Journal of Theoretical Biology 317 (2013) 1–10

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

The peaks and geometry of fitness landscapes

Kristina Crona*, Devin Greene, Miriam Barlow

University of California at Merced, 5200 Lake Road, Merced, CA 95343, United States

HIGHLIGHTS

- We study qualitative aspects of gene interactions and fitness landscapes.
- A sufficient local condition for multiple peaks is given.
- The fitness graph reveals sign epistasis and other coarse properties.
- The shape, as defined in the geometric theory, reveals all gene interactions.
- Fitness graphs and shapes provide complementary information.

ARTICLE INFO

Article history:
Received 2 May 2012
Received in revised form
20 September 2012
Accepted 21 September 2012
Available online 2 October 2012

Keywords:
Fitness landscape
Epistasis
Geometric classification
Directed acyclic graph (DAG)
Polytope

ABSTRACT

Fitness landscapes are central in the theory of adaptation. Recent work compares global and local properties of fitness landscapes. It has been shown that multi-peaked fitness landscapes have a local property called reciprocal sign epistasis interactions. The converse is not true. We show that no condition phrased in terms of reciprocal sign epistasis interactions only, implies multiple peaks. We give a sufficient condition for multiple peaks phrased in terms of two-way interactions. This result is surprising since it has been claimed that no sufficient local condition for multiple peaks exist. We show that our result cannot be generalized to sufficient conditions for three or more peaks. Our proof depends on fitness graphs, where nodes represent genotypes and where arrows point toward more fit genotypes. We also use fitness graphs in order to give a new brief proof of the equivalent characterizations of fitness landscapes lacking genetic constraints on accessible mutational trajectories. We compare a recent geometric classification of fitness landscape based on triangulations of polytopes with qualitative aspects of gene interactions. One observation is that fitness graphs provide information that are not contained in the geometric classification. We argue that a qualitative perspective may help relating theory of fitness landscapes and empirical observations.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

We will study qualitative aspects of gene interactions. In particular, it is of interest to what extent beneficial mutations combine well. This question relates to the concept epistasis. Absence of epistasis means that the fitness effects of mutations sum, where fitness is defined as the expected reproductive success (different definitions of these concepts occur in the literature Mani et al., 2008). It is immediate that beneficial mutations combine well if there is no epistasis. However, it is well known that double mutants which combine beneficial single mutations may have very low fitness. Several examples from different species are given in Weinreich et al. (2005). Put briefly, "good+good=better" if there is no epistasis, but sometimes "good+good=not good" in nature. By a qualitative perspective we understand that one considers fitness ranks of genotypes, but not necessarily more fine scaled information, such as relative fitness values.

Fitness landscapes are central in the theory of adaptation and we will focus on the qualitative perspective. The fitness landscape was initially introduced as a metaphor for adaptation (Wright, 1931). Informally, the surface of the landscape consists of genotypes, where similar genotypes are close to each other, and the fitness of a genotype is represented as a height coordinate. Adaptation can then be pictured as an uphill walk in the fitness landscape.

A qualitative analysis is sufficient for several theoretical aspects of fitness landscapes. Coarse properties of fitness landscapes, such as the number of peaks, depend on fitness ranks of genotypes only. The relation between global and local properties can be analyzed from a qualitative perspective as well. From a more practical point of view, the qualitative perspective has several advantages. Fitness ranks are usually easier to determine as compared to relative fitness values. Fitness ranks tend to be stable under small variations in the environment. Moreover, fitness data of qualitative nature are already available.

---

* Corresponding author. Tel.: +1 209 2339601.
E-mail address: kcrona@ucmerced.edu (K. Crona).

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.09.028

In particular, medical records on HIV drug resistance and antibiotic resistance provides indirect information about fitness ranks (see Section 5). It is frequently claimed that we know virtually nothing about fitness landscapes in nature. In our view, better methods for interpretation of fitness data are at least as important as new fitness measurements.

The concept of a fitness landscapes has been formalized in different ways. Conventionally, as a string in the 20, 4 or 2 letter alphabet, depending on if one considers the amino acids, the base pairs or *biallelic system*. In many real systems at most two alternative alleles occur at each position (or locus), resulting in a biallelic system. Alternatively, a biallelic assumption may be a reasonable simplification. For simplicity, we will consider biallelic populations throughout the paper. Let $\Sigma = \{0,1\}$ and let Σ^L denote bit strings of length L. The *zero-string* denotes the string with zero in all L positions, and the *1-string* denotes the string with 1 in all L positions. We define the *fitness landscape* as a function $w : \Sigma^L \mapsto \mathbb{R}$, which assigns a fitness value to each genotype. The metric we use is the Hamming distance, meaning that the distance between two genotypes equals the number of positions where the genotypes differ. In particular, two genotypes are adjacent, or *mutational neighbors*, if they differ at exactly one position.

A walk in the fitness landscape has a precise interpretation. Consider a population after a recent change in the environment. Assume that the wild-type no longer has optimal fitness. If we assume the strong-selection weak-mutation (SSWM) regime, then a beneficial mutation is assumed to go to fixation in the population before the next mutation occurs (Gillespie, 1983, 1984). The population is monomorphic for most of the time, so that one genotype dominates the population at a particular point in time. It follows that we can think of a Darwinian process as an adaptive walk in the fitness landscape, where each step represents that a beneficial mutation goes to fixation in the population. The described model of adaptation has been widely used and relies on work by Gillespie (1983, 1984). The sequence-based model of adaptation was introduced by Maynard Smith (1970). For more background and references, see also Orr (2002, 2006).

