# Multiallelic models of genetic effects and variance decomposition in non-equilibrium populations

**José M. Álvarez-Castro · Rong-Cai Yang**

**Abstract** Quantitative genetics stems from the theoretical models of genetic effects, which are re-parameterizations of the genotypic values into parameters of biological (genetic) relevance. Different formulations of genetic effects are adequate to address different subjects. We thus need to generalize and unify them under a common framework for enabling researchers to easily transform genetic effects between different biological meanings. The Natural and Orthogonal InterActions (NOIA) model of genetic effects has been developed to achieve this aim. Here, we further implement the statistical formulation of NOIA with multiple alleles under Hardy–Weinberg departures (HWD). We show that our developments are straightforwardly connected to the decomposition of the genetic variance and we point out several emergent properties of multiallelic quantitative genetic models, as compared to the biallelic ones. Further, NOIA entails a natural extension of one-locus developments to multiple epistatic loci under linkage equilibrium. Therefore, we present an extension of the orthogonal decomposition of the genetic variance to multiple epistatic, multiallelic loci under HWD. We illustrate this theory with a graphical interpretation and an analysis of published data on the human acid phosphatase (*ACP1*) polymorphism.

**Keywords** Models of genetic effects · Hardy–Weinberg disequilibrium · Multiple alleles · Variance decomposition · Acid phosphatase polymorphism

J. M. Álvarez-Castro (✉)
Department of Genetics, University of Santiago de Compostela, Avda. Carvalho Calero s/n, 27002 Lugo, Galiza, Spain
e-mail: jose.alvarez.castro@usc.es
URL: http://webspersoais.usc.es/persoais/jose.alvarez.castro/index_en.html

R.-C. Yang
Research and Innovation Division, Alberta Agriculture and Rural Development, Edmonton, AB T6H 5T6, Canada

R.-C. Yang
Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada
e-mail: rong-cai.yang@ualberta.ca

## Introduction

The models of genetic effects are re-parameterizations of the genotypic values into parameters entailing clearer biological (genetic) meanings (Kempthorne 1957; Fisher 1918, 1930). A general form of such reparameterizations can be written in matrix notation as $\mathbf{G} = \mathbf{SE}$, where the vector of genotypic values, $\mathbf{G}$, is expressed as a linear function of the vector of genetic effects, $\mathbf{E}$ (Cockerham and Zeng 1996). Cockerham (1954) expressed statistical models of genetic effects in terms of orthogonal scales, also known as contrasts, whose coefficients are used to form the genetic-effect design matrix, $\mathbf{S}$. Tiwari and Elston (1997) showed that this notation enormously facilitates the extension of the models of genetic effects to several loci under linkage equilibrium (LE) and Zeng et al. (2005) have further shown the convenience of this notation for statistical analyses of two-allele equilibrium populations.

The $\mathbf{G} = \mathbf{SE}$ matrix notation also provides an appropriate theoretical framework to unify different models of genetic effects (Álvarez-Castro and Carlborg 2007). Some issues in quantitative genetics need to be addressed using different mathematical formulations and it is thus necessary to unify all different uses of genetic effects under a

single theoretical perspective (Phillips 2008). The natural and orthogonal interactions (NOIA) model has indeed been developed for this purpose by using the matrix notation (Álvarez-Castro and Carlborg 2007).

The initial setting of the NOIA model provided a framework under which genetic effects can be defined on the basis of allele substitutions from the reference of individual genotypes (functional effects) or average effects of substitutions in populations (statistical effects) for the case of two alleles per locus. As an example to show the difference between these two formulations, the additive functional effect is always equal to half the difference between the performance of the two homozygotes whereas the additive statistical effect is the coefficient of weighted regression on the allele content—the number of one of the alleles—in the different genotypes present in a population. We have further extended the functional formulation of the NOIA model to the case of multiple alleles and explored the relationship between multiallelic functional and statistical genetic effects (Yang and Álvarez-Castro 2008). In that article, we have also shown some new properties of the multiallelic models as compared to the biallelic ones.

We here extend the NOIA statistical formulation to the multiallelic framework in non-equilibrium populations. To do this, we first obtain the multiallelic genetic effects as explicit functions of the genotypic values. This enables us to obtain the scales for the multiallelic statistical genetic-effect design matrix, $\mathbf{S}$. Next, we show that this $\mathbf{G} = \mathbf{SE}$ matrix notation makes it possible to provide the components of the genetic variance with multiple alleles also as explicit functions of the genotypic values through relatively simple expressions. The major focus of this communication is on one multiallelic locus under Hardy–Weinberg dissequilibrium (HWD). Under LE, extensions of one-locus formulations to an arbitrary number of multiallelic, epistatic loci is straightforward. We illustrate the theory developed in this communication with a graphical interpretation and an application to published experimental data on the polymorphism at the human acid phosphatase locus (*ACP1*) (Greene et al. 2000).

## The biallelic statistical formulation of NOIA

Before embarking on our new multillelic model, we provide a brief overview on the biallelic model as described by Álvarez-Castro and Carlborg (2007). For a one-locus two-allele ($A_1$ and $A_2$) genetic system, the general expression for the statistical formulation as a function of the genotypic frequencies of the population, $p_{ij}$, $ij = 11, 12, 22$, can be expressed in matrix notation as $\mathbf{G} = \mathbf{SE}$. This matrix expression can be expanded as:

$$\begin{pmatrix} G_{11} \\ G_{12} \\ G_{22} \end{pmatrix} = \begin{pmatrix} 1 & -2p_2 & -\frac{p_{12}p_{22}}{2p_1p_2 - 1/2p_{12}} \\ 1 & p_1 - p_2 & \frac{p_{11}p_{22}}{p_1p_2 - 1/4p_{12}} \\ 1 & 2p_1 & -\frac{p_{11}p_{12}}{2p_1p_2 - 1/2p_{12}} \end{pmatrix} \cdot \begin{pmatrix} \mu \\ \alpha \\ \delta \end{pmatrix}, \quad (1)$$

where the reference point of the model is the population mean, $\mu$, $p_i$, $i = 1, 2$, are the frequencies of the alleles, $p_i = p_{ii} + \frac{1}{2}p_{ij}$, $j \neq i$, and $\alpha$ and $\delta$ are, respectively, the additive and dominant statistical genetic effects.

The use of the $\mathbf{G} = \mathbf{SE}$ matrix notation enables us to readily obtain the genetic effects defined under one reference population as a function of the genetic effects defined under another reference population. A well-known example is the translation of the genetic effects defined under $F_2$ and $F_\infty$ populations (Yang 2004; Van der Veen 1959). Given two different decompositions (*i.e.* different formulations and/or from a different reference point) $\mathbf{G} = \mathbf{S}_1\mathbf{E}_1$ and $\mathbf{G} = \mathbf{S}_2\mathbf{E}_2$, we can obtain $\mathbf{E}_2$ from $\mathbf{E}_1$ by (Álvarez-Castro and Carlborg 2007):

$$\mathbf{E}_2 = (\mathbf{S}_2)^{-1}\mathbf{S}_1\mathbf{E}_1. \quad (2)$$

For obtaining a vector of genetic effects fitting to a particular biological meaning, $\mathbf{E}_2$, from another vector of genetic effects or genotypic values, it is necessary to have an expression for its corresponding genetic-effect design matrix, $\mathbf{S}_2$. In other words, expression (2) can be applied only when the appropriate genetic-effect design matrices are available. It is thus desirable to obtain expressions of the form $\mathbf{G} = \mathbf{SE}$ describing the properties of all possible genetic systems and populations. Hereafter, we develop genetic-effect design matrices fitting to statistical genetic effects and the decomposition of the genetic variance of multiallelic loci, whether their population frequencies are under HWD or not.

## The multiallelic statistical formulation of NOIA

We will now describe a one-locus multiallelic statistical formulation of genetic effects. We will first use the matrix notation to express the average genetic effects in terms of the genotypic values, particularly in the form $\mathbf{E} = \mathbf{S}^{-1}\mathbf{G}$. We will thus provide algorithms to build inverse genetic-effect design matrices that fit to the definitions of the genetic effects for any population frequencies. We start from the decomposition of the genotypic values into statistical genetic effects as expressed by Kempthorne (1957):

$$\mathbf{G} = \mathbf{1}\mu + \mathbf{N}\alpha_g + \delta_G. \quad (3)$$

This expression entails a multiple regression in which the additive (average) effects of the $r$ alleles, $\alpha_g = (\alpha_i)$, $i = 1,\ldots,r$, are the regression coefficients and the dominance deviations of the genotypes, $\delta_G = (\delta_{ij})$, $i,j = 1,\ldots,r$, $i > j$, are the residuals. Note that we use a capital $G$ in the subscripts to indicate genotypes and a lower–case $g$ to

indicate alleles. The vector $\mathbf{1}$ is an $n \times 1$ vector of ones, where $n = r(r + 1)/2$ is the number of possible genotypes, and $\mathbf{N}$ is the genetic content matrix, as detailed in expression (16), "Appendix A". The product $\mathbf{N}\boldsymbol{\alpha}_g = (\alpha_{ij}) = \boldsymbol{\alpha}_G$ is thus the vector of the additive components of the genotypic values, often called the breeding values.

The solution of regression (3) can be obtained from its normal equation (Kempthorne 1957) as we recapitulate in "Appendix A" for completeness. From that solution (18), we have obtained the genetic effects explicitly, as a linear function of the genotypic values. To do so we have factored the vector $\mathbf{G}$ as:

$$\boldsymbol{\alpha}_g = \left( (\mathbf{N}^T \mathbf{P}_{diag} \mathbf{N})^{-1} \mathbf{N}^T \mathbf{P}_{diag} (\mathbf{I} - \mathbf{1}\mathbf{P}_G^T) \right) \mathbf{G}, \tag{4}$$

where $\mathbf{I}$ is the identity matrix of dimension $n \times n$, $\mathbf{P}_G = (\begin{array}{cccccc} p_{11} & p_{12} & p_{22} & p_{13} & \dots & p_{rr} \end{array})^T$, $\mathbf{P}_{diag} = \text{Diag}(\mathbf{P}_G)$, "Diag" of a vector generates a diagonal matrix with that vector in the diagonal and the superscript "T" stands for the transposition operation. We can thus rewrite expression (4) as:

$$\boldsymbol{\alpha}_g = \mathbf{A}\mathbf{G}. \tag{5}$$

The additive genetic effects are just subtractions of the allelic effects in $\boldsymbol{\alpha}_g$, $\alpha^{ij} = \alpha_j - \alpha_i$ (we justify the use of these superscripts in "Appendix B"). Therefore, it is possible to operate from matrix $\mathbf{A}$ to obtain these genetic effects, which are the additive-related rows of the inverse genetic-effect design matrix $\mathbf{S}^{-1}$. We build an operator $\mathbf{B}$ adding a column filled with $-1$ to the left of an identity matrix so that we can obtain the additive genetic effects as $\mathbf{BAG}$. This product provides $(r - 1)$ genetic additive effects, the remaining ones being easily retrieved from them (cf. multiallelic functional genetic-effect design matrices in Yang and Álvarez-Castro 2008).

