Supplementary information outline

1) **Methods:**
    a. **Library generation**
    b. **Experimental conditions**
    c. **DNA extraction, amplification, and sequencing**
    d. **Analysis pipeline**
    e. **Validation, controls, and noise floor**
    f. **Replicates**
2) **Models**
    a. **Neutral model**
    b. **Selective model**
    c. **Transition from selective to neutral**
    d. **Maximizing effective fitness**
3) **Data and comparisons**
    a. **Species vs. time plots**
    b. **Muller plots**
    c. **Residence time**
    d. **Rank abundance plots**
    e. **Species relative abundance plots**
4) **Stochasticity**
    a. **Sample preparation and sequencing noise**
    b. **Biological noise**
    c. **Small bottlenecks**
5) **Distribution and nature of selective advantage**
6) **Key to experimental data**
7) **References**

www.pnas.org/cgi/doi/10.1073/pnas.1808118115

1a) Library generation

We established a barcoded library of cells as follows. Random genetic barcodes of the form NNNNACNNNNATNNNNGTNNNNCGNNNN were introduced upstream of a chloramphenicol resistance cassette by a PCR reaction which used the chloramphenicol resistance cassette on pSB1C3 as template.

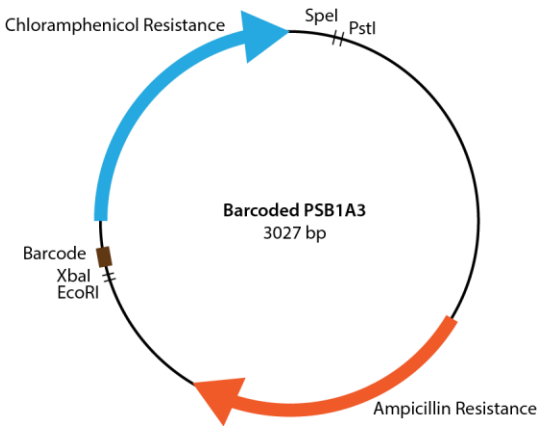
Primers:

5’ - GACTGGTCTAGAGNNNNACNNNNATNNNNGTNNNNCGNNNNGGTCATTACTGGATCTATCAACAGG - 3'

5’ - ATCAGTCTGCAGCGGCCGCTACTAGTATGGTTTCTTAGACGTCAGGTG - 3'

NNN = Species barcode

This construct was cloned using XbaI and PstI restriction digestion into backbone vector pSB1A3 which contains a pUC19-derived pMB1 replication origin, an ampicillin resistance cassette, and previously harbored the CcdB suicide gene[1] (BBa_P1010) in the cloning site, resulting in the construct in Supplementary figure 1. See the Registry of Standard Biological Parts for more information on these constituent parts.



Supplementary figure 1. Barcoded plasmid construct.

Ligated products were transformed into chemically competent TOP10 *E. coli* cells (ThermoFisher, genotype: F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G) by heat shock and plated on LB agar plates containing chloramphenicol (35 µg/mL) and ampicillin (100 µg/mL). Colonies were picked, grown separately in LB broth supplemented with chloramphenicol (35 µg/mL) and ampicillin (100 µg/mL) banked as glycerol freezer stocks at -80 ^{o}C, and the barcode region of the plasmid from each clone was Sanger sequenced.

Based on the Sanger sequencing results, clones were chosen which carried plasmids with barcodes that met the following criteria: 1) a single barcode was present in the clone 2) the sequence of the barcode and flanking regions matched the expected sequence 3) The barcode was at least a Hamming distance of 3 away from any other barcode chosen.

456 clones which met these criteria were grown separately to saturation and combined in equal volume ratio to create the initial community.

This barcoding and cloning strategy was devised to maximize the number of successful barcode insertions and ensure that barcodes remained associated with the cells as reliable indicators of abundance over time. Including known bases in the barcode region as opposed to a continuous string of random barcode bases helped prevent unintentional restriction digest of the barcode during cloning which could occur in a longer random sequence and might result in an inadvertently truncated barcode. Including the CcdB gene in the source vector prevented vector religation from a single restriction digest cut which, when paired with a random mutation causing chloramphenicol resistance, could otherwise have allowed growth of a strain with no barcode. To limit the opportunity for spontaneous mutations in the background genome, which could confer resistance and allow plasmid loss, our construct contained resistance genes to two different antibiotics.

We found the Sanger sequencing validation step important for a few reasons. First, we found a substantial fraction (~25%) of clones sequenced had gotten two different barcodes during transformation. We could not improve this ratio below 10%, even when titrating the input DNA concentration of the transformation down to the minimum amount required to get any transformants out of a transformation. Having two barcodes in a cell could have caused incorrect inferences about interactions between cell types in an experiment, so measures should be taken to prevent double transformation when using barcoded plasmids. A smaller fraction of cells had sequences which did not match the expected sequence form by having, for instance, an extra base in the random barcode section, or a mutation at the restriction digest site. These may have introduced biases in amplification or during analysis that would result in an incorrect measurement of abundance. Performing Sanger sequencing to establish the library allowed us to discard these undesirably labeled cells. Finally, having pools of known barcodes also allowed us to run controls (Supplementary section 1e) which would not have been possible without knowledge of the barcode sequence.

We placed these barcodes on high copy number plasmids as opposed to integrating them into the chromosome to enable analysis of extremely small populations of cells, though this benefit was not utilized in these experiments. In the ideal case, every template molecule is extracted, amplified, and has a chance to be sequenced, however there is inevitably loss during extraction and sample preparation. For small populations of cells, a high copy plasmid ensures that each cell has many template molecules each with some probability of being amplified and sequenced, and that no cell goes completely undetected.

Despite these advantages, there are disadvantages to placing the barcodes on plasmids as opposed to the chromosome which should be taken into account. Fluctuations in plasmid copy number may introduce additional noise to the measurement, and chromosomal integrations may be more stable in situations where supplementation with antibiotic is not possible or desired. Placing the barcodes on the chromosome may also prevent the double transformation events we observed with plasmids. This could abolish the need to validate the entire library by Sanger sequencing and allow for creation of larger libraries.

1b) Experimental conditions

The bacterial community was grown in 2 mL of LB Lenox broth (EMD lot number: VM576547 616) supplemented with chloramphenicol (35 µg/mL) and ampicillin (100 µg/mL) in 13 x 100 mm glass tubes in an incubator at 37 ^{o}C shaking at 300 rpm.

Every 24 hours cells were diluted to desired final concentrations then immigrant cells were added. This new population was placed in fresh media and allowed to grow another 24 hours before repeating the passaging procedure.

The initial community consisted of all 456 species in roughly equal abundance and was the same across all experiments. Each round fresh immigrants with the same composition as the initial community were thawed from glycerol freezer stocks at -80 ^{o}C and grown for one day overnight to saturation, then diluted appropriately and added to the communities. This maintained an unbiased naïve pool of immigrants going into each experiment. A key of parameters for each experiment is given in Supplementary section 6.

We measured concentration of cells at saturation with dilution plating at the start of the experiment and after the last round for experiment E6. We found that the initial saturated concentration of cells was (3.25 ± 0.40)x10^{9} cells/mL and the final saturated concentration of cells was (3.26 ± 0.63)x10^{9} cells/mL (mean ± standard deviation of ten plates), suggesting that the saturated cell density remained relatively constant through the experiment.

1c) DNA extraction, amplification, and sequencing

Plasmid DNA was extracted by miniprep kit according to the manufacturer's instructions (Qiagen). Template DNA was amplified by a two stage Phusion polymerase (NEB) PCR reaction similar to approaches used elsewhere[2–4]. In a first one cycle PCR reaction, adapter regions and unique molecular identifiers (UMIs) were added to each template (First PCR) This step generates one uniquely labeled functional template per initial template plasmid molecule. The initial primers were then removed using Ampure XP Beads at a ratio of 1:1. A second PCR reaction was then run for 35 cycles to amplify these labeled templates (Second PCR). During this reaction known indices were attached to allow informatic demultiplexing of pooled libraries.

First PCR

98 ^{o}C for 3 min, 2x(55 ^{o}C for 30 s 72 ^{o}C for 2 min), hold at 10 ^{o}C

Primers:

5' - ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTCGCTAAGGATGATTTCTGGA - 3'

5' - GACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNTCGCTTGGACTCCTGTTGAT - 3'

NNN = UMIs

Second PCR

98 ^{o}C for 30 s, 40x(98 ^{o}C for 20 s, 72 ^{o}C for 30 s, 72 ^{o}C for 30 s), 70 ^{o}C for 3 min, Hold at 4 ^{o}C

Primers:

5’ - AATGATACGGCGACCACCGAGATCTACACNNNNNNAAACACTCTTTCCCTACACGACGCTCTTCCGATCT - 3'

5’ - CAAGCAGAAGACGGCATACGAGATAANNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT - 3'

NNN = Multiplexing indices

Successful reactions were confirmed by agarose gel electrophoresis and pooled in equal abundance. The sequencing library was then finalized by purification again with Ampure XP Beads at a ratio of 1:1.

Using this two stage PCR scheme allows correction for bias incorporated during PCR amplification, since every unique template molecule should be tagged with a unique molecular barcode. Bias that happens during additional rounds of amplification can be removed by collapsing each unique molecule into a single read. Sequencing was performed on an Illumina miseq instrument with read lengths of 2x75.

1d) Analysis pipeline

Analysis was greatly simplified by using known barcode sequences. Reads took the form:

NNNNNNNNTCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAGNNNNACNNNNATNNNNGTNNNNCGNNNNGG
TCATTACTGGATCTATCAACAGGAGTCCAAGCGANNNNNNNNNNNNNNNN

NNN = Species barcode

NNN = UMIs

First, the pooled paired reads from the sequencer were demultiplexed into different experimental time points to be processed in parallel. Within each experimental time point forward and reverse reads were aligned with each other using the program seqprep (github.com/jstjohn/SeqPrep) with input arguments [-q 10 -o 40 -n 0.85 -m 0.5] to form a consensus read.

Next, the locations within the read of the regions flanking the barcodes,

TCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAG

and

GGTCATTACTGGATCTATCAACAGGAGTCCAAGCGA,

were determined by a custom python script which first checked reads for exact matches (fast), then, if an exact match was not found, implemented an exact Needleman-Wunsch[5] algorithm (slower) to allow looser global alignments which exceeded a threshold alignment score of 55 for the longer consensus region or 47 for the shorter consensus region with scoring criteria [match = 2, mismatch = -0.9, gapopen

= -1, gapextend = -2]. Based on the position of these flanking alignments, the sequences of both the species barcode and the unique molecular identifiers were extracted. All aforementioned thresholds were chosen to be intentionally loose to avoid biasing that might occur from discarding reads early in the pipeline.

The species barcodes (along with their associated UMIs) were then sorted into 456 bins (one for each species potentially present) by a custom python script, which again first sorted based on exact matches to the extracted species barcodes, then by an exact Needleman-Wunsch algorithm requiring a threshold score of 53, using the same scoring criteria.

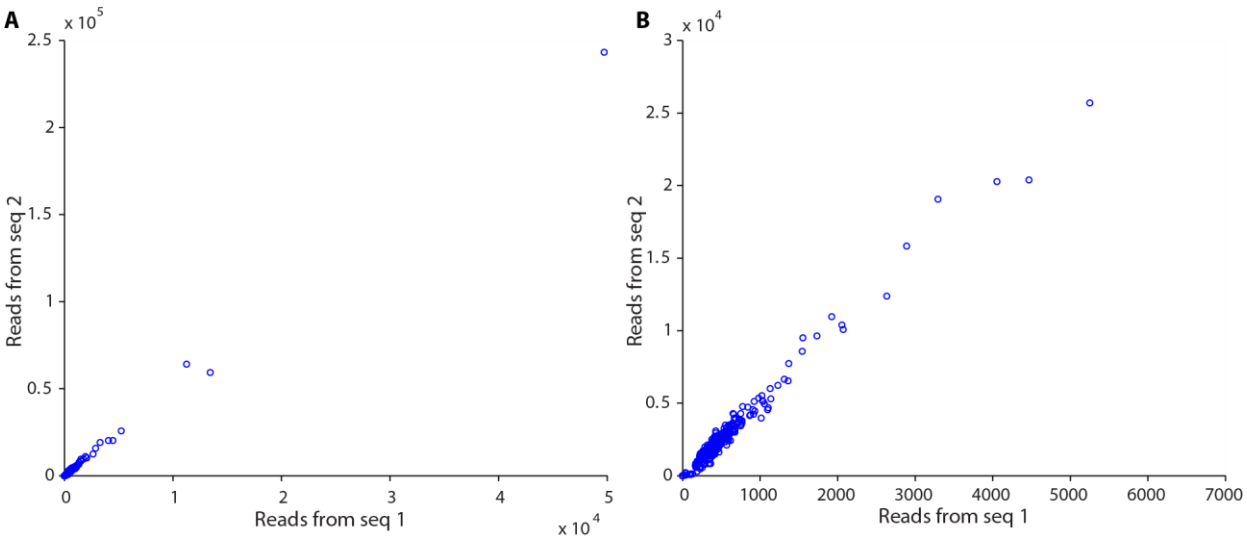
In the last step, the UMIs for each sequence assigned to a particular species within an experimental time point were compared to each other. All exact matches between these UMIs were collapsed into a single read, thereby reducing PCR bias by avoiding counting reads which arose from the same template molecule multiple times, though we note that this step did not seem to have a large effect on measured relative abundances in these experiments.

After this analysis, the number of reads remaining in each species bin is proportional to the number of cells of that type in the experimental time point.

Though we implemented this thorough analysis pipeline to avoid wasting reads, introducing bias during analysis, and to eliminate any PCR bias, it is relevant to note that we also obtained similar results by simply counting the number of reads which contained exact matches to each known barcode in each experimental time point.

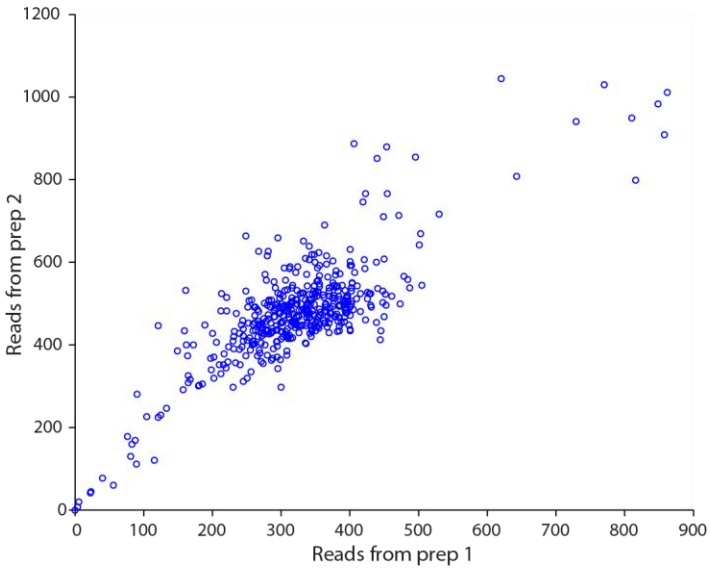
1e) Validation, controls, and noise floor

We ran several controls to verify the final barcode counts were indeed reflective of the concentration of cells as expected. First, we evaluated the noise associated with sequencing by comparing two sequencing runs on the same library and found the measured relative abundance was highly reproducible.



