doi:10.1038/nature11879

Supplementary Information: Methods

More detailed experimental protocols (including our modified allelic exchange protocol), all data, all media recipes, all computer code, and miscellaneous information for this project can be found at:

http://depts.washington.edu/kerrpost/Public/RifRampProject

Drug concentrations for the evolution experiment

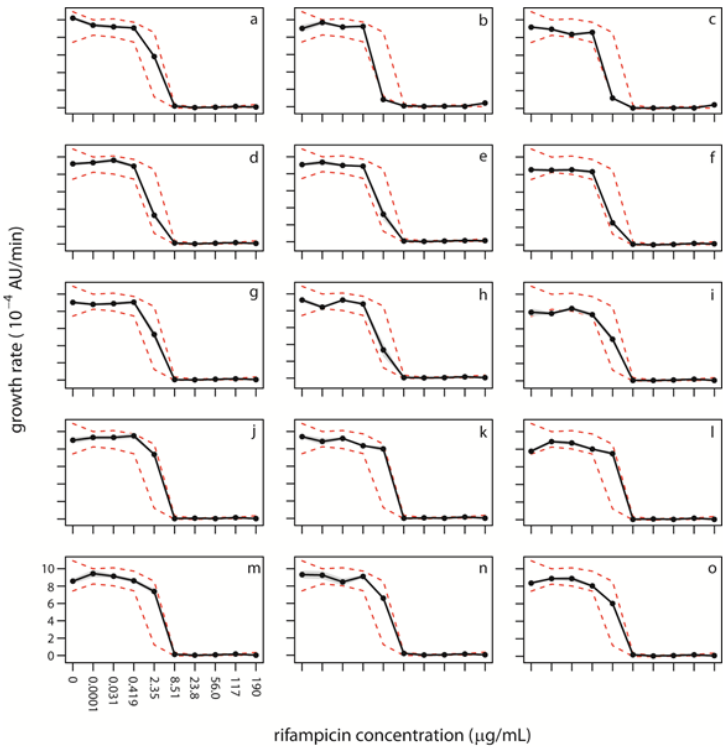
The following table lists rifampicin concentrations at each transfer in each treatment (see also Figure 1).

Transfer	*Gradual*	*Moderate*	*Sudden*
0	***0***	0	0
1	0.00000315	0.000109	190
2	***0.000109***	0.00383	190
3	0.000877	0.0306	190
4	0.00383	0.134	190
5	0.0120	0.419	190
6	***0.0306***	1.07	190
7	0.0674	2.35	190
8	0.134	4.66	190
9	0.244	8.51	190
10	***0.419***	14.6	190
11	0.683	23.8	190
12	1.07	37.2	190
13	1.61	56.0	190
14	***2.35***	81.9	190
15	3.34	117	190
16	4.66	162	190
17	6.35	190	190
18	***8.51***	190	190
19	11.2	190	190
20	14.6	190	190
21	18.8	190	190
22	***23.8***	190	190
23	29.9	190	190
24	37.2	190	190
25	45.8	190	190
26	***56.0***	190	190
27	68.0	190	190
28	81.9	190	190
29	98.0	190	190
30	***117***	190	190
31	138	190	190
32	162	190	190
33	***190***	190	190

Supplementary Table 1. Rifampicin concentration (μg/mL) at each transfer for the three experimental treatments. The *Gradual* treatment reaches the minimum inhibitory concentration (MIC) for the sensitive ancestor midway through the series and the highest concentration of rifampicin on the last transfer. Populations in the *Moderate* treatment reached the MIC at the quarter point and the highest concentration at the midpoint, while populations in the *Sudden* treatment were immediately exposed to the highest rifampicin concentration. The ten concentrations of rifampicin used in the growth and competition assays are given in bold-italic font and were chosen to be evenly spaced across the *Gradual* treatment.

Checking the allelic exchange protocol

For each of the *rpoB* mutants created by our allelic exchange protocol, we reintroduced the ancestral sequence into the mutant and compared this "re-engineered" (R) ancestor with the untouched (U) ancestor. In Supplementary Figure 1, we show the results of spectrophotometric assays for growth rate at a series of different rifampicin concentrations. First, we note that there is variation in growth rates of the U ancestor (red dotted lines in all the plots). The R ancestors behaved similarly to the U ancestor as rifampicin was increased. In nearly all cases, the R growth rates measured lay within the range measured for the U ancestor. In all cases the growth rate of the R ancestor dropped precipitously between 0.419 μg/mL and 8.51 μg/mL of rifampicin, just as it did for the U ancestor. While all competitions occurred between R ancestors and engineered mutants to control for any effects of the allele exchange protocol, the data in Supplementary Figure 1 suggests that the protocol does not introduce large deviations in growth.



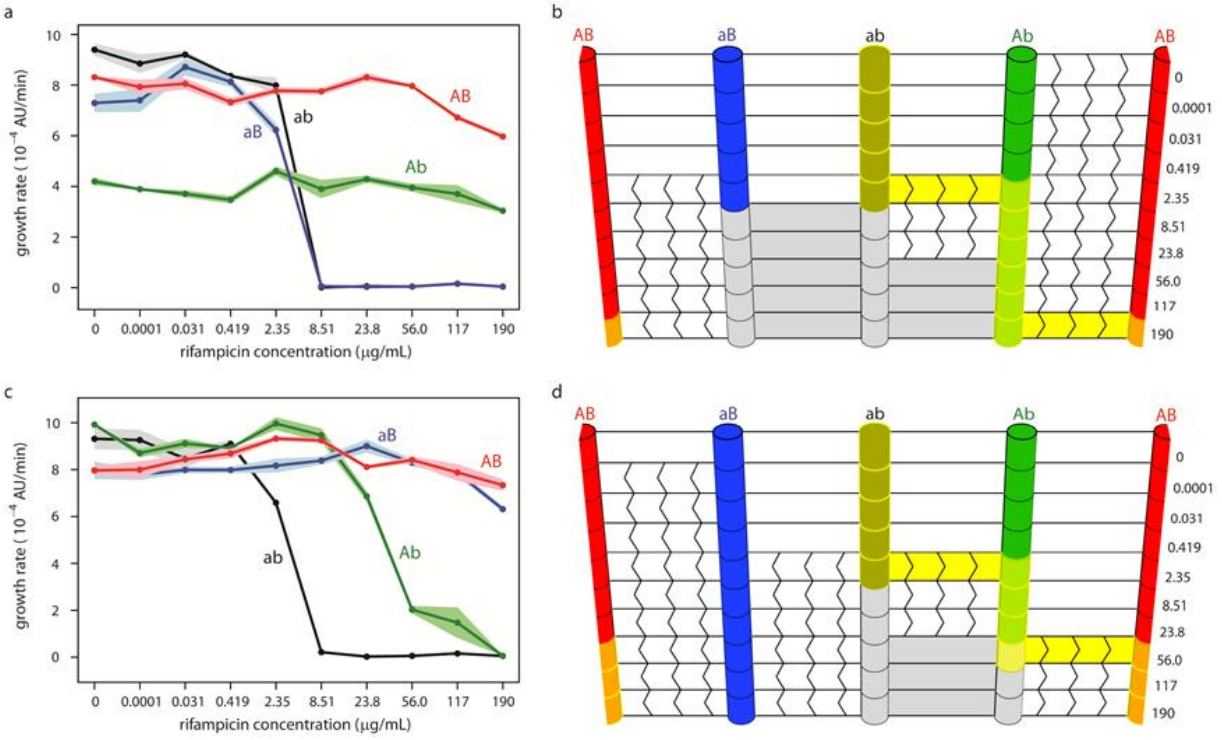
Supplementary Figure 1: Spectrophotometric growth assays of re-engineered ancestors were performed across a gradient of rifampicin concentrations. For each plot, the values on x-axis and y-axis are the same as in part m of this figure. (a) The mean growth rate of the untouched ancestor is shown (shading gives standard error of the mean and the maximum and minimum growth rates are shown with dotted red lines). For the rest of the graphs, the dotted lines are left as reference and the growth rates for the ancestral reconstruction of the following engineered strains are shown: (b) the S497 genotype with base t1546, (c) the S1236 genotype with base t436, (d) the G13 genotype with bases a428 and c1721, (e) the G13 genotype with bases g428 and t1721, (f) the G13 genotype with bases a428 and t1721, (g) the G18 genotype with bases t443 and c1527, (h) the G18 genotype with bases a443 and a1527, (i) the G18 genotype with bases t443 and a1527, (j) the M358 genotype with bases a437 and a1685, (k) the M358 genotype with bases t437 and c1685, (l) the M358 genotype with bases a437 and c1685, (m) the M367 genotype with bases g1532 and g1546, (n) the M367 genotype with bases t1532 and a1546, and (o) the M367 genotype with bases g1532 and a1546.

