Current Topics in Genetics
Vol. 3, 2008

Original Communication

Functional and statistical genetic effects with multiple alleles

Rong-Cai Yang[1,2,*] and José M. Álvarez-Castro[3]

[1]Agriculture Research Division, Alberta Agriculture and Rural Development, Edmonton, Alberta, T6H 5T6 Canada, [2]Department of Agricultural, Food and Nutritional Science, 410 Agriculture/Forestry Centre, University of Alberta, Edmonton, Alberta, T6G 2P5 Canada, [3]Department of Genetics, University of Santiago de Compostela, 27002 Lugo, Spain

ABSTRACT

Mapping quantitative trait loci (QTL) is often based on a functional or statistical model of gene action involving two alleles per locus. Such model is adequate for mapping populations derived from a cross between two inbred lines as in many plants and some laboratory animals and for di-allelic molecular markers (e.g., single nucleotide polymorphisms, SNPs). However, many mapping populations (e.g., those derived from crosses between more than two inbred lines) and multi-allelic molecular markers (e.g., microsatellites) require a model to describe functional genetic effects for multiple alleles. This paper has two objectives. The primary objective is to develop a set of new models of functional genetic effects with multiple alleles. The secondary objective is to establish the relationship between these functional genetic effects with well-known statistical genetic effects. Our multi-allelic models reveal three new features that do not exist in the di-allelic models. First, with r (>2) alleles, there are $r(r\text{-}1)/2$ functional additive effects and $r(r\text{-}1)/2$ functional dominance effects, but only $(r\text{-}1)$ functional additive effects need to be specified and the remaining $(r\text{-}1)(r\text{-}2)/2$ additive effects can be derived. Second, the presence of functional dominance effect for one pair of alleles is sufficient to cause the presence of statistical dominance deviations for all the genotypes. Third, the equality of gene frequencies is no longer a sufficient condition for any direct relationship between physiological and statistical genetic effects in the multi-allelic case. Thus, our new multi-allelic models will have a wider range of applications to QTL mapping and quantitative genetic studies.

KEYWORDS: functional genetic effects, multiple alleles, additive effects, dominance deviations, change-of-reference transformation

ABBREVIATIONS

GF, general function of genotypic values; HWD, Hardy-Weinberg disequilibrium; LD, linkage disequilibrium; NOIA, natural and orthogonal interactions; QTL, quantitative trait loci; UWR, Unweighted Regression.

INTRODUCTION

Genome-wide scans for quantitative trait loci (QTL) are now a routine strategy for identifying effects of individual QTL and interactions between QTL. With the availability of ever increasing marker density across the entire genome, most of the QTL effects will be picked up by the tightly linked markers so that the marker effects can serve as the reliable surrogates of QTL effects [1, 2]. Consequently, while location of a particular QTL and estimation of its effect remain to be a major effort of QTL mapping, a growing focus is now on modeling actions and interactions of detected QTL effects [e.g., 3, 4, 5, 6, 7, 8, 9].

*Corresponding author
rong-cai.yang@ualberta.ca

QTL effects are often defined in two ways using the two classic quantitative genetic models of Fisher [10], one being the model for defining genotypic values (often called functional model) and the other being the model for defining the average effect of a single gene (often called statistical model). For a locus (say locus *A*) with two alleles, A_1 and A_2, the values of three possible genotypes, A_1A_1, A_1A_2 and A_2A_2, are $m + a$, $m + d$ and $m - a$, respectively, where m is the homozygote mean located exactly at the mid-way between the values of the two homozygotes, a is the functional additive effect measured as half the difference between the values of two homozygotes and d is the functional dominance effect measured as the deviation of heterozygote from the homozygote mean. This model is subsequently known as the F_∞ model [11, 12] and can be transformed into other equivalent models including F_2 model [8, 13], Unweighted Regression (*UWR*) model [6] and a model of arbitrary reference [3]. The important feature of this model is that the ratio of d/a allows for discussing the level of dominance in the sense of elementary Mendelian genetics or simply physiological (functional) dominance [3, 6]. For example, the special cases of $d/a = 0$, $d/a = \pm 1$, $d/a < -1$ and $d/a > +1$ would represent no dominance, complete dominance, underdominance and overdominance, respectively. This functional model has been extended to include physiological or functional epistasis at two or more loci [e.g., 3, 5, 6, 8, 9]. However, to the best of our knowledge, all single-locus and multilocus functional models are limited to the case of two alleles per locus or to highly constrained genetic architectures.

The functional genetic effects (a and d) defined above are simply comparisons among genotypic values without reference to any population and thus they stay invariant from one population to another. However, a diploid individual passes on its genes, not its genotype, to offspring, and measures of such genic effects are statistical because they depend on both frequencies and values of genotypes in a given population. It has been well known since Fisher [10] that the average effect of a gene is a partial regression coefficient from the linear regression of the genotypic values on the number of that gene weighted by genotypic frequencies in a population. The average effects of two genes in an individual are added to the population mean to give individual's 'predicted' genotypic value (or breeding value) and deviations from the predicted value are due to the interaction of the two genes or to the presence of dominance [11]. This model has been extended to include statistical epistatic effects at two or more loci with two or an arbitrary number of alleles per locus [e.g., 7, 14, 15].

Recognizing that the same genotypic values can be expressed in terms of either functional or statistical genetic effects, Yang [8] suggested a connection between functional and statistical genetic effects as defined under F_2 model and Cockerham's model, in the same manner as the transformation of Van Der Veen [13] that allows for the translation of functional genetic effects between F_2 model and F_∞ model. Alvarez-Castro and Carlborg [3] generalized this suggestion for the two-allele framework in the natural and orthogonal interactions (NOIA) model that unifies the functional and the statistical formulations with the theory necessary to translate between functional and statistical genetic effects for multiple unlinked loci with arbitrary epistasis. This generalization enables, for instance, to obtain multilinear estimates of genetic effects from QTL data by embedding Hansen and Wagner's [16] multilinear model into NOIA [17]. However, it remains to be seen how functional and statistical effects are related to each other for the multi-allelic case. With increasing number of alleles, it is expected that the functional and statistical genetic effects are related in a more complicated manner. For example, C. C. Li in Kempthorne [18] questioned why in a three-allele case statistical dominance effects are present when apparently there is lack of functional dominance for some genotype pairs. The question remains unanswered since then.

In this paper, we will first extend the currently used two-allele models of functional genetic effects to include the cases of more than two alleles. We will then develop the relationships between functional and statistical effects with the presence of multiple alleles. Numerical analysis is carried out to illustrate the applications of the theory. There is an obvious need for modeling multi-allele functional genetic effects for mapping populations derived from multi-way crosses between more than two inbred lines or for the use of multi-allelic molecular markers for QTL mapping.

Functional effects of multiple alleles

Review of models for two alleles

A vector of genotypic values at a given locus, say locus A ($\mathbf{G}_A$) can be transformed to obtain a vector of genetic effects ($\mathbf{E}_{X.A}$) through a genetic-effect design matrix ($\mathbf{S}_{X.A}$), $\mathbf{G}_A = \mathbf{S}_{X.A}\mathbf{E}_{X.A}$, where subscript X represents a transformation operator. The dimensions of $\mathbf{G}_A$ and $\mathbf{E}_{X.A}$ are the same (say $q \times 1$) and $\mathbf{S}_{X.A}$ is a $q \times q$ square matrix. The q value represents the number of possible genotypes which equal to $r(r+1)/2$ if there are r ($r \geq 1$) alleles at locus A. The special case of $r = 1$ and $q = 1$ (all genotypes are one type of homozygotes) is trivial and thus $\mathbf{G}_A = \mathbf{E}_{X.A}$. The most discussed is the case of two alleles ($r = 2$ and $q = 3$) [e.g., 3, 5, 8, 9]. In this case, $\mathbf{G}_A = [G_{11}\ G_{12}\ G_{22}]'$ with the prime (′) denoting matrix or vector transposition and G_{ij} being the value of the genotype carrying alleles A_i and A_j. The genetic-effect vector is $\mathbf{E}_{X.A} = [R_{X.A}\ a_A\ d_A]'$, where $R_{X.A}$ is the reference point for a given transformation operator, a_A is the additive effect measured and d_A is the dominance effect. Since the functional additive and dominance effects are independent of the transformation operator, the reference point can be an arbitrary value including a mean of all genotypic values, a single arbitrarily chosen genotypic value or any numerical number. In particular, if $R_{X.A} = 1$ then $\mathbf{E}_{1.A} = [1\ a_A\ d_A]'$. The **S**-matrices derived from the well-known transformation operators, F_∞-metric, F_2-metric, *UWR* and NOIA from an individual reference genotype, G_{11}, are, respectively:

