nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-019-0952-9

In the format provided by the authors and unedited.

Uncovering the rules of microbial community invasions

Jean C. C. Vila[1,2,3]*, Matt L. Jones[1], Matishalin Patel[1,4], Tom Bell[1] and James Rosindell[1]

[1]Silwood Park Campus, Department of Life Sciences, Imperial College London, Ascot, UK. [2]Microbial Sciences Institute, West Campus, Yale University, West Haven, CT, USA. [3]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [4]Department of Zoology, University of Oxford, Oxford, UK. *e-mail: jeanccvila@gmail.com

Supplementary Materials for Uncovering the Rules of Microbial Community Invasion

Jean C C Vila, Matt L Jones, Matishalin Patel, Tom Bell, and James Rosindell

This file contains Supplementary Methods, Supplementary Table 1 and Supplementary Figures 1-7.

Supplementary Methods

Source community initial conditions

Simulations are initiated from a monomorphic community consisting of *R* individuals. Each individual is arbitrarily assigned a fitness category of 100 and the same unique genotype identifier. Simulations are run for 4*R* generations by which point the distribution of fitness categories, genotype richness and fitness category variance have all reached a dynamic equilibrium, see supplementary figure 6 for an example. 4*R* generations represent an extremely liberal burn-in time for the source community.

Selection algorithm

The probability that any particular individual will reproduce is proportional to its fitness weight, w, which reflects an additive model of selection where $w = 1 + sc$. At each time step, an individual's exact probability of reproducing is equal to its fitness weight divided by the sum of fitness weights of all individuals in the community. This means that an individual's probability of reproducing changes at each time step. Repeatedly re-deriving the probability of replacement for all individuals in a community would be computationally inefficient so instead we implement selection using an equivalent rejection algorithm that only needs to keep track of the maximum genotype fitness weight of the community.

The algorithm works as follows:

1. Sample a random number *x* from a uniform distribution between 0 and 1.
2. Chose an individual *A* at random from the community according to a uniform distribution.
3. If (the fitness weight of individual *A*)/(max fitness weight across the community) is larger than *x*, this individual reproduces and we continue the simulation. Otherwise go back to step 2.

Mutations and the distribution of mutational effects

Newly born individuals are mutated with probability μ. If a mutation does not occur, the new individual inherits its parent's fitness category and genotype. If a mutation does occur, the new individual is assigned a new genotype and it inherits it's parent fitness category shifted by randomly sampled value that in principle can be taken from any defined distribution. For the simulations of invasion outlined in the main text, we have follow previous work and adopted an equal discrete distribution with 0.5 probability of +1 and 0.5 probability of -1; i.e. new mutations have an equal probability of increasing or decreasing the fitness category (*c*) by 1. To qualitatively investigate the effect of other choices on our main results, we also report additional simulations of the source community that we conducted using a number of other distributions of mutational effects. Specifically, we ran simulations of the source community using 4 alternative distributions of mutational effects:

1. An asymmetric discrete distribution with 0.1 probability of +1 and 0.9 probability of - 1
2. A uniform distribution between +1 and -1;
3. A standard normal distribution;
4. An inverted gamma distribution with shape parameter 8 and scale parameter 0.1, shifted by 0.5.

The inverted gamma distribution is motivated by empirically observed distributions of fitness effects; the parameters were selected so that the shape of distribution reflects distributions

that have been reported in the literature. (Perfeito *et al.*, 2007, Eyre-Walker and Keightley, 2007).

Our analyses of distributions of mutational effects should not be taken as a comprehensive exploration of the effect of the different fitness distributions on community level eco-evolutionary dynamics (which is beyond the scope of this paper). However, we are reassured to find that distribution chosen has little effect on the qualitative relationship between model parameters and the community level properties that we focus on in the main text (Supplementary Fig. 7).

Assembly of invader and resident communities.

