Chapter 3

Solutions to Exercises

Exercise 3.1. Explain why for a fitness landscape such that all arrows point up, there is no sign epistasis.
Solution. Recall that sign epistasis occurs when $w_{AB} < w_{aB}$ or $w_{AB} < w_{Ab}$. In the first case, this means that the fitness of the genotype corresponding to AB is lower than the fitness of the genotype aB. This results in an arrow from AB *to* aB, that is, a "down" arrow. (This is illustrated in graphs (b) and (c) in Figure 3.3.) The condition $w_{AB} < w_{Ab}$ results in a similar picture. □

Exercise 3.2. Explain why fitness graphs are acyclic.

Solution. If there is a cycle, then we have a sequence of arrows starting and ending with the same genotype. Each genotype in the sequence has greater fitness than the previous one and thus has greater fitness than all the genotypes that come before it in the cycle. But this would mean that any genotype in the sequence must have greater fitness than itself, which is not possible. □

Exercise 3.3. *Choose the correct statement and explain your answer*: Antimicrobial drug resistance usually depends on several mutations. Sometimes there are constraints for the order in which mutations occur. For instance, a mutation A may only be beneficial if a mutation B has already occurred. One can take advantage of such constraints for managing resistance problems. Constraints on the order in which mutations occur require

(A) epistasis but not sign epistasis
(B) sign epistasis, but not reciprocal sign epistasis
(C) reciprocal sign epistasis.

Solution. The situation described is displayed in graph (c) of Figure 3.3. □

Exercise 3.4. *Choose the correct statement and explain your answer*: Some fitness landscapes have multiple peaks. Restrict to three-loci systems. Multiple peaks require

(A) epistasis but not sign epistasis
(B) sign epistasis, but not reciprocal sign epistasis
(C) reciprocal sign epistasis.

The result you showed for three-loci systems holds true in general. See Project 3.1 for a study of the theory.

Solution. Consider fitness graphs for three loci, such that the landscape has at least two peaks. By relabeling nodes if necessary, we may assume that 000 is at a peak, so that all three arrows between level 1 and level 0 point toward 000. It follows that 100, 010, and 001 are not at peaks. If one of the double mutants is at a peak, then one immediately gets a case of reciprocal sign epistasis (e.g., if 110 is at the peak, consider the nodes 000, 100, 010, and 110). It only remains to consider fitness graphs with peaks 000 and 111. The direction of six arrows is determined by our assumptions. Assume that there is no case of reciprocal sign epistasis, for a contradiction. It remains to determine the direction of six arrows. Again by relabeling the nodes if necessary and by the symmetry of the arrows established so far, we may assume that the arrow between 110 and 100 points down. From sketching the graph, one can determine the direction of the remaining five arrows. Indeed, first one concludes that the arrow between 100 and 101 points up because otherwise one gets reciprocal sign epistasis. Using this observation, the arrow between 101 and 001 points down by the same argument, and so forth. The resulting graph has a cycle

Algebraic and Discrete Mathematical Methods for Modern Biology. http://dx.doi.org/10.1016/B978-0-12-801213-0.00022-8

$$110 \mapsto 100 \mapsto 101 \mapsto 001 \mapsto 011 \mapsto 110,$$

which is a contradiction. □

Exercise 3.5. How many fitness graphs can one construct for two loci?

Solution. The fitness graph is determined by the direction of four arrows. Out of the $2^4 = 16$ possibilities, exactly two choices give us cyclic graphs. Each one of the remaining 14 alternatives could be a fitness graph. □

Exercise 3.6. How many mathematically different types of fitness graphs are there for two loci? (Graphs belong to the same type whenever they differ only in the labeling of the vertices.)

Solution. The three types correspond to no sign epistasis, sign epistasis but not reciprocal sign epistasis, and reciprocal sign epistasis. These are displayed in Figure 3.3 by graphs (a), (b) and (c), and (d), respectively. □

Exercise 3.7. How many biologically distinct fitness graphs are there for two loci? That is, how many types of graphs are there, if the special status of 00 as the wild type is taken into consideration? Motivate your answer.

Solution. Two fitness graphs are biologically equivalent if one can relabel the genotypes by changing places of 10 and 01. □

Exercise 3.8. Explain why additive fitness landscapes are all arrows up landscapes.

Solution. To answer this question, we work locally; that is, we can consider pairs of mutations as illustrated in Figure 3.3. In graph (b), for example, we have that $w_{ab} < w_{Ab}$ and $w_{AB} < w_{aB}$. Therefore, $w_{ab} + w_{AB} < w_{aB} + w_{Ab}$. Graphs (c) and (d) can be analyzed in a similar way. □

Exercise 3.9. Explain why any shortest walk to the peak is an adaptive walk in an all arrows up landscape.

