Novel artificial selection method improves function of simulated microbial communities

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

[1]Département de Microbiologie Fondamentale, Université de Lausanne, 1015 Lausanne, Switzerland
[2]BIH Center for Regenerative Therapies (BCRT), Charité - Universitätsmedizin Berlin, Berlin, Germany
[3]Swiss Institute of Bioinformatics
[*]These authors contributed equally

There is increasing interest in artificially selecting or breeding microbial communities, but experiments have reported modest success and it remains unclear how to best design such a selection experiment. Here, we develop computational models to simulate two previously known selection methods and compare them to a new "disassembly" method that we have developed. Our method relies on repeatedly competing different communities of known species combinations against one another, and sometimes changing the species combinations. Our approach significantly outperformed previous methods that could not maintain enough between-community diversity for selection to act on. Instead, the disassembly method allowed many species combinations to be explored throughout a single selection experiment. Nevertheless, selection at the community level in our simulations did not counteract selection at the individual level. Species in our model can mutate, and we found that they evolved to invest less into community function and more into growth. Increased growth compensated for reduced investment, however, and overall community performance was barely affected by within-species evolution. Our work provides important insights that will help design community selection experiments.

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, 3), 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 (4). The problem arises when a single community-level "generation", which we will call a "round of selection" to avoid confusion, can comprise several generations of cells, each belonging to different genotypes (i.e. species and strains), (Fig. 1A). 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 genetically distinct constituent members and their proportions, an "offspring" community may not resemble its "parent" (4–6). Another issue with group-level 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, 7). 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 (4). 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.

In the earliest community breeding experiments, Swenson *et al.* selected microbial communities to yield plants with high and low biomass and to control pH (8). In two out of three experiments, the communities selected for high vs. low function differed significantly from each other, but were not significantly different from the starting communities. The results were also noisy and inconsistent across experimental systems (8, 9). Many attempts have been made since, aiming to optimize several microbial community traits, including increased microbial biomass production (10), the stimulation of various plant properties (10–15), chitin degradation (16), the stimulation of fruit fly development (17), to reduce wastewater CO_2 emissions (18), and to hydrolyze starch (19). 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 (8, 16, 17). Overall, community breeding experiments have shown mixed success (3, 20), but computer simulations have provided some clues on how to improve them (5, 6, 21–23).

All 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 (Fig. 1B), 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 (Fig. 1C). While both selection methods have achieved some success, they suffer from a rapid decrease in between-community variability (24), such that selection has little to act on. Intuitively, the loss of variability arises firstly because only a fraction of community members are selected and replicated for the next round. Second, species composition can only change through loss of members when the communities are diluted, meaning that the communities evaluated throughout the whole experiment can only be sub-communities of the initial ones. Given that finding the right species composition is one of the goals of community-level selection, this suggests that we need novel selection methods that can better explore the search space of species combinations (23).

In this manuscript, we propose a new selection method that we call "disassembly selection" (DS), that is designed to maintain heritability as well as 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 (Fig. 1D). We construct two computational models of microbes in a well-mixed liquid culture, one individual-based and one based on differential equations, to systematically compare our new approach to the classical propagule selection (PS) and migrant pool selection (MS) methods.

Inspired by a four-species community that degrades an industrial pollutant (25), we aim to select for microbial communities with improved degradation capabilities. Based on this experimental system, the microbes in our models face a dilemma: whether to invest consumed nutrients into growth or into degradation of toxic compounds that would otherwise cause cell death. The populations evolve by random mutations to this relative investment. We evaluate the selection methods by comparing how the degradation scores change over several rounds of growth and selection starting from the same initial communities. We simulate community selection in both models separately, to test whether our results depend on the choice of model framework.

Our results confirm our intuition that propagule and migrant pool selection do not maintain enough variability to explore many different species combinations, which means that the communities can only improve by mutation. In contrast, our new 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 in group selection: competition within species leads to the dominance of strains that invest less into the function and more into growth. Our work thereby suggests a new method to find species combinations whose community function is high, but in which between-individual competition may be inevitable.

Results

Simulating community-level selection. In either model (see Methods for details), each species is described by its growth and uptake rates for each of 4 available nutrients, and its death and degradation rates for each of 10 toxic compounds. We assume that interactions between cells occur only via nutrients and toxic compounds, as cells of type i invest a fraction f_{ik} into degradation of the toxic compounds and the rest into growth. Cells of the same species differentiate by accumulating "mutations" as they grow and divide, that alter the total investment $f_{i\cdot} = \sum_k f_{ik}$. All other species properties remain unchanged throughout the simulations.

The simulations start with 21 communities of 4 species each, chosen at random with replacement from a set of 15 initial species, that are described by randomly drawn model parameters. The 21 communities are grown in simulated batch cultures containing defined initial concentrations of nutrients and toxic compounds for a fixed number of time-steps (Fig. 1F). At the end of each round, the 21 communities are 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: communities are diluted in propagule and migrant pool, whereas they are re-inoculated to a defined population size with equal proportions in the disassembly method (Fig. 1B-D). In disassembly, communities are penalized by species extinctions, and

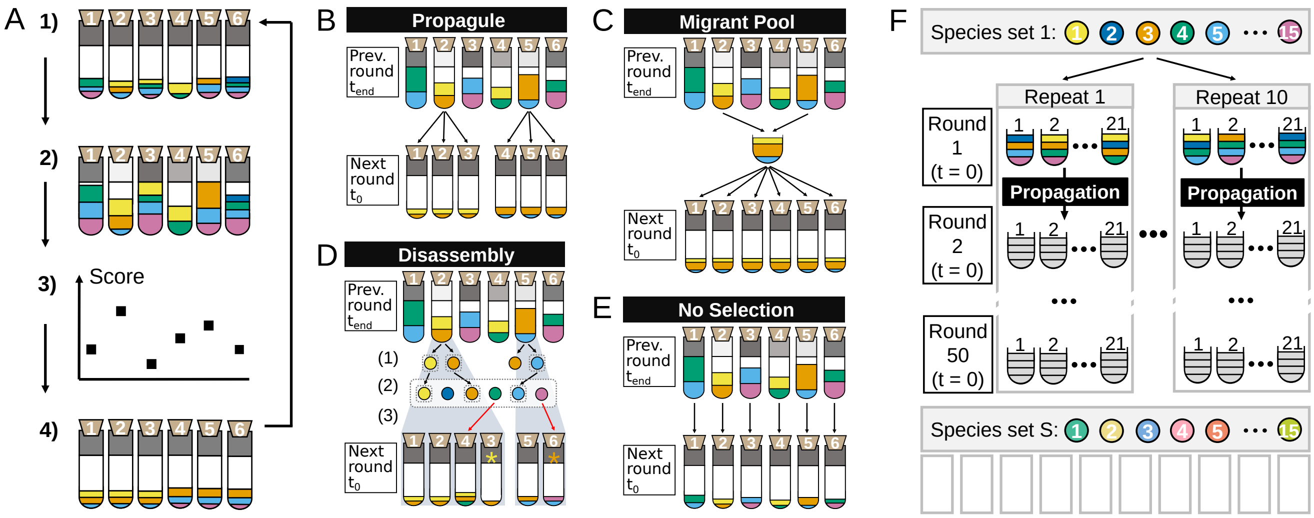

Fig. 1. (A) Overall method for artificial selection of microbial communities. Communities are illustrated as test tubes with bacterial "species" in different colors (white represents empty "space"). The concentration of toxic compounds is shown in shades of gray in the upper part of each tube (darker is more toxic). The inoculated communities (1) grow until the measurement (2) of toxic compound concentration, from which we (3) calculate a score for each community. (4) The highest-scoring communities are selected for propagation into offspring communities and the process is repeated. **(B)** Propagule: each selected community from the previous round is diluted to form the same number of communities for the next round. **(C)** Migrant pool: selected communities are merged before dilution. **(D)** Disassembly: Microbes are (1) isolated from the chosen communities and (2) saved in a repository (dotted rectangle). Each selected community contributes offspring communities in proportion to their degradation score (3). A fraction of the new communities receive new species (red arrows) or lose members from the previous round (asterisk in color of removed species). **(E)** No-selection control: each community is diluted into a new tube. Propagule, migrant pool and disassembly have selection treatments (PS, MS and DS) and random treatments (PR, MR and DR), where community scores are ignored (see Methods). **(F)** A "species set" consists of 15 randomly generated species. From this set, we draw 21 initial communities of 4 randomly chosen species each and for each of five species sets, simulate 10 repeats from different initial communities over 50 rounds of selection under each of the propagation methods (B-E).

communities are randomly chosen to receive or lose species (Fig. 1D). We compare each method to a corresponding random control line (e.g. random propagule: PR) where 7 communities are chosen at random instead of according to their score, and to a no-selection control (NS) where every community is diluted without selection (Fig. 1E). This last control forms a baseline for how communities change due to species interactions (23, 26). To achieve statistical power, 5 species sets were generated, each with a new set of 15 species. From each species set, we then sampled the 21 communities 10 times to run 10 replicate simulations, which were all subjected to 50 rounds of selection. The same initial conditions were used for the different selection methods to allow for a fair comparison (Fig. 1F).

Disassembly finds communities whose degradation ranks in the top percentile of all possible communities. All simulated selection methods succeeded in improving the median degradation score across the 21 communities between round 0 and 50 (Fig. S1), which is consistent with previous work (3, 23). 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, $p < 10^{-9}$ for both IBM and ODE, Fig. 2A, C). The increase in maximum score in DS (0.22 ± 0.06, 0.14 ± 0.08 for IBM, ODE), was also significantly different from the classical selection methods (-0.03 ± 0.06 and -0.12 ± 0.09 for PS and MS in the IBM, and -0.1 ± 0.08 for PS in the ODE), from its own random control (DR), and from NS (all two-sided Wilcoxon tests of diff. in max. degradation between DS and other methods, $n = 50$, $p < 10^{-9}$, for IBM and ODE). For comparison, 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 in both our models, finding the very best community out of 32767 (Fig. 2D, F) in 17 out of 50 runs in the IBM and 23 out of 50 in the ODE. We next investigate what distinguishes these high-ranking communities.

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 dif-

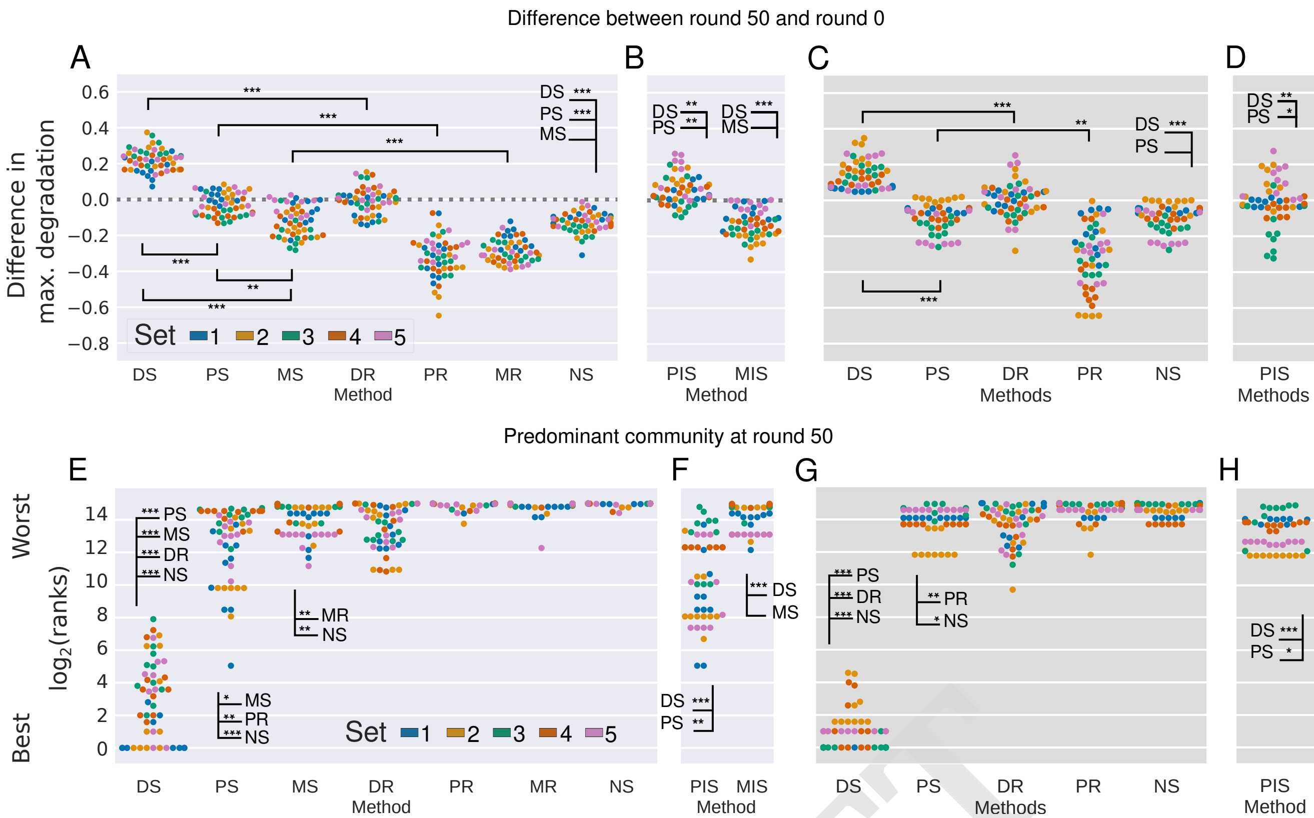

Fig. 2. Degradation scores and ranking of selected communities. Panels A, B, E, F with lighter background show results from the IBM, while panels C, D, G, H with darker background show results from the ODE model. Asterisks show the significance of a Wilcoxon signed-rank test for difference in degradation between methods (*: $p < 10^{-3}$, **: $p < 10^{-6}$, ***: $p < 10^{-9}$). **(A-D)** The difference in maximum degradation score between round 50 and round 0 over the 21 communities is shown as one dot for each 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. **(E-H)** 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. As above, each of the 50 dots marks 1 out of 10 repeated runs of 1 out of 5 sets of species.

ferent nutrients while the degradation score of a community can increase if its members specialize on degrading different toxic compounds (Eq. 1).

