RECOMBINATION PEAKS AND FAST EVOLUTION

KRISTINA CRONA

ABSTRACT. Genetic recombination can speed up or slow down evolution toward the global optimum. For analyzing mutational trajectories to the optimum it is conventional to study peaks in the fitness landscapes. However, such mutational peaks cannot provide a complete picture of the interplay between mutations, recombination and selection. Here we use recombination peaks as an additional tool. Informally, a recombination peak is a set of genotypes such that no allele shuffling among members can improve average fitness. If conventional peaks are optimal with respect to mutations, then recombination peaks are optimal with respect to recombination. The approach applies to general fitness landscapes with an arbitrary number of alleles and loci, and our experiments involve up to five alleles.

One of the most frequently observed patterns of epistasis in nature is diminishing returns epistasis, i.e., the combined effect of beneficial mutations is smaller than a multiplicative expectation would predict. Recombination tends to speed up adaptation toward a peak whenever there is a systematic pattern of diminishing returns epistasis near the peak. The opposite is true for increasing returns epistasis. Mixed curvature is more challenging to analyze. Recombination peaks provide a language for describing effects of recombination, and threshold results are discussed accordingly. The experiments suggest that recombination is more likely to speed up adaptation for larger fitness landscapes.

1. INTRODUCTION

Evolutionary aspects of recombination have been studied for over a hundred years with the goal to understand why recombination is widespread in nature. The famous paradox of recombination refers to that gene shuffling is nearly universal, although nobody has found a universal advantage of recombination. A related and in a sense more practical problem is the effect of recombination. Recombination can speed up or slow down evolution. If a new viral or bacterial variant has a higher recombination rate than the original type, it would be valuable to know whether one should expect faster or slower evolution, or perhaps no change at all. It should be noted that fast evolution does not imply that recombination as a mechanism is selected for. The evolutionary advantage or disadvantage of recombination is usually modeled by recombination modifiers, for more background on recombination see e.g., Otto and Lenormand (2002). Here, the goal is to analyze the effect of recombination, regardless of its evolutionary origin or any other evolutionary aspects of the mechanism as such.

Key words and phrases. Fitness landscape, recombination, polytope, triangulation.

A fundamental observation about recombination is that the mechanism has no effect in the absence of linkage disequilibria i.e., for an already well-mixed population additional gene shuffling has no impact. The terminology can be illustrated for a biallelic two-locus system with genotypes $00, 10, 01, 11$. For a population in equilibrium the product of the proportions of the extreme genotypes (00 and 11) equals the product of the proportions of the intermediate genotypes (10 and 01). If the extreme genotypes are underrepresented, then recombination tends to generate the extreme genotypes, and the opposite is true if the extreme genotypes are overrepresented. A common reason for linkage disequilibria is epistasis, or gene interactions. Epistasis measures deviations from multiplicative fitness, where the fitness of a genotype is defined as its expected contribution to the next generation.

For an evolutionary process from the wild-type 00 to the double mutant 11, where each mutation $0 \to 1$ increases fitness, recombination tends to slow down adaptation for increasing returns epistasis, i.e., if the fitness of 11 exceeds a multiplicative expectation (evolutionary speed is measured as the expected number of generations from the wild-type population 00 until fixation at the optimal genotype 11). That should be intuitively clear from considering the linkage disequilibria caused by the epistasis. By a similar argument, recombination tends to speed up adaptation for diminishing returns epistasis.

The impact of recombination is more challenging to describe for several alleles and loci. For an overview of the problem it is natural to consider fitness landscapes. A fitness landscape for an L-locus system with k alleles at each locus is determined by the fitness values of all k^L genotypes. If the fitness of a genotype is interpreted as a height coordinate, then an evolutionary process for a population can be represented as a walk in the fitness landscape where each step increases the height, i.e., each step represents a single point mutation that increases fitness. A genotype is a peak if it has higher fitness than its mutational neighbors.

For predicting the effect of recombination, one needs to consider accessible walks in the fitness landscape. The rank order of genotypes with respect to fitness determines possible walks. The second key factor is the curvature of the landscape. An obvious complication is that the majority of larger fitness landscapes have mixed curvature with both diminishing and increasing returns epistasis. In addition to rank orders and curvature, population parameters can make a difference. For instance, the effect of recombination is highly sensitive to population size for some landscapes with poor peak accessibility (Crona, 2018).

2. Results

2.1. **Rank orders and Curvature.** The biallelic two-locus landscape discussed in the introduction has a simple rank order. As a complement, one can consider the effect of recombination for chaotic rank orders.

Example 2.1. Consider adaptation from 000 to 111 for a finite population with fitness values

$$\begin{aligned} &w_{000} = 1 \\ &w_{100} = w_{010} = w_{001} = 1 + \alpha, \quad \alpha > 0 \\ &w_{110} = w_{101} = w_{011} = \beta, \qquad 0 < \beta < 1 \\ &w_{111} = \gamma > w_g \text{ for } g \neq 111. \end{aligned}$$

Figure 2.1 shows the fitness graph (Crona et al., 2013; Krug and De Visser, 2009) for the system. Adaptation from 000 to 111 is difficult because of the suboptimal peaks $100, 010, 001$, at least if double mutations are rare. Recombination has the potential to speed up adaptation because of *peak jumping*, i.e., that local peaks are replaced by the global peak, depending on α, β, γ and population parameters. Not surprisingly, for simulations where α is replaced by $-\alpha$, recombination is no longer beneficial. Specifically, that is the case for $\alpha = \beta = 0.1, \lambda = 3$ (Crona, 2018). It is easy to verify that the replacement of α by $-\alpha$ leaves the curvature unchanged. (See Section 5.1 for a discussion about curvature for biallelic 3-locus systems.)

The example shows that the effect of recombination cannot be predicted from curvature type alone. As clear from the introduction, neither can the effect be determined from rank orders of genotypes with respect to fitness. A complete analysis would need to capture the interplay between rank order aspects and curvature.