For the qualitative perspective on fitness landscapes, one needs a refined version of the concept epistasis. According to our definition, fitness is *additive* or non-epistatic if fitness effects of mutations sum. (In the literature non-epistatic fitness is sometimes defined as multiplicative fitness.) Suppose that

$$w(00) = 1, \quad w(10) = 1.04, \quad w(01) = 1.02.$$

If one considers 00 as a starting point, then the fitness effect of a mutation at the first locus is $+0.04$, and at the second $+0.02$. If fitness is additive, then $w(11) = 1.06$ since $0.04+0.02 = 0.06$, meaning that the fitness effects sum. Epistasis exists if $w(11) \neq 1.06$. *Sign epistasis* means that a particular mutation is beneficial or deleterious depending on genetic background. For example, if $w(11) = 1.03$, then there is sign epistasis. Indeed, in this case a mutation at the second locus is beneficial for the genotype 00 since $w(01) > w(00)$, and deleterious for the genotype 10 since $w(11) < w(10)$. In contrast, if $w(11) = 1.05$ there is epistasis, but no sign epistasis since fitness increases whenever a 0 at some locus is replaced by 1. For more background about epistasis, see e.g. Weinreich et al. (2005), Beerenwinkel et al. (2007b), Poelwijk et al. (2007, 2011) and Kryazhimskiy et al. (2011). Recent work that considers qualitative properties of fitness landscapes includes Weinreich et al. (2005) and Poelwijk et al. (2007, 2011). A central theme is how global properties of the fitness landscape, such as the number of peaks, relate to local properties, such as sign epistasis (see Sections 2 and 3). A related field is the study of constraints for orders in which mutations accumulate (see e.g. Desper et al., 1999; Beerenwinkel et al., 2007a). It is well known that a drug resistance mutation is sometimes selected for, only if a different mutation has already occurred. Such a phenomenon requires sign epistasis. Indeed, if a particular mutation is beneficial regardless of background, then it can occur before or after other mutations.

We will give an overview of classical models of fitness landscapes, and then compare with recent approaches and the qualitative perspective.

1.1. Classical models of fitness landscapes

Several models of fitness landscapes have had a broad influence in evolutionary biology, primarily additive fitness landscapes, random fitness landscapes, the block model and Kaufman's NK model. Additive fitness landscapes are single peaked. In contrast, for a random (uncorrelated or rugged) fitness landscape (see e.g. Kingman, 1978; Kauffman and Levin, 1987; Flyvbjerg and Lautrup, 1992; Rokyta et al., 2006; Park and Krug, 2008) there is no correlation between the fitnesses of mutational neighbors, or genotypes that differ at one locus only. Random fitness landscapes tend to have many peaks. Random fitness and additivity can be considered as two extremes with regard to the amount of structure in the fitness landscapes.

For the block model (Macken and Perelson, 1995; Orr, 2006) the string representing a genotype can be subdivided into blocks, where each block makes an independent contribution to the fitness of the string. Each block has random fitness, and the fitness of the string is the sum of contributions from each block. In particular, if there is only one block, then the block model coincides with a random fitness landscape.

Kaufmann's NK model (see e.g. Kauffman and Weinberger, 1989) is defined so that the epistatic effects are random, whereas the fitness of a genotype is the average of the "contributions" from each locus. More precisely, for the NK model the genotypes have length N (in our notation $L=N$), and the parameter K, where $0 \leq K \leq N-1$, reflects interactions between loci. The fitness contribution from a locus is determined by its state and the states at exactly K other loci. The key assumption is that this contribution, determined by the 2^{K+1} states (since we assume biallelic systems), is assigned at random from some distribution. The fact that the fitness of the genotype is the average of these N contributions, means that fitness effects of non-interacting mutations sum. Several important properties of NK landscapes depend mainly on N and K, rather than the exact structure of the epistatic interactions.

Notice that the NK model, as well as the block model includes additive landscapes and random landscapes as special cases. More importantly, the models are similar in that there is a sharp division between effects which are completely random and effects which are additive.

In contrast to the models discussed, the Orr-Gillespie theory (e.g. Orr, 2002) depends on the strategy to make minimal assumptions about the underlying fitness landscape, motivated by the fact that our knowledge about fitness landscapes is limited. The theory focuses on properties that hold for a broad category of fitness landscapes. Most results depend on extreme value theory. For more background and references on fitness landscapes in evolutionary biology, see e.g. Weinreich et al. (2005), Beerenwinkel et al. (2007b) and Kryazhimskiy et al. (2009). Fitness landscapes have been used in chemistry, physics and computer science, in addition to evolutionary biology. For a survey on combinatorial landscapes in general see Reidys and Stadler (2002). In combinatorial optimization the fitness function is referred to as the cost function.

1.2. New approaches to the theory of fitness landscapes

The classical theory of fitness landscapes has been criticized for the lack of contact with empirical data (Kryazhimskiy et al., 2009). One sometimes encounters the misunderstanding that the

Download English Version:

https://daneshyari.com/en/article/6370890

Download Persian Version:

https://daneshyari.com/article/6370890

Daneshyari.com