We need two additional easy-to-build matrix operators for constructing the inverse genetic-effect design matrix $\mathbf{S}^{-1}$. The first operator, $\mathbf{C}$, places the rows of $\mathbf{BA}$ in the appropriate positions of an inverse genetic-effect design matrix $\mathbf{S}^{-1}$. The other operator, $\mathbf{S}_{\mu\delta}$, adds the rows with the appropriate values that define the mean and dominance deviations in $\mathbf{S}^{-1}$. We refer to "Appendix B" for a detailed definition of the three operators and we define:

$$\mathbf{S}^{-1} = \mathbf{CBA} + \mathbf{S}_{\mu\delta}. \tag{6}$$

Finally, the inverse of this full-rank matrix is the genetic-effect design matrix $\mathbf{S}$ and we can obtain the desired statistical decomposition of genotypic values through the expression $\mathbf{G} = \mathbf{SE}$.

### Allelic effects from genetic effects

Vectors of genetic effects can be obtained, for instance, as the output of a typical QTL analysis of a line-cross experiment (see e.g. Zeng et al. 2005). With only two alleles, it is straightforward to obtain the additive (average) effects of the alleles from the statistical formulation $\mathbf{G} = \mathbf{SE}$. This expression entails the decomposition of genetic values, from which the additive (average) effects can be computed just as $\alpha_i = (\frac{1}{2}) \alpha_{ii}$, $i = 1, \dots, r$ (in general, $\alpha_{ij} = \alpha_i + \alpha_j$). However, in the multiallelic case, the additive contributions and the dominance deviations are the summation of several products of scalars from $\mathbf{S}$ and genetic effects from $\mathbf{E}$. Thus, we here provide an automated procedure for obtaining the additive (average) effects and average excesses from vectors of genetic effects, $\mathbf{E}$. Expressions (4–6) straightforwardly provide such a procedure. Indeed, combining expression (5) with formulations of the type $\mathbf{G} = \mathbf{SE}$ it follows:

$$\boldsymbol{\alpha}_g = \mathbf{ASE}, \tag{7}$$

The statistical decomposition of the genotypic values, $\mathbf{G} = \mathbf{SE}$ (1, 6), enables us to express each genotypic value as $G_{ij} = \mu + \alpha_{ij} + \delta_{ij}$, as in expression (3). Cancelling out the coefficients of the mean and the additive effects in $\mathbf{S}$ leads to the removal of the terms $\mu$ and $\alpha_{ij}$ in this expression, thereby leaving the dominance effects $\delta_{ij}$ alone. We thus express that column vector $\boldsymbol{\delta}_G = (\delta_{ij})$ as:

$$\boldsymbol{\delta}_G = \mathbf{S}\text{Diag}(\Delta_i)\mathbf{E}, \tag{8}$$

where $(\Delta_i)_{i=1,\dots,n}$ is a vector of index variables for dominance-related columns of $\mathbf{S}$, which coincide with the positions of the dominance effects in the vector $\mathbf{E}$ (an automated algorithm to obtain this vector of indexes is provided in "Appendix B").

### Hardy–Weinberg equilibrium

Under the Hardy–Weinberg proportions, the expressions within the matrices of the statistical formulation can be more easily illustrated than in the general (non-equilibrium) case. In particular, the matrix $\mathbf{A}$ for three alleles can be expressed as:

$$\mathbf{A} = \begin{pmatrix} p_1(p_2 + p_3) & p_2(1 - 2p_1) & -p_2^2 & p_3(1 - 2p_1) & -2p_2p_3 & -p_3^2 \\ -p_1^2 & p_1(1 - 2p_2) & p_2(p_1 + p_3) & -2p_1p_3 & p_3(1 - 2p_2) & -p_3^2 \\ -p_1^2 & -2p_1p_2 & -p_2^2 & p_1(1 - 2p_3) & p_2(1 - 2p_3) & p_3(p_1 + p_2) \end{pmatrix}. \tag{9}$$

Using (7) and (27), the three-allele statistical formulation under Hardy–Weinberg proportions, $\mathbf{E} = \mathbf{S}^{-1}\mathbf{G}$, is expanded to:

effects and average excesses of alleles (see e.g. Lynch and Walsh (1998) and (23) in "Appendix A"). From that expression, but using just vectors, the additive variance can

$$
\begin{pmatrix} \mu \\ \alpha^{12} \\ \delta^{12} \\ \alpha^{13} \\ \delta^{13} \\ \delta^{23} \end{pmatrix} = \begin{pmatrix} p_1^2 & 2p_1p_2 & p_2^2 & 2p_1p_3 & 2p_2p_3 & p_3^2 \\ -p_1 & p_1-p_2 & p_2 & -p_3 & p_3 & 0 \\ -\tfrac{1}{2} & 1 & -\tfrac{1}{2} & 0 & 0 & 0 \\ -p_1 & -p_2 & 0 & p_1-p_3 & p_2 & p_3 \\ -\tfrac{1}{2} & 0 & 0 & 1 & 0 & -\tfrac{1}{2} \\ 0 & 0 & -\tfrac{1}{2} & 0 & 1 & -\tfrac{1}{2} \end{pmatrix} \cdot \begin{pmatrix} G_{11} \\ G_{12} \\ G_{22} \\ G_{13} \\ G_{23} \\ G_{33} \end{pmatrix}, \tag{10}
$$

As a check, we get $\mathbf{G} = \mathbf{SE}$ as:

$$
\begin{pmatrix} G_{11} \\ G_{12} \\ G_{22} \\ G_{13} \\ G_{23} \\ G_{33} \end{pmatrix} = \begin{pmatrix} 1 & -2p_2 & -2(1-p_1)p_2 & -2p_3 & -2(1-p_1)p_3 & 2p_2p_3 \\ 1 & 1-2p_2 & 2p_1p_2+p_3 & -2p_3 & 2p_1p_3-p_3 & 2p_2p_3-p_3 \\ 1 & 2(p_1+p_3) & -2p_1(1-p_2) & -2p_3 & 2p_1p_3 & -2(1-p_2)p_3 \\ 1 & -2p_2 & 2p_1p_2-p_2 & 1-2p_3 & 2p_1p_3+p_2 & 2p_2p_3-p_2 \\ 1 & 1-2p_2 & 2p_1p_2-p_1 & 1-2p_3 & 2p_1p_3-p_1 & 2p_2p_3+p_1 \\ 1 & -2p_2 & 2p_1p_2 & 2(p_1+p_2) & -2p_1(1-p_3) & -2p_2(1-p_3) \end{pmatrix} \cdot \begin{pmatrix} \mu \\ \alpha^{12} \\ \delta^{12} \\ \alpha^{13} \\ \delta^{13} \\ \delta^{23} \end{pmatrix}. \tag{11}
$$

In "Appendix B", we justify the notation of the scalars within the $\mathbf{E}$ vector in expressions (10, 11). Expressions (9–11) are, as expected, equivalent to previous expressions of multiallelic genetic effects under the Hardy–Weinberg proportions (Kempthorne 1957; Wang and Zeng 2006, 2009). We here provide them using matrix notation as a particular case of a general formulation including also HWD—expressions (4–6).

**The decomposition of the genetic variance**

In expressions (4–8), we have formulated statistical multiallelic genetic effects as explicit functions of the genotypic values. Hereafter we use that formulation for the decomposition of the variance components. We indeed derive a method that provides the classical decomposition of the genetic variance of a trait in a population with the additional advantage of enabling a straightforward extension to genetic systems with multiple epistatic loci under LE, while allowing arbitrary numbers of alleles and arbitrary departures from the Hardy–Weinberg proportions at all loci.

The average excess of one allele, $\alpha_i^*$, is the difference by which the average of genotypes carrying that allele exceeds the average of genotypes carrying the alternative alleles (Fisher 1941). A common way to obtain the additive variance under HWD involves both the allelic additive (average)

be expressed as $V_A = 2\mathbf{P}_g^T\left(\boldsymbol{\alpha}_g \circ \boldsymbol{\alpha}_g^*\right)$, where $\boldsymbol{\alpha}_g^* = (\alpha_i^*)$, $\mathbf{P}_g$ is the column vector of the gene frequencies and "$\circ$" is the Hadamard product (just the pairwise product of the elements at the same position in the two vectors). It is also possible to compute the additive variance with HWD without the need of the use of the average excesses (see e.g. Bürger 2000). This can be done by using instead the additive genotypic components, often called the breeding values (25 in "Appendix A"). This way of computing the additive variance can also be expressed through just vectors as:

$$V_A = \mathbf{P}_G^T(\boldsymbol{\alpha}_G \circ \boldsymbol{\alpha}_G), \tag{12}$$

where $\boldsymbol{\alpha}_G = (\alpha_{ij}) = \mathbf{N}\boldsymbol{\alpha}_g$ is the vector of the additive components of the genotypic values. Similarly, from expression (24) in "Appendix A" the dominance variance can also be expressed as:

$$V_D = \mathbf{P}_G^T(\boldsymbol{\delta}_G \circ \boldsymbol{\delta}_G). \tag{13}$$

Now we provide a way to compute expressions (12, 13) simultaneously, which will conveniently have a straightforward extension to multiple loci. We first provide the decomposition of the genotypic values in matrix form as:

$$\mathbf{G}_{dec} = \mathbf{S}\mathrm{Diag}(\mathbf{E})\mathbf{H}, \tag{14}$$

where $\mathbf{H}$ is an operator that sums the additive-related and the dominance-related columns of the matrix to the left of

it. Therefore, with only two alleles, $\mathbf{H}$ is just the identity matrix (see "Appendix B" for details).