Supplementary Information: Results

Analysis of additional evolved isolates

In addition to the two isolates from distinct *Gradual* populations, we analyzed two isolates from distinct *Moderate* populations and two isolates from distinct *Sudden* populations. Both *Moderate* isolates had two mutations in *rpoB*, while both *Sudden* isolates had only a single mutation. All combinations of mutations for each isolate were engineered into a common background.

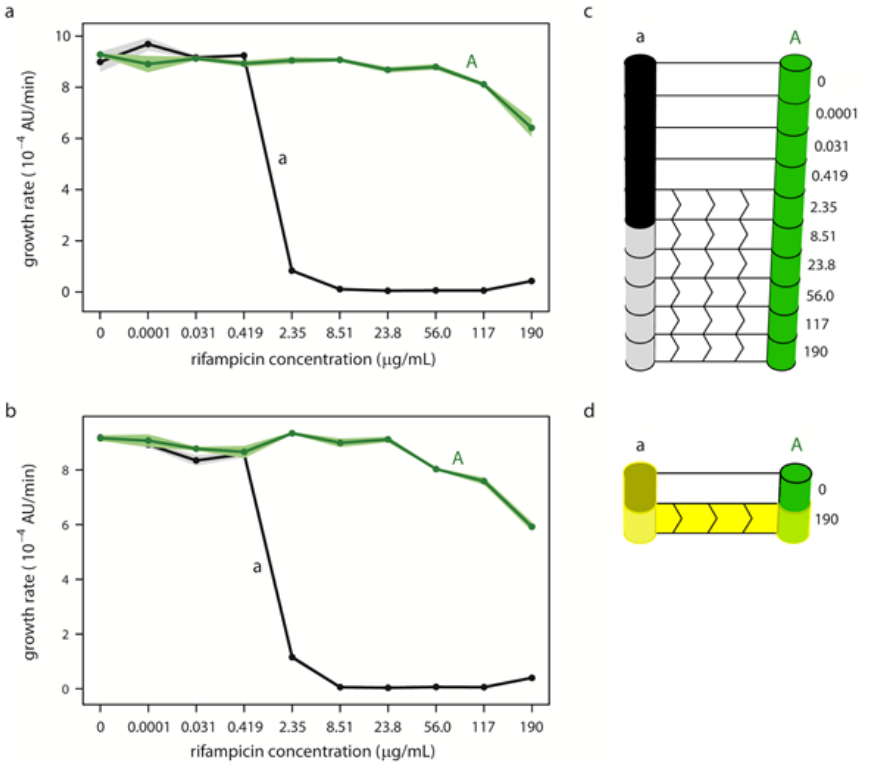
Supplementary Figure 2 shows the growth rate of the engineered strains corresponding to the *Moderate* isolates across a gradient of rifampicin concentrations. Regarding the paths actually taken (see highlighting in Supplementary Figures 2b,d), there is historical contingency upon intermediate environments for both isolates. For the lineage yielding the first *Moderate* isolate (Supplementary Figures 2a,b), the mutation ab→Ab becomes selectively accessible only for intermediate concentrations of rifampicin. For the lineage evolving the second *Moderate* isolate (Supplementary Figures 2c,d),



Supplementary Figure 2: Selective accessibility from the ancestral genotype to two evolved genotypes in two different *Moderate* populations. The evolved isolate in the first *Moderate* population has two mutations in *rpoB*, t437a and a1685c, yielding the amino acid substitutions V146D and E562A, respectively. The four engineered genotypes are denoted ab (bases t437 and a1685), Ab (bases a437 and a1685), aB (bases t437 and c1685) and AB (bases a437 and c1685). The evolved isolate in the second *Moderate* population also has two mutations in *rpoB*, t1532g and g1546a, yielding the amino acid substitutions L511R and D516N, respectively. The four engineered genotypes for the second isolate case are denoted ab (bases t1532 and g1546), Ab (bases g1532 and g1546), aB (bases t1532 and a1546) and AB (bases g1532 and a1546). **(a)** The maximum population growth rate for each of the engineered genotypes corresponding to the first isolate across a gradient of rifampicin concentrations. The points are means and the shading gives the standard error. **(b)** The "accessibility wall" for the case of the first isolate (see Figure 4 for a full description). **(c)** The growth rates and **(d)** selective accessibility for the case of the second Moderate isolate.

the path actually taken (ab→ Ab→AB) must be completed piece-wise in different environments. For the second isolate, if the ab→ aB→AB path had been taken (it was not), it would *not* have been historically contingent upon environments with intermediate drug concentrations (as the entire path is selectively accessible at the maximal concentration of rifampicin). For both lineages, the moderately slow change in concentration of rifampicin was critical to the actual evolutionary sequence: if the environment changed from an absence of antibiotic to its maximal concentration abruptly (as in the *Sudden* treatment), the path taken under moderate change would not be available (Figures 2b,d). Again, for the second isolate, the path not taken (ab→ aB→AB) would have been available under rapid change.