Solution. By relabeling nodes if necessary, we may assume that $1, \dots, 1$ is at the peak. By definition, for each step in a shortest walk to the peak, the Hamming distance to the peak increases. With our assumptions, each step toward the peak follows the direction of the arrow in the fitness graph. Consequently any shortest walk is an adaptive walk. □

Exercise 3.10. Consider single-peaked landscapes. Find an example where there is an adaptive walk to the peak which contains a reversion.

Solution. For three loci, construct a fitness graph where all arrows are up except the one between 010 and 110. The walk

$$000 \mapsto 100 \mapsto 110 \mapsto 010 \mapsto 011 \mapsto 111$$

contains a reversion. □

Exercise 3.11. Discuss the role of reversed mutations in adaptation in view of the previous two problems.

Solution. It seems like reversed mutations are most likely to be important for complex fitness landscapes. For all arrows up landscapes, there do not even exist walks to the peak with reversions. □

Exercise 3.12. Consider an all arrows up graph for $L = 2$. We say that the corresponding fitness landscape has *negative epistasis* if $w_{11} < w_{10} + w_{01} - w_{00}$. Positive epistasis is defined similarly. Give examples showing that the fitness landscape may have positive, negative, or no epistasis. You may use relative fitness values in your examples.

Solution. Each of the following fitness landscapes, where w_g denotes relative fitness, gives an all arrows up graph:

- no epistasis: $w_{00} = 1$, $w_{01} = 1.1$, $w_{10} = 1.1$, $w_{11} = 1.21$
- positive epistasis: $w_{00} = 1$, $w_{01} = 1.1$, $w_{10} = 1.1$, $w_{11} = 2$
- negative epistasis: $w_{00} = 1$, $w_{01} = 1.1$, $w_{10} = 1.1$, $w_{11} = 1.15$ □

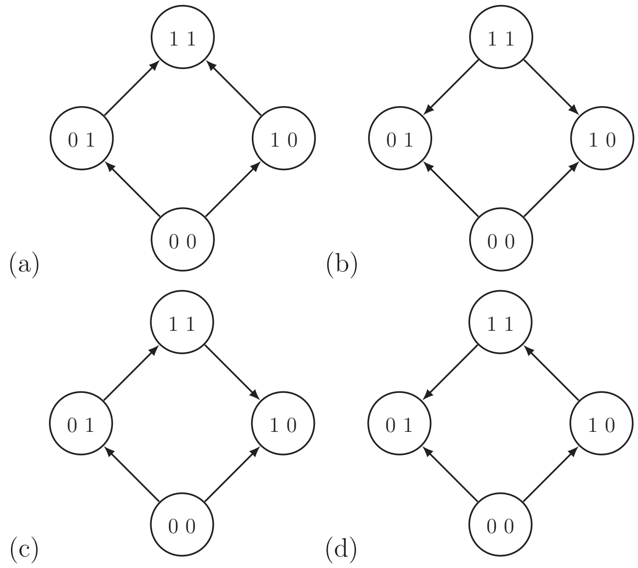


FIGURE S3.1

Exercise 3.13. Assume that $L = 2$ and both single mutants have higher fitness as compared to the wild type. Construct all possible fitness graphs under the assumption that epistasis is negative (in the sense of the previous problem) and give possible relative fitness values for each graph.

Solution. Such a situation can occur for any graph with up arrows between level 0 and level 1.

Possible fitness values for

- graph (a): $w_{00} = 1$, $w_{01} = 1.2$, $w_{10} = 1.2$, $w_{11} = 1.21$
- graph (b): $w_{00} = 1$, $w_{01} = 1.2$, $w_{10} = 1.2$, $w_{11} = 1.1$
- graph (c): $w_{00} = 1$, $w_{01} = 1.1$, $w_{10} = 1.21$, $w_{11} = 1.2$
- graph (d): $w_{00} = 1$, $w_{01} = 1.21$, $w_{10} = 1.1$, $w_{11} = 1.2$. □

Exercise 3.14. Express the relation between absolute fitness, growth rate, and generation time (i.e., the time between generations), based on the formula

$$N = N_0 e^{\alpha t}.$$

Growth rates of bacteria can be measured in the laboratory, so the relation is important for interpretations.

