Genet. Res., Camb. (2008), **90**, *pp.* 209–216. © 2008 Cambridge University Press
doi:10.1017/S0016672307009068 Printed in the United Kingdom

Deviations from Hardy–Weinberg proportions for multiple alleles under viability selection

GONZALO ALVAREZ*
Departamento de Genética, Facultad de Biología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain

(*Received 12 September 2007 and in revised form 16 November 2007*)

Summary

Departures of genotype frequencies from Hardy–Weinberg proportions (HWP) for a single autosomal locus due to viability selection in a random mating population have been studied only for the two-allele case. In this article, the analysis of deviations from HWP due to constant viability selection is extended to multiple alleles. The deviations for an autosomal locus with k alleles are measured by means of k f_{ii} fixation indices for homozygotes and $k(k-1)/2$ f_{ij} fixation indices for heterozygotes, and expressions are obtained for these indices (F_{IS} statistics) under the multiallele viability model. Furthermore, expressions for f_{ii} and f_{ij} when the multiallele polymorphism is at stable equilibrium are also derived and it is demonstrated that the pattern of multiallele Hardy–Weinberg deviations at equilibrium is characterized by a global heterozygote excess and a deficiency of each of the homozygotes. This pattern may be useful for detecting whether a given multiallelic polymorphism is at stable equilibrium in the population due to viability selection. An analysis of Hardy–Weinberg deviations from published data for the three-allele polymorphism at the β-globin locus in human populations from West Africa is presented for illustration.

1. Introduction

The departures of genotype frequencies from Hardy–Weinberg proportions (HWP) for a given locus provide relevant information for understanding genetic characteristics of populations, such as deviations from random mating, population subdivision, asymmetric allelic contributions of the sexes, or viability selection. Furthermore, the analysis of deviations from HWP is one of the few ways to identify systematic genotyping errors, so that at present it is a fundamental tool for genotyping quality control in large-scale studies of molecular markers (Hare *et al*., 1996; Gomes *et al*., 1999; Xu *et al*., 2002; Hosking *et al*., 2004; Chen *et al*., 2005; Zou & Donner, 2006; Teo *et al*., 2007). Furthermore, in the context of studies of association between human diseases and molecular markers, the analysis of deviations from HWP is important for distinguishing those deviations in patients and control samples that could be attributed to the underlying genetic disease model at the susceptibility locus from those due to genotyping errors, chance and/or violations of the assumptions of Hardy–Weinberg equilibrium (Wittke-Thompson *et al*., 2005).

Natural selection operating through differential survival of genotypes is probably one of the most important mechanisms disturbing HWP in random mating populations, particularly when genotypes are recorded at the adult stage of the life cycle. Departures from HWP for a single autosomal locus produced by viability selection have been investigated for the two-allele case, especially as regards statistical tests for detecting natural selection (Lewontin & Cockerham, 1959; Li, 1959; Workman, 1969; Brown, 1970; Hedrick, 2005, pp. 150–152). However, the study of deviations from HWP caused by viability selection acting on multiple alleles has received very little attention. In contrast, analysis of deviations from HWP for multiple alleles has been performed for

* Tel.: +34 981563100, ext 13261. Fax: +34 981596904. e-mail: bfjurado@usc.es

models of subdivided populations (Nei, 1965; Li, 1969), inbreeding (Li & Horvitz, 1953; Yasuda, 1968; Curie-Cohen, 1982; Robertson & Hill, 1984; Hill *et al.*, 1995; Rousset & Raymond, 1995) and differential selection between the sexes (Purser, 1966; Ziehe & Gregorius, 1981). Consequently, viability selection is the only basic model of deviations from HWP for which the multiple-allele case has not been investigated. This is rather striking given that the classical model of multiallele viability selection has been extensively studied, in particular with respect to conditions for the stability of multiallelic polymorphisms (Mandel, 1959, 1970; Weir, 1970; Lewontin *et al.*, 1978; Karlin, 1981; Karlin & Feldman, 1981). On theoretical grounds, multiallelic polymorphisms are expected to be easily maintained in natural populations by viability selection since, although the proportion of the viability parameter space permitting stable polymorphisms becomes extremely small as the number of alleles increases (Lewontin *et al.*, 1978; Karlin, 1981; Karlin & Feldman, 1981), models based on Monte Carlo simulations in which a series of new mutations are introduced into the population show that viability selection is capable of maintaining a large number of alleles (up to 38 in some cases) (Spencer & Marks, 1988, 1992; Marks & Spencer, 1991). In the present article, expressions for departures of genotype frequencies from HWP, as measured by means of fixation indices (F_{IS} statistics), are obtained for an autosomal locus with multiple alleles under a deterministic model of constant viability selection and random mating. Special attention is devoted to characterizing the multiallelic pattern of deviations from HWP exhibited by the population when it attains a stable equilibrium due to viability selection.