In order to isolate the effect of curvature it is useful to consider rank orders that are as simple as possible. To avoid some subtle aspects due to stochasticity (Barton and Otto, 2005; Hill and Robertson, 1966), populations are also assumed to be infinite and well mixed throughout the remainder of the paper. The biallelic two-locus system discussed in the introduction, with genotypes $00, 10, 01, 11$ and where each mutation $0 \mapsto 1$ increases fitness, is an important reference for larger fitness landscapes with simple rank orders. In more detail, let f_g denote fitness and w_g log fitness for a genotype g. Then

$$\epsilon := w_{00} + w_{11} - w_{10} - w_{01}$$

is zero if fitness is multiplicative, the system has diminishing returns epistasis if $\epsilon < 0$, and increasing returns epistasis if $\epsilon > 0$. Under the stated assumptions recombination speeds up adaptation from 00 to 11 for $\epsilon < 0$ and slows down adaptation for $\epsilon > 0$ whereas recombination has no effect if $\epsilon = 0$ (as explained in the introduction). The (Hamming) distance between two genotypes is defined as the number of loci where the genotypes differ, and two genotypes are called mutational neighbors if their distance is one.

In general, the rank orders are assumed to satisfy that the landscapes are single peaked and that each step toward the global peak increases fitness. Since larger fitness landscapes tend to have mixed curvature, i.e., both diminishing returns and increasing returns epistasis, it is more challenging to analyze the effect of recombination. Figure 2.1 shows a fitness graph for a two-locus system with three alleles. Each arrow in

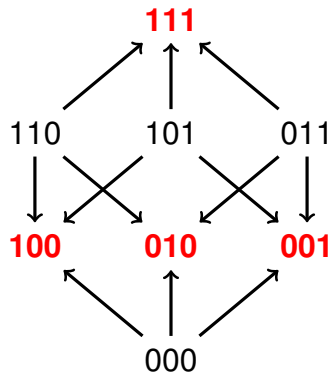

FIGURE 1. The three suboptimal peaks $100, 010, 001$ make all paths from 000 to 111 inaccessible. Depending on population parameters, recombination can be beneficial because of peak jumping, meaning that the suboptimal peaks are replaced by the global peak through repeated allele shuffling.

the figure points toward a genotype of higher fitness. The peak 00 (marked red) has mutational neighbors $10, 20, 01, 02$ (marked yellow), and the genotypes on distance two (marked blue) are $11, 12, 21, 22$. By assumption, there are no obstacles in the landscape that prevent walks from an arbitrary genotype to the peak 00. For a clean presentation only arrows between genotypes of different distances to the peak are included in the figure, whereas a complete fitness graph has arrows between all pairs of mutational neighbors.

For the class of fitness landscapes described by the figure, assume that the initial population is equally distributed on the genotypes of distance two from the peaks, and for simplicity that such genotypes have fitness 1. Extrapolating from the biallelic case, one expects that recombination speeds up adaptation for low values of f_{00} and slows it down for high values of f_{00}. If all intermediates have exactly the same fitness (Table 1), then the effect of recombination is indeed very similar to the biallelic two-locus case. However, if the intermediates have different fitness values, then the curvature is mixed for some values of f_{00}. For the experiments (Table 3) systematic diminishing returns epistasis implies a speed advantage for recombination and increasing returns epistasis a disadvantage, which should not be surprising. The more remarkable finding is that the threshold is relatively high within the interval for mixed curvature, specifically about 3/4 from the left end of the mixed curvature interval (similar results can be found in Tables 4-10). An intuitive explanation for the relatively high thresholds is that evolution tends to use high fitness intermediates rather than low fitness intermediates. The finding suggests that recombination is more likely to speed up adaptation for larger fitness landscapes as compared to the biallelic two-locus case. The threshold problem discussed for two loci can be extended to general landscapes with simple rank orders.

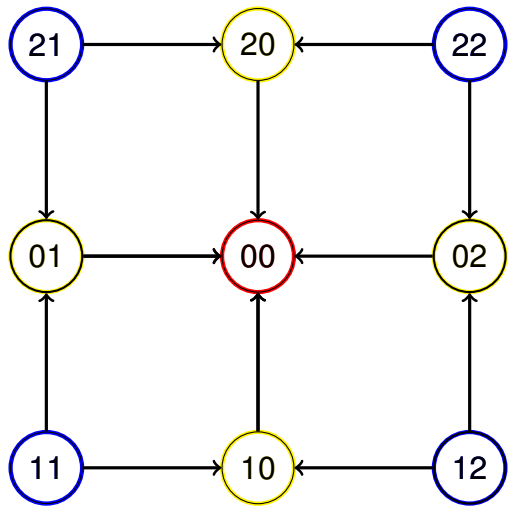

FIGURE 2. Fitness increases for each mutation $1 \mapsto 0$ and $2 \mapsto 0$. Initially the population is distributed on the blue marked genotypes $21, 22, 11, 12$, and eventually the population goes to fixation at the peak 00. The genotype 00 is the global peak. Each arrow points toward the genotype with higher fitness. For clarity, arrows are included for mutational neighbors with exception for the ones that have the same distance to the global peak (for instance, no arrow connects 10 and 20).

Definition 2.2. *The threshold problem* is defined for a single peaked fitness landscape that satisfy the following conditions:

(i) The initial population consists of genotypes of distance L from the peak p, and the population is uniformly distributed over such genotypes.
(ii) (a) If there are more than two alleles at each locus, then each mutation $a \mapsto 0$ for $a \neq 0$ increases fitness,
(b) for biallelic systems, each mutation $0 \mapsto 1$ increases fitness.
(iii) Unless otherwise stated, we also assume $f_g = 1$ for all g of distance L from the peak.

Under the stated assumptions, the goal is to find the threshold T such that recombination speeds up adaptation if $f_p < T$ and slows it down if $f_p > T$ (if such a threshold exists).