The growth rate of the engineered strains corresponding to the *Sudden* isolates across a gradient of rifampicin concentrations is shown in Supplementary Figure 3. We note that the mutation in each



Supplementary Figure 3: Selective accessibility from the ancestral genotype to two evolved genotypes in two different *Sudden* populations. The evolved isolate in the first *Sudden* population has a single mutation in *rpoB*, g1546t, yielding the amino acid substitution D516Y. The two engineered genotypes are denoted a (base g1546) and A (t1546). The evolved isolate in the second *Sudden* population also a single mutation in *rpoB*, g436t, yielding the amino acid substitutions V146F. The two engineered genotypes for the second isolate case are denoted a (base g436), and A (base t436). **(a)** The maximum population growth rate for each of the engineered genotypes corresponding to the first isolate across a gradient of rifampicin concentrations. The points are means and the shading gives the standard error. **(b)** The growth rates for the second isolate. **(c)** The "accessibility wall" was identical for both isolates (see Figure 4 for a full description of this wall diagram). We note that all intermediate concentrations of rifampicin were not experienced by these evolving lineages. **(d)** This abridged wall shows the only two concentrations actually experienced by the evolving *Sudden* populations. The actual evolutionary path (for both populations) is highlighted.

lineage is selectively accessible at both intermediate and high concentrations of the drug. Thus, the single mutations that arose in different *Sudden* populations could also be selected under gradual or moderate rates of change. We note that several isolates from the *Moderate* and *Gradual* treatments had a mutation in the same amino acid residue as one of the *Sudden* isolates (see Figure 2a and Supplementary Table 2). Some of these *Moderate* and *Gradual* isolates had only a single mutation. In cases where evolutionary rescue is due to such single mutations, more populations can survive under a slower rate of change because there is more time for these single mutations to arise. Specifically, such a mutation must have arisen prior to exposure to rifampicin in the *Sudden* treatment because after the first transfer, the ancestral genotype cannot survive. However, in the *Moderate* and *Gradual* treatments, there is a range of intermediate drug concentrations where the ancestral genotype survives and where the single mutant is selectively advantageous.

Differences in the rate of evolutionary rescue between treatments in such a case are due to demographic features, rather than the specific experience of an intermediate environment. This suggests one (tentative) way to separate support for demographic versus contingent forms of rescue. When the mutations in isolates from populations evolving under gradual change are the same (or similar) to the mutations in isolates from populations evolving under rapid change, it is plausible that demographic features could underlie differences in evolutionary rescue between treatments. However, when gradual isolates contain multiple mutations not found in any rapid isolate, it is plausible that evolutionary rescue was contingent upon certain intermediate environments. We note a few immediate caveats for this sorting algorithm. First, the mutational patterns we have outlined do not exhaust the set of possible patterns (e.g., there may be gradual isolates that contain a hybrid of mutations inside and outside the set of mutations from rapid isolates). Second, these two forms of rescue are not mutually exclusive (indeed, we have both types of mutational patterns in our isolates). Proper confirmation of different forms of rescue requires construction of intermediate genotypes and assays of fitness across an environmental gradient. It is possible that isolates that seem to support one form of rescue actually do not, when fully inspected.

Full list of mutations from the evolved isolates

Below is a list of all mutations found in isolates from selected populations across the three treatments. Every mutation discovered was a non-synonymous base substitution in *rpoB*. In the event of multiple mutations for an isolate, the first mutation to fix was determined by sequencing the population at different time points (from samples stored during the evolution experiment). The ancestor (*E. coli* B REL606: *rpoB* NC_012967.1) was the same for all populations.

Treatment	Population	Mutation location	Ancestral base	Mutant base	First to fix in population?	Residue location	Ancestral amino acid	Mutant amino acid
Gradual	13	428	G	A	No	143	Arginine	Histidine
Gradual	13	1721	C	T	Yes	574	Serine	Phenylalanine
Gradual	14	1691	C	T	Yes	564	Proline	Leucine
Gradual	17	1546	G	A	No	516	Aspartic Acid	Asparagine
Gradual	17	1687	A	C	Yes	563	Threonine	Proline
Gradual	18	443	A	T	No	148	Glutamine	Leucine
Gradual	18	1527	C	A	Yes	509	Serine	Arginine
Gradual	35	443	A	T	Yes	148	Glutamine	Leucine

Gradual	35	1534	T	C	No	512	Serine	Proline
Gradual	41	443	A	T	No	148	Glutamine	Leucine
Gradual	41	1714	A	C	Yes	572	Isoleucine	Leucine
Gradual	42	443	A	C	Yes	148	Glutamine	Proline
Gradual	42	1525	A	C	No	509	Serine	Arginine
Gradual	46	443	A	T	Yes	148	Glutamine	Leucine
Gradual	46	1715	T	G	No	572	Isoleucine	Serine
Gradual	49	443	A	T	No	148	Glutamine	Leucine
Gradual	49	1532	T	C	Yes	511	Leucine	Proline
Gradual	49	1546	G	A	No	516	Aspartic Acid	Asparagine
Gradual	66	1547	A	G	Yes	516	Aspartic Acid	Glycine
Gradual	69	1534	T	C	No	512	Serine	Proline
Gradual	69	1597	C	T	Yes	533	Leucine	Phenylalanine
Gradual	70	1714	A	C	No	572	Isoleucine	Leucine*
Gradual	70	1715	T	A	Yes	572	Isoleucine	Asparagine
Gradual	71	1585	C	T	Yes	529	Arginine	Cysteine
Gradual	77	1703	A	G	No	568	Asparagine	Serine
Gradual	77	1721	C	T	Yes	574	Serine	Phenylalanine
Gradual	94	1535	C	T	Yes	512	Serine	Phenylalanine
Gradual	95	1535	C	T	Yes	512	Serine	Phenylalanine
Gradual	96	427	C	T	No	143	Arginine	Cysteine
Gradual	96	1547	A	G	Yes	516	Aspartic Acid	Glycine
Gradual	99	1592	C	T	Yes	531	Serine	Phenylalanine
Gradual	112	1527	C	A	Yes	509	Serine	Arginine
Gradual	112	1577	A	G	No	526	Histidine	Arginine
Gradual	119	443	A	T	No	148	Glutamine	Leucine
Gradual	119	1715	T	A	Yes	572	Isoleucine	Asparagine
Gradual	123	1546	G	A	Yes	516	Aspartic Acid	Asparagine
Gradual	126	1538	A	G	No	513	Glutamine	Arginine
Gradual	126	1601	G	A	Yes	534	Glycine	Aspartic Acid
Gradual	131	1527	C	A	No	509	Serine	Arginine
Gradual	131	1687	A	C	Yes	563	Threonine	Proline
Gradual	135	1534	T	C	No	512	Serine	Proline
Gradual	135	1545	G	A	Yes	515	Methionine	Isoleucine
Gradual	140	1592	C	T	Yes	531	Serine	Phenylalanine
Gradual	141	1532	T	C	Yes	511	Leucine	Proline
Gradual	141	1546	G	A	No	516	Aspartic Acid	Asparagine
Gradual	143	1714	A	C	Yes	572	Isoleucine	Leucine
Gradual	165	1552	A	T	No	518	Asparagine	Tyrosine
Gradual	165	1715	T	G	Yes	572	Isoleucine	Serine
Gradual	179	1534	T	C	Yes	512	Serine	Proline
Gradual	183	1538	A	C	Yes	513	Glutamine	Proline
Gradual	183	1690	C	T	No	564	Proline	Serine
Moderate	185	1532	T	G	Yes	511	Leucine	Arginine
Moderate	185	1546	G	A	No	516	Aspartic Acid	Asparagine
Moderate	190	443	A	T	Yes	148	Glutamine	Leucine
Moderate	190	1546	G	A	No	516	Aspartic Acid	Asparagine
Moderate	196	1547	A	G	No	516	Aspartic Acid	Glycine
Moderate	196	1687	A	C	Yes	563	Threonine	Proline
Moderate	217	1592	C	T	Yes	531	Serine	Phenylalanine
Moderate	242	443	A	T	No	148	Glutamine	Leucine
Moderate	242	1715	T	A	Yes	572	Isoleucine	Asparagine
Moderate	244	1513	T	C	No	505	Phenylalanine	Leucine
Moderate	244	1534	T	C	Yes	512	Serine	Proline
Moderate	250	1532	T	A	Yes	511	Leucine	Glutamine
Moderate	250	1538	A	T	No	513	Glutamine	Leucine
Moderate	275	1546	G	A	Yes	516	Aspartic Acid	Asparagine
Moderate	281	443	A	T	No	148	Glutamine	Leucine
Moderate	281	1714	A	C	Yes	572	Isoleucine	Leucine
Moderate	307	1547	A	G	No	516	Aspartic Acid	Glycine
Moderate	307	1589	T	C	Yes	530	Isoleucine	Threonine
Moderate	315	437	T	C	No	146	Valine	Alanine
Moderate	315	1535	C	T	Yes	512	Serine	Phenylalanine
Moderate	318	1592	C	T	Yes	531	Serine	Phenylalanine
Moderate	326	1592	C	T	Yes	531	Serine	Phenylalanine
Moderate	326	1715	T	A	No	572	Isoleucine	Asparagine

