Theoretical Population Biology 82 (2012) 209–221

Contents lists available at SciVerse ScienceDirect

Theoretical Population Biology

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

Epistasis can increase multivariate trait diversity in haploid non-recombining populations

Cortland K. Griswold *, Thomas A. Henry

Department of Integrative Biology, University of Guelph, Canada N1G 2W1

ARTICLE INFO

Article history:
Received 27 April 2012
Available online 4 July 2012

Keywords:
Coalescent
Epistasis
Neutral
Quantitative genetics
RNA
Selection

ABSTRACT

We evaluate the effect of epistasis on genetically-based multivariate trait variation in haploid non-recombining populations. In a univariate setting, past work has shown that epistasis reduces genetic variance (additive plus epistatic) in a population experiencing stabilizing selection. Here we show that in a multivariate setting, epistasis also reduces total genetic variation across the entire multivariate trait in a population experiencing stabilizing selection. But, we also show that the pattern of variation across the multivariate trait can be more even when epistasis occurs compared to when epistasis is absent, such that some character combinations will have more genetic variance when epistasis occurs compared to when epistasis is absent. In fact, a measure of generalized multivariate trait variation can be increased by epistasis under weak to moderate stabilizing selection conditions, as well as neutral conditions. Likewise, a measure of conditional evolvability can be increased by epistasis under weak to moderate stabilizing selection and neutral conditions. We investigate the nature of epistasis assuming a multivariate-normal model genetic effects and investigate the nature of epistasis underlying the biophysical properties of RNA. Increased multivariate diversity occurs for populations that are infinite in size, as well as populations that are finite in size. Our model of finite populations is explicitly genealogical and we link our findings about the evenness of eigenvalues with epistasis to prior work on the genealogical mapping of epistatic effects.

1. Introduction

This paper addresses the question of whether epistasis increases or decreases multivariate trait variation within a population. The focus of analysis is non-recombining haploid populations. Haploid species have been shown to diversify in their multivariate phenotype in the lab (Rainey and Travisano, 1998) and there is evidence that they can locally adapt in nature where locations differ ecologically at the multivariate level (Belotte et al., 2003). Yet, a theoretical understanding of the effect of epistasis on multivariate trait variation in haploid species is lacking.

Multivariate trait variation is measured by a covariance matrix (discrete characters) or a covariance function (function-valued characters). In this paper, we focus on discrete traits. A common approach to measure the amount of variation in a multivariate trait is to perform a principal components analysis on the genetic covariance matrix for the trait. A principal components analysis involves finding the eigenvalues and eigenvectors of a covariance matrix. This analysis decomposes variation into independent directions (eigenvectors), where these eigenvectors explain proportionally less and less of the variation. The variation explained by an eigenvector is its associated eigenvalue. The sum of the eigenvalues and the degree of evenness of eigenvalues indicates the amount of multivariate trait variation (e.g. Wagner, 1984; Kirkpatrick and Lofsvold, 1992; Mezey and Houle, 2005). The sum of the eigenvalues is a measure of the total genetic variation of a multivariate trait and the degree of evenness of eigenvalues is a measure of how variation is distributed across the multivariate trait. Evenness is the extent to which eigenvalues are the same. If all of the eigenvalues are the same, then each principal component has an equal amount of variation and all directions of variation contribute equally to multivariate diversity.

For the univariate case, Hermisson et al. (2003) using a multilinear model of epistasis (Hansen and Wagner, 2001), showed that epistasis decreases genetic variance (additive plus epistatic) in a diploid population undergoing stabilizing selection and assuming linkage equilibrium. This result is suggestive that epistasis may have a negative effect on genetic variation under stabilizing selection. Here we determine whether epistasis reduces the genetic variance of a multivariate trait under stabilizing selection in haploid populations.

Hermisson et al. (2003) modeled a population that is infinite in size and measured the statistical properties of the entire population. Here we model both infinite and finite populations. In the infinite population case we measure the statistical properties of the entire population. In the finite population case, we study the

* Corresponding author.
E-mail address: cgriswol@uoguelph.ca (C.K. Griswold).

0040-5809/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2012.06.007

statistical properties of a sample from this finite population. There are three important differences between the infinite population analysis and the finite population analysis that may affect multivariate trait variation. First, the finite population experiences random genetic drift, which reduces genetic variation. Since there may be fewer polymorphic sites in the population at the time of sampling, there may be a smaller number of additive and epistatic effects that contribute to multivariate trait variation. Second, random sampling further reduces the number of polymorphic sites; thus reducing the number of additive and epistatic effects that contribute to multivariate trait variation. Lastly, the finite population has genealogical structure.

Genealogical structure has been shown to cause the genotypic values of haplotypes to covary (Cavalli-Sforza and Piazza, 1975). Griswold et al. (2007) showed that genealogical structure causes the eigenvalues of multivariate traits to be less even, in principle because of the correlation in genotypic values caused by shared common ancestry. More recently, Griswold and Eisner (2012) showed that epistasis reduces the correlation in genotypic values caused by genealogical structure.

In this paper we make use of a generalized measure of multivariate diversity which is the determinant of a covariance matrix (Wilks, 1932). The determinant of a covariance matrix is equal to the product of its eigenvalues, which is directly related to the geometric mean of the eigenvalues of a covariance matrix. The determinant and the geometric mean capture the evenness of a set of random variables better than the arithmetic mean. For instance, in signal processing, the geometric mean is used as a measure of spectral flatness (Gray and Markel, 1974). The geometric mean of the eigenvalues of **G** has been proposed as a measure of evolvability and is a relatively good predictor of evolvability for low dimensional traits, but may be less so for high dimensional traits (Hansen and Houle, 2008). Similarly, Cheverud et al. (1983) proposed and Wagner (1984) studied the geometric mean of the eigenvalues of the genetic or phenotypic correlation matrix as a measure of phenotypic integration. Generally, if epistasis increases multivariate diversity, we expect the determinant of the covariance matrix to be greater with epistasis compared to non-epistatic cases.

