selected communities are diluted, contributing offspring communities in proportion to their degradation score. Migrant pool (MS): selected communities are merged and then diluted. Disassembly (DS), the method introduced in this paper, highlighted with a red border: Each species from the selected communities is isolated and saved in a repository (dotted rectangle). Each selected community contributes offspring communities in proportion to their degradation score (grey areas). A fraction of the new communities receive new species (red arrows) or lose members from the previous round (asterisk in color of removed or added species). No-selection (NS): As a control treatment, each community is diluted into a new tube. Propagule, migrant pool and disassembly have corresponding random treatments (PR, MR and DR), where community scores are ignored (see Methods).

In the disassembly method (DS), the 7 best parent communities are chosen based on their degradation score, but we also penalize communities if species extinctions occur. This is meant to favor more cooperative communities. To choose the species composition of each offspring, we replicate the best communities proportionally to their score, and randomly choose 5 offspring tubes to receive or lose a species, respectively. In some cases this can result in species exchange. When adding new species, we select them to ensure that each species is present in at least one of the 21 tubes at each round. For each of the chosen species, we place 10 cells, making sure to represent all the mutant lineages from the parent population (see Methods, Fig. 1B, red box).

In addition to the selection treatments, we also implemented corresponding random controls (e.g. random propagule: PR) where 7 communities are chosen at random instead of according to their score, and a no-selection control (NS) where every community is diluted without selection (Fig. 1B). This last control forms a baseline for how communities change due to species interactions (25, 31), which can only occur through nutrient consumption and toxic compound degradation in our model. We ran 10 replicate simulations for each of the 5 species sets for each propagation method for 50 rounds of selection. The same initial conditions were used for the different selection methods to allow for a fair comparison.

Disassembly selection finds communities whose degradation ranks in the top percentile of all possible

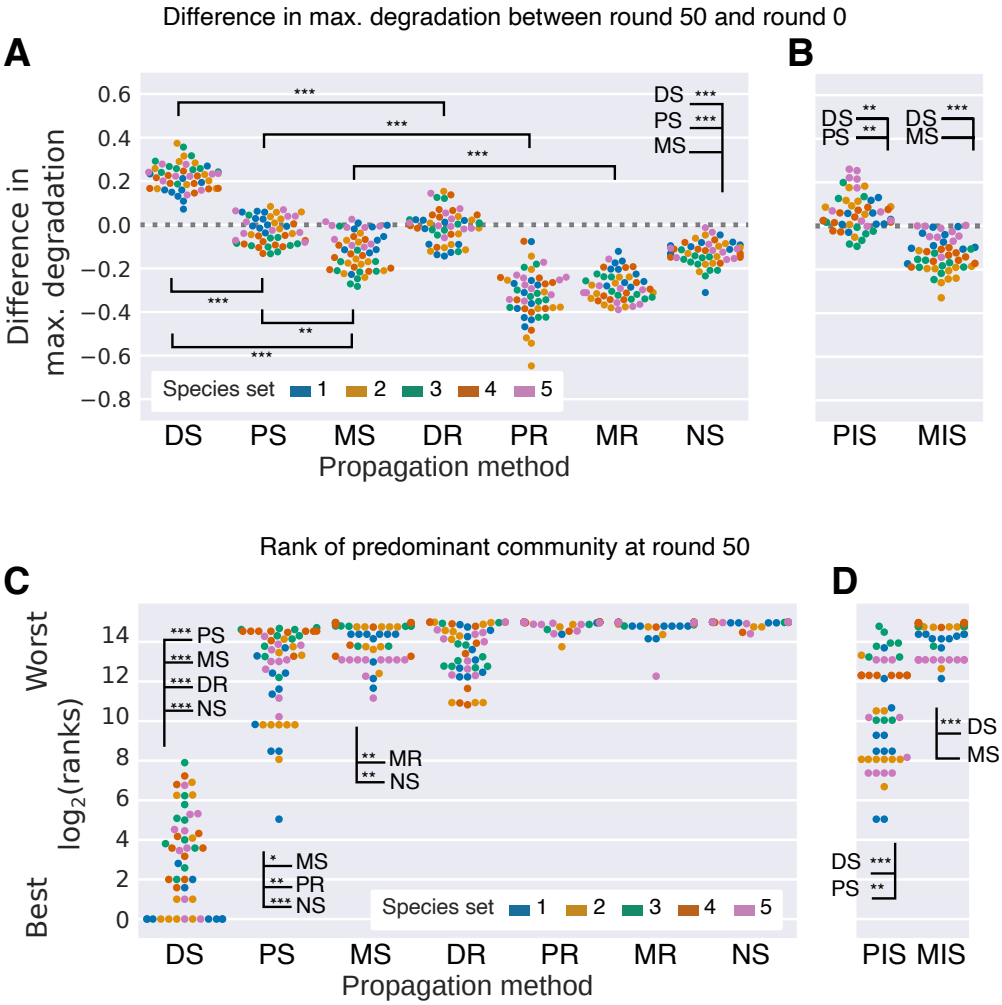

Fig. 2. Degradation scores and ranking of selected communities. **(A, B)** The difference in maximum degradation score between round 50 and round 0 over the 21 communities. **(C, D)** The rank of the predominant community (the most common combination of species among the 21 communities in the last round of selection, not counting sub-communities) in terms of its degradation score compared to all of the 32767 possible combinations of 1, 2, . . . , 15 ancestral species. **(B, D)** We reran the PS and MS propagation methods but where some communities receive invader species (PIS and MIS, respectively). This method was previously suggested by Chang et al. (25). In all plots, each dot represents one of 10 repeated runs, colored by species set, with 50 dots in total. As each run starts from identical communities for all methods, we have compared pairs of runs between the selection methods. Asterisks show the significance of a Wilcoxon signed-rank test for difference in degradation between methods (*: $p < 10^{-3}$, **: $p < 10^{-6}$, ***: $p < 10^{-9}$). Corresponding plots generated using the ODE model are shown in Fig. S3.

communities. All simulated selection methods succeeded in improving the median degradation score across the 21 communities between round 0 and 50 (Fig. S2), which is consistent with previous work (3, 25). However, DS was the only propagation method to significantly and consistently improve the maximum degradation score, meaning that on average, the best community in round 50 degraded significantly better than the best community in round 0 (one-sided Wilcoxon signed rank-test $n = 50$, 10 repeated runs of 5 species sets, $p < 10^{-9}$, Fig. 2A). The increase in maximum score in DS (0.22 ± 0.06), was also significantly different from the classical selection methods (-0.03 ± 0.06 and -0.12 ± 0.09 for PS and MS), from its own random control (DR), and from NS (all two-sided Wilcoxon tests of difference in max. degradation between DS and other methods, $n = 50$, $p < 10^{-9}$).

As an absolute performance measure, we computed the degradation scores of all $2^{15} - 1 = 32767$ possible communities consisting of 1 up to 15 species for each species set and sorted them from best to worst. The communities found by DS ranked among the best few hundred, finding the very best community out of 32767 in 17 out of 50 runs (Fig. 2C).

Communities selected by disassembly invest more into degradation and are composed of diverse species with complementary phenotypes. In our model, community performance depends on (a) the overall investment into degradation of toxic compounds relative to growth, and (b) how well community members complement each other. Community members will compete less if they take up different nutrients while the degradation score of a community

can increase if its members specialize on degrading different toxic compounds (Eq. 4).

To understand how these two properties changed over time, we first quantified the fraction $f_\cdot$ of resources invested into degradation summed over all toxic compounds, averaged over the species in each community. Starting from an average investment of 0.5, DS finds communities that invest significantly more resources into degradation at round 50 than in the first round (one-sided Wilcoxon test, all $p < 10^{-9}$, $n = 50$, Fig. 3A). This is not due to any single species with unusually high degradation capabilities, but rather because DS finds a combination of species with high investment. The average within-community species diversity increases over the 50 rounds (Fig. 3C), which means that the communities consist of an increasing number of species and/or that the communities are increasingly even. Accordingly, in DS, the effective number of consumed nutrients and toxic compounds increases over the 50 rounds (Fig. 3D, E). This increase in coverage and community diversity was not observed for the other selection methods (Fig. 3C, Fig. S5).