2. Hardy–Weinberg deviations under the multiallele viability model

(i) *Model and notation*

An autosomal locus with k alleles (denoted as $A_1, A_2, \ldots, A_k$) is considered, where p_i is the frequency of the A_i allele at the zygotic stage. Assuming random mating, the frequency of the A_iA_i homozygote is p_i^2 and the frequency of the A_iA_j heterozygote is $2p_ip_j$. Under the standard one-locus multiallele viability selection model, with fitness values w_{ii} for the A_iA_i homozygote and w_{ij} for the A_iA_j heterozygote, the adult frequencies for the A_iA_i and A_iA_j genotypes, A_{ii} and A_{ij} respectively, are

$$A_{ii}=p_i^2w_{ii}/W$$
$$A_{ij}=2p_ip_jw_{ij}/W \qquad (1)$$

and the allele frequencies in adults are

$$p_i'=p_iw_i/W$$
$$p_j'=p_jw_j/W \qquad (2)$$

where w_i is the marginal fitness of allele A_i and W is the mean fitness of the population, given by

$$w_i=p_iw_{ii}+\sum_{j\neq i}p_jw_{ij}=\sum_j p_jw_{ij}$$

$$W=\sum_i p_i^2w_{ii}+\sum_{i<j}2p_ip_jw_{ij}=\sum_i p_iw_i=\sum_i\sum_j p_ip_jw_{ij}. \qquad (3)$$

The departures of adult genotype frequencies from HWP for multiple alleles can be expressed in terms of either k f_{ii} fixation indices (F_{IS} statistics) or, alternatively, $k(k-1)/2$ f_{ij} fixation indices, as

$$f_{ii}=\frac{A_{ii}-p_i'^2}{p_i'(1-p_i')} \qquad (4)$$

$$f_{ij}=1-\frac{A_{ij}}{2p_i'p_j'} \qquad (5)$$

taking into account that f_{ii} and f_{ij} are functionally related by

$$f_{ii}=\frac{\sum_{j\neq i}p_j'f_{ij}}{(1-p_i')} \qquad (6)$$

(Weir, 1996, p. 94). In this formulation, the f_{ii} coefficients can be considered as allele-specific F_{IS} statistics (Chakraborty & Danker-Hopfe, 1991).

(ii) *Hardy–Weinberg deviations*

Expressions for the f_{ii} and f_{ij} fixation indices under the model of viability selection are obtained by substituting the allele and genotype frequencies in (4) and (5) by their values given by (1) and (2), and they are

$$f_{ii}=\frac{p_i(w_{ii}W-w_i^2)}{w_i(W-p_iw_i)} \qquad (7)$$

$$f_{ij}=1-\frac{w_{ij}W}{w_iw_j}. \qquad (8)$$

In these expressions for the deviations from HWP for multiple alleles, the terms $(w_{ii}W-w_i^2)$ for the homozygote A_iA_i and $(w_iw_j-w_{ij}W)$ for the heterozygote A_iA_j determine the sign of the deviation. Thus, when the fitness of a particular genotype multiplied by the mean fitness is equal to the product of the marginal fitnesses of the alleles forming that genotype, a deviation from HWP is not expected to occur for that genotype. This is the case for multiplicative or

https://doi.org/10.1017/S0016672307009068 Published online by Cambridge University Press

geometric fitnesses where $w_{ii}=a_i^2$ and $w_{ij}=a_ia_j$ since, in this case, marginal and mean fitnesses take the form

$$w_i=a_i\sum_j p_j'a_j$$

$$w_j=a_j\sum_i p_i'a_i$$

$$W=\sum_i\sum_j p_i'p_j'a_ia_j$$

and substituting these expressions in (7) and (8), we have

$$f_{ii}=f_{ij}=0.$$

Therefore, under multiplicative viability fitnesses, the genotype frequencies at a multiallelic locus, after the operation of selection, are in accordance with Hardy–Weinberg expectations. This result was demonstrated for a two-allele locus by Lewontin & Cockerham (1959) and extended to the three-allele case by Li (1959), and here is generalized for a k-allele system.

