Supplementary Material

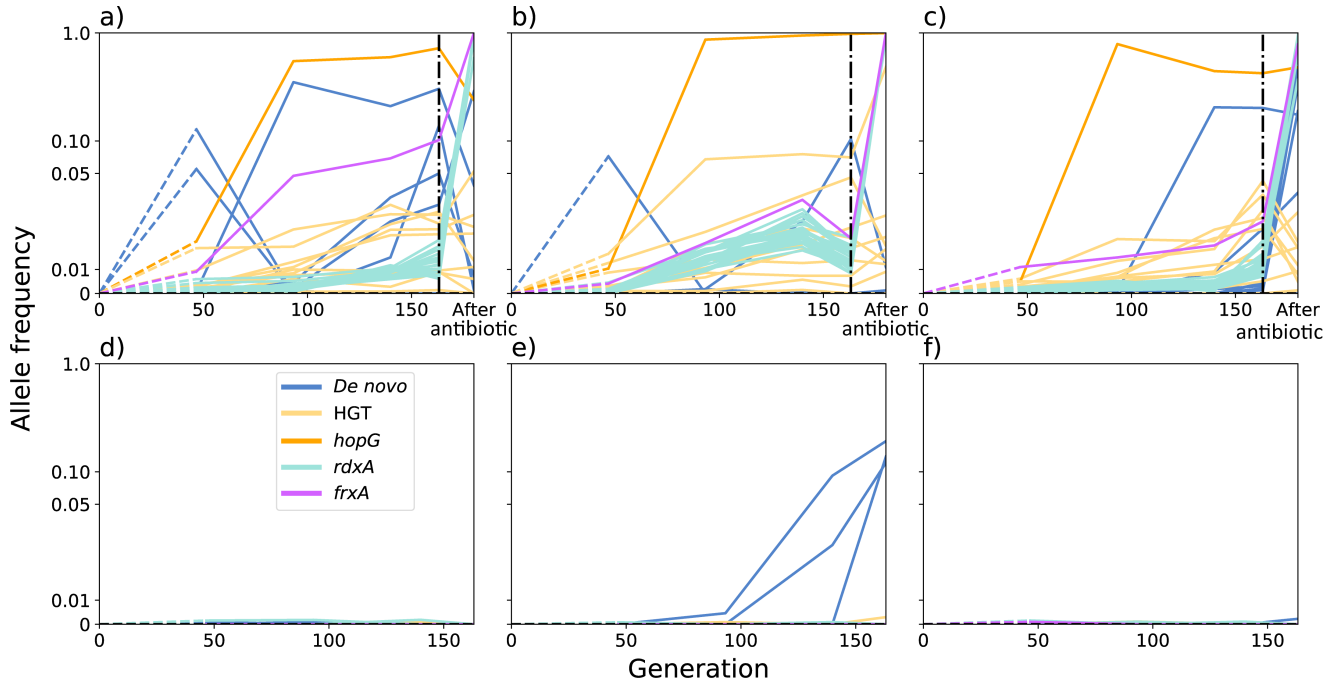


Figure S1. Trajectories of allele frequencies during evolution, as determined by whole-population sequencing. This figure is identical to Figure 2 in the main text, except that standing genetic variants are not shown, and the zeroth-time point is shown for the *de novo* and HGT genetic variants. Each line depicts an individual allele segregating in three HGT (**a-c**) and three non-HGT control (**d-f**) populations plotted on a symlog scale (linear for values below 0.05). The dashed vertical line after 161 generations shows the frequencies of all alleles at the end of the no-antibiotic evolution experiment before exposure to metronidazole (8 μg/ml). The frequencies at the end of the plot show allele frequencies after the addition of metronidazole. The HGT alleles are shown in light orange except for the *rdxA* (turquoise), *frxA* (magenta) and *hopG* (bright orange) HGT alleles. Inactivation of the *rdxA* (turquoise) and *frxA* (magenta) genes is required for metronidazole resistance. Dashed lines show the frequencies of all *de novo* and HGT alleles at time point zero.