The definition is intended to be relevant for the end behavior of adaptation. A rugged landscape can still be smooth near a peak, and then the conditions (i) and (ii) apply for the last few steps of the adaptation. The somewhat inconsistent notation in condition (ii) depends on that it is convenient to use zeros for the global peak in the multiallelic case, whereas the convention in the biallelic case is that zeros denote the wild-type alleles. Note that Condition (iii) is not of any theoretical importance, but included for easy comparisons of experiments. Throughout the remainder of the paper the landscapes are assumed to satisfy Conditions (i)-(iii), unless otherwise stated. The main purpose

TABLE 1. Initially the population consists of 00 genotypes, and eventually the population goes to fixation at 11 after the numbers of generation shown with and without recombination (bottom line). Recombination speeds up adaptation for diminishing returns epistasis (left) and slows it down for increasing returns epistasis (middle), whereas there is no effect if $\epsilon = 0$.

g	f_g
00	1
10	1.1
01	1.1
11	1.14
Epistasis	$\epsilon < 0$
with rec.	391
no rec.	690

g	f_g
00	1
10	1.1
01	1.1
11	1.3
Epistasis	$\epsilon > 0$
with rec.	236
no rec.	172

g	f_g
00	1
10	1.1
01	1.1
11	1.21
Epistasis	$\epsilon = 0$
with rec.	274
no rec.	274

TABLE 2. The experiment is similar to Table 1 except that there are three alleles $0, 1, 2$. Here 00 is the global optimum and the starting point is a population equally distributed on $11, 12, 21, 22$. The results differ only marginally from Table 1.

g	w_g
00	1.14
10	1.1
01	1.1
20	1.1
02	1.1
11	1
12	1
21	1
22	1
Epistasis	$\epsilon < 0$
with rec.	399
no rec.	710

g	f_g
00	1. 3
10	1.1
01	1.1
20	1.1
02	1.1
11	1
12	1
21	1
22	1
Epistasis	$\epsilon > 0$
with rec.	243
no rec.	177

g	f_g
00	1.21
10	1.1
01	1.1
20	1.1
02	1.1
11	1
12	1
21	1
22	1
Epistasis	$\epsilon = 0$
with rec.	281
no rec.	281

with the experiments was to better understand recombination thresholds. For additional details about the the assumptions for the experiments, see Section 5.2

2.2. **Recombination peaks for genotypes of distance two.** Landscapes that satisfy the conditions in Definition 2.2, have either systematic diminishing returns epistasis, systematic increasing returns epistasis, or mixed curvature (as illustrated for landscapes with two loci and three alleles). For a more detailed analysis it is convenient to consider sets of genotypes that are optimal with respect to recombination. We call such sets

recombination peaks. The word "peak" merely refers to optimality (no mountain top metaphor is intended).

For minimal landscapes, the set $\{10, 01\}$ is a recombination peak exactly if $w_{11}+w_{00} < w_{10} + w_{01}$. That should make sense since allele shuffling $10 + 01 \mapsto 11 + 00$ decreases mean fitness. Similarly, increasing returns epistasis is equivalent to that $\{00, 11\}$ is a recombination peak.

Analogously, for biallelic 3-locus landscapes (systematic) diminishing returns epistasis is equivalent to that

$$\{100, 010\}, \{100, 001\}, \{010, 001\}, \{110, 101\}, \{110, 011\}, \{101, 011\}$$

are recombination peaks (notice that the genotypes in each pair have distance two, and that they have same distance to the peak 111), and increasing returns epistasis to that

$$\{110, 000\}, \{101, 000\}, \{011, 000\}, \{111, 100\}, \{111, 010\}, \{111, 011\}.$$

are recombination peaks. For two locus and three alleles, diminishing returns epistasis means that $\{10, 01\}, \{10, 02\}, \{01, 20\}, \{02, 20\}$ are recombination peaks, and increasing returns epistasis that $\{00, 11\}, \{12, 01\}, \{21, 00\}, \{22, 00\}$ are recombination peaks. Diminishing and increasing returns epistasis can be expressed similarly for any number of loci and alleles. Note also that a pair of genotypes with distance two constitutes a recombination peak exactly if exchanging alleles at the two loci where they differ does not result in higher mean fitness (a general definition of recombination peaks will be discussed in Section 2.3).

The following claim appears obvious. However, we do not have a theoretical proof.

Conjecture 1. *Assume that an L-locus fitness landscape with peak p and with an arbitrary number of alleles satisfies the conditions in Definition 2.2.*

Consider the pairs of genotypes such that

(1) the distance to p is the same for both genotypes within the pair, and it is less than L,
(2) the distance between the genotypes in the pair is two.

If all such pairs are recombination peaks, then recombination speeds up adaptation. If none of the pairs are recombination peaks, then recombination slows down adaptation.

For biallelic landscapes, the condition that all the pairs which satisfy (1) and (2) are recombination peaks is equivalent to that the landscape has universal negative epistasis (as defined in Krug and Oros (2024)), and the opposite extreme to that the landscape has universal positive epistasis (Crona et al., 2023).

Example 2.3. Assume that the $f_{11} = f_{12} = f_{21} = f_{22} = 1$, and that the intermediates have fitness

$$1 + a, 1 + b, 1 + c, 1 + d \quad \text{where } 0 < a \leq b \leq c \leq d.$$

By assumption, $f_{00} > 1 + d$. Let

$$L = (1 + a) \times (1 + b), \quad H = (1 + c) \times (1 + d).$$

Then all pairs $\{10, 01\}, \{10, 02\}, \{01, 20\}, \{02, 20\}$ are recombination peaks if $f_{00} < L$ and that none of them if $f_{00} > H$. According to the conjecture, recombination speeds up adaptation if $f_{00} < L$ and slows it down if $f_{00} > H$. However, the conjecture does not apply if $L < f_{00} < H$. For an analysis of the exact recombination threshold it seems natural to consider some notion of average curvature. A simple measure [which agrees with the threshold in the case $a = b = c = d$] is

$$\big((1+a)(1+b)(1+c)(1+d)\big)^{1/2} * .$$

Experiments show that the threshold depends on how the genotypes are positioned in the fitness graph (Table 5). Specifically, the threshold is greater for

$$f_{10} = 1 + c, f_{01} = 1 + d,$$

as compared to

$$f_{10} = 1 + a, f_{01} = 1 + d,$$

provided $a < c < d$. Intuitively, if the highest fitness values are associated with a pair of genotypes that can produce the global peak through recombination, then recombination is more likely to speed up adaptation.

For instance, with $a = 0.3, b = 0.4, c = 0.5$ and $d = 0.6$, the recombination threshold is 2.32 or 2.16, depending on how the intermediates are placed in the fitness graph. For comparison, the lower bound from the conjectures is $L = 1.82$ and $H = 2.4$, and the measure $*$ equals 2.08. The experiments illustrate that the thresholds (2.32 or 2.16) in both experiments are relatively high in comparison with H, L and $*$, and secondly that the positions of the intermediates in the fitness graph have an impact.