Given the complementarity in nutrient uptake and toxic compound degradation, one might expect species to grow and degrade better together compared to when they are alone, as they may be facilitated by other species that remove toxic compounds that they themselves cannot. We use "synergy" to quantify whether a community property (e.g. degradation) is greater than that of its member species together (Fig. 3F, G). Against a baseline of all possible species combinations for a given community size – richness in our models increases niche overlap and competition for resources, which decreases synergy – communities selected by DS have significantly higher synergy, for both degradation and cumulative biomass (Kruskal–Wallis H test, $p < 10^{-9}$, Fig. 3F, G).

In sum, communities selected by DS invest more into degradation compared to communities from other methods. These communities are diverse in composition, consist of species with minimal niche overlap, and cover the toxic compounds evenly (Fig. 3F).

Disassembly can explore more species combinations by diversifying the selected communities. Seeing that communities selected by DS are diverse and efficient degraders, we now investigate how the method finds these communities. First, DS explores more species combinations than the other methods (Fig. 4A, Fig. S6). The classical propagule method (PS) can only find sub-communities of the species combinations present in round 0. Similarly, while migrant pool (MS) is in principle able to search all sub-communities of the first set of selected communities, they are in practice limited to a smaller subset as species tend to go extinct due to the toxic compounds, inter-species competition and/or the dilution bottleneck at each round. Accordingly, most communities available for selection by PS or MS resemble one another, seen as a rapid drop in between-community (or beta) diversity (Fig. 4B, Fig. S7). In contrast, changing the species composition of some selected communities by inserting or removing species at random, DS can search a larger number of communities and the drop in beta diversity resulting from selection is not as steep. The beta diversity of the no-selection control depends on the diversity of the initial communities.

Propagule but not migrant pool selection performs better by periodically adding species to communities. In the disassembly method, more and better communities can be found by randomly adding and removing species in some of the communities. To explore whether species introduction could improve PS and MS in our models (previously shown for PS (25, 27)), we implemented two new versions (PIS and MIS), where in each round, a fixed number of communities chosen at random will receive one or more "invader" species (also chosen at random) with a defined initial population size (see Methods).

With this modification, PIS increases the maximum degradation (one-sided Wilcoxon signed-rank test of D in round 50 versus 0, $p < 10^{-3}$, $n = 50$, Fig. 2B) and improves upon the standard PS method (two-sided Wilcoxon signed-rank test, $p < 10^{-6}$, $n = 50$). In our parallel implementation in the ODE model, however, the PIS method still improved upon the PS method (two-sided Wilcoxon signed-rank test, $p < 10^{-3}$, $n = 50$), but we did not find any significant improvements in the maximum degradation score compared to round 0 ($p = 0.9$, $n = 50$, Fig. S3B). PIS finds higher-ranking communities than PS in both models (two-sided Wilcoxon signed-rank test for differences in ranks between PIS and PS, IBM: $p < 10^{-6}$, $n = 50$, Fig. 2D, ODE model: Fig. S3D) and can explore more combinations than the regular PS. Finally, the initial drop in beta diversity is less severe (Fig. 4B), indicating that there is more variability for selection to act on. In contrast, MIS does not improve significantly on MS, either in terms of degradation, ranks or investment. Even though MIS explores more species combinations than MS, the beta diversity rapidly drops (Fig. 4B), and the introduced species do not contribute much to diversity or degradation of the resulting communities.

Mutation and selection can decrease per-species investment, but this increases biomass, maintaining

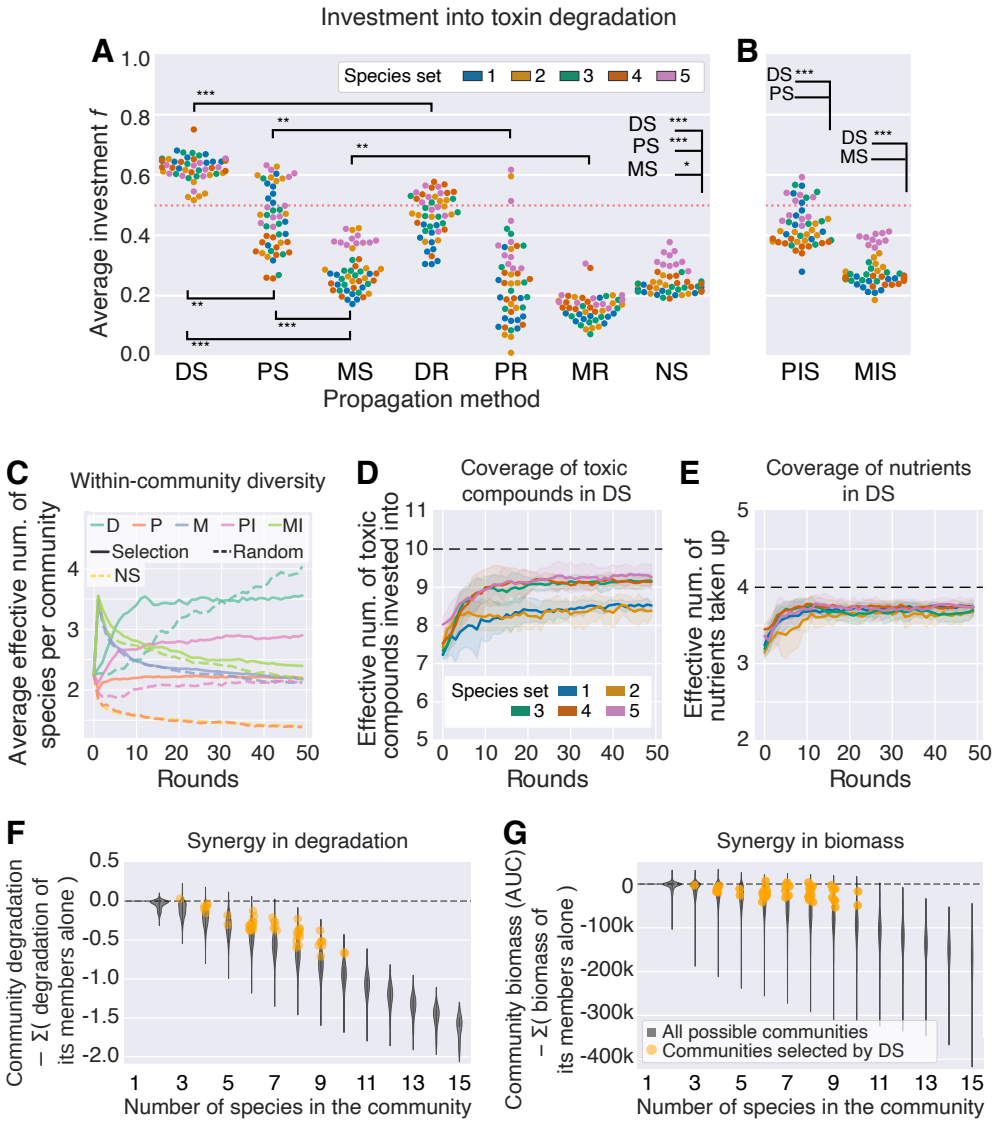

Fig. 3. Total investment into degradation, diversity, coverage and synergy. **(A, B)** Average total investment in a community at round 50, averaged for the 21 communities in a run. For each species we calculate the average investment weighted by strain population and then we do the unweighted average of all the species in the community. The red dotted line at 0.5 indicates the theoretical mean investment at round 0. One dot for each of 10 repeated runs, colored by the 5 species sets, with 50 dots in total. The asterisks indicate the results of a Wilcoxon signed-rank test with $n = 50$ (*: $p < 10^{-3}$, **: $p < 10^{-6}$, ***: $p < 10^{-9}$). Corresponding plots generated using the ODE model are shown in Fig. S4. **(C)** Within-community species diversity measured as the effective number of species Eq. (7), averaged over all 21 communities in a run. The line shows the average over the 10 repeats and the 5 sets of species, with error bars per set of species in Fig. S5. **(D)** The coverage of toxic compounds and **(E)** nutrients in communities selected by DS, measured as the effective number invested into (f_{ik}) or taken up (n_{ij}) respectively within a community (mean $\pm$ s.d. over the 10 repeated runs for a given set of species), Table 1 (see Methods). **(F, G)** Synergy in communities selected by DS at round 50, grouped by community richness. Synergy is the difference in **(F)** degradation scores or **(G)** biomass between a co-culture and the sum of the values of the corresponding monocultures. The violin plot shows the distribution of synergy for each possible community of that richness level after one round of growth. The dots show the average synergy per repeated run in the last round for the 5 species sets selected by DS. The average species richness per repeated run is rounded to obtain an exact value. For visibility, we have plotted all species sets in the same color.

community degradation. We have shown that DS can improve degradation by exploring many different species combinations and find ones that rank highly. Shuffling species around is, however, not the only way to improve degradation scores. Our models allow for mutations to the parameter f that determines the trade-off between investment into degradation and biomass production for a cell. If a mutant is more competitive than its parent, it can replace the original type in future rounds, even as other species come and go around it. To investigate the effect of mutations, we compare the investment into degradation of species at round 50 to that of their ancestors from round 0, and analyze how these changes affect degradation at the community level.