Supplementary figure 2. Sequencing replicates. The same sample was sequenced twice and reads detected from one sequencing run are plotted against reads from the other. Each circle is a single species. B) is a zoom of plot A).

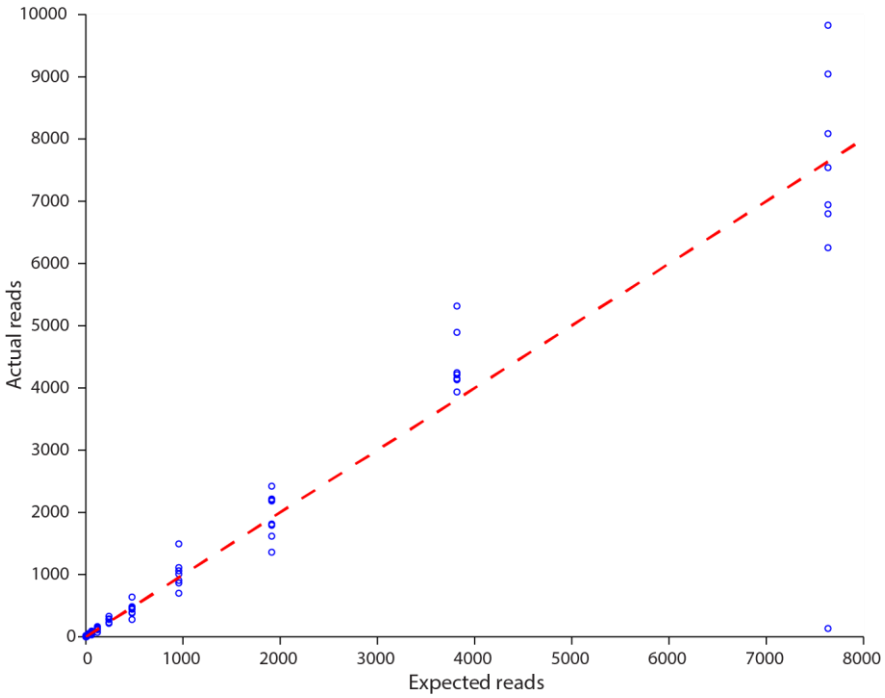
Next, we prepared and sequenced the same sample twice to evaluate the noise associated with sample preparation and sequencing. We noticed that the sample preparation introduced additional noise, but the measured relative abundances were still highly correlated.



Supplementary figure 3. Sample preparation and sequencing replicates. We prepared separate libraries from the same sample and sequenced them both in a single sequencing run. Reads detected from one preparation are plotted against reads from the other. Each circle is a single species.

Next, we created a community with members of known concentration by diluting cultures grown to saturation by various amounts. This allowed us to assess whether the measurement after the entire

workflow from sample processing, to sequencing, through analysis reflected the expected abundances. We found the measured reads were highly correlated with expected reads.

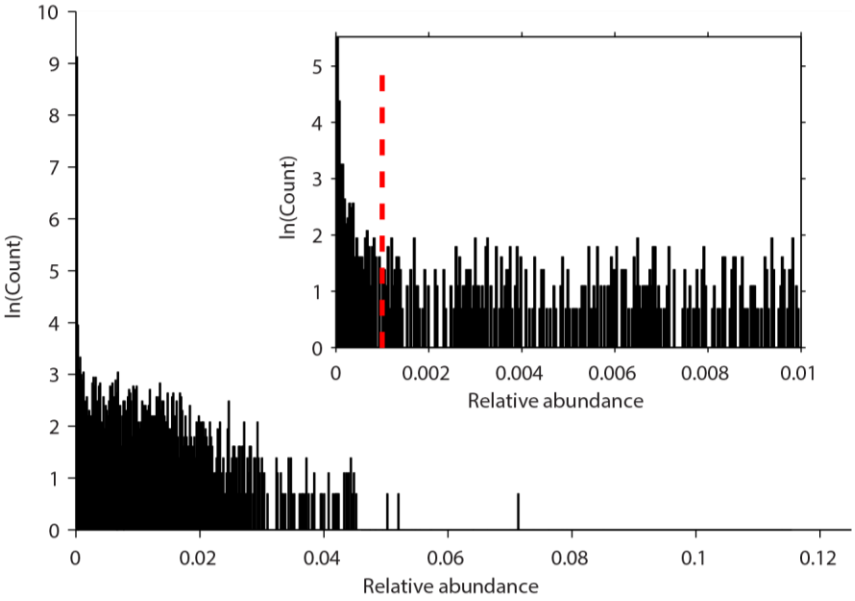


Supplementary figure 4. Validity of relative abundance measurements. Actual reads as measured by the full workflow discussed in "Methods" were plotted against expected reads, as determined by the known dilution factor. Each circle is a single species (multiple species were diluted to the same frequency). The red dashed line reflects unity for reference.

Despite this high correlation, we noticed two things. First, the abundance of one barcode, that of species #1 (lower rightmost point in Supplementary figure 4), was an outlier over an order of magnitude below the expected abundance. This species and around two other species depending on the experiment were also over an order of magnitude less abundant than average in the initial populations and metacommunity. Though we are not sure of the reason for this discrepancy, these species make up <1% of the total number of species, and we have no reason to expect they strongly influence the results. Second, some barcode sequences which were not included at all in this control experiment had a small number of reads assigned to them. Setting more stringent thresholds, for example by counting only exact matches to species barcodes, did not remedy this apparent noise, so it is unlikely that this noise is due to sequencing errors in the species barcodes resulting in misclassification of one species barcode as another. This noise arises from the from the "index hopping" that occurs when multiplexing samples on Illumina sequencers.

This resulted in a noise floor, below which reads could not be reliably detected. We analyzed experiment E1 which had a bottleneck size of 3.25 cells and immigration of 55 cells to help set this threshold. We took advantage of the fact that most of the 456 species are not expected to appear in the experiment at any given time point with such a small starting population and so many possible species. We plotted

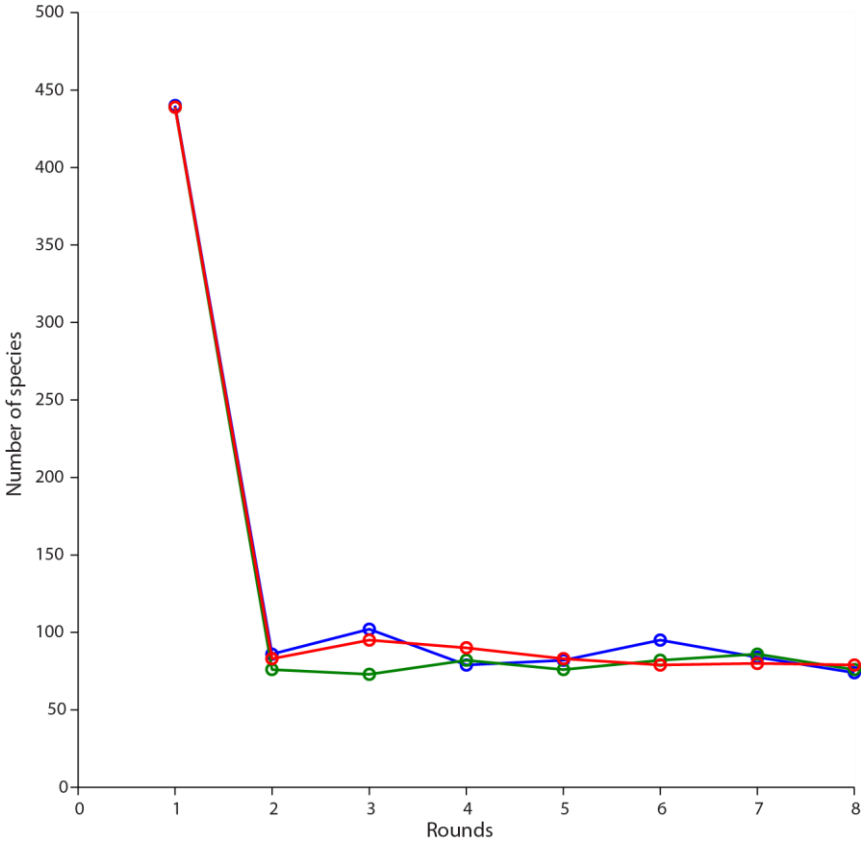
histograms of the relative abundance over each round for this experiment. These histograms should be the sum of two distributions: one representing the measured relative abundances of species actually present in the experiment, and one representing the noise associated with species which have a true abundance of zero. This distribution of zero abundance noise should peak at zero and fall rapidly as measured relative abundance increases. Indeed, we see that the histogram of relative abundance peaks at zero, and falls rapidly to an apparent baseline. We chose a 1/1000 threshold based on this data. Anything less than this threshold was not counted as a read. We erred on the side of what we think is a conservative threshold, since commonly used ecological measures such as total number of species present are sensitive to such noise, and we wanted to be sure that detected species were indeed actually present in the experiment.



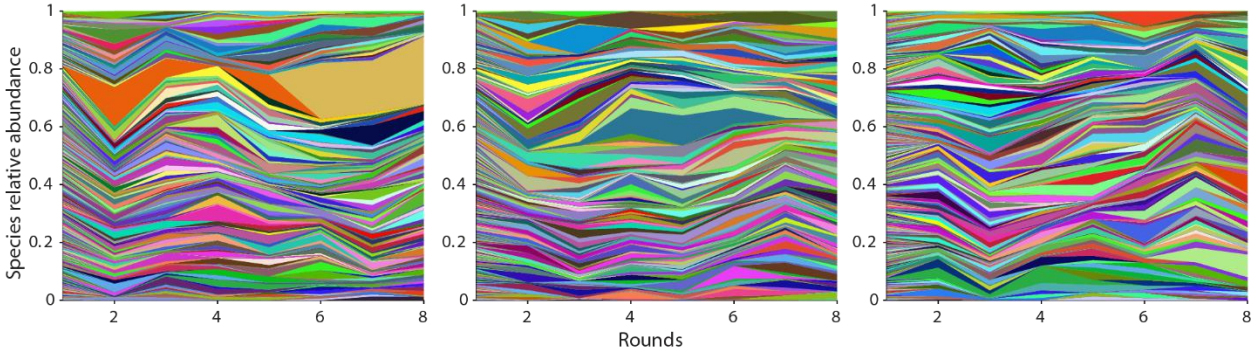
Supplementary figure 5. Noise floor. Here we plot a log scale histogram of relative abundance for rounds 2-25 of experiment E1. Inset is a zoom on the larger plot. The red line represents the detection threshold.

1f) replicates

We chose a strategy which allowed us cover a wide range of bottleneck sizes, as a result, we had fewer replicate experiments. We did run the experiment with bottleneck size 32.5 in triplicate for 8 rounds which revealed similar numbers of detected species across rounds (Supplementary figure 6) and similar stochasticity in relative abundance with different species competing most successfully in each replicate experiment (Supplementary figure 7).



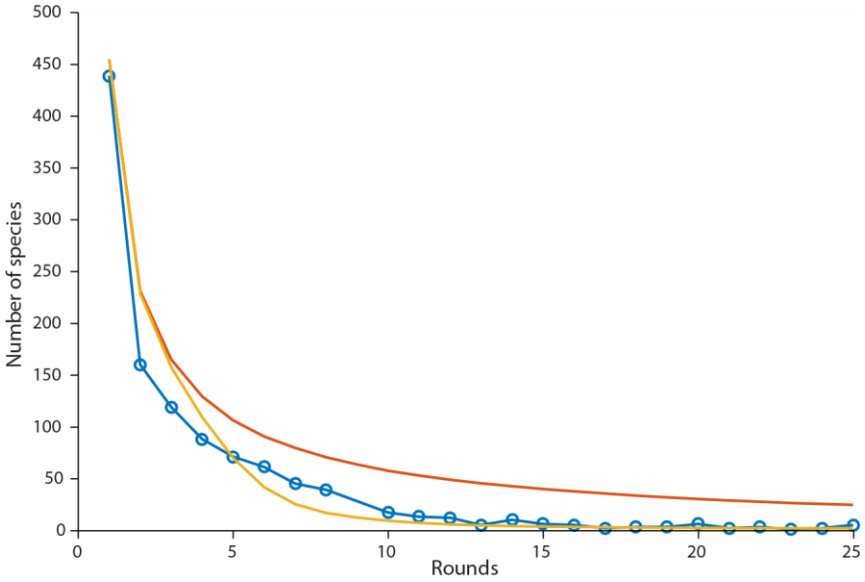
Supplementary figure 6. Species vs time replicates. The number of species present each round for triplicate experiments of bottleneck size 32.5 cells, immigration 55 cells/round.



Supplementary figure 7. Replicate Muller plots. Muller plots showing the relative abundance of all 456 strains over 8 rounds for triplicate experiments of bottleneck size 32.5 cells, immigration 55 cells/round. Each species is represented by a different color where the proportion of each color represents the relative abundance of that species. Color/species pairings are consistent from plot to plot.

We also included one experiment with no immigration. In this experiment the community steadily lost diversity (Supplementary figure 8). With no immigration and this single bottleneck size, it is difficult to assess the validity of each model type since they make similar predictions for metrics such as the number of species present over time. Including a range of bottleneck sizes and immigration increased

the discrepancy between the model predictions and allowed us to more easily assess which best captured the system.

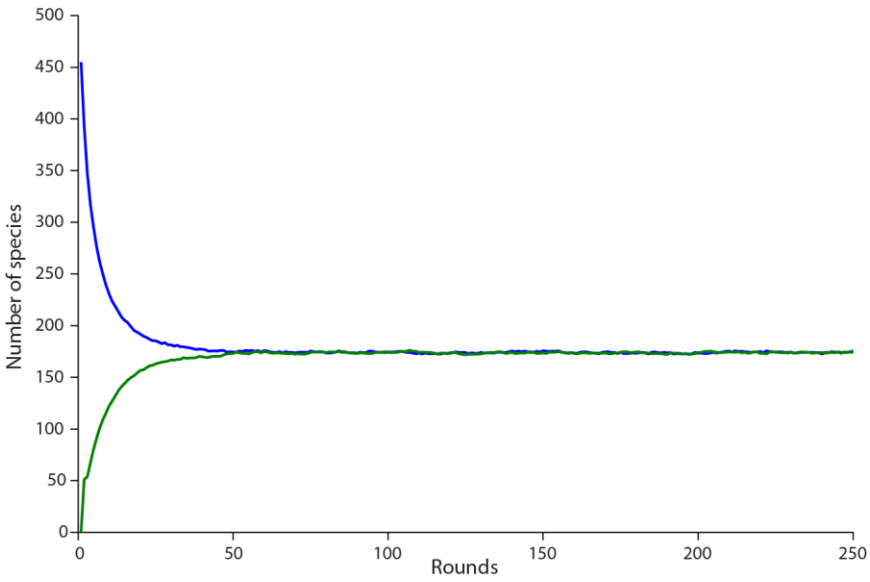


Supplementary figure 8. Community dynamics with no immigration, bottleneck size 325 cells. Blue line: experimental data. Red line: neutral simulation, average of 100 trials, gold line: selective simulation, average of 100 trials. Note: experimental point 9 was not included, since this point had only around 2% of the total reads compared to other time points.

