Spatial microbial community dynamics using a continuous species interaction model

Anshuman Swain*[1], Levi Fussell*[2], and William F Fagan[1]

[1]Department of Biology, University of Maryland, College Park, MD, USA

[2]Institute of Perception, Action and Behaviour, University of Edinburgh, Edinburgh, UK

**contributed equally*

Abstract

Comprehending the assembly and maintenance of microbial diversity in natural communities, despite the abundance of antagonistic interactions, is a major problem of interest in biology. A common framework to study the problem is through cyclic dominance games over pairwise interactions. Recent papers incorporating higher-order interactions in these models have successfully explained high diversity of microbes, especially in communities where antibiotic producing, sensitive, and resistant strains co-exist. But most of these models are based on a small number of discrete species, assume a notion of pure cyclic dominance, and focus on low mutation rate regimes, none of which best represents the highly interlinked, quickly evolving and continuous nature of microbial phenotypic space. Here, we present a model of species in a continuous space, with mutual higher order interactions, to examine the assembly and stability of microbial communities. Specifically, we focus on toxin production, vulnerability, and inhibition among the simulated species. We observe intricate interaction between certain parameters that generates highly divergent patterns of diversity and spatial community dynamics. We find that spatial properties are better predicted by species interaction constraints rather than mobility, and that community formation time, mobility, and mutation rate best explain the patterns of diversity.

Keywords: Cyclic dominance, continuous species model, community assembly, higher order interactions, eco-evolutionary dynamics

Significance Statement:

Understanding the assembly and maintenance of diverse microbial communities in nature is a question of great interest to theoretical biologists. Previous works, utilizing evolutionary game theory and other techniques, have explained the role of higher order interactions for the coexistence of diverse microbes in different kinds of environments. But these models are usually based on a small number of discrete species and low//no mutation rate, which is not how many natural microbial communities function. In this work, we explore a new framework which incorporates a continuous species model along with a wide range of mutation rates to comprehend the process of microbial community formation.

1. Introduction

Natural microbial communities possess high species diversity, often much higher than expected from a 'simple' interpretation of species interactions through a Darwinian lens (Reichenbach et al., 2007). This is especially intriguing given the ubiquity of antagonistic interactions among microbes, such as production of antibiotics/toxins against one another (Hibbing et al., 2010; Cordero et al., 2012; Ghoul and Mitri, 2016, García-Bayona and Comstock, 2018). Therefore, explaining the origin and maintenance of microbial species diversity has long been a major question in theoretical biology (Hutchinson, 1961; May, 1972; Shoresh et al., 2008), specifically the complex spatio-temporal co-existence of antibiotic/toxin producing, sensitive and resistant species, which has been well documented in numerous controlled and natural settings (Kerr et al., 2002; Hibbing et al., 2010; Kelsic et al., 2015). This complexity of interactions has prompted much research to understand the impacts on the eco-evolutionary maintenance of diverse microbial communities, including both pure theoretical works (Chesson, 2000; Czárán et al., 2002; Reichenbach et al., 2007; Frey, 2010; Szolnoki et al., 2014) and combined theoretical and empirical approaches (Kerr et al., 2002; Kelsic et al., 2015). For example, even if species interactions are important for maintaining diversity, their random nature can destabilize speciose communities (Hutchinson, 1961; Levin, 1970; McCann, 2000; Gross et al., 2009). Similarly, species richness is inversely related to the strength of random pairwise interactions in stable communities (May 1972), and communities constructed with random pairwise interactions feature an upper bound on diversity (Sinha and Sinha, 2005; Gross et al., 2009; Allesina and Tang, 2015).

To understand the nature of these interactions and their impact on community coexistence, researchers have relied on the framework of cyclic dominance (also termed non-transitive systems or rock-paper-scissors systems) and related techniques in evolutionary game theory

(Czárán et al., 2002; Frey, 2010; Szolnoki et al., 2014; Kelsic et al., 2015). These works have highlighted the importance of species properties such as mobility, and competitiveness (Frey, 2010; Szolnoki et al., 2014). Models with cyclic dominance feature community stabilization, increased co-existence, and maintenance of diversity in spatially structured environments (Durrett and Levin, 1998; Czárán et al., 2002; Kerr et al., 2002). Such effects occur in certain microbial communities (Kerr et al., 2002; Kirkup and Riley, 2004). However, most of these models did not account for intermixed coexistence of antibiotic producing, resistant and sensitive species at very small spatial scales as found in a large number of microbial systems (Green and Bohannan, 2006; Reichenbach et al., 2007; Vetsigian et al., 2011; Raynaud and Nunan, 2014; Ghoul and Mitri, 2016; García-Bayona and Comstock, 2018).

Historically, research on cyclic dominance has focused on pairwise interactions where toxin/antibiotic producing species can inhibit toxin/antibiotic sensitive species. In contrast, species are routinely involved in multi-species interactions in higher-order combinations such that the interactions between two species can be modulated by other species (Bairey et al., 2016; Grilli et al., 2017). One simple example of such a behavior occurs when an antibiotic produced by one species, that inhibits the growth of a competing (sensitive) species, can be attenuated by a third species that can degrade the antibiotic (Perlin et al., 2009; Kelsic et al., 2015; Abrudan et al., 2015). Here, the third species modulates the interaction between the antibiotic-producing and sensitive species, without impacting either of them directly (Bairey et al., 2016). Cyclic dominance-based models with higher order interactions feature increased stability and diversity in both spatially structured and well-mixed communities, and the findings from such models have been verified empirically (Kelsic et al., 2015). Cyclic dominance frameworks successfully explain many properties of microbial systems, especially the maintenance of high levels of diversity, but such feature few avenues for increased biological realism.

Theoretical studies of the diversity-stability relationship have tended to focus on the maintenance of high diversity level. In contrast, studies have seldom investigated the origin of such diversity, and when they do, they emphasize very slow mutation regimes and assume a separation of ecological and evolutionary timescales (Tarnita, 2018). Such assumptions do not work very well for microbial systems, where high mutation rates and extensive horizontal gene transfers are commonplace, creating an overlap of ecological and evolutionary timescales (Tarnita, 2018; Kotil and Vetsigian, 2018). Recent exploration of these fast-evolutionary regimes has uncovered the mechanisms responsible for de novo assembly of diverse microbial communities with higher order interactions, but such results are qualitatively different from what is possible when mutation is rare (Tarnita, 2018; Kotil and Vetsigian, 2018).

Models exploring cyclic dominance typically assume a small number of discrete interacting species (Durrett and Levin, 1998; Czárán et al., 2002; Kerr et al., 2002; Reichenbach et al., 2007; Kelsic et al., 2015; Kotil and Vetsigian, 2018). In contrast, horizontal gene transfer and mutation blur the boundaries among strains and species, reflecting the continuous nature of phenotypes. Incorporation of more than a few discrete species in models of cyclic dominance makes the dynamics extremely complex and hard to interpret systematically (Kelsic et al., 2015).

Moreover, pure cyclic dominance, in which there is a ‘closed’ loop of dominance, is rare in nature (Park et al., 2020). In contrast, rapid evolution routinely leads to model systems featuring semi-cyclic or non-cyclic patterns. Moreover, experimental studies have indicated the lack of pure cyclic interaction patterns in natural microbial communities (Wright and Vetsigian, 2016; Higgins et al., 2017; Friedman et al., 2017), although higher order interactions, such those involving antibiotic producing, sensitive and resistant species, are common (Grilli et al., 2017; Park et al., 2020).

Here, we merge these themes about cyclic dominance in a new theoretical framework to study higher level interactions in microbial communities of toxin/antibiotic producing, sensitive, and resistant species. We use a continuous species parametrization model with a wide range of mutational rates, spatial and species-level properties (see Figure 1 and methods) to uncover patterns of spatial and temporal assembly, notions of stability in these communities, and their diversity dynamics. The parameters are defined at a global level and are divided into two categories – spatial (which define rules and properties of spatial interactions) and species-level (which define the toxin interactions in the species space). The continuous species space generalizes both intra-species relationships of toxin (or antibiotic) production, sensitivity, and resistance that arise randomly in populations through mutation and interspecies relationships arising due to genetic diversity.