In DS, the total per-species investment $\sum_k f_{ik}$ (abbreviated as $f_{i\cdot}$) into degradation was significantly lower after 50 rounds of selection than that of the corresponding ancestral species (one-sided Wilcoxon signed-rank test of $f_{i\cdot}$ in round 0 versus 50, $p < 10^{-6}$, $n = 50$, Fig. 5A, Fig. S9). Given

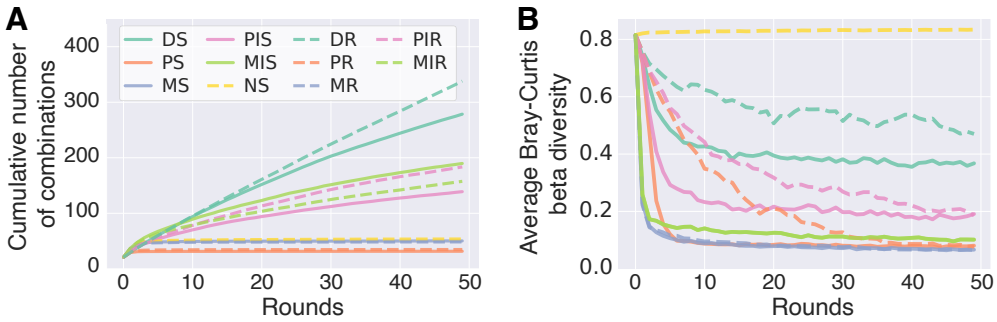

Fig. 4. Cumulative number of communities found by the selection methods and between-community diversity explain how DS can find better communities. We show the mean over the 10 repeats and 5 species sets for each propagation method and refer to Fig. S6-S7 for the full results. Data shown here for the IBM model and in Fig. S8 for the ODE model. **(A)** Time series of cumulative number of unique communities for each selection method. **(B)** Between-community or beta diversity, calculated as the average Bray–Curtis distance of each pair of communities.

the trade-off between investing into growth versus degradation, the communities made up of evolved species had greater total biomass than communities composed of the corresponding ancestral species (one-sided Wilcoxon signed-rank test of total AUC in round 0 versus 50, $p < 10^{-9}$, $n = 50$, Fig. 5B), such that overall, the degradation of the evolved communities was marginally but significantly higher ($+6 \times 10^{-3}$ units, averaged over all species sets, one-sided Wilcoxon signed-rank test $p < 10^{-3}$, $n = 50$) than that of communities made up of their ancestors (Fig. 5C, Fig. S10). Compared to the improvement in degradation due to finding better species combinations, the improvement due to species evolution is very small and is not likely to have a large effect on the outcome of selection. To further quantify this effect, we reran our algorithm, setting mutation rate to 0. We observed no significant difference in the performance of our algorithm with and without mutation (comparing the difference in maximum degradation between generations 50 and 0 with mutation rate $\mu = 0.1$ and $\mu = 0$, one-sided Wilcoxon signed-rank test : $p = 0.19$ and $p = 064$ for DS and DR respectively, $n = 50$, Fig. S11).

In summary, the disassembly method improved the degradation scores over the 50 rounds of selection by finding better species combinations. Within those communities, individual species evolved to invest less into degradation and more into biomass production. As an effect of the trade-off between degradation and growth, the communities still maintain their degradation capabilities and the most efficient communities are the species combinations found in round 50, composed of either their ancestral or evolved genotypes. Whether or not species were allowed to evolve had no measurable effect on performance.

Communities selected by DS are less stable than those selected by PS and MS. The disassembly method has features to ensure heritability and promote within-community diversity: we re-inoculate species in fixed and equal abundances, punish extinctions and re-inoculate extinct species. Controlling the ecological dynamics so tightly means that if we were to simply transfer these communities without adjusting relative abundances and without selection, as in the no-selection treatment, they could drift towards a different equilibrium with a lower degradation score. To assess the ecological stability of the selected communities, we transferred the best communities from round 50 for an additional 25 rounds of growth and dilution, this time without selection (Fig. 6A) and found that the degradation scores of communities selected by DS dropped by -0.21 ± 0.14 on average when left to their natural dynamics, close to how much the selection method increased the degradation (0.22 ± 0.06). This indicates that the high performance of these communities relied on controlling the ecological dynamics. Once ecologically stable, the communities converge to a degradation score that is not significantly different to the average of the initial communities (one-sided Wilcoxon signed-rank test, $p = 0.24$, $n = 50$, Fig. 2A, B).

In contrast, the degradation does not drop as much in communities selected by the classical methods PS and MS (-0.02 ± 0.03 and -0.03 ± 0.03 in max degradation, respectively, Fig. 6A). The methods are stable in the sense that the communities do not change much after the first few rounds of selection, either in terms of composition (Fig. 3E) or degradation (Fig. S12). The methods with invasion, PIS and MIS, show an intermediate drop in degradation (-0.07 ± 0.07 and -0.07 ± 0.04) indicating that the invasion step has an effect on community stability. In order to remain effective, the communities found by DS should be grown in the same conditions as they were selected, i.e. (i) from equal abundance, (ii) without any intermediate rounds of growth in between rounds of selection. The latter has been suggested to stabilize the dynamics and improve community selection (25).