The invader community is assembled by sampling (I) individuals from the source community containing (*R*) individuals. If (*I*) is less than or equal to (*R*) this is done simply by sampling I individuals from the source community without replacement using the Fisher-Yates shuffle. If $I > R$ then additional community members are generated by sampling the rest of the invader community from the source community with replacement. This would be analogous to a rapid expansion in community size during which neutrality is imposed. The resident community is the same size as the source community and so will have the same composition as the source community. After the resident and invader communities have been assembled, each undergoes independent acclimatization, but for the same period of generations (*G*) before the invasion takes place.

Linear regressions used to estimate 'effects' on internal variables

At multiple points throughout this paper we have quantified i) how the internal variables that emerge from simulations depend on the parameters used for the simulation and ii) how those internal variables may depend on other internal variables that differ stochastically across replicate runs. For the example in main text figure 3, we quantify how both propagule pressure (a model parameter) and the number of invading genotypes (an internal variable) impact invasion success.

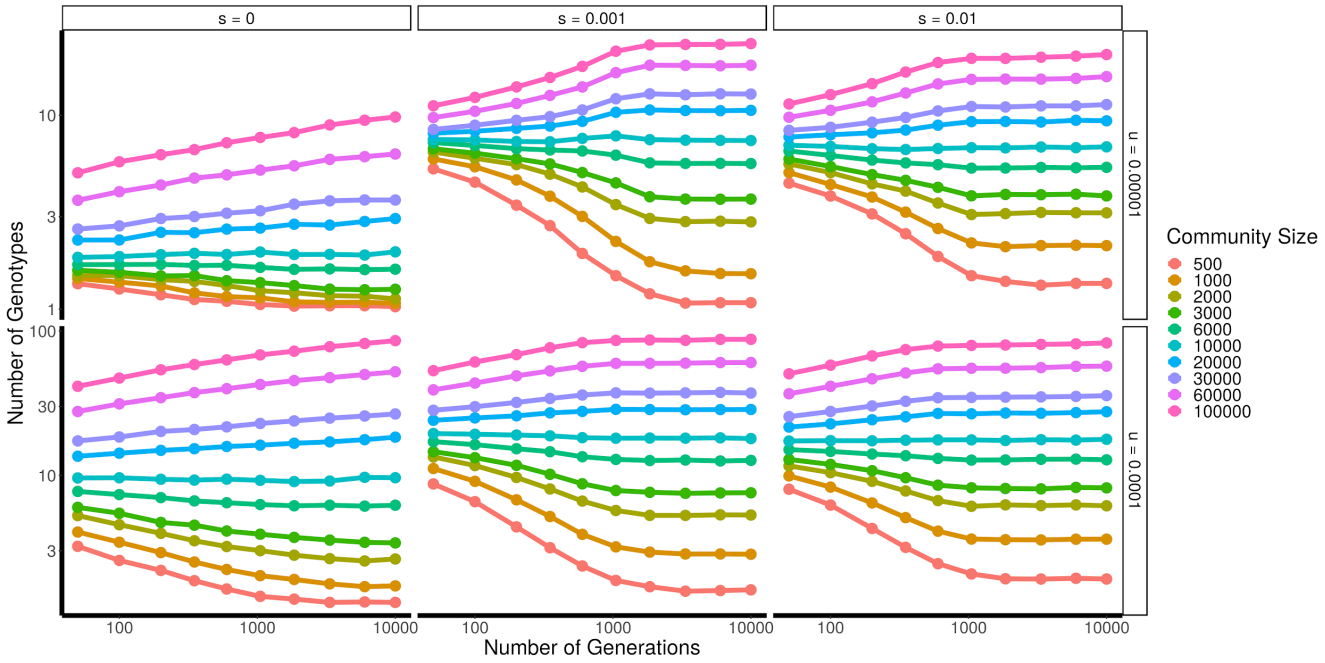
We use the slope of an ordinary least squares regression as a consistent metric for ‘effect size’. When a free parameter such as propagule pressure is the subject of interest, the fit is conducted over all replicate simulations that were carried out using multiple discrete values of the parameter of interest, but where other parameters were kept constant. For example, the fit for propagule pressure (*p*) would be over all simulations that had the same unique combination of selection strength, mutation rate, community size and generation time. The fit would be thus conducted over simulations that used 10 discrete values of *p* (1, 2, 3, 5, 10, 20, 30, 50, 100, 200) with 100 replicates for each value. When the parameter is an ‘internal parameter’ such as the number of invading genotypes, we fit the regression to all 100 replicate simulations for each unique parameter combination.