The expression $*$ can be generalized to an arbitrary fitness landscape that satisfies Definition 2.2, specifically

$$T_{G_{L-1}} := (\prod_{g \in G_{L-1}} f_g)^{\frac{L}{|G_{L-1}|}},$$

where G_k denote the genotypes of distance k. One can perhaps consider $T_{G_{L-1}}$ as the best estimate if the only available information is fitness values of genotypes in G_{L-1} and G_L. For a biallelic landscape with multiplicative fitness for all genotypes except the peak, the recombination threshold is exactly $T_{G_{L-1}} = \prod_{g \in G_{L-1}} f_g$. However, the estimate is a poor predictor unless fitness is nearly multiplicative.

Example 2.4. Under the assumption that

$$f_{00000} = 1, f_{10000} = f_{01000} = f_{00100} = f_{00010} = f_{00001} = 1.1,$$

and that the remaining intermediates follow multiplicative expectations except for that

$$f_{11110} = f_{11101} = f_{11011} = f_{11101} = f_{11110} = 20,$$

the recombination threshold is $f_{11111} = 25.8$. In other words, recombination speeds up adaptation if $20 < f_{11111} < 25.8$, and slows down adaptation if $f_{11111} > 25.8$. For comparison $T_{G_{L-1}} = (\prod_{g \in G_{L-1}} f_g)^{\frac{L}{|G_{L-1}|}} = 1.61$.

An estimate based on a comparison between f_p and the wild-type versus *all intermediates* has a better chance to be relevant for the recombination threshold. To that end it makes sense to consider recombination peaks for an arbitrary set of genotypes.

2.3. **General recombination peaks.** A recombination peak is a set of genotypes that is optimal with respect to recombination. One can verify that the set $\{111, 000\}$ is as a recombination peak if

$$\begin{aligned}
&w_{111} + w_{000} > w_{110} + w_{001}\\
&w_{111} + w_{000} > w_{101} + w_{010}\\
&w_{111} + w_{000} > w_{011} + w_{100}\\
&w_{111} + 2w_{000} > w_{100} + w_{010} + w_{001}\\
&w_{000} + 2w_{111} > w_{011} + w_{101} + w_{011}.
\end{aligned}$$

It is immediate that the three first inequalities are necessary. The bottom inequality is also necessary since one example of (repeated) allele shuffling starting out with the three genotypes 000, 111, 111 is

$$000 + 111 + 111 \mapsto 110 + 001 + 111 \mapsto 110 + 101 + 011,$$

and the remaining inequality can be explained similarly. More background on how to determine if a set constitutes a recombination peak is provided Section 2.5.

For fitness landscapes that satisfy i-iii) such that $\{111, 000\}$ is a recombination peak, it is fair to say that the peak genotype 111 has high fitness as compared to all the intermediates. Our experiments supports that recombination slows down adaptation whenever $\{111, 000\}$ is a recombination peak (Tables 6-9).

One can define recombination peaks for an arbitrary number of alleles and loci. We define the allele counter $C^j(\sum w_g)$ as the vector that counts how many times the allele j occurs in total.

(1) $C^j(w_g)$ is the vector that shows how many time allele j occurs (0 or 1) at each locus of g,
(2) $C^j(\sum w_g) = \sum C^j(w_g)$.

For instance, $C^0(w_{10}) = [0, 1]$ and $C^0(w_{01}) = [1, 0]$, and from Condition 2,

$$C^0(w_{10} + 2w_{01}) = [0, 1] + 2[1, 0] = [2, 1],$$

and similarly $C^1(w_{10} + 2w_{01}) = [1, 2]$ and $C^2(w_{10} + 2w_{01}) = [0, 0]$. For determining if a set of genotypes is a recombination peak one needs to compare sets of genotypes that have the same allele counts.

Definition 2.5. A set $S \subset \{0, \ldots, k\}^L$ is a recombination peak if for any genotypes $g_1, \ldots, g_m \in S$ and $g'_1, \ldots, g'_m \in \{0, \ldots, k\}^L$ the following condition holds:

$$\text{If } C^j(\sum w_{g_i}) = C^j(\sum w_{g'_i}) \text{ for all } j, \text{ then } \sum w_{g_i} > \sum w_{g'_i}$$

In other words, the sum of the fitnesses for genotypes from a recombination peak exceeds the sum from genotypes with the same allele count. Readers familiar with Beerenwinkel et al. (2007) should notice that recombination peaks are closely related to authors' concept of a fittest populations (however, note that recombination peaks are *sets* of genotypes).

The following result is immediate:

Observation 2.6. If the set of all genotypes from a population constitutes a recombination peak, then the population cannot increase its average fitness through a sequence of recombination events.

The observation is more or less a tautology but it is less obvious how to determine whether or not a set of genotypes is a recombination peak. For solving the problem one can apply theory on triangulations of polytopes (see Section 2.5).

The following claim, stated as an open problem has some support from experiments. If correct, the claim provides an upper bound for the recombination threshold.

2.4. **Open problem.** *Proposed upper bound for the recombination threshold: For a fitness landscape that satisfies the conditions in Definition 2.2 let p denote the global peak of the fitness landscape. If for all $g \in G_L$ the pair $\{p, g\}$ is a recombination peak, then recombination slows down adaptation.*

The condition stated in the proposed bound simply means that f_p must be sufficiently high as compared to intermediates. For the threshold experiments on mixed curvature (Tables 3-10) the thresholds were close to the proposed upper bound in several cases, and sometimes considerably lower. The threshold problem, at least in the form stated here, is primarily interesting for relatively few loci (otherwise the assumptions would seem less natural). If the proposed bound is not correct, it may still work as a rule of thumb for relatively small fitness landscapes.

Theoretical work on recombination for several loci has primarily considered selection of recombination modifiers (Barton, 1995), which is a problem different from the effect of recombination (Otto and Lenormand, 2002). Some of the fundamental insights from classical work could still be relevant for theoretical questions on recombination thresholds.

2.5. **Recombination and triangulations of polytopes.** The following observation is one of the most precise results on recombination for a general fitness landscapes. The observations rephrases theory developed in (Beerenwinkel et al., 2007). In brief, one can determine whether or not a population is "perfectly combined" by using theory on triangulations (De Loera et al., 2010). .

Observation 2.7. Let w be a fitness landscape.

(i) One can determine if a population is a recombination peak by checking the signs of a set of forms in the variables w_g. In particular, for a biallelic 2-locus system one needs to check the sign of the form $\epsilon = w_{00} + w_{11} - w_{10} - w_{01}$. The forms are known as circuits (see the appendix for a more detailed discussion).