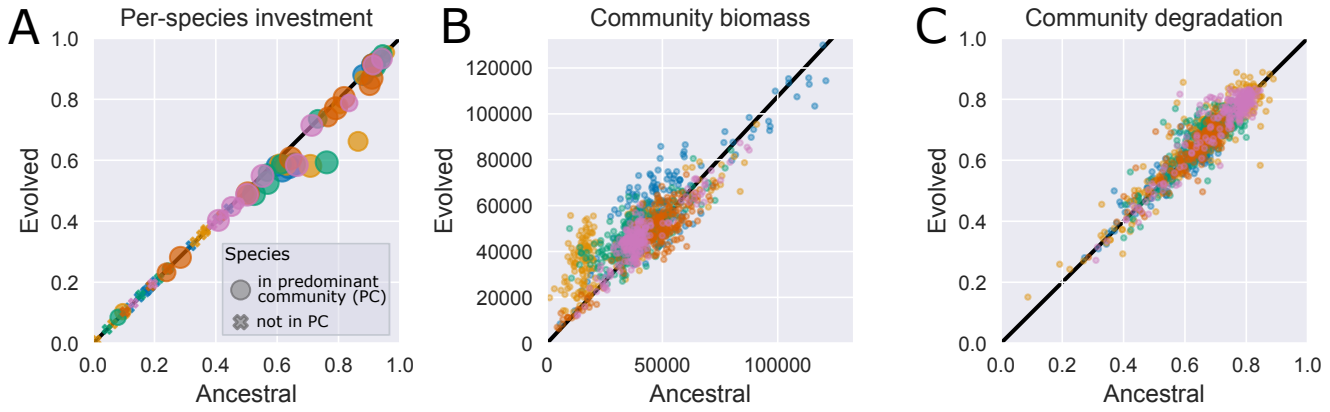

Fig. 5. Change in degradation investment per species and the effect of such change in the biomass and degradation of the communities in the last round, colored by species sets. All results from the IBM. **(A)** Total investment $f_i = \sum_k f_{ik}$ for each species $i = 1, 2, \ldots, 15$. There are 75 dots: 15 species, colored by the species sets 1-5. For evolved condition, the markers show the average total investment of the species, weighted by population size for each occurrence in communities from the last round of selection where the species is present and not weighted between repeated runs; while ancestral condition corresponds to the total investment of each ancestral species prior to any growth or evolution. Species that were present in a predominant community (P.C., see definition in the caption of Fig. 2) in the last round are shown as circles, where the radius is proportional to the number of repeated runs where the species appear in a P.C. Species that were never present in a P.C. are represented by crosses. We summarize the p-values of Wilcoxon tests of whether the investment is different in the last round of selection compared to the first in Fig. S9. **(B)** Total biomass per community, measured as the sum of the area under the growth curves (AUC) for species in the community. The initial AUC is calculated from one round of growth, where the community is composed of ancestral strains of the same species in the same proportions as the last community. There are 1050 dots: 21 communities per 10 repeated runs, for each of the 5 species sets. **(C)** Degradation scores of the same 1050 communities in (B). The degradation of the initial community is calculated over one round of growth when the community is composed of ancestral strains of the same species.

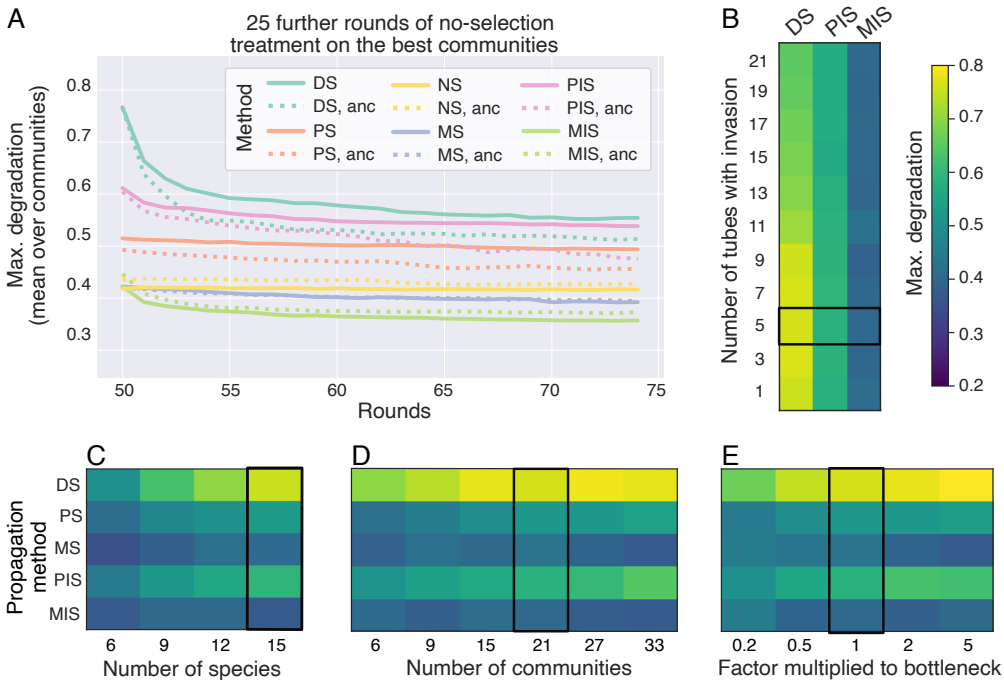

Fig. 6. **(A)** Stability in degradation over 25 additional rounds after releasing the selection pressure, calculated for the highest-scoring community for each selection method. Results show the average degradation for each repeat. The dotted lines show the corresponding degradation when the community is composed of ancestral species. **(B-E)** Max. degradation score in the last round of selection as a function of the experimental parameters: **(B)** the number of tubes to receive or lose a species, **(C)** the number of species in the initial set, **(D)** the number of tubes or communities and **(E)** a scaling factor for the dilution bottleneck. The heat maps show the median degradation score over all species sets and repeated runs (in (C) also over sub-samples) for each selection method. Fig. S13-S16 show the full data set. The color bar is the same for panels B-E. The black outline marks the parameter value used throughout the rest of the paper. All data from the IBM.

Varying experimental parameters to decrease the size of the experiment. Our model shows that DS can outperform other propagation methods, as long as the ecological dynamics of the communities are controlled. However, DS is more cumbersome than the other methods from an experimental perspective: constantly dis- and re-assembling communities and having to adjust the population sizes of each species at every round can cost a lot of time and resources. We now investigate how four experimental parameters impact the degradation scores, and affect experiment size. We focus on DS but also compare it to the other methods (Fig. 6B-E, Fig. S13-S16).

The parameter with the strongest effect on experiment size and the maximum degradation score is the number of species in the initial set (Spearman's rank correlation coefficient $\rho = 0.67$, $p < 10^{-9}$, Fig. S14). This means that the meta-community needs to be as rich as possible to efficiently improve degradation, and the main effort should be invested into managing a larger number of species, ideally by adding species that have positive effects on degradation or the growth of others. In contrast, the number of communities clearly affects experiment size, but it had a weaker correlation to degradation for DS ($\rho = 0.35$, $p < 10^{-9}$), meaning that the number of communities could be decreased, which would reduce effort with a limited effect on community performance. The number of communities receiving an invading species is negatively correlated with degradation ($\rho = -0.29$, $p < 10^{-9}$). Introducing species to a smaller number of communities should improve the final degradation score (Fig. S13), without affecting experiment size. Finally, the dilution factor (i.e. how large a fraction of the culture to re-inoculate for the next round of growth) is positively correlated to degradation scores ($\rho = 0.52$, $p < 10^{-9}$).

Discussion

The major challenges for community breeding are ensuring (i) that the community function is heritable, (ii) that within-community selection does not dominate over between-community selection and (iii) that communities differ sufficiently in phenotype to allow selection to act. While other theoretical studies have investigated heritability and the balance between within- and between-community selection (9, 23–25), our "disassembly" method contributes to improving the third point: how to maintain variability between communities.

In our models, disassembly improved significantly upon the maximum degradation scores of simulated synthetic communities, compared to a random line and a no-selection control. The method further outperformed the classical propagule and migrant pool methods, which could only improve the maximum function for some initial combination of species, confirming previous findings (3, 25). These methods suffer from the loss of between-community variability, which can be circumvented by periodically invading them with new species, as proposed previously (25, 27) and confirmed here. The maintenance of diversity in the disassembly method is achieved by exchanging species between communities, an idea inspired by the crossover operator in genetic algorithms in the field of evolutionary computing (26, 32, 33), whereby building blocks are recombined between digital individuals and generate variability for selection to act on. Of course, crossover in genetic algorithms is itself inspired by recombination in sexual organisms (34, 35). While interesting, these parallels do not map directly to species exchange in community breeding, where the units that are subject to exchange are well-defined and have their own ecology and within-species evolution.

Our approach was also designed to address the second challenge of community breeding: reducing within-community and within-species selection compared to between-community selection. Beyond the practical goal of improving community performance, this question is also a fundamental one: can one apply selection such that group-level selection dominates over individual-level selection? We designed our algorithm for high group-level selection in two ways: first by competing communities against one another in each selection round and second by penalizing between-species competition by reducing the score of communities in which species were driven extinct. These decisions favored communities with more synergistic effects (Fig. 3F, G). To follow how selection within species affects the outcome, we allowed species to acquire mutations in their investment into degradation (unlike other theoretical studies (25)). We found that competition between individuals is indeed strong in our simulated communities, leading species to evolve reduced investment into degradation in favor of growth. Interestingly, though, the faster growth rate of these evolved species compensated for the reduced investment by producing larger population sizes of degraders, such that the difference in degradation between ancestral and evolved communities was negligible. We expect this outcome to apply whenever community function is coupled to growth. Whether such a trade-off exists in experimental systems is therefore an important consideration to take into account.