When the genotype fitnesses do not follow a geometric progression, the pattern of Hardy–Weinberg deviations is difficult to specify since f_{ii} and f_{ij}, as expressed by (7) and (8), are dependent on the marginal and mean fitness which are changing along generations. However, a particular and relatively simple pattern of Hardy–Weinberg departures is expected to occur for a multiallelic locus when a stable equilibrium is attained in the population by the operation of viability selection. At the equilibrium, f_{ii} and f_{ij}, as expressed by (7) and (8), reduce to

$$f_{ii}^*=\frac{p_i^*}{(1-p_i^*)}\left(\frac{w_{ii}}{W^*}-1\right) \tag{9}$$

$$f_{ij}^*=1-\frac{w_{ij}}{W^*} \tag{10}$$

where * denotes equilibrium values, since the condition for equilibrium in the multiallele viability model is simply $w_i=w_j=\ldots\ldots=W$ (Lewontin *et al.*, 1978). Consequently, the departure from HWP for a given genotype is basically determined, at the stable equilibrium, by the ratio of the genotype fitness to the mean fitness of the population. Given that the homozygote and heterozygote finesses must satisfy two inequalities with respect to the mean fitness of the population as necessary conditions for the existence of a stable multiallele polymorphism, which are $W^*>w_{ii}$ for all $i=1, 2, \ldots, k$ and $W^*<\tilde{w}_{ij}^*$, where $\tilde{w}_{ij}^*$ is the weighted mean fitness of heterozygotes at the equilibrium (Mandel, 1959; Ginzburg, 1979), it follows that all homozygotes must present a deficiency with respect to HWP, that is

$$f_{ii}^*<0 \qquad (\text{all } i=1, 2, \ldots, \text{k})$$

and an excess must be present in many but not necessarily all heterozygote classes. In the three-allele case, for example, it has been shown that at most one heterozygous viability may fall below that of at most two homozygotes (Mandel, 1959) and therefore, in this case, the f_{ij}^* corresponding to that particular heterozygote will be positive.

For a two-allele locus, the expression for departures from HWP under viability selection, obtained by Workman (1969) and Brown (1970), is a particular case of expressions (7) and (8). At equilibrium, the Hardy–Weinberg deviation for a diallelic locus as given by Workman (1969) is a particular case of expressions (9) and (10).

(iii) *Estimation of* F_{IS} *statistics*

The model for statistical estimation of deviations from HWP for multiple alleles under viability selection is a model where either k f_{ii} parameters, or alternatively $k(k-1)/2$ f_{ij} parameters, must be independently estimated, in addition to the allele frequencies. At first sight, this model is more complicated than the model for the estimation of the inbreeding coefficient under regular inbreeding, in which only one f value needs to be estimated in addition to the allelic frequencies, and which has been extensively studied (Li & Horvitz, 1953; Curie-Cohen, 1982; Robertson & Hill, 1984; Hill *et al.*, 1995; Rousset & Raymond, 1995). However, in the framework of maximum likelihood theory, the estimation of both the set of parameters f and the allele frequencies is straightforward. Consider a random sample of n adults in which the observed numbers of A_iA_i and A_iA_j genotypes are n_{ii} and n_{ij}, respectively, and the observed allele frequency of A_i is p_i'. The likelihood of a sample of n individuals composed of n_{ii} genotypes A_iA_i and n_{ij} genotypes A_iA_j can be expressed in terms of a set F of $k(k-1)/2$ f_{ij} parameters and a set P of k parameters of allele frequencies as

$$L_{(P,F)}=\frac{n!}{n_{11}!n_{12}!\ldots\ldots n_{kk}!}\left(p_1'^2+p_1'\sum_{j\neq 1}p_j'f_{1j}\right)^{n_{11}}\times(2p_1'p_2'(1-f_{12}))^{n_{12}}\ldots\ldots\left(p_k'^2+p_k'\sum_{j\neq k}p_j'f_{kj}\right)^{n_{kk}}.$$

