



# Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes

Frank J. Poelwijk, Sorin Tănase-Nicola<sup>1</sup>, Daniel J. Kiviet, Sander J. Tans\*

AMOLF Institute, Science Park 104, 1098 XG, Amsterdam, the Netherlands

## ARTICLE INFO

### Article history:

Received 10 June 2010

Received in revised form

9 December 2010

Accepted 9 December 2010

Available online 16 December 2010

### Keywords:

Epistasis

Fitness landscape

Adaptation

Evolution

Specific interactions

## ABSTRACT

Having multiple peaks within fitness landscapes critically affects the course of evolution, but whether their presence imposes specific requirements at the level of genetic interactions remains unestablished. Here we show that to exhibit multiple fitness peaks, a biological system must contain reciprocal sign epistatic interactions, which are defined as genetic changes that are separately unfavorable but jointly advantageous. Using Morse theory, we argue that it is impossible to formulate a sufficient condition for multiple peaks in terms of local genetic interactions. These findings indicate that systems incapable of reciprocal sign epistasis will always possess a single fitness peak. However, reciprocal sign epistasis should be pervasive in nature as it is a logical consequence of specificity in molecular interactions. The results thus predict that specific molecular interactions may yield multiple fitness peaks, which can be tested experimentally.

© 2010 Elsevier Ltd. All rights reserved.

## 1. Introduction

Organisms are highly integrated functional systems, where change in one component often affects the functioning of other components and the system as a whole. That this interdependence of parts presents challenges for evolutionary processes was already remarked by Darwin (1859). Manifestations of these dependencies at the genetic and physiological level are pleiotropy, where a single gene or mutation has an effect on multiple phenotypic traits, and epistasis, which classifies the interaction between genes or mutations in their effect on fitness. The molecular basis for epistatic effects often lies in the physical interactions within or between gene products (Kogenaru et al., 2009), but it can for example also result from the interplay between protein stability and catalytic activity (DePristo et al., 2005), from Watson–Crick base pairing within an RNA molecule (Kirby et al., 1995), or from interactions between modular metabolic networks (Segre et al., 2005).

Epistatic interactions between parts may constrain the evolutionary optimization of a biological system. For instance, they may limit the number of trajectories towards a fitness optimum that are selectively accessible, with the latter meaning that each mutational step along the trajectory must be fitness increasing (Weinreich et al., 2005). It has been shown that a necessary and sufficient condition for such a reduced accessibility is the presence

of so-called sign-epistatic interactions (Weinreich et al., 2005). Sign-epistasis refers to the situation where the fitness effect of a mutation  $A \rightarrow a$  can have a different sign depending on the genetic background  $B$  or  $b$  (Fig. 1, left). As a consequence, the mutations must occur in a specific order for the paths to be selectively accessible. Sign-epistatic interactions between mutations have been observed empirically, by reconstructing neighboring genotypes and determining their fitness (Lunzer et al., 2005; Bridgham et al., 2006; Weinreich et al., 2006).

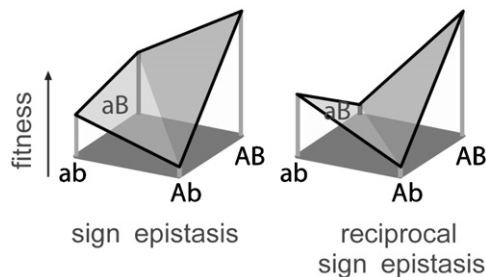
A different situation arises when the fitness landscape contains multiple peaks. By definition, not a single path is selectively accessible by single-mutation steps between two peaks (i.e., all paths exhibit fitness decreases). Escape from a suboptimal peak may still be possible, but it involves significantly increased waiting times as specific combinations of genetic changes must appear in a population (Phillips, 1996; Weissman et al., 2009), or requires large population sizes that support many mutant genotypes at a given time (Weinreich et al., 2005). Entrapment on sub-optima and its relation to landscape ruggedness has principally been studied theoretically, for instance within the mathematical framework of NK landscape models that were pioneered by Kauffman (1993). Empirical evidence for multiple adaptive peaks is scarce, though indications have been found in experimental evolution studies (Korona et al., 1994; Salverda, 2008), and by genotype–phenotype mapping (Poelwijk et al., 2006, 2007; de Visser et al., 2009; Lozovsky et al., 2009).