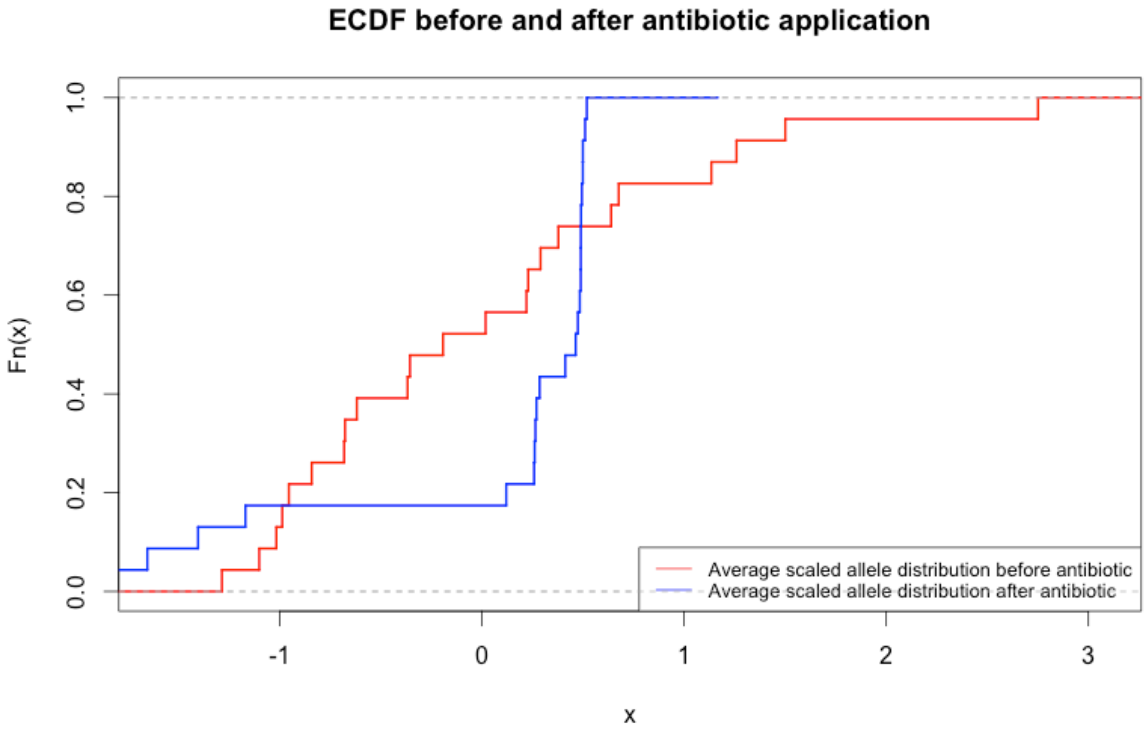
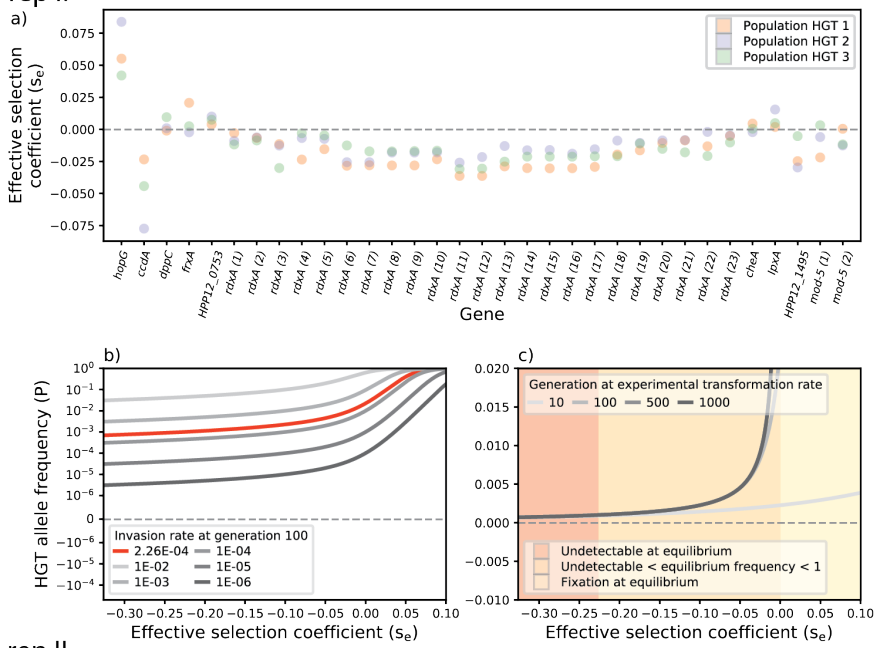
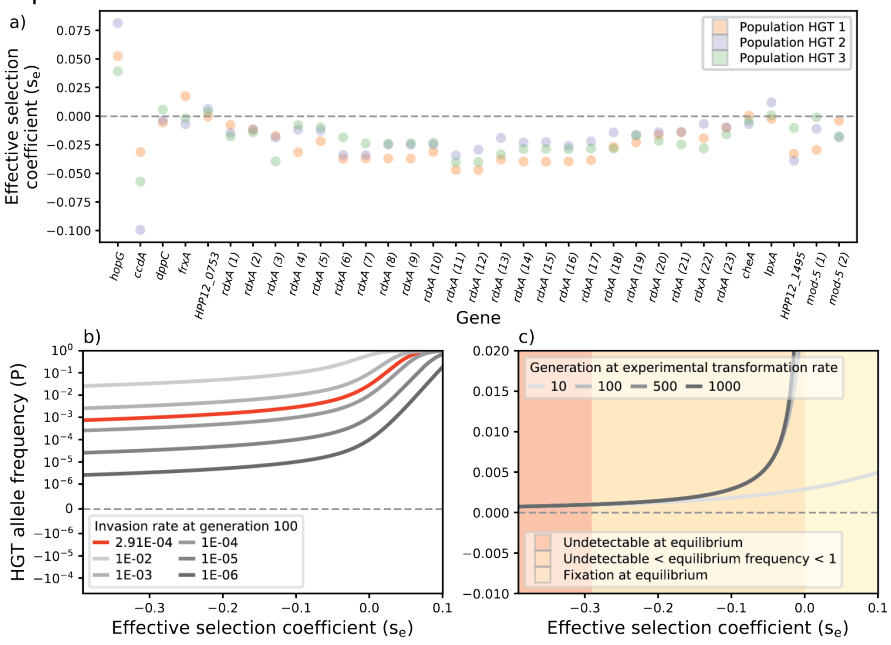


Figure S2. Empirical cumulative distribution function (ECDF) for allele distributions before and after antibiotic selection. This figure corresponds to figure panels 3c and 3d, showing the empirical cumulative distribution function (ECDF) for the scaled *rdxA* allele distributions before and after the addition of antibiotic. The ECDF curves before and after antibiotic selection were found to be significantly different (Kolmogorov-Smirnov test, D criterion = 0.3913, $p = 0.001103$).

rep I.



rep II.



rep III.