Under this formulation the f_{ii} parameters are not taken into consideration and therefore the number of independent parameters to be estimated equals the number of degrees of freedom in the data, so that Bailey's method (Bailey, 1951; Weir, 1996, pp. 63–66)

https://doi.org/10.1017/S0016672307009068 Published online by Cambridge University Press

Table 1. *Genotype distribution, allele frequencies and Hardy–Weinberg deviations at the β-globin locus in samples from West African populations*

POPULATION Age group	n	AA	AS	AC	SS	SC	CC	$\hat{p}_A$	$\hat{p}_S$	$\hat{p}_C$	$\hat{f}_T$	$\hat{f}_{AA}$	$\hat{f}_{SS}$	$\hat{f}_{CC}$
JOLA[1]														
Adults	312	252	53	7	0	0	0	0·904	0·085	0·011	−0·052	−0·105*	−0·091	−0·010
Infants[a]	104	86	15	2	1	0	0	0·909	0·082	0·010	0·019	0·020	0·044	−0·005
FULA[1]														
Adults	127	100	24	3	0	0	0	0·894	0·094	0·012	−0·056	−0·115	−0·101	−0·008
Infants[a]	69	55	12	1	1	0	0	0·891	0·101	0·007	0·026	0·035	0·054	0·000
YORUBA[2]														
Adults	259	168	64	25	0	2	0	0·820	0·127	0·052	−0·103**	−0·165**	−0·144*	−0·053
Children[b]	202	142	49	9	0	2	0	0·847	0·126	0·027	−0·083*	−0·103	−0·142*	−0·026
Children[c]	178	128	44	5	1	0	0	0·857	0·129	0·014	−0·056	−0·119	−0·096	−0·012
Children[d]	267	204	50	8	4	1	0	0·873	0·110	0·017	0·008	0·025	0·030	−0·015
MOSSI[3]														
>6 years	3513	2333	335	763	1	23	58	0·820	0·051	0·128	−0·028**	−0·060***	−0·048**	0·0004

[1] Allison (1956); [2] Roberts & Boyo (1962); [3] Modiano *et al*. (2001).
[a] 2 months to 1 year; [b] 6 years to 12 years; [c] 21 months to 6 years; [d] 4 months to 28 months.
*$P<0·05$; ** $P<0·01$; *** $P<0·001$.

can be applied. Consequently, the maximum likelihood estimates of the parameters are simply their observed values and, therefore, both the f_{ij} obtained from (5) and the allele frequencies computed by gene counting are maximum likelihood estimates. With regard to the f_{ii} fixation indices, their estimates from (4) are also maximum likelihood estimates since each particular f_{ii} corresponds to the f estimate that results from grouping all the alleles into two categories, i versus non-i, and for a diallelic system both $f_{11}=f_{12}=f_{22}=f$ and the allele frequency are maximum likelihood estimates (Li & Horvitz, 1953; Weir, 1996, pp. 64–65). In this way, a k-allele system can be split into k diallelic systems each leading to maximum likelihood estimates of both f_{ii} and p_i'. Note that all this estimation procedure is valid not only for the particular case of deviations from HWP produced by viability selection, but for any case where each specific genotype has a specific departure from HWP, as for example population subdivision or different allelic frequencies between the sexes.

3. Hardy–Weinberg deviations for the β-globin locus in human populations from West Africa

The β-globin gene is one of the most thoroughly studied polymorphisms in man, since it is an adaptive polymorphism involved in resistance against *Plasmodium falciparum* malaria (Cavalli-Sforza & Bodmer, 1971; Vogel & Motulsky, 1997). An analysis of multiallelic deviations from HWP for this locus in human populations has been performed in the present study using published data (Allison, 1956; Roberts & Boyo, 1962; Modiano *et al*., 2001). The populations considered belong to the geographical area of West Africa where this locus presents three alleles with detectable frequencies: the *HbA* allele that gives rise to the normal haemoglobin, the *HbS* allele responsible for the sickle haemoglobin, and the *HbC* responsible for haemoglobin *C*. Samples of adults and infants from the Jola and Fula populations (The Gambia), and of adults and children from the Yoruba population (Nigeria), were analysed. A very large sample ($n=3513$) from the Mossi population (Burkina Faso) was also included in the analysis: this is a control sample from a large case–control study performed in Burkina Faso to investigate the protective role against severe malaria of genotypes at the β-globin locus, and was composed mainly of healthy subjects more than 6 years old (87% children aged 6–15 years, and 8·4% individuals more than 15 years old), though a small number of children aged 1–5 years (4·6%) was also included (Modiano *et al*., 2001).