Solution.

$$\alpha = \frac{\log W}{T}.$$

Our starting point is the formula

$$N = N_0 e^{\alpha t}.$$

For a population where there is no overlap between generations, one can express the number of individuals N as:

$$N = N_0 W^n$$

where W the absolute fitness, and n the number of generations. However, $n = \frac{t}{T}$, where T denotes the generation time and consequently, $N = N_0 W^{\frac{t}{T}}$. It follows that

$$\alpha = \log W^{\frac{1}{T}} = \frac{\log W}{T}.$$

□

Exercise 3.15. The introduction to Section 3.3 described the following situation: Two single mutations are deleterious, but the double mutant combining the single mutations has very high fitness. Draw a fitness graph representing this two-loci system.

Solution. The two arrows between level 0 and 1 point down, whereas the arrows between level 1 and 2 point up. □

Exercise 3.16. Analyze the argument in Section 3.3 described as a “widespread fallacy” by Joseph Muller. Explain why the argument seems fairly weak in most situations.

Solution. If 10 and 01 have lower fitness than the wild type, one would expect them to be rare. The probability that two rare genotypes encounter each other and then recombine seems exceedingly low. □

Exercise 3.17. In Section 3.3, we discussed the impact of recombination for a special four-loci system. An analogous situation for three loci would be a system such that 100, 010, and 001 have higher fitness than the wild type, the double mutants have lower fitness than the single mutants, and 111 has maximal fitness in the system. Analyze the fitness graph in this situation, and explain why recombination is not an obvious advantage.

Solution. The fitness graph for this system is given below.

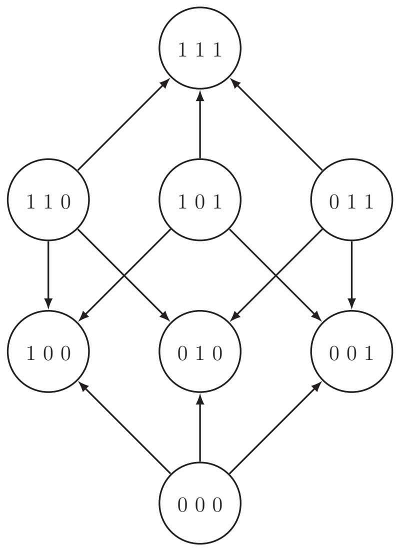


FIGURE S3.2

A single recombination event cannot combine the three different genotypes 100, 010, and 001 to form the 111 genotype. Suppose we have subpopulations consisting of 100, 010, and 001 genotypes. The one could get double mutants as a result of migration and recombination. However, the double mutants would be selected against. For that reason, it is not obvious that migration and recombination would result in the 111 genotype within a short time frame. □

Exercise 3.18. Verify the claim about the four-loci system consisting of TEM-1, TEM-50, and intermediates. List the single and double mutants using standard notation.

Solution. The exercise is straightforward from the record of clinically found TEM mutants provided by the Lahey Clinic at: http://www.lahey.org/Studies/temtable.asp. □

Exercise 3.19. The probabilities for mutations can be expressed in matrix form, as described in the text. Explain why all entries on the main diagonal are 0 or 1.

Solution. An entry on the main diagonal is the probability that a given genotype does not mutate. According to our assumption, this probability is either 0 or 1 and is 1 exactly when the genotype is a peak. □

Exercise 3.20. Recall that the trace of a matrix is the sum of the entries on the main diagonal. Interpret the trace in terms of the fitness landscape associated with the drug.

Solution. According to our assumptions, each entry of the main diagonal is 0 or 1. Consequently, the trace equals the number of peaks. □

Exercise 3.21. Consider four different drugs and a two-loci system. Construct fitness graphs so that a sequence of four drugs induce the mutational trajectory

$$00, 10, 11, 01, 00.$$

This means that successive mutations lead back to the wild type.

Solution. Label the four drugs (in the order that they will be applied) by A, B, C, and D.

We need a drug A such that 10 is more fit than 00, a drug B such that 11 is more fit than 10, a drug C such that 11 is more fit than 10, and a drug D such that 00 is more fit than 10. It is straightforward to construct fitness graphs subject to the conditions listed. Therefore, the fitness graph associated with each drug could be