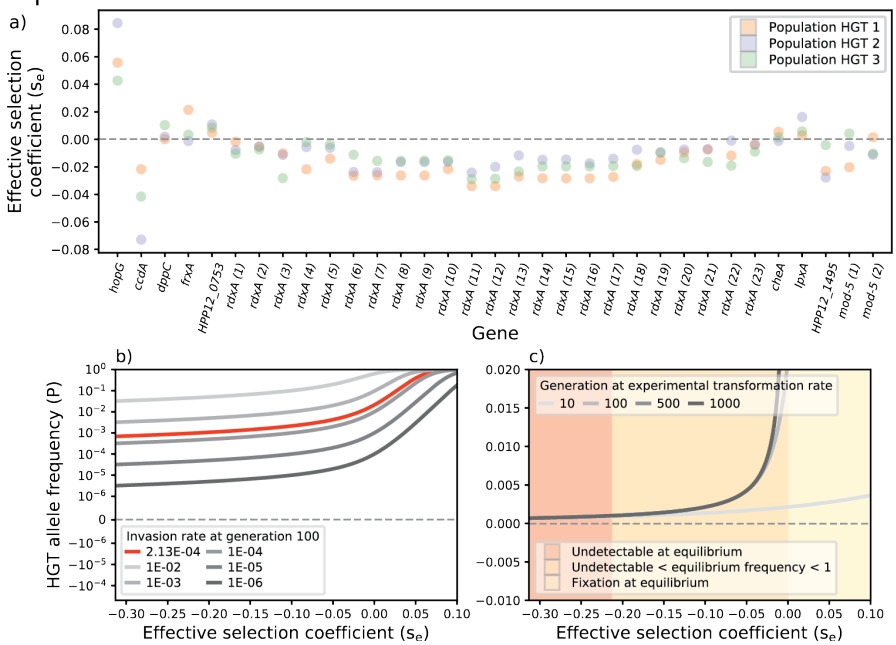


Figure S3. Maintenance of genetic variation with HGT and fitness estimates from sequence data in replicate populations I, II and III. This figure corresponds to figure 4 in the main text, where values were calculated from the average trajectory of the synonymous HGT allele (HPP12_0753) that segregated in each replicate population. This figure shows alternative numerical evaluations of the model, based on each individual replicate population. For each replicate (I-III); panel (**a**) shows the estimates of the effective selection coefficient for donor-derived alleles. Calculations were carried out for each of the three independently evolving populations, HGT rep 1 (orange bar), HGT rep 2 (purple bar) and HGT rep 3 (green bar). Panel (**b**) shows the expected frequency of horizontally-acquired alleles after 100 generations of evolution for several values of γ. The red curve corresponds to the HGT invasion rate (γ) estimated from our experiments. Panel (**c**) shows the expected frequencies of horizontally-acquired alleles after 10, 100, 500 and 1000 generations, given the HGT invasion rate (γ) estimated from the evolution experiment. The shaded areas correspond to three equilibrium frequency states for donor alleles (undetectably rare, detectable and polymorphic, and fixed).

Table S1

Position	Gene	Description	Population(s)	Frequency
24,931	*hopD*	Outer membrane protein HopD	HGT replicate 3	0.0270
35,661	*HPP12_0032*	Hypothetical protein	HGT replicate 1	0.0370
375,321	*HPP12_0359*	Hypothetical protein	HGT replicate 2	0.1040
643,280	*flaA*	Flagellin A	HGT replicate 3	0.2020
814,592	*flaG-2*	Polar flagellin	HGT replicate 1	0.3040
1,001,701	*dadA*	D-amino acid dehydrogenase subunit	control replicate 2	0.1200
1,098,003	*cheA*	Autophosphorylating histidine kinase	HGT replicate 1	0.1380
1,376,228	*HPP12_1299*	Hypothetical protein	HGT replicate 1	0.0500
1,437,437	*mod-3*	Type III R-M system methyltransferase	control replicate 2	0.1380
1,591,063	*HPP12_1502*	Periplasmic competence protein ComH	control replicate 2	0.1920
272,419	*ccdA*	Cytochrome C-type biogenesis protein CcdA	HGT replicate 1	0.0090
			HGT replicate 2	0.0029
			HGT replicate 3	0.0051
807,151	*HPP12_0753*	Hypothetical protein	HGT replicate 1	0.0328
			HGT replicate 2	0.0483
			HGT replicate 3	0.0412

1,099,702	*cheA*	Autophosphorylating histidine kinase	HGT replicate 1	0.0340
			HGT replicate 2	0.0231
			HGT replicate 3	0.0269
1,580,122	*HPP12_1495*	Hypothetical protein	HGT replicate 1	0.0086
			HGT replicate 2	0.0074
			HGT replicate 3	0.0196
1,585,223	*mod-5a*	Type III R-M system methyltransferase	HGT replicate 1	0.0095
			HGT replicate 2	0.0189
			HGT replicate 3	0.0315
1,586,017	*mod-5b*	Type III R-M system methyltransferase	HGT replicate 1	0.0268
			HGT replicate 2	0.0138
			HGT replicate 3	0.0143
1,657,367	*HPP12_1559*	Hypothetical protein	HGT replicate 1	0.0014
260,922	*hopG*	Outer membrane protein HopG	HGT replicate 1	0.7225
			HGT replicate 2	0.9800
			HGT replicate 3	0.4238
315,474	*dppC*	Dipeptide transport system permease protein	HGT replicate 1	0.0247
			HGT replicate 2	0.0275
			HGT replicate 3	0.0469
1,444,863	*lpxA*	UDP-N-acetylglucosamine acyltransferase	HGT replicate 1	0.0291
			HGT replicate 2	0.0704
			HGT replicate 3	0.0347
1,010,298	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0223
			HGT replicate 2	0.0163
			HGT replicate 3	0.0143