Moderate	357	1534	T	C	Yes	512	Serine	Proline
Moderate	358	437	T	A	Yes	146	Valine	Aspartic Acid
Moderate	358	1685	A	C	No	562	Glutamic Acid	Alanine
Moderate	362	1546	G	A	Yes	516	Aspartic Acid	Asparagine
Moderate	367	1532	T	G	Yes	511	Leucine	Arginine
Moderate	367	1546	G	A	No	516	Aspartic Acid	Asparagine
Moderate	368	1714	A	C	Yes	572	Isoleucine	Leucine
Moderate	368	1721	C	T	No	574	Serine	Phenylalanine
Moderate	377	1592	C	T	No	531	Serine	Phenylalanine
Moderate	377	1600	G	C	Yes	534	Glycine	Arginine
Moderate	382	1547	A	G	No	516	Aspartic Acid	Glycine
Moderate	382	1601	G	T	Yes	534	Glycine	Valine
Moderate	385	1592	C	T	Yes	531	Serine	Phenylalanine
Moderate	387	1547	A	G	Yes	516	Aspartic Acid	Glycine
Moderate	387	1565	C	T	No	522	Serine	Phenylalanine
Moderate	389	1703	A	G	No	568	Asparagine	Serine
Moderate	389	1715	T	G	Yes	572	Isoleucine	Serine
Moderate	408	80	T	C	No	27	Leucine	Proline
Moderate	408	1600	G	T	Yes	534	Glycine	Cysteine
Moderate	408	1703	A	G	No	568	Asparagine	Serine
Moderate	424	1547	A	G	Yes	516	Aspartic Acid	Glycine
Moderate	424	1703	A	G	No	568	Asparagine	Serine
Moderate	433	1547	A	G	No	516	Aspartic Acid	Glycine
Moderate	433	1703	A	G	No	568	Asparagine	Serine
Moderate	433	2059	C	T	Yes	687	Arginine	Cysteine
Moderate	434	443	A	T	Yes	148	Glutamine	Leucine
Moderate	434	1538	A	C	No	513	Glutamine	Proline
Moderate	434	1572	T	G	No	524	Isoleucine	Methionine
Moderate	437	443	A	T	Yes	148	Glutamine	Leucine
Moderate	437	1535	C	T	No	512	Serine	Phenylalanine
Moderate	443	437	T	C	No	146	Valine	Alanine
Moderate	443	1715	T	G	Yes	572	Isoleucine	Serine
Moderate	459	1527	C	A	Yes	509	Serine	Arginine
Moderate	459	1610	G	A	No	537	Glycine	Aspartic Acid
Sudden	477	1714	A	T	Yes	572	Isoleucine	Phenylalanine
Sudden	497	1546	G	T	Yes	516	Aspartic Acid	Tyrosine
Sudden	504	1592	C	A	Yes	531	Serine	Tyrosine
Sudden	513	1546	G	T	Yes	516	Aspartic Acid	Tyrosine
Sudden	582	1546	G	T	Yes	516	Aspartic Acid	Tyrosine
Sudden	593	1592	C	T	Yes	531	Serine	Phenylalanine
Sudden	742	1592	C	T	Yes	531	Serine	Phenylalanine
Sudden	762	1592	C	A	Yes	531	Serine	Tyrosine
Sudden	767	1714	A	T	Yes	572	Isoleucine	Phenylalanine
Sudden	878	1592	C	A	Yes	531	Serine	Tyrosine
Sudden	1223	1592	C	T	Yes	531	Serine	Phenylalanine
Sudden	1236	436	G	T	Yes	146	Valine	Phenylalanine
Sudden	1334	436	G	T	Yes	146	Valine	Phenylalanine

Supplementary Table 2. A complete list of all mutations from sequenced isolates from the three treatments.

*For *Gradual* population 70, the two mutations that occurred were in the same codon. Our table describes each of these mutations as if they happened in the absence of the other. However, the order of mutations in this codon was ATC→AAC→CAC, thus, the amino acid sequence was Isoleucine→Asparagine→Histidine. Therefore, the "Leucine" listed in this table was not actually present at this residue in this evolving lineage (as the ATC→CTC mutation did not occur).

Supplementary Information: Mathematical framework and proofs

Model and terminology

Here we use an approach inspired by the analysis of Weinreich *et al*. (2005). Let there be L loci in our genetic system, with two alleles each. Thus, a genotype can be described as a bit-string of length L:

$$\vec{g} = \langle a_1, a_2, a_3, \ldots a_L \rangle ,$$

where $a_i \in \{0,1\}$, for all $i \in \{1,2,3, \ldots L\}$. Let the set of all genotypes be denoted $\mathbf{G}$ (we note $|\mathbf{G}| = 2^L$).