To understand how these two properties changed over time, we first quantified the "total investment", i.e. the fraction $\sum_{k=1}^{10} f_{ik} < 1$ of resources invested into degradation of all toxic compounds k, 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 of average total investment, all $p < 10^{-9}$, $n = 50$ for both IBM and ODE, Fig. 3A, C). 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. 3E), 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. 3F). This increase in coverage and community diversity was not observed for the other selection methods (Fig. 3E, Fig. S2).

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 degrade 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. 3G). 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, $p < 10^{-9}$ in either case, Fig. 3G).

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, B, Fig. S3). 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. 4C, D, Fig. S4). 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 resulting drop in beta diversity is not as steep. The beta diversity of the no-selection control depends on the diversity of the initial communities.

Propagule selection —but not migrant pool— performs better by periodically adding species to selected 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 (23)), 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. With this modification, PIS increases the maximum degradation (one-sided Wilcoxon signed-rank test of degradation scores 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 the IBM. The results are however model-dependent. While the PIS method still improved upon the PS method in the ODE model (two-sided Wilcoxon signed-rank test, $p < 10^{-3}$, $n = 50$), we did not find any significant improvements in the maximum degradation score compared to round 0 ($p = 0.9$, $n = 50$, Fig. 2D). Further, PIS finds higher-ranking communities than PS in both the IBM (two-sided Wilcoxon signed-rank test for differences in ranks between PIS and PS, $p < 10^{-6}$, $n = 50$, Fig. 2F) and the ODE model ($p < 10^{-3}$, $n = 50$, Fig. 2H) over the 50 rounds. PIS can explore more combinations than the regular PS, and the initial drop in beta diversity is less severe in both models (Fig. 4A-D), 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. 4C), 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 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_{ik} 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 now 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 $\sum_{k=1}^{10} f_{ik}$ into degradation was significantly lower after 50 rounds of selection than that of the corresponding ancestral species (one-sided Wilcoxon signed-rank test of total investment in initial vs final round of selection, $p < 10^{-6}$, $n = 50$, Fig. 5A, Fig. S5). Given 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 communities, initial vs final round, $p < 10^{-9}$, $n = 50$, Fig. 5B), such that overall, the degradation of the evolved communities was marginally but significantly higher (+6×10^{-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. S6). 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.

In summary, the disassembly method improved the degradation scores over the 50 rounds of selection by finding better

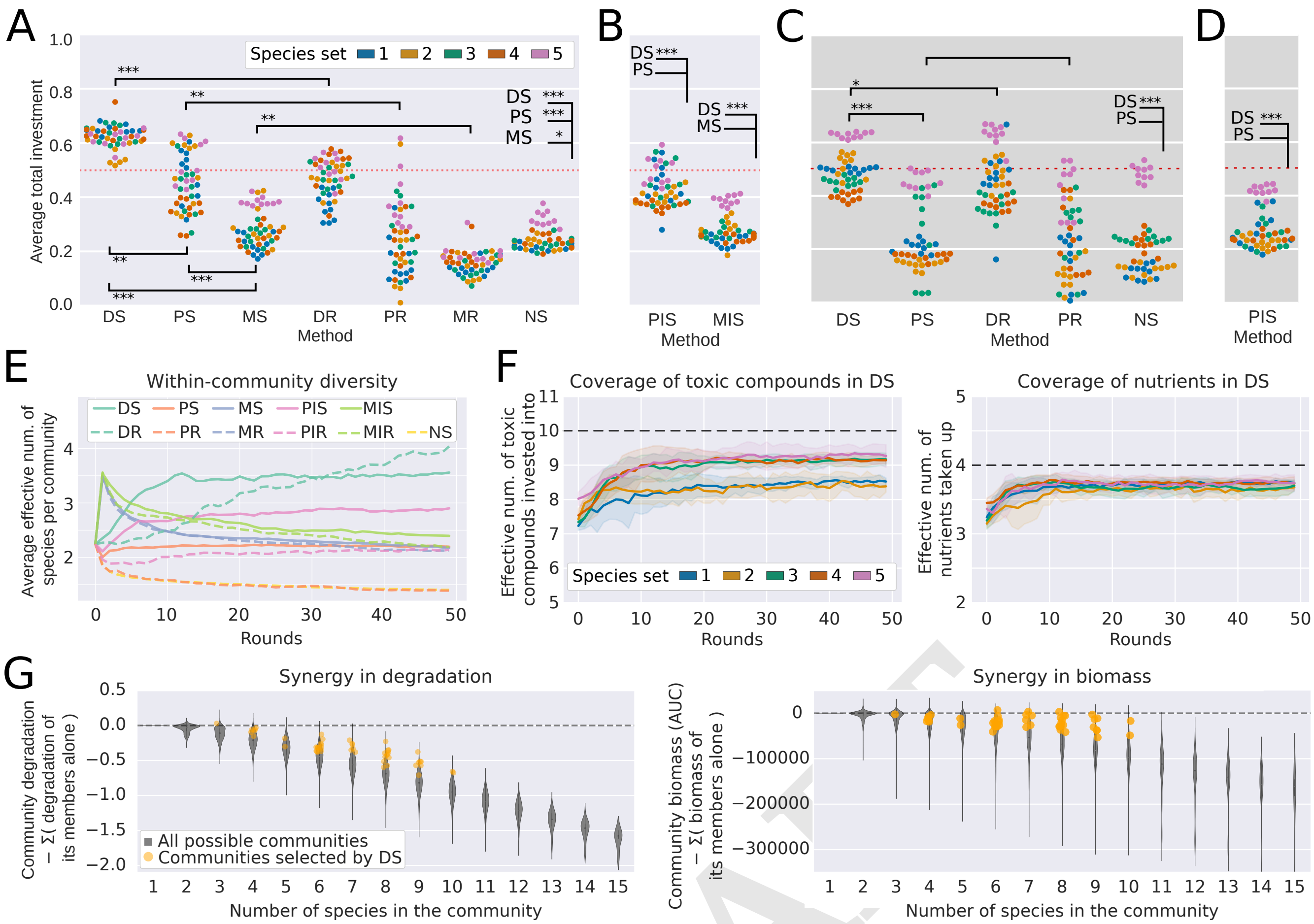

Fig. 3. Total investment into degradation, diversity, coverage and synergy in degradation. Data from the IBM shown on lighter background everywhere except for panels CD, which has data from the ODE model on darker background. **(A-D)** Average total investment in a community at round 50, averaged for 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. IBM results in panels A, B and from the ODE in C and D. One dot for each out 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 $n = 50$ (*: $p < 10^{-3}$, **: $p < 10^{-6}$, ***: $p < 10^{-9}$). **(E)** Within-community species diversity measured as the effective number of species Eq. (13), 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. S2. **(F)** The coverage of toxic compounds and 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). Results from the IBM. **(G)** Synergy in communities selected by DS at round 50, grouped by community richness. Synergy is the difference in degradation scores (left panel) or biomass (right) 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. 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.

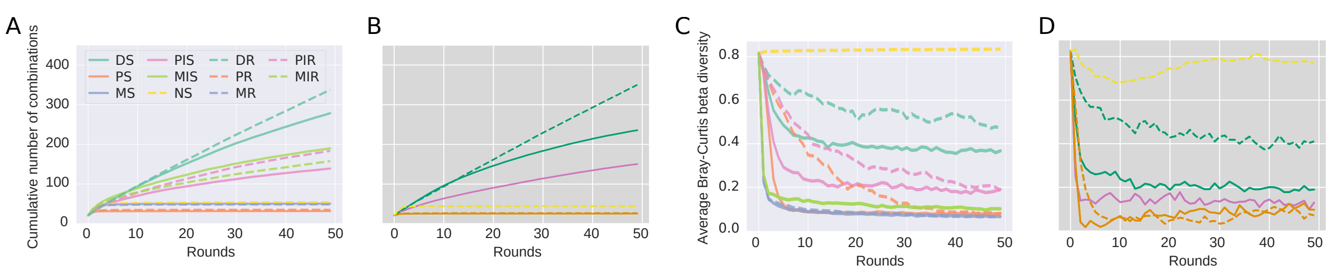

Fig. 4. Cumulative number of communities found by the selection methods and between-community diversity explain how DS can find better communities. We show the mean over the 10 repeats and 5 species sets for each propagation method and refer to Fig. S3-S4 for the full results. Results for the IBM and ODE model are shown against a light or a dark background, respectively. **(A, B)** Time series of cumulative number of unique communities for each selection method. **(C, D)** Between-community or beta diversity, calculated as the average Bray–Curtis distance of each pair of communities.

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.

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 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. This means that the communities converge, once ecologically stable, 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. S7). 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 the community selection (23).

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 could cost a lot of time and resources. We now investigate how four experimental parameters impact the degradation scores in DS, and affect experiment size. We focus on DS but also compare it to the other methods (Fig. 6B-E, Fig. S8-S11).
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. S9). This means that the metacommunity 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. Next, we turn to two parameters that affect the degradation scores but not size of the experiment. 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. S8). 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. While other theoretical studies have investigated heritability and the balance between within- and between-community selection (6, 21–23), our "disassembly" method contributes to improving the third point: how to maintain variability between communities.
We have shown that disassembly can improve 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

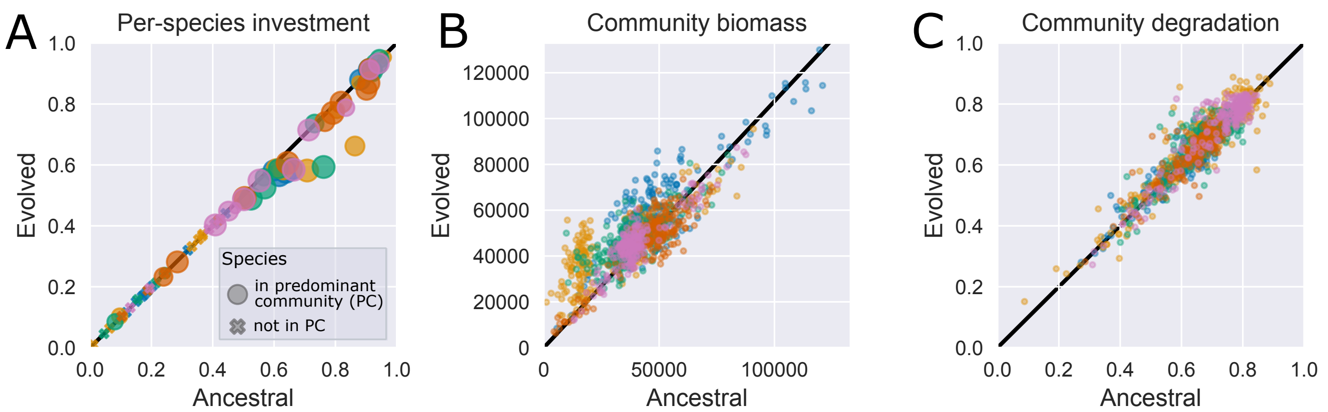

Fig. 5. Change in degradation investment per species and the effect of such change in the biomass and degradation of the communities in the last round, colored by species sets. All results from the IBM. **(A)** Total investment $f_i = \sum_k f_{ik}$ for each species $i = 1, 2, \ldots, 15$. There are 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. S5. **(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.

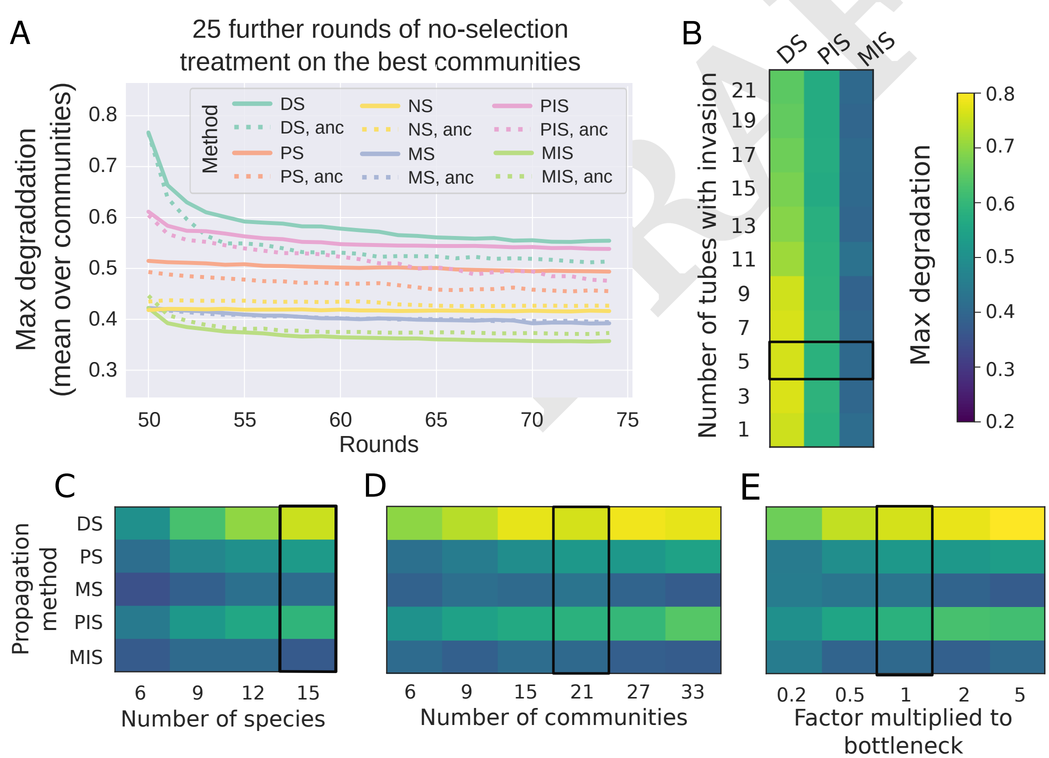

Fig. 6. **(A)** Stability in degradation over 25 additional rounds after releasing the selection pressure, calculated for the highest-scoring community for each selection method in the IBM. Results show the average degradation for each repeat. The dotted lines show the corresponding degradation when 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. S9-S11 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.

improve the maximum function for some initial combination of species, confirming previous findings (3, 23).