We find that species-level properties, such as the constraints on higher level interactions among species, affect the spatial structure of microbial dynamics more than spatial factors like mobility and the effective distance over which individuals kill or inhibit each other.. In contrast, community diversity reflects complex interactions among species-level and spatial parameters, especially mobility and mutation rate, plus community formation time (see Kotil and Vetsigian, 2018). Community formation time (the time it takes to form a stable community) is a good predictor of diversity dynamics, but this time duration cannot be estimated well from other parameters – pointing at emergent complex properties of community assembly.

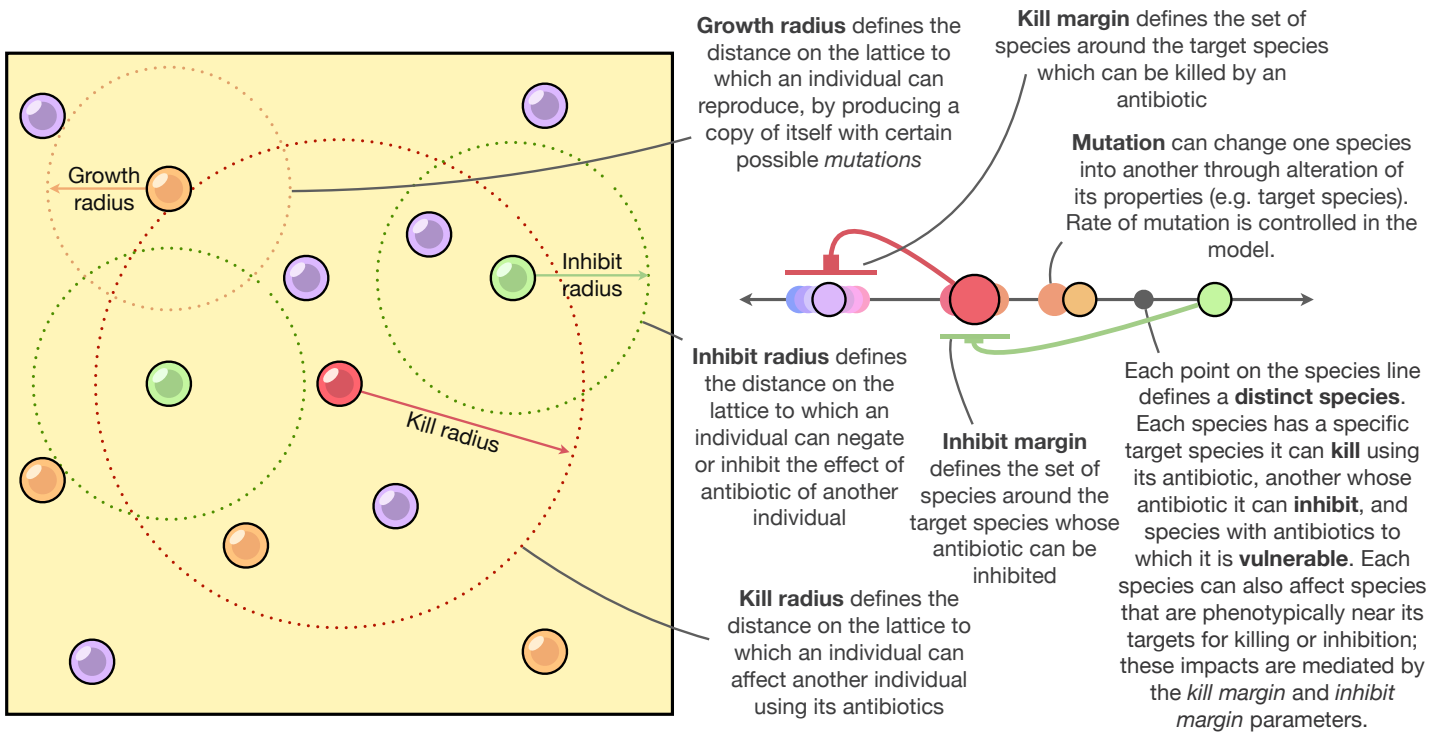


Figure 1: A conceptual representation of the model, with descriptions of all the parameters involved. See Methods for details about implementation.

2. Results

After running 10.49 million simulations across a broad range of parameter combinations, we observed strong differences in spatial heterogeneity among runs. We classified this spatial disturbance behavior into 3 major categories (low, medium, and high) depending only on the level of spatial disturbance that was observed (Figure 1, Methods). These categories had both between and within category differences in the spatio-temporal dynamics of the communities but had similar disturbance regimes within the same categorization (Figure 1). The time that each simulation took to stabilize (i.e., to form a stable community of 'species') is termed as the community formation time (CFT; see Methods and Kotil and Vetsigian, 2018) and the community that is thus assembled is called an eco-evolutionary stable community (ESC).

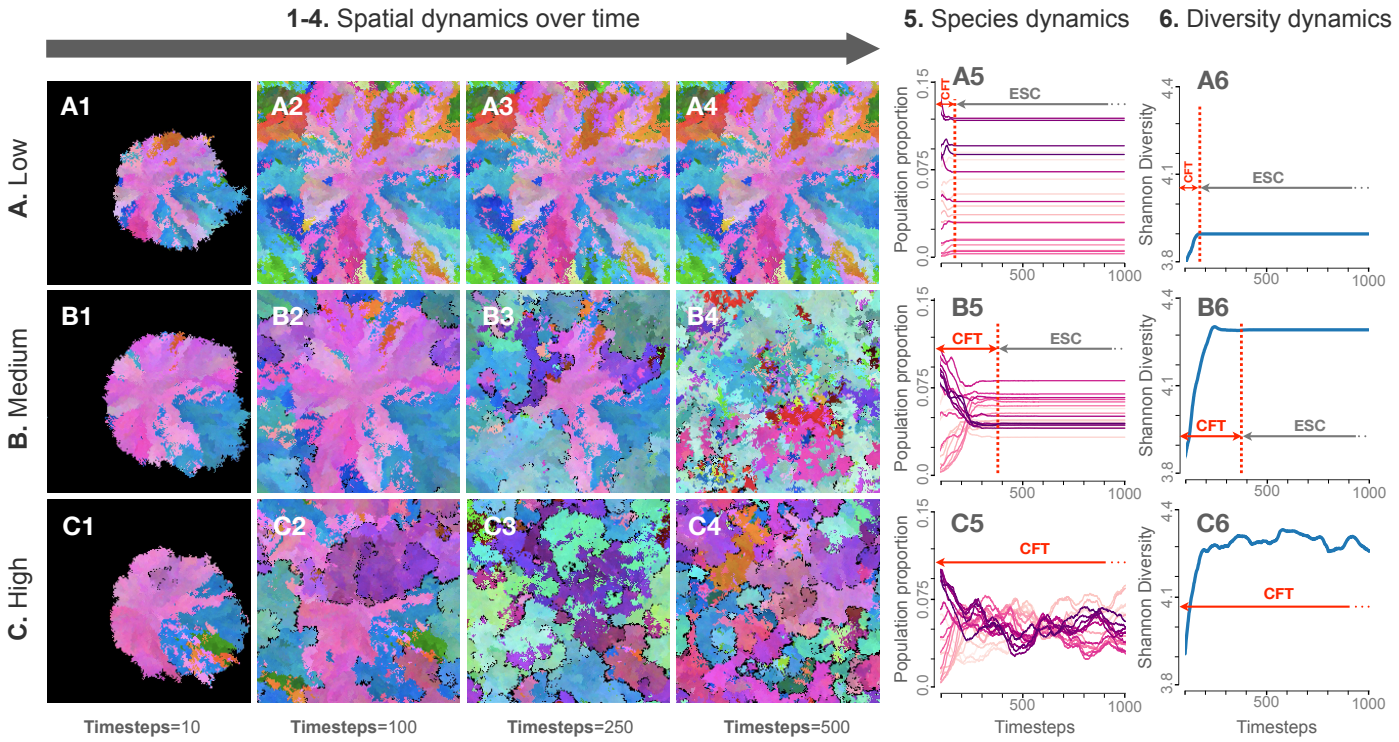


Figure 2: System dynamics in three example conditions of initial conditions pertaining to different regimes of spatial disturbance behavior (Low, Medium and High) (rows A-C). Snapshots of each simulation at different timesteps (50, 100, 250 and 500) are shown in columns 1 through 4. The respective species dynamics (where different colors represent different species 'bins', see Methods) and diversity dynamics are represented in columns 5 and 6. In columns 5 and 6, the time taken for the simulation to stabilize has been denoted with a vertical red dotted line and termed as the Community formation time (CFT) to the left, and to denote the stable community phase, we use the term eco-evolutionary stable community (ESC) in our work.