2) Models

2a) Neutral model

We established and simulated a neutral model to make predictions about the community composition by accounting for random changes in relative abundance due to bottlenecking and immigration. In this model, the initial community was seeded by randomly assigning each starting species' abundance from a Poisson distribution with mean of 22 individuals. This resulted in an initial distribution of relative abundances comparable to the initial experimental distribution (see Supplementary figures 21-23). Bottlenecking and immigration were then simulated in series each round. We determined the number of individuals of each species after a bottleneck by drawing from a Poisson distribution with mean equal to the previous relative abundance multiplied by the bottleneck size. Immigrants were then added also by a random Poisson process for each species with mean equal to 1/456 (1/number of species) multiplied by the number of individuals immigrating. Rounds were sequentially simulated to determine community changes over time. Note that our model has obvious parallels with previous neutral models in ecology[6], and species influx due to immigration and species loss from random sampling balance each other to create an equilibrium number of species present in the simulations (Supplementary figure 9).



Supplementary figure 9. Stability of the neutral simulation. With immigration (here on average 55 immigrant cells per round) and a bottleneck size of 1625 cells) a neutral model predicts an equilibrium number of species present, independent of the number of species present as a starting condition here 456 (blue) or 0 (green), averages of 100 trials.

2b) Selective model

We also constructed and simulated a model where species had differences in fitness (selective model). This model contained bottleneck and immigration elements as in the neutral model, but also contained an additional phase for "growth". During growth, after bottlenecking and immigration, each species' abundance was multiplied by that species' per round relative (Darwinian) fitness to give a new abundance. Relative fitness values were assigned randomly to each species from an exponential distribution and scaled according to the number of replication rounds for each experiment, as discussed in Supplementary section 5. These relative fitness values were static throughout the simulation and this model might be further refined by accounting for dynamic adaptive changes to fitness throughout the experiment.

Since our experimental detection threshold was an abundance of greater than 1/1000, all simulation results (both neutral and selective) were also thresholded to enable direct comparison unless otherwise noted.

2c) Transition from selective to neutral

In addition to the simulations, we investigated a simple scenario to understand the transition from neutral to selective, considering a single species with a selective advantage. Say that this species has fitness R compared to the neutral species such that its frequency changes as $f_t = f_0 e^{Rt}$ when there is no slowing of growth due to the population's carrying capacity.

We will describe the dynamics of the population using a continuous time approximation which results in somewhat simpler equations than discrete cycles. The continuous time results capture the same qualitative features and become exact when the changes in frequency over a cycle are small. In the continuous time picture, there is a total population of $N + M$ and individuals are replaced by immigrants at a rate $m = \frac{M}{N+M}$. The change in frequency for the fit species can be written as

$$df = Rdtf(1-f) + mdt(f_m - f) + \eta\sqrt{\frac{2}{N_e}dtf(1-f)} \quad (1)$$

The first term describes the change in frequency due to selection. The second is the net change due to immigration from an immigrant pool with the fit species present at frequency f_m. The last term has a random Gaussian variable, η, with variance one. This term describes the effects of stochasticity which is parameterized by N_e which is on the order of $N + M$.

This model, originally by Kimura[7], describes dynamics as a diffusive process and is valid when changes in frequency are small, i.e. N_e is large. It has been previously applied to ecological community models, for example in Sloan *et al.*[8]. The probability distribution for the fit species frequency is

$$p(f) = Ae^{N_e Rf} f^{N_e m f_m - 1}(1-f)^{N_e m(1-f_m)-1} \quad (2)$$

where A is a normalization constant. The effect of selection is simply to make larger frequencies exponentially more likely but this effect is only strong if $R > \frac{1}{N_e}$, i.e. if changes due to selection are larger than changes due to stochasticity.

If selection is strong, then the dynamics can still appear neutral if there is enough immigration. In this regime ($R \gg \frac{1}{N_e}$ and $m \gg \frac{1}{N_e}$) the ratio R/m controls whether the population looks neutral or not. In this approximation, the most likely frequency will correspond to the deterministic equilibrium, f_{eq}, which satisfies

$$0 = df = \left[Rf_{eq}\left(1 - f_{eq}\right) + m(f_m - f_{eq})\right]dt \quad (3)$$

with a solution

$$f_{eq} = \frac{R - m + \sqrt{(R-m)^2 + 4Rmf_m}}{2R} \quad (4)$$

$$\approx \begin{cases} 1 - (1-f_m)\frac{m}{R}\ldots & \text{for } R \gg m \\ f_m + f_m(1-f_m)\frac{R}{m}\ldots & \text{for } R \ll m \end{cases} \quad (5)$$

We see that for $R \ll m$ the results are consistent with the neutral case ($R = 0$) which has an equilibrium frequency of f_m. Otherwise, selection will shift the equilibrium closer to a frequency of one. This result can be understood simply. Being replaced by new immigrants decreases the effective growth rate such that $df \approx (R - m)f$. If the effective growth rate is positive, $R - m > 0$, then selection is important and

the dynamics appear non-neutral. If it is negative, then the immigration dominates the selection and the dynamics appear neutral with immigration.

For discrete cycles this condition is slightly different. During a cycle a fraction of the fit species is replaced by immigrants, so $\tilde{f} = \frac{N}{N+M} f_1$. Then growth due to selection increases the frequency by a factor W. The overall change in frequency is $f_2 = W\tilde{f} = \frac{N}{N+M} f_1$. The dynamics appear non-neutral if the frequency increases over a cycle, so the condition for discrete cycles is

$$W \frac{N}{N+M} > 1 \tag{6}$$

It is convenient to work with log-variables so let $W \equiv e^R$ and $\frac{N}{N+M} \equiv e^{-\delta}$. The condition in eq. (6) becomes $R - \delta > 0$. If R and δ are small, then the discrete case is close to the continuous limit. The condition for non-neutrality is the same as before since

$$\frac{N}{N+M} = e^{-\delta} \tag{7}$$

$$1 - m \approx 1 - \delta \tag{8}$$

so $m \approx \delta$. This reproduces the condition $R - m > 0$ from before.

2d) Maximizing effective fitness

Different experiments grow to the same saturation density but start at different bottleneck sizes. This means that cells in different experiments experience a different amount of time in the various phases of bacterial growth and therefore the same cell may have a different per round fitness, R, in each experiment. The simplest selective advantage is an increase in the doubling rate by r. The per round fitness then scales with the number of doublings, T, since the overall change in frequency goes as

$$W = 2^R = 2^{rT} \tag{9}$$

where r is measured in units of the doubling time. To reach the final number of cells N_f from the bottleneck size N, it takes $T = \log_2 \frac{N_f}{N+M}$ doublings. The fit species will increase in frequency across multiple cycles according to its effective growth rate:

$$R - \delta = rT - \delta = r \log_2 \frac{N_f}{N+M} + \ln \frac{N}{N+M} \tag{10}$$

The effective growth rate is maximized when

$$N = \frac{M}{r} \ln 2 \tag{11}$$

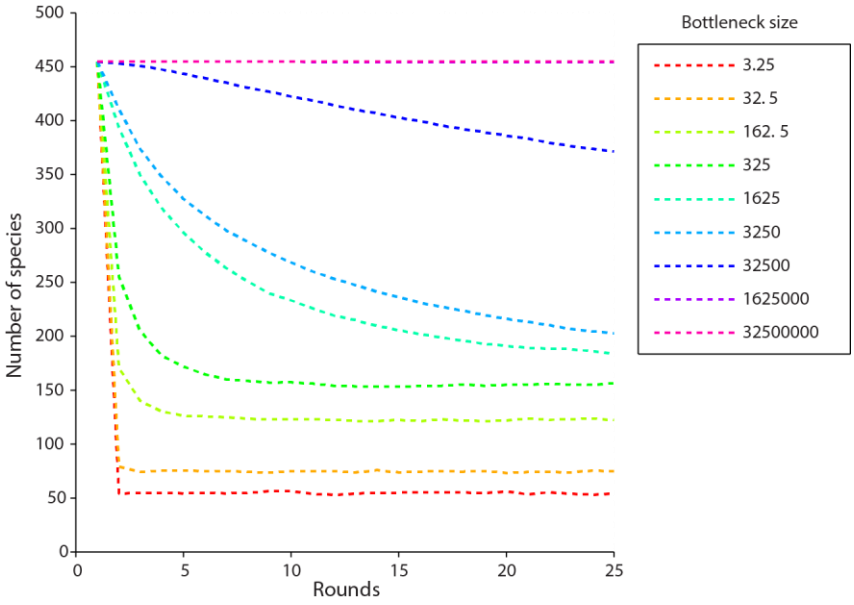
This explains why in the experiment larger fitness differences and a faster drop in diversity are seen at intermediate bottleneck sizes. As the bottleneck size decreases there is a trade-off between growing for a longer time and being replaced by more immigrants and this trade-off results in a maximum at an intermediate bottleneck size.

3) Data and comparisons

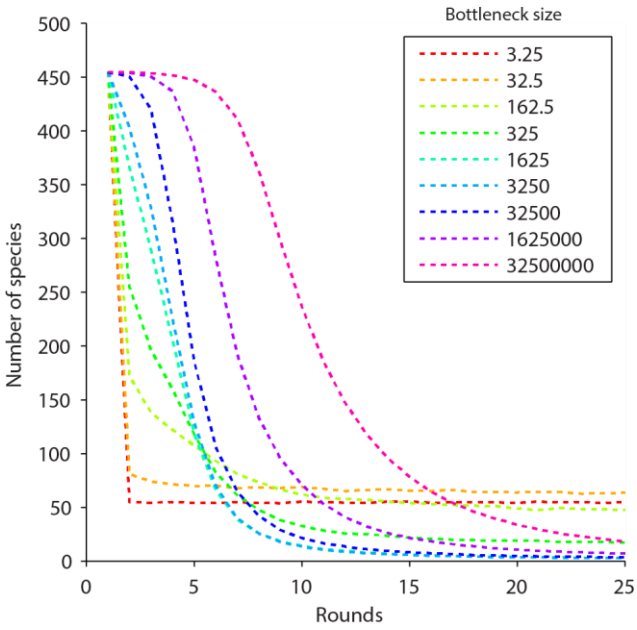
Here we visualize the experimental data and simulation results in a number of ways to enable comparisons between predictions and experimental results.

3a) Species vs. time plots

Here we show mean species vs. time results for both models with the same parameters used in each experimental condition. For clarity of visualization a subset of these results are displayed in the main text. The full species vs. time plot for the experimental data can be found in the main text, figure 2A.



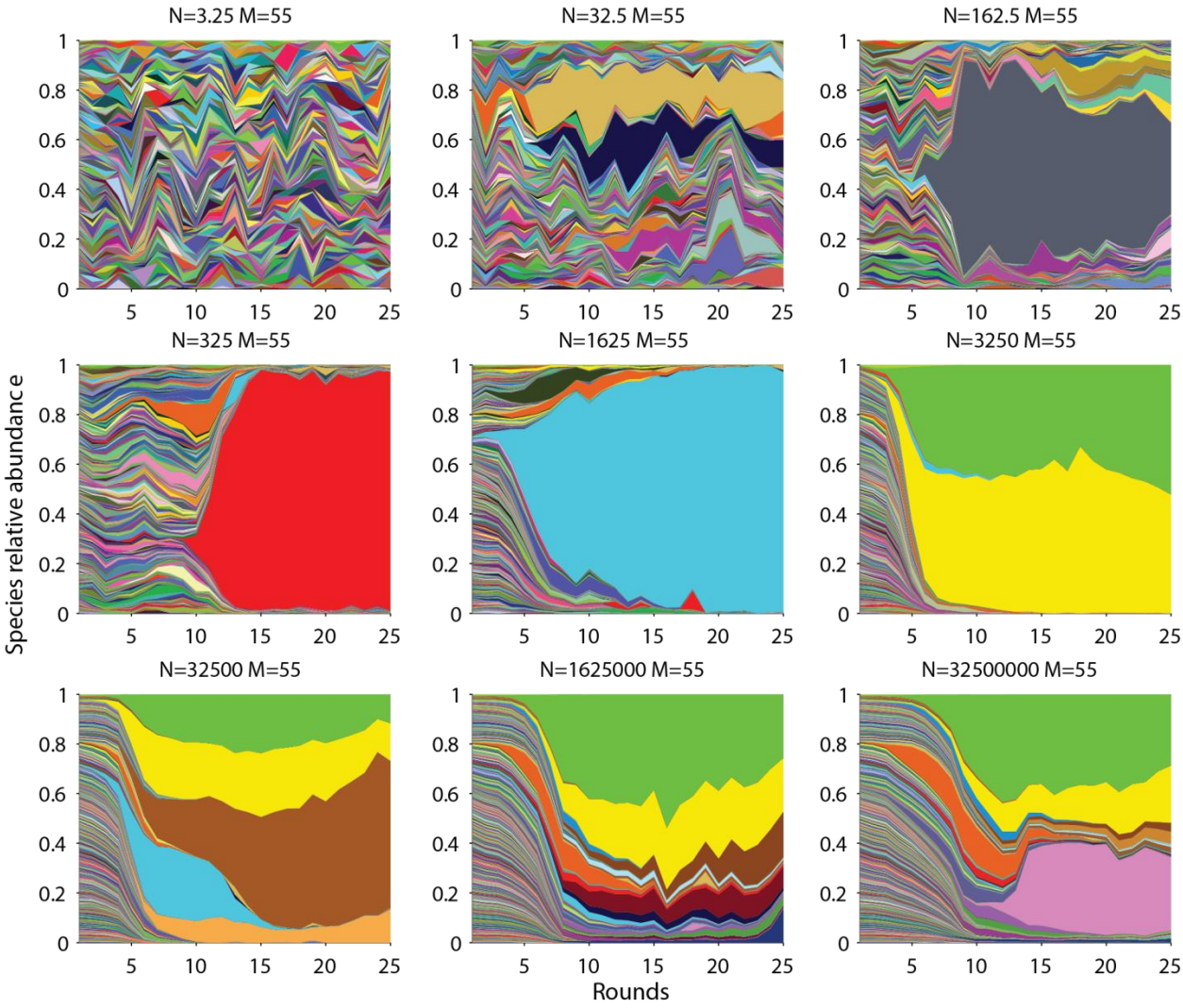
Supplementary figure 10. Number of species present over time in a neutral model. Averages of 100 simulations displayed for indicated bottleneck sizes with immigration of 55 cells/round.