(ii) A population is a recombination peak exactly if its genotypes belong to the same simplex in the triangulation induced by the fitness landscape w, in the sense described in Beerenwinkel et al. (2007) (see also below for a brief description).

For a biallelic two-locus population one can describe the result in a simple way. Observation (i) translates to that the form

$$\epsilon = w_{00} + w_{11} - w_{10} - w_{01}$$

determines if a population is a recombination peak; specifically $\epsilon > 0$ implies that $\{11, 10, 00\}$ and $\{11, 01, 00\}$ are recombination peaks, whereas $\epsilon < 0$ implies that $\{11, 10, 01\}$ and $\{10, 01, 00\}$, are recombination peaks. Observation (ii) translates to that the square with vertices labelled by the genotypes $00, 10, 01, 11$ is subdivided to the triangles with vertices $11, 10, 00$ and $11, 01, 00$ if $\epsilon > 0$, and with vertices $11, 10, 00$ and $11, 01, 00$ if $\epsilon < 0$. A population being a recombination peak is equivalent to all genotypes belonging to the same triangle in the induced triangulation.

More generally, for a biallelic L-locus system, the genotypes can be considered vertices in an L-cube. If w_g is the height coordinate for the genotype g and the polytope P the convex hull of the resulting point configuration, then the projection of the upper faces of P onto the L-cube describes the triangulation induced by w. As in the two-locus case, the population is a recombination peak if its genotypes belong to the same simplex (the same tetrahedron for $L = 3$, and so forth). Biallelic L-locus systems are also discussed in some detail in Crona et al. (2023); Crona (2014).

Here we also consider also $k > 2$ alleles. In brief, let Δ_k be the $k - 1$ dimensional simplex

$$\Delta_k = \{(p_1, ..., p_l) \in [0, 1] : p_1 + \cdots + p_l = 1\}$$

and let Δ_k^L be the L-fold direct product of simplices Δ_k . Then the genotypes can be identified with the vertices of the polytope Δ_k^L. Similar to the biallelic case, w induces a triangulation of the Δ_k^L.

Consider systematic increasing returns epistasis for biallelic L-locus system in our setting. The condition implies that the fitness landscape induces a specific triangulation known as the stair-case triangulation (Crona et al., 2023) For $L = 3$, the triangulation is

$$\{000, 100, 110, 111\}, \{000, 100, 101, 111\}, \{000, 010, 110, 111\}$$

$$\{000, 010, 011, 111\}, \{000, 001, 101, 111\}, \{000, 010, 011, 111\}.$$

One can ask if a similar result holds for $k > 2$ alleles, Below is a counterexample.

Example 2.8. The following fitness landscapes have systematic increasing returns epistasis.

$$w_1 : w_{00} = 20, w_{01} = 7, w_{02} = 5, w_{10} = 9, w_{20} = 3, w_{11} = 1, w_{12} = 1, w_{21} = 2, w_{22} = 1$$

$$w_2 : w_{00} = 20, w_{10} = 5, w_{02} = 7, w_{10} = 3, w_{20} = 9, w_{11} = 2, w_{12} = 1, w_{21} = 1, w_{22} = 2$$

The triangulation for w_1 is

$$\{00, 01, 02, 12, 22\},\ \{00, 01, 11, 12, 21\},\ \{00, 01, 12, 21, 22\}$$
$$\{00, 10, 11, 12, 21\},\ \{00, 10, 12, 21, 22\},\ \{00, 10, 20, 21, 22\}$$

The triangulation for w_2 is

$$\{00, 01, 02, 11, 21\}, \{00, 02, 11, 12, 22\},\ \{00, 02, 11, 21, 22, \}$$
$$\{00, 10, 11, 12, 20\},\ \{00, 11, 12, 20, 22\},\ \{00, 11, 20, 21, 22\}.$$

The two triangulations are of different combinatorial types (as explained below), which shows that the condition of increasing returns epistasis does not determine the triangulation type in general.

For analyzing the triangulations induced by w_1 and w_2 in the example, one can use theory that relates triangulations of $\Delta^2 \times \Delta^2$ and lozenge tilings of triangles, see Chapter 9.2 in De Loera et al. (2010). The connection to tilings is known as the Cayley-trick. Figure 9.57 in Chapter 9 shows all combinatorial types of triangulations of $\Delta^2 \times \Delta^2$ as lozenge tilings of a triangle. The triangulation of the fitness landscape w_1 from the example corresponds to the second type from the top in the figure, and the triangulation of w_2 to the fourth type from the top.

3. DISCUSSION

Recombination can speed up or slow down adaptation. Our experiments suggest that recombination speeds up adaptation for many larger fitness landscapes. Several studies have shown that the impact of recombination is difficult to predict from conventional statistics on pairwise epistasis and similar summary statistics (Kouyos et al., 2006). We propose new tools for relating the curvature of the fitness landscape to the impact of recombination. We have demonstrated that neither rank orders alone, nor curvature alone, can predict the impact. For many chaotic rank orders, recombination is not very sensitive for curvature, whereas curvature is the determining factor for simple rank orders. In addition to rank orders and curvature, the impact of recombination also depends on population parameters.

In order to isolate the effect of curvature, our experiments use a rank order that is as simple as possible, i.e., fitness increases for each step toward a global peak, and infinite population size is assumed. In contrast to most studies, the experiments include landscapes with several alleles and loci and allow for initial populations with several genotypes. Already for two-locus landscapes the effect of several alleles is that recombination is more likely to speed up adaptation. One reason is that evolution tends to use high fitness intermediates rather than low fitness intermediates if there are several alternatives. The finding is yet another example of the recurrent observation in evolutionary biology that the walks chosen by adapting populations differ from arbitrary walks in the fitness landscape (Greene and Crona, 2014; Moradigaravand et al., 2014; Draghi and Plotkin, 2013).

The impact of recombination is clear-cut for minimal landscapes in the setting, i.e., for landscapes with four genotypes $00, 10, 01, 11$ such that each mutations $0 \mapsto 1$ increases

fitness. Recombination slows down adaptation if the landscape has increasing returns epistasis (that is if the fitness of 11 exceeds a multiplicative expectation) and speeds it up for diminishing returns epistasis. Our experiments support that the analogous result holds for larger landscapes if the curvature is consistent (either diminishing or increasing returns epistasis). However, for most larger landscapes the curvature is mixed which makes the effect of recombination more challenging to predict. A striking observation from experiments is that the outcome does not depend on the average curvature (of any sort). The speed advantage appears to discount low fitness intermediates if there are high fitness alternatives. Another argument in favor of that recombination tends to speed up adaptation is that diminishing returns epistasis have been observed in many empirical studies (Stround and Ratcliff, 2025; Wei and Zhang, 2019; Otwinowski et al., 2018; Lenski, 2017; Kryazhimskiy et al., 2014).