The problem with the classical methods is that they rapidly lose between-community diversity, which selection acts upon to improve community function. In these methods, diversity only arises through mutations, or through loss of species due to competition and dilutions between rounds of growth, while directed selection reduces between-community variability by only propagating a small subset of high-performing communities. When communities become increasingly similar, fitness differences become increasingly random which makes selection less effective (23). By removing and introducing species, the disassembly method reshuffles the species composition to access new communities that sometimes outperform the original best community. As proposed in (23), we show that the classical methods can be improved by periodically invading them with new species, which allows them to maintain some variability. This approach still under-performs compared to disassembly, however.

In a sense, the disassembly method is inspired by the crossover operator used in genetic algorithms in the field of evolutionary computing (27–29), 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 (30, 31). 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.

Next, heritability is crucial for evolution (32) and a major challenge for community selection (6). In the disassembly method, as in (6), we sidestep the issue of heritability and ecological stability by re-assembling communities in a fixed abundance and equal species proportions. In our models, this allows community dynamics to unfold in almost the same way after each transfer. The disadvantage of this approach is that we cannot guarantee that communities selected by disassembly would maintain either their community composition or function when propagated by regular dilution. One way to overcome this would be to include a few rounds of transferring without selection to allow the communities to equilibrate before each selection step (23). This would, however, make disassembly quite inefficient, in which case, propagule with invasion can be more stable than disassembly (Fig. 6A).

Another factor that we do not explore in detail here is the relationship between community stability and the timing of selection. While our model is robust to timing errors, a community that degrades the toxic compounds too quickly can in principle be invaded by cheater cells that profit from the initial degradation and out-compete degraders in the absence of toxic compounds (6). Such ecological succession was found in a chitin-degrading community (16, 33).

While we have focused on increasing community performance from an applied perspective, our study provides important insights into discussions on the levels of selection and whether group selection can dominate over individual-level selection. Our choice to implement mutations (unlike other theoretical studies (23)) means that selection will also act at the level of individual genotypes. We found that competition between individuals is indeed strong in our simulated communities, leading individual species to reduce their investment into degradation in favor of growth. Interestingly, though, the faster growth rate of these evolved species compensates for the reduced investment by producing larger population sizes of degraders, such that the difference in degradation between ancestral and evolved communities was negligible. Our approach does not overcome within-community selection for growth, but this may not impact community function as much as one might expect when the community function is indirectly coupled to growth. Going back to the practical perspective, in a lab setting, it may be worth investigating whether the best selected species composition is more efficient when assembled from ancestral or evolved strains.

Ultimately, the goal of our investigation is to help design experiments rather than just computer simulations. While we found that our method can efficiently search for better species combinations, its scope is limited to communities that can be disassembled in the lab, and the effort needed to find out how to isolate species from the communities should not be underestimated. We have, however, confirmed that a periodic introduction of species (which is possible independently of whether we can remove species or disassemble the communities) improves the propagule method (23) to balance experimental feasibility and improvement of community functions.

From an experimenter's perspective, it is also useful to consider where one can reduce the size and complexity of the experiment. We find 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). Disassembling rich communities may however prove quite challenging in practice, and future experimental work will aim to make this step more efficient. We also 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 assumptions for 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 (25).

Taken together, we have introduced a new 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 been able to improve community function with respect to randomly assembled communities, but show that genetic mutation can contribute to reduced investment by individual strains into community function. We are testing this approach experimentally in parallel work.

Materials and methods

Growth and community function. 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. Both models simulate different microbial species growing together in batch culture, in a medium containing 4 types of nutrients that allow cells to grow, and 10 toxic compounds that cause cell death. Species are described by how quickly they grow, the rate at which they convert nutrients into biomass and how they are affected by toxic compounds. We describe the details of both models in the following sections. All interspecies interactions are due to the consumption of nutrients and degradation of the toxic compounds. The concentration of each nutrient $j = 1,\ 2,\ 3,\ 4$ is denoted by N_j and the concentration of each toxic compound $k = 1,\ 2, \ldots,\ 10$ is denoted by T_k. The removal of toxic compounds in either model is determined by the parameters f_{ik} that denote the fraction of resources that a cell (IBM) or population (ODE) of type i invests into the degradation of toxic compound k. The remaining fraction $1 - \sum_k f_{ik}$ is invested into growth. As we describe in the following section, f_{ik} is the only parameter that is subject to mutation and therefore gives rise to different strains of the same species.

Communities are created in two steps: First we draw a library of 15 species, each with unique (random) combinations of model parameters sampled as described in Table 1 and 2. These species are then randomly assigned (with replacement) to 21 communities of four species each, such that each species is present in at least one community.

In each round, each community grows and degrades the toxic compounds independently of the other communities, for 80 time-steps in the IBM and 100 in the ODE model. The root-mean-square decrease in T_k from the initial time point t_0 to the last t_{end} forms the basis of our community function, the degradation score D:

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

These scores are used to rank the communities, so that we can propagate the best ones by the different selection methods (elaborated below). To compare the selection methods, we simulate 50 rounds of growth, degradation and selection with the different methods. Upon propagating communities to fresh medium, we only transfer the cells and no leftover media. To reduce bias due to initial conditions, we run 10 simulations with different initial communities for each out of 5 sets of 15 species (Fig. 1F). The same exact set of initial conditions is used for all selection methods to make comparisons fair.

Within-species evolution. In both models, a strain i of the ancestral species l_i ("l" as in lineage) invests a fraction $0 \leq f_{ik} \leq 1$ of the resources that they take up into degradation of toxic compound k. These fractions can mutate, creating new strains of the same species that invest different amounts into degradation and the rest, $1 - \sum_k f_{ik}$, into population growth. The mutations take place at cell division in the IBM and during community replication in the ODE. The distinction between "species" of a fixed set of 15 and "strains", that emerge from an ancestral strain of each species as mutations arise, is important. For example, in the disassembly method described below, we inoculate a fixed concentration of each *species* in each round, which can consist of different *strains* of the species, depending on their frequency in the previous round.

To evaluate the evolution of the total investment $f_{i\cdot} = \sum_k f_{ik}$ (Fig. 5A) of species l, we compare the investment of the ancestral strain i 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 strains i of species l,

$$\tilde{f}_{l\cdot} = \sum_{i \text{ of species } l} S_i f_{i\cdot} / \sum_i S_i \qquad \textbf{(2)}$$

to emphasize the investment of abundant strains instead of small recent mutants that have not contributed as much to the community function. The value $\tilde{f}_{l\cdot}$ 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.

Individual-based model (IBM). We simulate well-mixed batch cultures that initially contain 10 cells per species in the communities. All cells are initially equally exposed to $T_k(t_0) = 700$ units of each of the 10 toxic compounds, which can cause cell death (see below), and have equal access to $N_j(t_0) = 2000$ units of each of the 4 nutrients that allow the cells to divide and reduce the concentrations of toxic compounds (see below). These processes occur according to the parameters of each strain i (Alg. 2, Table 1, Fig. S12).

Cell division consists of two steps, here called "activation" and "replication", that respectively involve the acquisition of some nutrients by the cell and their utilization for cell reproduction (34). Cells of strain i share their parameters (Table 1) and can be in either of two states: the initial "inactivated" with population size p_{i0}, or the "activated" with size p_{i1}. Every activation, replication, mutation and death event occurs with a given probability by random sampling from a Poisson distribution with rate $population \cdot probability$ (22).

Degradation and activation are costly and are carried out first. At each time step, a cell of type i can take up an amount n_{ij} of the nutrient j with current concentration N_j, with

$$max\_uptake = \sum_j (n_{ij} \text{ if } N_j > n_{ij}). \qquad \textbf{(3)}$$

The max_uptake scales down the amount of degradation or the probability to activate when not all the nutrients consumed by the cell are present.

Regardless of their activation state, at each time step, all the cells of a strain (i.e. the total population size $S_i = p_{i0} + p_{i1}$) degrade the amount $f_{ik} \cdot max\_uptake$ units of each toxic compound k, consuming $f_{i\cdot}$ units of nutrients in total. If there are not enough nutrients, a smaller fraction of S_i degrades, determined by the amount of nutrients available. When a toxic compound is depleted, its degradation and thus the corresponding nutrient consumption does not occur.

Following degradation, an inactive cell can activate with probability

$$a_i \cdot (1 - \sum_k f_{ik}) \cdot max\_uptake \cdot \sum_j \left(\hat{n}_{ij} \frac{N_j(t)}{N_j(t_0)} \right), \qquad \textbf{(4)}$$

where $\hat{n}_{ij}$ is a re-scaled version of n_{ij}: if some nutrients are depleted, their n_{ij} are set to 0 and the n_{ij} of the remaining nutrients are re-scaled such that $\sum_j n_{ij} = 1$. To activate, a cell of type i consumes in total $1 - f_{i\cdot}$ units of nutrients and if the it does not replicate, it needs the same amount of nutrients in subsequent time steps to remain activated.

The amount of nutrient of type j that is consumed for degradation and activation depends on the parameters n_{ij} and on the amount of each nutrient that is available. When one type of nutrient gets depleted, cells will take up more of the other available nutrients that they require. At every time step we check how many cells of strain i can degrade and activate based on the scarcest nutrient. Thus, when a nutrient j is nearly depleted, fewer cells that require this nutrient degrade and have the chance to activate in the current time step. In following time steps, when nutrient j is depleted, cells consume the remaining required nutrient types. So, although now cells would use nutrients less efficiently, if these nutrients are sufficiently abundant, a greater population can degrade and activate. This models a pause in degradation and division caused by a metabolic shift towards consuming fewer nutrients. These events are stochastic and lead to noise between runs of the model with the same starting conditions. When we calculate the growth and degradation of specific communities such as the 32767 possible combinations of species (Fig. 2 E-F, Fig. 3 G), we average the results over three replicates of the simulations. We use the same seed for the random number generators for consistency.

At each time step, activated cells divide with probability $r_i \cdot (1 - f_{i\cdot})$ without additional cost, resulting in two inactivated daughter cells: one daughter maintains the parameter values, and the other is susceptible to mutation with probability $\mu = 0.01$. Upon mutation, the previous value of at least one f_{ik}

Parameter	Description	Randomly sampled from
l_i	Species ID of strain 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 $\sum_j n_{ij} = 1$
f_{ik}	Fraction of consumed nutrients invested into degradation of toxic compound k	Uni(0,1), Sparse, Rescaled so that $\sum_k f_{ik} = Uni(0,1)$
m_{ik}	Death rate of strain i due to toxic compound k	Uni(0.001,0.02), Sparse

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 $[0,\ 1]$.

is multiplied by a random number from the lognormal ($\mu = 0, \sigma^2 = 0.4$) distribution, making sure the total investment $f_{i\cdot}$ falls in the $[0,\ 1]$ interval. As a result, a new strain of the same species with population $p_{i0} = 1$ is introduced.

At each time step, activated and inactivated cells may die with a probability determined by the following Hill function:

$$\sum_k m_{ik} \frac{T_k^2}{T_k^2 + K^2} \tag{5}$$

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

Population-level model. As the IBM above, the ODE model simulates well-mixed batch cultures with nutrients and toxic compounds, extending a previous model (25). In the model, the population size S_i of each strain i in a community grows in relation to the concentrations of nutrients N_j and decline by the toxic compounds T_k (Fig. S13) by the model parameters in Table 2. Growth, death, nutrient uptake and degradation is described by the following ODE system:

$$\frac{dS_i}{dt} = \left((1 - \sum_k f_{ik}) \rho_i(\mathbf{N}) - \mu_i(\mathbf{T}) \right) S_i \tag{6}$$

$$\frac{dN_j}{dt} = -\sum_i \frac{\rho_i(N_j)}{Y_i} S_i \tag{7}$$

$$\frac{dT_k}{dt} = -T_k \sum_i f_{ik} \delta_i \rho_i(\mathbf{N}) S_i \tag{8}$$

The bold-face $\mathbf{N}$, $\mathbf{T}$ denote the vectors of all nutrients and toxic compounds, respectively. We assume Monod and Hill functions for the per-capita growth and death rates ρ_i, μ_i.

$$\rho_i(\mathbf{N}) = \sum_j r_{ij} \frac{N_j}{N_j + K_N} \tag{9}$$

$$\mu_i(\mathbf{T}) = \sum_k m_{ik} \frac{T_k^2}{T_k^2 + K_T^2} \tag{10}$$

The system of equations Eq. (6)–Eq. (8) is solved with a standard ODE solver (`dopri5`, (35, 36)) for 100 time steps with initial conditions $S_i(t_0) = 100$, $N_j(t_0) = 100$ and $T_k(t_0) = 100$ for all i, j, k.