To understand how parameters relate to the spatial disturbance categories, we performed a random forest classification using the six model parameters (*kill radius, inhibit radius, growth radius, kill margin, inhibit margin* and *mutation size*). The model had an out-of-box error of 10.2% and identified *kill margin* and *inhibit margin* as the best predictors of the disturbance classification (Figure 3A). Note that *kill margin* and *inhibit margin* are global species-specific parameters.

The next most important parameter was the *growth radius*, which is a proxy for mobility in our model, followed by *mutation size*, *kill radius* and *inhibit radius*. These findings showed that except for *growth radius*, other spatial parameters played a minor role in determining the structure of spatial disturbance in the systems denoted by our model.

We performed a random forest regression on the Shannon diversity (SD) mean (i.e., mean Shannon diversity (SD) for the period from timestep=0 to the CFT) and the associated SD variance using the same six simulation parameters. That random forest model explained 68.4% of the variance for the SD mean and 70.2% for the SD variance. M*utation size* and *growth radius* were the two best predictors of both the outcome variables (Figure 3B, 3C). Among the less important parameters that affect SD mean and variance, we can a small distinction that *kill margin* and *kill radius* play a larger role in determining SD mean whereas inhibition processes better predict SD variance (Figure 3B, 3C, S1).

From a naïve interpretation, one would expect *mutation size* to be a good predictor of increasing diversity. However, after overlaying the plot of SD mean versus variance with the respective *mutation size* values, the picture is largely muddled (Figure 3D). We saw even more mixed results on overlaying the growth radius values on the SD mean-variance plot (Figure 3E). Overlaying other less important model parameters on the SD mean-variance plot showed no straightforward pattern at all (Figure S2).

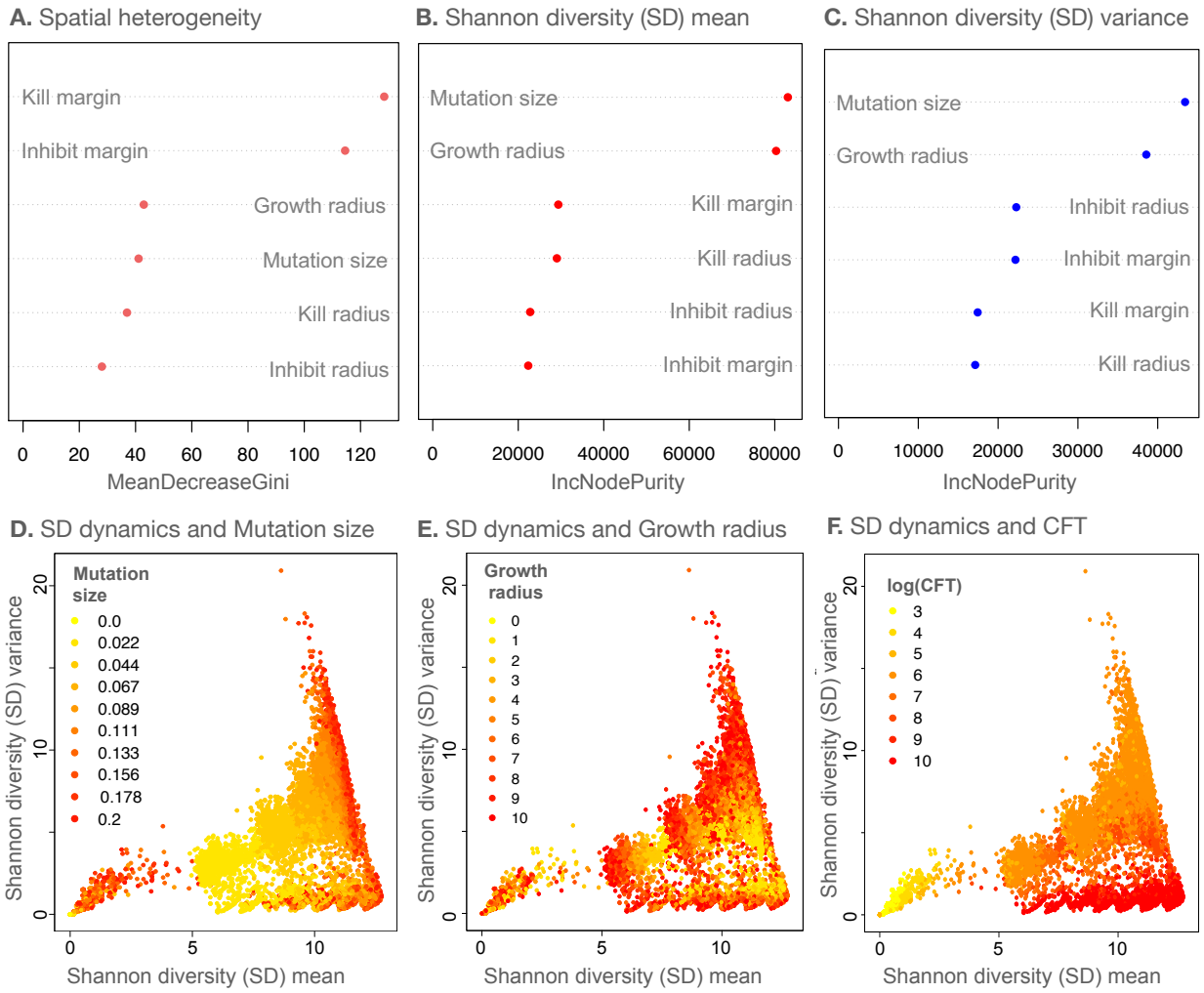
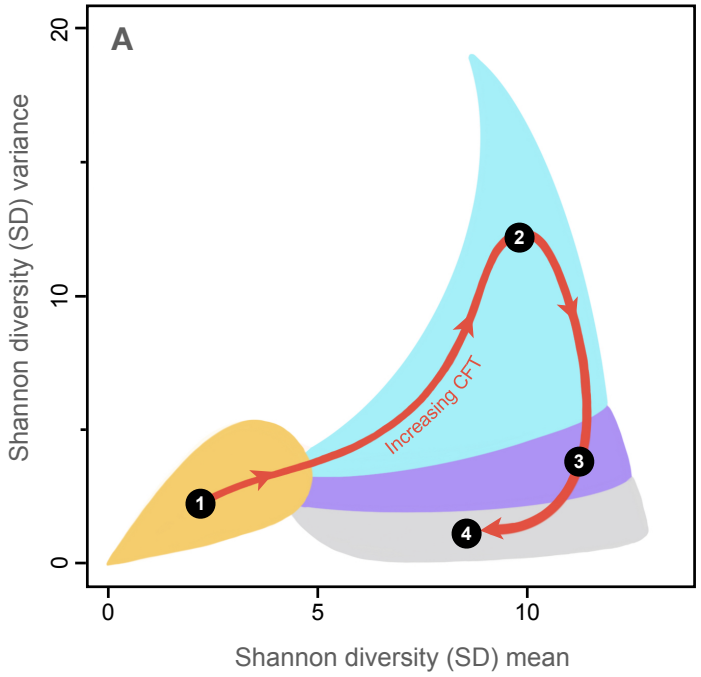