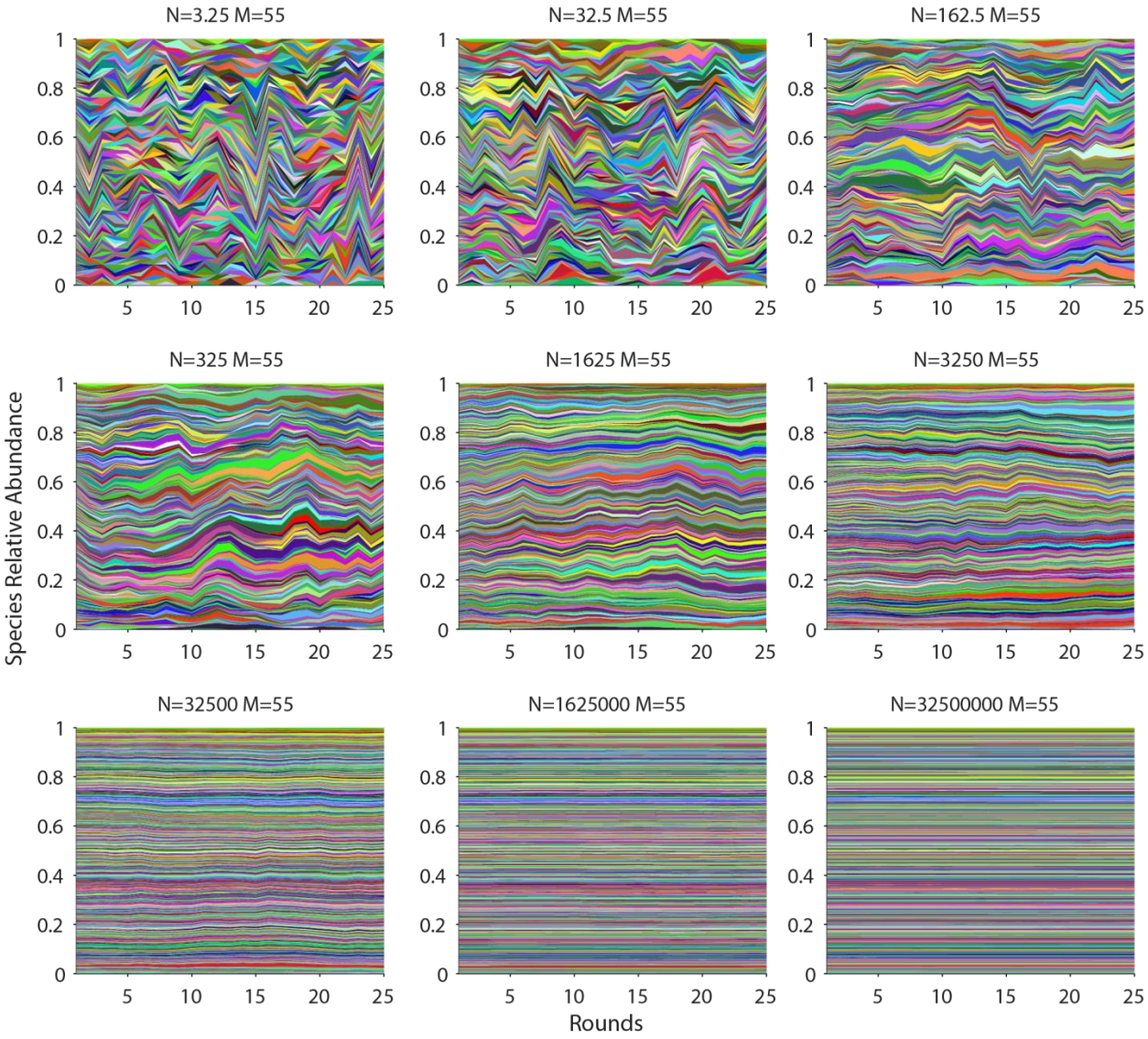
Supplementary figure 11. Number of species present over time in a selective model. Averages of 100 simulations displayed for indicated bottleneck sizes with immigration of 55 cells/round.

3b) Muller plots

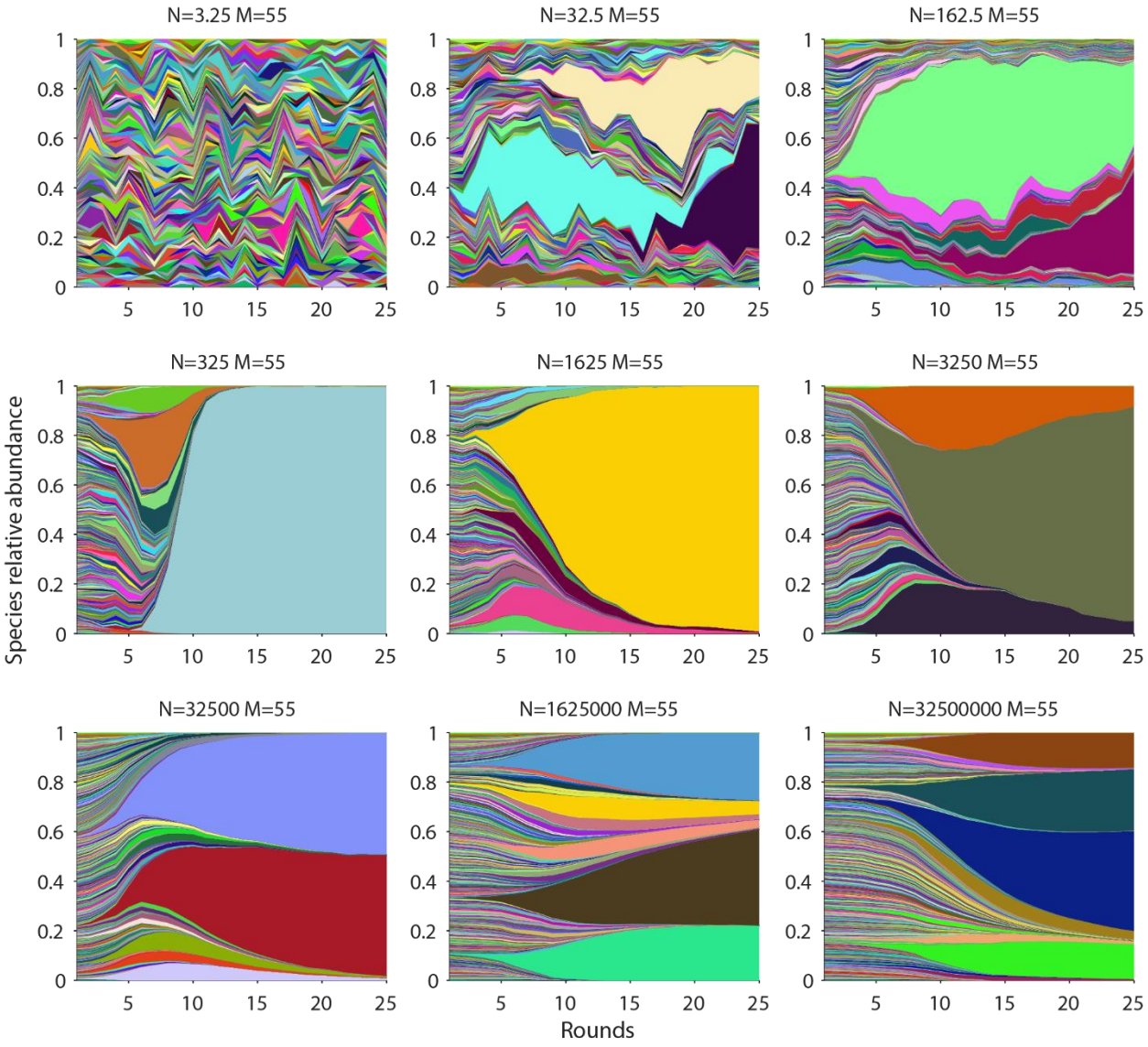
Here we show Muller plots for all the experimental data as well as representative plots for each model. We find that the neutral model is unable to capture the differences in growth seen in all experiments but those with the smallest bottleneck sizes.



Supplementary figure 12. Experimental Muller plots. Muller plots showing the relative abundance of all 456 strains over time for nine experiments. In each subplot title, N indicates the bottleneck size and M indicates the number of immigrants. Each species is represented by a different color where the proportion of each color represents the relative abundance of that species. Color/species pairings are consistent from plot to plot.



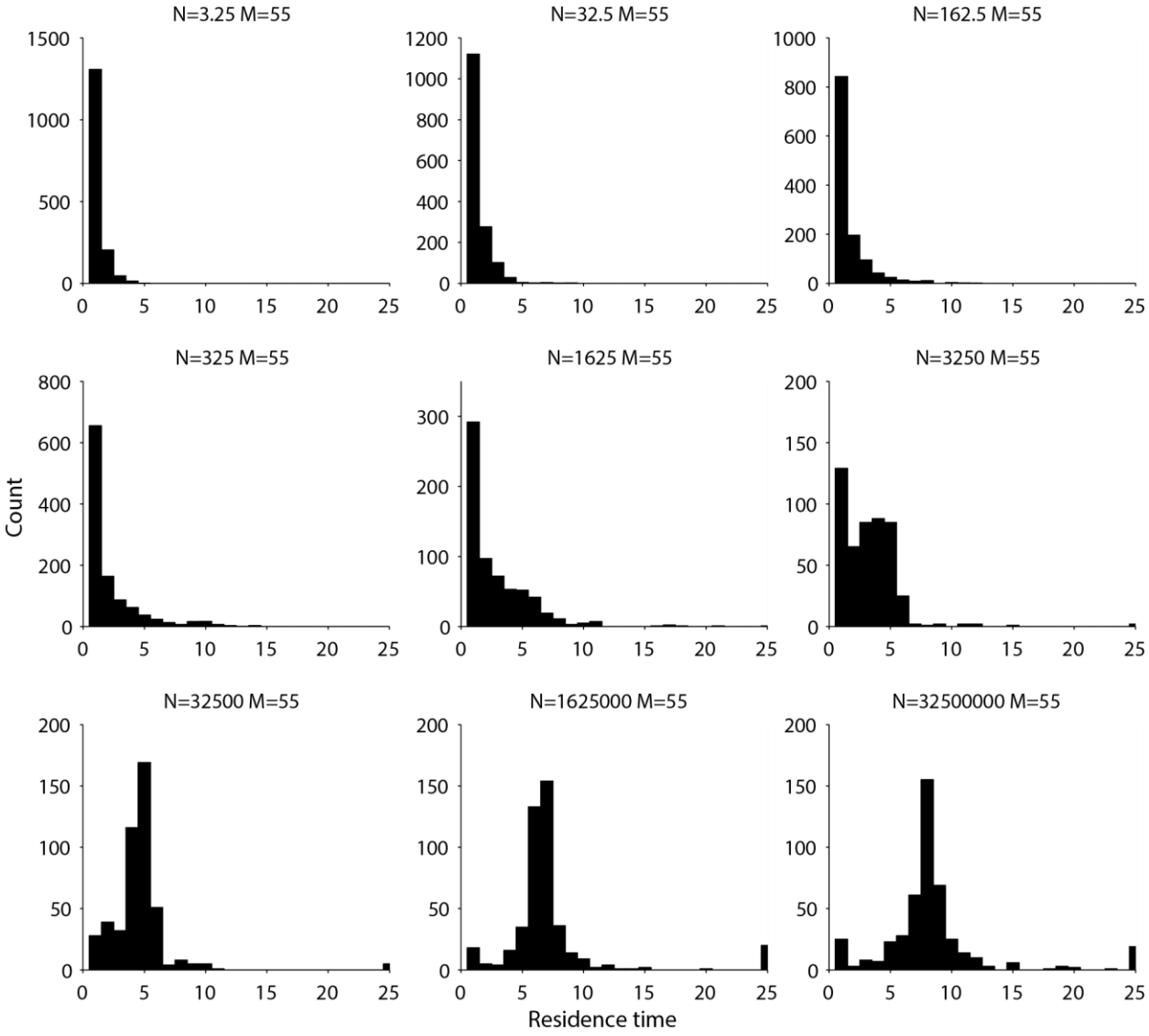
Supplementary figure 13. Neutral Muller plots. Muller plots showing the relative abundance of all 456 strains over time for nine neutral simulations. In each subplot title, N indicates the bottleneck size and M indicates the number of immigrants. Each species is represented by a different color where the proportion of each color represents the relative abundance of that species. Color/species pairings are consistent from plot to plot, representative simulations displayed.



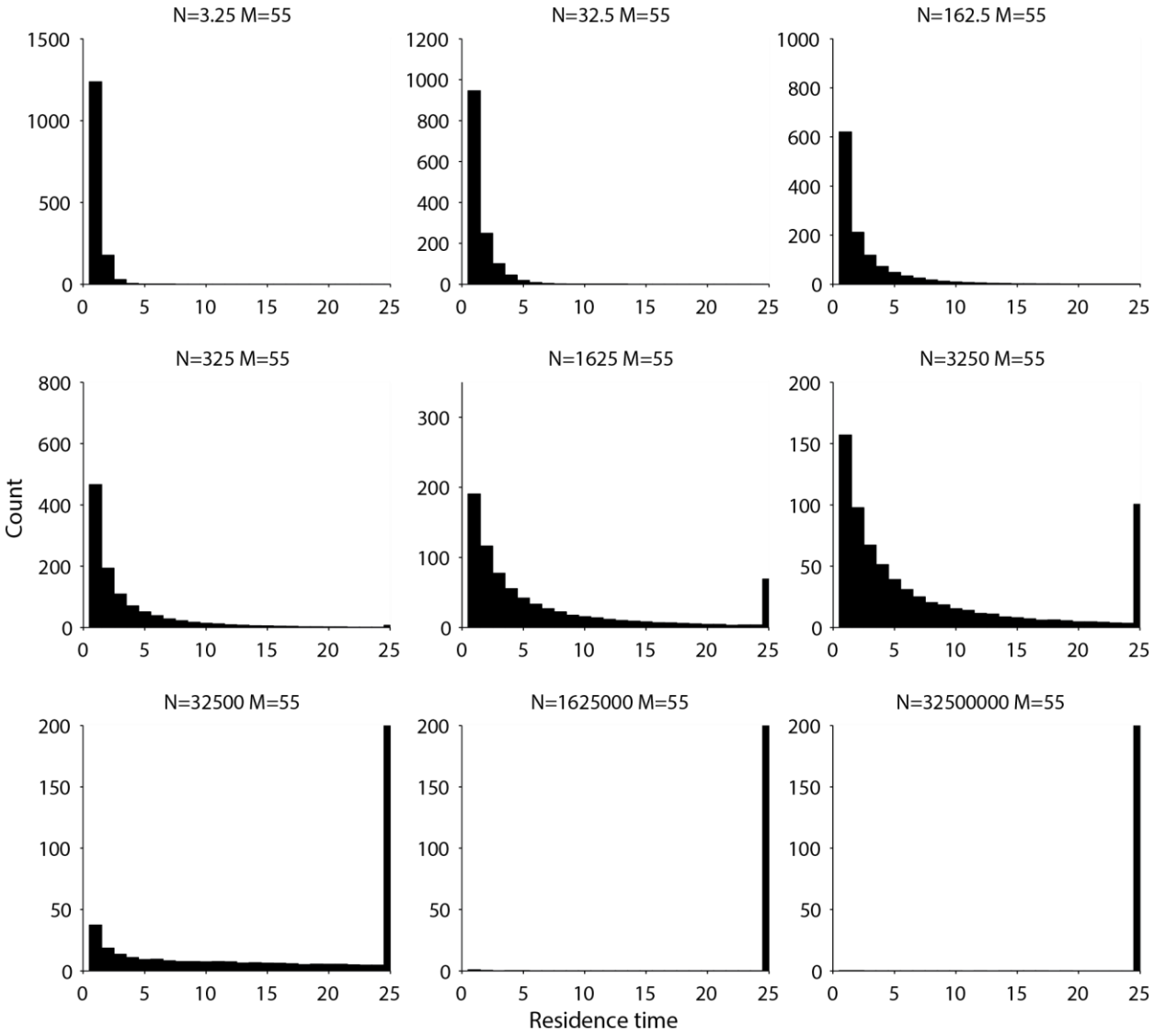
Supplementary figure 14. Selective model Muller plots. Muller plots showing the relative abundance of all 456 strains over time for nine simulations incorporating selection. In each subplot title, N indicates the bottleneck size and M indicates the number of immigrants. Each species is represented by a different color where the proportion of each color represents the relative abundance of that species. Color/species pairings are consistent from plot to plot, representative simulations displayed.