The investment f_{ik} can mutate to form different strains of the same species. When this happens, we add a new population equation of the type Eq. (6) to the ODE system, with the same parameters r_{ij}, m_{ik}, Y_i and d_i as the ancestor but with the modified f_{ik}. To not make the system of equations too large, we have limited the number of strains to 28 per community. We estimate this to be enough since we expect mutants to rapidly replace their ancestral strains if their growth rate is higher, and otherwise disappear rapidly. If there are already 28 strains in a community, then no more mutants are allowed. Otherwise, when communities are propagated to the next round of growth, any surviving strain can have a mutant with probability 0.05. Having chosen which strains i to mutate, we pick one or more traits f_{ik} at random and multiply them by numbers drawn at random from lognormal$(0,\ 0.4)$ and ensure that both the mutated traits f_{ik} and the total investment $f_{i\cdot}$ falls in the $[0,\ 1]$ interval. The mutant receives the same r_{ij}, m_{ik}, Y_i and d_i parameters as its ancestor and is introduced with population size 100, the same as the initial population before the first round of growth. This population size is chosen relatively high, in order to speed up the competition between ancestor and mutant strain.

Parameter	Description	Sampled from
l_i	Species ID of strain i	
r_{ij}	Maximum growth rate with respect to nutrient j	$\mathrm{Uni}(0.01,\ 0.1)$ Sparse
K_N	Half-saturation constant for nutrients	$K_N = 10$ (fixed)
m_{ik}	Maximum death rate with respect to toxic compound k	$\mathrm{Uni}(10^{-4},\ 10^{-3})$ Sparse
K_T	Half-saturation constant for toxic compounds	$K_T = 10$ (fixed)
f_{ik}	Fraction of amassed nutrients that are invested into degradation of toxic compound k	$\mathrm{Uni}(0,1)$, Sparse Rescaled so that $\sum_k f_{ik} = \mathrm{Uni}(0,1)$
Y_i	(Average) biomass yield with respect to the nutrients	$\mathrm{lognormal}(\log(10^{-3}), \log(5))$
δ_i	(Average) degradation efficiency with respect to the toxic compounds	$\mathrm{lognormal}(\log(10^{-4}), \log(5))$

Table 2. Parameters defining a microbial strain i in the ODE model. Growth and degradation parameters in relation to nutrients and toxic compounds N_j and T_k. All parameters are assumed to be positive, and the investment f_{ik} is limited to the interval $[0,\ 1]$. The matrices of growth rates, death rates and degradation investment r_{ij}, m_{ik} and f_{ik} are made sparse by multiplying them by matrices drawn from $\mathrm{Bernoulli}(0.5)$, i.e. flipping a coin for each entry. In this way, each species takes up approximately half of the nutrients, is affected by half of the toxic compounds and degrades half of the toxic compounds.

Artificial selection methods. After scoring all communities in a given round (see above), a fraction of these "parent" communities is propagated to "offspring" for the next round of growth (Fig. 1A, Alg. 1). Here we implement several methods of propagating the selected communities as described below (Fig. 1B, C), and compare them to a no-selection control (NS) where each community is propagated by 100-fold or approximately 20-fold (stochastic process based on Poisson distribution) dilution, respectively for the ODE model or IBM (Fig. 1E). NS shows the baseline change in community function due to interspecies interactions and changes to the species composition (23).

Propagule selection. In the propagule method, the 7 communities with the highest scores are propagated to the next round by dilution (10, 24) (Fig. 1C, Alg. 7, Alg. 8). The communities are populated uniformly such that each selected parent contributes 3 offspring communities for the next round. The important design parameters are the dilution factor and the fraction of communities to propagate, i.e. the selection bottleneck. In previous experimental studies, dilution factors between 5 and 30 have been used for bottlenecks between 1/10 and 1/3 of parent communities (8–10, 17, 19). We keep a wide bottleneck, selecting 7 out of 21 communities, before diluting them by a factor 100 (ODE model) or approximately 20 in the IBM (where population sizes are smaller and we sample the cells to propagate at random, according to a Poisson distribution) for the next round. See also simulation studies in (5, 6, 22, 23, 37). We compare PS to the random control PR, where the communities are selected at random without regards to 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 to 5 out of the 21 offspring communities (chosen at random with uniform probability) (23).

Migrant pool selection. Here, selected parent communities are mixed in a *migrant pool* before new offspring communities are formed by taking samples from the pool (10, 24) (Fig. 1D, Alg. 9). Previous experiments have used microbial communities from wastewater (18), soil and rhizospheres (8, 10–14), marine environments (16) and other strain collections (38), selecting between 1/10 and 2/7 of communities and diluting them by factors between 1/100 and 1/2. The method has also been subject to at least one simulation study (5). We select 7 communities out of 21, merge them in a pool and create 21 new communities by sampling without replacement cells from the pool with an approximately 20-fold dilution (stochastic process according to Poisson distribution). We compare MS to a random control (MR), where we select communities with uniform probability without regards to their degradation scores. We also implement a version of migrant pool selection where we introduce one or more species 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. Our proposed method is intended for synthetic communities, where each species can be grown separately and isolated from a multispecies community, such that the communities can be disassembled between transfers. We select 7 out of 21 parent communities by degradation scores (Fig. 1D, Alg. 10). By disassembling these communities, we maintain a record of samples of each species that were present in at least one selected community in each round. If a species is present in more than one selected community, we sample from the highest-scoring community that this species was part of. In this way, we are able to re-introduce any species that went extinct.

To select against extinctions and communities whose members out-compete one another, we scale the degradation score D by the fraction of surviving species at the end of a round of growth as follows:

$$\hat{D} = D \times \frac{\text{number of surviving species}}{\text{number of species in the community}}. \quad \textbf{(11)}$$

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:

$$\frac{1}{\sum_k \hat{D}_k} \hat{D}_n, \quad \textbf{(12)}$$

for each selected community. In this way, parent communities with (i) high degradation scores and (ii) low or no extinctions will have more offspring. The offspring communities have the same species composition as their parents, but we reset the initial abundance of each species to a specified population size (100 in the ODE, around 10 cells in the IBM by a random sampling with replacement from a Poisson distribution) to standardize the growth conditions between rounds, i.e. to maintain heritability.

To introduce variability between communities, we change the species composition of a few of the 21 offspring communities. First, we choose 5 offspring communities at random and remove one or more species, always one species plus an additional number drawn from $\text{Poisson}(0.5)$. 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 introduce one or more invader species —as above, $1+\text{Poisson}(0.5)$— chosen with uniform probability from the frozen stock, to 5 randomly chosen communities. 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. See Alg. 10 for more details.

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 (36) implementations for all three methods.

We quantify species diversity within a community (Fig. 3E, Fig. S2) as the Hill number of order 1 or the average effective number of species present in the community (39, 40), which is based on the Shannon index H':

$$\exp(H') = \exp(-\sum_{l=1} p_l \log p_l), \quad \textbf{(13)}$$

which in turn depends on the species' relative abundances

$$p_l = S_l / S_{tot} \quad \textbf{(14)}$$

where we divide the population size S_l of each species by the total population size in the community $S_{tot} = \sum_l S_l$. If more than one strain 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. 4 C, D, Fig. S4) 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. 3F) 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 $\tilde{f}_{\cdot k} = \sum_l \tilde{f}_{lk}$ of total investment into degrading toxic compound k in a given community. The 'tilde' indicates that we have scaled the f_{lk} for each strain of a species by the corresponding population sizes of the strains in the community as in Eq. (2). In this way, we emphasize the most relevant strain of each species and do not bias the result to the number of competing strains 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

$$\exp(H') = \exp\left(-\sum_{k=1} \tilde{f}_{\cdot k} \log \tilde{f}_{\cdot k}\right), \qquad \textbf{(15)}$$

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 strains 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 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. S9), 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. S11), 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. 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.
5. 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. .
6. 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.
7. Melanie Ghoul, Ashleigh S. Griffin, and Stuart A. West. Toward an Evolutionary Definition of Cheating. *Evolution*, 68(2):318–331, 2014.
8. W. Swenson, D. S. Wilson, and R. Elias. Artificial ecosystem selection. *Proceedings of the National Academy of Sciences*, 97(16):9110–9114, 2000.
9. William Swenson, Jeff Arendt, and David Sloan Wilson. Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. *Environmental Microbiology*, 2(5):564–571, 2000.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. Chang-Yu Chang, Melisa L. Osborne, Djordje Bajic, and Alvaro Sanchez. Artificially selecting bacterial communities using propagule strategies†. *Evolution*, 74(10):2392–2403, 2020.
20. U.G. Mueller and J.L. Sachs. Engineering Microbiomes to Improve Plant and Animal Health. *Trends in Microbiology*, 23(10):606–617, 2015.
21. Li Xie and Wenying Shou. Steering ecological-evolutionary dynamics to improve artificial selection of microbial communities. *Nature Communications*, (12):264697, 2021.
22. 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.
23. 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.
24. 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.
25. 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.
26. Michael J Wade. An Experimental Study of Group Selection. *Evolution*, 31(1):134–153, 1977.

27. 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.
28. Melanie Mitchell. *An Introduction to Genetic Algorithms.* MIT Press, Cambridge, MA, USA, .
29. John H. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, MI, USA, .
30. Sarah P. Otto and Thomas Lenormand. Resolving the paradox of sex and recombination. *Nature Reviews Genetics 2002 3:4*, 3(4):252–261, 2002.
31. 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.
32. R. C. Lewontin. The Units of Selection. *Annual Review of Ecology and Systematics*, 1(1): 1–18, 1970.
33. Manoshi S Datta, Elzbieta Sliwerska, Jeff Gore, Martin F Polz, and Otto X Cordero. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature Communications*, 7(1):11965, 2016.
34. 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.
35. 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, .
36. Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python. [Online; accessed 2019-10-21].
37. 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.
38. 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.
39. M. O. Hill. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology*, 54(2):427–432, 1973.
40. 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.

Supplementary tables

Method 1	Method 2	p (IBM)	p (ODE)
DS	$\emptyset$	4×10^{-10}	4×10^{-10}
DS	PS	8×10^{-10}	8×10^{-10}
DS	MS	8×10^{-10}	-
DS	PIS	1×10^{-9}	2×10^{-8}
DS	MIS	8×10^{-10}	-
DS	DR	8×10^{-10}	8×10^{-10}
DS	NS	8×10^{-10}	8×10^{-10}
PS	$\emptyset$	1.0	1.0
PS	MS	1×10^{-7}	-
PS	PIS	4×10^{-9}	7×10^{-5}
PS	PR	8×10^{-10}	3×10^{-9}
PS	NS	8×10^{-10}	0.1
MS	$\emptyset$	1.0	-
MS	MIS	0.67	-
MS	MR	8×10^{-10}	-
MS	NS	0.3	-
PIS	$\emptyset$	9×10^{-6}	0.8
MIS	$\emptyset$	1.0	-

Table S1. P-values for Fig.2A-D a Wilcoxon signed-rank test of difference in maximum degradation between methods.

Method 1	Method 2	p (IBM)	p (ODE)
DS	PS	8×10^{-10}	7×10^{-10}
DS	MS	8×10^{-10}	-
DS	PIS	8×10^{-10}	7×10^{-10}
DS	MIS	8×10^{-10}	-
DS	DR	8×10^{-10}	8×10^{-10}
DS	NS	8×10^{-10}	7×10^{-10}
PS	MS	1×10^{-4}	-
PS	PIS	3×10^{-8}	7×10^{-4}
PS	PR	2×10^{-9}	7×10^{-7}
PS	NS	8×10^{-10}	3×10^{-6}
MS	MIS	8×10^{-2}	-
MS	MR	2×10^{-8}	-
MS	NS	2×10^{-9}	

Table S2. P-values for Fig.2E-H a Wilcoxon signed-rank test of difference in degradation ranks between methods.

Method 1	Method 2	p (IBM)	p (ODE)
DS	PS	6×10^{-9}	8×10^{-10}
DS	MS	8×10^{-10}	-
DS	PIS	8×10^{-10}	8×10^{-10}
DS	MIS	8×10^{-10}	-
DS	DR	8×10^{-10}	6×10^{-4}
DS	NS	8×10^{-10}	8×10^{-10}
PS	MS	1×10^{-9}	-
PS	PIS	0.1	0.01
PS	PR	2×10^{-8}	0.5
PS	NS	8×10^{-10}	0.3
MS	MIS	0.13	-
MS	MR	1×10^{-9}	-
MS	NS	1×10^{-5}	-

Table S3. P-values from a Wilcoxon signed-rank test of difference in total investment in communities, between methods for Fig.3A-D.