1,010,349	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0183
			HGT replicate 2	0.0186
			HGT replicate 3	0.0166
1,010,412	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0145
			HGT replicate 2	0.0138
			HGT replicate 3	0.0073
1,010,524	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0090
			HGT replicate 2	0.0182
			HGT replicate 3	0.0220
1,010,535	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0122
			HGT replicate 2	0.0176
			HGT replicate 3	0.0201
1,010,558	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0077
			HGT replicate 2	0.0084
			HGT replicate 3	0.0138
1,010,560	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0078
			HGT replicate 2	0.0084
			HGT replicate 3	0.0114
1,010,564	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0077
			HGT replicate 2	0.0110
			HGT replicate 3	0.0114
1,010,566	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0077
			HGT replicate 2	0.0110
			HGT replicate 3	0.0115
1,010,570	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0091

			HGT replicate 2	0.0111
			HGT replicate 3	0.0116
1,010,577	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0061
			HGT replicate 2	0.0083
			HGT replicate 3	0.0071
1,010,578	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0061
			HGT replicate 2	0.0096
			HGT replicate 3	0.0072
1,010,583	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0076
			HGT replicate 2	0.0136
			HGT replicate 3	0.0085
1,010,601	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0073
			HGT replicate 2	0.0118
			HGT replicate 3	0.0097
1,010,605	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0072
			HGT replicate 2	0.0119
			HGT replicate 3	0.0097
1,010,607	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0072
			HGT replicate 2	0.0106
			HGT replicate 3	0.0097
1,010,619	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0075
			HGT replicate 2	0.0122
			HGT replicate 3	0.0098
1,010,627	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0104
			HGT replicate 2	0.0165

			HGT replicate 3	0.0099
1,010,643	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0117
			HGT replicate 2	0.0151
			HGT replicate 3	0.0147
1,010,647	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0150
			HGT replicate 2	0.0166
			HGT replicate 3	0.0123
1,010,654	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0166
			HGT replicate 2	0.0168
			HGT replicate 3	0.0111
1,010,679	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0134
			HGT replicate 2	0.0233
			HGT replicate 3	0.0099
1,010,703	*rdxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.0197
			HGT replicate 2	0.0201
			HGT replicate 3	0.0154
696,808	*frxA*	Oxygen-insensitive NADPH nitroreductase	HGT replicate 1	0.1020
			HGT replicate 2	0.0230
			HGT replicate 3	0.0300

Table S1. Frequencies of HGT and *de novo* variants after 161 generations of evolution without antibiotic. *De novo* mutations (blue) evolved during the evolution experiment in both HGT and control lines, but did not fix by the end of the evolution experiments. All but one HGT-derived allele appeared in all HGT replicate populations. HGT mutations listed are either unrelated to antibiotic resistance (orange) or confer resistance through disruptions in the *rdxA* (turquoise) or *frxA* (purple) gene.

Table S2

Generation	Population	Resistance (%)
46	HGT rep.1	0.000572
-	HGT rep.2	0.000984
-	HGT rep.3	0.000648
-	control rep.1	0
-	control rep.2	0
-	control rep.3	0
69	HGT rep.1	0.002277
-	HGT rep.2	0.007767
-	HGT rep.3	0.003064
-	control rep.1	0
-	control rep.2	0
-	control rep.3	0
92	HGT rep.1	0.000000
-	HGT rep.2	0.003448
-	HGT rep.3	0.008649
-	control rep.1	0
-	control rep.2	0
-	control rep.3	0
115	HGT rep.1	0.007387
-	HGT rep.2	0.037812
-	HGT rep.3	0.008656
-	control rep.1	0
-	control rep.2	0
-	control rep.3	0
138	HGT rep.1	0.015221
-	HGT rep.2	0.059899
-	HGT rep.3	0.011566
-	control rep.1	0
-	control rep.2	0
-	control rep.3	0
161	HGT rep.1	0.02397
-	HGT rep.2	0.03515
-	HGT rep.3	0.01305
-	control rep.1	0
-	control rep.2	0
-	control rep.3	0

TableS2:Metronidazole (8 µg/ml) resistance in evolving populations during the evolution experiment, without selection on antibiotic. Populations were plated on agar plates with and without metronidazole to determine frequency of resistance by CFU counts.

Table S3.

Population	GC (CFU/ml)	$GCM_{(0.19)}$ (CFU/ml)	$GCM_{(1)}$ (CFU/ml)	$GCM_{(4)}$ (CFU/ml)	$GCM_{(16)}$ (CFU/ml)
P12 (ancestor)	9.00×10^8 1.40×10^9	no single colonies	no single colonies	no single colonies	no single colonies
HGT rep 1	7.50×10^8 8.00×10^8	-	1.25×10^7 8.50×10^6	3.40×10^5 3.25×10^5	2.65×10^5 1.85×10^5
HGT rep 2	4.50×10^8 8.00×10^8	-	4.50×10^6 1.00×10^7	8.00×10^4 8.00×10^4	7.50×10^4 7.50×10^4
HGT rep 3	1.25×10^9 9.50×10^8	-	3.50×10^7 6.00×10^7	2.00×10^5 2.25×10^5	7.00×10^4 4.50×10^4
control rep 1	6.00×10^8 9.50×10^8	no single colonies	no single colonies	no single colonies	20 30
control rep 2	5.00×10^8 No single colonies	no single colonies	no single colonies	no single colonies	no single colonies
control rep 3	1.30×10^9 1.30×10^9	no single colonies	no single colonies	5 10	no single colonies

Table S3. Metronidazole resistance in evolved populations. To determine the frequency of metronidazole resistant colonies in control and HGT populations after 161 generations of evolution, cells were plated onto GC agar plates (See Methods) containing metronidazole (GCM), where the subscripts indicate the concentration of metronidazole in μg/ml. Numbers are counts for each replicate. We found a significant difference in level of resistance between HGT populations and non-HGT controls (Mann-Whitney U Test, $p < 0.05$).