Given two genotypes, $\vec{g}_u$ and $\vec{g}_v$, we define the set

$$\mathbf{D}(\vec{g}_u, \vec{g}_v) = \{i | a_{i,u} \neq a_{i,v}\},$$

where $a_{i,u}$ is the allele of the i^{th} locus of genotype $\vec{g}_u$. That is, $\mathbf{D}(\vec{g}_x, \vec{g}_y)$ gives all the indices of the alleles that differ between $\vec{g}_u$ and $\vec{g}_v$. The (Hamming) distance between two genotypes is simply:

$$\Delta(\vec{g}_u, \vec{g}_v) = |\mathbf{D}(\vec{g}_u, \vec{g}_v)|.$$

We define the set

$$\mathbf{N}(\vec{g}_u, \vec{g}_v) = \{\vec{g} \notin \{\vec{g}_u, \vec{g}_v\} | \Delta(\vec{g}_u, \vec{g}) + \Delta(\vec{g}, \vec{g}_v) = \Delta(\vec{g}_u, \vec{g}_v)\}.$$

That is, $\mathbf{N}(\vec{g}_u, \vec{g}_v)$ is the set of genotypes "near" both genotypes (these are genotypes that agree in allelic state with $\vec{g}_u$ or $\vec{g}_v$ at every locus). This is the set of "intermediate" genotypes.

We define a "flip" function F_i (mapping one bitstring to another) that flips the i^{th} locus to the other allelic state, while leaving all other loci alone. Thus,

$$F_i[\vec{g}] = F_i[\langle a_1, a_2, a_3, \ldots a_L \rangle] = \langle a_1, a_2, \ldots, a_{i-1}, (a_i + 1) (\text{mod } 2), a_{i+1}, \ldots, a_L \rangle.$$

Each of the shortest paths between $\vec{g}_u$ and $\vec{g}_v$ can be defined by a vector of consecutive genotypes each differing by a single base. One of the paths from $\vec{g}_u$ to $\vec{g}_v$ can be written as:

$$\vec{p}_j(\vec{g}_u, \vec{g}_v) = \langle \vec{g}_u, F_{i_1}[\vec{g}_u], (F_{i_2} \circ F_{i_1})[\vec{g}_u], (F_{i_3} \circ F_{i_2} \circ F_{i_1})[\vec{g}_u], \ldots, \left(F_{i_{\Delta(\vec{g}_u, \vec{g}_v)}} \circ \ldots \circ F_{i_3} \circ F_{i_2} \circ F_{i_1}\right)[\vec{g}_u] \rangle , \quad [S1]$$

where "$\circ$" represents functional composition and $\{i_1, i_2, i_3, \ldots, i_{\Delta(\vec{g}_u, \vec{g}_v)}\} \equiv \mathbf{D}(\vec{g}_u, \vec{g}_v)$. Thus,

$$\left(F_{i_{\Delta(\vec{g}_u, \vec{g}_v)}} \circ \ldots \circ F_{i_3} \circ F_{i_2} \circ F_{i_1}\right)[\vec{g}_u] = \vec{g}_v .$$

The index j in $\vec{p}_j(\vec{g}_u, \vec{g}_v)$ corresponds to one ordering of the elements of $\mathbf{D}(\vec{g}_u, \vec{g}_v)$. There are $\{\Delta(\vec{g}_u, \vec{g}_v)\}!$ shortest paths between $\vec{g}_u$ and $\vec{g}_v$, so the index j runs from 1 to $\{\Delta(\vec{g}_u, \vec{g}_v)\}!$. We let the set of shortest paths from $\vec{g}_u$ to $\vec{g}_v$ be denoted by $\mathbf{P}(\vec{g}_u, \vec{g}_v)$.

The fitness of any genotype can depend on environment. Let there be N environmental states. The x^{th} environmental state is denoted e_x and the set of environments is denoted $\mathbf{E}$. Let the fitness of genotype $\vec{g}$ in environment e be given by $\Omega[\vec{g}, e]$. The fitness effect of a mutation at the i^{th} locus on the background of genotype $\vec{g}$ in environment e is given by the function:

$$\omega_i[\vec{g}, e] = \Omega[F_i[\vec{g}], e] - \Omega[\vec{g}, e].$$

Thus, the mutation from $\vec{g}$ to $F_i[\vec{g}]$ is beneficial in environment e if $\omega_i[\vec{g}, e] > 0$. Also

$$\omega_i[\vec{g}, e] = -\omega_i[F_i[\vec{g}], e]. \qquad [S2]$$

We will call genotype $\vec{g}$ a peak ($P(e)$) or a valley ($V(e)$) in environment e if all mutations are detrimental or beneficial, respectively (namely, $\omega_i[P(e), e] < 0$ and $\omega_i[V(e), e] > 0$, for all $i \in \{1,2,3, \dots L\}$).

Consider the path described above in [S1] from genotype $\vec{g}_u$ to $\vec{g}_v$. We will call this path "selectively accessible" in environment e if all mutational steps along the path are beneficial*. That is, if

$$\omega_{i_1}[\vec{g}_u, e] > 0$$

$$\omega_{i_2}\left[F_{i_1}[\vec{g}_u], e\right] > 0$$

$$\omega_{i_3}\left[\left(F_{i_2} \circ F_{i_1}\right)[\vec{g}_u], e\right] > 0$$

$$\vdots$$

$$\omega_{i_{\Delta(\vec{g}_u,\vec{g}_v)}}\left[\left(F_{i_{\Delta(\vec{g}_u,\vec{g}_v)-1}} \circ \dots \circ F_{i_3} \circ F_{i_2} \circ F_{i_1}\right)[\vec{g}_u], e\right] > 0.$$

Let $\mathbf{B}_e(\vec{g}_u, \vec{g}_v)$ be the set of all selectively accessible paths from $\vec{g}_u$ to $\vec{g}_v$ in environment e. We note that $\mathbf{B}_e(\vec{g}_u, \vec{g}_v) \subseteq \mathbf{P}(\vec{g}_u, \vec{g}_v)$.