Supplementary figures

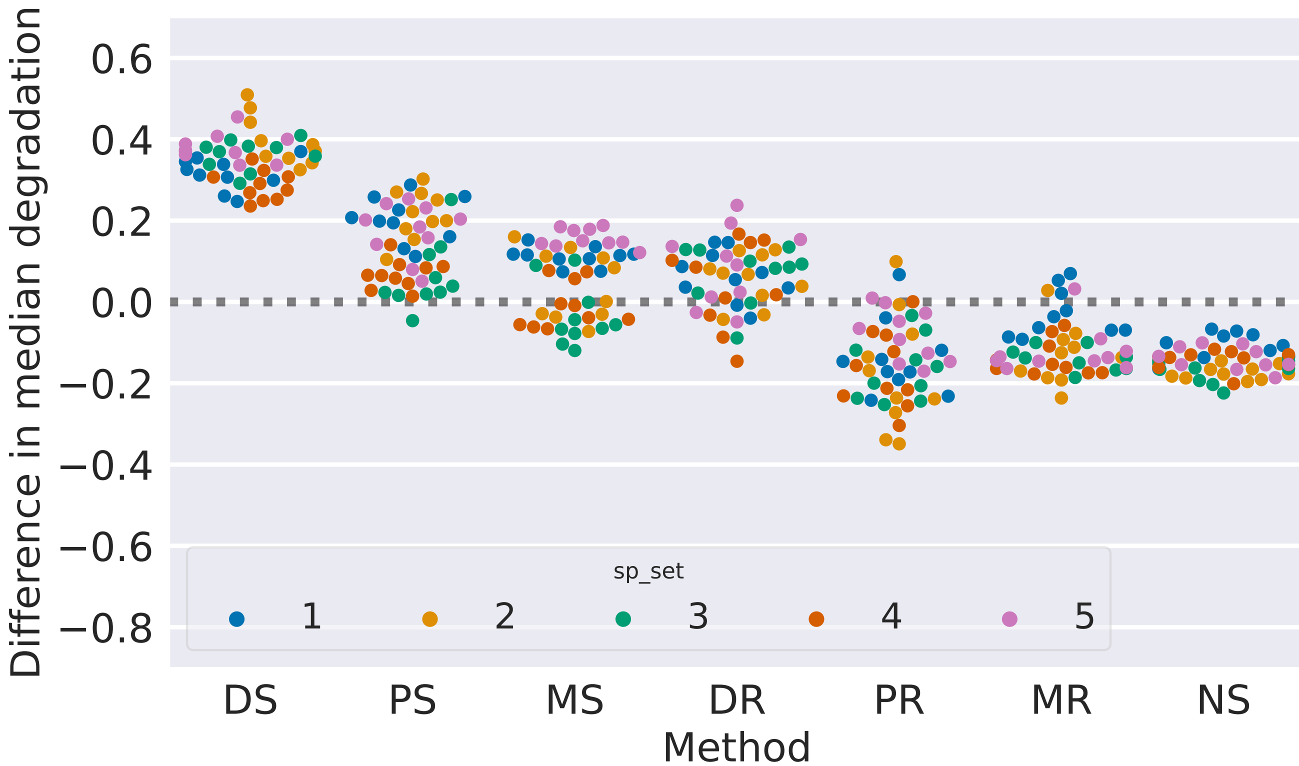

Fig. S1. The difference in median degradation between round 50 and round 0 for each propagation method, corresponding to Fig. 2. The two-sided Wilcoxon test for difference in degradation against the no-selection control is significant for the selection methods DS, PS and MS. Data generated by the IBM.

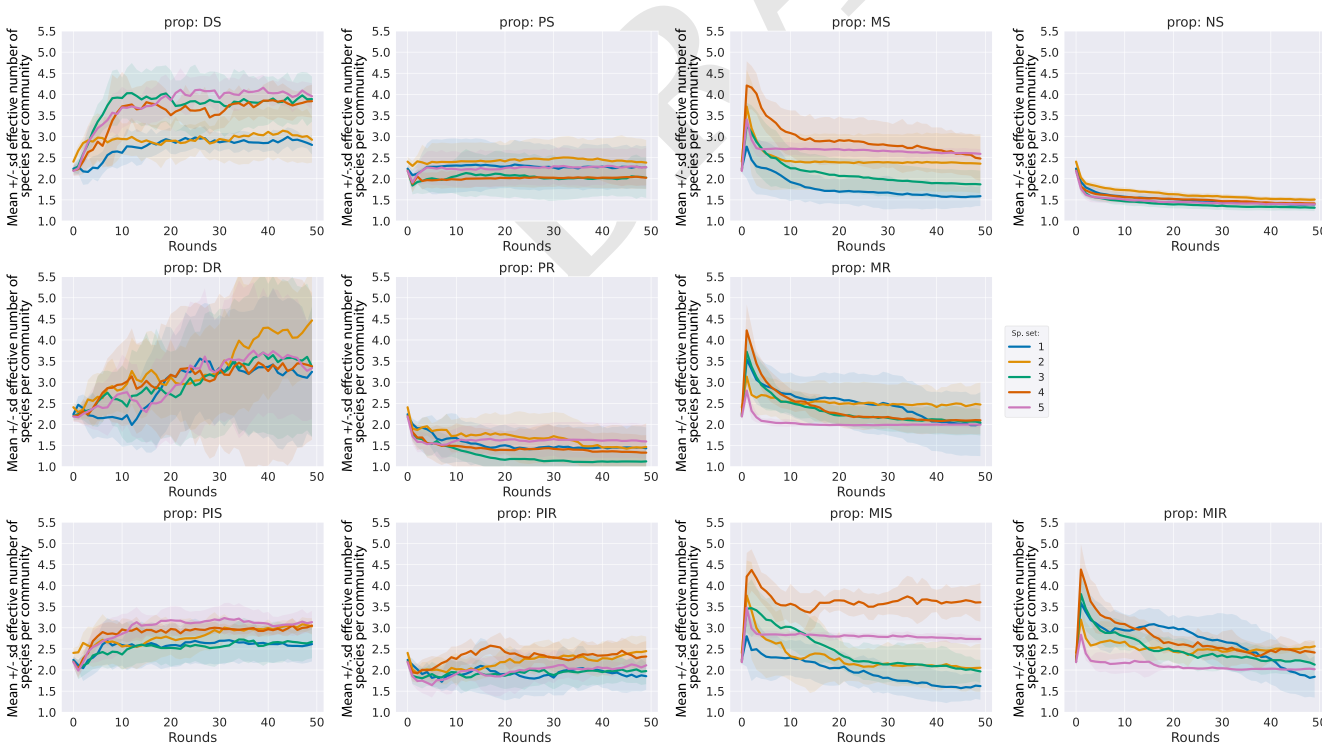

Fig. S2. Time-series of the species diversity (effective number of species per community) corresponding to Fig. 4D. Each panel shows the mean ± standard deviation over the 10 repeated runs, for each species set 1-5, for one propagation method. Data generated by the IBM.

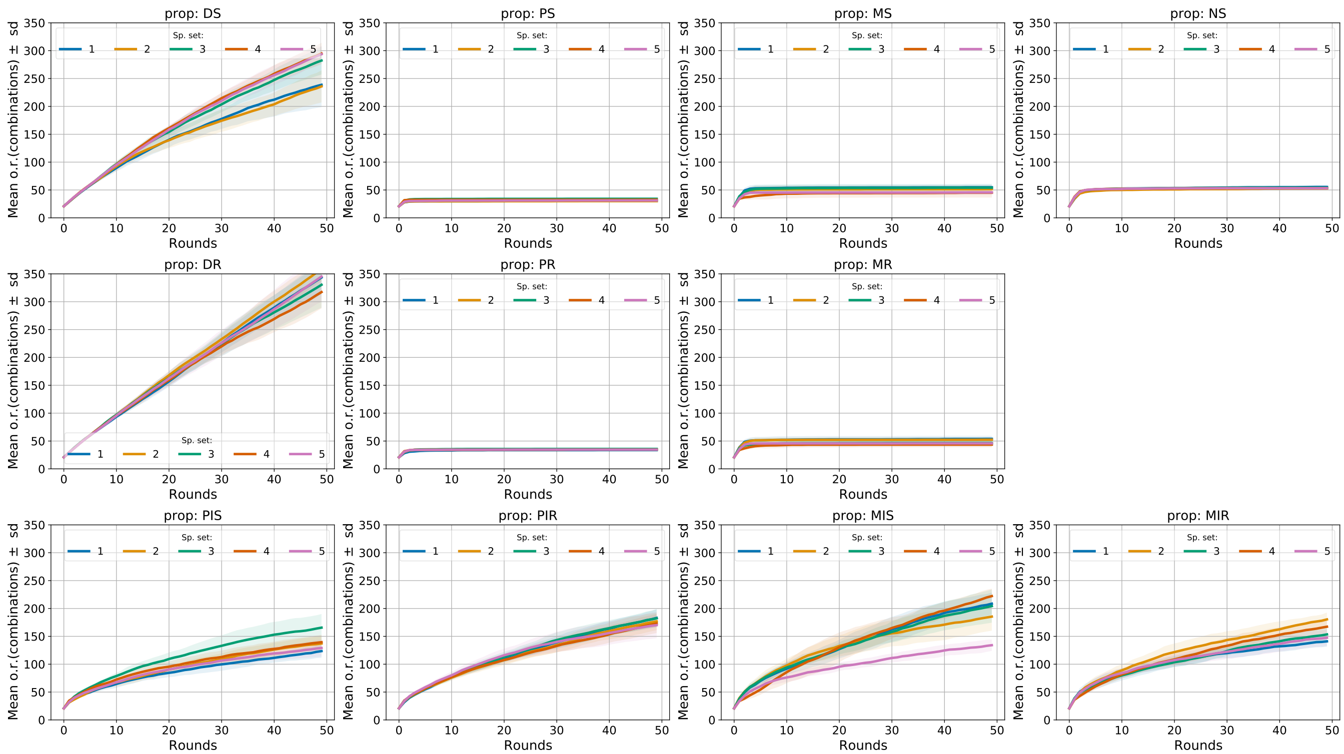

Fig. S3. Time-series of the number of explored communities, corresponding to Fig. 4B. Each panel shows the mean ± standard deviation over the 10 repeated runs, for each species set 1-5, for one propagation method. Data generated by the IBM.

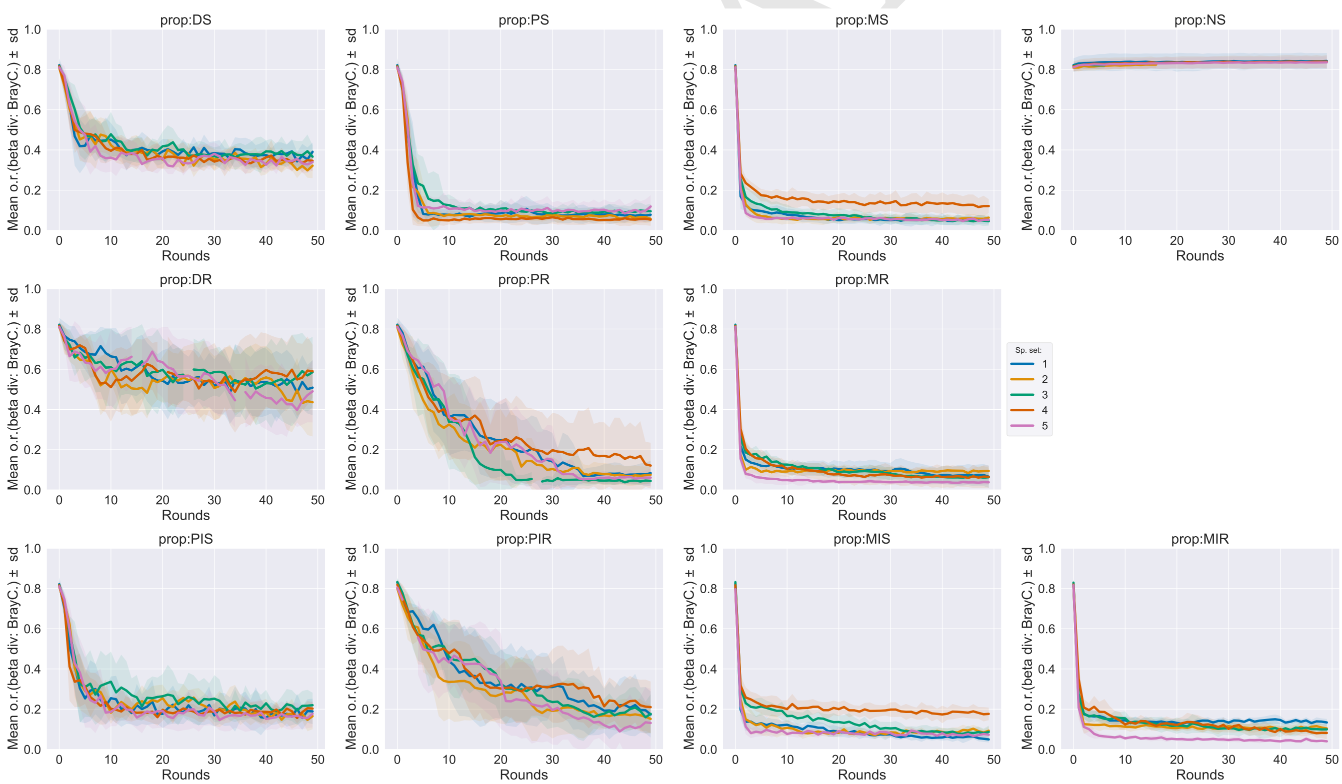

Fig. S4. Time-series of the beta diversity corresponding to Fig. 4F. Each panel shows the mean ± standard deviation over the 10 repeated runs, for each species set 1-5, for one propagation method. Data generated by the IBM.