Genotype distribution, allele frequencies and deviations from HWP for each of the samples analysed are given in Table 1. Deviations from HWP were measured by means of the f_{ii} estimators of Robertson & Hill (1984), giving estimates for the three homozygous genotypes ($\hat{f}_{AA}$ for the homozygote *HbAA*, $\hat{f}_{SS}$ for *HbSS*, and $\hat{f}_{CC}$ for *HbCC*) and a global estimate of deviation from HWP at the locus ($\hat{f}_T$) obtained from the weighted average of the f_{ii} estimates. The variance of f_{ii} estimates equals $1/n$ for $f_{ii}=0$ and the ratio of the squared estimate to its variance will be approximately distributed as a chi-square variable with one degree of freedom, leading to a two-tailed test of the null hypothesis $H_0: f_{ii}=0$ (Elandt-Johnson, 1971, pp. 355–356; Robertson & Hill, 1984). One-tailed tests can be performed from the ratio of the

https://doi.org/10.1017/S0016672307009068 Published online by Cambridge University Press

estimate to its standard error, which is approximately distributed as a standard normal variable (Elandt-Johnson, 1971, pp. 355–356). The two-tailed test of f_{ii} is equivalent to the Hardy–Weinberg test of a single homozygous genotype recently proposed by Chen *et al*. (2005), since the chi-square statistic given by Chen *et al*. (2005, p. 1440) is simply nf^{2}_{ii}. The analysis of deviation from HWP for multiple alleles by means of f_{ii} fixation indices and/or single genotype tests gives a complete view of the distribution of deviations among particular genotypes at the given locus in contrast to the overall tests such as the chi-square goodness-of-fit test or the exact test (Louis & Dempster, 1987; Guo & Thompson, 1992; Chakraborty & Zhong, 1994; Rousset & Raymond, 1995). A very regular pattern of deviations from HWP for the β-globin locus is observed in the adult samples from West Africa populations. First, a global heterozygote excess is found in all adult samples: $\hat{f}_T$ ranges from $-0{\cdot}103$ to $-0{\cdot}052$ with an average of $-0{\cdot}070 \pm 0{\cdot}016$. This heterozygote excess is statistically significant by one-sided tests in the Yoruba sample. Second, the distribution of Hardy–Weinberg deviations among particular homozygotes is clearly uneven in the adult samples, since homozygotes for the *HbA* and *HbS* alleles show a clear deficiency with respect to HWP, whereas the frequency of the homozygote *HbCC* is very close to Hardy–Weinberg expectations. Specifically, $\hat{f}_{AA}$ ranges from $-0{\cdot}165$ to $-0{\cdot}105$ with a mean value of $-0{\cdot}128 \pm 0{\cdot}019$, these deviations being statistically significant in two of the three adult samples analysed; similarly, f_{SS} ranges from $-0{\cdot}144$ to $-0{\cdot}091$ with a mean value of $-0{\cdot}112 \pm 0{\cdot}016$, these deviations being statistically significant in the Yoruba sample. In contrast, $\hat{f}_{CC}$ ranges from $-0{\cdot}053$ to $-0{\cdot}008$ with a mean value of $-0{\cdot}024 \pm 0{\cdot}015$ and these negative estimates are associated with the absence of *HbCC* homozygotes in the adult samples due to the low frequency of the *HbC* allele (see expression (4)). In addition, these deviations are not statistically significant in any of the three adult samples studied. A substantial number of *HbCC* homozygotes is present in the large sample from the Mossi population which probably represents a partially selected stage since it is composed by individuals older than 6 years and, in this case, $\hat{f}_{CC}$ takes a positive value ($\hat{f}_{CC} = 0{\cdot}0004$). As a whole, these results do not support the idea that this three-allele polymorphism is at stable equilibrium in the West African populations due to viability selection, since stable equilibrium would require all three homozygotes to present a deficiency with respect to Hardy–Weinberg expectations, as already demonstrated. Obviously, a large number of West African populations must be analysed in order to confirm these results but it is interesting to point out that the analysis of multiallelic deviations from HWP presented here is in accordance with recent evidence based on epidemiological and fitness data which suggests that this three-allele system may be a transient polymorphism in West African populations (Modiano *et al*., 2001; Hedrick, 2004, 2005, pp. 161–163).