Software and hardware details

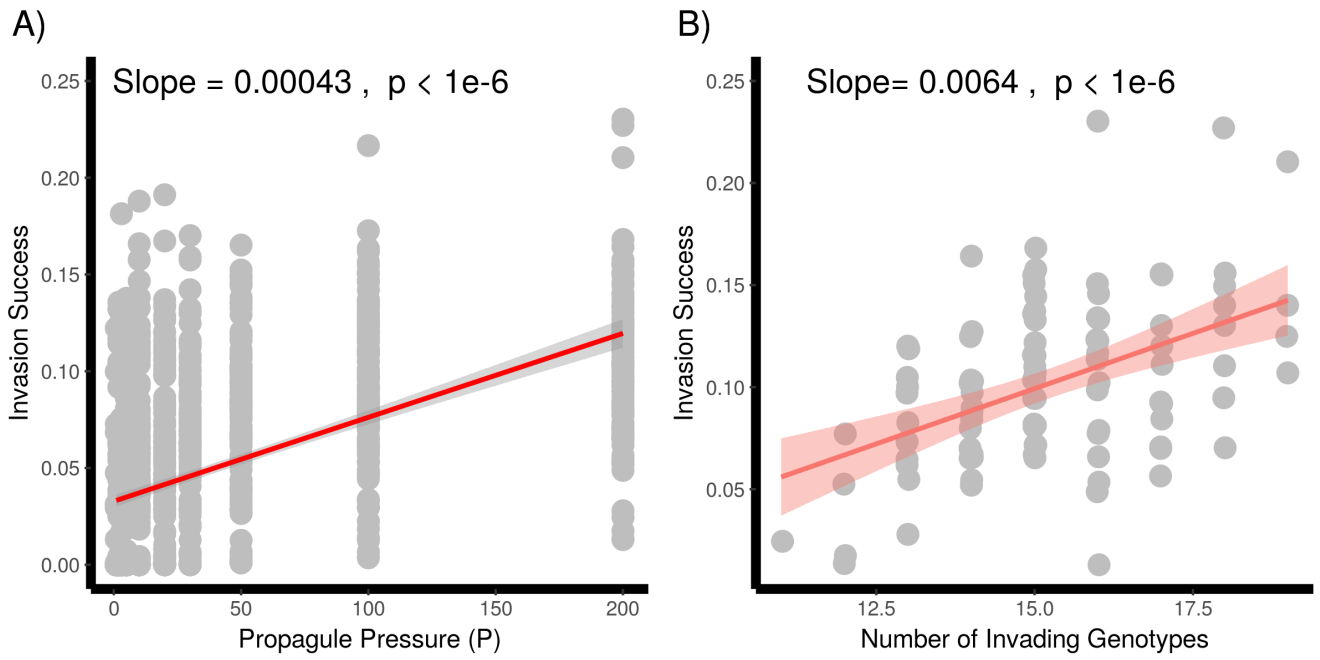
We carried out 4,500,000 Simulations of invasion using C++. Invasions were bundled into 450 groups of 10,000, each group utilized 3 days of CPU time via high throughput computing. Statistical analyses, data-processing and data visualization were carried out locally using R version 3.4.4.

Supplementary Table 1: full set of parameter values used for simulations presented in the main body of the paper.