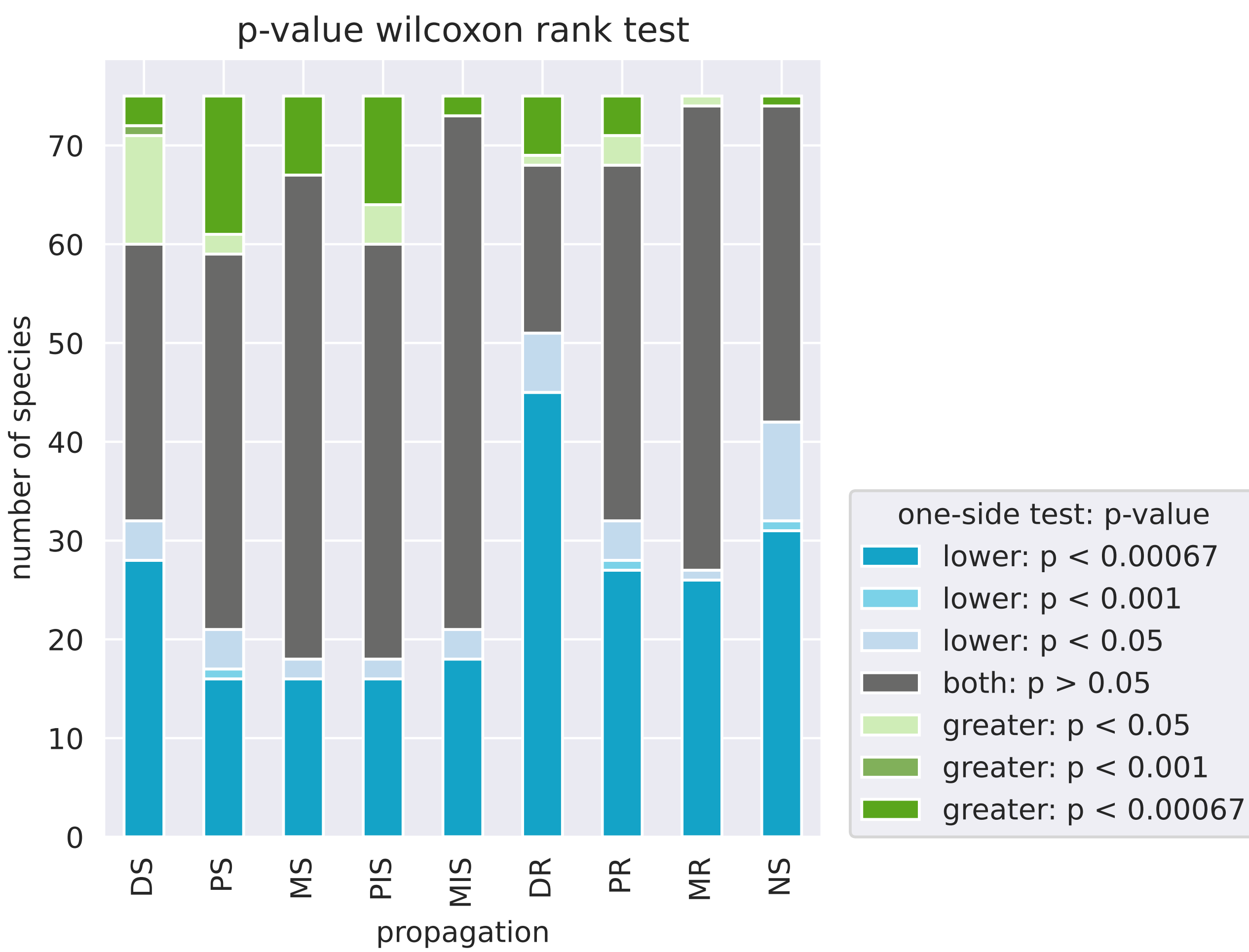

Fig. S5. Distribution of p-values from a one-sided Wilcoxon signed-rank test of whether the total investment $f_{l\cdot}$ of a species is larger/smaller in the last round where a species survived, than the investment of the ancestral species. There is one bar for each selection method, with 15 species x 5 sets of species for each bar. The alternative hypothesis is that difference in investment (ancestral-evolved) is greater (green) or less (blue) than zero. Data generated by the IBM. Data for DS is shown in Fig. 5.

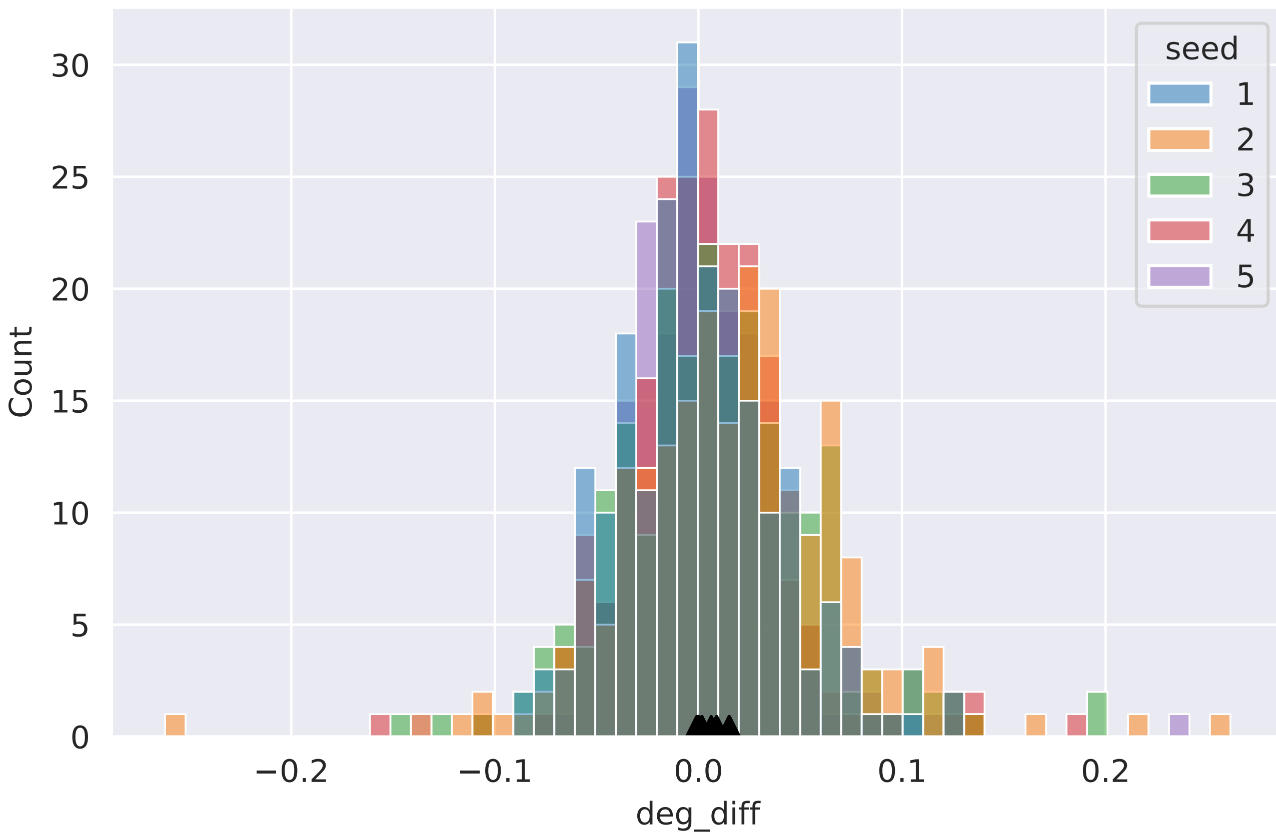

Fig. S6. Histogram of difference in max degradation between evolved and ancestral communities. Triangles indicate the mean values for each species set. Data generated by the IBM.

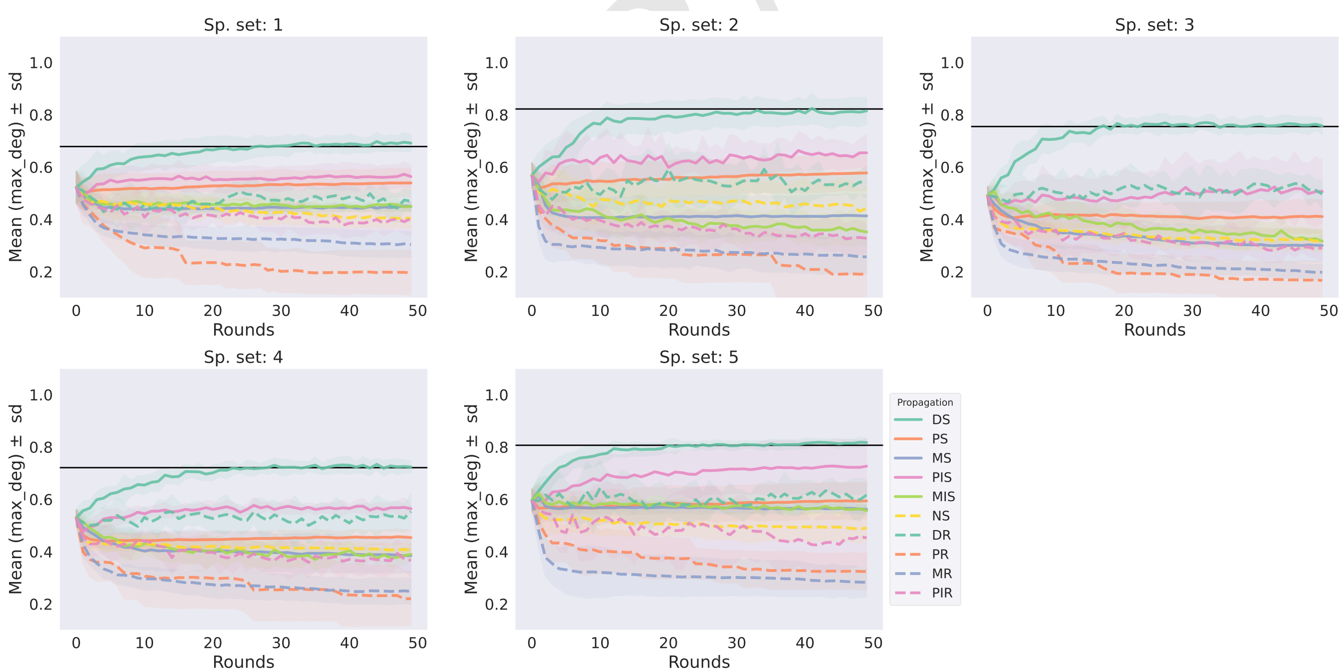

Fig. S7. Time-series of max. degradation over 50 rounds of selection, for the different propagation methods. Each plot corresponds to one species set and shows the maximum degradation in the meta-community averaged over repeats, with the standard deviation in shades of the corresponding color. For each species set, each repeat etc, the degradation score at transfer 50 forms the swarms in Fig. 2. The black line shows the degradation score of the best ancestral community out of the 32767 combinations. Data generated by the IBM.

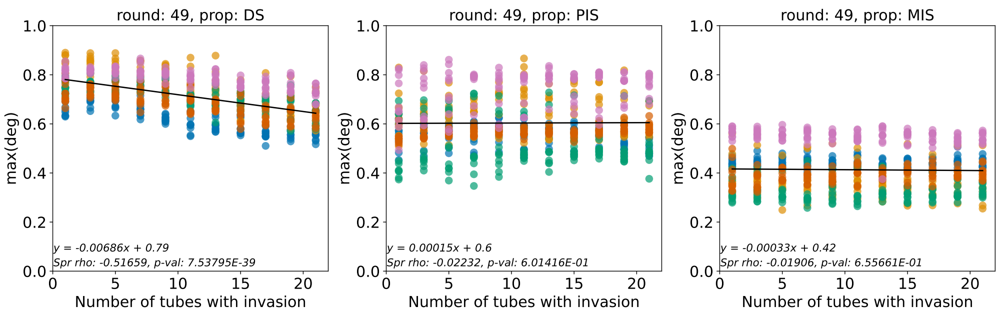

Fig. S8. Effect on max community degradation score from changing the number of communities to receive a migrating species. Data generated by the IBM.

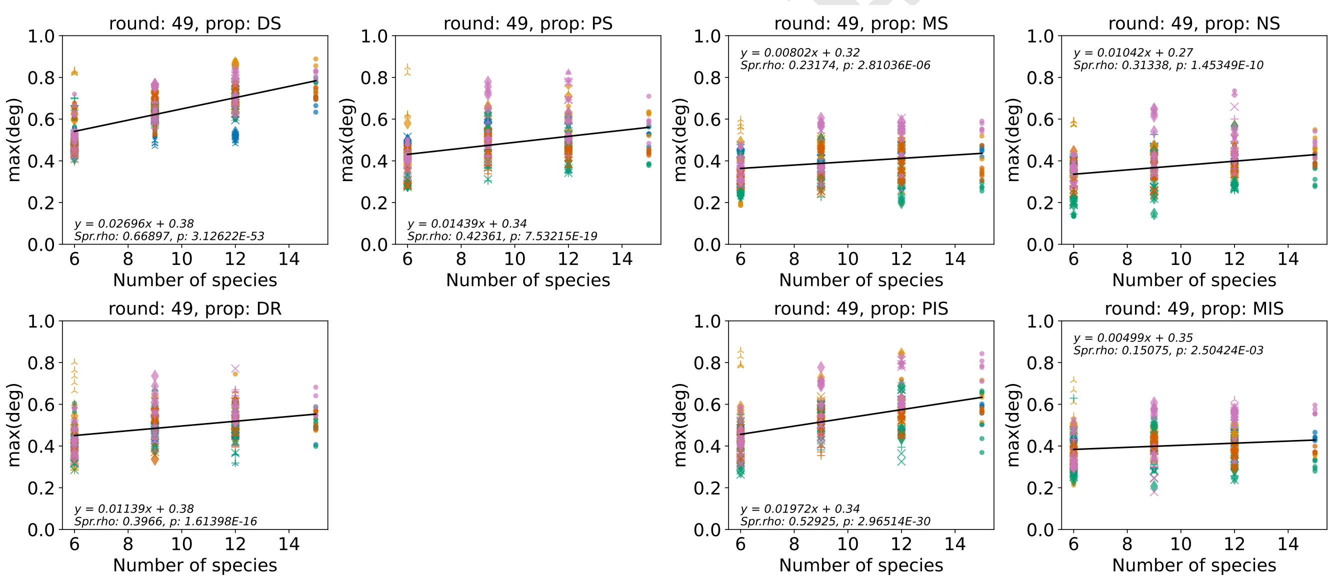

Fig. S9. Effect on max community degradation score from changing the number of species in the ancestral community. Different marker shape indicates sub-sample (1-5) for each species group of size 6, 9, 12. For 15 species we keep the original species sets. Data generated by the IBM.