The analysis of deviations from HWP for infants (2 months to 1 year) and very young children (4–28 months) shows that their genotypic frequencies are very close to Hardy–Weinberg expectation and, thus, the pattern of deviations from HWP observed in these age groups is very different from that found in adult samples (Table 1): mean values for $\hat{f}_T$, $\hat{f}_{AA}$, $\hat{f}_{SS}$ and $\hat{f}_{CC}$ are $0{\cdot}018 \pm 0{\cdot}005$, $0{\cdot}027 \pm 0{\cdot}004$, $0{\cdot}043 \pm 0{\cdot}007$ and $-0{\cdot}007 \pm 0{\cdot}004$, respectively, in the three samples analysed. These results reveal that the heterozygote excess observed in adult samples is not a consequence of asymmetric allelic contributions of the sexes due to differential selection in the two sexes or to chance, since in this case the heterozygote excess would be present at the zygotic stage (see Section 4). On the contrary, our findings indicate that the heterozygote excess observed in adult samples is probably due to the operation of viability selection. In older children (21 months to 6 years, 6–12 years and >6 years), the pattern of deviations from HWP observed is very close to that seen in the adult samples: mean values for $\hat{f}_T$, $\hat{f}_{AA}$, $\hat{f}_{SS}$ and $\hat{f}_{CC}$ in the three older-children samples are $-0{\cdot}056 \pm 0{\cdot}016$, $-0{\cdot}094 \pm 0{\cdot}018$, $-0{\cdot}095 \pm 0{\cdot}027$ and $-0{\cdot}013 \pm 0{\cdot}008$, respectively. This heterozygote excess is statistically significant by one-sided tests for both f_T and f_{SS} in two of the three samples analysed (Yoruba and Mossi) and for f_{AA} in the Mossi sample. This result is consistent with evidence indicating that differential mortality among genotypes at the β-globin locus due to death from either sickle-cell anaemia or malaria occurs mainly in young children (Allison, 1956; Roberts & Boyo, 1960; Cavalli-Sforza & Bodmer, 1971, Greenwood *et al*., 1987; Vogel & Motulsky, 1997).

4. Discussion

The effect of viability selection on the distribution of genotypes for a multiallelic polymorphism in a random mating population is effectively identified through the departures of genotype frequencies from Hardy–Weinberg proportions (HWP) (expressions (7) and (8)). Furthermore, a genetic polymorphism for multiple alleles maintained by balancing viability selection will show, at equilibrium, both a global heterozygote excess and a deficiency of each of the homozygotes (expressions (9) and (10)). This pattern of Hardy–Weinberg deviations is a consequence of the relationship between the genotype fitnesses and the mean fitness when the population reaches stable equilibrium, since, at this point, the mean fitness of the population must be higher than the fitness of each

homozygote and lower than the weighted mean fitness of heterozygotes (Mandel, 1959; Ginzburg, 1979). This 'footprint' of selection on the genotypic distribution may be useful for detecting whether a given multiallele polymorphism is at stable equilibrium in the population due to viability selection, since it can be easily distinguished from other potential causes of deviations from HWP. Inbreeding and subdivision or admixture of populations will give rise to heterozygote deficiency although, under population subdivision for multiple alleles, some particular heterozygote might be in excess due to a positive covariation of allelic frequencies (Nei, 1965; Li, 1969). On the other hand, heterozygote excess can also be caused by differences in allelic frequencies between the sexes. These differences might arise either by chance or by differential selection between the sexes. These two different mechanisms are formally analogous in terms of deviations from HWP, since in both cases the deviation is dependent on the difference in allele frequencies in the two sexes, irrespective of the process generating these differences. Differences in allelic frequencies between sexes may well arise by chance if the number of parents is small, and will cause an excess of heterozygotes in the progeny (Robertson, 1965). Heterozygote excess as a consequence of asymmetric allelic contributions of the sexes due to differential viability or fertility selection in the two sexes has been characterized for the two-allele case (Bundgaard & Christiansen, 1972; Andresen, 1978) and for multiallelic systems (Purser, 1966; Ziehe & Gregorius, 1981). For multiple alleles, differential allelic contributions from each sex lead to a deficiency of each homozygote and an excess of the sum of all heterozygotes. Therefore, differences in allelic frequencies between sexes will give rise to a pattern of Hardy–Weinberg deviations very similar, at first sight, to that produced by balancing viability selection at equilibrium. There is, however, a striking difference between these two patterns of Hardy–Weinberg deviations as regards the specific stage of the life cycle in which they originate. Thus, differences in allelic frequencies between sexes will produce deviations from HWP apparent at the zygotic stage; in contrast, under viability selection genotypic frequencies at the zygotic stage are expected to show HWP as a consequence of random mating, and the deviations generated by the operation of the viability selection will be mainly observed in the adult phase of the life cycle. Moreover, under differential selection between the sexes, the deviations of genotype frequencies from HWP become very small after several generations and a strong affinity of these frequencies for HWP is observed at the equilibrium (Ziehe & Gregorius, 1981). In contrast, under a viability selection model, large deviations from HWP may be seen at the stable equilibrium, at least for those genotypes showing larger departures from the mean fitness of the population (expressions (9) and (10)).