The three columns of $\mathbf{G}_{dec}$ give, respectively, the mean and the additive and dominance components of the genotypic values. This decomposition (14) is meant to be the one referred to in expression (3) and, thus, the second and third columns of the matrix $\mathbf{G}_{dec}$ are the vectors of additive and dominance effects, $\boldsymbol{\alpha}_G = \mathbf{N}\boldsymbol{\alpha}_g$ and $\boldsymbol{\delta}_G$ respectively. Using this information about expression (14) and expressions (12, 13), it is easy to see that the two components of the genetic variance can be obtained simultaneously, just as:

$$\mathbf{V} = \mathbf{P}_G^T(\mathbf{G}_{dec} \circ \mathbf{G}_{dec}). \tag{15}$$

The first scalar of $\mathbf{V}$ is actually the squared population mean, being the remaining two scalars the additive and the dominance variances, i.e. $\mathbf{V} = (\mu^2, V_A, V_D)$.

The properties of the statistical models of genetic effects for multiple alleles are not exactly the same as the ones of the two-allele case. With only two alleles, all genetic effects are orghogonal, despite HWD (Álvarez-Castro and Carlborg 2007; Yang 2004; Cockerham 1954). With three or more alleles there are several genetic effects accounting for additive (average) effects of allele substitutions. Since the effect of each allele substitution depends on the frequencies of all other alleles present in that locus, those additive parameters are dependent on each other. We want our parameters to reflect effects of substitutions and some of them will thus not be orthogonal to each other, even without HWD.

We can explain the same fact mathematically. The multiallelic genetic effects come from a multiple regression (3) and some of them will necessarily be statistically dependent on each other. Indeed, the statistical formulation of NOIA for multiple alleles (5) is not fully orthogonal. It is nevertheless orthogonal by blocks that gather effects of the same type (additive- or dominant-related effects) for different pairs of alleles, as illustrated in "Appendix A" (26). Therefore, the decomposition of genotypic values given by (14) is not fully orthogonal, but again orthogonal by blocks of effect-types. Conveniently, though, the variance decomposition performed from expression (15) is fully orthogonal. This is so because the components of variance are computed using the effects within each of the orthogonal blocks of the statistical formulation of NOIA. Thus, the variance decomposition provided by the NOIA model (15) is orthogonal even under departures from the Hardy–Weinberg proportions. Furthermore, expression (15) can be straightforwardly extended to an arbitrary number of loci with arbitrary numbers of alleles and arbitrary HWD, under LE.

## Multiple loci

The NOIA formulations can conveniently allow for straightforward extensions to multiple loci under LE. To distinguish the expressions of each of the locus, we implement the notation used so far with appropriate indicators for each locus name and number of alleles. We do this using subscripts and superscripts, respectively, in all matrices and vectors. For a locus $A$ with three alleles, for instance, we enunciate the statistical formulation as: $\mathbf{G}_A^{(3)} = \mathbf{S}_A^{(3)}\mathbf{E}_A^{(3)}$. We now consider a two-locus genetic system in which there is, in addition, a locus $B$ with two alleles. Assuming LE between the two loci, we use the Kronecker product of the one-locus genetic-effect design matrices (as in Álvarez-Castro and Carlborg 2007), $\mathbf{S}_{AB}^{(3,2)} = \mathbf{S}_B^{(2)} \otimes \mathbf{S}_A^{(3)}$, to obtain the two-locus statistical formulation as $\mathbf{G}_{AB}^{(3,2)} = \mathbf{S}_{AB}^{(3,2)}\mathbf{E}_{AB}^{(3,2)}$. The equivalent expression $\mathbf{E}_{AB}^{(3,2)} = \left(\mathbf{S}_{AB}^{(3,2)}\right)^{-1}\mathbf{G}_{AB}^{(3,2)}$ can be obtained by computing the inverse of the two-locus genetic-effect design matrix $\mathbf{S}_{AB}^{(3,2)}$, or equivalently using the inverses of the one-locus matrices as $\left(\mathbf{S}_{AB}^{(3,2)}\right)^{-1} = \left(\mathbf{S}_B^{(2)}\right)^{-1} \otimes \left(\mathbf{S}_A^{(3)}\right)^{-1}$. Either way, this entails the extension of the solution to Eq. (4) to multiple loci under LE. The general expression to obtain the statistical genetic-effect design matrix for an arbitrary number of loci, $l$, is: $\overset{1}{\underset{k=l}{\otimes}}\left(\mathbf{S}_{Lk}^{(rk)}\right)$, where $r_k$ is the number of alleles at locus $L_k$.

Once the multiallelic $\mathbf{S}$ matrix has been obtained, the decomposition of the genetic variance through expressions (14, 15) naturally holds for multiple loci. The multilocus operator $\mathbf{H}$ can be built from the single-locus ones as $\overset{1}{\underset{k=l}{\otimes}}\left(\mathbf{H}_{Lk}^{(rk)}\right)$, to account for the additional (due to interactions among loci) variance components. Applying expressions (14, 15) to the system of loci $A$ and $B$ leads to a vector of variance components $\mathbf{V}_{AB}^{(3,2)}$ in which the order of the variance components is the same as the one of the genetic effects in the vector $\mathbf{E}$ for two alleles (Álvarez-Castro and Carlborg 2007).

For the multilocus genetic systems, it is possible to test for orthogonality in the same way as shown for the one-locus case at the end of "Appendix A". By doing so at the system of loci $A$ and $B$ considered just above, we have obtained a matrix with nine independent non-zero blocks for the mean, additive effect of locus $A$, additive effects of locus $B$, dominance effect of locus $A$, dominance effects of locus $B$, and the additive-by-additive, additive-by-dominance, dominance-by-additive and dominance-by-dominance interactions—i.e. an analogous matrix to (26), although larger and having more independent blocks (not shown). These separate blocks reflect the orthogonality of all the variance components. Thus, expressions (14, 15) comprise a straightforward routine to perform orthogonal decomposition of variance from the NOIA model for

genetic systems of arbitrary numbers of alleles at multiple epistatic loci under LE.

## Applications

Here we show a graphical interpretation of the decomposition of the genotypic values for a three-allele case under HWD and we make an analysis of real data on the human *ACP1* polymorphism.

### Graphical interpretation

Álvarez-Castro and Carlborg (2007, Figures 2 and 3A) have represented graphically the biallelic statistical formulation of NOIA, although here we point out a misprint in that representation—there lacks a factor 2 adjacent to every $\alpha_i$, which also applies to every $a_i$ in Figures 2 and 3B of that article. On the other hand, Kempthorne (1957) has provided a graphical interpretation of the statistical decomposition of the genotypic values for the one-locus three-allele case. Although Kempthorne (1957) illustrated his graphical interpretation under the Hardy–Weinberg proportions, an analogous interpretation holds when, like in the case we are dealing with here, there are departures from these proportions. In Fig. 1 we actually provide a graphical interpretation that fits this case, based on the formulation of the NOIA model for non-equilibrium populations (6).

Figure 1 is produced using the one-locus three-allele example taken from Yang and Álvarez-Castro (2008) where the genotypic values $\mathbf{G} = (10, 30, 50, 36, 46, 42)^{\mathrm{T}}$ and the genotypic frequencies $\mathbf{P} = (0.12, 0.06, 0.195, 0.1, 0.15, 0.375)$. The genotypic values are represented by globes whose sizes are in accordance with their genotypic frequencies. The decomposition of the genotypic value $G_{11} = \mu + \alpha_{11} + \delta_{11}$ is marked by vertical grey arrows that represent the additive expectation $\alpha_{11}$ (from the mean to the value predicted by the regression plane) and the dominance deviation $\delta_{11}$ (the departure between the multilinear prediction and the true genotypic value). The genotypic values are expressed as functions from the two-dimensional domain defined by the axis of the gene content of alleles $A_2$ and $A_3$, $G(c_2, c_3)$, and so it is the regression plane, $\hat{G}(c_2, c_3) = 18.\overline{3}c_2 + 15c_3 + 13$. The intercept in this regression, 13, is the predicted value for the genotype $A_1A_1$, $\alpha_{11}$—for which the allele contents of the alleles $A_2$ and $A_3$ are zero. Under the Hardy–Weinberg proportions, the predictions of the regression plane are the breeding values. In Fig. 1 it is possible to observe that, although only the genotypic value of the heterozygote $A_1A_3$ lies outside the midpoint of the two flanking homozygotes, the regression plane does not meet, for instance, the genotypic value $G_{12}$.
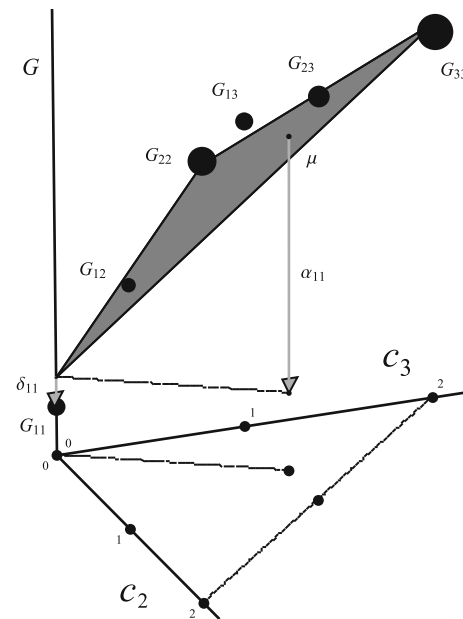


**Fig. 1** Graphical interpretation of the statistical decomposition of the genotypic values as a function of the content of alleles "2" and "3", $G(c_2, c_3)$, for the one-locus three-allele example given by the genotypic values $\mathbf{G} = (10, 30, 50, 36, 46, 42)^{\mathrm{T}}$ and the genotypic frequencies $\mathbf{P} = (0.12, 0.06, 0.195, 0.1, 0.15, 0.375)$, which is one of the instances we have considered in a previous publication (Yang and Álvarez-Castro 2008). The genotypic values are represented by globes whose sizes are in accordance with their genotypic frequencies. The decomposition of the genotypic value $G_{11} = \mu + \alpha_{11} + \delta_{11}$ is marked by vertical grey arrows that represent the additive expectation $\alpha_{11}$ (from the mean to the value predicted by the regression plane) and the dominance deviation $\delta_{11}$ (the departure between the multilinear prediction and the true genotypic value)

In other words, the presence of functional dominance interaction at one only heterozygote suffices to cause non zero dominance deviations, $\delta_{ij}$, for all genotypes—as already noted by Yang and Álvarez-Castro (2008).