$$\mathbf{S}_{F_\infty \cdot A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix},\ \mathbf{S}_{F_2 \cdot A} = \begin{bmatrix} 1 & 1 & -\tfrac{1}{2} \\ 1 & 0 & \tfrac{1}{2} \\ 1 & -1 & -\tfrac{1}{2} \end{bmatrix},\ \mathbf{S}_{UW \cdot A} = \begin{bmatrix} 1 & 1 & -\tfrac{1}{3} \\ 1 & 0 & \tfrac{2}{3} \\ 1 & -1 & -\tfrac{1}{3} \end{bmatrix} \text{ and } \mathbf{S}_{G_{11} \cdot A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$$

and their respective inverses are:

$$\mathbf{S}^{-1}_{F_\infty \cdot A} = \begin{bmatrix} \tfrac{1}{2} & 0 & \tfrac{1}{2} \\ \tfrac{1}{2} & 0 & -\tfrac{1}{2} \\ -\tfrac{1}{2} & 1 & -\tfrac{1}{2} \end{bmatrix},\ \mathbf{S}^{-1}_{F_2 \cdot A} = \begin{bmatrix} \tfrac{1}{4} & \tfrac{1}{2} & \tfrac{1}{4} \\ \tfrac{1}{2} & 0 & -\tfrac{1}{2} \\ -\tfrac{1}{2} & 1 & -\tfrac{1}{2} \end{bmatrix},\ \mathbf{S}^{-1}_{UW \cdot A} = \begin{bmatrix} \tfrac{1}{3} & \tfrac{1}{3} & \tfrac{1}{3} \\ \tfrac{1}{2} & 0 & -\tfrac{1}{2} \\ -\tfrac{1}{2} & 1 & -\tfrac{1}{2} \end{bmatrix} \text{ and } \mathbf{S}^{-1}_{G_{11} \cdot A} = \begin{bmatrix} 1 & 0 & 0 \\ \tfrac{1}{2} & 0 & -\tfrac{1}{2} \\ -\tfrac{1}{2} & 1 & -\tfrac{1}{2} \end{bmatrix}$$

It is evident from the first row of the inverse matrices that these transformation operators differ only in the choice of the reference point, with the reference point being either the mean of the population or an arbitrary genotypic value.

The genetic-effect design matrix for two unlinked loci (say A and B), $\mathbf{S}_{X.AB}$, is a direct product (Kronecker product) of two single-locus design matrices, $\mathbf{S}_{X.A}$ for locus A, and $\mathbf{S}_{X.B}$ for locus B, i.e., $\mathbf{S}_{X.AB} = \mathbf{S}_{X.B} \otimes \mathbf{S}_{X.A}$ [3]. The two-locus genetic-effect vector $\mathbf{E}_{X.AB}$ can be obtained by replacing appropriate terms in the direct product,

$$\mathbf{E}_{1.AB} = \mathbf{E}_{1.B} \otimes \mathbf{E}_{1.A} = [1\ a_A\ d_A\ a_B\ d_B\ a_A a_B\ a_A d_B\ d_A a_B\ d_A d_B]'.$$

Specifically, we replace 1 by $R_{X.AB}$, the product terms by single terms (e.g., the product of two additive, effects, $a_A a_B$, by a single additive × additive effect, aa_{AB}). In this way, we obtain the two-locus genetic-effect vector

$$\mathbf{E}_{X.AB} = [R_{X.AB}\ a_A\ d_A\ a_B\ d_B\ aa_{AB}\ ad_{AB}\ da_{AB}\ dd_{AB}]'$$

Once again, subscript X is a transformation operator (e.g., $X = G_{11}$, F_∞, F_2, or *UWR*). The vector of two-locus genotypic values $\mathbf{G}_{AB}$ can be similarly constructed. The two-locus genetic effects are obtained by solving the system equations,

$$\mathbf{E}_{X.AB} = \mathbf{S}^{-1}_{X.AB}\mathbf{G}_{AB} = (\mathbf{S}^{-1}_{X.B} \otimes \mathbf{S}^{-1}_{X.A})\mathbf{G}_{AB}$$

Similarly, the genetic-effect design matrix for three or more unlinked loci (*A, B, C, ..., L*), $\mathbf{S}_{X.ABC...L}$, is obtained from the reverse-order direct products of multiple single-locus design matrices, $\mathbf{S}_{X.A}$ for locus *A*, and $\mathbf{S}_{X.B}$ for locus *B* and $\mathbf{S}_{X.C}$ for locus *C*, ..., $\mathbf{S}_{X.L}$ for locus *L*, i.e., $\mathbf{S}_{X.ABC\cdots L} = \mathbf{S}_{X.L} \otimes \cdots \otimes \mathbf{S}_{X.C} \otimes \mathbf{S}_{X.B} \otimes \mathbf{S}_{X.A}$.

The construction of $\mathbf{E}_{X.ABC...L}$ and $\mathbf{G}_{ABC...L}$ is similar to that of $\mathbf{E}_{X.AB}$ and $\mathbf{G}_{AB}$ by the extended use of the reverse-order direct products for multiple vectors and appropriate subsequent substitutions. The multilocus genetic effects are obtained by solving the system equations,

$$\mathbf{E}_{X.ABC\cdots L} = \mathbf{S}^{-1}_{X.ABC\cdots L}\mathbf{G}_{ABC\cdots L} = (\mathbf{S}^{-1}_{X.L} \otimes \cdots \otimes \mathbf{S}^{-1}_{X.C} \otimes \mathbf{S}^{-1}_{X.B} \otimes \mathbf{S}^{-1}_{X.A})\mathbf{G}_{ABC\cdots L}$$

Clearly obtaining the single-locus genetic-effect matrices or their inverses is fundamental for the above multilocus generalization. Thus, the subsequent development of models for multiple alleles focuses on one locus only (locus *A*).

Extension to an arbitrary number of alleles

We now show how to obtain the expressions of a functional formulation of genetic effects with more than two alleles. We first extend the two-allele NOIA model [3] to the multi-allelic case. We then develop a direct approach based on meaningful comparisons that need to be made given available homozygotes and heterozygotes for multiple alleles.

From the expressions of the functional formulation of NOIA for two alleles as described above, we proceed by recurrence showing how to extend an (r – 1)-allele case to an r-allele case, for $r > 2$. By applying this procedure recursively, functional genetic effect models for any number of alleles can be reached. The extensions of the $\mathbf{G}_A$ and $\mathbf{E}_{X.A}$ vectors are simply achieved by appending more terms arising from the presence of multiple alleles to the end of the respective vectors, with the subscribed numerals for the alleles in an ascending order. For example, for $r = 2$, there are 3 possible genotypes: A_1A_1, A_1A_2 and A_2A_2, but for $r = 3$, there are three additional genotypes: A_1A_3, A_2A_3 and A_3A_3. Thus, the vector $\mathbf{G}_A$ is expanded from $\mathbf{G}_A = [G_{11}\ G_{12}\ G_{22}]'$ for $r = 2$ to $\mathbf{G}_A = [G_{11}\ G_{12}\ G_{22}\ G_{13}\ G_{23}\ G_{33}]'$ for $r = 3$. Correspondingly, $\mathbf{E}_{X.A}$ is expanded from $[R_{X.A}\ a_{12}\ d_{12}]'$ to $[R_{X.A}\ a_{12}\ d_{12}\ a_{13}\ d_{13}\ d_{23}]'$, where additive and dominance effects are simply comparisons between pairs of homozygotes and deviations of a given heterozygote from the average of the two corresponding homozygotes, $a_{ij} = (G_{ii} - G_{jj})/2$ and $d_{ij} = G_{ij} - (G_{ii} + G_{jj})/2$. Notice that a_{23} is missing from the $\mathbf{E}_{X.A}$ vector and can be recovered from the relation, $a_{23} = a_{12} - a_{13}$. This example illustrates a distinct feature with the presence of more than two alleles that the additive effects derived from individual comparisons between pairs of homozygotes are linearly dependent of each other. In general, for $r > 2$, there are r homozygotes and (r-1) basic additive effects are defined as the differences between the values of the reference homozygote (G_{11}) and the remaining (r-1) homozygotes. The remaining (r-1)(r-2)/2 additive effects can be recovered from the basic additive effects as $a_{ji} = a_{1j} - a_{1i}$, $i, j = 2,\ldots,r$.