In this paper we investigate the relation between epistasis and multiple peaks. It has long been recognized that interaction between mutations can cause landscape ruggedness (Wright, 1932), and the involvement of epistasis in multiple peaks has been

\* Corresponding author.

E-mail address: [tans@amolf.nl](mailto:tans@amolf.nl) (S.J. Tans).

<sup>1</sup> Present address: Dept. of Physics, Emory University, Atlanta, Ga 30322, USA.



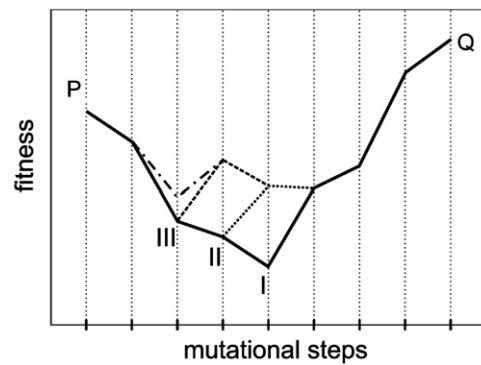
**Fig. 1.** Different manifestations of epistasis along a path from a suboptimal allele  $ab$  towards the optimal  $AB$ . *Left:* Sign epistasis: the fitness effect of a mutation from  $a$  to  $A$  differs in sign (is either beneficial or deleterious) depending on whether the other locus is  $b$  or  $B$ . *Right:* Reciprocal sign epistasis: the fitness effect of a mutation from  $a$  to  $A$  differs in sign depending on whether another locus is  $b$  or  $B$ , and vice versa, the effect of mutation  $b$  to  $B$  differs in sign depending on whether the other locus is  $a$  or  $A$ .

discussed (Weinreich et al., 2005). Here we show that when the landscape harbors multiple peaks, it necessarily contains an epistatic motif that is referred to as reciprocal sign epistasis (Fig. 1, right) (Poelwijk et al., 2007). Put differently, multiple fitness peaks cannot be formed by regular sign-epistasis or magnitude epistasis alone. Reciprocal sign epistasis occurs when two mutations are separately deleterious but jointly enhance fitness: mutation  $A \rightarrow a$  can be advantageous or deleterious, depending on the genetic background  $B$  or  $b$ , while conversely mutation  $B \rightarrow b$  can also be advantageous or deleterious, depending on the genetic background  $A$  or  $a$ . We refer to the appendix for a more formal definition of reciprocal sign epistasis. An intuitive physical model for a reciprocal sign-epistatic interaction is that of a lock and a key: if one changes either the lock or key, the key will no longer fit, but changing them both can produce a new matching lock and key combination. Reciprocal sign epistasis thus implies the existence of underlying functional and structural elements within the biological system that are distinct from those that give rise to regular sign-epistasis. Specific lock–key recognition capabilities are pervasive in organisms, and we therefore anticipate that reciprocal sign-epistasis is a crucial property of biological systems.

## 2. Results

To demonstrate that reciprocal sign epistasis is a necessary condition for the existence of multiple peaks, we have to show that multiple peaks imply the existence of a reciprocal sign epistasis motif somewhere in the landscape. Here we describe a graphical method to determine these motifs, of which there may be multiple within the landscape, by considering direct paths from one maximum  $P$  to another maximum  $Q$ , in a bi-allelic multi-locus system. We consider a path between  $P$  and  $Q$  to be direct when the number of single mutations along it is equal to the Hamming distance  $d_H$ , the number of sequence positions where  $P$  and  $Q$  differ. In the appendix we provide a more rigorous version of this proof.