### Analysis of the *ACP1* polymorphism

The acid phosphatase multiallelic polymorphism, *ACP1*, has been discovered in Europe almost half of a century ago (Hopkinson et al. 1963) and extensively studied ever since. Three alleles were found at different frequencies in northern European populations, *ACP1\*A*, *ACP1\*B* and *ACP1\*C* (hereafter A, B and C, respectively), for which no significant deviations from an additive inheritance of enzyme activity have been found (Spencer et al. 1964; Greene et al. 2000; Eze et al. 1974). Indeed, taking a vector of genotypic values, $\mathbf{G}^{\mathrm{ac}}$, from Greene et al. (2000, reproduced in our Table 1), and using the multiallelic functional formulation of NOIA (Yang and Álvarez-Castro 2008), we obtain functional genetic effects from the reference of the genotype AA, $\mathbf{E}_{\mathrm{AA}}^{\mathrm{ac}}$, showing that dominance effects are very small compared to the additive effects

**Table 1** Genotypic values for the traits "*ACP1* enzyme activity" and "*ACP1* enzyme inhibition", observed frequencies used in our analyses and results obtained for the different geontypes—fixation indexes, fitness values minimizing $V_A/V_G$ and equilibrium frequencies of these fitnesses

|  | ACP1 genotypes | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | AA | AB | BB | AC | BC | CC |
| ACP1 activity[a], $\mathbf{G}^{ac}$ | 122.4 | 153.9 | 188.3 | 183.6 | 212.3 | 240.0 |
| ACP1 inhibition[b], $\mathbf{G}^{in}$ | 41.2 | 37.9 | 34.4 | 58.7 | 53.1 | 76.0 |
| SD of ACP1 inhibition[b] | 4.3 | 5.2 | 5.2 | 8.3 | 3.8 | 15.5 |
| Observed frequencies[c], $p_{ij}$ | 0.1242 | 0.4139 | 0.3349 | 0.0445 | 0.0799 | 0.0025 |
| Observed fixation indexes, $F_{ij}$ | −0.003 | −0.006 | −0.015 | 0.028 | −0.061 | −0.027 |
| Fitness values, $\omega_{ij}$ | 0.9303 | 1 | 0.9667 | 1.0465 | 0.9658 | 0.5832 |
| Equilibrium frequencies, $p^*_{ij}$ | 0.1115 | 0.4323 | 0.3766 | 0.0295 | 0.0491 | 0.0001 |

[a] Greene et al. (2000). Expressed as μmol of *p*-niotrophenol liberated in 0.5 h at 37°C per g hemoglobin

[b] Sensabaugh and Golden (1978). Enzyme inhibition by 0.1 folic acid, expressed as μmol of *p*-niotrophenol liberated

[c] Brinkmann et al. (1971)

**Table 2** Genetic effects for the traits "*ACP1* enzyme activity" (ac), "*ACP1* enzyme inhibition" (in, with confidence intervals, CI) and for the stabilizing (st) genotype-fitness (GF) map of fitnesses minimizing the ratio $V_A/V_G$. Genotypic values, fitness values and observed frequencies as in Table 1

|  | Reference point and genetic effects[a] | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | $R$ | $a^{AB}$ | $a^{AC}$ | $a^{BC}$ | $d^{AB}$ | $d^{AC}$ | $d^{BC}$ |
| Enzyme activity, $\mathbf{E}^{ac}_{AA}$ | 122.4 | 35.95 | 58.80 | −23.85 | −1.45 | 2.40 | −1.85 |
| Enzyme activity, $\mathbf{E}^{ac}_{\mu}$ | 167.7 | 33.00 | 59.19 | 26.19 | −1.45 | 2.40 | −1.85 |
| Enzyme inhibition, $\mathbf{E}^{in}_{AA}$ | 41.2 | −3.4 | 17.4 | 20.8 | 0.1 | 0.1 | −2.1 |
| CI for $\mathbf{E}^{in}_{AA}$ | ±1.99 | ±0.85 | ±5.87 | ±5.72 | ±2.30 | ±10.56 | ±7.58 |
| Enzyme inhibition, $\mathbf{E}^{in}_{\mu}$ | 39.4 | −3.57 | 16.05 | 19.62 | 0.1 | 0.1 | −2.1 |
| Stabilizing GF map, $\mathbf{E}^{st}_{AA}$ | 0.930 | 0.018 | −0.174 | −0.192 | 0.052 | 0.290 | 0.191 |
| Stabilizing GF map, $\mathbf{E}^{st}_{\mu}$ | 0.978 | 0.000[b] | 0.000[b] | 0.000[b] | 0.052 | 0.290 | 0.191 |

[a] We use only Latin letters in the headings just for simplicity. It is understood that statistical genetic effects are named with the corresponding Greek letters, $\alpha$ instead of $a$, $\delta$ instead of $d$ and $\mu$ instead of $R$

[b] These values are zero to eight decimal places

(Table 2). The transformation tool of NOIA (2) enables us to obtain a vector of statistical genetic effects from a functional one (Álvarez-Castro and Carlborg 2007). Any statistical genetic effects are associated to certain population frequencies. We take them from a study with a sample of 7059 individuals from a German population (Brinkmann et al. 1971, see our Table 1). Using these frequencies (referred to as the observed frequencies hereafter) we obtain the vector of genetic effects, $\mathbf{E}^{ac}_{\mu}$, (Table 2). From this vector and (15) we obtain the following variance decomposition: $V_A = 658.43$, $V_D = 0.97$, $V_A/V_G = 0.999$, showing the very small contribution of the dominance effects of this trait to the genetic variance.

The high proportion of additive variance indicates that directional selection on *ACP1* activity would have a high response of the trait in this population, which would actually lead to the fixation of one of the alleles.

Sensabaugh and Golden (1978) inspected whether the maintenance of the polymorphism could be explained by a different trait—the inhibition of *ACP1* by folic acid. The genotypic values obtained for this trait, $\mathbf{G}^{in}$, and their standard deviations are shown in Table 1. From them we computed the functional and statistical genetic effects, $\mathbf{E}^{in}_{AA}$ and $\mathbf{E}^{in}_{\mu}$, which are again largely additive (Table 2). The dominance effects are actually not significantly different from zero. From the observed frequencies, the decomposition of the genetic variance for this trait is: $V_A = 44.54$, $V_D = 1.15$, $V_A/V_G = 0.997$. To determine the significance of the genetic effects, we have computed 95% confidence intervals of these estimates (Table 2). These just come from transforming the standard errors of the genotypic values into the standard errors of the genetic effects (Le Rouzic and Álvarez-Castro 2008). Then we have inspected
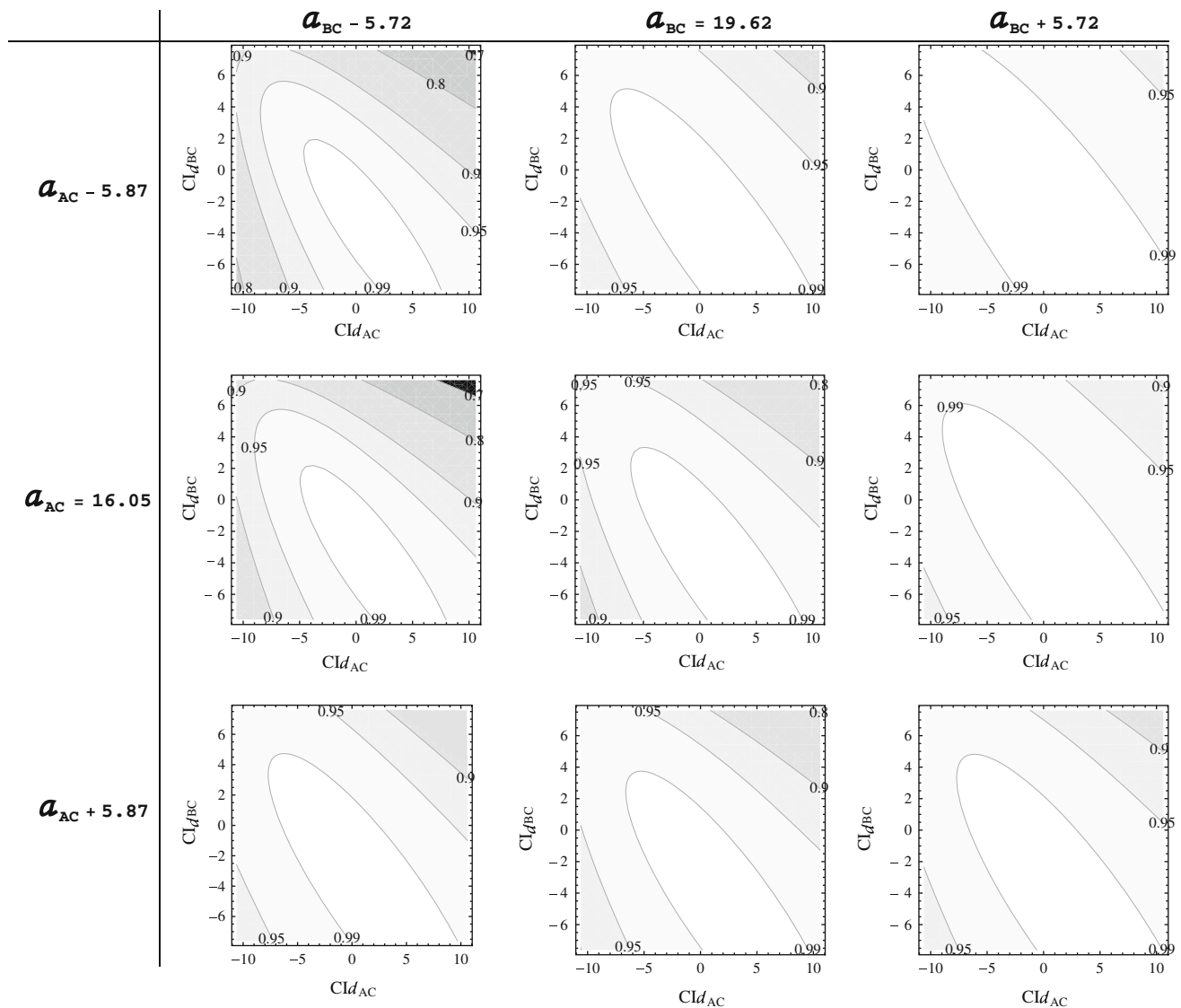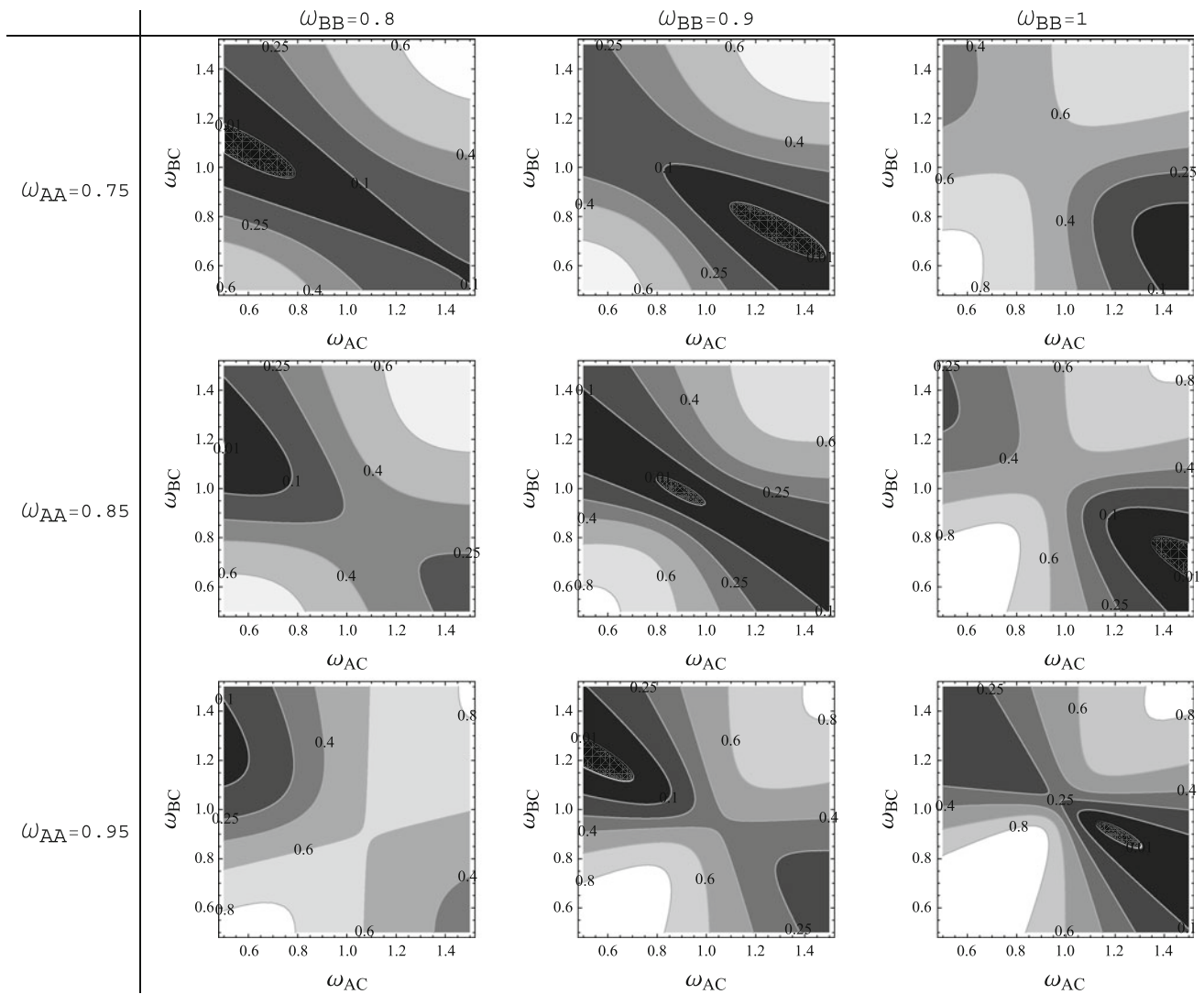
**Fig. 2** Contour plots of the ratio $V_A/V_G$ for *ACP1* activity (from Greene et al. 2000) with the observed frequencies (from Brinkmann et al. 1971) for the less accurate genetic effects estimates ranging within their confidence intervals