The **S**-matrix can be expanded similarly. In general, the new **S**-matrix shall have i rows below and i columns to the right of the previous matrix, $\mathbf{S}_{i\text{-}1}$, so that it can be expressed as

$$\mathbf{S}_i = \begin{bmatrix} \mathbf{S}_{i-1} & \mathbf{M}_{i1} \\ \mathbf{M}_{i2} & \mathbf{M}_{i3} \end{bmatrix} \quad (1)$$

The scalars in the new blocks are the ones for keeping on describing the genetic system as a set of allele substitutions from the reference of G_{11}. This procedure can be automated using the following algorithm:

- Block $\mathbf{M}_{i1}$, to the right of the previous matrix: All zeros.
- Block $\mathbf{M}_{i2}$, below the previous matrix: Ones in the first column (these reflect the reference point) and in positions $1+\sum_{k=1}^{i}(k-1)$ of every row but the first and the last ones (these reflect additive effects of previous alleles in new genotypes due to the new allele) and zeros in the rest of the positions.
- Block $\mathbf{M}_{i3}$, square $i \times i$ remaining block: First column of ones, except from the last one, which shall be a two (these reflect the additive effects of the new allele).

Zeros in the last row (except from the first position, which is already filled by a two). And the identity matrix in the remaining $(i-1) \times (i-1)$ block (these ones reflect the new dominant effects).

Following these steps, a functional formulation of genetic effects from the reference of $R_1 = G_{11.A}$ can be obtained for a one-locus genetic system with any number of alleles. For example, the three-allele genetic system with the reference of G_{11} can be obtained as $\mathbf{G}_A = \mathbf{S}_{G11.A}\mathbf{E}_{G11.A}$ expanding to:

$$\begin{bmatrix} G_{11} \\ G_{12} \\ G_{22} \\ G_{13} \\ G_{23} \\ G_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} R_{G_{11}.A} \\ a_{12} \\ d_{12} \\ a_{13} \\ d_{13} \\ d_{23} \end{bmatrix} \quad (2)$$

Evidently the choice of G_{11} as a reference point is somewhat arbitrary. In fact, like in the di-allelic case, any other genotypic value or any linear combination of genotypic values can be used as a reference point. The switch from one reference point to the other can also be easily achieved by a change-of-reference operation that works in the same way as for the two-allele case [3]. Suppose that the reference is changed from G_{11} to a general function (GF) of all genotypic values,

$$GF = \sum_{u,v \geq u}^{r} P_{uv} G_{uv} \quad (3)$$

where P_{uv} is the frequency of the genotype carrying alleles A_u and A_v with $\sum_{u,v \geq u}^{r} P_{uv} = 1$. Following [3], the **S**-matrix for GF as the reference is then given by,

$$\mathbf{S}_{GF.A} = \mathbf{S}_{G11.A} - \mathbf{P}_{GF.A} \cdot \mathbf{S}_{G11.A} \cdot \mathbf{I}^{*} \quad (4)$$

where $\mathbf{P}_{GF.A}$ is a square matrix in which each column is filled with one of the coefficients of the linear combination of genotypic values (GF) as given in (3) and $\mathbf{I}^*$ is the identity matrix except the first scalar of the matrix being zero. For example, with three alleles and six possible genotypes, $\mathbf{P}_{GF.A}$ is given by,

$$\mathbf{P}_{GF.A} = [\mathbf{1}P_{11} \quad \mathbf{1}P_{12} \quad \mathbf{1}P_{22} \quad \mathbf{1}P_{13} \quad \mathbf{1}P_{23} \quad \mathbf{1}P_{33}],$$

where $\mathbf{1}$ is a 6×1 vector of ones. Plugging $\mathbf{P}_{GF.A}$, $\mathbf{S}_{G11.A}$ and $\mathbf{I}^*$ into (4), we obtain the general expression for the three-allele functional genetic-effect design matrix,

$$\mathbf{S}_{GF.A} = \begin{bmatrix} 1 & -2p_2 & -P_{12} & -2p_3 & -P_{13} & -P_{23} \\ 1 & 1-2p_2 & 1-P_{12} & -2p_3 & -P_{13} & -P_{23} \\ 1 & 2-2p_2 & -P_{12} & -2p_3 & -P_{13} & -P_{23} \\ 1 & -2p_2 & -P_{12} & 1-2p_3 & 1-P_{13} & -P_{23} \\ 1 & 1-2p_2 & -P_{12} & 1-2p_3 & -P_{13} & 1-P_{23} \\ 1 & -2p_2 & -P_{12} & 2-2p_3 & -P_{13} & -P_{23} \end{bmatrix} \quad (5)$$

where $p_2 = \frac{1}{2}P_{12} + P_{22} + \frac{1}{2}P_{23}$ and $p_3 = \frac{1}{2}P_{13} + \frac{1}{2}P_{23} + P_{33}$ are frequencies of alleles A_2 and A_3 with $p_1 = 1 - p_2 - p_3 = P_{11} + \frac{1}{2}P_{12} + \frac{1}{2}P_{13}$ being the frequency of allele A_1, and P_{12}, P_{13} and P_{23} are frequencies of heterozygotes A_1A_2, A_1A_3 and A_2A_3, respectively. In Appendix, we provide more details on the use of the general result in (5) for some special three-allele examples.

Direct approach

The above approach to calculating genetic effects involves two steps, first constructing an **S**-matrix and then obtaining its inverse so that $\mathbf{E}_{GF.A} = \mathbf{S}_{GF.A}^{-1}\mathbf{G}_A$. Here we develop a one-step approach that allows for a direct calculation of functional genetic effects without the need to find the **S**-matrix. Once again, of $r(r+1)/2$ possible genotypes with r alleles at locus A, there are r homozygotes (A_1A_1, A_2A_2, …, and A_rA_r), and $r(r-1)/2$ heterozygotes (A_1A_2, A_1A_3, …, and $A_{r-1}A_r$). Numerous comparisons among these genotypes are possible. In particular, two sets of comparisons can be meaningfully made corresponding to the following two sets of hypotheses: (i) all r homozygotes are functionally equivalent (i.e., all homozygotes have the same genotypic values, $G_{11} = G_{22} = \ldots = G_{rr}$) and (ii) a heterozygote is functionally equivalent to the average of the two corresponding homozygotes (i.e., $G_{uv} = (G_{uu} + G_{vv})/2$). Testing for hypotheses (i) embodies $(r-1)$ comparisons between a base homozygote value (say G_{11}) and each of the remaining $(r-1)$ homozygote values (i.e., $G_{11} = G_{22}$; $G_{11} = G_{33}$; …; $G_{11} = G_{rr}$) whereas testing for hypotheses (ii) requires $r(r-1)/2$ comparisons, each being between a heterozygote and the average of the two corresponding homozygotes.