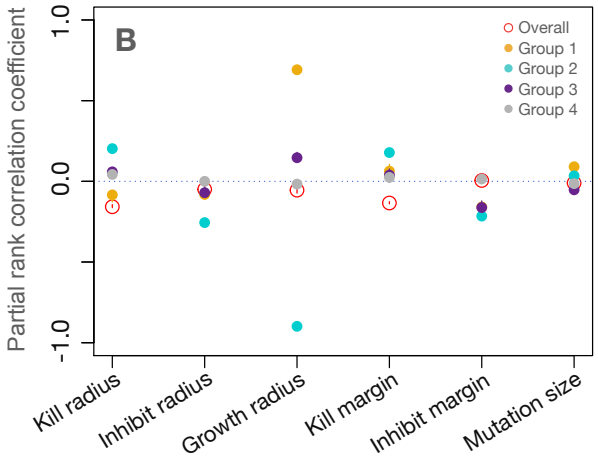
Figure 3: Dependence of model dynamics on different parameters. (A) shows the variable importance score for all the parameters from a random forest classification (with 10,000 trees) on the three categories of spatial disturbance (OOB=10.2%). (B) and (C) denote the variable importance scores from random forest regression models (with 10,000 trees) for Shannon diversity (SD) mean and variance respectively. (D)-(E) show the SD mean-variance plots overlaid with *mutation size*, *growth radius*, and community formation time (CFT) values of the respective simulations. For (A)-(C), IncNodePurity (used in regression random forests) and MeanDecreaseGini (used in classification random forests) refer to how well does a parameter explain the prediction variable in question.

Interestingly, upon overlaying the SD mean-variance plot with the values of the respective CFTs, a stronger pattern emerged. Note the clustering by color gradient of the points due to CFT in SD

mean-variance space (Figure 3F). Although CFT is not a parameter, it is an emergent outcome of the community assembly process and exploring it further gave us insights into the diversity formation process itself.

To formally detect clusters in the SD mean-variance plot with respect to CFT, we used k-means clustering optimized for the number of clusters using multiple methods (Methods) and identified four clusters (Figure 4A, S3). These clusters pertain to (1) Low SD mean and SD variance, (2) High SD mean and SD variance, (3) High SD mean and medium SD variance and, (4) High SD mean and low SD variance, in order of increasing values of CFT.





C	Group 1	Group 2	Group 3	Group 4
Low	0.872			0.128
Medium		0.943	0.057	
High			0.315	0.685

Figure 4: Effects of Community formation time (CFT) on the spatial diversity dynamics. (A) depicts the four optimal clusters (groups) of SD mean-variance space found using k-means clustering (see Figure S3). (B) shows the results from the partial rank correlation coefficient (PRCC) analyses of the four groups and the overall data. (C) shows the proportion of spatial disturbance regimes present in different groups.

On performing random forest regression predicting CFT using the six-parameters for the whole dataset, we found a low explanation of variance (22.3%) with the *kill parameters* (*kill radius,* and *kill margin*) being the strongest predictors. However, upon performing the analysis separately on the four CFT groups, a stronger pattern emerged (Figure S4). For group 1, the regression explained 30.2% of the variance with *growth radius* being the strongest predictor, and other variables playing a smaller part, but nevertheless the second and third best predictors are *kill parameters* (Figure S3A). Group 2 regression explained 90.6% of the variance with again *growth radius* as the strongest predictor with other parameters making almost no contributions (Figure S4B). For group 3 and 4, the regression explains 27.2% and 6.0% of the variance respectively, and therefore with *growth radius* again having most impact on prediction (Figure S3C,D). This pattern of results is unusual because the overall regression did not show any strong dependence on *growth radius* (Figure S3E).

To explore this dilemma further, we performed a partial rank correlation coefficient analysis on this data pertaining to CFT (Figure 4B). We found that the sign of the dependence of *growth radius for CFT* groups varies widely; hence, the random forest regression on the whole dataset misses this signal. We observed that *growth radius* had a strong positive PRCC with CFT in group 1, a very strong negative PRCC with CFT in group 2, a weakly positive PRCC in group 3 and a negligible one in group 4. Other parameters had relatively smaller PRCC values as compared to *growth radius* values for group 1 and 2. *Kill radius* and *kill margin* had overall higher PRCC values than the other parameters and thus were identified by the random forest algorithm (Figure 4B, S4E). Moreover, on calculating PRCC of parameters on SD mean and SD variance, we found that the general values are similar for both SD metrics but what differentiates SD mean and SD variance values the most is *growth radius* (Figure S1).

To link the dependence of diversity dynamics on CFT with the spatial disturbance regimes, we compared what proportion of the three regimes (low, medium and high) fall into the four CFT-classified groups from SD mean-variance space (Figure 4C). We found that the low disturbance regime was most associated with group 1 and, to a much lesser extent, group 4. The medium disturbance regime was primarily dominated by group 2, with a small contribution from group 3. Lastly, the high disturbance regime was dominated by groups 3 and 4. Collectively, these results highlight some important, and novel, connections between spatial patterns and diversity dynamics, especially through transient phenomena (see Hastings, 2001; Hastings et al., 2018).

3. Discussion

A continuous species model with simple rules for interactions based on antibiotic production, inhibition, and vulnerability can produce a wide array of complex dynamics and generate diverse, persistent communities. Despite enormous heterogeneity in behavior, certain general patterns in the assembly and stability of these hypothetical microbial communities stand out.

The diversity patterns seen across time were expressed in the form of mean Shannon diversity (SD) and its associated variance. The phase space of these two metrics shows a non-linear relationship between the metrics. Random forest regressions revealed that *mutation size* and *growth radius* were the best predictors of mean diversity and its variance (Figure 3B,C). In other words, the mutation rate and mobility of species dominate the outcome of how diversity dynamics will play out for any given system equivalent to our model.

Mobility controls diversity dynamics in numerous previous studies (Reichenbach et al., 2007; Frey, 2010; Szolonki et al., 2014; Kelsic et al., 2015). In particular, small amounts of mobility enhance diversity, whereas large amounts of mobility jeopardize it (Reichenbach et al., 2007;

Szolonki et al., 2014). Reduced mobility results in spatially structured populations where co-existence is easier to maintain, as seen in experimental works on competing *E.* coli strains (Kerr et al., 2002, 2006; Weber et al., 2014). Recent work has shown that even in cases of high mobility (i.e., well-mixed communities), one can observe co-existence if higher order interactions, such as antibiotic production and degradation, are considered (Kelsic et al., 2015). This is the case for *in vivo* experiments with bacterial colonies in the intestines of co-caged mice; these systems can be considered locally well-mixed and have high levels of coexistence (Kirkup and Riley, 2004).

High mutation rates, and processes such as horizontal gene transfer (HGT), have been long known experimentally to affect diversity in microbial populations, especially during the initial phase of community assembly (Giraud et al., 2001; Datta and Pan, 2002). But cyclic dominance models have tended to overlook the role of mutations and HGT in coexistence. Such studies have focused on detailed understanding of a small number of discrete species under a low mutation regime (Tarnita, 2018). Recently, studies have focused on understanding the effects of high mutational regimes in community assembly, which better represent microbial systems (Kotil and Vetsigian, 2018), and can generate frequent non-cyclic interactions (Park et al., 2020). Our efforts have considered an array of mutational regimes and characterized the importance of non-cyclic patterns of interactions in the assembly and maintenance of microbial communities.

M*utation size* and *growth radius* (a proxy for mobility) were good predictors of SD mean and SD variance but yielded no discernable patterns in the phase space (Figure 3D,E). In contrast, community formation time (CFT) partitioned the space cleanly into four regions with distinct properties (Figure 3F, 4A). Communities that have low diversity (SD mean) and small fluctuations in it (i.e., low SD variance) (group 1) assemble quickly, and that mobility enhances diversity and co-existence in this scenario (Figure 4A,B). Mobility affects the diversity and coexistence negatively for communities with high diversity (SD mean) and larger fluctuations (high SD variance) (group 2) and these communities assemble slower than group 1. For communities that

take a long time to stabilize (groups 3 and 4), diversity is generally high (SD mean), but it fluctuates less and mobility has a negligible effect.