If we take a direct path between two maxima in a bi-allelic multi-locus system (Fig. 2, solid line), we will find the fitness minimum (indicated with I in Fig. 2) at some position, defined by the number of mutations away from the first peak. We denote the mutation leading into this minimum by  $A \rightarrow a$ , and the following mutation leading away from it by  $B \rightarrow b$ . We can now try to optimize the minimum fitness of this path by exchanging these two mutations, so  $B \rightarrow b$  first, followed by  $A \rightarrow a$ . In many cases the reversal will cause the minimum to shift to a different position along the path between  $P$  and  $Q$  (e.g. to position II in Fig. 2). If it does, another mutation ( $C \rightarrow c$ ) will now lead into or out of the new minimum, and one can thus perform another reversal. This optimization process can be



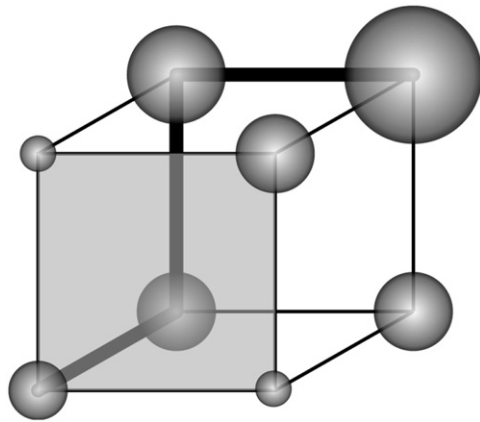
**Fig. 2.** Procedure for finding the reciprocal sign epistasis motif. A random direct path between peaks  $P$  and  $Q$  has a minimum at location I. By reversing the order of the mutations leading to and from this minimum, a new minimum occurs at location II. Again the order of the two mutations around this minimum are reversed. A new minimum occurs at III. When the relevant mutations are reversed here, however, the minimum does not change its location. At this location a reciprocal sign epistasis motif is found.

continued until the minimum remains located at the same position (position III in Fig. 2). At this point, no new reversal can be applied around the minimum, and the optimization procedure thus comes to an end. As the graph illustrates, these two mutations exhibit reciprocal sign epistasis by definition, thus proving the presence of this motif.

In the above method, different starting points may locate different reciprocal sign epistasis motifs that can exist within the landscape. To establish the presence of one specific reciprocal sign epistasis motif, namely the one that is highest in fitness, one can follow a simpler argument. Among the paths connecting two peaks, there always exists one unique path that has the highest fitness minimum. When the two mutations around this minimum are exchanged, the new minimum must be lower. Any other outcome would contradict with the assumption that the unique starting path has the highest fitness minimum. Hence reciprocal sign epistasis is established.

We also explored whether the presence of reciprocal sign epistasis implies the existence of multiple peaks, and more generally, whether a sufficient criterion for multiple peaks can be formulated in terms of local two-way genetic interactions. For this purpose we first consider the simple case of a two-allele system detailed in Fig. 3, where each axis represents one locus, the two points along each axis represent the two alleles, and the size of the points denotes the fitness. The two loci indicated by the grayed front face exhibit reciprocal sign epistasis, with the high fitness bottom-left and double mutant top-right separated by low-fitness single mutants. However, a mutation in the third locus does provide an accessible path to higher fitness from the bottom-left, indicating that it does not represent a peak in this landscape, which thus has just a single peak. This simple example shows that reciprocal sign epistasis does not imply the presence of multiple peaks.

The question remains whether other criteria can be formulated that indicate the presence of multiple peaks in terms of local two-way interactions. The generic possibility of having bridging paths along unexplored loci suggests that this may be difficult, and that only a global mapping of the relevant genotype space with all its higher-order genetic interactions provides a sufficient criterion. Insight into the possibility of defining a sufficient criterion can be found using Morse theory (Milnor, 1963). Morse theory is a powerful tool to investigate the topology of a smooth manifold (surface or space) using the properties of differentiable functions on the manifold. In particular it provides a set of inequalities relating the number of critical points of the function (minima,



**Fig. 3.** Schematic diagram illustrating that reciprocal sign epistasis is not a sufficient criterion for having multiple peaks. The corners of the cube represent the genotypes of a 2-allelic, 3-locus system. The sizes of the spheres represent fitness. The front face indicates the presence of reciprocal sign epistasis and hence two maxima, but the third dimension provides a bridge between them, thus limiting the number of peaks to one.

maxima, saddles, etc.). For example, for functions defined on the continuous real manifold of  $N$  dimensions, a maximum is a critical point with  $N$  negative directions and zero positive directions. Morse theory indicates that the number of saddles with only one positive direction is larger than the number of maxima minus one. This inequality can be compared to our proof (Fig. 2 and appendix) that at least one reciprocal sign epistasis motif, which has similar properties as a saddle, must exist when there are two maxima.