The problem that we have not yet addressed then is heritability, which is crucial for evolution (36) and a major challenge

for community selection (9). As in (9), we sidestep the issue of heritability and ecological stability in the disassembly method by always re-assembling communities in a fixed abundance and equal species proportions. This allows community dynamics to unfold in almost the same way in each round during selection. But once the winning communities were found and propagated by regular growth and dilution, we found that they were not ecologically stable and changed in both composition and function. One way to overcome this would be to include a few rounds of transferring between the rounds of selection to allow the communities to equilibrate before each selection step (25). This would, however, make disassembly quite inefficient, in which case, propagule with invasion can be more stable than disassembly (Fig. 6A). Another solution could be to bias our search space towards communities that have higher heritability from the start, i.e. communities that behave more like "evolutionary individuals" (37). In this way, one would aim to achieve individuality first, and selection second, increasing the likelihood that the winning community would be one that does not lose its properties after the end of the selection experiment (37).

Ultimately, the goal of our investigation was to help design community selection experiments. To explore all the factors that might matter in such experiments, we opted here to build models that are rather complex and strive to overcome this complexity by building two different models (IBMs and ODEs) to support the generality of our conclusions. The decision to allow for genetic mutations, for example, may seem excessive, but what we learned from it is important: in experimental systems where a trade-off between function and growth exists, within-species evolution may play a minor role in affecting performance. This means that experimentally, one need only work with ancestral species, avoiding the need to disassemble communities into their individual members at every round. This conclusion is also supported by our experimental work (30). Taken together, this suggests that future implementations of this algorithm can avoid the experimentally impractical disassembly step altogether, making the approach significantly more accessible. We also found that the parameter with the biggest effect on the method is the size of the initial species set and that the disassembly method is more efficient for larger sets of species (Fig. 6). If disassembling communities is not necessary, increasing the species pool would be much more feasible. Finally, we find that for disassembly, a higher number of communities grown in each selection round can compensate for the number of selection rounds needed since we are then able to search more distinct communities per round of selection.

We have made a number of simplifying assumptions in our models. First, we assume a well-mixed liquid culture, where in reality, clumps may affect species interactions and community function. We also assumed a trade-off between degradation and growth in both models and that the toxic compounds cannot be used as nutrients. Both assumptions serve to decouple the community phenotype from population growth and while they are not completely independent, a smaller contribution from growth on the community phenotype should make artificial selection more difficult. In a sense, this is the more interesting problem to explore, since all the methods we explored are expected to improve a community phenotype that is aligned with population growth. Further, we have assumed that species diversity is key to functional diversity: each species can only degrade a subset of the toxic compounds in our model and complete removal of the compounds depends on finding other species with complementary degradation capabilities. Within-community diversity is in this way fundamental for community success, and also decreases competition, as each species only uses a subset of the nutrients. In our simulations, it is therefore unlikely that any mono-culture scores higher than a multi-species community, and the median size of a selected community was 6.9 ± 1.8 species after 50 rounds of selection by disassembly (in the IBM model). This optimum will, however, differ for each system (29).

Taken together, we have computationally explored an approach to community selection, where species composition is shuffled between competing communities, allowing for a greater exploration of the space of possible communities to find the best performing ones. In doing so, we have shown that community function can improve with respect to randomly assembled communities, but that genetic mutation can contribute to reduced investment by individual strains into community function.

Materials and methods

We separately implement an individual-based model (IBM) of individual cell growth and a system of ordinary differential equations (ODEs) that model population-level dynamics. We describe the IBM here and the ODE model in supplementary methods. All parameters are listed in Tables 1 and 2.

Setting up the simulations. A set or pool of 15 species (the "metacommunity") is generated prior to the start of the simulations by fixing parameters for each species i describing its intrinsic growth rate (a_i, r_i), its preference for each nutrient j (n_{ij}), how it is affected by (m_{ik}) and degrades (f_{ik}) each

toxic compound k (Alg. 2, Table 1, Fig. 1A).

We then prepare 21 simulated well-mixed batch culture tubes. At $t = 0$ in the very first round, each culture tube contains $S_i = 10$ identical cells of each of 4 microbial species, $T_k(t_0) = 700$ units of each of the 10 toxic compounds, which can cause cell death, and $N_j(t_0) = 2000$ units of each of the 4 nutrients that allow the cells to divide and reduce the concentrations of toxic compounds (Fig. 1A). We assign the 4 species in each tube by selecting randomly with replacement from the pool of 15 species, making sure that each species is present in at least one of the 21 communities.

Toxic compound degradation, cell division, nutrient consumption, and death. Each round of selection consists of 80 simulated time-steps. At each time-step, growth, cell division, and death occur according to probabilities that are fixed for each species i when the initial species set is generated (Table 1), and these probabilities are used for random sampling from a Poisson distribution (24). How much a cell of type i invests into toxic compound degradation (from the nutrients it consumes) is determined by $0 \leq \sum_k f_{ik} \leq 1$ (we abbreviate $\sum_k f_{ik}$ as $f_{i\cdot}$). These f_{ik} parameters can change if a cell mutates, creating new lineages i of the same ancestral species that are identical in all other parameters. The remaining fraction of nutrients a cell takes up $1 - f_{i\cdot}$ is invested into population growth, creating a trade-off between growth and toxin degradation.

At each time-step, toxic compound degradation occurs first, followed by cell growth, then death. Cell growth consists of two steps, "activation" and "replication" (38). At a given time-point t, a culture contains a population of size p_{i0} of inactivated cells, and p_{i1} of activated cells of each lineage i ($S_i = p_{i0} + p_{i1}$). At $t = 0$, all cells are inactive, and every time a cell divides, its daughter cells are both inactive. Only degradation and activation require nutrient consumption (i.e. they are costly).

We first calculate how much nutrient will be consumed by each species for degradation and activation, which depends on the preferences n_{ij} of species i ($\sum_j n_{ij} = 1$), and the availability N_j of each nutrient j. If for at least one nutrient, $N_j < n_{ij}$, we set that $n_{ij} = 0$ for the rest of the round, compute the maximum amount of energy that the species can take up at that time-step $u_i = \sum_j n_{ij}$ and re-scale the preferences of that species to $\hat{n}_{ij}$:

$$\hat{n}_{ij} = \frac{n_{ij}}{\sum_j (n_{ij} \text{ if } N_j > n_{ij})}, \quad \textbf{(1)}$$

such that $\sum_j \hat{n}_{ij} = 1$. This implementation ensures that even if some nutrient types get depleted, cells consume the remaining required nutrients, even if less efficiently because $u_i < 1$. Next, we calculate the maximum number of cells of that species that can afford to consume nutrients, based on current availability: $S_i^{max} = \min_j (N_j / \hat{n}_{ij})$.

Toxic compound degradation occurs first. For each toxin k, S_i^{max} cells of each lineage i (or S_i if it is $< S_i^{max}$) degrade the amount $f_{ik} \cdot u_i$ units, consuming $\hat{n}_{ij} \cdot f_{i\cdot}$ units of nutrients. When a toxic compound is depleted, its degradation and the corresponding nutrient consumption do not occur.

The next step is cell activation. An inactive cell can activate with probability

$$a_i(1 - f_{i\cdot})u_i \sum_j \left(\hat{n}_{ij} \frac{N_j(t)}{N_j(t_0)} \right), \quad \textbf{(2)}$$

where a_i is the fixed activation probability for a given species. This expression is used to sample from S_i^{max} (or p_{i0} if it is $< S_i^{max}$), resulting in the number of cells that activate, thereby consuming $\hat{n}_{ij}(1 - f_{i\cdot})$ units of each nutrient, and joining the active population p_{i1}. All cells that activated in previous time-steps but did not replicate continue to consume the same amount of nutrients to remain activated.

Next, activated cells p_{i1} divide with probability $r_i(1 - f_{i\cdot})$ without additional nutrient consumption, each yielding two inactivated daughter cells: one daughter maintains the parameter values, and the other is susceptible to mutation with probability $\mu = 0.01$ to mutate each of its f_{ik} values. Upon mutation, the previous value of f_{ik} is multiplied by a random number from the lognormal (mean$= 0, \sigma^2 = 0.4$) distribution, making sure the total investment $f_{i\cdot}$ falls in the $[0,\ 1]$ interval. As a result, a new lineage of the same species with population size $p_{i0} = 1$ is introduced.