Now consider an ordered series of environmental states: $\vec{e} = \langle e_{s_1}, e_{s_2}, e_{s_3}, \dots, e_{s_Z} \rangle$. We will say there exists a selectively accessible path from genotype $\vec{g}_u$ to genotype $\vec{g}_v$ over this environmental series if there exists a set of genotypes $\vec{g}_{n_j} \in \mathbf{N}(\vec{g}_u, \vec{g}_v) \cup \{\vec{g}_u, \vec{g}_v\}$, such that

$$\mathbf{B}_{e_{s_1}}\left(\vec{g}_u, \vec{g}_{n_1}\right) \neq \emptyset$$

$$\mathbf{B}_{e_{s_2}}\left(\vec{g}_{n_1}, \vec{g}_{n_2}\right) \neq \emptyset$$

$$\mathbf{B}_{e_{s_3}}\left(\vec{g}_{n_2}, \vec{g}_{n_3}\right) \neq \emptyset$$

$$\vdots$$

$$\mathbf{B}_{e_{s_Z}}\left(\vec{g}_{n_{Z-1}}, \vec{g}_v\right) \neq \emptyset$$

* Here, we do not consider neutral mutations as part of selectively accessible paths (we ignore neutrality in this supplement for simplicity).

where $\Delta\left(\vec{g}_{n_j}, \vec{g}_v\right) \geq \Delta\left(\vec{g}_{n_{j+1}}, \vec{g}_v\right)$. Because we allow for the possibility of $\vec{g}_{n_j} = \vec{g}_{n_{j+1}}$, if $\mathbf{B}_{e_{s_j}}(\vec{g}_u, \vec{g}_v) \neq \emptyset$, for any e_{s_j} element of the series $\vec{e}$, then there exists a selectively accessible path from $\vec{g}_u$ to $\vec{g}_v$ over the environmental series.[†]

We will label a selectively accessible path "contingent on environment e_{s_k}" if there exists a selectively accessible path from $\vec{g}_u$ to $\vec{g}_v$ over the environmental series $\vec{e} = \langle e_{s_1}, e_{s_2}, e_{s_3}, \ldots, e_{s_Z} \rangle$, but there exists no selectively accessible path over the environmental series $\vec{e}' = \langle e_{s_1}, e_{s_2}, e_{s_3}, \ldots e_{s_{k-1}}, e_{s_{k+1}}, \ldots, e_{s_Z} \rangle$. Similarly, we will say that a selectively accessible path is contingent on a set of environments if their joint removal from the environmental series prohibits any selectively accessible path.

We now define environmental epistasis. Let e_x and e_y denote two distinct environment states. There exists environmental epistasis if, for some genotype $\vec{g}$ and some allele a_i:

$$\omega_i[\vec{g}, e_x] \neq \omega_i[\vec{g}, e_y]$$

If $\omega_i[\vec{g}, e_x]$ and $\omega_i[\vec{g}, e_y]$ are of the same sign (but different values), we call this "magnitude environmental epistasis" and if they are of opposite signs, we call this "sign environmental epistasis." A landscape has no environmental epistasis if

$$\omega_i[\vec{g}, e_x] = \omega_i[\vec{g}, e_y]$$

for all $\vec{g} \in \mathbf{G}$, for all e_x, $e_y \in \mathbf{E}$ (where $x \neq y$), and for all $i \in \{1,2,3, \ldots L\}$.

We now move to genetic epistasis. Let $\mathbf{G}_{i=0}$ and $\mathbf{G}_{i=1}$ be the sets of all genotypes with $a_i = 0$ and $a_i = 1$, respectively. Let $\vec{g}_{u,i=0}$, $\vec{g}_{v,i=0} \in \mathbf{G}_{i=0}$, where $\vec{g}_{u,i=0} \neq \vec{g}_{v,i=0}$. There exists genetic epistasis if

$$\omega_i[\vec{g}_{u,i=0}, e] \neq \omega_i[\vec{g}_{v,i=0}, e]$$

If $\omega_i[\vec{g}_{u,i=0}, e]$ and $\omega_i[\vec{g}_{v,i=0}, e]$ are of the same sign (but different values), this is "magnitude genetic epistasis" and if they are of opposite signs, this is "sign genetic epistasis." A landscape has no genetic epistasis if

$$\omega_i[\vec{g}_{u,i=0}, e] = \omega_i[\vec{g}_{v,i=0}, e],$$

for all $e \in \mathbf{E}$, for all $\vec{g}_{u,i=0}$, $\vec{g}_{v,i=0} \in \mathbf{G}_{i=0}$ (where $\vec{g}_{u,i=0} \neq \vec{g}_{v,i=0}$), and for all $i \in \{1,2,3, \ldots L\}$. By [S2], this means

$$\omega_i[\vec{g}_{u,i=1}, e] = \omega_i[\vec{g}_{v,i=1}, e],$$

for all $e \in \mathbf{E}$, for all $\vec{g}_{u,i=1}$, $\vec{g}_{v,i=1} \in \mathbf{G}_{i=1}$ (where $\vec{g}_{u,i=1} \neq \vec{g}_{v,i=1}$), and for all $i \in \{1,2,3, \ldots L\}$.

[†] We note that just because a path is selectively accessible over a series of environments does not guarantee that an evolving population can realistically follow it. This will depend on whether sufficient time is spent in each environment (which itself will depend on the strength of selective differences, population size, and the mutation rate). For simplicity, we will ignore these issues here and implicitly assume that there is sufficient time in each environment for any selective path to be evolutionarily traversed (however, none of the theorems below make explicit evolutionary claims).

Results

Theorem 1*: Consider two genotypes* $\vec{g}_u$ *and* $\vec{g}_v$ *in an environment* e *in which there is no sign genetic epistasis. Each of the shortest paths from* $\vec{g}_u$ *to* $\vec{g}_v$ *is selectively accessible if and only if* $\Omega[\vec{g}_u, e] < \Omega[\vec{g}_n, e] < \Omega[\vec{g}_v, e]$ *for all* $\vec{g}_n \in \mathbf{N}(\vec{g}_u, \vec{g}_v)$.

Proof:

(1) $\Omega[\vec{g}_u, e] < \Omega[\vec{g}_n, e] < \Omega[\vec{g}_v, e] \Rightarrow$ *all paths are selectively accessible*

We simplify the notation of [S1]:

$$\vec{p}_j(\vec{g}_u, \vec{g}_v) = \langle \vec{g}_{u_0}, \vec{g}_{u_1}, \vec{g}_{u_2}, \vec{g}_{u_3}, \dots, \vec{g}_{u_{\Delta(\vec{g}_u, \vec{g}_v)}} \rangle,$$

where $\vec{g}_{u_0} = \vec{g}_u$, $\vec{g}_{u_{\Delta(\vec{g}_u, \vec{g}_v)}} = \vec{g}_v$, and $F_{i_k}[\vec{g}_{u_{k-1}}] = \vec{g}_{u_k}$. Assume that this path is not selectively accessible.

If this path is not accessible, then there exists at least one value $k \in \{0,1,2, \dots, \Delta(\vec{g}_u, \vec{g}_v) - 1\}$, where

$$\omega_{i_{k+1}}[\vec{g}_{u_k}, e] < 0$$

Case 1: $\Delta(\vec{g}_u, \vec{g}_v) = 1$.

There is only one mutational step to consider.

$$0 > \omega_{i_1}[\vec{g}_{u_0}, e] = \Omega[F_{i_1}[\vec{g}_{u_0}], e] - \Omega[\vec{g}_{u_0}, e] = \Omega[\vec{g}_{u_1}, e] - \Omega[\vec{g}_{u_0}, e] = \Omega[\vec{g}_v, e] - \Omega[\vec{g}_u, e].$$

But we have assumed $\Omega[\vec{g}_u, e] < \Omega[\vec{g}_v, e]$, which gives a contradiction (epistasis does not matter here).

Case 2: $\Delta(\vec{g}_u, \vec{g}_v) > 1$.