In addition to measuring the total genetic variation and generalized variance of a multivariate trait, we also measure the conditional evolvability of a population (Hansen and Houle, 2008). The conditional evolvability of a population measures the constrained response of a population to selection in the direction of a random selection gradient, where constraint arises from pleiotropically associated characters or genetic polymorphisms in linkage disequilibrium that are under purifying selection.

The outline of the paper is as follows. First we review the univariate model of epistasis first introduced by Griswold and Eisner (2012) and then outline the multivariate extension of this model. Next we define the genetic covariance matrix of a population. We then define the determinant of the genetic covariance matrix as a generalized measure of multivariate diversity and Hansen and Houle's (2008) unconditional and conditional measures of evolvability. Fourth, we present in succession two models of evolution, an infinite population model and a finite population model. We have chosen to consider both infinite and finite populations because in the infinite population there is no genealogical structure, whereas in the finite population there is. For both models, we derive the expected genetic variance of a trait experiencing no selection. These genetic variances are used as a standard to parameterize the results and allows for direct comparisons between results. Fifth, we describe the methods that are used for two case studies. In one case study we assume naively that genetic effects are multivariate and normally distributed. The second case study involves using the biophysical properties of RNA as a model system to understand the nature of genetic effects and multivariate trait evolution. Lastly we follow-up by presenting results and discuss their context and implications.

2. Model components and measures of multivariate diversity

2.1. The univariate and multivariate models of epistasis

Here we extend the epistatic model introduced by Griswold and Eisner (2012) to the multivariate case. The genotype of each individual is haploid and each haplotype consists of L sites. For simplicity, we assume a site can take on one of two states. The model makes use of the concept of a reference haplotype (Hansen and Wagner, 2001). As in Griswold and Eisner (2012), the reference haplotype is represented by the empty set {}. A haplotype that differs in sequence from the reference haplotype is given by a set of numbers that indicate the sites where a substitution occurs relative to the reference haplotype. For instance, the haplotype with substitutions at sites 1, 10 and 23, relative to the reference haplotype is represented by the set {1, 10, 23}.

In the univariate case, the genotypic value of a haplotype H (G_H) is the sum of the set of its genetic effects (g_i). The set of genetic effects for haplotype H is g mapped to the power set ($W(H)$) of H. For instance, the power set of the haplotype $H = \{1, 10, 23\}$ is

$$\langle\langle\rangle, \langle 1\rangle, \langle 10\rangle, \langle 23\rangle, \langle 1, 10\rangle, \langle 1, 23\rangle, \langle 10, 23\rangle, \langle 1, 10, 23\rangle\rangle.$$

We distinguish a set representing a haplotype using "{" and "}" and the set of genetic effects using "⟨" and "⟩". A genetic effect involving a single mutation is called "1st order", two mutations "2nd order", etc. Genetic effects that are 2nd order or higher are epistatic because these effects indicate a deviation from the sum of the genetic effects of each mutation by themselves, i.e. additivity. The genotypic value of a haplotype H is then

$$G_H = \sum_{i \in W(H)} g_i. \tag{1}$$

In the multivariate case, the genotypic value of a haplotype H $\left(\vec{G}_H\right)$ is the sum of the set of its genetic effects $\left(\vec{g}_i\right)$, where $\vec{g}_i$ is a vector in which each element of the vector indicates the genetic effect on a character that makes up a multivariate trait. The genotypic value of a haplotype H is then

$$\vec{G}_H = \sum_{i \in W(H)} \vec{g}_i. \tag{2}$$

The genetic effect $\vec{g}_{\langle 1\rangle}$ is found by comparing the average phenotype of individuals with the reference haplotype to individuals with a haplotype with a substitution at site 1, such that $\vec{g}_{\langle 1\rangle} = \vec{G}_{\{1\}} - \vec{g}_{\langle\rangle}$. Higher order effects are determined in a similar way, for instance the effect $\vec{g}_{\langle 1,10\rangle}$ is calculated as $\vec{g}_{\langle 1,10\rangle} = \vec{G}_{\{1,10\}} - \vec{g}_{\langle\rangle} - \vec{g}_{\langle 1\rangle} - \vec{g}_{\langle 10\rangle}$, where $\vec{g}_{\langle 10\rangle} = \vec{G}_{\{10\}} - \vec{g}_{\langle\rangle}$. Note that our model of epistasis, although similar to Hansen and Wagner (2001), in that we use a reference haplotype, is different in that we do not assume their multilinear form of genetic effects that involves a rescaling lower order effects. Griswold and Eisner (2012) showed how to transform genetic effects from one reference haplotype to another; as part of this analysis, they showed that the relative scaling of genetic effects is preserved after a transformation from one reference to another reference. In a genealogical context it is convenient and informative to use the most recent common ancestor (MRCA) of a sample as the reference haplotype. Using the MRCA as the reference captures the chronological accumulation of genetic effects through time.

2.2. Genetic covariance

The vectors of genotypic values of individuals determines the genetic covariance of a population or sample. A genetic covariance matrix is an $K \times K$ matrix, where K is the number of characters