Let $R_{G11.A} = G_{11}$, $a_{1v} = (G_{vv} - G_{11})/2$ for $v = 2, 3, \ldots, r$, and $d_{uv} = G_{uv} - (G_{uu} + G_{vv})/2$ for $u < v$, $u = 1, 2, \ldots, r\text{-}1$. The a's are obviously not independent of each other because unspecified comparisons among homozygotes would be functions of the comparisons specified under hypotheses (i). For example, the comparison between homozygotes, A_uA_u and A_vA_v, can be obtained from the relation $a_{uv} = a_{1u} - a_{1v}$. Collecting all $r(r+1)/2$ possible equations from $R_{G11.A}$, $(r\text{-}1)$ a's and $r(r\text{-}1)/2$ d's directly leads to $\mathbf{E}_{G11.A} = \mathbf{S}^{-1}_{G11.A}\mathbf{G}_A$ without *a priori* specifying the genetic-effect design matrix $\mathbf{S}_{G11.A}$. For the tri-allelic case, we have

$$\begin{bmatrix} R_{G11\cdot A} \\ a_{12} \\ d_{12} \\ a_{13} \\ d_{13} \\ d_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -½ & 0 & ½ & 0 & 0 & 0 \\ -½ & 1 & -½ & 0 & 0 & 0 \\ -½ & 0 & 0 & 0 & 0 & ½ \\ -½ & 0 & 0 & 1 & 0 & -½ \\ 0 & 0 & -½ & 0 & 1 & -½ \end{bmatrix} \begin{bmatrix} G_{11} \\ G_{12} \\ G_{22} \\ G_{13} \\ G_{23} \\ G_{33} \end{bmatrix} \tag{6}$$

The numerical matrix in (6) obviously is the inverse of the **S**-matrix, $\mathbf{S}^{-1}_{G11.A}$ in (2). As mentioned earlier, the choice of G_{11} as a reference point is somewhat arbitrary and any of the $r(r+1)/2$ genotype values can serve as the reference point. With the change in the reference point, functional additive and dominance effects would need to be redefined as comparisons of the new referenced genotype with all other genotypes. The *GF* function in (3) can also serve as a reference point. Some *GF*-based examples are those equivalent to F_∞-metric, F_2-metric, *UWR* models in the case of two alleles. Thus, for the case of three alleles, we can also have one of the following averages,

$$R_{GF.A} = \begin{cases} ⅓(G_{11} + G_{22} + G_{33}), & \text{if } GF = F_\infty \\ ⅛(G_{11} + 2G_{12} + G_{22} + 2G_{13} + 2G_{23} + G_{33}), & \text{if } GF = F_2 \\ 1/6(G_{11} + G_{12} + G_{22} + G_{13} + G_{23} + G_{33}), & \text{if } GF = UW \end{cases} \tag{7}$$

as the reference point. If *GF* is equal to the population mean (μ) and G_{11} is used for the referenced genotype for all genotypic comparisons, then the genetic effects can be directly written as

$$\mathbf{E}_{\mu.A} = \mathbf{S}^{-1}_{\mu.A}\mathbf{G}_A$$

where $\mathbf{E}_{\mu.A} = [R_{\mu.A}\ a_{12}\ d_{12}\ a_{13}\ d_{13}\ d_{23}]'$, and

$$\mathbf{S}^{-1}_{\mu.A} = \begin{bmatrix} P_{11} & P_{12} & P_{22} & P_{13} & P_{23} & P_{33} \\ -½ & 0 & ½ & 0 & 0 & 0 \\ -½ & 1 & -½ & 0 & 0 & 0 \\ -½ & 0 & 0 & 0 & 0 & ½ \\ -½ & 0 & 0 & 1 & 0 & -½ \\ 0 & 0 & -½ & 0 & 1 & -½ \end{bmatrix} \tag{9}$$

is the inverse of $\mathbf{S}_{\mu.A}$ which is the same as $\mathbf{S}_{GF.A}$ given in (5). If the deviations from μ rather than genotypic values per se are used, then the reference point is zero,

$$\mathbf{E}_{0.A} = \mathbf{S}^{-1}_{\mu.A}(\mathbf{G}_A - \mathbf{1}\mu) \tag{10}$$

with $\mathbf{E}_{0.A} = [0\ a_{12}\ d_{12}\ a_{13}\ d_{13}\ d_{23}]'$. These results can be easily generalized for more than three alleles and an algorithm such as the one described above for obtaining the **S**-matrices can be designed accordingly.

Relations to statistical effects

Statistical effects

The statistical additive and dominance effects for multiple alleles are defined for a Hardy-Weinberg equilibrium (HWE) population. Thus, the value of genotype A_uA_v in the population, G_{uv}, can be expressed as,

$$G_{uv} = \mu + \alpha_u + \alpha_v + \delta_{uv} \tag{11}$$

where $\mu = \Sigma^r_{u=1}\Sigma^r_{v\geq u}P_{uv}G_{uv}$ is the population mean, $\alpha_u = \Sigma^r_{v=1}p_vG_{uv} - \mu$ is the average (additive) effect of the uth allele with $\Sigma^r_{u=1}p_u\alpha_u = 0$ and δ_{uv} is the dominance deviation, i.e., $\delta_{uv} = G_{uv} - \mu - \alpha_u - \alpha_v$, with $\Sigma^r_{u=1}P_{uv}\delta_{uv} = \Sigma^r_{v=1}P_{uv}\delta_{uv} = \Sigma^r_{u=1}\Sigma^r_{v=1}P_{uv}\delta_{uv} = 0$. In the matrix form, (11) can be written as,

$$\mathbf{G} = \mathbf{1}\mu + \mathrm{N}\boldsymbol{\alpha} + \boldsymbol{\delta}, \tag{12}$$

where **1** is a $(t \times 1)$ vector of ones,

$$\mathbf{G}_{t\times 1}=\begin{bmatrix} G_{11}\\ G_{12}\\ G_{22}\\ \vdots\\ G_{(r-1)r}\\ G_{rr}\end{bmatrix},\ \mathbf{N}_{t\times r}=\begin{bmatrix} 2 & 0 & \cdots & 0\\ 1 & 1 & \cdots & 0\\ 0 & 2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\\ 0 & 0 & \cdots & 2\end{bmatrix},\ \boldsymbol{\alpha}_{r\times 1}=\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_r\end{bmatrix},\ \text{and}\ \boldsymbol{\delta}_{t\times 1}=\begin{bmatrix}\delta_{11}\\ \delta_{12}\\ \delta_{22}\\ \vdots\\ \delta_{(r-1)r}\\ \delta_{rr}\end{bmatrix} \tag{13}$$

The weighted least squares solution of $\boldsymbol{\alpha}$ in the linear model (12) is well known [e.g., 15, 19],

$$\boldsymbol{\alpha}=(\mathbf{N}'\mathbf{P}_{\mathrm{HWE}}\mathbf{N})^{-1}\mathbf{N}'\mathbf{P}_{\mathrm{HWE}}(\mathbf{G}_A-\mathbf{1}\mu) \tag{14}$$

where $\mathbf{P}_{\mathrm{HWE}}$ = diag{ $p_1^2, 2p_1p_2, p_2^2,\cdots, p_r^2$ } is the diagonal matrix whose diagonal elements are the HWE genotypic frequencies. The dominance deviations are,

$$\boldsymbol{\delta}=\mathbf{G}_A-\mathbf{1}\mu-\mathbf{N}\boldsymbol{\alpha} \tag{15}$$

Transformation matrices

We will now proceed to find a pair of transformation matrices that allow for converting functional genetic effects into statistical genetic effects. Suppose that statistical additive effects may be obtained by finding a matrix, $\mathbf{T}_\alpha$, such that $\boldsymbol{\alpha}=\mathbf{T}_\alpha\mathbf{E}_{0.A}$. Since $\mathbf{E}_{0.A}=\mathbf{S}_{\mu.A}^{-1}(\mathbf{G}_A-\mathbf{1}\mu)$ (cf. equation (10)), it follows that $\mathbf{T}_\alpha$, can be expressed as,

$$\mathbf{T}_\alpha=(\mathbf{N}'\mathbf{P}_{\mathrm{HWE}}\mathbf{N})^{-1}\mathbf{N}'\mathbf{P}_{\mathrm{HWE}}\mathbf{S}_{\mu.A} \tag{16}$$