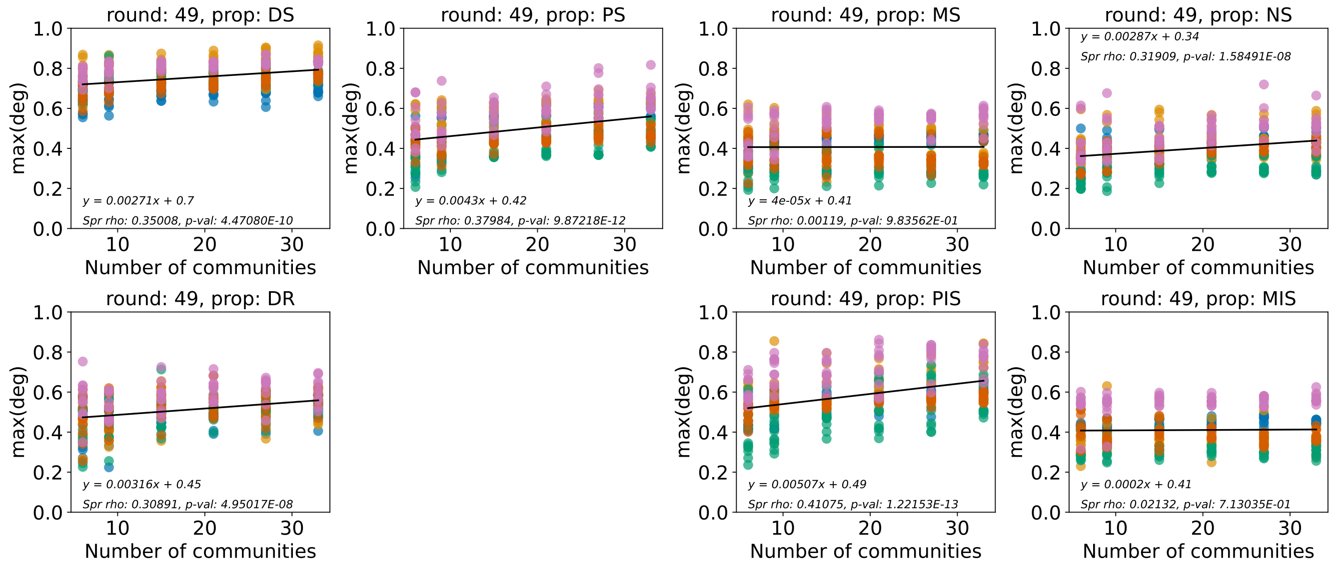

Fig. S10. Effect on max community degradation score from changing the number of communities. Data generated by the IBM.

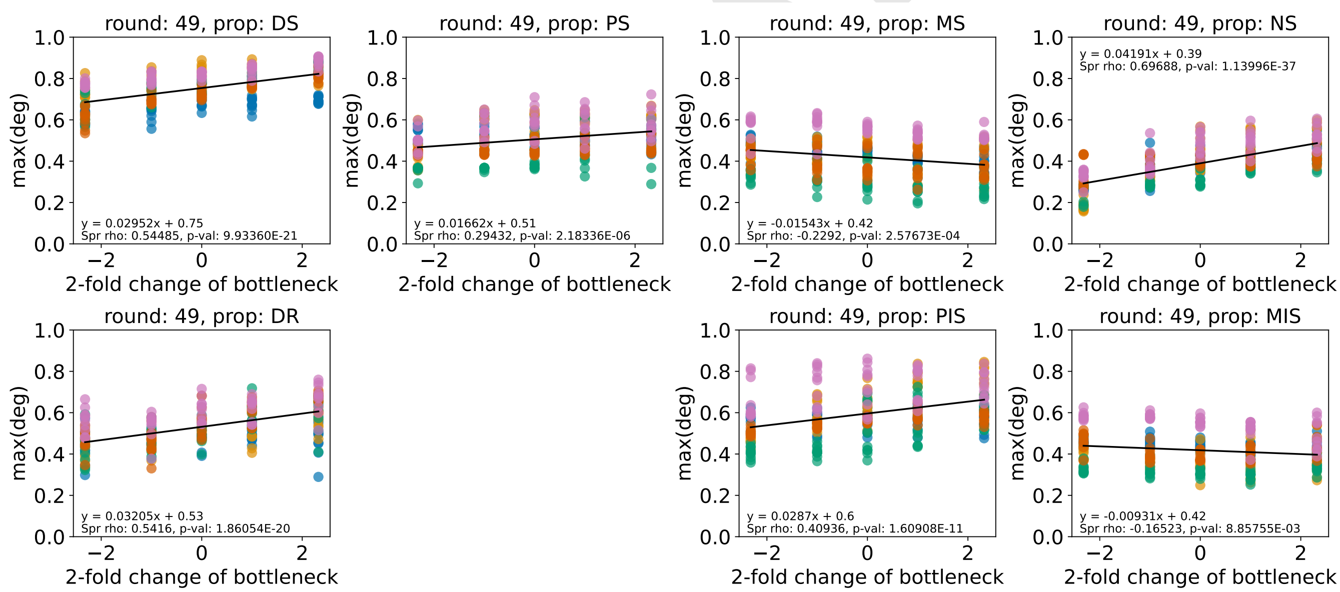

Fig. S11. Effect on max community degradation score from scaling the dilution factor or inoculum size. Data generated by the IBM.

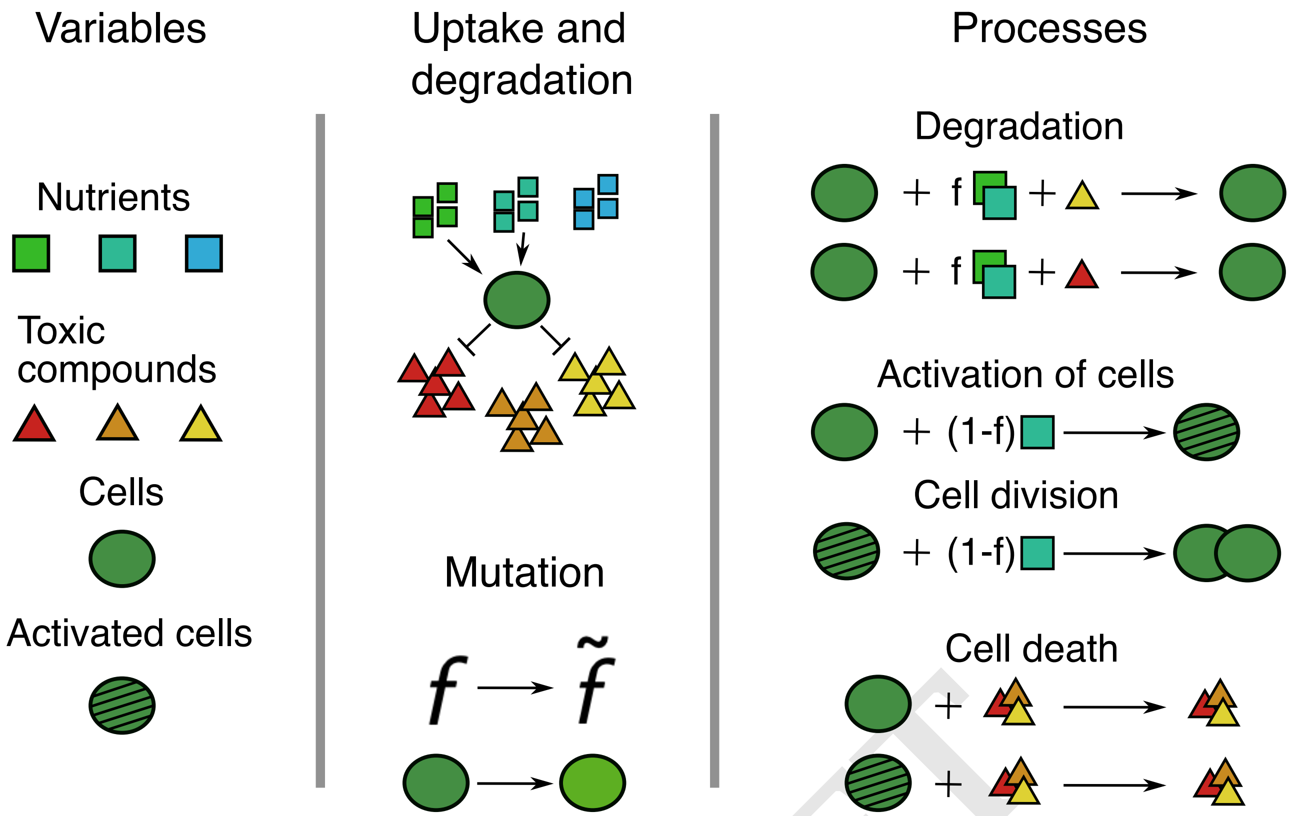

Fig. S12. Illustration of the variables and processes in the individual-based model. Cells of a certain species vary in their preferences for nutrients and degradation capabilities. Cells use the available nutrients to degrade the toxic compounds and for cell division.

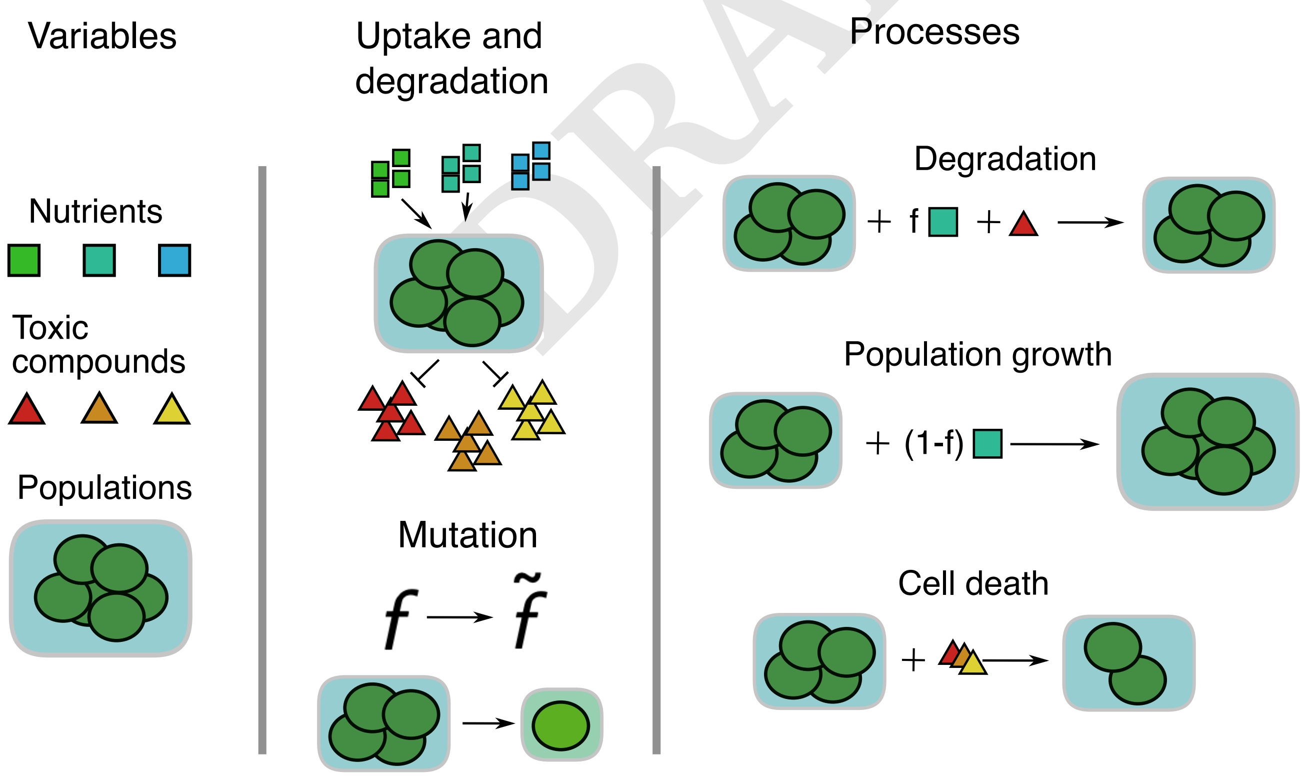

Fig. S13. Illustration of the variables and processes in the ODE model. Populations of cells vary in their preferences for nutrients and degradation capabilities. The populations use the available nutrients to degrade the toxic compounds and to grow.

1037 ## Pseudo-code for implementation of models and selection methods

Input: A set of 15 species, defined by model parameters.
Input: Experimental parameters: Number of communities, time span $[t_0,\ t_{end}]$ of growth in batch. Community bottleneck $\beta = 1/3$, dilution ratio $d \in [0,\ 1]$. Initial conditions $S_i(t_0)$, $N_j(t_0)$, $T_k(t_0)$.
Assemble 21 communities by randomly drawing 4 species with replacement from the species set. Ensure that each species is present in at least one initial community.
for Each round of selection **do**
    // ***Population growth, interspecies competition and invasion of mutants***
    for Each community **do**
        Grow the communities for a time span $[t_0,\ t_{end}]$. (IBM implementation: Alg. 2, ODE implementation: Alg. 6).
        Save the population sizes $S_i(t_{end})$ for each strain i in the community
        Save the end-state concentrations $T_k(t_{end})$ for each toxic compound k
        Compute the degradation score D from $T_k(t_{end})$ by Eq. (1)
    end
    // **Propagate the communities by the chosen selection method**
    // Required parameters: the community bottleneck β, dilution ratio d
    // Required variables: degradation scores D for each community
    // *For propagule method, follow Algs. 7 and 8*
    // *For migrant pool method, IBM only, follow Alg. 9*
    // *For disassembly method, follow Alg. 10*
    // **Replenish the substrates**
    Set $N_j(t_0) = N_0$ and $T_k(t_0) = T_0$ for all j, k
end

Algorithm 1: Overall flow of the selection simulations from population growth, community dynamics and mutations to propagation by the different selection methods.