3c) Residence time

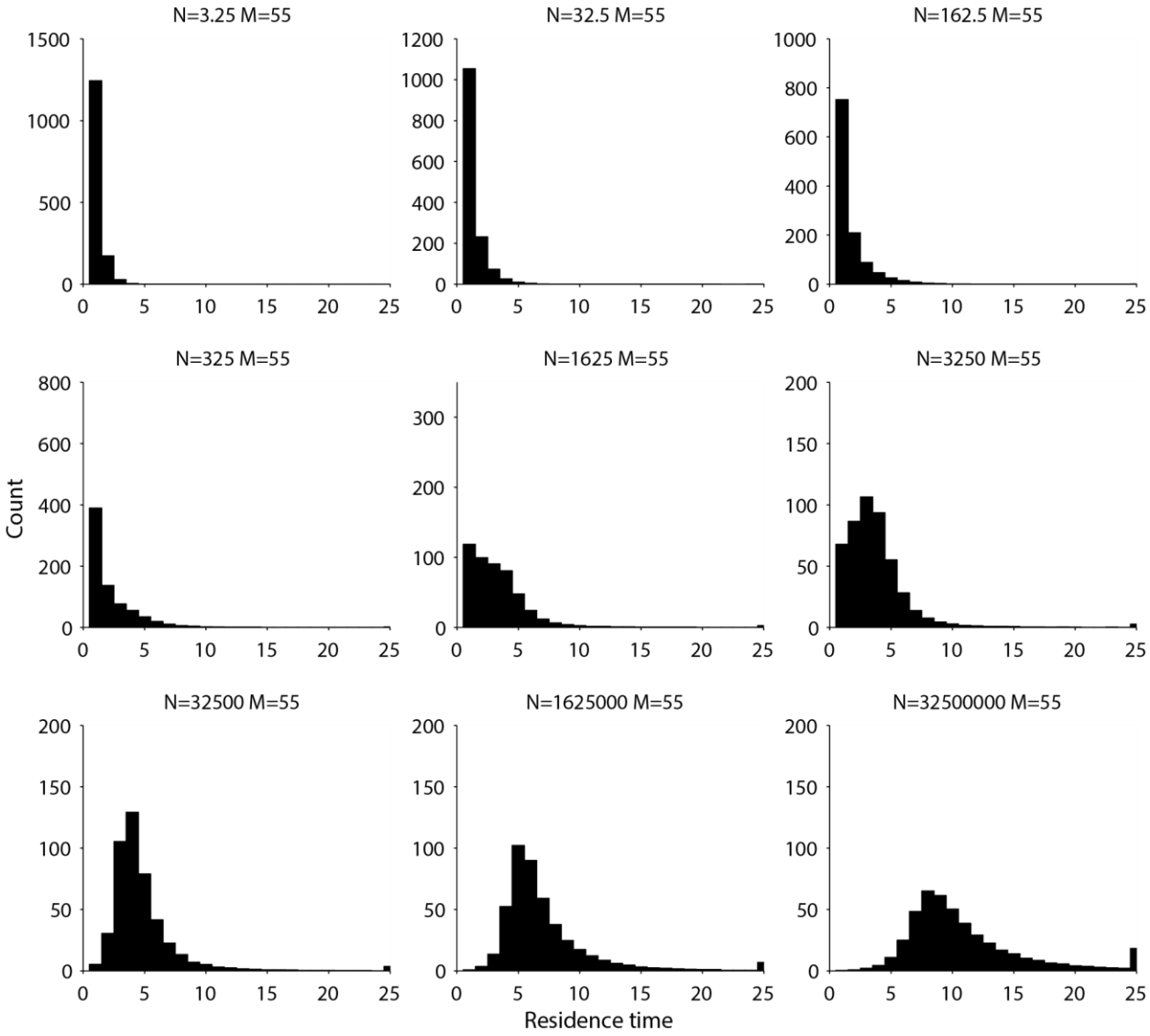
Residence time, also referred to as extinction time, is the length of time a species is present in the community. In addition to commonly used species relative abundance distributions, residence time has been suggested as a metric that can be used to assess whether a system is neutral.[9,10] Here we compare residence time in experiments vs models. We find that the neutral theory overestimates residence times at larger bottlenecks, and a selective model performs much better. We note that here, given the starting conditions, many of these residence time histograms are dominated by the speed at which the initial community loses diversity.



Supplementary figure 15. Experimental residence times. The rightmost histogram bin indicates the number of species present through the entire time range.



Supplementary figure 16. Residence time in neutral simulations. The rightmost histogram bin indicates the number of species present through the entire time range, note that this bar exceeds the axis for the largest three bottlenecks. Bin counts are means of 100 trials.

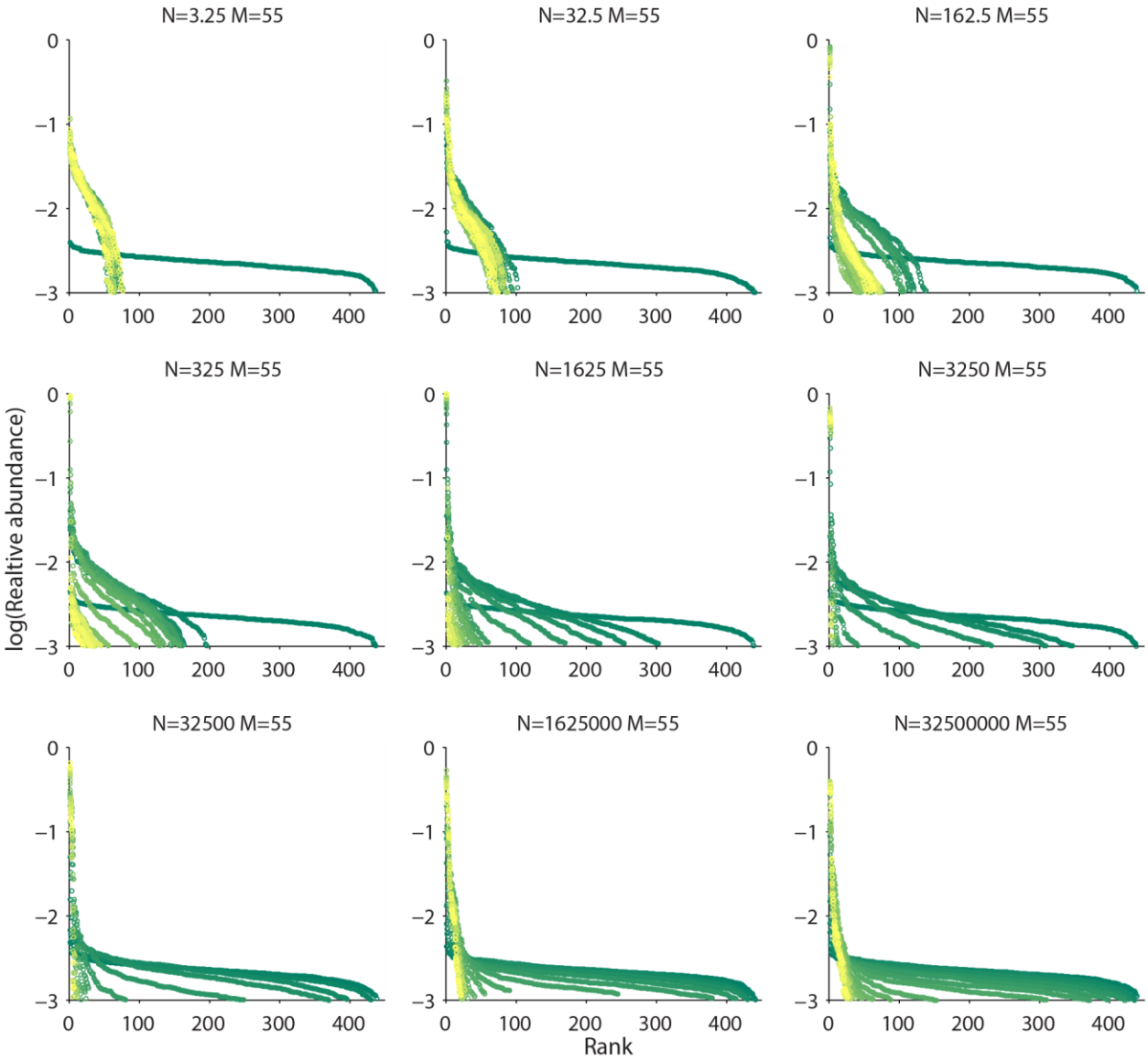


Supplementary figure 17. Residence times in simulations incorporating selection. In each subplot title, N indicates the bottleneck size and M indicates the number of immigrants. The rightmost histogram bin indicates the number of species present through the entire time range. Bin counts are means of 100 trials.

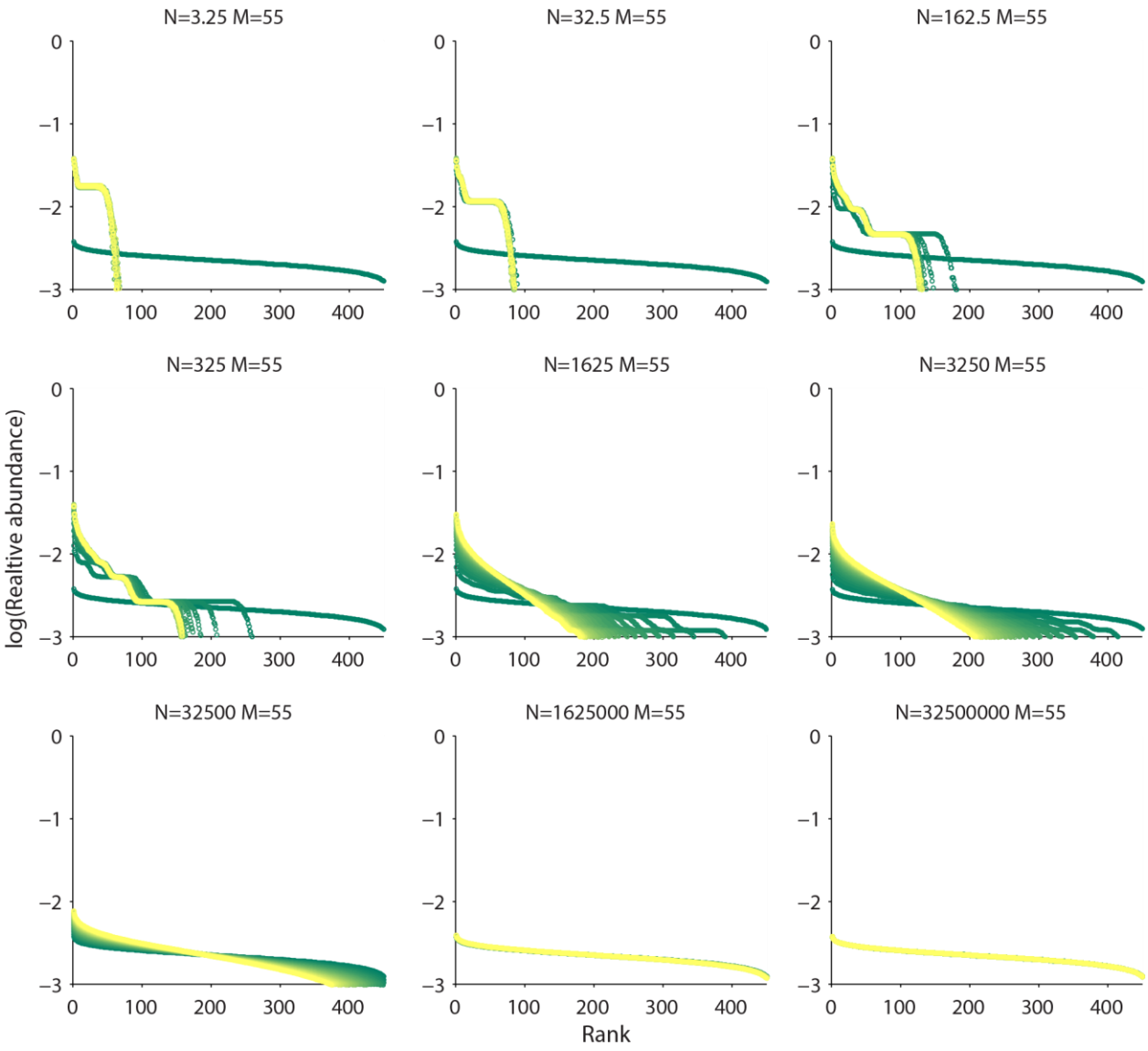
3d) Rank abundance distributions

Data represented as rank abundance plots and species relative abundance distributions are frequently compared to predictions from neutral models[11]. Natural distributions of these types can often be recreated with neutral models by fitting a few parameters. Since the underlying mechanisms which these parameters represent can be difficult to measure in natural systems, the necessity of this fitting has been one criticism of neutral models. Indeed, unconstrained models of different types (selective and neutral) can make similar predictions.[12] In this work we tightly controlled the mechanistic details of the experiment, so the parameter inputs to each model are known and need not be fit. Below we compare