Similarly, we can obtain statistical dominance effects by finding another matrix, $\mathbf{T}_\delta$,

$$\mathbf{T}_\delta=[\mathbf{I}-\mathbf{N}(\mathbf{N}'\mathbf{P}_{\mathrm{HWE}}\mathbf{N})^{-1}\mathbf{N}'\mathbf{P}_{\mathrm{HWE}}]\mathbf{S}_{\mu.A} \tag{17}$$

such that $\boldsymbol{\delta}=\mathbf{T}_\delta\mathbf{E}_{0.A}$. For the case of three alleles, we have,

$$\mathbf{T}_\alpha=\begin{bmatrix} ½ & -p_2 & (1-2p_1)p_2 & -p_3 & (1-2p_1)p_3 & -2p_2p_3\\ ½ & 1-p_2 & p_1(1-2p_2) & -p_3 & -2p_1p_3 & (1-2p_2)p_3\\ ½ & -p_2 & -2p_1p_2 & 1-p_3 & p_1(1-2p_3) & p_2(1-2p_3)\end{bmatrix} \tag{18}$$

and

$$\mathbf{T}_\delta=\begin{bmatrix} 0 & 0 & -2(1-p_1)p_2 & 0 & -2(1-p_1)p_3 & 2p_2p_3\\ 0 & 0 & 2p_1p_2+p_3 & 0 & 2p_1p_3-p_3 & 2p_2p_3-p_3\\ 0 & 0 & -2p_1(1-p_2) & 0 & 2p_1p_3 & -2(1-p_2)p_3\\ 0 & 0 & 2p_1p_2-p_2 & 0 & 2p_1p_3+p_2 & 2p_2p_3-p_2\\ 0 & 0 & 2p_1p_2-p_1 & 0 & 2p_1p_3-p_1 & 2p_2p_3+p_1\\ 0 & 0 & 2p_1p_2 & 0 & -2p_1(1-p_3) & -2p_2(1-p_3)\end{bmatrix} \tag{19}$$

Just like in the two-allele case, the statistical additive effects depend on both functional additive and dominance effects and the statistical dominance deviations depend only on the functional dominance effects. However, a new feature emerging from equations (18) and (19) is that any one nonzero functional dominance effect between a particular pair of alleles is sufficient to cause nonzero values of all statistical dominance deviations.

The presence of Hardy-Weinberg disequilibrium (HWD)

In a HWD population, the average effects (statistical effects) of the alleles, A_1, A_1, …, A_r can only be given implicitly [e.g., 15, 19]

$$\alpha_u+\sum_{v=1}^{r}\left(P_{uv}/p_u\right)\alpha_v=\alpha_u^*, u=1,2,...,r \tag{20}$$

where α_u^* is the average excess of allele A_u defined as the deviation of the mean values of genotypes carrying one or two copies of A_u weighted by its allele frequency from the population mean,

$$\alpha_u^*=\frac{\sum_{v=1}^{r}P_{uv}G_{uv}}{\sum_{v=1}^{r}P_{uv}}-\mu=\frac{\sum_{v=1}^{r}P_{uv}G_{uv}}{p_u}-\mu \tag{21}$$

In the HWE population, the average effect and average excess of a given allele are the same (i.e., $P_{uv}=p_up_v$ and thus $\alpha_u^*=\alpha_u=\Sigma_{v=1}^{r}p_vG_{uv}-\mu$).

However, when mating is not random, α_u and α_u^* are generally different from each other. It is difficult to explicitly establish the relations of the average excesses in a HWD population to functional genetic effects. Therefore, a numerical evaluation is provided in the following section to assess the impact of HWD on the average excesses of multiple alleles.

Numerical analysis

We will now provide a reanalysis of the example of a three-allele population taken from C.C. Li in Kempthorne [18]. This example is chosen for the following reasons. First, it has special features (e.g., functional dominance effects appear only in some genotypic comparisons but not in others). Second, our reanalysis of the example actually provides a clear answer to C.C. Li's original question of why statistical dominance effects are present in all genotypes when apparently there is lack of functional dominance effects for some genotypic comparisons. In response, Kempthorne [18] did not really answer the question. And the question has left unanswered since then. Third, our reanalysis has much broader scope than does C.C. Li's original analysis which only considered the simplest case of a HWE population with a specific set of allele frequencies, thereby clearly demonstrating that a set of genotypic values there are only one set of functional genetic effects but numerous sets of statistical genetic effects depending on gene and genotypic frequencies.

C.C. Li's analysis

C.C. Li in Kempthorne [8] used a hypothetical example of three-allele HWE population to calculate functional and statistical effects of individual alleles and to ask the question of why the two types of effects are different. In Li's example, there are three alleles (A_1, A_2, A_3) and six possible genotypes: A_1A_1, A_1A_2, A_2A_2, A_1A_3, A_2A_3 and A_3A_3 with genotypic values of $\mathbf{G}_A$ = [10 30 50 36 46 42]'. Li assumed a HWE population with the frequencies of the three alleles being chosen as $p_1 = 0.2$ for A_1, $p_2 = 0.3$ for A_2, and $p_3 = 0.5$ for A_3 so that the population mean is $\mu = \Sigma_{u=1}^{3}\Sigma_{v=1}^{3} p_u p_v G_{uv} = 40$. Using the direct approach, the functional additive effects are $a_{12} = (50 - 10)/2 = 20$, $a_{13} = (42 - 10)/2 = 16$ and $a_{23} = a_{12} - a_{13} = 4$ whereas the functional dominance effects are $d_{12} = 30 - (10+50)/2 = 0$, $d_{13} = 36 - (10+42)/2 = 10$ and $d_{23} = 46 - (50+42)/2 = 0$. These functional effects can of course be easily obtained using the population mean as the reference point (cf. equation (12)), $\mathbf{E}_{\mu.A} = \mathbf{S}_{\mu.A}^{-1}\mathbf{G}_A$ = [40 20 0 16 10 0]'. The average effects of alleles, A_1, A_2, and A_3 are $\alpha_1 = \Sigma_{v=1}^{3} p_v G_{1v} - \mu$ = -11, $\alpha_2 = \Sigma_{v=1}^{3} p_v G_{2v} - \mu = 4$ and $\alpha_3 = \Sigma_{v=1}^{3} p_v G_{3v} - \mu = 2$. The dominance deviations of six possible genotypes: A_1A_1, A_1A_2, A_2A_2, A_1A_3, A_2A_3 and A_3A_3 based on $\delta_{uv} = G_{uv} - \mu - \alpha_u - \alpha_v$ are -8, -3, 2, 5, 0 and -2, respectively.

New analysis

In Li's original analysis as recapitulated above, the allele frequencies (= 0.2 for A_1, $p_2 = 0.3$ for A_2, and $p_3 = 0.5$ for A_3) were intentionally chosen to have the mean of μ = 40 in the HWE population. The genotypic values corrected for the mean ($\mathbf{G}_A$ - **1** μ) = [-30 -10 10 -4 6 2]'.The functional effects using these deviations from the mean is $\mathbf{E}_{0.A} = \mathbf{S}_{\mu.A}^{-1}(\mathbf{G}_A - 1\mu)$ = [0 20 0 16 10 0]' as in [10]. The statistical additive effects can be obtained from their relations with the functional effects $\boldsymbol{\alpha} = \mathbf{T}_{\alpha}\mathbf{E}_{0.A}$, i.e.,

$$\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix} = \begin{bmatrix}½ & -0.3 & 0.18 & -0.5 & 0.3 & -0.3\\ ½ & 0.7 & 0.08 & -0.5 & -0.2 & 0.2\\ ½ & -0.3 & -0.12 & 0.5 & 0 & 0\end{bmatrix}\begin{bmatrix}0\\20\\0\\16\\10\\0\end{bmatrix} = \begin{bmatrix}-11\\4\\2\end{bmatrix}$$

Similarly, the statistical dominance effects can be obtained from their relations with the functional effects $\boldsymbol{\delta} = \mathbf{T}_{\delta}\mathbf{E}_{0.A}$, i.e.,

$$\begin{bmatrix}\delta_{11}\\ \delta_{12}\\ \delta_{22}\\ \delta_{13}\\ \delta_{23}\\ \delta_{33}\end{bmatrix} = \begin{bmatrix}0 & 0 & -0.48 & 0 & -0.8 & 0.3\\ 0 & 0 & 0.62 & 0 & -0.3 & -0.2\\ 0 & 0 & -0.48 & 0 & 0.2 & -0.7\\ 0 & 0 & -0.18 & 0 & 0.5 & 0\\ 0 & 0 & -0.08 & 0 & 0 & 0.5\\ 0 & 0 & 0.12 & 0 & -0.2 & -0.3\end{bmatrix}\begin{bmatrix}0\\20\\0\\16\\10\\0\end{bmatrix} = \begin{bmatrix}-8\\-3\\2\\5\\0\\-2\end{bmatrix}$$

This directly answers Li's original question of why the functional dominance only occurs for the allele pair A_1A_3 (i.e., A_1A_3 is not at the mid-point of A_1A_1 and A_2A_2) but the statistical dominance deviations are nonzero in all but one genotypes.