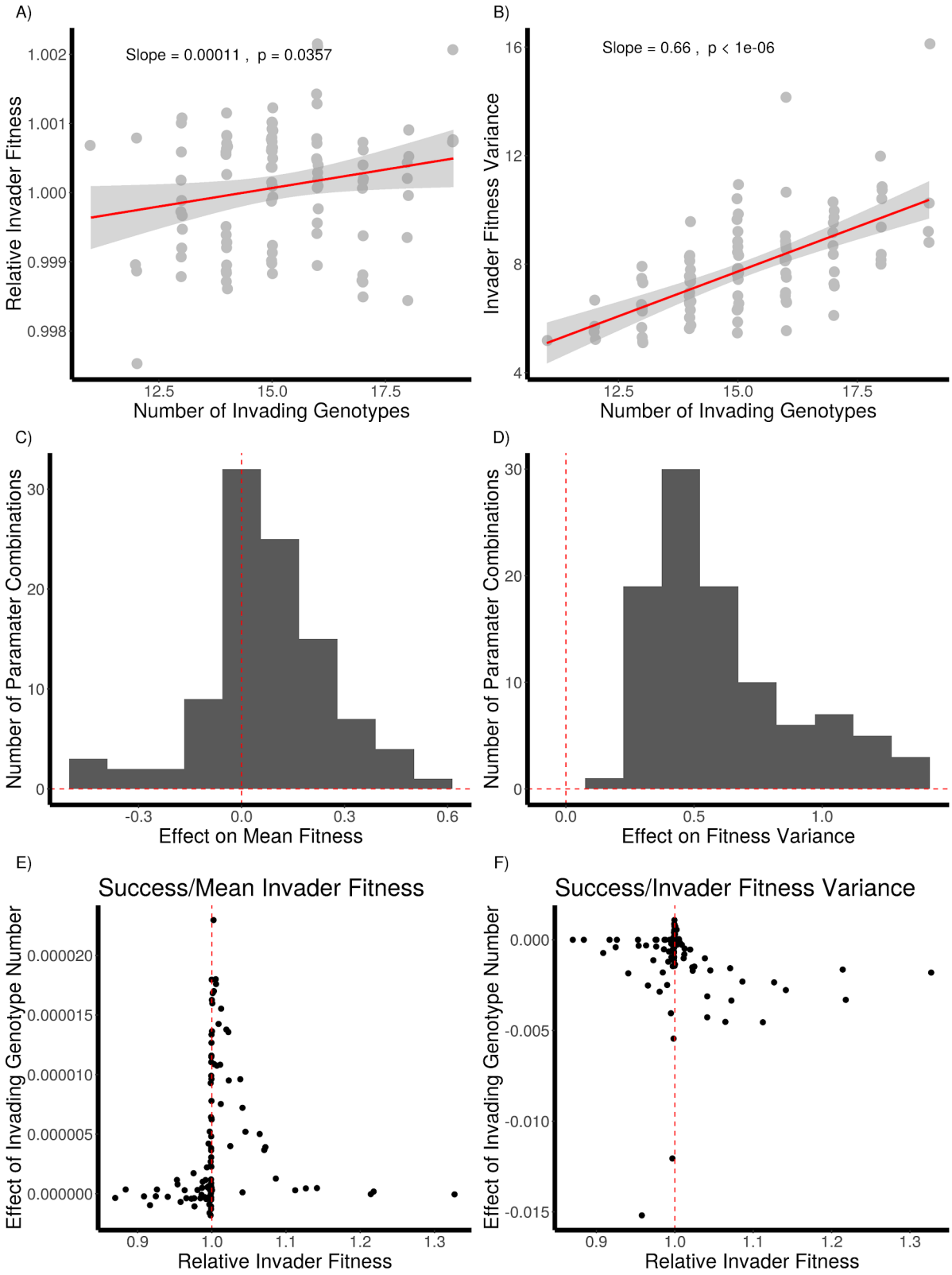
Parameter	**Values**
Invader community size (I) (measured in numbers of individual organisms).	500, 1000 2000, 3000, 6000, 10000, 20000, 30000, 60000, 100000
Resident community size (R) (measured in numbers of individual organisms). Note that the source community always had the same size as the resident community in our simulations.	1000, 5000, 10000, 50000, 100000
Pre-invasion generations (G) (measured in generations where one generation is the amount of time during which every individual is expected to reproduce or die once).	50, 100, 200, 350, 600, 1050, 1850, 3350, 6000, 10000
Propagule pressure (*p)* (measured as total number of individual organisms from the invader community invading the resident community).	1, 2, 3, 5, 10, 20, 30, 50, 100, 200
Mutation rate (μ) (per individual, per birth probability of mutation).	0, 0.00001, 0.0001
Selection strength (s) (fitness weights of individuals are all of form *1+ns* where *n* is a trait of the genotype).	0, 0.001, 0.01



Supplementary Figure 1: effect of model parameters (mutation rate (μ), selection strength (s), invader community size (I), number of generations (G)) on genotypic richness after the invading community is assembled.