whether the proportion of additive variance varies significantly for parameter values within the confidence intervals of the estimates—we have used the theory provided in this communication (4–6, 14, 15) to plot the ratio $V_A/V_G$ within those intervals. We have obtained that at least three estimates have to reach values at the edges of their intervals for the ratio $V_A/V_G$ to be lower than 0.7 (Fig. 2). We can therefore conclude that this genotype-phenotype map is largely additive and that directional selection on the phenotype would lead to fixation of one of the alleles.

Alternatively, we have computed the fixation indexes of the observed frequencies (Weir 1996), which are shown in Table 1. These fixation indexes reflect HWD with a deficiency of all homozygotes and heterozygote AC and an

excess of heterozygotes AB and BC. According to the observations of Alvarez (2008), it is possible that the observed frequencies are equilibrium frequencies under viability selection acting on the *ACP1* locus. However, this selection pressure cannot be directional selection acting either on enzyme activity or on inhibition by folic acid, which in the absence of significant dominance interactions would lead to fixation of one of the alleles—as illustrated above. In fact, Greene et al. (2000) suggested that this polymorphism could instead be maintained by stabilizing selection due to the balance of two forces. On the one hand, *ACP1* genotypes with high enzyme-activity (particularly genotype CC) would not be well adapted to cold environments and therefore they could be selected against in

**Fig. 3** Contour plots of the ratio $V_A/V_G$ for various fitness values in accordance with the verbal model by Greene et al. (2000) with the observed frequencies (from Brinkmann et al. 1971). For all panels, $\omega_{AB} = 1$ and $\omega_{CC} = 0.6$

Northern Europe. On the other hand, phenotypes with the lowest enzyme activity levels (particularly genotype AA) were found to be associated with risk of macrosomia and adult obesity. Genotypes showing intermediate enzyme activities would thus in average perform better than the most extreme ones.

Selection acts on the fitnesses of the different genotypes present in a population. With directional selection on the trait value, the fitnesses of the individuals with a particular genotype directly reflect their phenotypes—their genotypic values. This is why the decomposition of the genetic variance of a trait is informative about the response of that trait to directional selection (e.g. Falconer and MacKay 1996). This does however not hold for other selection regimes. With stabilizing selection the genotype-phenotype

map is not a linear transformation of the genotype-fitness map and thus the former one cannot be used in substitution of the latest one, as it was the case for directional selection. Therefore, we hereafter address the variance decomposition of stabilizing genotype-fitness maps to analyze the effect of stabilizing selection on the *ACP1* polymorphism.

First, we have considered several *ad-hoc* genotype-fitness maps in accordance with the verbal model by Greene et al. (2000). In particular, we have fixed the fitness of the CC genotype at $\omega_{CC} = 0.6$ relative to the fitness of AB, $\omega_{AB} = 1$, and considered a variety of values for the fitnesses of the other genotypes. Then we have inspected whether they could explain the maintenance of the *ACP1* equilibrium. To do so, we have plotted the ratio $V_A/V_G$ for those genotype-fitness maps, using the observed

frequencies. These results are shown in Fig. 3, where there appear many regions with very low ratios $V_A/V_G$. These ratios leave very little room for selection to act. Therefore, they point us to sets of fitness values that could maintain the *ACP1* equilibrium at frequencies similar to the observed ones.

Next, assuming the observed frequencies, we have used Mathematica (Wolfram_Research_Inc. 2010) to find a minimum of the $V_A/V_G$ ratio—which we have expressed the as an explicit function of the fitness values using the theory provided in this communication (4–6, 14, 15)—with some constraints given by the observations of Greene et al. (2000). We have in particular set the reference fitness as $\omega_{AB} = 1$ without loss of generality and imposed two constraints: (1) the fitnesses reflect enzyme activity values in the absence of macrosomia ($\omega_{CC} < \omega_{BC} < \omega_{BB}$, $\omega_{AC}$) and (2) the extreme enzyme activities, particularly the lowest one, perform worse than the reference fitness but not extremely bad ($0.5 < \omega_{CC} < \omega_{AA} < \omega_{AB}$). The set of minimizing fitnesses we have obtained using this procedure (Table 1) leads to a ratio of $V_A/V_G$ of virtually zero.

As a next step, we have checked that the minimizing fitnesses we have obtained actually fulfill the necessary conditions for the maintenance of the multiallelic equilibrium, derived by Lewontin et al. (1978). Since these are, however, not sufficient conditions, we have run deterministic simulations in order to find out the outcome of the genetic system—we have run recursions of selection assuming random mating, no drift and non-overlapping generations (see e.g. Li 1976)—using the minimizing fitnesses and starting from the observed frequencies. After less than 1000 generations the population reaches a stable equilibrium (the frequencies stay accurately constant to five decimal places) with a set of frequencies close to the observed ones (Table 1). We have then used the multiallelic NOIA (Yang and Álvarez-Castro 2008 and expression (6) in this communication) to compute the functional and statistical genetic effects for the genotype-fitness map given by the minimizing fitnesses (Table 2). Additive and dominance functional effects are similar whereas statistical additive effects get to be virtually zero for the observed frequencies, which is in accordance with having obtained these fitnesses by minimizing the additive contribution in the genetic variance for those frequencies.

## Discussion

In this communication, we have provided algorithms for multiallelic formulations of statistical genetic effects using the $\mathbf{G} = \mathbf{SE}$ matrix notation. This notation has the major advantages of straightforwardly extending the models to multiple independent loci and connecting formulations that

entail different meanings of the parameters. This makes it possible to express the estimates of genetic effects detected in a QTL mapping experiment as average effects over a population of interest—with different genotypic frequencies than the one under study—and, thus, to easily obtain the decomposition of the genetic variance at that population (Álvarez-Castro and Carlborg 2007). Those estimates can also be expressed as effects of allele substitutions from a reference genotype, which are appropriate to assist the study of different evolutionary phenomena (Álvarez-Castro et al. 2008; Besnier et al. 2010; Le Rouzic and Álvarez-Castro 2008; Le Rouzic et al. 2008). Of course, the outcome of these transformations will depend on how accurate the estimates of genetic effects could be performed initially, e.g. how affected the genetic effects of the detected QTL were by non-detected QTL (see e.g. Zeng et al. 2005).

Connecting formulations of genetic effects

There are two essential ways of parameterizing models of genetic effects for diploid species. First, they can describe the properties of the genotypes of the individuals through the genotypic genetic effects (1, 11). Second, they can also describe the properties of the haploid genotypes through the allelic genetic effects (5). As noted by e.g. Templeton (1987), this is needed for comprehensive genetic analyses when diploid individuals pass on haploid gametes to their offspring. We observe, however, that the genotypic frequencies are implicit in the computation of the allelic additive (average) effects under HWD. Therefore, independently on whether focusing on diploid or haploid genotypes, we deem all formulations of genetic effects— the functional and the statistical formulations from any reference point—to be genotype-based models whenever applied to non-equilibrium populations. Indeed, using the matrix notation, the parameters of all formulations are visibly connected by linear transformations—expression (2)—, which reflects that they share analogous properties.