Impacts of HWD on statistical genetic effects

We now conduct further analyses of Li's example by considering different sets of allele frequencies and varying levels of HWD to demonstrate the effects of allele frequencies and HWD on statistical genetic effects (average effects and average excesses in particular) and to gain insights into behaviors of average effects and average excesses of individual alleles. We construct a large number of HWD populations with different gene and genotypic frequency distributions but all

with the same genotypic values of $\mathbf{G}_A$ = [10 30 50 36 46 42]'. For a given set of allele frequencies, we calculate six genotypic frequencies as the sums of HWE frequencies and HWD deviations (*D*s) following [20],

$$P_u^u = p_u^2 + D_u^u$$
$$P_v^u = p_u p_v - D_v^u$$

There are a range of values that each HWD coefficient (D_{uv}) can take, $D_{\max}^{-} \le D_{uv} \le D_{\max}^{+}$, where $D^{-}_{\max} = -\min[p_u^2, p_v^2]$ and $D_{\max}^{+} = p_u p_v$. It is evident that if $D_v^u \to D_{\max}^{-}$ (e.g., populations undergoing selection for heterozygotes or negative assortative matings) then the population is approaching to a complete heterozygosity; on the other hand, if $D_{uv} \to D_{\max}^{+}$ (e.g., populations consisting of inbred lines) then the populaion is approaching to a complete homozygosity. To avoid the possibility of undefined negative genotypic frequencies with the presence of two extreme disequilibria, we choose five levels of HWD away from the two extremes ($D_{\max}^{-}$ and $D_{\max}^{+}$): D_{uv} = $\frac{1}{2}D_{\max}^{-}$, $\frac{1}{4}D_{\max}^{-}$, 0, $\frac{1}{4}D_{\max}^{+}$ and $\frac{1}{2}D_{\max}^{+}$. We then create the two extreme cases, one with complete heterozygosity by equaling the frequencies of three alleles to those of the three heterozygotes and the other with complete homozygosity by equaling the frequencies of three alleles to those of the three homozygotes.

Presented in Table 1 are the values of average excess (α_u^*) and additive effect (α_u) for seven selected sets of allele frequencies each for the

Table 1. The statistical additive effects and average excesses of three alleles in Li's example* for the populations constructed with varying levels of Hardy-Weinberg disequilibrium for seven sets of allele frequencies.

Allele frequency array ($p_1 : p_2 : p_3$)	Additive effect or average excess	Level of Hardy-Weinberg disequilibrium						
		All Hetero.	$D_v^u = \frac{D_{\max}^{-}}{2}$	$D_v^u = \frac{D_{\max}^{-}}{4}$	$D_v^u = 0$	$D_v^u = \frac{D_{\max}^{+}}{4}$	$D_v^u = \frac{D_{\max}^{+}}{2}$	All Homo.
0.05 : 0.1 : 0.85	α_1^*	-1.033	-6.825	-7.388	-7.950	-13.763	-19.575	-31.200
	α_1	-8.350	-7.274	-7.622	-7.950	-11.010	-13.050	-15.600
	α_2^*	3.967	3.075	3.313	3.550	4.863	6.175	8.800
	α_2	1.650	3.375	3.468	3.550	3.890	4.117	4.400
	α_3^*	0.353	0.040	0.045	0.050	0.238	0.425	0.800
	α_3	7.650	0.031	0.040	0.050	0.190	0.283	0.400
0.1 : 0.1 : 0.8	α_1^*	-1.067	-6.200	-7.300	-8.400	-13.700	-19.000	-29.600
	α_1	-8.200	-7.080	-7.786	-8.400	-10.960	-12.667	-14.800
	α_2^*	1.600	2.300	2.950	3.600	5.300	7.000	10.400
	α_2	1.800	2.920	3.295	3.600	4.240	4.667	5.200
	α_3^*	0.711	0.488	0.544	0.600	1.050	1.500	2.400
	α_3	7.800	0.520	0.561	0.600	0.840	1.000	1.200
0.2 : 0.3 : 0.5	α_1^*	-3.514	-6.800	-8.900	-11.000	-15.250	19.500	-28.000
	α_1	-8.900	-9.079	-10.191	-11.000	-12.200	-13.000	-14.000
	α_2^*	1.800	1.667	2.833	4.000	6.000	8.000	12.000
	α_2	1.100	2.491	3.387	4.000	4.800	5.333	6.000
	α_3^*	1.950	1.720	1.860	2.000	2.500	3.000	4.000

Table 1. continued..

Allele frequency array ($p_1 : p_2 : p_3$)	Additive effect or average excess	Level of Hardy-Weinberg disequilibrium						
		All Hetero.	$D_v^u = \frac{D_{\max}^-}{2}$	$D_v^u = \frac{D_{\max}^-}{4}$	$D_v^u = 0$	$D_v^u = \frac{D_{\max}^+}{4}$	$D_v^u = \frac{D_{\max}^+}{2}$	All Homo
	α_3	7.100	2.137	2.044	2.000	2.000	2.000	2.000
0.4 : 0.3 : 0.3	α_1^*	-4.029	-5.925	-8.063	-10.200	-13.050	-15.900	-21.600
	α_1	-8.300	-9.480	-9.923	-10.200	-10.440	-10.600	-10.800
	α_2^*	0.257	2.300	4.550	6.800	9.700	12.600	18.400
	α_2	1.700	3.320	5.551	6.800	7.760	8.400	9.200
	α_3^*	4.400	5.600	6.200	6.800	7.700	8.600	10.400
	α_3	7.700	9.320	7.680	6.800	6.160	5.733	5.200
0.5 : 0.3 : 0.2	α_1^*	-4.286	-6.760	-7.980	-9.200	11.500	-13.800	-18.400
	α_1	-8.000	-9.063	-9.156	-9.200	-9.200	-9.200	-9.200
	α_2^*	0.000	5.133	6.967	8.800	12.000	15.200	21.600
	α_2	2.000	7.291	8.187	8.800	9.600	10.133	10.800
	α_3^*	6.000	9.200	9.500	9.800	10.750	11.700	13.600
	α_3	8.000	11.722	10.609	9.800	8.600	7.800	6.800
0.3 : 0.6 : 0.1	α_1^*	-8.700	-9.867	-11.533	-13.200	-16.700	-20.200	-27.200
	α_1	-10.100	-12.950	-13.094	-13.200	-13.360	-13.467	-13.600
	α_2^*	0.467	4.167	4.983	5.800	7.550	9.300	12.800
	α_2	-0.100	5.555	5.690	5.800	6.040	6.200	6.400
	α_3^*	4.371	4.600	4.700	4.800	4.800	4.800	4.800
	α_3	5.900	5.522	5.142	4.800	3.840	3.200	2.400
0.1 : 0.85 : 0.05	α_1^*	-11.900	-16.100	-16.750	-17.400	-21.950	-26.500	-35.600
	α_1	-11.950	-17.275	-17.339	-17.400	-17.560	-17.667	-17.800
	α_2^*	0.416	1.952	2.026	2.100	2.675	3.250	4.400
	α_2	-1.950	2.069	2.085	2.100	2.140	2.167	2.200
	α_3^*	1.544	-0.975	-0.938	-0.900	-1.575	-2.250	-3.600
	α_3	4.050	-0.626	-0.766	-0.900	-1.260	-1.500	-1.800

*In Li's example (Kempthorne, 1955), there are three alleles (A_1, A_2, A_3) and six possible genotypes: A_1A_1, A_1A_2, A_2A_2, A_1A_3, A_2A_3 and A_3A_3 with genotypic values of $\mathbf{G}_A$ = [10 30 50 36 46 42]'. With the exception of the reference, all functional effects based on this set of genotypic values remain the same across different sets of allele frequencies and different levels of Hardy-Weinberg disequilibrium. These effects are a_{12} = 20, d_{12} = 0, a_{13} = 16, d_{13} = 10 and d_{23} = 0.

seven HWD coefficients. The two special features of the general relationship between α_u^* and α_u as given in (20) are evident. First, the average effects and average excesses of the same alleles are the same in the HWE populations ($\alpha_u^* = \alpha_u$). In this case, the average effects of the three alleles obtained in the Li's original analysis (α_1 = -11; α_2 = 4; and α_3 = 2) are recovered. Second, the average excess of a given allele is twice the average effects in the populations with complete homozygosity, i.e., completely inbred populations ($\alpha_u^* = 2\alpha_u$). In all other cases, the differences between α_u^* and α_u increase with the levels of HWD in either positive or negative direction, but no clear patterns exist in the direction of such differences across different levels of HWD and sets of allele frequencies. The deviations of α_u^* and α_u from their HWE expectations increase with the rate of approach to either complete homozygosity or complete heterozygosity.