Mutation rates were weakly correlated with CFT, as observed previously observed (Kotil and Vetsigian, 2018), but exhibited substantial variation in trends due to interactions with other model parameters. For low SD mean and variance (group 1) and high SD mean and variance (group 2), *mutation size* positively affects CFT (Figure 4B). This implies that communities with a high rate of mutations take longer to stabilize, but only when the diversity outcome corresponds to these two groups (Figure 4B). If SD mean is high but SD variance is low, mutation affects CFT negatively, i.e., higher mutation helps communities converge faster (Figure 4B).

Community spatial structure is also important for coexistence (Szolonki et al., 2014). Rather than focusing on the final spatial structure, we examined the degree of spatial disturbance that occurred throughout the assembly process, which provides insight on the extent of mixing/migration that might occur in these communities. The categorization of spatial disturbance regimes was best predicted by species-level parameters (*kill margin* and *inhibit margin*), rather than by other spatial parameters in our model. This is somewhat surprising, but higher order interactions are known to influence spatial patterns of coexistence (Kelsic et al., 2015). Importantly, we found a strong correspondence between the regimes of spatial disturbance and the groups produced by CFTs. The communities that formed quickly (group 1) tended to have low spatial disturbance, but in some cases can have very high disturbances (Figure 4C). Communities that took more time to stabilize and ended up with high diversity through large overall diversity fluctuations (high SD mean and SD variance, group 2) have primarily a medium level of spatial disturbance. The communities that took longest to stabilize were characterized by high diversity (SD mean) and lower average fluctuations in diversity (SD variance) (groups 3 and 4) but have larger spatial disturbances. All these observations point towards a complex interplay of mutation

and mobility, which affects the assembly time of a community (CFT), and in turn controls the diversity of the assembled community.

Collectively, these results demonstrate the rich spatiotemporal dynamics that are possible when large numbers of microbial species with limited but heterogeneous rules for aggressive, inhibitory, and sensitive interactions live in a common space. Although these findings are focused primarily on antagonistic interactions between microbes, such dynamics may also be relevant for coral communities and other spatially structured systems featuring diverse types of interspecific interactions (see Karlson and Jackson, 1981; Precoda et al., 2017; Laird and Schamp, 2018). In addition, these results point towards the important yet often-overlooked role that transient dynamics play in the behavior and structure of ecological systems (see Hastings et al., 2018). Overall, studying the impact of higher order intransitive interactions on community stability, and diversity should provide insights on the complex dynamics of co-existence and the role of transient phenomena in community assembly of natural ecosystems.

4. Model

4.1 Model Description

Here we outline an algorithmic presentation of the model and describe our methods for technical implementation. The term *simulation* and *model* will be used interchangeably. The parameters of the model and their corresponding symbols are detailed in Table 1. The simulation took place on a 200 x 200 square lattice with toroidal boundary conditions. Each cell of the lattice is stateful, represented as a five-dimensional vector $\mathbb{R}^5$ where all elements of the vector are a value in the set $\{-1, [0,1]\}$. -1 represents a cell that is 'empty' in the corresponding element and $[0,1]$ is a bounded real value. The five elements of each cell's five-vector are, in order, the individual's species vulnerability value (S_v), the individual's species inhibition value (S_i), the individual's

species antibiotic production value (S_a), the antibiotic occupying that cell space (C_a), and the inhibitor occupying that cell space (C_i). The first three elements define the individual in that space, hence these three values will simultaneously be empty or non-empty and we can refer to this state with a single empty or non-empty species value (S_s). The last two parameters describe what antibiotic and inhibitor are actually present in that cell. Note that this setup means that no more than one individual, antibiotic, or inhibitor can occupy one cell at a given time, and hence the discretization of our simulation is considered the smallest unit of diffusion.

At fixed-interval time steps, the state of the five-vector of each cell is updated synchronously, for four sub-steps defining a single update step. First, the antibiotic release sub-step: each cell which has a non-empty species value (S_s) is randomly assigned the inhibit value (C_i) of one other cell that is within its antibiotic radius (K_r). Note that rewrites are possible during this sub-step, so the order of cell updates is randomized. Second, the inhibition step: for every cell with a non-empty species value, the inhibition radius (I_r) around that cell defines an area that is checked for other cells which contain an antibiotic value (C_a) where $I_v - \epsilon_i \leq C_a \leq I_v + \epsilon_i$. If this is the case, that cell's antibiotic value (C_a) is assigned a value of -1 (the 'empty' value). The fourth sub-step is the kill step: every cell is iterated over, and its antibiotic value (C_a) and individual vulnerability value (S_v) are compared. If $C_a - \epsilon_a \leq S_v \leq C_a + \epsilon_a$, then that cell's first three parameters are assigned a value of -1 (the individual in that cell is killed). The final sub-step is the growth step: every cell is iterated over, and if it contains a non-empty individual value (S_s), another cell that is within that cell's growth radius (G_r) is randomly selected. If the selected cell has an empty individual value (S_s), the three individual (species) values of the current cell are copied to the empty values of the selected cell, which might be modified according to a mutation size parameter (μ). This modification is a non-correlated additive perturbation of the organism's three species-level parameters by a uniform value sampled from $U(-\mu, \mu)$ and the modified/perturbed value is then kept in the range [0,1] by adding or subtracting 1 if it is outside this range. This ensures that all

species have equal likelihood of interactions with other species by putting the species parameters on a three-dimensional toroid. From a theoretical perspective, this growth can be seen as a model of growth probability, where the probability that an individual will reproduce is equal to $E(C)/F(C)$ where $E(C)$ is the number of empty cells within radius G_r of the current cell C and $F(C) = floor(2\pi\, G_r^2)$ is equal to the total number of cells within radius G_r of cell C.

Parameter	Type	Definition
Species antibiotic production value (S_a)	Species-level, Species-specific	Defines what species are vulnerable to the antibiotic that this individual produces
Species vulnerability value (S_v)	Species-level, Species-specific	Defines what antibiotics (and hence species) this individual is vulnerable to
Species inhibition value (S_i)	Species-level, Species-specific	Defines the species whose antibiotics this individual can inhibit
Kill radius (K_r)	Spatial, Global	Maximum spatial distance of diffusion of the antibiotics produced by an individual
Inhibit radius (I_r)	Spatial, Global	Maximum spatial distance of diffusion of the inhibitors to antibiotics of produced by an individual
Growth radius (G_r)	Spatial, Global	Maximum spatial distance at which an individual can reproduce by producing a copy of itself with certain possible mutations
Kill margin (ϵ_a)	Species-level, Global	The total span centered around the specified *Species antibiotic production value* (S_a), which can be affected/killed by the antibiotic of a given species
Inhibit margin (ϵ_i)	Species-level, Global	The total span centered around the specified *Species inhibition value* (S_i), whose antibiotic can be inhibited by a given species
Mutation size (μ)	Species-level, Global	Maximum value by which the species-specific parameters can be additively altered during reproduction/copying

Table 1: Brief description of model parameters

We implemented the simulation in *C++* using the *SFML* graphics library for rendering. It is possible to parallelize the update sub-steps of the cells due to the synchronous nature of the model, but we chose to use a simpler sequential algorithm that looped through each cell one at a time. We did, however, parallelize the running of multiple simulations when sweeping parameter spaces. This allowed us to reduce the total computation time required by two orders of magnitude. Each simulation was run for 2000 timesteps or until stability/stagnation (also termed the Community Formation Time (CFT)). Stagnation/stability was determined by comparing the state of all cells between two consecutive time steps and if it remained unchanged for a total of five-time steps, the simulation was terminated. This helped to quickly remove a large number of simulations that terminated early due to poor parameter settings. Because the simulation was so large, collecting per-cell data across many parameters was not practical. Alternatively, we chose to compute population-level statistics online during the simulation using Shannon Diversity mean and variance (over time) and then from the analysis of this lower-granularity data, we selected individual parameter settings to explore population-level statistics. Further, we also built a qualitative-based classification application to allow quick classification of simulations, assigning them to a class based on their spatial heterogeneity/disturbance dynamics. Simulations were assigned to one of three classes according to how the spatial dynamics evolved.