Activated and inactivated cells may die with a probability determined by the following Hill function:

$$\sum_k m_{ik} \frac{T_k^2}{T_k^2 + K^2} \quad \textbf{(3)}$$

Where T_k is the current concentration of toxic compound k and the constant K = 700.

Degradation score. Each round consists of $t_{end} = 80$ time-steps, in which the cells in each culture tube grow and degrade the toxic compounds independently of the others. At the end of the round, we calculate the degradation score D as the root-mean-square decrease in T_k from the initial time point t_0 to the last t_{end}:

$$D = 1 - \sqrt{\frac{1}{10} \sum_{k=1}^{10} \left(\frac{T_k(t_{end})}{T_k(t_0)} \right)^2}. \quad \textbf{(4)}$$

Parameter	Description	Randomly sampled from
l_i	ID of lineage i	
a_i	Activation probability	Beta(2,2)
r_i	Replication probability	Beta(2,2)
n_{ij}	Consumption rate of nutrient j	Uni(0,1), Sparse, Rescaled so that $\sum_j n_{ij} = 1$
m_{ik}	Death rate of strain i due to toxic compound k	Uni(0.001,0.02), Sparse
f_{ik}	Fraction of consumed nutrients invested into degradation of toxic compound k	Uni(0,1), Sparse, Rescaled so that $\sum_k f_{ik} = Uni(0,1)$

Table 1. Parameters defining a microbial strain i in the IBM. Growth rates, death rates and degradation investment vectors r_{ij}, m_{ik} and f_{ik} are made sparse by multiplying them by a vector drawn from Bernoulli(0.5). Each species can this way only take up a random fraction of nutrients, be affected by a random fraction of the toxic compounds and degrade another fraction of the toxic compounds. Despite changing by mutation, the total investment $f_{i\cdot}$ is limited to the interval $[0,\ 1]$.

Description	Value	Relevant propagation method
Mutation rate μ	0.01	all
Mutation magnitude	lognormal [0, 0.4]	all
Number of tubes	21	all
Selected parent tubes in each round	7	all except NS
Size of species pool	15	all
Number of nutrients	4	all
Initial nutrient concentrations $[N_j(t=0)]$	2000	all
Number of toxic compounds	10	all
Initial toxic compound concentrations $[T_k(t=0)]$	700	all
Toxic effect Hill function coefficient K (Eq. 3)	700	all
Initial number of cells (round 0) $[S_i = pi0 = 10]$	10	all
Initial number of cells (all rounds) $[S_i = pi0 = 10]$	10	DS, DR
Bottleneck size (number of cells transferred)	1/20 of prev. round	PS, PIS, PR, PIR
Bottleneck size (number of cells transferred)	1/21 of prev. round, pooled	MS, MIS, MR, MIR
Time steps t per round (t_{end})	80	all
Number of selection rounds	50	all
Number of replicate simulations per initial condition	10	all
Number of tubes to receive an invader	5	PIS, MIS, DS, DR
Number of tubes to lose a member	5	DS, DR

Table 2. Other parameters used in the paper, and relevance to the different propagation methods. DS = Disassembly Selection, DR = Disassembly Random, PS = Propagule Selection, PR = Propagule Random, PIS = Propagule Selection with Invasion, PIR = Propagule Random with Invasion, MS = Migrant pool Selection, MR = Migrant pool Random, MIS = Migrant pool Selection with Invasion, MIR = Migrant pool Random with Invasion, NS = No Selection. We explore variations to some of these in Fig. 6.

These scores are used to rank the communities, so that we can select the best ones and use them to start a new round. The "offspring" culture tubes in the new round again contain $T_k(t_0) = 700$ of each toxic compound, $N_j(t_0) = 2000$ of each nutrient and are seeded with cells from the selected tubes from the previous round – the "parent" communities – according to the propagation method (see below). We simulate a total 50 rounds of growth, degradation and selection.

Simulation replicates. The events occurring at each timestep are stochastic and lead to noise between runs of the model with the same starting conditions. The outcome also depends strongly on the parameters of the set of species that we sample from. We therefore generate 5 sets of 15 species each (we sample $5 \times 15 = 75$ times from Table 1). For each set, we then run 10 simulations, each starting with different combinations of the 15 species. The same exact initial conditions are used for all selection methods to make comparisons fair. When we calculate the growth and degradation of specific communities such as the 32767 possible combinations of species (Fig. 2E-F, Fig. 3G), we average the results over three replicates of the simulations. We use the same seed for the random number generators throughout all simulations for consistency.

Propagation methods. After scoring all communities in a given round (Eq. 4), we select 7 out of 21 parent communities that are propagated to create offspring for the next round of growth (Fig. 1A, Alg. 1). Here we implement several methods of propagating the selected communities, and compare them to one another (Fig. 1B). The choice of bottleneck size (7/21) is relatively wide, as compared to similar studies in the literature (8, 9, 11–16, 18–21, 24, 25, 39–41).

No Selection (NS). We implement the no-selection treatment as a control or a baseline, to evaluate how the community function (toxic compound degradation) changes simply due to interspecies interactions and the resulting ecological and evolutionary dynamics (25). Note that all interspecies interactions are due to the consumption of nutrients and degradation of toxic compounds. NS is the only treatment where we do not select 7 of the parent communities. Instead, each community is propagated by approximately 20-fold dilution at the end of each round. Dilution occurs by stochastically sampling cells from the parent population with all cell types combined according to a Poisson distribution and placing them in the offspring tubes. The dilution factor is approximate as we have a discrete number of cells in each population, which is not always divisible by 20.

Propagule selection (PS). In propagule selection, the 7 communities with the highest scores are propagated to the next round by dilution (12, 28) (Fig. 1B, Alg. 7, Alg. 8). Each selected parent contributes equally to 3 offspring communities. We then seed each offspring by diluting the corresponding parent community by approximately 20, as in NS. This means that following the first round (where we inoculate 10 cells of each of 4 species), every tube may contain a different number of species, each with a different number of cells, and a different total population size. We compare PS to the random control PR, where 7 communities are selected at random, ignoring their degradation scores. We also compare PS to a version that we call propagule with invasion (PIS) and its corresponding random control (PIR). In this version, we introduce at least one species (chosen at random with uniform probability out the set of 15), to 5 out of the 21 offspring communities (chosen at random with uniform probability) (25).

Migrant pool selection (MS). In MS, the 7 selected parent communities are mixed in a *migrant pool* before new offspring communities are formed by sampling from the pool (12, 28) (Fig. 1B, Alg. 9). To create the 21 offspring communities, cells are sampled without replacement from the pool with an approximately 20-fold dilution (stochastic process according to Poisson distribution). In MS, as with PS, the species composition and cell number can vary compared to the initial 10 cells of 4 species in each tube of the first round. In contrast to PS, though, all tubes in following rounds are expected to have similar total population sizes, as they are all seeded from the same pool.

We compare MS to a random control (MR), where we select the 7 communities with uniform probability ignoring their degradation scores. We also implement a version of migrant pool selection where we introduce one or more species (chosen at random with uniform probability out the set of 15) to 5 out of the 21 offspring communities (chosen at random with uniform probability), and call this migrant pool with invasion (MIS and the random control MIR).

Disassembly selection (DS). We have designed this selection method to overcome some of the short-comings of the other selection methods (red box in Fig. 1B, Alg. 10). Rather than diluting each parent to create an offspring, DS disassembles each parent community into its member species (including each species' different lineages) and uses these isolated species to recompose communities that resemble the parents in species composition, with some variability introduced by removing species that were present in the parent or introducing new member species from the species pool. By disas-

sembling communities, we maintain a record of samples of each species and their evolved lineages that were present in at least one selected community in each round (box outlined in grey in Fig. 1B, DS). If a species is present in more than one selected community, we sample from the highest-scoring community that this species was part of.