Table S4. Primer sequences used for amplification of *rdxA* and *frxA* genes.

Primer	Oligo sequence
RdxA_FWD	5'-TTGCTCGGACTCATGGAATTGC-3'
RdxA_REV	5'-AGAGAGCCAGATAGCCAAATGG-3'
FrxA_FWD	5'-GCAGGAGAGGCGATAAAGTTGC-3'
FrxA_REV	5'-TCTTTGTCCGTGTCTTCAATGC-3'

APPENDIX

Novozhilov et al., (1) developed a single-locus model with two alleles (allele 1 and allele 2), with unidirectional mutation and unidirectional horizontal gene transfer (HGT). Below, we will first extend their single-locus model to allow for bidirectionality of mutation and HGT, and temporal fluctuations in "invasion", which in the current empirical study should be proportional to the concentration of donor DNA in the medium of HGT populations. Building on the single-locus model, we subsequently explore the effects of partial linkage between alleles within the donor strain, and characterize the evolutionary dynamics of a pair of partially linked loci experiencing both HGT and epistatic fitness interactions.

Single locus model with bi-directional mutation and HGT

We follow the evolutionary dynamics of a single locus with two allele types, A_1 and A_2, and individuals of the population are haploid. Allele A_1 is carried by individuals and DNA from the "donor" strain, and p and q represent the allele frequencies of A_1 and A_2 in our focal population. We suppose, that the focal population is exponentially growing with population size of $N = n_1 + n_2$, where n_1 and n_2 are the numbers of individuals carrying alleles 1 and 2, respectively. The population dynamics of individuals carrying each of the two allele types are described by the differential equations:

$$\frac{dn_1}{dt} = m_1 n_1 - u_1 n_1 + u_2 n_2 + \gamma N - \theta_1 n_1 \frac{n_2}{n_1 + n_2} + \theta_2 n_2 \frac{n_1}{n_1 + n_2}$$

$$\frac{dn_2}{dt} = m_2 n_2 - u_2 n_2 + u_1 n_1 + \theta_1 n_1 \frac{n_2}{n_1 + n_2} - \theta_2 n_2 \frac{n_1}{n_1 + n_2}$$

where γ represents the "invasion rate" of donor alleles, which in the context of the experiment represents the rate with which donor DNA is incorporated into the genomes of individuals from the focal population; m_i is the Malthusian fitness, u_i is the mutation rate, and θ_i is the "infection rate" (*i.e.*, rate of HGT conversion), of individuals carrying allele i. The rate of frequency change for A_1 in the focal population is:

$$\frac{dp}{dt} = \frac{d}{dt}\left(\frac{n_1}{n_1 + n_2}\right) = \frac{1}{N}\left(\frac{dn_1}{dt} - p\left(\frac{dn_1}{dt} + \frac{dn_2}{dt}\right)\right)$$
$$= p(m_1 - \bar{m}) + u_2(1 - p) - u_1 p + (1 - p)\gamma + p(1 - p)(\theta_2 - \theta_1)$$

where $p = n_1/N = 1 - n_2/N$, and $\bar{m} = p_1 m_1 + p_2 m_2$. Letting $m_1 - m_2 = s$, where s is the selection coefficient for the A_1 allele, and assuming that the mutation rate is sufficiently small given the timescale of the experiment that it can be ignored, the above simplifies to:

$$\frac{dp}{dt} = (s + \theta_2 - \theta_1)p(1 - p) + (1 - p)\gamma = s_e p(1 - p) + (1 - p)\gamma$$

where $s_e = s + \theta_2 - \theta_1$ is the effective selection coefficient in favour (when $s_e > 0$) or against the A_1 allele (when $s_e < 0$); s_e takes into account effects of the allele on population growth (s) and any biases in the rates of HGT between individuals of the focal population with different genotypes (biases arise when $\theta_2 \neq \theta_1$).

A stability analysis of the differential equation shows that there are three possible equilibria for the locus:

1. $\hat{p} = 1$, which is stable when $\gamma + s_e > 0$ (*i.e.*, the invasion rate is strong and/or there is positive selection on the donor allele) and unstable when $\gamma + s_e < 0$ (*i.e.*, purifying selection is stronger than the invasion rate: $|s_e| > \gamma > 0 > s_e$).
2. $\hat{p} = 0$, which is valid when $\gamma = 0$, is stable when $s_e < 0$ and unstable when $s_e > 0$
3. $\hat{p} = -\gamma/s_e$, which is valid (in which case, it will be stable) when $s_e < 0 < \gamma < |s_e|$.

If we assume that the parameters of the model are constant over time (*i.e.*, s, θ_1, θ_2, γ are constant), the general solution for p is:

$$p_t = \frac{(p_0 s_e + \gamma)e^{(s_e+\gamma)t} - \gamma(1-p_0)}{(p_0 s_e + \gamma)e^{(s_e+\gamma)t} + s_e(1-p_0)}$$

where t refers to time in units of generations.