To track population-level statistics, it was necessary to discretize the continuous species space. l-discretization (or l-binning) assigned individuals to the same species identity if the three parameters of those individuals had the same integer values after performing a binning transformation on them like so: $\lfloor\frac{S_v}{l}\rfloor$, $\lfloor\frac{S_i}{l}\rfloor$, $\lfloor\frac{S_a}{l}\rfloor$. When recording the population-level data we stored the number of individuals per species every time step, using a standard 0.05-discretization. We did not explore this parameter further and leave that for future work.

For every parameter setting, we repeated the simulation ten times, seeding the pseudo-random number generator with a unique value for each repetition. Every simulation began by assigning every cell vector the all-empty state $(-1, -1, -1, -1, -1)$ except the center cell which is assigned a random value in the range $[-1, 1]$ for the first three elements of its vector. This initial cell represented a randomly created individual. One potential issue is that if the center cell was assigned a random value such that it was vulnerable to its own antibiotic, it would kill itself and the simulation would terminate immediately. Although this was possible, it was rarely the case and did not significantly affect our results due to our use of multiple runs per parameter setting.

4.2 Model implementation and data analysis

We implemented the model in *C++* and performed data analysis with *Python 3.7* and the *R* statistical language. We ran a total of 10.49 million runs of our model with various parameter initializations. For these runs, we obtained a manual-classification of the spatial disturbance (low, medium and high) based on the size of the spatial disturbance. The simulations which stabilized really quickly (less than 20 time-steps) with no further changes in spatial patterns and the ones which took longer than 20 time-steps and but had more-or less stabilized by then, except for certain point fluctuations in the grid (only), were categorized as 'low' regime (Figure 2 A1). The ones which, after 20 time-steps, had not stabilized but had small waves of diversity replacements which diminished over time were termed as 'medium' disturbance (and all of them stabilized before 500 time-steps) (Figure 2 B1). The rest did neither stabilize even after 500 time-steps, nor had reduced amplitude of waves of diversity replacements, and were categorized as 'large' disturbance regime (Figure 2 C1).

We also calculated the mean Shannon diversity over time (until it reached stability, i.e. Community formation time (CFT)), and the Shannon diversity variance over the same time period. These

three variables (disturbance classification, SD mean, and SD variance) were our outcomes of interest.

To understand how parameters affect diversity dynamics, we determined the partial rank correlation coefficient (PRCC) of various parameters with respect to the outcomes of interest, using the *sensitivity* package (Iooss et al., 2020) in R. We also performed Random Forest (RF) regressions and classification, using the *randomForest* package (Liaw and Wiener, 2002) in R, to identify which parameters are the strongest predictors of the patterns in predicting the outcomes of interest. We optimized the number of parameters available for splitting at each tree node in the RF using out-of-box error (OOB) (Liaw and Wiener, 2002). Please note that the overlay figures (Figure 3 D-F, Figure S2) have been plotted with pruned samples for better visualization.

For clustering values on the SD mean versus SD variance space with respect to CFT, we used k-means clustering in *R* and optimized for the number of clusters using the elbow method, the D index, and the Hubert indices using the package *NbClust* (Charrad et al., 2014). For each obtained cluster in the Shannon diversity space, we performed a random forest analysis to calculate differences in variable importance scores across these clusters. For each cluster, we also performed PRCC and random forest analyses, in order to delineate the effects that individual parameters had on the different subsets of the diversity space.

5. Acknowledgements

NSF DMS-1853465 (to WFF) and the University of Maryland supported this work.

6. Code and Data Availability

All code and required data to replicate our results are available in the following repository: https://github.com/levifussell/MicroEvo.

7. References

1. Abrudan, M.I., Smakman, F., Grimbergen, A.J., Westhoff, S., Miller, E.L., Van Wezel, G.P. and Rozen, D.E., 2015. Socially mediated induction and suppression of antibiosis during bacterial coexistence. Proceedings of the National Academy of Sciences, 112(35), pp.11054-11059.
2. Allesina, S. & Tang, S. The stability–complexity relationship at age 40: a random matrix perspective. Popul. Ecol. 57, 63–75 (2015).
3. Chesson, P., 2000. Mechanisms of maintenance of species diversity. Annual review of Ecology and Systematics, 31(1), pp.343-366.
4. Cordero, O.X., Wildschutte, H., Kirkup, B., Proehl, S., Ngo, L., Hussain, F., Le Roux, F., Mincer, T. and Polz, M.F., 2012. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. Science, 337(6099), pp.1228-1231.
5. Czárán, T.L., Hoekstra, R.F. and Pagie, L., 2002. Chemical warfare between microbes promotes biodiversity. Proceedings of the National Academy of Sciences, 99(2), pp.786-790.
6. Dutta, C. and Pan, A., 2002. Horizontal gene transfer and bacterial diversity. Journal of biosciences, 27(1), pp.27-33.
7. Durrett, R. and Levin, S., 1998. Spatial aspects of interspecific competition. Theoretical population biology, 53(1), pp.30-43.
8. Friedman, J., Higgins, L.M. and Gore, J., 2017. Community structure follows simple assembly rules in microbial microcosms. Nature ecology & evolution, 1(5), pp.1-7.

9. García-Bayona, L. and Comstock, L.E., 2018. Bacterial antagonism in host-associated microbial communities. Science, 361(6408).
10. Ghoul, M. and Mitri, S., 2016. The ecology and evolution of microbial competition. Trends in microbiology, 24(10), pp.833-845.
11. Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M. and Taddei, F., 2001. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. science, 291(5513), pp.2606-2608.
12. Green, J. and Bohannan, B.J., 2006. Spatial scaling of microbial biodiversity. Trends in ecology & evolution, 21(9), pp.501-507.
13. Grilli, J., Barabás, G., Michalska-Smith, M.J. and Allesina, S., 2017. Higher-order interactions stabilize dynamics in competitive network models. Nature, 548(7666), pp.210-213.
14. Hastings, A., 2001. Transient dynamics and persistence of ecological systems. Ecology Letters, 4(3), pp.215-220.
15. Hastings, A., Abbott, K.C., Cuddington, K., Francis, T., Gellner, G., Lai, Y.C., Morozov, A., Petrovskii, S., Scranton, K. and Zeeman, M.L., 2018. Transient phenomena in ecology. Science, 361(6406).
16. Hibbing, M.E., Fuqua, C., Parsek, M.R. and Peterson, S.B., 2010. Bacterial competition: surviving and thriving in the microbial jungle. Nature reviews microbiology, 8(1), pp.15-25.
17. Higgins, L.M., Friedman, J., Shen, H. and Gore, J., 2017. Co-occurring soil bacteria exhibit a robust competitive hierarchy and lack of non-transitive interactions. BioRxiv, p.175737.
18. Hutchinson, G.E., 1961. The paradox of the plankton. The American Naturalist, 95(882), pp.137-145.
19. Iooss, B., Da Veiga, S., Janon, A., Pujol, G., et al., 2020. sensitivity: Global Sensitivity Analysis of Model Outputs. R package version 1.20.0. https://CRAN.R-project.org/package=sensitivity

20. Karlson, R.H. and Jackson, J.B., 1981. Competitive networks and community structure: a simulation study. Ecology, 62(3), pp.670-678.
21. Kelsic, E.D., Zhao, J., Vetsigian, K. and Kishony, R., 2015. Counteraction of antibiotic production and degradation stabilizes microbial communities. Nature, 521(7553), pp.516-519.
22. Kerr, B., Riley, M.A., Feldman, M.W. and Bohannan, B.J., 2002. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. Nature, 418(6894), pp.171-174.
23. Kerr, B., Neuhauser, C., Bohannan, B.J. and Dean, A.M., 2006. Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. Nature, 442(7098), pp.75-78.
24. Kirkup, B.C. and Riley, M.A., 2004. Antibiotic-mediated antagonism leads to a bacterial game of rock–paper–scissors in vivo. Nature, 428(6981), pp.412-414.
25. Kotil, S.E. and Vetsigian, K., 2018. Emergence of evolutionarily stable communities through eco-evolutionary tunnelling. Nature ecology & evolution, 2(10), pp.1644-1653.
26. Laird, R.A. and Schamp, B.S., 2018. Calculating competitive intransitivity: Computational challenges. The American Naturalist, 191(4), pp.547-552.
27. Levin, S.A., 1970. Community equilibria and stability, and an extension of the competitive exclusion principle. The American Naturalist, 104(939), pp.413-423.
28. Liaw, A. and Wiener, M., 2002. Classification and Regression by randomForest. R News 2(3), 18--22.
29. May, R.M., 1972. Will a large complex system be stable?. Nature, 238(5364), pp.413-414.
30. McCann, K.S., 2000. The diversity–stability debate. Nature, 405(6783), pp.228-233.
31. Perlin, M.H., Clark, D.R., McKenzie, C., Patel, H., Jackson, N., Kormanik, C., Powell, C., Bajorek, A., Myers, D.A., Dugatkin, L.A. and Atlas, R.M., 2009. Protection of Salmonella by ampicillin-resistant Escherichia coli in the presence of otherwise lethal drug