For more quantitative precision, one needs some notion of curvature and to that end we introduced recombination peaks. A recombination peak is a set of genotypes such that no allele shuffling among members of the set can improve average fitness. For instance, for minimal landscapes diminishing returns epistasis means that $\{10, 01\}$ is a recombination peak and increasing returns epistasis that $\{11, 00\}$ is a recombination peak. If one keeps the fitness values of all genotypes fixed, except for the peak, then one can ask how the impact of recombination will change by peak fitness. The threshold experiments assume that the initial population is equally distributed on the same distance from the peak (for more details see Section 5.2). For all landscapes, the experiments identified a threshold such that recombination slows down adaptation if the peak fitness exceeds the threshold and otherwise speeds it up. We found that if $\{g, p\}$ is a recombination peak for all g on maximal distance from the peak p, then recombination slows down adaption. The condition, if true in general, would provide an upper bound for the recombination threshold. Not surprisingly, the threshold was considerably lower for some landscapes.

It would be interesting to have a more precise understanding of recombination thresholds. One open problem is if recombination thresholds exist for all landscapes in the setting. Perhaps more empirically important, the implications of the observed relatively high recombination thresholds for larger fitness landscapes would be interesting to quantify. For most fitness landscapes it seems nearly impossible to predict the recombination thresholds. However, we conjecture that recombination speeds up adaptation for systematic diminishing returns epistasis and slows it down for increasing returns epistasis, regardless of the number of loci and alleles.

4. Acknowledgements

Devin Greene produced all code for the evolutionary experiments and for determining triangulations of polytopes.

REFERENCES

Barton, N. H. A general model for the evolution of recombination. *Genetics Research*, 65(2), 123-144.

Barton, N. H., and Otto, S. P. (2005). Evolution of recombination due to random drift. *Genetics*, 169(4), 2353-2370.

Beerenwinkel, N., Pachter, L., and Sturmfels, B. (2007). Epistasis and shapes of fitness landscapes. *Statistica Sinica*, 1317-1342.

Brown, C. J., and Yahia, F. (2023). Virus-immune dynamics determined by prey-predator interaction network and epistasis in viral fitness landscape. *Journal of Mathematical Biology*, 86(1), 9.

Crona, K. (2020). Rank orders and signed interactions in evolutionary biology. *Elife*, 9, e51004.

Crona, K. Recombination and peak jumping. (2018). *Plos one*, 13(3), e0193123.

Crona, K. (2014). Polytopes, graphs and fitness landscapes. In Recent advances in the theory and application of fitness landscapes (pp. 177-205). Berlin, Heidelberg: Springer Berlin Heidelberg.

Crona, K., and Greene, D. (2024). Walsh Coefficieints and circuits for several alleles `arXiv:2401.00743`

Crona, K., Greene, D., and Barlow, M. (2013). The peaks and geometry of fitness landscapes. *Journal of theoretical biology*, 317, 1-10.

Crona, K., Krug, J., and Srivastava, M. (2023). Geometry of fitness landscapes: peaks, shapes and universal positive epistasis. *Journal of Mathematical Biology*, 86(4), 62.

De Loera, J., Rambau, J., and Santos, F. (2010). Triangulations: structures for algorithms and applications (Vol. 25). Springer Science and Business Media.

Draghi, J. A. and Plotkin, J. B. (2013). Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* doi: 10.1111/evo.12192.

Eble, H., Joswig, M., Lamberti, L., and Ludington, W. B. (2023). Master regulators of biological systems in higher dimensions. *Proceedings of the National Academy of Sciences*, 120(51), e2300634120.

Eble, H., Joswig, M., Lamberti, L., and Ludington, W. B. (2019). Cluster partitions and fitness landscapes of the Drosophila fly microbiome. *Journal of mathematical biology*, 79(3), 861-899.

Faure, A. J., Lehner, B., Miro Pina, V., Serrano Colome, C., and Weghorn, D. (2024). An extension of the Walsh-Hadamard transform to calculate and model epistasis in genetic landscapes of arbitrary shape and complexity. *PLOS Computational Biology*, 20(5), e1012132.

Greene, D. Multiallelic Walsh transforms `arXiv:2311.16925`

Greene, D. A Primer for the Walsh Transform `arXiv:2311.02177`

Greene, D., Crona, K. (2014) . The Changing Geometry of a Fitness Landscape Along an Adaptive Walk. *PLoS Comput Biol* 10(5): e1003520. doi:10.1371/journal.pcbi.1003520

Hill, W. G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, 8(3), 269-294.

De Visser, J. A. G., Park, S. C., and Krug, J. (2009). Exploring the effect of sex on empirical fitness landscapes. *the american naturalist*, 174(S1), S15-S30.

Kouyos, R. D., Otto, S. P.,and Bonhoeffer, S. (2006). Effect of varying epistasis on the evolution of recombination. *Genetics*, 173(2), 589-597.

Krug, J., and Oros, D. (2024). Evolutionary accessibility of random and structured fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(3), 034003.

Kryazhimskiy, S., Rice, D. P., Jerison, E. R., and Desai, M. M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344(6191), 1519-1522.

Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *The ISME journal*, 11(10), 2181-2194.

Moradigaravand, D., Kouyos, R., Hinkley, T., Haddad, M., Petropoulos, C. J., Engelstädter, J., and Bonhoeffer, S. (2014). Recombination accelerates adaptation on a large-scale empirical fitness landscape in HIV-1. *PLoS genetics*, 10(6), e1004439.

Otto, S. P., Lenormand, T. Resolving the paradox of sex and recombination. *Nature Reviews Genetics*. 2002; 3:252-261.

Otwinowski, J., McCandlish, D. M., and Plotkin, J. B. (2018). Inferring the shape of global epistasis. *Proceedings of the National Academy of Sciences*, 115(32), E7550-E7558.

Stroud, J. T. and Ratcliff, W. C. (2025). Long-term studies provide unique insights into evolution. *Nature*, 639(8055), 589-601.