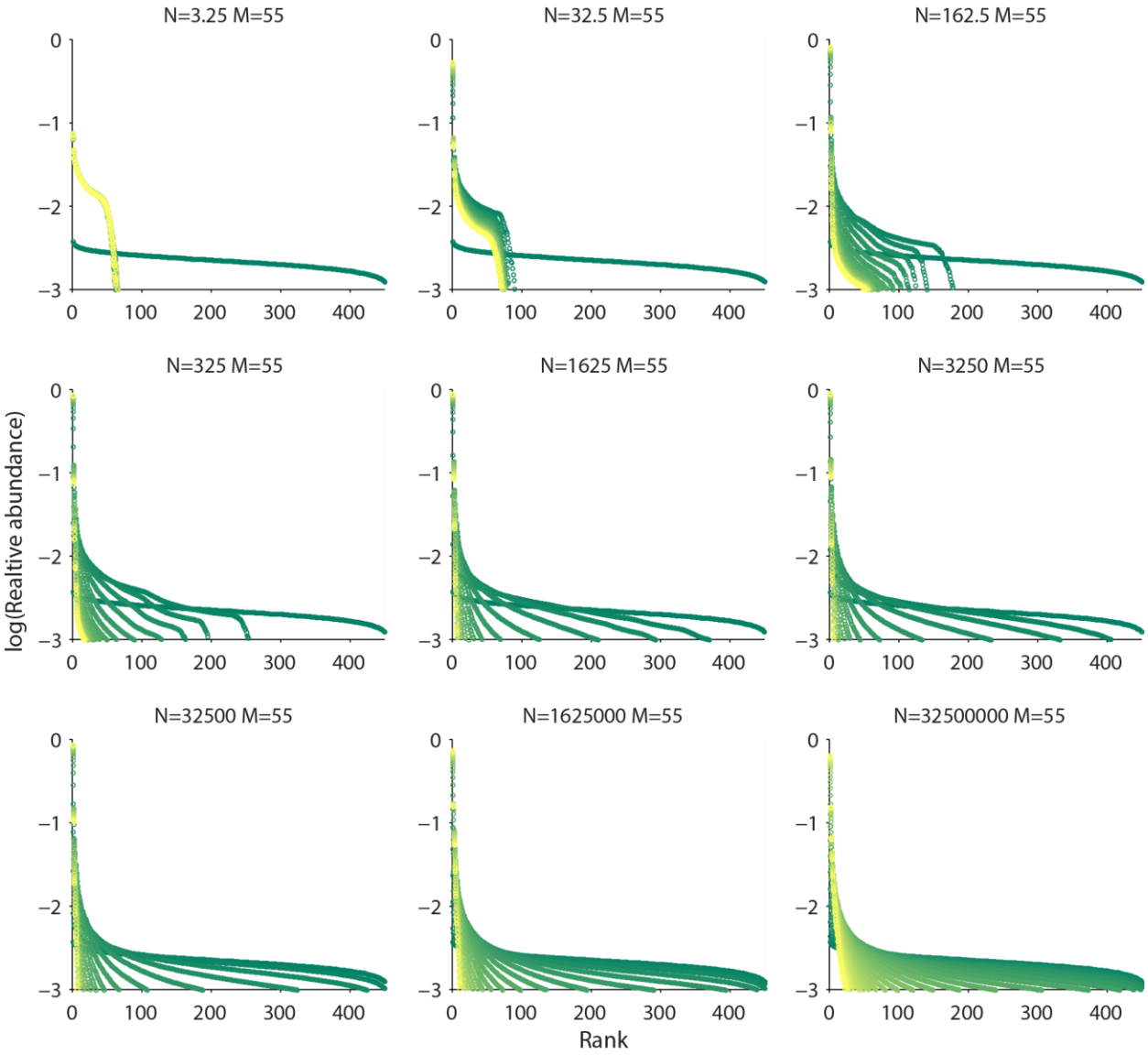
the resulting distributions from experiment and models using these known parameter inputs. We find that the neutral model performs quite poorly in all but the smallest bottleneck sizes. The selective model better captures how these distributions change over time and bottleneck size.



Supplementary figure 18. Experimental rank abundance plots. Colors indicate the round, where blue is the initial time point and yellow is the final round.

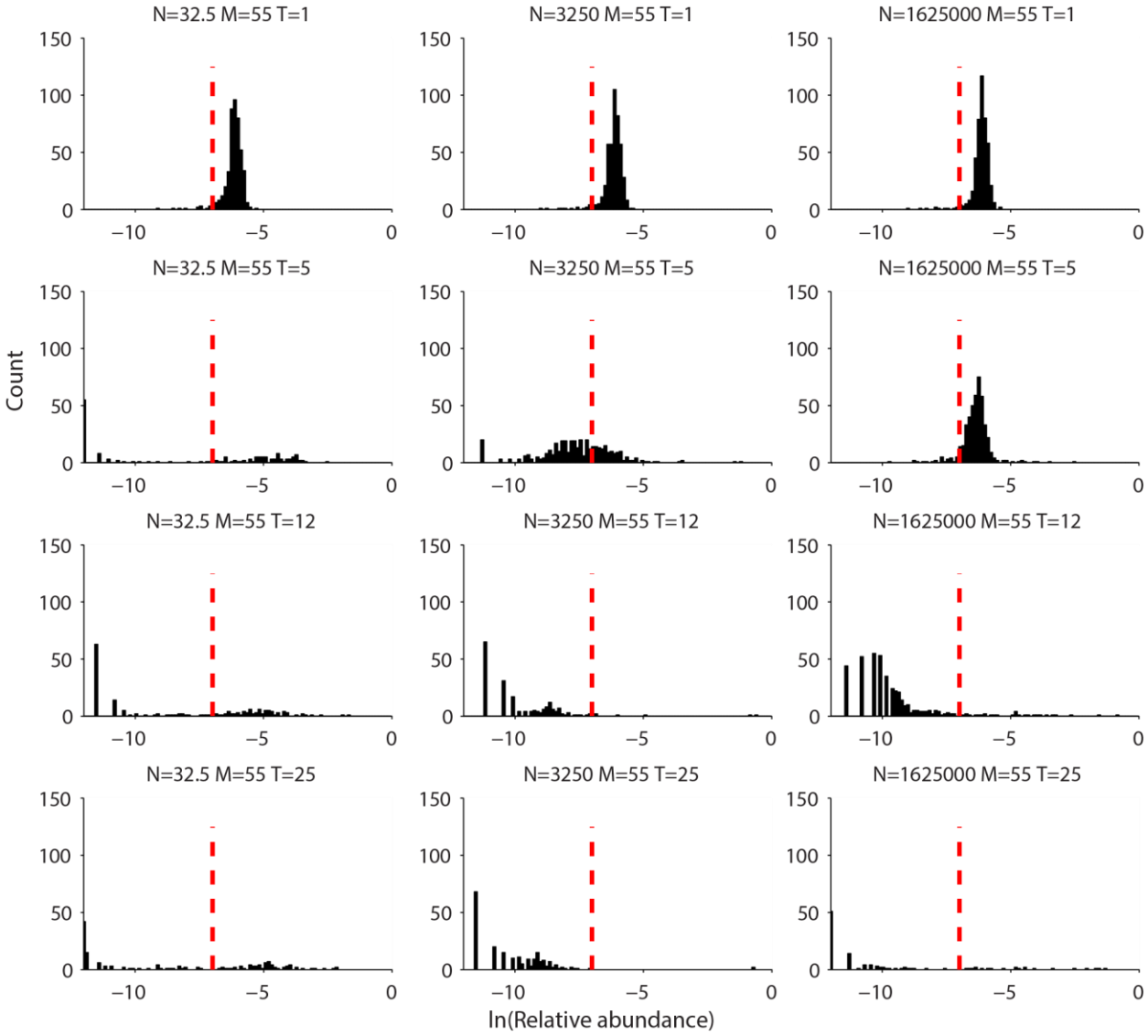


Supplementary figure 19. Neutral rank abundance plots. Colors indicate the round, where blue is the initial time point and yellow is the final round. Mean of 100 trials.

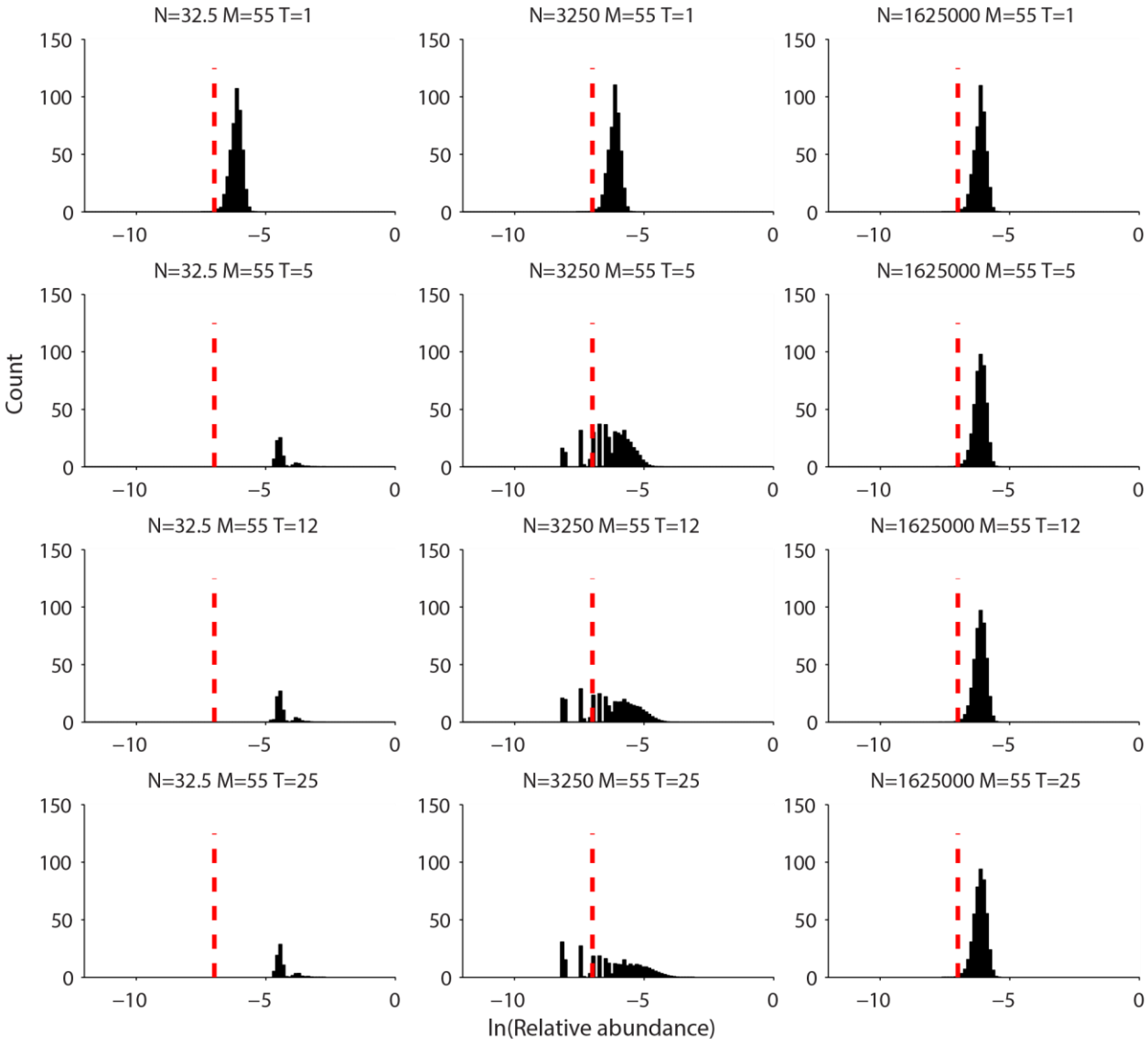


Supplementary figure 20. Rank abundance plots for a selective model. Colors indicate the round, where blue is the initial time point and yellow is the final round. Mean of 100 trials.

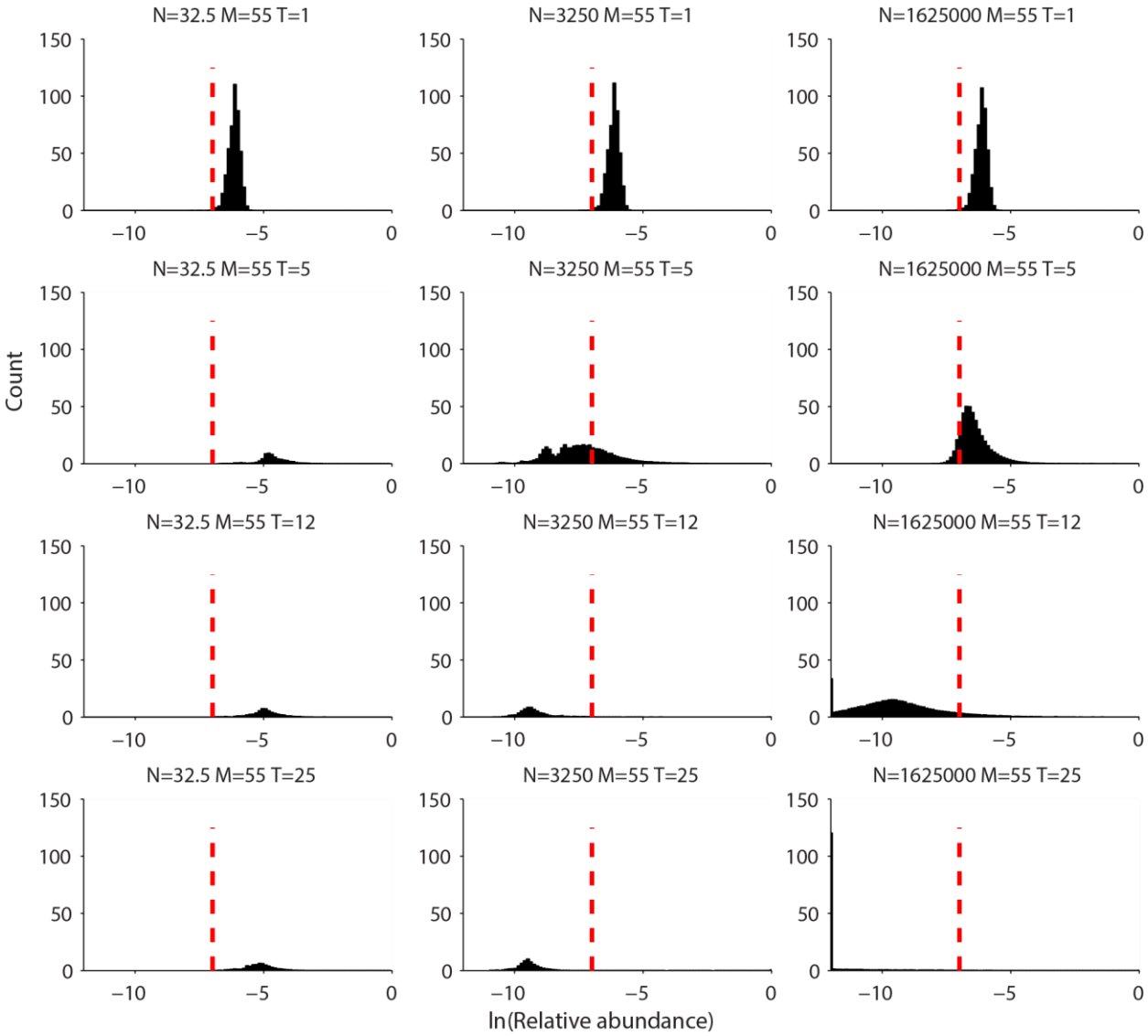
3e) Species relative abundance distributions



Supplementary figure 21. Experimental species relative abundance histograms. In each subplot title, N indicates the bottleneck size, M indicates the number of immigrants, and T indicates the time point. The red dashed line indicates the detection limit, below which species are not considered present (Supplementary section 1e).



Supplementary figure 22. Neutral species relative abundance histograms. In each subplot title, N indicates the bottleneck size, M indicates the number of immigrants, and T indicates the time point. The red dashed line indicates the detection limit for the experiment, below which species are not considered present. Bin counts are means of 100 trials.