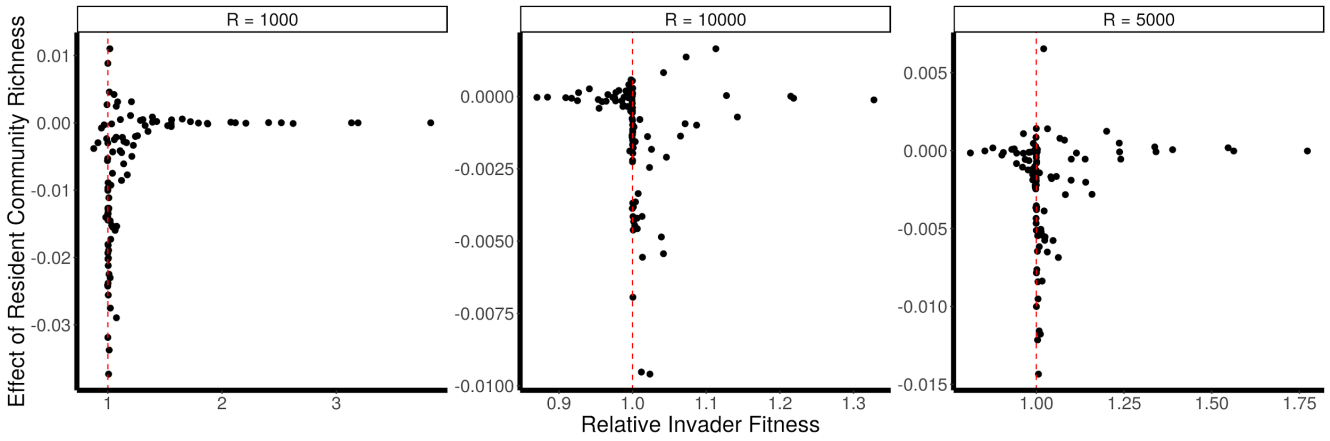


Supplementary Figure 2: example of linear regressions used to calculate the 'effects' shown in main text figure 4b and 4d. In panel a, invasion success is plotted as a function of propagule pressure $(s = 0.01, \mu = 0.0001, R = 10000, I = 100000, G = 100)$. In panel b, invasion success is plotted as a function of total propagule richness $(s = 0.01, \mu = 0.0001, R = 10000, I = 100000, G = 100, p = 200)$. The data plotted here corresponds to the red lines in main text figures 4a and 4c.