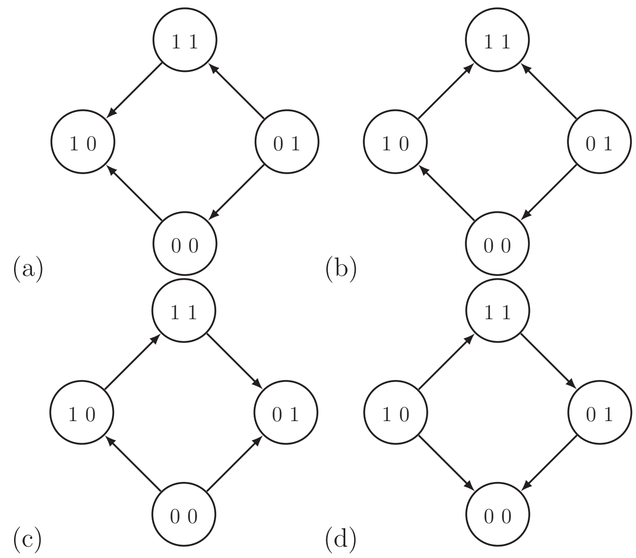


FIGURE S3.3 □

Exercise 3.22. Discuss why a treatment plan as in the previous problem may be valuable for managing resistance problems in hospitals.

Solution. Note that at each step in the mutational trajectory described, the genotype resulting from the application of the drug is at a peak and thus could represent a genotype that is resistant to the drug. However, with this pattern of drug cycling, whenever a microbe became resistant to a given drug, there would always be a different drug available that would be effective. □

Exercise 3.23. For any biallelic system and set of drugs, the maximum probabilities for returning to the wild type depend on how many steps one allows in the treatment plan. The following example demonstrates that the maximum probabilities may increase by the number of steps indefinitely.

Consider a three-loci system where the genotypes are ordered as

$$000, 100, 010, 001, 110, 101, 011, 111.$$

Suppose there are two available drugs, A and B, with transition matrices as given following. Starting at the genotype 100, we see that the probability for returning to 000 by applying A is 0.9. One can calculate that the probability for ending at 000 for A-B-A 0.99, for A-B-A-B-A 0.999, and so forth.

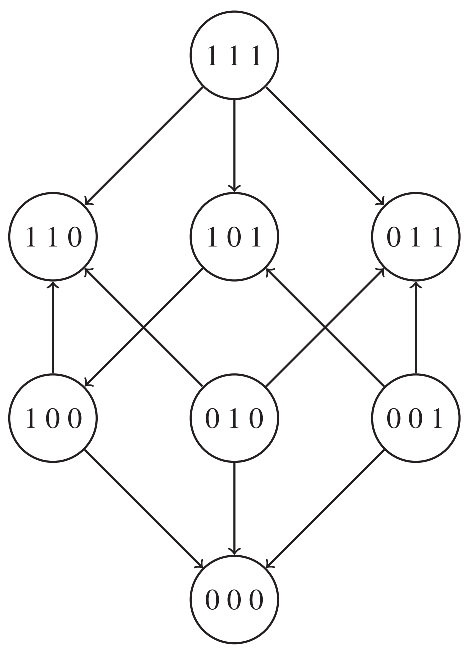
$$M(A) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.9 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$M(B) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 \end{bmatrix}$$

(i) Construct two fitness graphs that are compatible with the matrices A and B,
(ii) Describe the possible evolutionary scenarios in terms of the fitness graphs you found. Use 100 as starting point.

Solution.

(i) For two mutational neighbors u and v, if $u \to v$ in the fitness graph, then v has higher fitness than v. Assuming CPM, the mutation v is more likely to occur so that we expect a positive probability that u is replaced by v. Therefore, the following fitness graph is consistent with the transition matrix A:



Using the same reasoning, the fitness graph below is consistent with the transition matrix B:

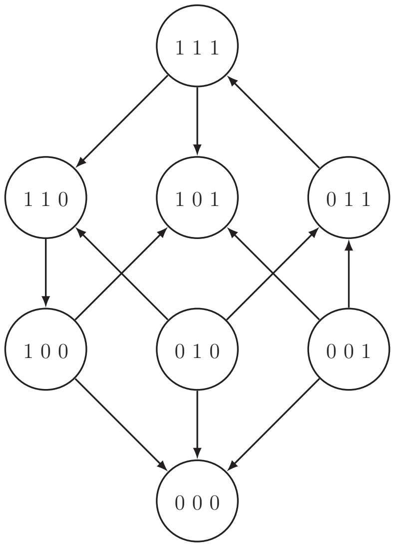


FIGURE S3.4

(ii) In the presence of drug *A*, 100 mutates to either 000 or 110. Note that 000 is a peak in both graphs, so that if 100 mutates to 000, the evolutionary path terminates. If 100 mutates to 110, then when drug *B* is applied, 110 mutates to 100. Then drug *A* is applied and we are in the same situation as initially described. □