For the special case of a neutrally evolving locus with symmetric rates of HGT ($s = 0$, $\theta_1 = \theta_2$; thus, $s_e = 0$), the above trajectory simplifies to:

$$p_t = 1 - (1-p_0)e^{-\gamma t}$$

Given empirical data on the frequency dynamics of a neutral donor allele, and provided the population is sufficiently large that frequency trajectories are approximately deterministic, we can infer the infection rate of donor DNA into the population (γ) by rearranging the neutral trajectory and solving for γ:

$$\gamma = \frac{1}{t}\ln\left(\frac{1-p_0}{1-p_t}\right) = -\frac{1}{t}\ln(1-p_t)$$

with the final result applying in the case where the population is initially fixed for the recipient allele ($p_0 = 0$).

If γ varies over time, the general solution for the neutral case ($s_e = 0$) becomes:

$$p_t = 1 - (1-p_0)e^{-\int_0^t \gamma(x)dx} = 1 - (1-p_0)e^{-t\langle\gamma_t\rangle}$$

where $\gamma(x)$ is a function describing change in infection rate over time, and $\langle\gamma_t\rangle$ is the average value of γ across the time interval (0, t).

Two-locus model with linkage during invasion

Single-locus models can characterize the dynamics of donor and recipient alleles provided epistasis between loci is negligible and linkage between donor alleles is loose with respect to the rates at which DNA from pairs of alleles are taken up from the medium. To explore the consequences of violating either assumption, we developed a two-locus population genetic model. Each locus is bi-allelic, and each experiences symmetric, bi-directional HGT (*i.e.*, values of θ for a locus are symmetrical between individuals carrying donor and recipient alleles). We assume that mutation rates are sufficiently low during the course of the experiment that we can neglect mutation between alleles. We explore the extreme case in

which invasion events involve pairs of donor alleles rather than either of the two donor alleles singly. Populations are assumed to grow exponentially.

Allele 1 at each locus represents the donor allele, and allele 2 represents the recipient allele. n_{11}, n_{12}, n_{21} and n_{22} represent the number of individuals with the four possible genotypes (11, 12, 21, 22). The population dynamics of the four genotypes are described by:

$$\frac{dn_{11}}{dt} = n_{11}m_{11} + \gamma N + \theta_A n_{21}\left(\frac{n_{11}}{N} + \frac{n_{12}}{N}\right) + \theta_B n_{12}\left(\frac{n_{11}}{N} + \frac{n_{21}}{N}\right) - \theta_A n_{11}\left(\frac{n_{21}}{N} + \frac{n_{22}}{N}\right) - \theta_B n_{11}\left(\frac{n_{12}}{N} + \frac{n_{22}}{N}\right)$$

$$\frac{dn_{12}}{dt} = n_{12}m_{12} + \theta_A n_{22}\left(\frac{n_{12}}{N} + \frac{n_{11}}{N}\right) + \theta_B n_{11}\left(\frac{n_{12}}{N} + \frac{n_{22}}{N}\right) - \theta_A n_{12}\left(\frac{n_{21}}{N} + \frac{n_{22}}{N}\right) - \theta_B n_{12}\left(\frac{n_{11}}{N} + \frac{n_{21}}{N}\right)$$

$$\frac{dn_{21}}{dt} = n_{21}m_{21} + \theta_A n_{11}\left(\frac{n_{21}}{N} + \frac{n_{22}}{N}\right) + \theta_B n_{22}\left(\frac{n_{21}}{N} + \frac{n_{11}}{N}\right) - \theta_A n_{21}\left(\frac{n_{11}}{N} + \frac{n_{12}}{N}\right) - \theta_B n_{21}\left(\frac{n_{12}}{N} + \frac{n_{22}}{N}\right)$$

$$\frac{dn_{22}}{dt} = n_{22}m_{22} + \theta_A n_{12}\left(\frac{n_{21}}{N} + \frac{n_{22}}{N}\right) + \theta_B n_{21}\left(\frac{n_{12}}{N} + \frac{n_{22}}{N}\right) - \theta_A n_{22}\left(\frac{n_{12}}{N} + \frac{n_{11}}{N}\right) - \theta_B n_{22}\left(\frac{n_{21}}{N} + \frac{n_{11}}{N}\right)$$

where m_{ij} is the Malthusian growth rate of genotype ij, θ_A and θ_B describe the rates of HGT at the first and second locus (labelled locus A and locus B, respectively), γN is the invasion rate of the donor alleles at both loci, and $N = n_{11} + n_{12} + n_{21} + n_{22}$ is the total population size.

Letting $p_A = (n_{11} + n_{12})/N$ and $q_A = (n_{21} + n_{22})/N$ be the frequencies of resistant and susceptible alleles at locus A, and $p_B = (n_{11} + n_{21})/N$ and $q_B = (n_{12} + n_{22})/N$ be the frequencies of resistant and susceptible alleles at locus B, the population dynamics simplify to:

$$\frac{dn_{11}}{dt} = n_{11}m_{11} + \gamma N + \theta_A n_{21}p_A + \theta_B n_{12}p_B - \theta_A n_{11}(1 - p_A) - \theta_B n_{11}(1 - p_B)$$

$$\frac{dn_{12}}{dt} = n_{12}m_{12} + \theta_A n_{22}p_A + \theta_B n_{11}(1 - p_B) - \theta_A n_{12}(1 - p_A) - \theta_B n_{12}p_B$$

$$\frac{dn_{21}}{dt} = n_{21}m_{21} + \theta_A n_{11}(1 - p_A) + \theta_B n_{22}p_B - \theta_A n_{21}p_A - \theta_B n_{21}(1 - p_B)$$

$$\frac{dn_{22}}{dt} = n_{22}m_{22} + \theta_A n_{12}(1 - p_A) + \theta_B n_{21}(1 - p_B) - \theta_A n_{22}p_A - \theta_B n_{22}p_B$$