Weinreich, D. M., Lan, Y., Wylie, C. S., and Heckendorn, R. B. (2013). Should evolutionary geneticists worry about higher-order epistasis? *Current opinion in genetics* & *development*, 23(6), 700-707.

Wei, X., and Zhang, J. (2019). Patterns and mechanisms of diminishing returns from beneficial mutations. *Molecular biology and evolution*, 36(5), 1008-1021.

5. Appendix

5.1. **Circuits and triangulations.** A general approach to gene interactions based on polytope theory was introduced in Beerenwinkel et al. (2007). For mathematical background, see De Loera et al. (2010).

The approach uses 20 forms in the variables w_g for analyzing the gene interactions for biallelic 3-locus systems, analogous to the sole form ϵ for a biallelic two-locus landscape discussed in the main text. For instance, the form

$$w_{110} + w_{000} - w_{100} - w_{010},$$

(with last coordinate zero for all genotypes) is relevant for an early phase of the adaptation, and the form

$$w_{111} + w_{001} - w_{101} - w_{011},$$

with last coordinate 1 for all genotypes) is relevant for a late phase of adaptation.

The 20 forms fall into three categories. Six forms, including

$$w_{000} - w_{010} - w_{100} + w_{110},$$

represent faces of the cube; six forms, including

$$w_{000} - w_{011} - w_{100} + w_{111},\ w_{000} - w_{101} - w_{010} + w_{111},\ w_{000} - w_{110} - w_{001} + w_{111}, \tag{1}$$

represent squares that cut the cube into two triangular prisms, and eight forms, including

$$w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000} \tag{2}$$

represent bipyramides. For an explicit list of the 20 forms see Beerenwinkel et al. (2007, Page 1352). If the sign of each form (positive or negative) is the same for two biallelic three-locus landscapes, then the gene interactions are considered similar.

Out of the 20 forms exactly five have support in both w_{000} and w_{111}, specifically the five listed in (1) and (2) above. Consequently, one needs to check five inequalities for determining if $\{000, 111\}$ constitutes a recombination peak, exactly as stated in Section 2.3. The 20 forms constitute the complete set of circuits for a biallelic three-locus system. For a general definition of circuits, let C denote the allele counter defined in the main text, and let $C^j(\sum a_i w_g) = \sum a_i C(w_g)$. For $\epsilon = w_{00} + w_{11} - w_{10} - w_{01}$ one gets

$$C^0(\epsilon) = [1,1] + [0,0] - [0,1] - [1,0] = [0,0], \quad C^1(\epsilon) = [0,0] + [1,1] - [1,0] - [0,1] = [0,0].$$

A circuit can be defined as a linear form in the variables w_g such that

(a) the allele counter C^j equals the zero-vector for all j,
(b) its support (the w_g which appear with non-zero coefficients) is non-empty but minimal with respect to inclusion.

Recent applications of the described approach to interactions in evolutionary biology and ecology includes (Brown et al., 2023; Crona et al., 2023; Eble et al., 2023; Crona, 2020; Eble et al., 2019). A different approach to gene interactions for biallelic systems relies on Walsh coefficients (Weinreich et al., 2013), for mathematical background see also (Greene, 2023), with recent extensions to the multiallelic case (Faure et al., 2024; Greene, 2023; Crona and Greene, 2024).

5.2. **Experiments.** All experiments concern infinite populations that are well mixed. We use a Fisher-Wright type of model where relative offspring proportions are determined by fitness. Mutation is followed by recombination in the reproductive step. Finally the offspring are chosen to be the parents of the subsequent generation. Recombination is modeled so that for the resulting genotype each locus is equally likely to agree with either parent's allele. The experiment halts if the proportion of the peak genotype exceeds 99.99 percent of the population. The mutation rate is 10^{-9} (however, most results are not sensitive for the mutation rate). The speed of adaptation is measured as the number of generation from wild-type population to fixation at the global peak.

Recall that G_L denotes genotypes of distance L from the the peak p. The assumptions are the same as the conditions (i)-(iii) stated in Definition 2.2, i.e.,

(i) The initial population consists of genotypes of distance L from the peak p, and the population is uniformly distributed over such genotypes.

(ii) (a) If there are more than two alleles at each locus, then each mutation $a \mapsto 0$ for $a \neq 0$ increases fitness,
(b) for biallelic systems, each mutation $0 \mapsto 1$ increases fitness.
(iii) Unless otherwise stated, we also assume $f_g = 1$ for all g of distance L from the peak.

The experiments in Tables 3-5 and 7-10 study the threshold T such that recombination slows down recombination if $f_p > T$ and speeds up recombination if $f_p < T$. The main purpose with the experiments was to gain some understanding of the recombination threshold (r-threshold).

For Table 3 the landscape has two loci and three alleles. Two intermediates have fitness 1.1 and the other two 1.2. It follows that $f_{00} < 1.21$ implies systematic diminishing returns epistasis and $f_{00} > 1.44$ increasing returns epistasis. Extrapolating from the biallelic two-locus case, one expects that recombination speeds up adaptation if $f_{00} < 1.21$ and slows down adaptation if $f_{00} < 1.44$, but the outcome is less obvious for values inbetween. The recombination threshold is 1.38, clearly much closer to 1.44 (difference - 0.6) than to 1.21 (difference +0.17). As explained in the main text, an intuitive explanation for the relatively high recombination threshold is that evolution tends to use high fitness intermediates rather than low fitness intermediates.

For Table 4 the landscape has two loci and five alleles. By the general assumptions, the original population is distributed equally on the 16 genotypes with maximal distance from the peak 00, all of them having fitness 1. All the intermediates have different fitness values, the two lowest 1.1 and 1.2, and the highest 1.7 and 1.8. It follows that the recombination threshold is expected to fall between $1.1 \cdot 1.2 = 1.32$ and$1.7 \cdot 1.8 = 3.06$. The recombination threshold is 2.6, again relatively high.

It should be noted that the outcome is sensitive for details. Table 5 investigates the effect of a permutation of fitness values for the intermediates. In brief, the table shows that if the peak can be produced directly from high fitness intermediates, then the recombination threshold is higher.