Input: A community where each microbial strain i is defined by the parameters in Tab. 1. Size of the inactive and active sub-populations p_{i0}, p_{i1} and total population size $S_i = p_{i0} + p_{i1}$. Nutrient concentrations N_j, toxic compound concentrations T_k.
Input: Mutation parameters: mutation rate μ_{mut}, trait deviation σ_m.
for Each time step **do**
  for Each community $1, \ldots, 21$ **do**
    for Each strain i **do**
      // Maximum uptake of nutrients of this strain, used to scale the growth and degradation according to the concentration of nutrients
      $max\_uptake := \sum_j (n_{ij} \text{ if } N_j > n_{ij})$
      // If some nutrients are depleted we re-scale n_{ij} to consume the remaining nutrients in higher amounts
      if $N_j < n_{ij}$ **then**
        $n_{ij} := 0$
      end
      $\hat{n}_{ij} := \frac{n_{ij}}{\sum_j (n_{ij} \text{ if } N_j > n_{ij})}$
      // The largest number of cells that can consume the scarcest nutrient at this time step, we just take its integer part
      $S_i^{max} := int(min(N_j/\hat{n}_{ij} \text{ if } N_j > \hat{n}_{ij}))$
      // Toxic compound degradation
      if $S_i^{max} > 0$ **then**
        // Maximal population that can degrade
        $P_{i,tot} := \min(S_i^{max},\ S_i)$
        for Each compound T_k **do**
          if T_k *cannot be completely degraded in this time step* **then**
            $T_k := T_k - P_{i,tot} \cdot max\_uptake \cdot f_{ik}$
            for Each nutrient N_j **do**
              $N_j := N_j - \hat{n}_{ij} \cdot P_{i,tot} \cdot f_{ik}$
            end
          end
          else
            Degrade the remaining toxic compounds and consume the corresponding nutrients
          end
        end
      end
      // Cell division step 1: Costly activation
      Alg. 3.
    end
  end
  // Cell division step 2: Replication
  Alg. 4.
  // Cell death
  Alg. 5.
end

Algorithm 2: Implementation of population growth, competition and mutations in the IBM model described in the section *Individual-based model*.

Input: Communities where each strain i is defined by parameters in Tab. 1. Inactive and active sub-populations p_{i0}, p_{i1}. The maximal population S_i^{max} that can afford to consume nutrients, based on their current availability. Re-scaled nutrient consumption rates $\hat{n}_{ij}$. Current nutrient concentrations N_j.

Input: Parameters: Initial nutrient concentration N_0.

// Cell activation

if $S_i^{max} > 0$ **then**
  // Already activated cells consume nutrients
  if $S_i^{max} \geq p_{i1}$ **then**
    $N_j := N_j - \hat{n}_{ij} \cdot p_{i1} \cdot (1 - \sum_k f_{ik})$
    $S_i^{max} := S_i^{max} - p_{i1}$
  end
  else
    Deactivate cells that cannot afford to stay activated, consume the corresponding nutrients for cells that remain activated, set $S_i^{max} := 0$
  end
  // Newly activated cells
  $cells\_activate :=$ Poisson$(a_i \cdot p_{i0} \cdot max_uptake \cdot (1 - \sum_k f_{ik}) \cdot \sum_j (\hat{n}_{ij} \cdot N_j / N_0))$
  if $cells\_activate > p_{i0}$ **then**
    $cells\_activate := p_{i0}$
  end
  if $cells\_activate > S_i^{max}$ **then**
    $cells\_activate := S_i^{max}$
  end
  $p_{i0} := p_{i0} - cells\_activate$
  $p_{i1} := p_{i1} + cells\_activate$
  for Each nutrient N_j **do**
    $N_j := N_j - \hat{n}_{ij} \cdot cells\_activate \cdot (1 - \sum_k f_{ik})$
  end
end
else
  Deactivate all the activated cells
end
return *Populations* p_{i0}, p_{i1} *for strain* i*, current nutrient concentrations* N_j

Algorithm 3: Activation of cells, the first step of cell division in the IBM described in Alg. 2.

Input: Communities where each strain i is defined by parameters in Tab. 1. Inactive and active sub-populations p_{i0}, p_{i1}.
for Each community **do**
    for Each strain i in the community **do**
        // Calculate the number of new cells appearing due to division
        new_cells :=Poisson $(p_{i1} \cdot r_i \cdot (1 - \sum_k f_{ik}))$
        if $new\_cells > p_{i1}$ **then**
            $new\_cells := p_{i1}$
        end
        // Calculate how many new_cells will carry mutations
        $mutants$:=Poisson $(new\_cells \cdot \mu_{mut})$
        if $mutants > new\_cells$ **then**
            $mutants := new\_cells$
        end
        $p_{i1} := p_{i1} - new\_cells$
        $p_{i0} := p_{i0} + new\_cells \cdot 2 - mutants$
        //Mutation
        for Each new mutant **do**
            Add a new strain i to the community, with the model parameters of the ancestor and set $p_{i0} := 1$ and $p_{i1} := 0$
            Decide which f_{ik} to mutate by drawing from Bernoulli$\left(\frac{1}{N_{tox}}\right)$ for each f_{ik}; ensure that at least one f_{ik} mutates
            for Each successful draw **do**
                Multiply the chosen f_{ik}, by a factor $x \sim lognormal(0.0, \sigma_m)$
                Re-scale so $\sum_k f_{ik} \leq 1$, if needed
            end
        end
    end
end
return *Populations p_{i0}, p_{i1} for each strain i, including the new ones resulted from mutation.*

Algorithm 4: Replication and mutation, second step of cell division of the IBM described in Alg. 2.

Input: Communities where each strain i is defined by parameters in Tab. 1. Inactive and active sub-populations p_{i0}, p_{i1}.
    Current concentrations T_k of toxic compounds. Death rates m_{ik} and K constant for the Hill function.
for Each community **do**
    for Each strain i, looping over the community in reverse order **do**
        $p_{i0} := p_{i0} - \text{Poisson}\left(p_{i0} \cdot \sum_k (m_{ik} \cdot \frac{T_k^2}{T_k^2 + K^2})\right)$; ensure that $p_{i_0} \geq 0$
        $p_{i1} := p_{i1} - \text{Poisson}\left(p_{i1} \cdot \sum_k (m_{ik} \cdot \frac{T_k^2}{T_k^2 + K^2})\right)$; ensure that $p_{i_1} \geq 0$
        if $p_{i0} + p_{i1} = 0$ **then**
            remove strain i from the community
        end
    end
    Randomly shuffle strains in community
end
return *Populations p_{i0}, p_{i1} for each strain i.*

Algorithm 5: Cell death of the IBM described in Alg. 2.

Input: Strains with model parameters from 2 for each population i.
Input: Mutation parameters: rate μ_{mut}, trait deviation σ_m.
Input: Experimental parameters: Number of toxic compounds N_{tox}, initial concentrations N_0, T_0 of nutrients and toxic compounds, initial population size S_0. Time span for growth $[t_0,\ t_{end}]$

```
for Each community do
    // Growth and competition within one round
    Solve the equations Eq. (6)–Eq. (8) for a time span [t_0, t_end].
    Save the end states S_i(t_end), N_j(t_end), T_k(t_end).
    // Mutations, ODE model
    for Each strain i, with probability μ_mut do
        Copy the species parameters to an empty place in the list of populations
        Choose a f_ik at random by drawing from Bernoulli(1/N_tox) for each k = 1, ..., 10
        For each chosen f_ik, multiply by a factor x_k ~ lognormal(0.0, σ_m)
        Set the inoculum size to S_0
    end
end
return S_i(t_end), N_j(t_end), T_k(t_end), f_ik
```

Algorithm 6: Implementation of population growth, competition and mutations in the ODE model described in the section *Population-level model*. To solve the equations, we use *dopri5* from the SciPy library (35, 36).

Input: Communities with populations S_i and degradation scores D. End states $T_k(t_{end})$.
Input: Experimental parameters: selection bottleneck $\beta = 1/3$, dilution ratio d.

```
Rank the communities by degradation D
Select the top N_β = 7 of communities with the highest ranks.
// Re-populate the new set of tubes
Allocate 1/β new tubes for each selected community
for Each selected community 1, 2, ..., 7 do
    for Each population S_i in the selected community do
        // Dilute the population
        Dilute S_i(t_0) := d · S_i(t_end)
        if S_i(t_0) < 1.0 then
            // The population is extinct
            Set S_i(t_0) := 0.0
            Remove all species parameters from the community
        end
        Copy model parameters and population sizes S_i(t_0) of each strain in the parent communities to each of the 1/β
          offspring communities
    end
end
```

Algorithm 7: Implementation of the propagule selection method for the ODE model.

```
Input: Communities with degradation scores D and strains S_i with total population S_i = p_i0 + p_i1. End-state concentration
       of toxic compounds T_k(t_end).
Input: Experimental parameters: selection bottleneck β = 1/3, dilution ratio d.
Rank the communities by D
Select the top N_β = 7 of communities with the highest ranks
// Re-populate the new set of tubes
Allocate 1/β new tubes for each selected community
for Each selected community 1, 2, ..., 7 do
    for Each strain i in the community do
        // Deactivate cells
        p_i0(t_end) = S_i(t_end)
        p_i1(t_end) = 0
        // Dilute the population, new cells will be inactivated
        p_i0(t_0) :=Poisson(d · S_i(t_end))
        if p_i0(t_0) > S_i(t_end) then
            p_i0(t_0) = S_i(t_end)
        end
        if p_i0(t_0) > 0 then
            Append strain i with population p_0(t_0) to the new tube
        end
        // Delete selected cells to not chose them again
        S_i(t_end) = S_i(t_end) − p_i0(t_0)
    end
end
```

Algorithm 8: Implementation of the propagule selection method for the IBM.

Input: Communities with degradation scores D and strains S_i with total population $S_i = p_{i0} + p_{i1}$. End-state concentration of toxic compounds $T_k(t_{end})$.
Input: Experimental parameters: selection bottleneck $\beta = 1/3$, dilution ratio d.

```
Rank the communities by D
Select the top N_β = 7 communities with the highest ranks, and pool their populations.
// Re-populate the new set of tubes
Allocate 21 new tubes
for Each offspring community 1, 2, ..., 21 do
    for Each strain i in the pool do
        // Deactivate cells
        p_i0(t_end) = S_i(t_end)
        p_i1(t_end) = 0
        // Dilute the population, new cells will be inactivated
        Draw p_i0(t_0) from Poisson (d/N_β · S_i(t_end))
        if p_i0(t_0) > S_i(t_end) then
            p_i0(t_0) = S_i(t_end)
        end
        if p_i0(t_0) > 0 then
            Append strain i with population p_i0(t_0) to the new tube
        end
        // Delete selected cells to not chose them again
        S_i(t_end) = S_i(t_end) − p_i0(t_0)
    end
end
```

Algorithm 9: Implementation of the migrant pool selection method for the IBM

Input: Communities with populations S_i (in the IBM $S_i = p_{i0} + p_{i1}$) with model parameters from (Tab. 1, Tab. 2) and degradation scores D.
Input: Experimental parameters: selection bottleneck $\beta = 1/3$, initial population size S_0 and number of new communities to emigrate species from $N_{emi} = 5$, and immigrate species to $N_{immi} = 5$.

```
// Rank the communities
Rank the communities by the degradation score D
// Update the fossil record with the top communities in this round
for Each selected community 1, 2, ..., 7 do
    for Each species l in the selected community do
        if The record of species l is not yet updated in this round of selection then
            Add all strains i of species l to the fossil record, including the corresponding parameters and population sizes
        end
    end
end
// Propose new communities in proportion to their degradation scores and survival
Follow Alg. 11
// Emigration
Draw N_emi = 5 communities with uniform probability
for Each chosen community 1, ..., N_emi do
    // Find emigrating species
    Number_of_emigrants = 1 + Poisson(0.5)
    Verify that at least one species will remain
    for Each emigrant do
        Choose the emigrant at random, with priority for species that occur in more than one community
        Remove all strains of this species from the community
    end
end
// Immigration
Draw N_immi = 5 communities with uniform probability
for Each chosen community 1, ..., N_immi do
    // Find immigrating species
    Number_of_immigrants = 1 + Poisson(0.5)
    for Each immigrant do
        if There are species that do not feature in any community then
            Choose one of them at random
        end
        else
            Choose a species that is not already in the community with uniform probability
        end
        Take all strains of the species from the species record, add them to the offspring community
        Set the population size to S_0 (approximately S_0 in the IBM, see Alg. 11), in proportion to strain relative abundance
    end
end
```

Algorithm 10: Implementation of the disassembly method.

Input: The subset of selected communities, their degradation scores D. Population sizes S_i of all strains (in the IBM, $S_i = p_{i0} + p_{i1}$).
Input: Parameters: Number of species N_{spc} that were inoculated in the different communities. Inoculum size per species S_0.

```
// Scale degradation scores by species extinctions
for Each selected community 1, 2, ..., 7 do
    // Count the number of surviving species in this community
    Set the extinction counter E = 0
    for Each species l in the community do
        if S_i = 0 for all strains of species l then
            Set E := E + 1
            // Re-introduce species l from the record
            Take all strains of species l from the most recent record
        end
    end
    // Scale the degradation score D by the fraction of surviving species
    Set D̂ := D · (N_spc − E)/N_spc
end
// Calculate a probability distribution based on degradation scores
Set p_n := D̂_n / Σ_m D̂_m for each community n
// Propose new communities randomly in proportion to p_n
for Each offspring community do
    Choose a parental community at random, by the probability distribution p_n
    for Each species l in the parental community do
        Copy the growth parameters and f_ik of all strains i to the offspring community
        // ODE: Set the population size of each species to S_0 in total, in proportion to the relative abundance of the strains.
        // IBM: Sample approximately S_0 cells with replacement as follows:
        for Each strain i (with population S_i) of species l (with population S_l) do
            //Deactivate cells
            p_i0(t_end) = S_i(t_end)
            p_i1(t_end) = 0
            // Draw new population
            Draw p_i0(t_0) from Poisson (S_0 · S_i/S_l)
            if p_i0(t_0) > 0 then
                Add strain i with population p_i0(t_0) to the new community
            end
        end
    end
end
```

Algorithm 11: Method to propose new communities based on their degradation scores from the previous round. Called by Alg. 10