Another difference with DS is that we modify the scoring method by penalizing communities in which extinctions occur. The rationale is to favor communities whose member species do not out-compete one another, which we expect to lead to more cooperative communities, potentially with better collective functions. We implement this by scaling the degradation score D by the fraction of surviving species at the end of a round of growth as follows:

$$\hat{D} = D \times \frac{\text{number of surviving species}}{\text{number of species in the community}}. \quad \textbf{(5)}$$

For example, if a 5-species community loses one member species, its degradation score is scaled by 0.8. Next, we draw 21 offspring communities from the 7 ($n = 1, \ldots, 7$) selected parent communities for the next round of growth, in proportion to their scores $\hat{D}$ with probability:

$$\frac{1}{\sum_k \hat{D}_k} \hat{D}_n, \quad \textbf{(6)}$$

with replacement.

We then modify the species composition of the 21 new offspring communities to introduce variability between communities. First, we choose 5 of the 21 offspring communities at random and remove one or more species, always one species plus an additional number drawn from Poisson(0.5). This a choice which means that we are more likely to remove few species but can also remove more from time to time. If the drawn number is equal to or higher than the number of species currently in the community, we leave one species to avoid emptying or completely changing the community composition. Having found a number of species to remove, we choose the species to remove with uniform probability, but avoid removing any species that is present in only this community. Next, we randomly chose 5 communities to which we introduce one or more invader species —as above, $1+$Poisson(0.5)— chosen with uniform probability from the species pool. These 5 are chosen anew and could be the same communities that we just removed species from, or not. In order to maintain diversity, we ensure that all species appear in at least one community by preferentially introducing species that are not currently present in any offspring community. If a species goes extinct in all 21 tubes, we re-introduce it from the most recent round in which it was present. See Alg. 10 for more details.

Once the final species composition of each offspring community has been determined, 10 cells of each species from the species record (box outlined in grey in Fig. 1B, DS) (sampled randomly with replacement from a Poisson distribution) are inoculated into the offspring tubes. The initial population size in all tubes is therefore always 10 times the number of species, which begins with 4 in round 0, but can vary over subsequent rounds. Keeping the initial population size of each species fixed allows to standardize the growth conditions between rounds, i.e. to maintain heritability (9).

Statistical and other analyses. Correlations are evaluated by the Spearman's rank correlation coefficient ρ. We compare selected communities to the set of all possible communities by a Kruskal–Wallis H test for differences in median. We use the scipy (42) implementations for both.

We quantify species diversity within a community (Fig. 3C, Fig. S5) as the Hill number of order 1 or the average effective number of species present in the community (43, 44), which is based on the Shannon index H':

$$\exp(H') = \exp(-\sum_{i=1} p_i \log p_i), \quad \textbf{(7)}$$

which in turn depends on the species' relative abundances

$$p_i = S_i / S_{tot} \quad \textbf{(8)}$$

where we divide the population size S_i of each species i by the total population size in the community $S_{tot} = \sum_i S_i$. If more than one lineage is present, we sum up their population sizes to find the species' total population size. We then average $\exp(H')$ over all communities to find the average effective number of species. The measure falls between 0 and 15 effective species in an average community.

Beta diversity (Fig. 4B, Fig. S7) is calculated by considering each community as a vector of the population sizes of the 15 species. Species absence from the community is marked by zero. We find the beta diversity as the average Bray-Curtis dissimilarity of each of the 210 possible pairs in the 21 communities.

The community coverage of nutrients and toxic compounds (Fig. 3D, E) is quantified similarly to species diversity, as the effective number of toxic compounds invested into or nutrients taken up. Toxic compound coverage is calculated from the vector $\tilde{f}_{\cdot k} = \sum_i \tilde{f}_{ik}$ of total investment into degrading toxic compound k in a given community. The 'tilde' indicates that we have scaled the f_{ik} for each lineage of a species by the corresponding population sizes of the lineage, as in Eq. (10). In this way, we emphasize the most relevant lineage of each species and do not bias the result to the number

of competing lineages within a species. Note that we do not scale $\tilde{f}_{\cdot k}$ by species abundance in the community. Once each investment $\tilde{f}_{\cdot k}$ is rescaled so that $\sum_k \tilde{f}_{\cdot k} = 1$, we calculate the effective number of toxic compounds invested into as

$$\exp(H') = \exp\left(-\sum_{k=1} \tilde{f}_{\cdot k} \log \tilde{f}_{\cdot k}\right), \quad \textbf{(9)}$$

and average this value over all 21 communities. The measure falls between 0 (no toxic compounds are invested into) and 10 (all compounds). For the nutrient coverage, we use the same calculation using the nutrient uptake rates n_{ij}, but without scaling between different lineages as this parameter does not mutate. The effective number of nutrients taken up takes values between 0 and 4, the number of different nutrients.

To evaluate the evolution of the total investment $f_{i\cdot} = \sum_k f_{ik}$ (Fig. 5A) of species i, we compare the investment of the ancestral lineage to that in the last round of selection (or as late as possible in case the species went extinct). The average per-species investment is weighted by the population size S_i of the different lineages l of species i,

$$\tilde{f}_{i\cdot} = \sum_{l \text{ of species } i} S_l f_{l\cdot} / \sum_l S_l \quad \textbf{(10)}$$

to emphasize the investment of abundant lineages instead of small recent mutants that have not contributed as much to the community function. The value $\tilde{f}_{i\cdot}$ is further averaged over the 10 repeated runs (recall: each run starts from the same set of species) for each of the 15 species, without weighting by population size. For the statistical comparison, we hence have 75 data points: 15 species per each of the 5 sets.

To evaluate the stability of the selection methods (Fig. 6A), we choose the highest-scoring community that each method found after 50 rounds of selection (one community for each repeated run of each species set), and seed 10 replicates with its identical initial composition. Then we grow and dilute them for a further 25 rounds, as we would do for the no-selection treatment. We do not allow mutations in these rounds, to focus only on the ecological stability of the found communities.

For the analysis of sensitivity to the number of species in the initial species pool (Fig. 6C, Fig. S14), we sample 5 random subsets of 6, 9, 12 from the original set of 15 species, and run for each of them 5 simulations with different 21 initial communities. Drawing new species sets of the corresponding size would introduce further variance, which we would rather avoid. For the effect of the dilution factor (Fig. 6E, Fig. S16), we multiply the 10-cell inoculum by a factor 0.2, 0.5, 1, 2 or 5 for the disassembly method and scale the 5% dilution fraction for the other methods by the same factor.

Contributions and acknowledgements

BV, FAS and SM conceived the project. PGF and BV developed, implemented and simulated the models and selection. PGF and BV analyzed the data. BV wrote the first draft. BV, PGF and SM wrote the manuscript. FAS commented on the manuscript. BV, FAS, PGF and SM were funded by European Research Council Starting Grant 715097, and SM by the NCCR Microbiomes (Swiss National Science Foundation). We thank the Mitri lab at the University of Lausanne for valuable discussions, in particular Afra Salazar de Dios and Shota Shibasaki for detailed comments on the manuscript.