The frequency dynamics of the pair of functional alleles and of linkage disequilibrium between them ($D = p_{11}p_{22} - p_{12}p_{21} = p_{11} - p_A p_B$) are described by:

$$\frac{dp_A}{dt} = \frac{1}{N}\left(\frac{dn_{11}}{dt} + \frac{dn_{12}}{dt}\right) - \frac{p_A}{N}\left(\frac{dn_{11}}{dt} + \frac{dn_{12}}{dt} + \frac{dn_{21}}{dt} + \frac{dn_{22}}{dt}\right)$$

$$\frac{dp_B}{dt} = \frac{1}{N}\left(\frac{dn_{11}}{dt} + \frac{dn_{21}}{dt}\right) - \frac{p_B}{N}\left(\frac{dn_{11}}{dt} + \frac{dn_{12}}{dt} + \frac{dn_{21}}{dt} + \frac{dn_{22}}{dt}\right)$$

$$\frac{dD}{dt} = \frac{1}{N}\left(p_{22}\frac{dn_{11}}{dt} + p_{11}\frac{dn_{22}}{dt} - p_{21}\frac{dn_{12}}{dt} - p_{12}\frac{dn_{21}}{dt}\right) - \frac{2D}{N}\left(\frac{dn_{11}}{dt} + \frac{dn_{12}}{dt} + \frac{dn_{21}}{dt} + \frac{dn_{22}}{dt}\right)$$

Assuming that the rates of HGT are the same at each locus ($\theta = \theta_A = \theta_B$), and noting that $p_{11} = p_A p_B + D$, $p_{12} = p_A q_B - D$, $p_{21} = q_A p_B - D$, $p_{22} = q_A q_B + D$, we have:

$$\frac{dn_{11}}{dt} = N[(p_A p_B + D)m_{11} + \gamma - 2\theta D]$$

$$\frac{dn_{12}}{dt} = N[(p_A q_B - D)m_{12} + 2\theta D]$$

$$\frac{dn_{21}}{dt} = N[(q_A p_B - D)m_{21} + 2\theta D]$$

$$\frac{dn_{22}}{dt} = N[(q_A q_B + D)m_{22} - 2\theta D]$$

$$\frac{dp_A}{dt} = (p_A p_B + D)m_{11} + (p_A q_B - D)m_{12} - p_A\overline{m} + \gamma(1 - p_A)$$

$$\frac{dp_B}{dt} = (p_A p_B + D)m_{11} + (q_A p_B - D)m_{21} - p_B\overline{m} + \gamma(1 - p_B)$$

$$\frac{dD}{dt} = (q_A q_B - D)(p_A p_B + D)m_{11} + (p_A p_B - D)(q_A q_B + D)m_{22} - (q_A p_B + D)(p_A q_B - D)m_{12} - (p_A q_B + D)(q_A p_B - D)m_{21} - 2\theta D + (q_A q_B - D)\gamma$$

where $\overline{m} = (p_A p_B + D)m_{11} + (p_A q_B - D)m_{12} + (q_A p_B - D)m_{21} + (q_A q_B + D)m_{22}$

Defining fitnesses as $m_{11} = m_{22} + s_A + s_B + \varepsilon$, $m_{12} = m_{22} + s_A$, and $m_{21} = m_{22} + s_B$, we have:

$$\frac{dp_A}{dt} = s_A p_A(1 - p_A) + Ds_B + (1 - p_A)(p_A p_B + D)\varepsilon + \gamma(1 - p_A)$$

$$\frac{dp_B}{dt} = s_B p_B(1 - p_B) + Ds_A + (1 - p_B)(p_A p_B + D)\varepsilon + \gamma(1 - p_B)$$

$$\frac{dD}{dt} = D(1 - 2p_A)s_A + D(1 - 2p_B)s_B + (q_A q_B - D)(p_A p_B + D)\varepsilon - 2\theta D + (q_A q_B - D)\gamma$$

where s_A and s_B are the fitness effects of donor alleles at locus *A* and *B*, respectively, and ε is the epistatic fitness interaction between donor alleles at both loci.

Linkage disequilibria between linked loci under selection

The dynamics of the donor allele are mediated by build-up of linkage disequilibrium (*D*) between loci. To explore how much disequilibrium to expect between the pair of loci, we carried out a quasi-equilibrium analysis of the two-locus system by first approximating quasi-equilibrium *D* between the loci, and then evaluating the allele frequency dynamics following the attainment of quasi-equilibrium *D* (see chapter 9 of Otto and Day 2007)(2). We first write the selection and epistasis parameters as functions of a term ζ which we assume is small:

$$s_A = \widetilde{s_A}\zeta$$

$$s_B = \widetilde{s_B}\zeta$$

$$\varepsilon = \tilde{\varepsilon}\zeta^2$$

Assuming that *D* converges rapidly to a quasi-equilibrium state (denoted D_{QLE}) relative to the rates of allele frequency change at each locus, we can write D_{QLE} as a polynomial to second order in ζ:

$$D_{QLE} = D_0 + D_1\zeta + D_2\zeta^2 + O(\zeta^3)$$

in which the individual terms are:

$$D_0 = \frac{q_A q_B \gamma}{\gamma + 2\theta}$$

$$D_1 = \frac{q_A q_B \gamma[(q_A - p_A)\widetilde{s_A} + (q_B - p_B)\widetilde{s_B}]}{(\gamma + 2\theta)^2}$$

$$D_2 = \frac{q_A q_B \gamma[(1-2p_A)\widetilde{s_A} + (1-2p_B)\widetilde{s_B}]^2}{(\gamma + 2\theta)^3} + \frac{1}{(2\theta + \gamma)}\left(\frac{q_A q_B \gamma(q_A q_B - p_A p_B)}{\gamma + 2\theta} + \left(\frac{q_A q_B \gamma}{\gamma + 2\theta}\right)^2 + p_A q_A p_B q_B\right)\tilde{\varepsilon}$$