There currently is no full consensus to name the different formulations of models of genetic effects. Here, we are using the label "functional" following Hansen and Wagner's (2001) conceptualization of it. However, some more recent uses of this label are restricted to reflect interactions among particular molecules in a physiological pathway (e.g. Boone et al. 2007). Indeed, following Phillips' (2008) classification of uses of epistasis, "functional epistasis addresses the molecular interactions that proteins (and other genetic elements) have with one another" and NOIA's functional formulation would instead fit to his label "compositional". Whichever label is used, we concur that it is crucial to typify and unify all uses of genetic effects under a single perspective. NOIA has actually been developed to achieve that task, with or without epistasis,

via its different mathematical formulations and its transformation tool (Álvarez-Castro and Carlborg 2007).

Variance decomposition

The general solution to the decomposition of the genotypic values into statistical genetic effects—see expressions (3, 4)—has often been considered to be rather involved for the general, multiallelic case (e.g. Lynch and Walsh 1998). We have shown that the decomposition of the genetic effects and the genotypic variance can be straightforwardly provided by the multiallelic statistical formulation here developed, which can be obtained by performing simple algebra with easy-to-build matrices. Thus, we have provided a new, convenient extension of genetic modelling to multiple epistatic multiallelic loci with HWD, which comprises both the decomposition of the genotypic values and the variance decomposition, including all epistatic components. In brief, we use matrix notation to merge the models of genetic effects by Kempthorne (1954) and Cockerham (1954), for we consider together arbitrary numbers of alleles and loci, arbitrary epistasis and HWD.

As an example, this extension allows researchers to easily perform thorough examinations of the portion of genetic variance of a trait in a population that is due to additive variance. We have illustrated this point through the analysis of the *ACP1* polymorphism. We note that, although out of the scope of our current communication, the theory we are presenting here can be used to inspect the multilocus variance decomposition of genetic architectures under more general situations than previously, which enables researchers to assess the effect of HWD on the variance components of multilocus epistatic systems.

QTL analysis

The advent of quantitative trait loci (QTL) analysis (e.g. Lander and Botstein 1989) has challenged the state of the art of genetic modeling. QTL analysis pursues the aim of determining the loci underlying heritable traits, for which it unavoidably relies on models of genetic effects. These models are used to infer the underlying genetics from the phenotypes of individuals for which the genotypes at certain marker locations are known. Consequently, any constraints of the models of genetic effects instantly preclude researchers to unravel any genetic architecture not fitting those constraints. QTL analysis has motivated additional developments in models of genetic effects. For instance, the $\mathbf{G} = \mathbf{SE}$ matrix notation has proven convenient for QTL analysis, for it comprises an optimal way for implementing models of genetic effects into QTL analyses using the regression approach (see e.g. Álvarez-Castro and Carlborg 2007; Álvarez-Castro et al. 2008; Zeng et al.

2005). Therefore, the extension of the NOIA formulations to the multiallelic framework facilitates a versatile implementation of QTL analysis for multiallelic genetic systems. The implementation of HWD is also significant in this regard since models adapting to (being orthogonal at) the empirical populations under study may assist the mapping procedure both by speeding up the performance of estimates of genetic effects and by facilitating model selection strategies (see e.g. Kao and Zeng 2002; Yang 2004; Zeng et al. 2005).

Recently, a variance component approach has been proposed to detect segregation of multiple alleles in line-cross experiments—the FIA method (Rönnegård et al. 2008, 2009). NOIA has already been used to perform estimates of general genetic effects of biallelic loci detected by FIA (Besnier et al. 2010), and it can now be used for the same task with multiallelic loci, which are the recurrent output of variance component mapping methods in general. As an example, this combination of tools may well aid the genetic analysis of hybrid zones (Besnier et al. 2010).

Yang and Álvarez-Castro (2008) have shown that the functional formulation of genetic effects fits a common statistical testing procedure and that, therefore, it makes sense to use it to obtain estimates of genetic effects in QTL analysis. For the two-allele case, the statistical formulation has also been proposed for this task due to two major reasons. First, orthogonality facilitates the estimation and in particular the model selection procedure. Second, the statistical estimates are directly related to the decomposition of the genetic variance at the population or population sample under study (e.g. Álvarez-Castro and Carlborg 2007; Yang 2004; Zeng et al. 2005). For the multiallelic framework, we have shown in this communication that, although the statistical formulation is not fully orthogonal, it conveniently is orthogonal by blocks. This keeps on conferring to it an advantage over non-orthogonal settings, both for obtaining estimates in QTL analyses and for automatically performing the orthogonal decomposition of the genetic variance from those estimates whenever necessary. In any case, whatever formulation is used to obtain estimates, we recall that those are easily transferable among formulations by expression (2), the transformation tool (see also Álvarez-Castro and Carlborg 2007).

Analysis of the *ACP1* polymorphism

We have shown that the flexibility of our theory makes it easy to examine the values of an index of interest (e.g. $V_A/V_G$) for a range of parameter values (e.g. for values within the confidence intervals of the estimates of genetic effects). Obtaining fitness values that explain polymorphisms at biallelic loci is easy using the classical studies of equilibrium for these systems, but not when multiple alleles are

present (see e.g. Li 1967). Using NOIA to express the ratio $V_A/V_G$ as an explicit function of the model parameters has also enabled us to obtain a set of fitness values that are in accordance with experimental observations (Greene et al. 2000) and explain the maintenance of the multiallelic *ACP1* polymorphism. To this aim, we have first set stabilizing phenotype-fitness maps on top of the genotype-phenotype map for enzyme activity, by following Greene et al. (2000) empirical observations. Next, we have checked that the resulting genotype-fitness maps under those assumptions can lead to very low values for $V_A$—*i.e.* to very little room for selection to act. Finally, we have obtained fitness values that actually minimize the ratio $V_A/V_G$ to virtually zero. Using this procedure, we have obtained a set of fitnesses that can explain the maintenance of the *ACP1* polymorphism, particularly for one of the populations studied in greater detail.

The equilibrium frequencies obtained using these fitnesses are very similar to the observed ones, although we have checked statistically that the equilibrium and the observed frequencies differ significantly (using a Chi-square test). We do not deem the observed frequencies to be the exact equilibrium frequencies because the fitness values could have changed recently—cold exposure and death by macrosomia are likely to have varied during the latest generations—and migration events could also have recently affected the observed polymorphism frequencies. The main result of our analysis is that the polymorphism can be explained by stabilizing selection, whether the fitnesses we have obtained by minimizing $V_A/V_G$ are very close to the real ones or not.

The C allele has also been interpreted as a recessive deleterious allele that has not yet been removed by natural selection (Wilder and Hammer 2004). This was argued by pooling alleles A and B into cluster X and then noting a significant excess of XC and a significant deficiency of CC, but no significant departure of "genotype" XX from its expected frequency under the Hardy–Weinberg proportions. This fits with C being deleterious with respect to "allele" X. However, we note that this observation also fits with a balance in XX between an excess of AB and deficiencies of AA and BB, which is in accordance with stabilizing selection favouring the heterozygotes over the homozygotes. This is actually the case for the observed frequencies we have used in our analysis. Therefore, we do not find Wilder and Hammer's (2004) argument to support that the C allele is just deleterious.

How much additional complexity?

Achieving generality enables us to increase the explanatory power of our models. Through the analysis of the *ACP1* data, we have just recalled that the properties of a multiallelic system cannot be explained by means of reductionistic approaches using biallelic models. This fact was noticed by geneticists long ago (Lewontin et al. 1978; e.g. Kempthorne 1957). Multiallelic models acquire emergent properties and it is thus not surprising that the extension of NOIA to multiple alleles requires having to take into account increasing numbers of parameters and addressing new notation issues (see "Appendix B"). The absence of full orthogonality of the genetic effects is an emergent fact of multiallelic models that does not preclude us from obtaining an orthogonal decomposition of the genetic variance. In any case, the variance decomposition of the multiallelic system is not the sum of the variance decompositions of all reduced biallelic systems. This is in connection with another emergent fact we have already noted in a previous publication—functional dominance interaction between two alleles generates statistical dominance deviations between all pairs of alleles (Yang and Álvarez-Castro 2008). In that publication we also discussed that in the multiallelic case redundant additive genetic effects appear. These are missing in the **E**-vectors but can easily be retrieved from the remaining ones.

We are aware that achieving generality comes at the cost of increasing complexity. As an example, the dimension of a genetic-effect design matrix for a non-equilibrium genetic system of three loci with two, three and four alleles and all levels of epistatic interactions is $180 \times 180$. Prohibitive amounts of data would be necessary for obtaining sound estimates of the 180 parameters within the genetic effects vector of such a system. In any case, the complexity of the general model cannot possibly be perceived in itself as a disadvantage. Indeed, it is straightforward to reduce the general model to fit any desired constraints, namely to the absence of third order genetic interactions, to the same kind of interactions among alleles of certain genes or to no epistasis at all between certain pairs of loci (e.g. the absence of epistatic interactions involving the third locus of the example mentioned just above would reduce the number of parameters from 180 to 28). Also the connection of the general setting of NOIA to the more constrained multilinear model has been described (Le Rouzic and Álvarez-Castro 2008).

To sum up, from the theoretical perspective, we are motivated to provide as general and unified as possible formalizations of genetic effects, from which to eventually consider more constrained cases. In other words, we pursue the situation in which the available data—as opposed to the developed theory—sets the constraints of a quantitative analysis of genetic effects. In particular—although out of the scope of this communication—imprinting and gene-by-environment interactions could naturally be implemented within the theoretical framework here presented. Also, although significant achievements have been reported for

considering linkage disequilibrium in genetic modelling (e.g. Mao et al. 2006; Wang and Zeng 2006; Yang 2004), this topic is to our view liable to further development.