More generally, Morse inequalities state that the number of maxima is larger than the number of saddles of order one (saddle with one positive direction) minus the number of saddles of order two (saddles with two positive directions). For instance, if there would be  $N$  saddles of order 1, then one could still have just one peak, as there could be also  $N$  saddles of order 2. This statement has a straightforward consequence, namely that to know the number of maxima one has to know *all* the saddles of order two, and hence have information on the complete surface. By analogy to fitness landscapes, this statement would imply that to determine the number of fitness peaks, one must have exhaustive information on the complete landscape, again indicating that information on local two-way interactions is not sufficient. This argumentation is intuitively captured in the illustration in Fig. 3: even if reciprocal sign epistasis exists, other parts of the landscape can bridge high points and hence smooth-out the ruggedness.

We are not aware of a generalization of Morse theory that applies directly to genotype spaces. Morse theory has been extended to discrete functions defined on spaces obtained by gluing together points, lines, triangles and larger dimensional counterparts (Forman, 1998), which represents a step towards this aim. It would be of great interest to formally extend Morse theory to genotype spaces. Taken together, the arguments suggest that it is not possible to derive a lower bound on the number of maxima based only on the presence or absence of a small number of local two-way interactions, without knowing higher-order genetic interactions.

### 3. Discussion

The presence of multiple peaks may be considered to be one of the features within fitness landscapes that most critically affect the course of evolution since it precludes adaptation by stepwise mutational optimization. In multi-peaked landscapes, evolutionary trajectories can become trapped on suboptimal peaks, which may

lead to (temporary) evolutionary stasis. While empirical information on the ruggedness of fitness landscapes remains scarce, significant progress has recently been made by systematically reconstructing neighboring genotypes that affect a specific molecular function, such as the resistance against an antibiotic (Burch and Chao, 1999; Buckling et al., 2003; Lunzer et al., 2005; Bridgham et al., 2006; Miller et al., 2006; Weinreich et al., 2006; Fernandez et al., 2007), which has yielded some indications for the presence of multiple peaks (Korona et al., 1994; Poelwijk et al., 2006, 2007; Salverda, 2008). However, conclusively demonstrating that two adaptive peaks are distinct – and not interconnected by a ridge that circumvents the adaptive valley – remains difficult if not impossible, due to the high-dimensional nature of sequence space (Whitlock et al., 1995). Moreover, global fitness maps do not directly provide insight into the elementary properties of the biological system that give rise to specific landscape features. Indeed, the extent to which system architecture and shape of the landscape are fundamentally related remains incompletely answered.

Our current investigation of the relation between the presence of multiple fitness peaks and epistatic interactions has been motivated by these unresolved issues. We find that the presence of multiple fitness peaks requires a specific form of epistatic interactions, where two mutations are separately unfavorable but jointly advantageous. This so-called reciprocal sign epistasis motif (Poelwijk et al., 2006, 2007; Salverda 2008) reflects a mutual dependence between parts at the functional or physico-chemical level that resembles the specific interaction between a lock and a key. Thus, our results imply that to escape from a suboptimal adaptive peak, selection must overcome at least one key–lock type interaction.

Our analysis also shows that the reverse is not true: the occurrence of reciprocal sign epistasis does not guarantee multiple fitness peaks. Moreover, using Morse theory we show that one cannot formulate a local condition in terms of epistatic interactions for the existence of multiple fitness peaks. In other words, local landscape topology (epistatic interactions) cannot conclusively inform about global landscape properties (the presence of multiple peaks). This means that rigorously proving the existence of multiple peaks requires a systematic mapping of genotype space. While this approach may be unfeasible for most biological functions as sequence spaces rapidly become too large for exhaustive experimentation, it may be achievable for specific biological functions that are confined to a small physical region, and hence entail a limited sequence space. Our findings suggest that multi-peaked fitness landscapes may be observed for biological functions involving specific interactions, such as the binding between transcription factor and DNA binding site or residue interactions within a protein fold, which can be tested experimentally.