References

1. Frances H Arnold. Protein engineering for unusual environments. *Proc Natl Acad Sci USA*, 4:450–455, 1993.
2. Flor I. Arias-Sánchez, Björn Vessman, and Sara Mitri. Artificially selecting microbial communities: If we can breed dogs, why not microbiomes? *PLoS Biology*, 17(8):1–8, 2019.
3. Álvaro Sánchez, Jean C.C. Vila, Chang-Yu Chang, Juan Diaz-Colunga, Sylvie Estrela, and María Rebolleda-Gomez. Directed evolution of microbial communities. *Annual Review of Biophysics*, 50(1):null, 2021. PMID: 33646814.
4. Shi Rui Yu, Yuan Ye Zhang, and Quan Guo Zhang. The effectiveness of artificial microbial community selection: a conceptual framework and a meta-analysis. *Frontiers in Microbiology*, 14:1257935, 2023.
5. Magdalena San Román, Andrea Arrabal, Belen Benitez-Dominguez, Isabel Quirós-Rodríguez, and Juan Diaz-Colunga. Towards synthetic ecology: strategies for the optimization of microbial community functions. *Frontiers in Synthetic Biology*, 3:1532846, 2025.
6. Juhee Lee, Wenying Shou, and Hye Jin Park. Rafting a waterfall: Artificial selection for collective composition can succeed or fail depending on the initial and target values.
7. C J Goodnight. Heritability at the ecosystem level. *Proceedings of the National Academy of Sciences of the United States of America*, 97(17):9365–6, 2000.
8. Hywel T. P. Williams and Timothy M. Lenton. Artificial ecosystem selection for evolutionary optimisation. In Fernando Almeida e Costa, Luis Mateus Rocha, Ernesto Costa, Inman Harvey, and António Coutinho, editors, *Advances in Artificial Life*, pages 93–102, Berlin, Heidelberg, 2007. .
9. Li Xie, Alex E. Yuan, and Wenying Shou. Simulations reveal challenges to artificial community selection and possible strategies for success. *PLOS Biology*, 17(6):e3000295, 2019.
10. Melanie Ghoul, Ashleigh S. Griffin, and Stuart A. West. Toward an Evolutionary Definition of Cheating. *Evolution*, 68(2):318–331, 2014.
11. W. Swenson, D. S. Wilson, and R. Elias. Artificial ecosystem selection. *Proceedings of the National Academy of Sciences*, 97(16):9110–9114, 2000.
12. Tiffany Raynaud, Marion Devers, Aymé Spor, and Manuel Blouin. Effect of the Reproduction Method in an Artificial Selection Experiment at the Community Level. *Frontiers in Ecology and Evolution*, 7:416, 2019.
13. Samuel Jacquiod, Aymé Spor, Shaodong Wei, Victoria Munkager, David Bru, Søren J. Sørensen, Christophe Salon, Laurent Philippot, and Manuel Blouin. Artificial selection of stable rhizosphere microbiota leads to heritable plant phenotype changes. *Ecology Letters*, 25(1):189–201, 2022.
14. Michael D. Jochum, Kelsey L. McWilliams, Elizabeth A. Pierson, and Young Ki Jo. Host-mediated microbiome engineering (HMME) of drought tolerance in the wheat rhizosphere. *PLoS ONE*, 14(12):1–15, 2019.
15. Ulrich G Mueller, Thomas E Juenger, Melissa R Kardish, Alexis L Carlson, Kathleen M Burns, Joseph A Edwards, Chad C Smith, Chi-chun Fang, and L Des Marais. Artificial Selection on Microbiomes To Breed Microbiomes That Confer Salt Tolerance to Plants. *mSystems*, 2021.
16. Kevin Panke-Buisse, Angela C Poole, Julia K Goodrich, Ruth E Ley, and Jenny Kao-Kniffin. Selection on soil microbiomes reveals reproducible impacts on plant function. *The ISME Journal*, 9(4):980–989, 2015.
17. Kevin Panke-Buisse, Stacey Lee, and Jenny Kao-Kniffin. Cultivated Sub-Populations of Soil Microbiomes Retain Early Flowering Plant Trait. *Microbial Ecology*, 73(2):394–403, 2017.

18. Robyn J Wright, Matthew I Gibson, and Joseph A Christie-Oleza. Understanding microbial community dynamics to improve optimal microbiome selection. *Microbiome*, 7(1):85, 2019.
19. Jigyasa Arora, Margaret Mars Brisbin, and Alexander S. Mikheyev. Effects of microbial evolution dominate those of experimental host-mediated indirect selection. *PeerJ*, (8):e9350, 2020.
20. Manuel Blouin, Battle Karimi, Jérôme Mathieu, and Thomas Z. Lerch. Levels and limits in artificial selection of communities. *Ecology Letters*, 18(10):1040–1048, 2015.
21. Chang-Yu Chang, Melisa L. Osborne, Djordje Bajic, and Alvaro Sanchez. Artificially selecting bacterial communities using propagule strategies†. *Evolution*, 74(10):2392–2403, 2020.
22. U.G. Mueller and J.L. Sachs. Engineering Microbiomes to Improve Plant and Animal Health. *Trends in Microbiology*, 23(10):606–617, 2015.
23. Li Xie and Wenying Shou. Steering ecological-evolutionary dynamics to improve artificial selection of microbial communities. *Nature Communications*, (12):264697, 2021.
24. Guilhem Doulcier, Amaury Lambert, Silvia De Monte, and Paul B Rainey. Eco-evolutionary dynamics of nested darwinian populations and the emergence of community-level heredity. *eLife*, 9:e53433, 2020.
25. Chang-Yu Chang, Jean C. C. Vila, Madeline Bender, Richard Li, Madeleine C. Mankowski, Molly Bassette, Julia Borden, Stefan Golfier, Paul Gerald L. Sanchez, Rachel Waymack, Xinwen Zhu, Juan Diaz-Colunga, Sylvie Estrela, Maria Rebolleda-Gomez, and Alvaro Sanchez. Engineering complex communities by directed evolution. *Nature Ecology Evolution 2021 5:7*, 5(7):1011–1023, 2021.
26. Alexander Lalejini, Emily Dolson, Anya E Vostinar, and Luis Zaman. Artificial selection methods from evolutionary computing show promise for directed evolution of microbes. *eLife*, 11, 2022.
27. Ashish B. George and Kirill S. Korolev. Ecological landscapes guide the assembly of optimal microbial communities. *PLOS Computational Biology*, 19:e1010570, 2023.
28. M. Slatkin and M. J. Wade. Group selection on a quantitative character. *Proceedings of the National Academy of Sciences of the United States of America*, 75(7):3531–3534, 1978.
29. Philippe Piccardi, Björn Vessman, and Sara Mitri. Toxicity drives facilitation between 4 bacterial species. *Proceedings of the National Academy of Sciences*, 116(32):15979–15984, 2019.
30. Flor I. Arias-Sánchez, Björn Vessman, Alice Haym, Géraldine Alberti, and Sara Mitri. Artificial selection improves pollutant degradation by bacterial communities. *Nature Communications*, 15:7836, 2024.
31. Michael J Wade. An Experimental Study of Group Selection. *Evolution*, 31(1):134–153, 1977.
32. Melanie Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, USA, .
33. John H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, USA, .
34. Sarah P. Otto and Thomas Lenormand. Resolving the paradox of sex and recombination. *Nature Reviews Genetics 2002 3:4*, 3(4):252–261, 2002.
35. J. Arjan G.M. De Visser and Santiago F. Elena. The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews Genetics 2007 8:2*, 8(2):139–149, 2007.
36. R. C. Lewontin. The Units of Selection. *Annual Review of Ecology and Systematics*, 1(1): 1–18, 1970.
37. Afra Salazar and Sara Mitri. Can a microbial community become an evolutionary individual? *Current Opinion in Microbiology*, 84:102596, 2025.
38. Jose Alvarez-Ramirez, M. Meraz, and E. Jaime Vernon-Carter. A theoretical derivation of the monod equation with a kinetics sense. *Biochemical Engineering Journal*, 150:107305, 2019.
39. William Swenson, Jeff Arendt, and David Sloan Wilson. Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. *Environmental Microbiology*, 2(5):564–571, 2000.
40. Alexandra S Penn and Inman Harvey. The role of non-genetic change in the heritability, variation and response to selection of artificially selected ecosystems. 2004.
41. Tiffany Raynaud, Marion Devers-Lamrani, Aymé Spor, and Manuel Blouin. Community diversity determines the evolution of synthetic bacterial communities under artificial selection. *Evolution*, 76(8):1883–1895, 2022.
42. Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python. [Online; accessed 2019-10-21].
43. M. O. Hill. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, 54(2):427–432, 1973.
44. Anne Chao, Chun-Huo Chiu, Lou Jost, A Chao, C.-H Chiu, and L Jost. Phylogenetic Diversity Measures and Their Decomposition: A Framework Based on Hill Numbers. *Topics in Biodiversity and Conservation*, 14:141–172, 2016.
45. Ernst Hairer, Syvert P. Norsett, and Gerhard Wanner. *Solving Ordinary Differential Equations i. Nonstiff Problems.*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, 2 edition, .