Supplementary figure 23. Species relative abundance histograms for a selective model. In each subplot title, N indicates the bottleneck size, M indicates the number of immigrants, and T indicates the time point. The red dashed line indicates the detection limit for the experiment, below which species are not considered present. Bin counts are means of 100 trials.

4) Stochasticity

Stochasticity in the number of cells from one round to the next can allow some lineages to reach high frequencies in neutral models. One way to rule out a neutral model is to estimate this stochasticity and show that it is unable to explain the rapid loss of diversity in the experiments.

If each cell has fluctuations in growth that are independent, the variance after one round of growth and dilution should scale as

$$\mathrm{Var}[n] = c\mathrm{E}[n] \tag{12}$$

where n is the number of cells immediately after the bottleneck and immigration and c is the variance per cell. This results in variance in frequency of

$$\mathrm{Var}[f] = \frac{c}{N_t}\mathrm{E}[f] \tag{13}$$

Where $N_t = N + M$ is the total population size after the bottleneck and immigration and $f \equiv n/N_t$.

There will be additional noise because of the library preparation and sequencing. A common form of this noise[13] is

$$\mathrm{Var}[f] \approx \left(\frac{1}{R} + \frac{1}{A}\right)\mathrm{E}[f] \tag{14}$$

The first term represents the sequencing noise associated with R total reads. If there is a single probability that a barcode is successfully sequenced then the noise would have a binomial distribution which we approximate as Poisson ($\mathrm{Var}[r] = \mathrm{E}[r]$ which implies $\mathrm{Var}[f] = \mathrm{E}[f]/R$). The second term represents the library preparation noise. It describes the multiple rounds of PCR as a sampling of A cells that are then amplified up. This introduces errors as $\mathrm{Var}[f] = \mathrm{E}[f]/A$.

4a) Sample preparation and sequencing noise

For the largest bottleneck sizes with $N = 3 \times 10^7$ and $N = 2 \times 10^6$ (Experiments E11 & E12), the number of cells is much larger than the number of reads, $R = 1 \times 10^5$. Unless the within experiment stochasticity, c, is very large, the variance in reads should be dominated by the sample preparation and sequencing noise. So we estimate this noise by determining how $\mathrm{Var}[f]$ scales with $\mathrm{E}[f]$.

We can estimate the expected frequency $\mathrm{E}[f_i]$ from the frequency at the previous time point, f_{i-1}. We focus on neutral lineages so

$$\mathrm{E}[f_i|f_{i-1}] = e^{-\mu} f_{i-1} \tag{15}$$

Where $e^{-\mu}$ accounts for the selective pressure due to any adaptive lineages. A complication is that f_{i-1} itself is a noisy estimator of the true frequency, $\hat{f}_{i-1}$, so we must account for this in the variance. The true frequency is

$$\hat{f}_{i-1} = f_{i-1} \pm \sqrt{\left(\frac{1}{R_{i-1}} + \frac{1}{A}\right) f_{i-1}} \tag{16}$$

The frequency at the next time point is then

$$f_i = e^{-\mu}\hat{f}_{i-1} \pm \sqrt{\left(\frac{1}{R_i} + \frac{1}{A}\right) f_i} \tag{17}$$

The combined noise in f_i and $\hat{f}_{i-1}$ gives

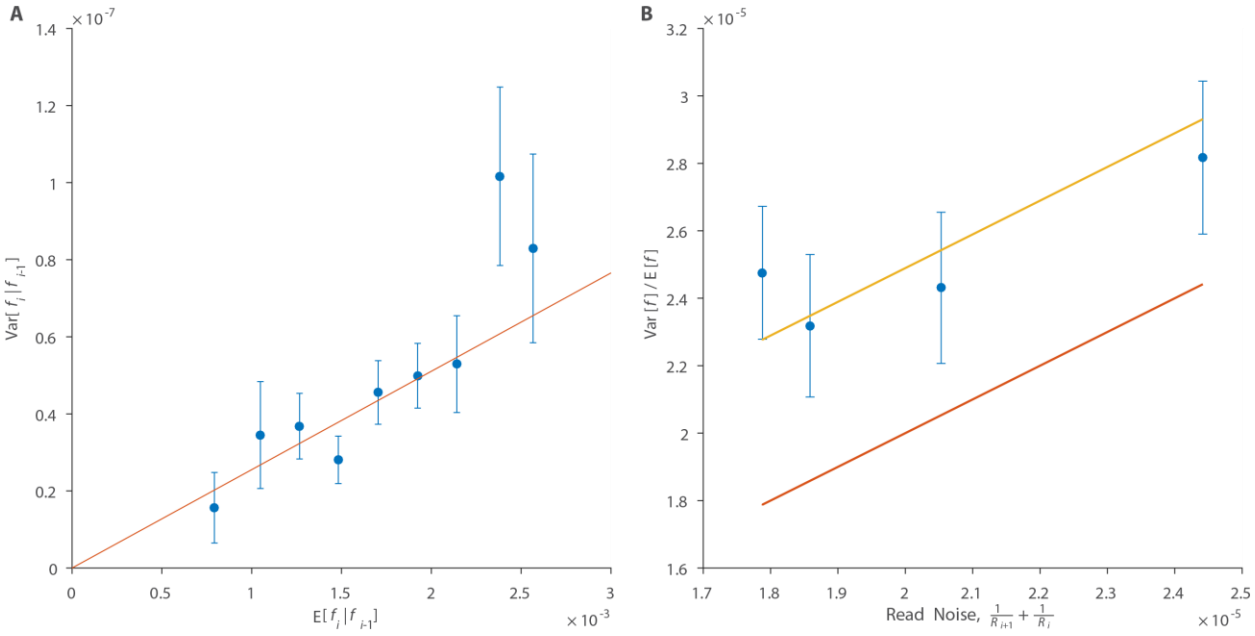
$$\mathrm{Var}[f_i|f_{i-1}] = \left(\frac{1}{R_i} + \frac{e^{-\mu}}{R_{i-1}} + \frac{1+e^{-\mu}}{A}\right)\mathrm{E}[f_i|f_{i-1}] \tag{18}$$

For the large bottlenecks the factor of $e^{-\mu}$ is close to one and its effect on our estimate of A is small compared to other sources of uncertainty.

The scaling relation in equation 18 can be used to estimate $\frac{1}{A}$ by comparing to the slope, $\mathrm{Var}[f]/\mathrm{E}[f]$. For each time point, we bin the frequencies, evaluate the variance, and determine the slope by linear regression. This slope should be linear in the read noise, $\frac{1}{R_{i+1}} + \frac{1}{R_i}$, with an offset that depends on $\frac{1}{A}$. Supplementary figure 24 shows the results of this process. The final estimation of the library preparation noise gives

$$A = (4\pm1) \times 10^5 \tag{19}$$

This value is several times larger than the typical number of reads, 1 x 10^5, which agrees with the intuition that A represents the size of some sampling.



Supplementary figure 24. Determination of the library preparation and sequencing noise for experiment E11. A) shows how the variance scales linearly with the expected frequency for a single time point. Also shown is the best fit line with slope determined by linear regression. B) shows the slope for four different time points. The red line plots the idealized read noise, $\frac{1}{R_{i+1}} + \frac{1}{R_i}$, for comparison. The yellow line is offset by the best fit value for $\frac{2}{A}$.

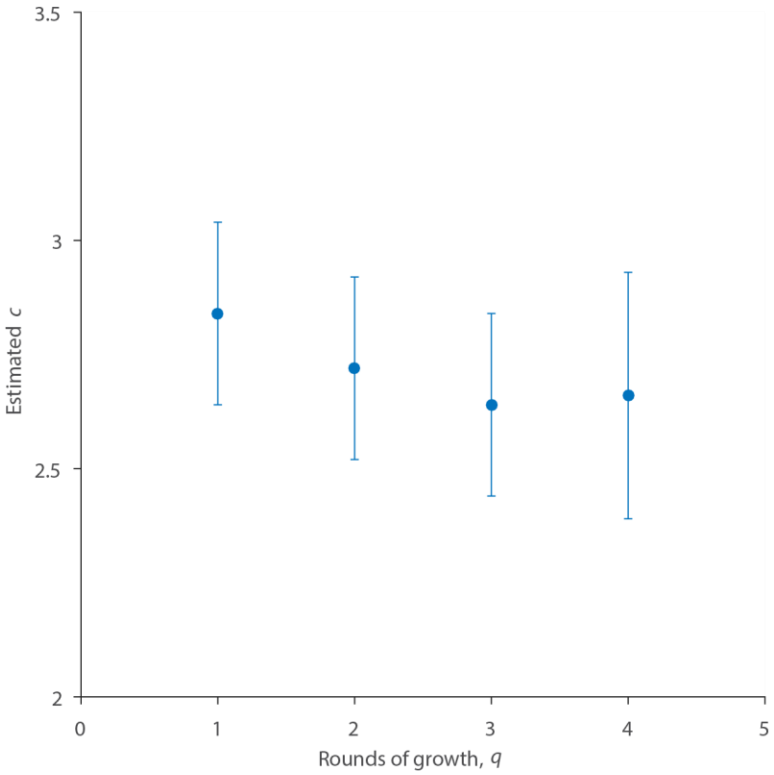
4b) Biological noise

At lower bottleneck sizes, the within experiment stochasticity due to growth and dilution becomes larger in comparison to the library preparation and sequencing noise. Experiment E10 with $N = 3 \times 10^4$ represents a nice window in the data for which the biological noise is large enough to be measured and the barcodes remain at reasonable frequencies for several time points.

Unlike the library preparation and sequencing noise, the within experiment stochasticity will add up after several rounds of growth and dilution. We can infer this stochasticity by seeing how the variance depends on the number of rounds, q, that have elapsed. In the simple case of zero mean fitness ($\mu = 0$), the variance would add linearly as $\mathrm{Var}[f] \approx q \frac{c}{N} \mathrm{E}[f]$.

If $\mu \neq 0$ we must consider how the contribution from earlier rounds scales. Growth in round i will introduce fluctuations in frequency of order

$$\delta f_i \approx \sqrt{\frac{c}{N_t} f_i} \tag{20}$$



Supplementary figure 25. Estimates of the variance in cell number at the bottleneck, c, as a function of the number of rounds of growth, q.

The resulting fluctuations in round q will be

$$\delta f_i \approx e^{-\mu_{i\to q}} \sqrt{\frac{c}{N_t} f_i} = \sqrt{e^{-\mu_{i\to q}} \frac{c}{N_t} \mathrm{E}[f_q]} \tag{21}$$

Where $e^{-\mu_{i\to q}}$ is the total drop in frequency from round i to q and therefore $\mathrm{E}[f_q] = e^{-\mu_{i\to q}} f_i$. The overall variance is

$$\mathrm{Var}[f_q|f_0] = \frac{c}{N_t} \left(\sum_{i=0}^{q-1} e^{-\mu_{i\to q}}\right) \mathrm{E}[f_q|f_0] + (Prep + Read\ Noise) \tag{22}$$

As before, we can determine $\mathrm{Var}[f]/\mathrm{E}[f]$ for different q and get multiple estimates of c as shown in Supplementary figure 25. For experiment 10, we find that

$$c = 2.7 \pm 0.1 \tag{23}$$

This quantity describes the variance in the number of cells at the bottleneck. Given n_{i-1} cells at the bottleneck, the variance at the next bottleneck is $\mathrm{Var}[n_i|n_{i-1}] = c\mathrm{E}[n_i|n_{i-1}]$.

The within experiment stochasticity, $c \approx 3$, is the sum of the effects due to dilution and the biological noise due to bacterial growth. For dilution, $c_{dil} = 1$ since $\mathrm{Var}[n] = n$ for a Poisson dilution. This leaves $c_{bio} \approx 2$. This biological noise could be attributed to different sources, either fluctuations in lag phase, exponential growth, or cell death in stationary phase. The biological noise could differ some across the different bottleneck sizes because each experiment will involve different amounts of time spent in each phase of growth.

The fact that c is order one means that the effective population size N_e is on the order of the bottleneck size N_t. The typical neutral theory result without immigration is that the coalescent time (i.e. the time to lose diversity) would be on the order of N_e rounds of growth-dilution. This is much longer than what is observed in the experiments which lose diversity within 5 to 10 rounds. For neutral models with immigration, stochasticity must overcome the influx of new migrants in order for a single species to reach a large frequency. This requires $N_e m \ll 1$ where m is the fraction of migrants. This is not the case for our experiments since $m = \frac{M}{M+N}$ so $N_e m \sim M = 55$.

4c) Small bottlenecks

At small bottleneck sizes the variance in frequency is dominated by the biological noise. This allows us to estimate the distribution of lineage sizes starting from a single cell. We can identify barcodes that are likely to start the cycle of growth at a single cell by determining which barcodes correspond to new immigrants.

Each barcode is immigrated via a Poisson dilution with a small mean, $Mf_m \approx M\frac{1}{S} \approx \frac{55}{456} \ll 1$. Given that a barcode is immigrated it is likely to have a single cell since

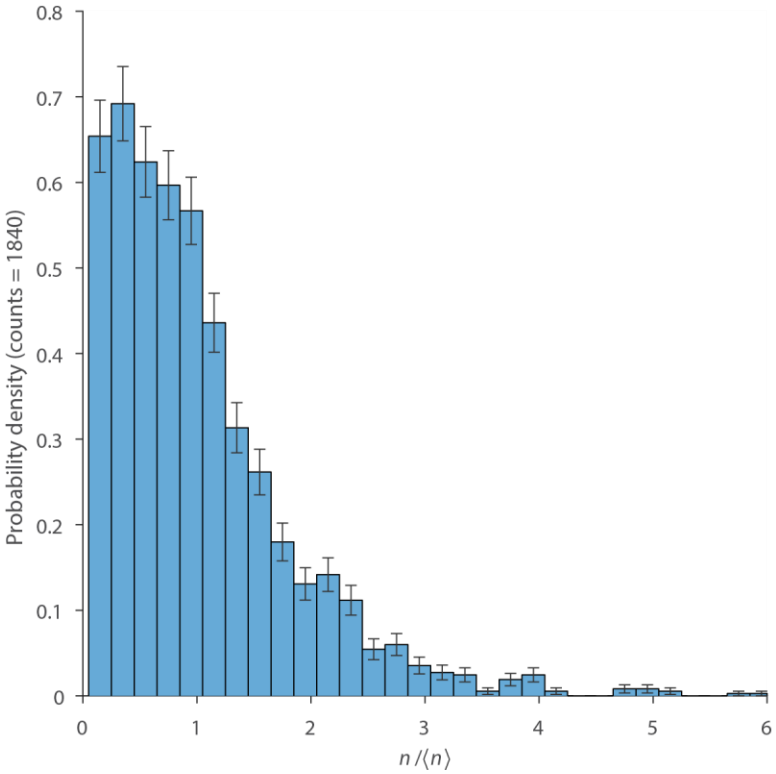
$$p(n = 1|n > 0) = \frac{e^{-Mf_m}}{1-e^{-Mf_m}}(Mf_m) \approx 0.94 \tag{24}$$

We know that a barcode corresponds to a migrant if it is not present at the previous time point. Instead of using frequency, we consider the effective number of cells at the bottleneck as $n \equiv N_t f$. In these units, we expect the variance to be order one. A few early divisions would make the lineage appear to grow up from $n = 2$ or 3 instead of a single cell.

