Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 243 (2006) 76–85
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Additive genetic variation and the distribution of QTN effects among sites

Xin-Sheng Hu[a,*], Bailian Li[b]

[a]Department of Renewable Resources, 751 General Services Building, University of Alberta, Edmonton, AB, Canada T6G 2H1
[b]Department of Forestry, North Carolina State University, P.O. Box 8002, Raleigh, NC 27695, USA

Received 20 February 2006; received in revised form 5 May 2006; accepted 7 June 2006
Available online 15 June 2006

Abstract

Early quantitative genetic theories emphasize the distribution of gene effects at individual loci or the distribution of mutant effects in maintaining quantitative genetic variation, but infrequently stress the distribution of gene effects among loci. In this study, we analyse the effects of the distribution of quantitative trait nucleotides (QTN) effects among sites under artificial and stabilizing selection. Wright's formula is applied to describing the density distribution of allele frequencies of multiple diallelic QTN at the equilibrium of mutation-drift-selection. Our results demonstrate that the distribution of QTN effects among sites can affect additive genetic variation in terms of total additive variance, average gene diversity, per-class contribution of QTN effects and per-QTN contribution. When the distribution of QTN effects among sites is changed from L-shaped to bell-shaped or to be a flatter, both the total additive variance and the average gene diversity are changed. Per-class and per-QTN contributions exhibit different distribution patterns. The L-shaped distribution indicates the predominant role of the aggregative effects from the QTN of small finite effects. The bell-shaped or flatter distributions indicate the predominance of the QTN of intermediate and large effects. These predictions highlight the significance of the distribution of QTN effects among sites in interpreting the maintenance of quantitative genetic variation at the fine genome scale.

Keywords: Genetic variation; QTN; Artificial selection; Stabilizing selection; Distribution of gene effects; L-shaped distribution

1. Introduction

Additive genetic variation plays a critical role in determining the evolutionary potential of a quantitative trait in a population (Fisher, 1958), and its composition is mainly determined by three factors: (i) the number of genes; (ii) the effect of each gene; and (iii) the allele frequencies at each locus. The number of genes each with an infinitesimal effect is assumed to be infinite, but the number of genes with finite effects should be finite although not all of them can be detected for a given sample size (Mitchell-Olds, 1995). Two issues relevant to the first and the second factors are the distribution of gene effects at individual loci and the distribution of gene effects among loci. Early theoretical studies mainly concentrate on the effects of the distribution of gene effects at individual loci (Hill, 1982; Hill and Rasbash, 1986; Zhang and Hill, 2002) other than the effects of the distribution of gene effects among loci. The current development of genome-wide screening technique facilitates the detection of quantitative trait loci (QTL) at a much fine scale, such as quantitative trait nucleotides (QTN; Syvänen, 2001; Flint and Mott, 2001; Mackay, 2004). This development makes the study on the impacts of the distribution of gene effects among loci in a complex trait more significant than the study on the impacts of the distribution of gene effects at individual loci at the fine genome scale. This is because the maximum number of alleles at a single nucleotide site is not more than four. The diallelic single nucleotide polymorphisms (SNPs) are the most common case while the tri-and tetra-allelic SNPs are infrequent in natural populations (Brookes, 1999). Furthermore, mutants at individual nucleotide sites have small effects, which otherwise will be quickly removed or fixed from their resident population owing to natural selection. Thus,

*Corresponding author. Tel.: +780 492 0715; fax: +780 492 4323.
E-mail address: xin-sheng.hu@ualberta.ca (X.-S. Hu).

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.06.009

diallelic QTN sites are expected to be the most frequently occurring case in reality.

The biological basis for the existence of multiple genes (or QTN) with finite effects is the network of physiological and biochemical pathways underlying the formation of the quantitative trait in question (Keightley, 1989; Byrne et al., 1996; McMullen et al., 1998). The shape of the distribution of gene effects among loci is essentially related to such networks of metabolic pathways, and the relative effects of individual loci are associated with their function and locations on these networks. For example, wood specific gravity (the ratio of the weight of a unit volume of wood to that of water) contains a complex network. It is influenced by length, diameter and wall thickness of xylem cells, relative proportion of early and late wood, cellulose and lignin content, and extractive content. The shape of the distribution of five detected QTL effects exhibits a slightly flat pattern in loblolly pine (Groover et al., 1994; Grattapaglia et al., 1996). In a separate study on a complex trait, the timing of spring bud flush in *Populus*, which is correlated with tolerance to late spring frosts, the shape of the distribution of the observed QTL effects exhibits a leptokurtic pattern (Bradshaw and Stettler, 1995). It is expected that a variety of distribution patterns of gene effects among loci may occur among quantitative traits differing in the network of metabolic pathways. Thus, understanding of the linkage of these patterns to genetic variation is of significance for gaining insights into how quantitative genetic variation is maintained in a population.