DISCUSSION

Recently, there is a considerable amount of discussion about the need to distinguish functional vs. statistical effects of alleles at one or more loci [e.g., 3, 5, 6, 8, 9, 16]. However, such discussion is often limited to the case of two alleles per locus. In the present study, we extend the formulation of functional effects for the di-allelic case to the general formulation for the multi-allelic case from any reference point and establish a relationship between functional and statistical effects of multiple alleles. Our extension reveals some new features that do not exist in the diallelic case. First, with only two alleles per locus, there is only one functional additive effect (one half the difference between the two homozygotes) and one functional dominance effect (the deviation of heterozygote from the average of two homozygotes) whereas with r(>2) alleles, there are $r(r\text{-}1)/2$ functional additive effects and $r(r\text{-}1)/2$ functional dominance effects, but only (r-1) functional additive effects need to be specified and the remaining $(r\text{-}1)(r\text{-}2)/2$ can be derived. Note that the di-allelic case actually fits to these expressions with r = 2 and thus one additive effect, one dominance effect and zero remaining additive effects, whereas when r>2 there will always be linearly dependent additive effects (i.e. $(r\text{-}1)(r\text{-}2)/2>0$). Second, the presence of functional dominance effect at only one allele pair is sufficient to cause the presence of statistical dominance deviations for all the genotypes [cf. equation (24)]. The reanalysis of LI's example in Kempthorne [18] eloquently shows this feature. In the one-locus di-allelic case, the dominance effect plays the role of a last-order deviation which is thus unaffected by the other estimates. The same situation occurs in the two-locus di-allelic case, being the dominance-by-dominance effect the last-order deviation. However, when several alleles are present in one locus, there is no single last-order deviation, but several ones at the same level that, therefore, generate the above emergent property by being affected by each other in the system. And finally, the need for the functional vs. statistical distinction is more pronounced in the presence of multiple alleles because the equality of gene frequencies is no longer a sufficient condition for any direct relationship between physiological and statistical genetic effects in the multi-allelic case, as opposed to the di-allelic case [3].

We provide a satisfactory answer to a longstanding question of why there is lack of direct correspondence between physiological and statistical genetic effects in the multi-allelic case as raised by C. C. Li in Kempthorne [18]. Both previous di-allelic models and present multi-allelic models have demonstrated that for a given set of genotypic values, there is only one set of functional genetic effects (measured as an array of genotypic comparisons) but numerous sets of statistical genetic effects, each varying with gene and genotypic frequencies. Therefore, there is little chance for functional and statistical effects to be coincident. In the presence of two alleles only (r = 2), there are special cases (e.g., an F_2 population) where the two alleles are equally frequent ($p_1 = p_2 = ½$) and functional and statistical additive effects are clearly related: $a_A = 2p_2\alpha_1 - 2p_1\alpha_2 = \alpha_1 - \alpha_2 = \alpha_A$, i.e., the functional additive effect (a_A) equals to the average effect of gene substitution (α_A). This is also evident from the definition of $\alpha_A = a_A + d_A(p_2 - p_1)$ that the functional dominance does not enter the

average effect of gene substitution, $\alpha_A = a_A$ for $p_1 = p_2 = ½$. However, this clear relationship disappears when the consideration is shifted to three or more alleles. For example, with three equally frequent alleles ($p_1 = p_2 = p_3 = 1/3$), the relationships between functional and statistical effects (cf. equation (18)), are no longer as simple as in the di-allelic case,

$$\begin{bmatrix}\alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix} = \begin{bmatrix} ½ & -⅓ & ⅑ & -⅓ & ⅑ & -⅔ \\ ½ & ⅔ & ⅑ & -⅓ & -⅔ & ⅑ \\ ½ & -⅓ & -⅔ & ⅔ & ⅑ & ⅑ \end{bmatrix} \begin{bmatrix} 0\\ a_{12}\\ d_{12}\\ a_{13}\\ d_{13}\\ d_{23}\end{bmatrix} = \begin{bmatrix} -⅓(a_{12}+a_{13})+⅑(d_{12}+d_{13}-2d_{23}) \\ ⅓(2a_{12}-a_{13})+⅑(d_{12}-2d_{13}+d_{23}) \\ -⅓(a_{12}-2a_{13})+⅑(-2d_{12}+d_{13}+d_{23}) \end{bmatrix}$$

Clearly, each statistical additive effect depends on both functional additive and dominance effects. Sometimes, confusion may arise from lack of the distinction between functional and statistical genetic effects. For example, when developing the general two-allele (G2A) model, Zeng *et al.* [9] made no distinction between a_A and α_A in their equation (20) where the second row of the $\mathbf{S}^{-1}_{\text{G2A}\cdot A}$ matrix multiplied by the column vector of genotypic values $\mathbf{G}_A$ gives,

$$\begin{aligned} p_1G_{11}+(1-2p_1)G_{12}-(1-p_1)G_{22} &= p_1(\mu+a_A)+(1-2p_1)(\mu+d_A)-(1-p_1)(\mu-a_A) \\ &= a_A+d_A(1-2p_1) \end{aligned}$$

which is equal to α_A as describe above, not to a_A (as Zeng *et al.* intended to show) unless $p_1 = p_2 = ½$. Thus, in actual QTL mapping, the distinction of functional and statistical genetic effects is needed to avoid any unintended interpretation of the effects.

Our models for functional effects of multiple alleles complement the existing diallelic models for quantitative genetic studies and QTL mapping. In the past, QTL mapping has most often been carried out using some special segregating populations derived from a cross between two inbred lines, such as F_2 or backcross (e.g., 15). In these populations, there are only two alleles at a locus with known gene frequencies. However, there are many other types of populations including mapping populations derived from multi-way crosses between more than two inbred lines or populations with an unknown population structure (e.g., natural populations). QTL mapping for these populations requires multi-allelic models. Models for statistical effects of multiple alleles are well known (e.g., 7, 15, 19), but these models are hardly used in practical QTL mapping efforts because of their complicated nature. We have shown through theoretical and numerical analyses that our straightforward formulation of functional effects of multiple alleles coupled with the establishment of its relations to the statistical genetic effects will facilitate the use of multi-allelic models for QTL mapping. For any of such efforts, detection and estimation QTL effects can first be based on a functional genetic model without regard of gene and genotypic frequencies and then estimated effects can be transformed into statistical genetic effects accounting for gene and genotypic frequencies through the use of equations (16) and (17).

Just like in the diallelic case, both functional and statistical effects of multiple alleles are needed for describing genotypic and genic effects and variation, respectively. The models for the statistical effects focus on the heritable part of genotypic values. A diploid parent transmits only one allele per locus to each of its progeny and the additive effect of the transmitted allele is expressed when combined with a gene from another randomly chosen parent in a particular population, but the parent's dominance deviation due to the interaction between its two alleles is immediately gone once meiosis takes place to produce gametes for next generation. Thus, the

statistical additive effect of an allele necessarily depends on the allele frequency in the population. On the other hand, the models for functional effects emphasize on describing the standing genetic effects without reference to any population. As opposed to the statistical formulations, the functional ones allow us to describe genetic effects as effects of allele substitutions preformed from an individual genotype. This kind of description is needed, for instance, for analyzing studying evolution processes in which new mutations appear from a monomorphic ancestral population [4]. Furthermore, in this communication we have also shown that the functional formulation conceptually fits to a common statistical testing procedure.