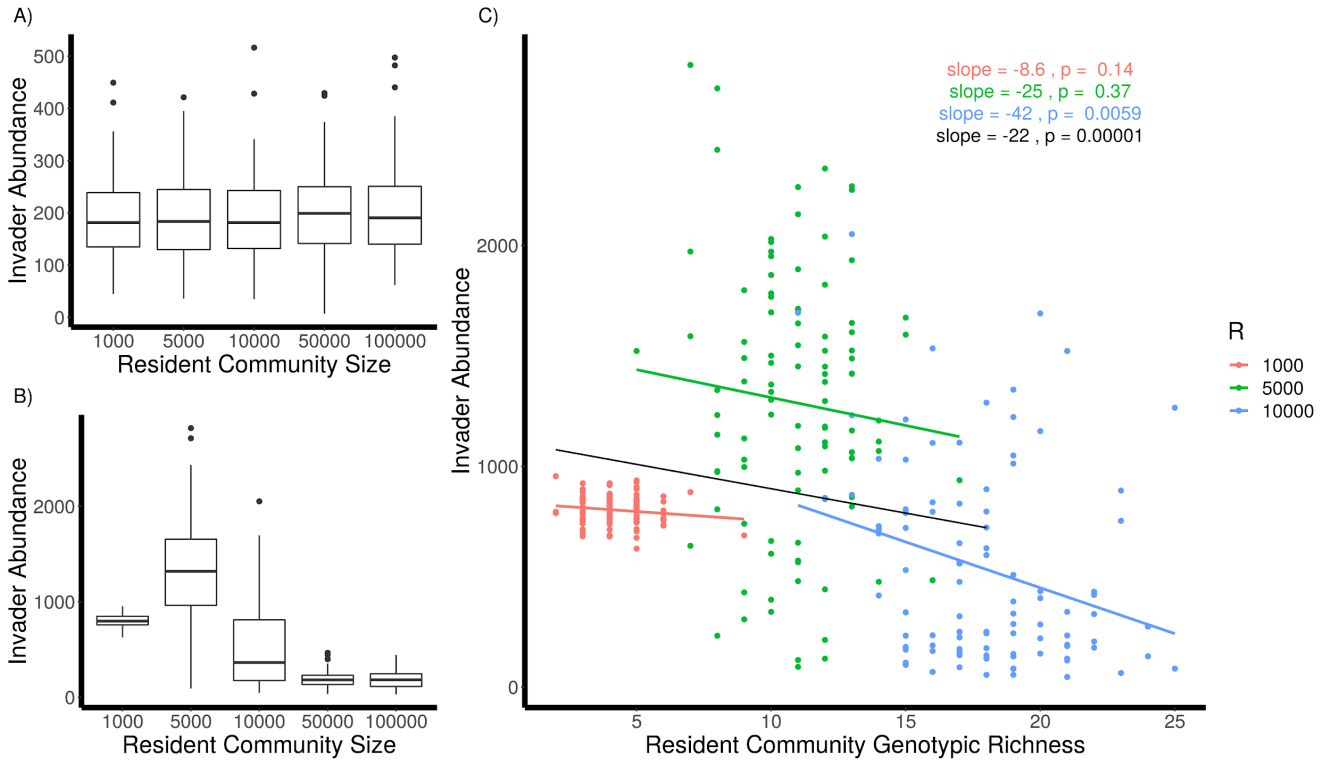


Supplementary Figure 3: evidence that the relationship between number of invading genotypes and invasion success outlined in main text figure 4d is due to increased fitness variance rather than

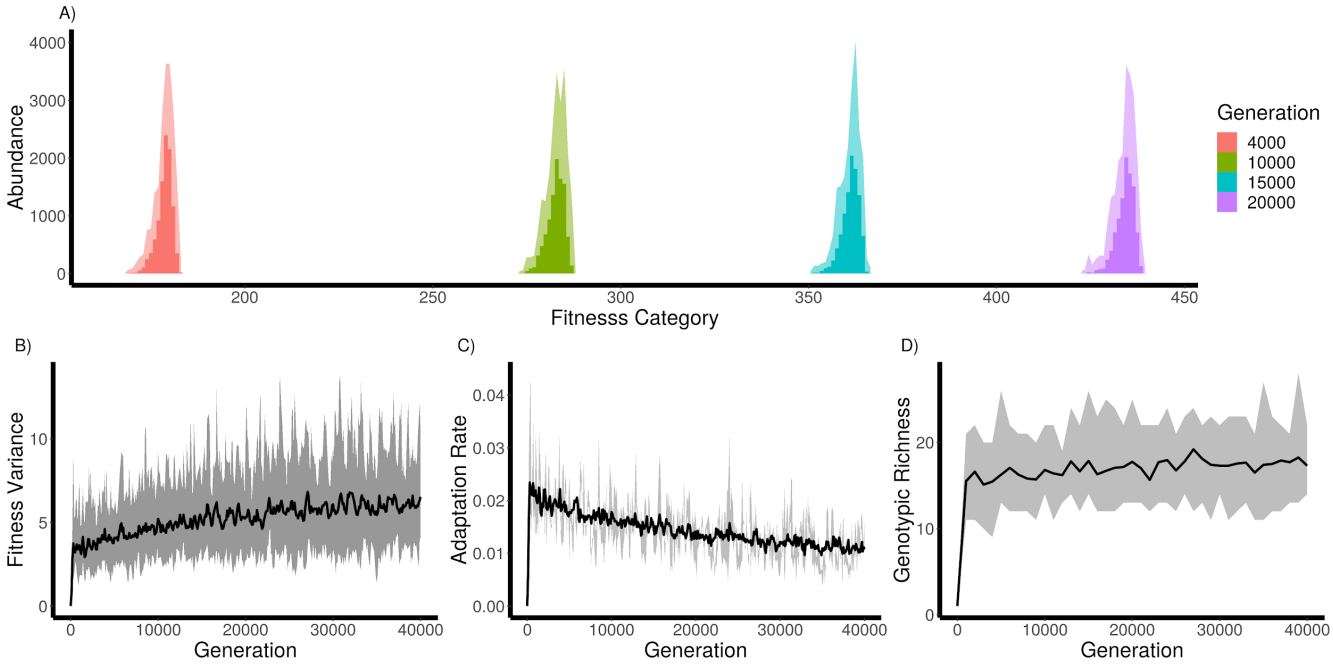
increased fitness means. For the set simulations presented in figure 4c and supplementary figure 2b, increasing the number of invading genotypes increases the variance in invader fitness (b), but has a negligible (though weakly significant) effect on the relative fitness of invaders (a). More generally, the distribution of regression slopes (c) between number of genotypes and mean invader fitness suggests that increasing the number of invading genotypes does not generally increase the fitness of invaders. In contrast, for all parameter combinations, increased invader richness is strongly correlated with increases invader fitness variance (d). Finally, we show that the relationship between genotype number and invasion success shown in main text figure 3d persists even when invader mean fitness is accounted for (by dividing invasion success by mean fitness in the regression). This does not hold true when mean fitness variance is accounted for (by dividing invasion success by fitness variance in the regression), suggesting that the observed dynamics are largely driven by variance.