Tables 6-10 concerns biallelic three-locus systems. For table 6 the single mutants $f_{100} = f_{010} = f_{00} = 1.1$, and the peak $f_{111} = 1.331$ are all fixed and agrees with a multiplicative expectation. The three double mutants have the same fitness, and it varies between 1.15 and 1.31 in the experiments. Consequently, the fitness landscape has mixed curvature for all fitness values of double mutants (except if the value is exactly 1.21). The table shows that recombination speeds up adaptation if $\{000, 111\}$ is a recombination peak (which implies that the double mutants have fitness below the multiplicative expectation), and slows down adaptation otherwise. Recombination speeds up adaptation considerably toward the right end of the interval (435 generations with recombination, versus 1024 generations without recombination).

Table 7 works with assumptions similar to Table 6, and computes recombination thresholds. Tables 7-10 support the claim that recombination slows down adaptation if $\{000, 111\}$ is a recombination peak, and shows that the threshold is considerably lower

TABLE 3. The table shows the fitness values for the genotypes in a two-locus system with three alleles. By assumption, the original population is distributed equally on the four genotypes with maximal distance from the peak 00. Experiments show that the recombination threshold for, i.e., the value such that evolution is equally fast with and without recombination, is $f_{00} = 1.38$. Systematic diminishing returns epistasis requires that $f_{00} < 1.1^2 = 1.21$, and systematic increasing returns epistasis that $f_{00} > 1.2^2 = 1.44$.

g	f_g
10	1.1
01	1.1
20	1.2
02	1.2
11	1
12	1
21	1
22	1
diminishing returns	$f_{00} < 1.21$
increasing returns	$f_{00} < 1.44$
r-threshold	$f_{00} = 1.38$

under some circumstances. Table 10 shows that recombination can speed up adaptation for landscapes with substantial increasing returns epistasis, as long as the very last step of adaptation is almost flat.

TABLE 4. The table shows the fitness values for the genotypes in a two-locus system with five alleles. By assumption, the original population is distributed equally on the 16 genotypes with maximal distance from the peak 00. Experiments show that the recombination threshold for, i.e., the value such that evolution is equally fast with and without recombination, is $f_{00} = 2.6$. From the theory, the threshold is between $1.1 \times 1.2 = 1.32$ and $1.7 \times 1.8 = 3.06$, so that the threshold falls closer to the upper bound. For a reference, the condition that all sets $\{00, g : g \in G_2\}$ are recombination peaks implies that $f_{00} \geq 3.06$.

g	f_g
10	1.1
01	1.2
20	1.3
02	1.4
30	1.5
03	1.6
40	1.7
04	1.8
$g \in G_2$	1
r-threshold	$f_{00} = 2.6$

TABLE 5. The fitness values for Tables a) and b) are the same, except for a permutation of fitness values for the intermediate genotypes, and similarly for Tables c) and d). The tables show that the permutations have a substantial impact on the recombination thresholds.

a)

g	f_g
10	1.4
01	1.4
20	1.8
02	1.8
11	1
12	1
21	1
22	1
r-threshold	$f_{00} = 2.93$

b)

g	f_g
10	1.4
01	1.8
20	1.4
02	1.8
11	1
12	1
21	1
22	1
r-threshold	$f_{00} = 2.52$

c)

g	f_g
10	1.3
01	1.4
20	1.5
02	1.6
11	1
12	1
21	1
22	1
r-threshold	$f_{00} = 2.32$

d)

g	f_g
10	1.3
01	1.6
20	1.4
02	1.5
11	1
12	1
21	1
22	1
r-threshold	$f_{00} = 2.16$

TABLE 6. The fitness values are fixed for $000, 100, 010, 001, 111$ (Table 6a). The fitness values for the three double mutants are the same in each experiment. The key assumption is that the fitness f_{111} equals the multiplicative expectation based on the fitness values $f_{000}, f_{100}, f_{101}, f_{001}$. Table 6b) shows how the effect of recombination varies depending on the fitness values for the three double mutants. The number of generations until fixation with and without recombination is given. The table also indicates whether or not $\{000, 111\}$ is a recombination peak. Notice that fitness is multiplicative exactly if the double mutations have fitness 1.21.

a)

g	f_g
000	1
100	1.1
010	1.1
001	1.1
111	$1.331 = 1.1^3$

b)

f_{110}, f_{101}, f_{011}	recomb	no recomb	$\{000, 111\}$
1.15	261	238	r-peak
1.20	273	258	r-peak
1.21	278	278	$1.21 = 1.1^2$
1.25	306	407	not r-peak
1.30	435	1024	not r-peak

TABLE 7. The fitness values are fixed for $000, 100, 010, 001$ (Table 7a). The three double mutants have the same fitness value in each experiment (Table 7b). The tables show the recombination threshold along with the threshold that makes $\{000, 111\}$ a recombination peak. Notice that fitness is multiplicative exactly if the double mutations have fitness 1.21.

a)

g	f_g
000	1
100	1.1
010	1.1
001	1.1

b)

f_{110}, f_{101}, f_{011}	r-threshold	$\{000, 111\}$ is an r-peak if $f_{111} >$
1.11	$f_{111} = 1.30$	1.331
1.15	$f_{111} = 1.30$	1.331
1.19	$f_{111} = 1.31$	1.331
1.21	$f_{111} = 1.331$	1.331
1.23	$f_{111} = 1.36$	1.36
1.5	$f_{111} = 1.67$	1.84
1.8	$f_{111} = 2.02$	2.41
2.0	$f_{111} = 2.25$	2.83

TABLE 8. The set $\{000, 111\}$ is a recombination peak if $f_{111} > 1.65$. The recombination threshold was also 1.65.

g	f_g
000	1
100	1.1
010	1.1
001	1.1
110	1.3
101	1.4
011	1.5
r-threshold	$f_{111} = 1.65$

TABLE 9. The set $\{000, 111\}$ is a recombination peak if $f_{111} > 2.02$. The recombination threshold was 1.88.

g	f_g
000	1
100	1.1
010	1.1
001	1.1
110	1.5
101	1.6
011	1.7
r-threshold	$f_{111} = 1.88$

TABLE 10. The set $\{000, 111\}$ is a recombination peak if $f_{111} > 2.52$. The landscape has substantial increasing returns epistasis except for the very last step.

g	f_g
0000	1
1000	1.1
0100	1.1
0010	1.1
0001	1.1
1100	1.3
1010	1.3
1001	1.3
0110	1.3
0101	1.3
0011	1.3
1110	2
1101	2
1011	2
0111	2
r-threshold	$f_{111} = 2.23$