Substituting values of D_0, D_1, and D_2 into the polynomial and dropping terms of $O(\zeta^3)$, we have:

$$D_{QLE} \approx \frac{q_A q_B \gamma}{\gamma + 2\theta} + \frac{q_A q_B \gamma[(1-2p_A)s_A + (1-2p_B)s_B]}{(\gamma + 2\theta)^2}\left(1 + \frac{[(1-2p_A)s_A + (1-2p_B)s_B]}{\gamma + 2\theta}\right) + \frac{\varepsilon}{2\theta + \gamma}\left(\frac{q_A q_B \gamma(q_A q_B - p_A p_B)}{\gamma + 2\theta} + \left(\frac{q_A q_B \gamma}{\gamma + 2\theta}\right)^2 + p_A q_A p_B q_B\right)$$

To isolate effects of epistasis and linkage between donor alleles during invasion and DNA uptake, we can explore two special cases of D_{QLE}. First, in a polymorphic population in which invasion ceases ($\gamma = 0$), the QLE approximation will converge to:

$$D_{QLE} \approx \frac{p_A q_A p_B q_B \varepsilon}{2\theta}$$

in which case the allele frequency dynamics at the pair of loci will be approximately independent of one another provided the rate of HGT between individuals is high relative to the strength of epistatic interaction (see chapter 9 of Otto and Day 2007 for similar results based on discrete-time model with meiotic recombination rather than HGT)(2).

Second, when invasion rates are high relative to the strength of selection, we have:

$$D_{QLE} \approx \frac{q_A q_B \gamma}{\gamma + 2\theta}$$

which illustrates that invasion tends to generate coupling disequilibrium between donor alleles, and which facilitates their removal when donor alleles are deleterious (as is readily confirmed by numerical evaluation of the differential equations presented above).

Evolutionary spread of an epistatic-beneficial donor allele combination

In a population initially fixed for the recipient genotype (A_2 at the A locus, and B_2 at the B locus) and no longer receiving donor genes ($\gamma = 0$), conditions for invasion of donor alleles or genotypes can be determined by a linear stability analysis of the system of differential equations for the two-locus system. The Jacobian matrix for the equilibrium $p_{22} = 1$ is:

	$\frac{\partial}{\partial p_{22}}\left(\frac{dp_{ij}}{dt}\right)$	$\frac{\partial}{\partial p_{12}}\left(\frac{dp_{ij}}{dt}\right)$	$\frac{\partial}{\partial p_{21}}\left(\frac{dp_{ij}}{dt}\right)$	$\frac{\partial}{\partial p_{11}}\left(\frac{dp_{ij}}{dt}\right)$
$\frac{\partial}{\partial p_{ij}}\left(\frac{dp_{22}}{dt}\right)$	$-m_{22}$	$-m_{12}$	$-m_{21}$	$-\theta_A - \theta_B - m_{11}$
$\frac{\partial}{\partial p_{ij}}\left(\frac{dp_{12}}{dt}\right)$	0	$m_{12} - m_{22}$	0	$\theta_A + \theta_B$
$\frac{\partial}{\partial p_{ij}}\left(\frac{dp_{21}}{dt}\right)$	0	0	$m_{21} - m_{22}$	$\theta_A + \theta_B$
$\frac{\partial}{\partial p_{ij}}\left(\frac{dp_{11}}{dt}\right)$	0	0	0	$m_{11} - m_{22} - \theta_A - \theta_B$

which has the following eigenvalues:

$$r_{22} = -m_{22}$$

$$r_{12} = m_{12} - m_{22}$$

$$r_{21} = m_{21} - m_{22}$$

$$r_{11} = m_{11} - m_{22} - \theta_A - \theta_B$$

Selection favours fixation of the recipient genotype (selection disfavours invasion of donor alleles/genotypes) when all four eigenvalues are negative (r_{22}, r_{12}, r_{21}, $r_{11} < 0$). Three conditions can favour invasion of a donor allele or genotype:

- The A_1 donor allele is beneficial ($m_{12} - m_{22} > 0$)
- The B_1 donor allele is beneficial ($m_{21} - m_{22} > 0$)
- The pair of donor alleles, A_1B_1, is beneficial and linkage between them is sufficiently tight that $m_{11} - m_{22} > \theta_A + \theta_B = 2\bar{\theta}$

The final scenario is reminiscent of the condition for a selectively favoured peak shift in classical population genetic models of selection with linkage(3).

References

1. A. S. Novozhilov, G. P. Karev, E. V. Koonin, Mathematical modeling of evolution of horizontally transferred genes. *Mol Biol Evol* **22**, 1721-1732 (2005).
2. S. P. Otto, T. Day, *A biologist's guide to mathematical modeling in ecology and evolution* (Princeton University Press, Princeton, 2007), pp. x, 732 p.
3. J. F. Crow, M. Kimura, Evolution in sexual and asexual populations. *The American naturalist* **99**, 439-450 (1965).