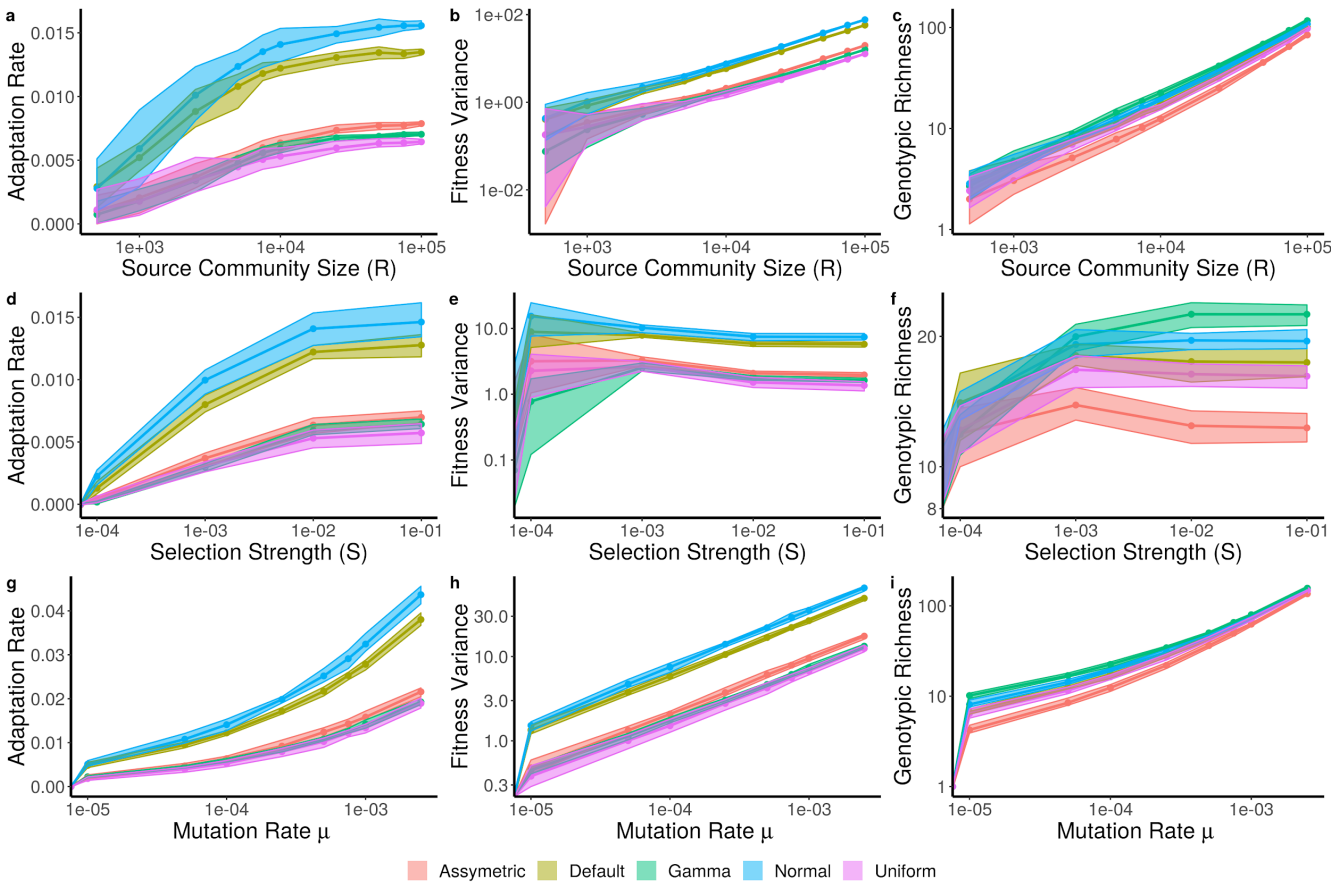
Supplementary Figure 4: independently of community size (R), more diverse communities are more resistant to invasion in the same 'neutral zone'. Effect of resident community richness on invasion success is quantified as the slope of a linear regression between resident community richness and invasion success. Each point represents 100 simulations where the combination of parameters was kept constant. Each data point corresponds to a unique combination of invader community size (I) and number of generations (G) as in main text figure 4(b, d). The other simulation parameters were $s = 0.01$, $\mu = 0.0001$ and $p = 200$.