Suppose the allelic index $i_{k+1} = m$. Because there is no genetic epistasis

$$\omega_m[\vec{g}_u, e] = \Omega[F_m[\vec{g}_u], e] - \Omega[\vec{g}_u, e] < 0.$$

However, since $F_m[\vec{g}_u] \in \mathbf{N}(\vec{g}_u, \vec{g}_v)$, we must have $\Omega[\vec{g}_u, e] < \Omega[F_m[\vec{g}_u], e]$, which leads to a contradiction.

Therefore, if there is no sign genetic epistasis and $\Omega[\vec{g}_u, e] < \Omega[\vec{g}_n, e] < \Omega[\vec{g}_v, e]$ for all $\vec{g}_n \in \mathbf{N}(\vec{g}_u, \vec{g}_v)$, all shortest paths must be selectively accessible.

(2) *All paths are selectively accessible* $\Rightarrow \Omega[\vec{g}_u, e] < \Omega[\vec{g}_n, e] < \Omega[\vec{g}_v, e]$.

Again, we use the simplified notation for a path:

$$\vec{p}_j(\vec{g}_u, \vec{g}_v) = \langle \vec{g}_{u_0}, \vec{g}_{u_1}, \vec{g}_{u_2}, \vec{g}_{u_3}, \dots, \vec{g}_{u_{\Delta(\vec{g}_u, \vec{g}_v)}} \rangle,$$

Because all paths are selectively accessible, we are guaranteed that

$$0 < \omega_{i_1}[\vec{g}_{u_0}, e] = \Omega[F_{i_1}[\vec{g}_{u_0}], e] - \Omega[\vec{g}_{u_0}, e] = \Omega[\vec{g}_{u_1}, e] - \Omega[\vec{g}_{u_0}, e],$$

for all $i_1 \in \mathbf{D}(\vec{g}_u, \vec{g}_v)$. Thus, all possible single mutants from $\vec{g}_u$ (along the shortest paths to $\vec{g}_v$) have higher fitness than $\vec{g}_u$. However, because there is no sign genetic epistasis, we are guaranteed

$$0 < \omega_{i_2}[\vec{g}_{u_1}, e] = \Omega[F_{i_2}[\vec{g}_{u_1}], e] - \Omega[\vec{g}_{u_1}, e] = \Omega[\vec{g}_{u_2}, e] - \Omega[\vec{g}_{u_1}, e],$$

for all $\{i_2|i_1\} \in \mathbf{D}(\vec{g}_u, \vec{g}_v) \cap \{i_1\}^C$, where $i_1 \in \mathbf{D}(\vec{g}_u, \vec{g}_v)$. This means that all double mutants from $\vec{g}_u$ (along the shortest paths to $\vec{g}_v$) have higher fitness than the single mutants, which means they have higher fitness than $\vec{g}_u$. The same argument can be used repeatedly. For instance, for the j^{th} mutational step

$$0 < \omega_{i_j}\left[\vec{g}_{u_{j-1}}, e\right] = \Omega\left[F_{i_j}[\vec{g}_{u_{j-1}}], e\right] - \Omega\left[\vec{g}_{u_{j-1}}, e\right] = \Omega\left[\vec{g}_{u_j}, e\right] - \Omega\left[\vec{g}_{u_{j-1}}, e\right],$$

for all:

$$\{i_j|i_{j-1}, i_{j-2}, \dots, i_1\} \in \mathbf{D}(\vec{g}_u, \vec{g}_v) \cap \{i_{j-1}, i_{j-2}, \dots, i_1\}^C,$$

where

$$\{i_{j-1}|i_{j-2}, i_{j-3}, \dots, i_1\} \in \mathbf{D}(\vec{g}_u, \vec{g}_v) \cap \{i_{j-2}, i_{j-3}, \dots, i_1\}^C,$$

$$\{i_{j-2}|i_{j-3}, i_{j-4}, \dots, i_1\} \in \mathbf{D}(\vec{g}_u, \vec{g}_v) \cap \{i_{j-3}, i_{j-4}, \dots, i_1\}^C,$$

$$\vdots$$

$$\{i_2|i_1\} \in \mathbf{D}(\vec{g}_u, \vec{g}_v) \cap \{i_1\}^C,$$

$$i_1 \in \mathbf{D}(\vec{g}_u, \vec{g}_v).$$

Therefore, each successive step along the shortest paths must increase in fitness. Now, the only question is whether the genotypes corresponding to the penultimate mutational step have a higher fitness than $\vec{g}_v$. Suppose this is the case:

$$\Omega\left[\vec{g}_{u_{\Delta(\vec{g}_u,\vec{g}_v)-1}}, e\right] - \Omega[\vec{g}_v, e] > 0,$$

for some $i_{\Delta(\vec{g}_u,\vec{g}_v)-1} \in \mathbf{D}(\vec{g}_u, \vec{g}_v)$. This implies

$$\Omega\left[F_{i_{\Delta(\vec{g}_u,\vec{g}_v)-1}}\left[\vec{g}_{u_{\Delta(\vec{g}_u,\vec{g}_v)-1}}\right], e\right] - \Omega\left[\vec{g}_{u_{\Delta(\vec{g}_u,\vec{g}_v)-1}}, e\right] = \omega_{i_{\Delta(\vec{g}_u,\vec{g}_v)-1}}\left[\vec{g}_{u_{\Delta(\vec{g}_u,\vec{g}_v)-1}}, e\right] < 0.$$

This would make this path selectively inaccessible, which contradicts our assumption. Therefore, if each of the shortest paths from $\vec{g}_u$ to $\vec{g}_v$ is selectively accessible, then $\Omega[\vec{g}_u, e] < \Omega[\vec{g}_n, e] < \Omega[\vec{g}_v, e]$ for all $\vec{g}_n \in \mathbf{N}(\vec{g}_u, \vec{g}_v)$, which completes the proof.

***Theorem 2**: Consider two genotypes $\vec{g}_u$ and $\vec{g}_v$ in an environment e in which there is no sign genetic epistasis. If $\vec{g}_u$ is a valley ($\vec{g}_u = V(e)$) or $\vec{g}_v$ is a peak ($\vec{g}_v = P(e)$), then all paths between these two genotypes are selectively accessible.*

Proof:

Again, we use the simplified notation for a path:

$$\vec{p}_j(\vec{g}_u, \vec{g}_v) = \langle \vec{g}_{u_0}, \vec{g}_{u_1}, \vec{g}_{u_2}, \vec{g}_{u_3}, \dots, \vec{g}_{u_{\Delta(\vec{g}_u,\vec{g}_v)}} \rangle,$$