We collected the data from experiments E2, E3, and E4 which all had a bottleneck size of $N = 32$ so $N_t = 87$. Since the bottleneck size is so small, fluctuations in the lineage size of non-migrants could change all the migrant frequencies by some multiplicative factor. For this reason, we normalize the frequencies as

$$\hat{n} \equiv \frac{n}{\langle n \rangle} = \frac{f}{\langle f \rangle} \tag{25}$$

Where $\langle n \rangle$ is the average for a single time point. This quantity is unchanged by an overall multiplicative factor, $f \to \alpha f$. The resulting distribution is shown in Supplementary figure 26.



Supplementary figure 26. The distribution of migrant lineage sizes. The variable $n/\langle n \rangle$ is the effective cell number at the bottleneck.

The distribution has a roughly exponential tail with some outliers with $\hat{n} > 8$. The variance of this distribution is an estimate of the biological noise:

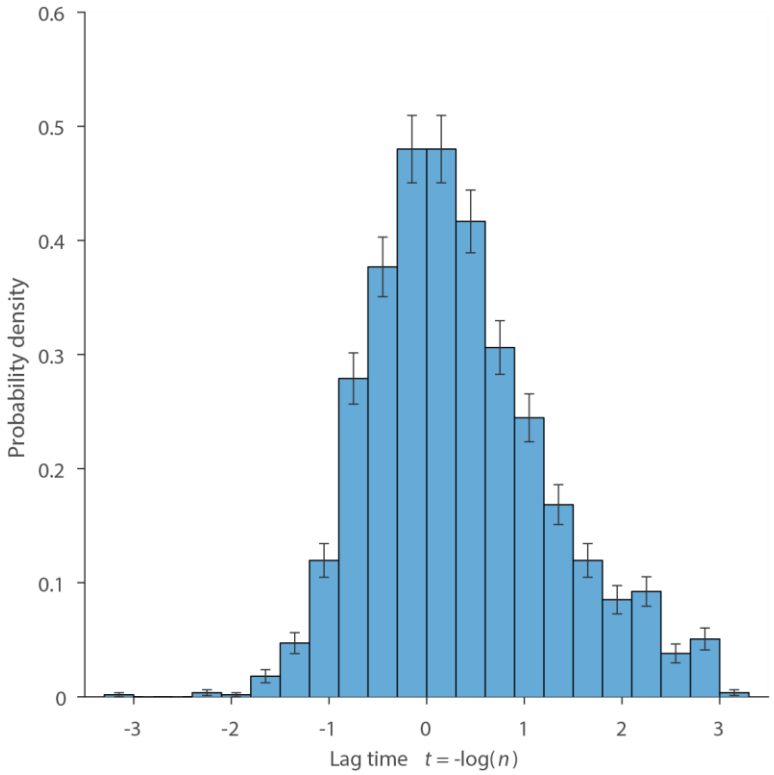
$$c_{bio} \approx \mathrm{Var}[n] = \langle n \rangle^2 Var[\hat{n}] = 0.7 \tag{26}$$

Where we have estimated $\langle n \rangle \approx 1$ since most migrants enter with $n = 1$. This is smaller than the previous estimate of $c_{bio} \approx 2$ for larger bottlenecks, consistent with other experiments which have shown higher variance in lag times for cells that spend longer times in stationary phase[14].

A natural interpretation of this distribution is that it arises from the variation in lag times. We can define an effective lag time, t_l, such that the final lineage size is $n_f = e^{T - t_l}$ and the frequency is $f = e^{-t_l}/N_t$. Therefore we can estimate

$$t_l = -\log(\hat{n}) \tag{27}$$

where we have set the units so that the growth rate is one. The distribution of effective lag times is shown in Supplementary figure 27. This effective lag time should be dominated by the true lag time distribution but will also include effects due to the variation in the next few division times. Each later division contributes about half as much variance as the previous division[13]. The standard deviation in units of the doubling time is $\frac{\sigma}{t_{doubling}} = 1.3$. Other studies using optical imaging have found that differences in lag time are on the order of the doubling time[15,16].



Supplementary figure 27. The distribution of effective lag times.

5) Distribution and nature of selective advantage

While it is not possible to determine the mechanistic causes of the selective advantages without more extensive experiments, which were not the focus of this work, here we attempt to make inferences about the nature of the fitness differences from the existing data.

We extracted the effective relative fitness, F, from the experimental data by fitting an exponential to the initial period of growth for each species whose relative abundance exceeded 10% at any point in each experiment. At the initial stages, before the fit species rise to high abundance, we assume the background population fitness is neutral and approximate $\frac{f_{fit}}{f_{neutral}}$ as $\frac{f_{fit}}{f_{neutral}+f_{fit}}$. To obtain the true relative Malthusian fitness from the apparent fitness we note that immigration decreases the true relative fitness by $-\delta = \ln\frac{N}{N+M}$ so we apply the correction, $R = F - \ln\frac{N}{N+M}$. Note that this analysis

works best when the population contains primarily neutral members, which is the case for most species which appear to rise in abundance from a fitness advantage in these experiments. However, species 277 in experiment E12 and species 39 in experiment E11 seem to rise in abundance at much later time points against a background with fitness higher than neutral. This method of fitness extraction would underestimate the size of their fitness advantages, and, as a result, we do not include these two species in the below analysis. Exponential expansion at later time points suggests these species may have acquired a fitness advantage by adapting during the experiment (ex: via *de novo* mutation). Adaptation during the experiment is not accounted for in our simple selective model incorporating preexisting fitness differences. We did not construct the more complex model that would be required to capture this effect given the low fraction of adaptive clones (2/20) which seem likely to have acquired fitness advantages during the experiment, and the success of the simple model at capturing key features of the data.

First we ask, were the selective advantages present before the experiment began?

It seems likely that at least some of the fitness differences were present in the original population. Many of the same species show similar fitness advantages in different experiments, for example, species 446 and 454 exhibit fitness advantages in the experiments with the largest four bottleneck sizes (Supplementary table 1 gives extracted fitness values for different species and experiments). That both species started equivalent to the other 456 species and developed fitness advantages by chance in all four independent experiments is exceedingly unlikely given the small fraction of species which have measured fitness advantages of this magnitude. We cannot rule out that these differences arose in the immigrant pool during a round of growth and were simultaneously carried to multiple experiments, but this also seems highly unlikely given the small number of immigrants and the limited opportunity for the immigrant population to adapt (grown fresh each round). We note that it is difficult to eliminate all preexisting fitness differences, since growth is required to generate the initial library before the experiment begins. A single colony contains around 10^9 cells and starts from a single cell. The required rounds of replication, when paired with the genome size and mutation rate, suggest there are bound to be genetic differences in the initial pool of cells, some of which likely lead to differences in fitness at the start of the experiment. (see Supplementary section 1a for discussion of library creation)

Where in the growth cycle does the fitness advantage occur?

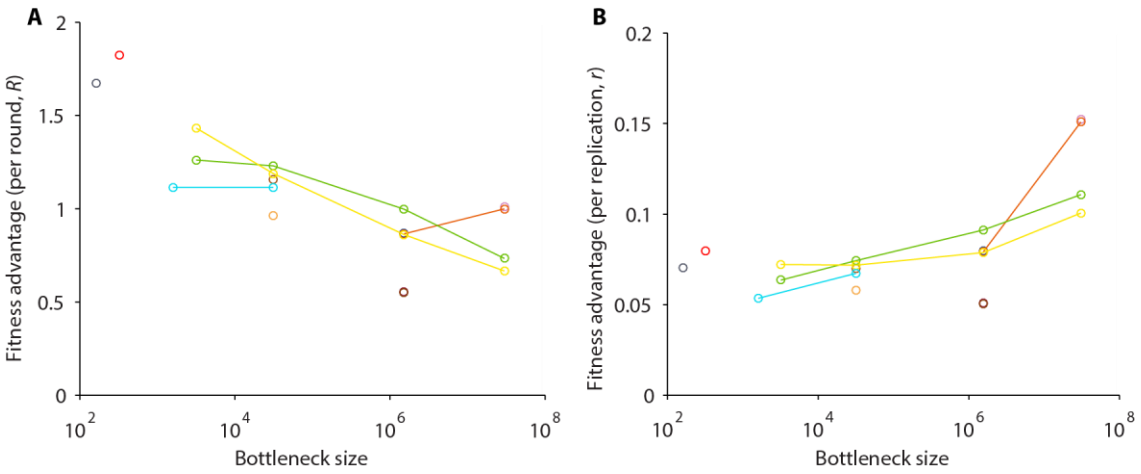
If the fitness advantage was entirely in the exponential growth phase, then the per round relative Malthusian fitness, R, would equal the number of replication cycles times the per replication relative Malthusian fitness, r, $R = xr$. The number of rounds of replication for each experiment are given by $x = \log_2 \frac{N_f}{N+M}$, where N_f is the final number of cells. Given the wide range of bottleneck sizes tested here we can assess whether the per round fitness advantage is consistent with the fitness advantage being in the exponential phase. If we assume the per replication relative fitness is the same, regardless of experiment, the per round relative fitness should decrease approximately linearly with the log of the bottleneck size, which is indeed what we see from our extracted relative fitness values up to around a bottleneck size of 10^6 cells, above which it seems to level out (Supplementary figure 28). This suggests that at most bottlenecks tested here the selective advantage comes from the exponential phase, but at

the largest bottlenecks a different portion of the growth cycle such as lag or stationary phase may contribute to the selective advantages.

This is perhaps not overly surprising, given that cells undergoing different bottlenecks spend different amounts of time in different portions of the growth cycle. For example, cells in the smallest bottleneck go through over 20 doublings and spend less time in stationary phase than cells in the largest bottleneck which go through less than 10 doublings and spend far less time in exponential phase. In other words, even though the media is the same, and the time between passaging is the same. The different bottleneck sizes might be thought of as distinct environments where a beneficial mutation at one size is not necessarily beneficial at all bottleneck sizes.

Further supporting the notion that the type of fitness advantage might be particular to the bottleneck size, we find that certain species seem to have advantages only in a certain range of bottleneck sizes. For example, species 446 and 454 exhibit fitness advantages only in the largest four bottleneck sizes while species 330 has an advantage only in two intermediate bottleneck sizes, though more experiments would be required to confirm these preliminary findings and elucidate the mechanistic details.

We encompass these observations in the model with fitness advantages by holding r constant across bottleneck sizes, but varying R according to $R = xr$. We select r for each species randomly from an exponential distribution with mean, $\lambda = 0.0127$, which yields a typical maximum r (across 456 species) of 8.5%, to match the mean maximum r measured in the different experiments.



Supplementary figure 28. Extracted relative fitness. Here we plot the relative Malthusian fitness against the bottleneck size on a A) per round basis, *R,* and a B) per replication basis, *r*. Colors indicate the species identity and are consistent with Supplementary figure 12, Figure 2 C-E, and Figure 3 D&G, and species which appear more than once are connected with lines.

Species number	Bottleneck size, *N*	Per round relative fitness, *R*	Per replication relative fitness, *r*	Max per replication relative fitness
243	163	1.69	0.071	0.071
173	325	1.84	0.080	0.080
330	1625	1.12	0.054	0.054
446	3250	1.44	0.072	0.072
454	3250	1.27	0.064	
287	32500	0.97	0.058	0.075
330	32500	1.12	0.068	
381	32500	1.16	0.070	
446	32500	1.20	0.072	
454	32500	1.24	0.075	
335	1625000	0.56	0.051	0.092
366	1625000	0.87	0.079	
438	1625000	0.55	0.051	
446	1625000	0.87	0.079	
454	1625000	1.00	0.092	
446	32500000	0.67	0.101	0.151
454	32500000	0.74	0.111	
366	32500000	1.01	0.151	

Supplementary table 1. Extracted Malthusian relative fitnesses.

6) Table of experimental data

Experiments were run at the dilutions presented in Supplementary table 2. For a full table of read counts for each species at each experimental time point for each experiment, see the appendix.

Experiment Title	Bottleneck size, *N*, (# cells)	Immigration Rate, *M*, (# cells/day)	Duration
E1	3.25	55	25
E2	32.5	55	25
E3	32.5	55	8
E4	32.5	55	8
E5	162.5	55	25
E6	325	0	25
E7	325	55	25
E8	1625	55	25
E9	3250	55	25
E10	32500	55	25
E11	1625000	55	25
E12	32500000	55	25
E13	variable*	55	*
E14	variable†	55	†

Supplementary table 2. Key to experimental data. Experimental parameters for each experiment are as listed. * E13 comes from E2; the first ten rounds match. After round ten, the bottleneck size was increased and held at 3250. † E14 comes from E2; the first ten rounds match. After round ten, the bottleneck size was gradually increased over six rounds according to the sequence (*N* = 70, 151, 325, 700, 1510, 3250), then held at 3250 for the remainder of the rounds.

7) References

1. Bernard, P., Gabarit, P., Bahassi, E. M. & Couturier, M. Positive-selection vectors using the F plasmid ccdB killer gene. *Gene* **148,** 71–74 (1994).
2. Fu, G. K., Hu, J., Wang, P.-H. & Fodor, S. P. A. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 9026–31 (2011).
3. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9,** 72–74 (2011).
4. Vollmers, C., Sit, R. V, Weinstein, J. a, Dekker, C. L. & Quake, S. R. Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 13463–8 (2013).
5. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similiarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48,** 443–453 (1970).
6. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography*. **17,** (Princeton University Press, 2001).
7. Kimura, M. Diffusion Models in Population Genetics. *J. Appl. Probab.* **1,** 177–232 (1964).
8. Sloan, W. T. *et al.* Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* **8,** 732–740 (2006).
9. Ricklefs, R. E. A comment on Hubbell's zero-sum ecological drift model. *Oikos* **100,** 185–192 (2003).
10. Ricklefs, R. E. The unified neutral theory of biodiversity: Do the numbers add up? *Ecology* **87,** 1424–1431 (2006).
11. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* **424,** 1035–1037 (2003).
12. Chave, J., Muller-Landau, H. C. & Levin, S. A. Comparing classical community models: theoretical consequences for patterns of diversity. *Am. Nat.* **159,** 1–23 (2002).
13. Levy, S. F. *et al.* Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519,** 181–6 (2015).
14. Levin-Reisman, I. *et al.* Automated imaging with ScanLag reveals previously undetectable bacterial growth phenotypes. *Nat. Methods* **7,** 737–9 (2010).
15. Métris, A., Le Marc, Y., Elfwing, A., Ballagi, A. & Baranyi, J. Modelling the variability of lag times and the first generation times of single cells of E. coli. in *International Journal of Food Microbiology* **100,** 13–19 (2005).
16. Niven, G. W., Morton, J. S., Fuks, T. & Mackey, B. M. Influence of environmental stress on distributions of times to first division in Escherichia coli populations, as determined by digital-image analysis of individual cells. *Appl. Environ. Microbiol.* **74,** 3757–3763 (2008).