Although there are extensive studies on the model of genes with equal effects (Wright, 1969; Falconer and Mackay, 1996), theoretical studies on the distribution of gene effects among loci are not much emphasized (Hill, 1982). According to the parent populations (P_1 and P_2) and their hybrids (F_1 and F_2) or backcross (B_1 and B_2), Wright (1968, pp. 381–391) proposed several distribution patterns of gene effects among loci, such as the geometric and arithmetic series. Using a chain of metabolic pathway, Bost et al. (1999) showed that L-shaped distribution of gene effects among loci can be generated. Barton and Keightley (2002) pointed out that an exponential distribution of gene effects among loci is expected when a population adapts by fixing sequence of additive mutations. Robertson (1967) argued that the distribution of gene effects among loci belongs to an exponential distribution where a few QTL have large effects and contribute most of variation but many more QTL have small effects. Mackay (2001) reviewed that the exponential type of distribution describes well many observations, although a rigorous theoretical analysis remains to be explored and the statistical test is needed in different quantitative traits.

The purpose of this study is to examine the relationships between genetic variation and the distribution of gene effects among loci. Throughout this study, we focus on the distribution of QTN effects among sites (not neutral QTN) and each QTN site has two alleles. The individual QTN effects are small, but their sum may be large (Mackay, 2004). Artificial (mass) and stabilizing selection on an additive polygenic trait are considered although other kinds of selection may be involved (Barton and Turelli, 1989; Zhang and Hill, 2002). In our analyses, Wright's formula is used to describe the multidimensional probability distribution of gene frequencies for the multiple diallelic QTN under the balance of selection-mutation-drift (Wright, 1969, p. 396). Simulation is conducted to evaluate additive genetic variation under different distributions of QTN effects among sites (from L-shaped to bell-shaped distributions).

2. Artificial selection

Consider K diallelic QTN sites with alleles A_{i1} and A_{i2} at the ith site ($i = 1, 2, \ldots, K$) in an additive polygenic trait in a random-mating diploid population with effective size N. The relative genotypic values at the ith site are expressed as $2a_i$ for $A_{i1}A_{i1}$, a_i for $A_{i1}A_{i2}$, and 0 for $A_{i2}A_{i2}$. Denote the allele frequencies by p_i for A_{i1} and $1-p_i$ for A_{i2} at the ith site. Note that in the case of tri- or tetra-alleles at a single QTN site, the most common allele is considered as A_{i1} while others are treated as A_{i2}. Fitness is assumed as $1+s_i$ for A_{i1} and 1 for A_{i2}. Without loss of generality, we assume that allele A_{i1} always has a selective advantage over A_{i2}, i.e. $s_i \geqslant 0$, in a given environment. According to the additive viability model, genotypic fitness at the ith site is set as $1+2s_i$ for $A_{i1}A_{i1}$, $1+s_i$ for $A_{i1}A_{i2}$, and 1 for $A_{i2}A_{i2}$. From previous studies (Robertson, 1960; Kimura and Crow, 1978; Falconer and Mackay, 1996), the relationship between s_i and a_i can be expressed as

$$s_i = Ia_i/\sigma, \tag{1}$$

where I is the selection intensity (the standardized selection differential) and σ is the phenotypic standard deviation. The linear relationship between s_i and a_i depends on the value of Ia_i/σ that is assumed to have the order similar to weak selection and is applied to link individual QTN effects to genotypic fitness (Robertson, 1960; Hill, 1982).

Assume that allele A_{i2} mutates to A_{i1} with a probability v_i and the reverse mutation with a probability u_i. The mutation effects on increasing genetic diversity are important even in process of artificial selection (Hill, 1982). According to Wright (1969, p. 396) the multidimensional joint probability distribution for allele frequencies at K sites at steady state is given by

$$\phi(p_1, p_2, \ldots, p_K) = C\bar{W}^{2N}\prod_{i=1}^{K} p_i^{\Theta_{1i}-1}(1-p_i)^{\Theta_{2i}-1}, \tag{2}$$

where $\Theta_{1i} = 4Nv_i$ and $\Theta_{2i} = 4Nu_i$. If the fitness is measured in Malthusian parameters (Fisher, 1958), $\bar{W}$ can be replaced with $\exp(\bar{r})$, where $\bar{r}$ is the mean fitness of the population in Malthusian parameter. According to Kimura (1964), Eq. (2) is the steady-state solution to