### Acknowledgement

This work is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie, which is financially supported by the Nederlandse Organisatie voor Wetenschappelijke Onderzoek.

### Appendix

In this appendix we prove the following theorem:

In a  $N$ -allelic  $L$  locus system, reciprocal sign epistasis is a necessary condition for the existence of multiple peaks in the fitness landscape.

Let us first define the different types of epistasis referred to in the main text.

**Definition 1.** Epistasis: means that the fitness effect of a mutation is conditional on the presence of other mutations (the ‘genetic background’).

$\Delta w_{ab \rightarrow Ab} \neq \Delta w_{aB \rightarrow AB}$ , where  $\Delta w$  is the fitness difference between two mutational states, and  $B$  and  $b$  can be considered to be the genetic background for states  $A$  and  $a$ .

For one mutation in different backgrounds, several classes of epistasis can be discerned.

**Definition 2.** Magnitude epistasis: means that the *magnitude* of the fitness effect of a mutation is conditional on the presence of other mutations.

$$\Delta w_{ab \rightarrow Ab} \neq \Delta w_{aB \rightarrow AB}$$

AND

$$|\Delta w_{ab \rightarrow Ab} + \Delta w_{aB \rightarrow AB}| = |\Delta w_{ab \rightarrow Ab}| + |\Delta w_{aB \rightarrow AB}|$$

**Definition 3.** Sign epistasis: means that the *sign* of the fitness effect of a mutation is conditional on the presence of other mutations.

$$|\Delta w_{ab \rightarrow Ab} + \Delta w_{aB \rightarrow AB}| < |\Delta w_{ab \rightarrow Ab}| + |\Delta w_{aB \rightarrow AB}|$$

**Definition 4.** Reciprocal sign epistasis: means that the sign of the fitness effect of mutation  $A \rightarrow a$  is conditional on whether the state of another locus is  $b$  or  $B$ , and vice versa.

$$|\Delta w_{ab \rightarrow Ab} + \Delta w_{aB \rightarrow AB}| < |\Delta w_{ab \rightarrow Ab}| + |\Delta w_{aB \rightarrow AB}|$$

AND

$$|\Delta w_{ab \rightarrow aB} + \Delta w_{Ab \rightarrow AB}| < |\Delta w_{ab \rightarrow aB}| + |\Delta w_{Ab \rightarrow AB}|$$

We start by narrowing down the scope of systems for which we need to prove the theorem:

1. If reciprocal sign epistasis must occur for *direct* paths between two peaks, then reciprocal sign epistasis must also occur for *all* paths, as they include both *direct* and *indirect* paths. Hence, we can restrict ourself to *direct* paths between two peaks.
2. The set of direct paths with length  $d$  between two points in a  $N$ -allelic  $L$  locus system is equal to the set of direct paths of length  $d$  in the bi-allelic  $L$  locus system. This is rather trivial: *direct* paths between two points in sequence space ( $d - d_H = 0$ ) require only *one* substitution at each locus, thus only *two* alleles. Hence we can restrict ourself to the bi-allelic case.
3. A bi-allelic  $L$  locus system may contain more than two peaks. However, we may consider the sub-space between any two peaks in the full landscape, defined by the loci where their sequences differ. If this sub-space should contains reciprocal sign epistasis, then the full landscape also contains it. Hence, we can restrict ourself to two peaks.

So finally, we are left to prove the following theorem:

In a bi-allelic  $L$  locus system, reciprocal sign epistasis is a necessary condition for the existence of two peaks at a distance  $L$  in the fitness landscape.

The following proof will show that along a path containing the highest fitness minimum in the landscape, at least two mutations exhibit reciprocal sign epistasis. Because all paths between two maxima necessarily go through a minimum, the landscape must contain reciprocal sign epistasis.

- (1) There are two maxima at distance  $L$  in a bi-allelic  $L$  locus system.
- (2) We consider a random direct path (length  $L$ ) from one maximum to the other one.

- (3) This path necessarily contains a minimum.
- (4) We change the path, by reversing the order of the two mutations leading to and away from the minimum.
- (5) Two cases are possible:
  - (a) If the fitness minimum of the path is still located between the same mutations, then these mutations exhibit reciprocal sign epistasis.
  - (b) If the fitness minimum is now located at another step along the path, go back to point (4), now for the two mutations around this new minimum.
- (6) The repeating of steps (4) and (5) necessarily leads to breaking the loop via case (5a), because the maximization of the minimum along the path is necessarily bounded (by the value of the lowest peak).