515 concentrations. Proceedings of the Royal Society B: Biological Sciences, 276(1674),
516 pp.3759-3768.
517 32. Precoda, K., Allen, A.P., Grant, L. and Madin, J.S., 2017. Using traits to assess
518 nontransitivity of interactions among coral species. The American Naturalist, 190(3),
519 pp.420-429.
520 33. Raynaud, X. and Nunan, N., 2014. Spatial ecology of bacteria at the microscale in soil.
521 PLoS One, 9(1), p.e87217.
522 34. Reichenbach, T., Mobilia, M. and Frey, E., 2007. Mobility promotes and jeopardizes
523 biodiversity in rock–paper–scissors games. Nature, 448(7157), pp.1046-1049.
524 35. Shoresh, N., Hegreness, M. and Kishony, R., 2008. Evolution exacerbates the paradox of
525 the plankton. Proceedings of the National Academy of Sciences, 105(34), pp.12365-
526 12369.
527 36. Sinha, S. and Sinha, S., 2005. Evidence of universality for the May-Wigner stability
528 theorem for random networks with local dynamics. Physical Review E, 71(2), p.020902.
529 37. Tarnita, C.E., 2018. Fast evolution unlocks forbidden communities. Nature ecology &
530 evolution, 2(10), pp.1525-1526.
531 38. Vetsigian, K., Jajoo, R. and Kishony, R., 2011. Structure and evolution of Streptomyces
532 interaction networks in soil and in silico. PLoS Biol, 9(10), p.e1001184.
533 39. Weber, M.F., Poxleitner, G., Hebisch, E., Frey, E. and Opitz, M., 2014. Chemical warfare
534 and survival strategies in bacterial range expansions. Journal of The Royal Society
535 Interface, 11(96), p.20140172.
536 40. Wright, E.S. and Vetsigian, K.H., 2016. Inhibitory interactions promote frequent bistability
537 among competing bacteria. Nature communications, 7(1), pp.1-7.
538
539

Supplementary Figures

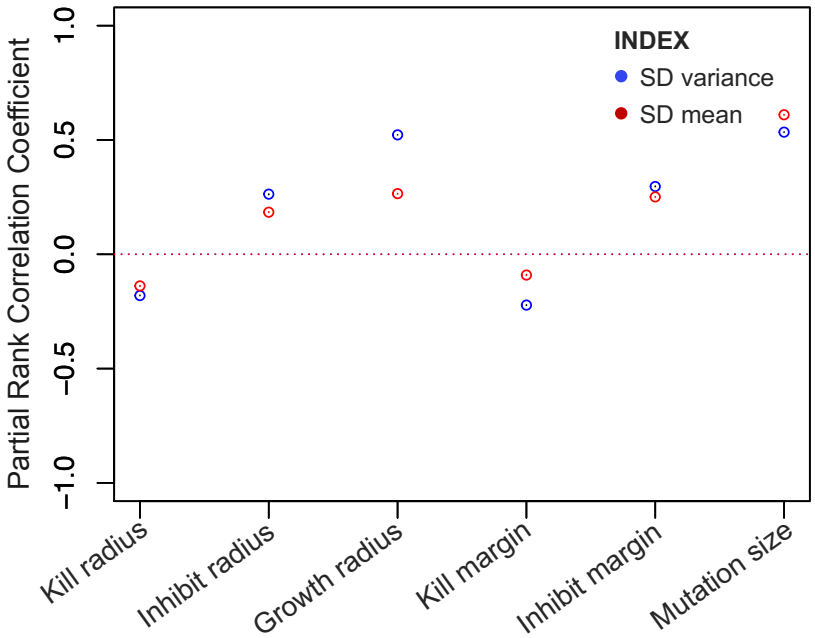


Figure S1: Partial rank correlation coefficients (PRCC) of various parameters in relation to the Shannon Diversity (SD) mean and variance. Please note that different parameters have similar values of PRCC for both the diversity metrics, and the greatest difference in the trends of the two metrics was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

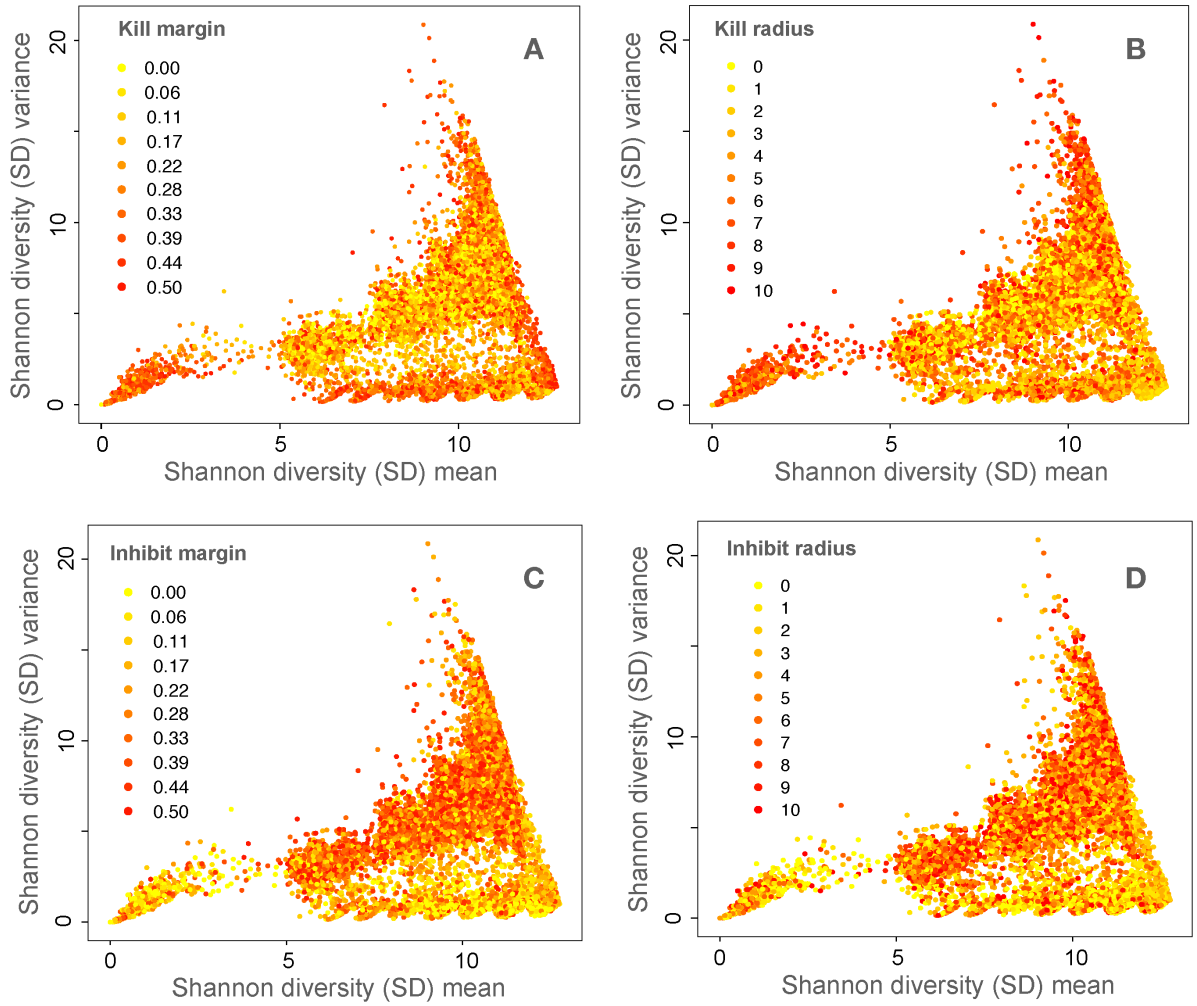


Figure S2: Overlaying values of parameters, (A) *Kill margin,* (B) *Kill radius*, (C) *Inhibit margin*, and (D) *Inhibit radius,* over the respective values on Shannon Diversity (SD) mean and SD variance space. No obvious patterns are discernable in how these parameters affect the diversity space as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