Supplementary Figure 5: reproduction of main text figure 5, except with invasion success measured as invader abundance after 40 generations (rather than frequency of invaders). In the neutral model (a), resident community size does not affect invader abundance, as invader abundance fluctuates neutrally around the propagule pressure ($p = 200$). The nearly neutral model predicts that invader abundance generally decreases with increasing community size (b) and diversity (c). However, when community size is very small, resident community size acts as an upper bound on invader abundance. This can result in invader abundance increasing as community size and diversity increases, see resident communities of size 1000 and 5000 in panels b and c.



Supplementary Figure 6: dynamics of adaptation in the source community for parameters $R = 10000$, $\mu = 0.0001$ and $s = 0.01$. Panel a shows the distribution of fitness categories at 4 time points. Panel b shows the fitness category variance through time. Panel c shows the adaptation rate through time (calculated over 100 generation intervals). Panel d shows the genotypic richness through time. Results are displayed as a mean across 20 independent simulations and the shaded region is bounded by maximum and minimum values. By ~2*R* Generations the Communities have reached a dynamic equilibrium in which the community-level fitness distribution forms a ‘traveling wave’ that has constant shape and increases in fitness with an approximately constant speed.



Supplementary Figure 7: simulations of the source community with quantified adaptation rate, fitness variance and genotypic richness of the community after 4*R* generations. These simulations were conducted under 5 distinct distributions of mutational effect (see supplementary methods). The default distribution corresponds to the distribution used throughout the rest of the main text. Panels a-f show that the relationship between community level properties and model parameters (such as community size or selection strength) are robust to the choice of distribution. The values of the relevant parameters are $R = 10000$, $\mu = 0.0001$ and $s = 0.01$ except when they vary along the x axis in the plots.

References

1. Eyre-Walker, Adam, and Peter D. Keightley. 2007. “The Distribution of Fitness Effects of New Mutations.” Nature Reviews. Genetics 8 (8): 610–18.
2. Perfeito, Lília, Lisete Fernandes, Catarina Mota, and Isabel Gordo. 2007. “Adaptive Mutations in Bacteria: High Rate and Small Effects.” Science 317 (5839): 813–15.