In the case that $\vec{g}_u$ is a valley, we are guaranteed that

$$0 < \omega_{i_1}[\vec{g}_{u_0}, e] = \Omega[F_{i_1}[\vec{g}_{u_0}], e] - \Omega[\vec{g}_{u_0}, e] = \Omega[\vec{g}_{u_1}, e] - \Omega[\vec{g}_{u_0}, e],$$

for all $i_1 \in \mathbf{D}(\vec{g}_u, \vec{g}_v)$. Because there is no sign genetic epistasis, we are guaranteed

$$0 < \omega_{i_2}[\vec{g}_{u_1}, e] = \Omega[F_{i_2}[\vec{g}_{u_1}], e] - \Omega[\vec{g}_{u_1}, e] = \Omega[\vec{g}_{u_2}, e] - \Omega[\vec{g}_{u_1}, e],$$

for all $\{i_2|i_1\} \in \mathbf{D}(\vec{g}_u, \vec{g}_v) \cap \{i_1\}^C$, where $i_1 \in \mathbf{D}(\vec{g}_u, \vec{g}_v)$. This means that all double mutants from $\vec{g}_u$ (along the shortest paths to $\vec{g}_v$) have higher fitness than the single mutants, which means they have higher fitness than $\vec{g}_u$. The same argument can be used repeatedly to show that each successive step along the shortest paths must increase in fitness. Is it possible for the genotypes corresponding to the penultimate mutational step to have a higher fitness than $\vec{g}_v$? Suppose this is the case:

$$\Omega\left[\vec{g}_{x_{\Delta(\vec{g}_u,\vec{g}_v)-1}}, e\right] - \Omega[\vec{g}_v, e] > 0,$$

for some $i_{\Delta(\vec{g}_u,\vec{g}_v)-1} \in \mathbf{D}(\vec{g}_u, \vec{g}_v)$. This implies

$$\Omega\left[F_{i_{\Delta(\vec{g}_u,\vec{g}_v)-1}}\left[\vec{g}_{u_{\Delta(\vec{g}_u,\vec{g}_v)-1}}\right], e\right] - \Omega\left[\vec{g}_{u_{\Delta(\vec{g}_x,\vec{g}_y)-1}}, e\right] = \omega_{i_{\Delta(\vec{g}_u,\vec{g}_v)-1}}\left[\vec{g}_{u_{\Delta(\vec{g}_u,\vec{g}_v)-1}}, e\right] < 0.$$

Let $i_{\Delta(\vec{g}_u,\vec{g}_v)-1} = m$. Because there is no sign epistasis, we know

$$\omega_m[\vec{g}_u, e] < 0.$$

But this would mean that $\vec{g}_u$ is not a valley, which is a contradiction. Therefore, we are guaranteed that $\Omega[\vec{g}_u, e] < \Omega[\vec{g}_n, e] < \Omega[\vec{g}_v, e]$ for all $\vec{g}_n \in \mathbf{N}(\vec{g}_x, \vec{g}_y)$. Theorem 1 guarantees that all paths are selectively accessible in this case. A similar approach can be used to show that if $\vec{g}_v$ is a peak, then all paths are selectively accessible.

***Theorem 3**: Consider a selectively accessible path from $\vec{g}_u$ to $\vec{g}_v$ over the environmental series $\vec{e} = \langle e_{s_1}, e_{s_2}, e_{s_3}, \ldots, e_{s_Z} \rangle$. Further, suppose that $\vec{g}_u = V(e_{s_1})$ or $\vec{g}_v = P(e_{s_Z})$. If this path is contingent upon environment e_{s_k} then both sign genetic epistasis and sign environmental epistasis exist.*

Proof:

If there is a selectively accessible path from $\vec{g}_u$ to $\vec{g}_v$ over this environmental series, then there exists a genotypes $\vec{g}_{n_{k-1}}, \vec{g}_{n_k} \in \mathbf{N}(\vec{g}_u, \vec{g}_v) \cup \{\vec{g}_u, \vec{g}_v\}$, such that

$$\mathbf{B}_{e_{s_k}}(\vec{g}_{n_{k-1}}, \vec{g}_{n_k}) \neq \emptyset .$$

However, if there is no selectively accessible path without environment e_{s_k}, then $\vec{g}_{n_{k-1}} \neq \vec{g}_{n_k}$ (if $\vec{g}_{n_{k-1}} = \vec{g}_{n_k}$, then selective accessibility with e_{s_k} would imply selective accessibility without e_{s_k}). Thus, in environment e_{s_k} there exists a selectively accessible path between two distinct genotypes $\vec{g}_{n_{k-1}}$ and $\vec{g}_{n_k}$. If there is no sign environmental epistasis, then this path must also be accessible in environments $e_{s_{k-1}}$ and $e_{s_{k+1}}$ (where such environments are defined). If $e_{s_{k-1}}$ is a defined environment (i.e., if $k > 1$), then $\mathbf{B}_{e_{s_{k-1}}}(\vec{g}_{n_{k-1}}, \vec{g}_{n_k}) \neq \emptyset$. Given that $\mathbf{B}_{e_{s_{k-1}}}(\vec{g}_{n_{k-2}}, \vec{g}_{n_{k-1}}) \neq \emptyset$, we must have

$$\mathbf{B}_{e_{s_{k-1}}}(\vec{g}_{n_{k-2}}, \vec{g}_{n_k}) \neq \emptyset .$$

This means that we do not need e_{s_k} to get from $\vec{g}_u$ to $\vec{g}_v$ (because we can get from $\vec{g}_u$ to $\vec{g}_{n_k}$ over environments $\langle e_{s_1}, e_{s_2}, e_{s_3}, \ldots, e_{s_{k-1}} \rangle$ and we can get from $\vec{g}_{n_k}$ to $\vec{g}_v$ over environments $\langle e_{s_{k+1}}, e_{s_{k+2}}, e_{s_{k+3}}, \ldots, e_{s_Z} \rangle$). If $e_{s_{k-1}}$ is not defined, then we can use $e_{s_{k+1}}$ and repeat the above argument. Thus, if there is no sign environmental epistasis, then there is a selective path available without environment e_{s_k}, which contradicts our assumption that the path is contingent on environment e_{s_k}. Therefore, there must be sign environmental epistasis present.

Suppose that there is no sign genetic epistasis. If $\vec{g}_u = V(e_{s_1})$, then Theorem 2 guarantees that all paths from $\vec{g}_u$ to $\vec{g}_v$ are selectively accessible in environment e_{s_1}, which would mean

$$\mathbf{B}_{e_{s_1}}(\vec{g}_u, \vec{g}_v) \neq \emptyset.$$

If $\vec{g}_v = P(e_{s_Z})$, then Theorem 2 guarantees that all paths from $\vec{g}_u$ to $\vec{g}_v$ are selectively accessible in environment e_{s_Z}, which would mean

$$\mathbf{B}_{e_{s_Z}}(\vec{g}_u, \vec{g}_v) \neq \emptyset.$$

Thus, if $\vec{g}_u = V(e_{s_1})$ or $\vec{g}_v = P(e_{s_Z})$ and if there is no sign genetic epistasis, then there is a selective path available without environment e_{s_k}, which contradicts our assumption that the path is contingent on environment e_{s_k}. Thus, there must be sign genetic epistasis present, which completes the proof.