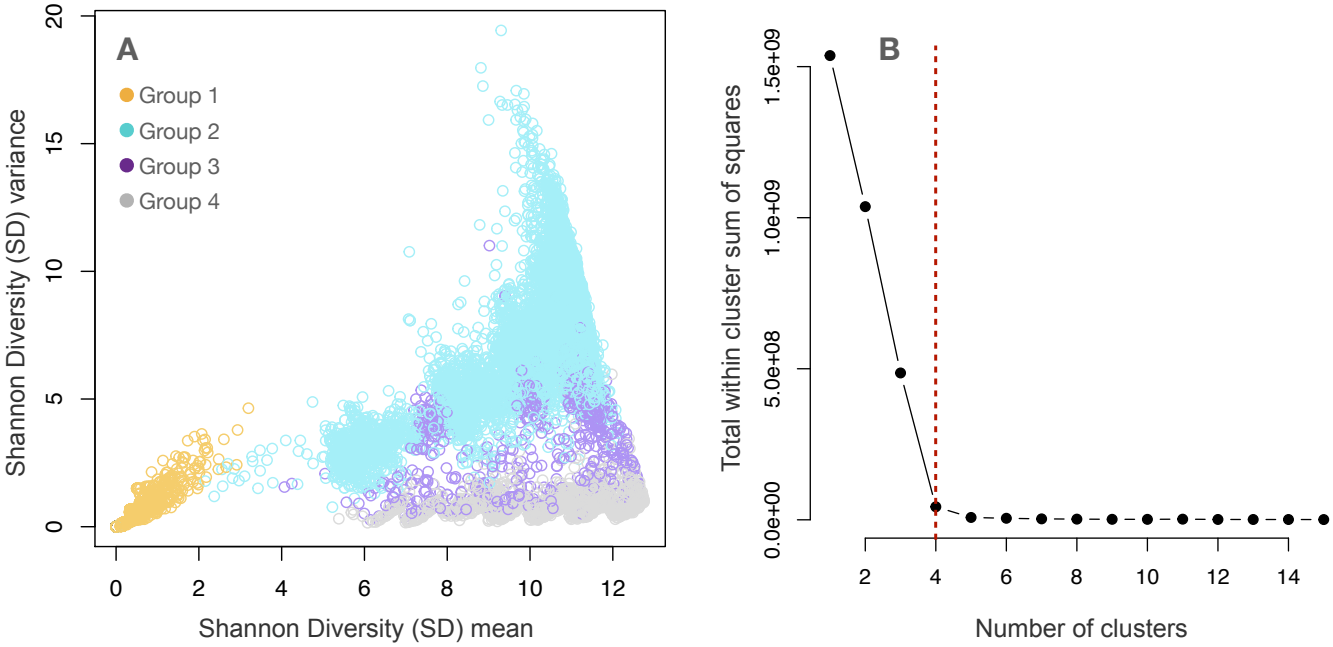


Figure S3: (A) k-means clustering of the Shannon Diversity (SD) mean-variance space by the community formation time (CFT) value. (B) shows the optimal number of clusters using the elbow method. We obtained similar results when optimizing clusters using the D index and the Hubert indices.

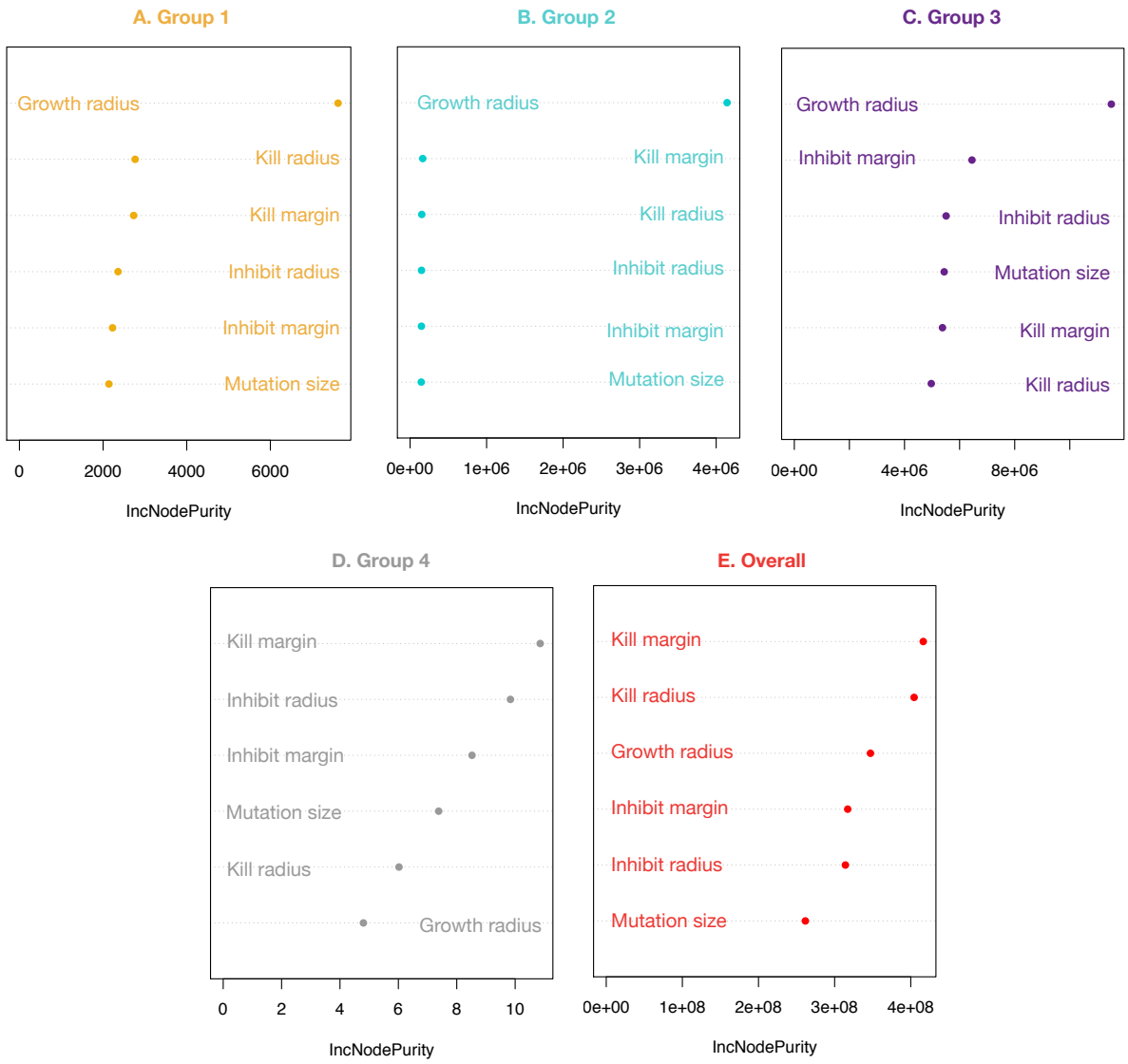


Figure S4: Random forest regressions (10,000 trees each with optimization for splits using Out of box (OOB) error) for community formation time (CFT) for various groups clustered using k-means. Variance explained for- Group 1: 30.2%, Group 2: 90.6%, Group 3: 27.3%, Group 4: 6.0%, Overall: 22.3%. Note: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