## Appendix A: Computations leading to the orthogonal decompositions of the genotypic values and the genetic variance

For simplicity, we abbreviate throughout this appendix the notation of some of the subscripts. In particular, we use $\mathbf{P}$ instead of $\mathbf{P}_{diag}$, $\boldsymbol{\alpha}$ instead of $\boldsymbol{\alpha}_g$, $\boldsymbol{\delta}$ instead of $\boldsymbol{\delta}_G$ and $\boldsymbol{\alpha}^*$ instead of $\boldsymbol{\alpha}_g^*$. The matrix $\mathbf{N}$ has as rows the vectors of the gene content of alleles $A_i$, for the genotypes in the vector of genotypic values, $\mathbf{G}$:

$$\mathbf{N} = \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 2 \end{pmatrix}. \tag{16}$$

The decomposition of the genotypic values

The least-squares solution to expression (3) is the set of values that minimizes:

$$\boldsymbol{\delta}^{\mathrm{T}}\mathbf{P}\boldsymbol{\delta} = (\mathbf{G} - \mathbf{1}\mu - \mathbf{N}\alpha)^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu - \mathbf{N}\alpha). \tag{17}$$

Thus, the solution comes from equalling to zero the derivative of (17) respect to $\boldsymbol{\alpha}$, $2\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu - \mathbf{N}\alpha) = 0$, which leads to the normal equations of the system:

$$(\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N})\alpha = \mathbf{N}^T\mathbf{P}(\mathbf{G} - \mathbf{1}\mu). \tag{18}$$

We can rewrite (18) in terms of average excesses as:

$$(\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N})\alpha = 2\left(\mathbf{P}_{\mathbf{g}}^{\mathrm{T}} \circ \boldsymbol{\alpha}^*\right). \tag{19}$$

The decomposition of the genetic variance

The total genetic variance is:

$$V_{\mathrm{G}} = \mathbf{G}^{\mathrm{T}}\mathbf{P}\mathbf{G} - \left(\mathbf{G}^{\mathrm{T}}\mathbf{P}\mathbf{1}\right)^2 = \mathbf{G}^{\mathrm{T}}\mathbf{P}\mathbf{G} - \mu^2, \tag{20}$$

which can be partitioned into components due to the additive and dominance effects given in (3). Indeed, from (20) and (3):

$$\begin{aligned} V_{\mathrm{G}} &= (\mathbf{1}\mu + \mathbf{N}\alpha + \delta)^{\mathrm{T}}\mathbf{P}(\mathbf{1}\mu + \mathbf{N}\alpha + \delta) - \mu^2 \\ &= \mathbf{1}^{\mathrm{T}}\mathbf{P}\mathbf{1}\mu^2 + \mathbf{1}^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha\mu + \mathbf{1}^{\mathrm{T}}\mathbf{P}\delta\mu + \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{1}\mu + \delta^{\mathrm{T}}\mathbf{P}\mathbf{1}\mu \\ &\quad + (\mathbf{N}\alpha + \delta)^{\mathrm{T}}\mathbf{P}(\mathbf{N}\alpha + \delta) - \mu^2. \end{aligned} \tag{21}$$

Now we consider separately several terms in this expression:

$$\mathbf{1}^{\mathrm{T}}\mathbf{P}\mathbf{1} = \sum_{i=1}^{r}\sum_{j=i}^{r} p_{ij} = 1,$$

$$\begin{aligned} \mathbf{1}^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha &= \sum_{i=1}^{r}\sum_{j=i}^{r} p_{ij}(\alpha_i + \alpha_j) = \sum_{i=1}^{r}\sum_{j=i}^{r} p_{ij}\alpha_i \\ &\quad + \sum_{j=1}^{r}\sum_{i=j}^{r} p_{ji}\alpha_j = \sum_{i=1}^{r} p_i\alpha_i + \sum_{j=1}^{r} p_j\alpha_j = 0, \end{aligned}$$

$$\mathbf{1}^{\mathrm{T}}\mathbf{P}\delta = \sum_{i=1}^{r}\sum_{j=i}^{r} p_{ij}\delta_{ij} = 0,$$

$$\alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{1} = \left(\mathbf{1}^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha\right)^{\mathrm{T}} = 0 \text{ and}$$

$$\delta^{\mathrm{T}}\mathbf{P}\mathbf{1} = \left(\mathbf{1}^{\mathrm{T}}\mathbf{P}\delta\right)^{\mathrm{T}} = 0,$$

which allows us to rewrite (21) as:

$$\begin{aligned} V_{\mathrm{G}} &= (\mathbf{N}\alpha + \boldsymbol{\delta})^{\mathrm{T}}\mathbf{P}(\mathbf{N}\alpha + \delta) \\ &= \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha + \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\delta + \boldsymbol{\delta}^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha + \boldsymbol{\delta}^{\mathrm{T}}\mathbf{P}\delta. \end{aligned} \tag{22}$$

Now we consider separately the four terms in this expression, having in mind (18) and (19):

$$\begin{aligned} \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha &= \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N}\left(\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N}\right)^{-1}\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu) \\ &= \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu) = 2\sum_{i=1}^{r} p_i\alpha_i\alpha_i^* = V_{\mathrm{A}} \end{aligned} \tag{23}$$

$$\begin{aligned} \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\delta &= \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu - \mathbf{N}\alpha) = \alpha^{\mathrm{T}}N^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu) \\ &\quad - \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha = \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu) \\ &\quad - \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N}\left(\mathbf{N}^{\mathrm{T}}\mathbf{P}\mathbf{N}\right)^{-1}\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu) \\ &= \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu) - \alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}(\mathbf{G} - \mathbf{1}\mu) = 0. \end{aligned}$$

$$\delta^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha = \left(\alpha^{\mathrm{T}}\mathbf{N}^{\mathrm{T}}\mathbf{P}\delta\right)^{\mathrm{T}} = 0.$$

$$\delta^T\mathbf{P}\delta = \sum_{i=1}^{r}\sum_{j=i}^{r} p_{ij}\delta_{ij}^2 = V_{\mathrm{D}}. \tag{24}$$

Thus, (22) provides the total variance as the sum of the additive and the dominance variances, $V_{\mathrm{G}} = V_{\mathrm{A}} + V_{\mathrm{D}}$, which are defined by (23) and (24) respectively. Note that (23) can also be expressed using the additive components of the genotypic values, $(\alpha_{ij}) = \mathbf{N}\alpha$ as:

$$V_{\mathrm{A}} = (\mathbf{N}\alpha)^{\mathrm{T}}\mathbf{P}\mathbf{N}\alpha = \sum_{i=1}^{r}\sum_{j=i}^{r} p_{ij}\alpha_{ij}^2. \tag{25}$$

Assessing orthogonality

We have expressed the decomposition of the genetic variance from the NOIA statistical formulation (6) in accordance to a proper decomposition of the genetic variance (22), leading to only the additive (23, 25) and dominance (24) summing terms being different from zero. Indeed, we hereafter provide a proof for this claim using a standard

method. We obtain matrix $\mathbf{S}_S$ from expression (6) for the three-locus case and change the order of its columns so that the additive and the dominance effects are pooled together—e.g. by just exchanging columns 3 and 4. We call $\mathbf{S}_{Sr}$ to the resulting reordered matrix and inspect its orthogonality by computing the following matrix product (see e.g. Álvarez-Castro and Carlborg 2007):

$$
\mathbf{S}_{Sr}^{\mathrm{T}} \cdot \mathbf{P} \cdot \mathbf{S}_{Sr}
$$
$$
= \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & f_{22}(p_{ij}) & f_{23}(p_{ij}) & 0 & 0 & 0 \\
0 & f_{23}(p_{ij}) & f_{33}(p_{ij}) & 0 & 0 & 0 \\
0 & 0 & 0 & f_{44}(p_{ij}) & f_{45}(p_{ij}) & f_{46}(p_{ij}) \\
0 & 0 & 0 & f_{45}(p_{ij}) & f_{55}(p_{ij}) & f_{56}(p_{ij}) \\
0 & 0 & 0 & f_{46}(p_{ij}) & f_{56}(p_{ij}) & f_{66}(p_{ij})
\end{pmatrix},
$$
$$
(26)
$$

where $f_{kl}(p_{ij})$ are functions of the genotypic frequencies.

Since expression (26) does not necessarily lead to a diagonal matrix, the formulation given by (6) is not completely orthogonal. More in detail, being $f_{23}(p_{ij}) \neq 0$ in (26) means that the additive effects are not orthogonal to each other. The remaining non-zero functions outside the diagonal of the matrix (26) mean that the dominance effects are not either orthogonal to each other. Finally, and interestingly, the upper-right and the bottom-left ($3 \times 3$)-blocks of zeros indicate that the dominance effects are orthogonal to the additive effects and to the reference point—which also holds for the additive effects and the reference point. This fact indeed shows that the variance due to the additive and the dominance effects are orthogonal—i.e. there is no covariance between them—, which extends to higher numbers of alleles. Therefore, (26) shows that the statistical formulation $\mathbf{G} = \mathbf{SE}$ coming from (6) leads to an orthogonal decomposition of the genetic variance. This result is a consequence of the additive and dominance parameters having been implemented in this formulation in accordance to expressions (3, 4).

## Appendix B: Technical details of the NOIA multiallelic formulation

Hereafter we provide detailed algorithms for some of the steps outlined in the main text, including notation issues.

Matrix-operators for the multiallelic statistical formulation

The multiallelic additive genetic effects of the vector of statistical genetic effects $\mathbf{E}_S$ are, as well as for the two-allele case, subtractions of the additive (average) effects of the alleles. Thus, we build an operator that performs those subtractions, $\mathbf{B} = (-\mathbf{1}_{r-1}|\mathbf{I}_{r-1})$ where $\mathbf{1}_{r-1}$ is a column-vector of $(r-1)$ unities and $\mathbf{I}_{r-1}$ is the identity matrix with dimension $(r-1)(r-1)$. The dimension of $\mathbf{B}$ is thus $(r-1)r$. Now we can express the statistical additive genetic effects of the $\mathbf{E}$ vector as $\mathbf{B}\boldsymbol{\alpha}_g$, not to be confused with $\boldsymbol{\alpha}_G = \mathbf{N}\boldsymbol{\alpha}_g$ (see comments on notation issues below). Equivalently, we can also express it from the genotypic values using (5) as $\mathbf{BAG}$. We note here that the vector $\mathbf{B}\boldsymbol{\alpha}_g$—as well as the $\mathbf{E}$ vector—does not include parameters for all the possible additive genetic effects since some of the genotypes are not represented. Both vectors include an independent set of these parameters from which the rest can be retrieved in the same way as for the analogous additive parameters of the multiallelic functional formulation of genetic effects (see Yang and Álvarez-Castro 2008 for details).