In this study, we have explicitly considered only the single-locus models for functional effects of multiple alleles. If there is no linkage disequilibrium and/or epistasis between different loci, then our results for a single locus are directly applicable. In this case, the value of a genotype at *m* loci is simply the sum of the single-locus genotypic values ($\Sigma_{j=1}^{m} {}^{j}G_{uv}$), where ${}^{j}G_{uv}$ is the value of genotype A_uA_v at the *j*th locus that can be described in terms of functional and statistical formulations. The presence of epistasis at unlinked loci can be accommodated using the Kronecker product of genetic-effect design matrices just as described for the diallelic case. The functional formulation with linkage disequilibrium (LD) presents no further complication because LD represents a statistical property of a population and functional genetic effects are invariant across populations with varying levels of LD. However, if epistasis is present as well, the use of the Kronecker product of genetic-effect design matrices for individual loci is no longer feasible because these loci are not independent. Moreover, it is not a trivial matter to establish the relations between multilocus statistical and functional genetic effects with the presence of LD.

ACKNOWLEDGMENTS

We thank A. Le Rouzic for helpful discussion and comments. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada grant OGP0183983.

APPENDIX

Special cases of equation (5)

Equation (5) gives the general expression for the three-allele functional genetic-effect design matrix. Here we describe some special cases of this general result that either provide further illustration of the change-of-reference operation or show the equivalence to some well-known **S**-matrices for two alleles. If the reference point is changed from G_{11} to G_{23}, then $P_{23} = 1$ and the rest of genotypic frequencies are zeros, and $p_1 = 0$, $p_2 = p_3 = ½$. Plugging these gene and genotypic frequencies into (5), we have,

$$\mathbf{S}_{G23.A} = \begin{bmatrix} 1 & -1 & 0 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & 1 & 0 & -1 & 0 & -1 \\ 1 & -1 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \end{bmatrix}$$

Similarly, different **S**-matrices can arise from changing the reference point to the values of other individual genotypes.

Alternatively, it is possible to use the mean of a specific population as a reference point. For example, equation (5) can be used to calculate tri-allelic equivalents of diallelic F_∞-metric, F_2-metric, *UWR* **S**-matrices if, like in the diallelic models, the equal allelic frequencies are assumed. Thus, for the tri-allelic equivalent of F_∞-metric model ($P_{11} = P_{22} = P_{33} = 1/3$; $P_{12} = P_{13} = P_{23} = 0$; $p_1 = p_2 = p_3 = 1/3$), the **S**-matrix is

$$\mathbf{S}_{F_\infty.A} = \begin{bmatrix} 1 & -⅔ & 0 & -⅔ & 0 & 0 \\ 1 & ⅓ & 1 & -⅔ & 0 & 0 \\ 1 & 1⅓ & 0 & -⅔ & 0 & 0 \\ 1 & -⅔ & 0 & ⅓ & 1 & 0 \\ 1 & ⅓ & 0 & ⅓ & 0 & 1 \\ 1 & -⅔ & 0 & 1⅓ & 0 & 0 \end{bmatrix} \qquad \text{(A1)}$$

Similarly, we can obtain the **S**-matrices for F_2-metric model ($P_{11} = P_{22} = P_{33} = 1/9$; $P_{12} = P_{13} = P_{23} = 2/9$; $p_1 = p_2 = p_3 = 1/3$):

$$\mathbf{S}_{F_2.A} = \begin{bmatrix} 1 & -\frac{2}{3} & -\frac{2}{9} & -\frac{2}{3} & -\frac{2}{9} & -\frac{2}{9} \\ 1 & \frac{1}{3} & \frac{7}{9} & -\frac{2}{3} & -\frac{2}{9} & -\frac{2}{9} \\ 1 & 1\frac{1}{3} & -\frac{2}{9} & -\frac{2}{3} & -\frac{2}{9} & -\frac{2}{9} \\ 1 & -\frac{2}{3} & -\frac{2}{9} & \frac{1}{3} & \frac{7}{9} & -\frac{2}{9} \\ 1 & \frac{1}{3} & -\frac{2}{9} & \frac{1}{3} & -\frac{2}{9} & \frac{7}{9} \\ 1 & -\frac{2}{3} & -\frac{2}{9} & 1\frac{1}{3} & -\frac{2}{9} & -\frac{2}{9} \end{bmatrix} \quad \text{(A2)}$$

and UWR model ($P_{11} = P_{22} = P_{33} = P_{12} = P_{13} = P_{23} = 1/6$; $p_1 = p_2 = p_3 = 1/3$):

$$\mathbf{S}_{UW.A} = \begin{bmatrix} 1 & -\frac{2}{3} & -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} \\ 1 & \frac{1}{3} & \frac{5}{6} & -\frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} \\ 1 & 1\frac{1}{3} & -\frac{1}{6} & -\frac{2}{3} & -\frac{1}{6} & -\frac{1}{6} \\ 1 & -\frac{2}{3} & -\frac{1}{6} & \frac{1}{3} & \frac{5}{6} & -\frac{1}{6} \\ 1 & \frac{1}{3} & -\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & \frac{5}{6} \\ 1 & -\frac{2}{3} & -\frac{1}{6} & 1\frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} \end{bmatrix} \quad \text{(A3)}$$

For the two-allele case, the functional formulations of the F_2, F_∞ and *UWR* models are orthogonal in populations whose frequencies fit the reference points of those models [3]. It is worth noting that this is no longer valid for the multi-allelic case. In particular, expressions (A1-A3) do not lead to orthogonal formulations of different additive and dominance genetic effects.

REFERENCES

1. Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. 2001, Genetics, 157, 1819.
2. Xu, S. 2003, Genetics, 163, 789.
3. Álvarez-Castro, J. M., and Carlborg, O. 2007, Genetics, 176, 1151.
4. Álvarez-Castro, J. M., Le Rouzic, A., and Carlborg, O. 2008, Plos Genetics, 4(5), E1000062.
5. Cheverud, J. M. 2000, Epistasis And The Evolutionary Process, Wolf, J. B., Brodie Iii, E. D., and Wade, M. J. (Eds.), Oxford University Press, New York, 58.
6. Cheverud, J. M., and Routman, E. J. 1995, Genetics, 139, 1455.
7. Wang, T., and Zeng, Z-B. 2006, BMC Genetics, 7, 9.
8. Yang, R-C. 2004**,** Genetics, 167, 1493.
9. Zeng, Z-B., Wang, T., and Zou, W. 2005**,** Genetics, 169, 1711.
10. Fisher, R. A. 1918, Trans. Roy. Soc. Edinburgh, 52, 399.
11. Falconer, D. S., and Mackay, T. F. C. 1996, Introduction to Quantitative Genetics, Ed. 4, Longman, Harlow, UK.
12. Mather, K., and Jinks, J. L. 1982. Biometrical Genetics, Ed. 3, Chapman & Hall, London.
13. Van Der Veen, J. H. 1959, Genetica, 30, 201.
14. Cockerham, C. C. 1954, Genetics, 39, 859.
15. Lynch, M., and Walsh, B. 1998, Genetics And Analysis Of Quantitative Traits, Sinauer Associates, Sunderland, Massachussetts, USA.
16. Hansen, T. F., and Wagner, G. P. 2001, Theor. Popul. Biol., 59, 61.
17. Le Rouzic, A., and Álvarez-Castro, J. M. 2008, Evolutionary Bioinformatics, In Press.
18. Kempthorne, O. 1955, Cold Spring Harbor Symp. Quant. Biol., 20, 60.
19. Kempthorne, O. 1957, An Introduction To Genetic Statistics, John Wiley & Sons, New York.
20. Weir, B. S. 1996. Genetic Data Analysis II, Sinauer Associates, Sunderland, Massachussetts, USA.