This completes the proof the theorem. ■

## References

- Bridgham, J.T., Carroll, S.M., Thornton, J.W., 2006. Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312, 97–101.
- Buckling, A., Wills, M.A., Colegrave, N., 2003. Adaptation limits diversification of experimental bacterial populations. *Science* 302 (5653), 2107–2109.
- Burch, C.L., Chao, L., 1999. Evolution by small steps and rugged landscapes in the RNA virus phi6. *Genetics* 151 (3), 921–927.
- Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection*. Murray, London.
- de Visser, J.A.G.M., Park, S.C., Krug, J., 2009. Exploring the effect of sex on empirical fitness landscapes. *Am. Nat.* 174, S15–S30.
- DePristo, M.A., Weinreich, D.M., Hartl, D.L., 2005. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* 6, 678–687.
- Fernandez, G., Clotet, B., Martinez, M.A., 2007. Fitness landscape of human immunodeficiency virus type 1 protease quasiespecies. *J. Virol.* 81 (5), 2485–2496.
- Forman, R., 1998. Morse theory for cell complexes. *Adv. Math.* 134 (1), 90–145.
- Kauffman, S.A., 1993. *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press, Oxford.
- Kirby, D.A., Muse, S.V., Stephan, W., 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc. Natl. Acad. Sci. USA* 92 (20), 9047–9051.
- Kogenaru, M., de Vos, M.G., Tans, S.J., 2009. Revealing evolutionary pathways by fitness landscape reconstruction. *Crit. Rev. Biochem. Mol. Biol.* 44 (4), 169–174.
- Korona, R., Nakatsu, C.H., Forney, L.J., Lenski, R.E., 1994. Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. *Proc. Natl. Acad. Sci. USA* 91, 9037–9041.
- Lozovsky, E.R., Chookajorn, T., Brown, K.M., Imwong, M., Shaw, P.J., Kamchonwongpaisan, S., Neafsey, D.E., Weinreich, D.M., Hartl, D.L., 2009. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci. USA* 106 (29), 12025–12030.
- Lunzer, M., Miller, S.P., Felsheim, R., Dean, A.M., 2005. The biochemical architecture of an ancient adaptive landscape. *Science* 310, 499–501.
- Miller, S.P., Lunzer, M., Dean, A.M., 2006. Direct demonstration of an adaptive constraint. *Science* 314, 458–461.
- Milnor, J., 1963. *Morse Theory*. Princeton University Press, Princeton.
- Phillips, P.C., 1996. Waiting for a compensatory mutation: phase zero of the shifting-balance process. *Genet. Res.* 67, 271–283.
- Poelwijk, F.J., Kiviet, D.J., Tans, S.J., 2006. Evolutionary potential of a duplicated repressor-operator pair: simulating pathways using mutation data. *PLoS Comput. Biol.* 2, e58.
- Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., Tans, S.J., 2007. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445, 383–386.
- Salverda, M., 2008. *On the Natural and Laboratory Evolution of an Antibiotic Gene*. Wageningen University, Wageningen.
- Segre, D., Deluna, A., Church, G.M., Kishony, R., 2005. Modular epistasis in yeast metabolism. *Nat. Genet.* 37 (1), 77–83.
- Weinreich, D.M., Delaney, N.F., Depristo, M.A., Hartl, D.L., 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312, 111–114.
- Weinreich, D.M., Watson, R.A., Chao, L., 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59, 1165–1174.
- Weissman, D.B., Desai, M.M., Fisher, D.S., Feldman, M.W., 2009. The rate at which asexual populations cross fitness valleys. *Theor. Populat. Biol.* 75 (4), 286–300.
- Whitlock, M.C., Phillips, P.C., Moore, F.B., Tonsor, S.J., 1995. Multiple fitness peaks and epistasis. *Ann. Rev. Ecol. Syst.* 26 (1), 601–629.
- Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: *Proceedings of the 6th International Congress of Genetics*, vol. 1, pp. 356–366.