The dimension of $\mathbf{BA}$ is $n(r-1)$, which in the two-allele case equals $3 \times 1$. In order to build the inverse genetic-effect design matrix for an arbitrary number of alleles, we just need to place the rows of $\boldsymbol{\alpha}_G$ at the correct position of a square matrix and to then add to it the rows for the mean and the dominance effects. These rows are actually the same as the ones in the functional formulation of NOIA from the reference of $\mu$, $(\mathbf{S}_F)^{-1}$ (see again Yang and Álvarez-Castro 2008 for details). We substitute the scalars of the rows for the additive effects in $(\mathbf{S}_F)^{-1}$ by zeros and call $\mathbf{S}_{\mu\delta}$ to the resulting matrix. More explicitly, $\mathbf{S}_{\mu\delta} = \mathrm{Diag}(\Delta_i^{\mu})(\mathbf{S}_F)^{-1}$, where $(\Delta_i^{\mu})$ is a vector with a unity in the first position and equal to $(\Delta_i)$ otherwise (see details on $(\Delta_i)$ below).

Matrix $\mathbf{C}$, of dimension $(r-1)n$, is built with unities at the positions $ij$, given that we want to place the $i$th row of $\mathbf{BA}$ as the $j$th row of $(\mathbf{S}_S)^{-1}$, and zeros otherwise. The algorithm to build the matrix $\mathbf{C}$ can more precisely be described recursively. Starting from the $\mathbf{C}$ matrix for two alleles, $(0, 1, 0)^{\mathrm{T}}$, we perform two additions in each step towards considering one more allele ($r$ from $r-1$). First, we add on zeros below the existing columns up to getting to have $n$ rows. Finally, we add on a new column to the right of the previous matrix, all positions in it holding zeros except from the one corresponding to the first new row of the new matrix, which must hold a unity.

We also note here that the matrix $\mathbf{H}$ needed for expression (14) can be easily obtained from matrix $\mathbf{C}$. In fact, the second column of matrix $\mathbf{H}$ equals the sum of the columns of matrix $\mathbf{C}$. The first column of $\mathbf{H}$ is a unity followed by zeros, and the remaining third column is such that there is one only unity in each row of $\mathbf{H}$. This third column of $\mathbf{H}$ actually gives the vector $(\Delta_i)$ for expression (8). For instance, for $r = 3$ alleles and thus $n = 6$ genotypes, the matrices $\mathbf{B}$, $\mathbf{C}$, $\mathbf{H}$ and $\mathbf{S}_{\mu\delta}$ are simply:

$$\mathbf{B} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \tag{27}$$

$$\mathbf{S}_{\mu\delta} = \begin{pmatrix} p_{11} & p_{12} & p_{22} & p_{13} & p_{23} & p_{33} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1/2 & 1 & -1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & 0 & -1/2 & 0 & 1 & -1/2 \end{pmatrix}.$$

### A matter of notation

Here, we discuss the need of superscripts for genetic effects with multiple alleles, in contrast with the case of two alleles. The statistical formulation of genetic effects (6) leads to the decomposition of each genotypic value in terms of an additive contribution from the population mean and a dominance deviation, $G_{ij} = \mu + \alpha_{ij} + \delta_{ij}$, as in expression (3). In the two-allele case (1), the values $\alpha_{ij}$ and $\delta_{ij}$ come from multiplying the corresponding scalar of the second and third columns of the matrix $\mathbf{S}$ times the additive and the dominance genetic effects of the $\mathbf{E}$ vector ($\alpha$ and $\delta$), respectively. More to the point, the additive contribution is $\alpha_{ij} = \alpha_i + \alpha_j$, which is, under random mating, the breeding value of the genotype $ij$.

For considering the decomposition of the genotypic values of a multiallelic genetic system, however, we have to take into account that the $\mathbf{E}$ vector comprises more than one additive and one dominance effects—it actually comprises the $n$ dominance effects and $(r - 1)$ of the existing $n$ additive genetic effects, the remaining $(r - 1)(r - 2)/2$ of them being linear combinations of these—for further details see Yang and Álvarez-Castro (2008). Here we want to stress that the additive and the dominance genetic effects are properties of pairs of alleles, which is unnecessary to make explicit in the notation for two alleles (since all genetic effects are related to the same pair), but necessary for more general formulations.

Now, the easiest way to denote the additive and dominance genetic effects for a particular pair of alleles, e.g. $i$ and $j$, would probably be $\alpha_{ij}$ and $\delta_{ij}$, but this is actually the way in which we are already denoting the breeding values and the dominance deviations in the decomposition of the genotypic values (3) in vectors $\boldsymbol{\alpha}_G = (\alpha_{ij}) = \mathbf{N}\boldsymbol{\alpha}_g$ and $\boldsymbol{\delta}_G = (\delta_{ij})$. Therefore, we denote the statistical additive and dominance genetic effects in $\mathbf{E}$ using superscripts instead of subscripts: $\alpha^{ij}$ and $\delta^{ij}$.

### References

Alvarez G (2008) Deviations from Hardy–Weinberg proportions for multiple alleles under viability selection. Genet Res (Camb) 90(2):209–216

Álvarez-Castro JM, Carlborg Ö (2007) A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. Genetics 176(2):1151–1167

Álvarez-Castro JM, Le Rouzic A, Carlborg Ö (2008) How to perform meaningful estimates of genetic effects. PLoS Genet 4(5): e1000062

Besnier F, Le Rouzic A, Álvarez-Castro JM (2010) Applying QTL analysis to conservation genetics. Conserv Genet 11(2):399–408

Boone C, Bussey H, Andrews BJ (2007) Exploring genetic interactions and networks with yeast. Nat Rev Genet 8(6):437–449

Brinkmann B, Hoppe HH, Hennig W, Koops E (1971) Red cell enzyme polymorphisms in a northern German population. Gene frequencies and population genetics of the acid phosphatase (AP), phosphoglucomutase (PGM), adenylate kinase (AK), adenosine deaminase (ADA) and 6-phosphogluconate dehydrogenase (6-PGD). Hum Hered 21(3):278–288

Bürger R (2000) The mathematical theory of selection, recombination and mutation. Wiley, Chichester

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859–882

Cockerham CC, Zeng ZB (1996) Design III with marker loci. Genetics 143(3):1437–1456

Eze LC, Tweedie MC, Bullen MF, Wren PJ, Evans DA (1974) Quantitative genetics of human red cell acid phosphatase. Ann Hum Genet 37(3):333–340

Falconer DS, MacKay TFC (1996) Introduction to quantitative genetics, 4th edn. Prentice Hall, Harlow

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinburgh 52:339–433

Fisher RA (1930) The genetical theory of natural selection. Clarendon, Oxford

Fisher RA (1941) Average excess and average effect of a gene substitution. Ann Eugen 11:53–63

Greene LS, Bottini N, Borgiani P, Gloria-Bottini F (2000) Acid phosphatase locus 1 (ACP1): possible relationship of allelic variation to body size and human population adaptation to thermal stress-A theoretical perspective. Am J Hum Biol 12(5):688–701

Hansen TF, Wagner GP (2001) Modeling genetic architecture: a multilinear theory of gene interaction. Theor Popul Biol 59(1):61–86. doi:10.1006/tpbi.2000.1508

Hopkinson DA, Spencer N, Harris H (1963) Red cell acid phosphatase variants: a new human polymorphism. Nature 199:969–971

Kao CH, Zeng ZB (2002) Modeling epistasis of quantitative trait loci using Cockerham's model. Genetics 160(3):1243–1261

Kempthorne O (1954) The correlation between relatives in a random mating population. Proc R Soc Lond B Biol Sci 143(910):102–113

Kempthorne O (1957) An introduction to genetic statistics. Wiley, New York

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121(1):185–199

Le Rouzic A, Álvarez-Castro JM (2008) Estimation of genetic effects and genotype-phenotype maps. Evol Bioinform 4:225–235

Le Rouzic A, Álvarez-Castro JM, Carlborg Ö (2008) Dissection of the genetic architecture of body weight in chicken reveals the impact of epistasis on domestication traits. Genetics 179:1591–1599

Lewontin RC, Ginzburg LR, Tuljapurkar SD (1978) Heterosis as an explanation for large amounts of genic polymorphism. Genetics 88(1):149–169

Li CC (1967) Genetic equilibrium under selection. Biometrics 23(3):397–484

Li CC (1976) First course in population genetics. The Boxwood Press, Pacific Grove

Lynch M, Walsh B (1998) Genetic analysis of quantitative traits. Sinauer, Sunderland

Mao Y, London NR, Ma L, Dvorkin D, Da Y (2006) Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. Physiol Genomics 28(1):46–52. doi:10.1152/physiolgenomics.00096.2006

Phillips PC (2008) Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet 9(11):855–867. doi:10.1038/nrg2452

Rönnegård L, Besnier F, Carlborg Ö (2008) An improved method for quantitative trait loci detection of within-line segregation in F2 intercross designs. Genetics 178:2315–2326

Rönnegård L, Besnier F, Carlborg Ö (2009) Modelling dominance in a flexible intercross analysis. BMC Genet 10:30. doi:10.1186/1471-2156-10-30

Sensabaugh GF, Golden VL (1978) Phenotype dependence in the inhibition of red cell acid phosphatase (ACP) by folates. Am J Hum Genet 30(5):553–560

Spencer N, Hopkinson DA, Harris H (1964) Quantitative differences and gene dosage in the human red cell acid phosphatase polymorphism. Nature 201:299–300

Templeton AR (1987) The general relationship between average effect and average excess. Genet Res Camb 49:69–70

Tiwari HK, Elston RC (1997) Deriving components of genetic variance for multilocus models. Genet Epidemiol 14(6):1131–1136

Van der Veen JH (1959) Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. Genetica 30:201–232

Wang T, Zeng ZB (2006) Models and partition of variance for quantitative trait loci with epistasis and linkage disequilibrium. BMC Genet 7:9. doi:10.1186/1471-2156-7-9

Wang T, Zeng ZB (2009) Contribution of genetic effects to genetic variance components with epistasis and linkage disequilibrium. BMC Genet 10:52

Weir BS (1996) Genetic data analysis II. Sinauer, Massachusetts

Wilder JA, Hammer MF (2004) European ACP1*C allele has recessive deleterious effects on early life viability. Hum Biol 76(6):817–835

Wolfram Research Inc (2010) Mathematica edition: version 80. Wolfram Research Inc, Campaign

Yang R-C (2004) Epistasis of quantitative trait loci under different gene action models. Genetics 167(3):1493–1505

Yang R-C, Álvarez-Castro JM (2008) Functional and statistical genetic effects with miltiple alleles. Curr Top Genet 3:49–62

Zeng ZB, Wang T, Zou W (2005) Modeling quantitative trait Loci and interpretation of models. Genetics 169(3):1711–1725