Heterozygote excess has been detected for some multiallelic polymorphisms such as the inversion polymorphism in *Drosophila* (Dobzhansky & Levene, 1948; Ruiz *et al*., 1986) or the polymorphisms of the human β-globin gene (Cavalli-Sforza & Bodmer, 1971, pp. 161–165) and HLA complex (Hedrick, 1990; Markov *et al*., 1993; Chen *et al*., 1999). This excess is thought to be the result of the operation of natural selection. However, analysis of multiallelic deviations from HWP through the estimation of f_{ii} fixation indices has rarely been carried out for such polymorphisms, since until now there was no reference model to interpret the observed patterns of multiallelic deviations generated by selection. Certainly, analysis of the heterozygote excess associated with adaptive polymorphisms in terms of multiallelic deviations may give valuable information on the mechanism of balancing selection responsible for the maintenance of such polymorphisms. As discussed above, the occurrence of a global heterozygote excess associated with a deficiency of each and every one of the homozygotes ($f^{*}_{ii} < 0$, for all $i = 1, 2.....k$) is strong evidence suggesting that a multiallelic polymorphism is at equilibrium due to viability selection. Otherwise, when such a pattern of multiallelic deviations is not seen, either the population is not at equilibrium, or some mechanism of balancing selection other than viability selection must be responsible for maintaining the observed multiallelic polymorphism.

References

Allison, A. C. (1956). The sickle-cell and haemoglobin C genes in some African populations. *Annals of Human Genetics* **21**, 67–89.

Andresen, E. (1978). A note on deviation from Hardy–Weinberg proportions due to differences in gene frequencies between parental males and females. *Animal Blood Groups and Biochemical Genetics* **9**, 55–58.

Bailey, N. T. J. (1951). Testing the solubility of maximum likelihood equations in the routine application of scoring methods. *Biometrics* **7**, 268–274.

Brown, A. H. D. (1970). The estimation of Wright's fixation index from genotypic frequencies. *Genetica* **41**, 399–406.

Bundgaard, J. & Christiansen, F. B. (1972). Dynamics of polymorphisms. I. Selection components in a experimental population of *Drosophila melanogaster*. *Genetics* **71**, 439–460.

Cavalli-Sforza, L. L. & Bodmer, W. F. (1971). *The Genetics of Human Populations*. San Francisco: Freeman.

Chakraborty, R. & Danker-Hopfe, H. (1991). Analysis of population structure: a comparative study of different estimators of Wright's fixation indices. In *Handbook of Statistics* (ed. C. R. Rao & R. Chakraborty), vol. 8, pp. 203–254. Amsterdam: Elsevier.

Chakraborty, R. & Zhong, Y. (1994). Statistical power of an exact test of Hardy–Weinberg proportions of genotypic data at a multiallelic locus. *Human Heredity* **44**, 1–9.

Chen, J. J., Hollenbach, J. A., Trachtenberg, E. A., Just, J. J., Carrington, M., Rønningen, K. S., Begovich, A., King, M. C., McWeeney, S., Mack, S. J., Erlich, H. A. & Thomson, G. (1999). Hardy–Weinberg testing for HLA class II (DRB1, DQA1, DQB1, and DPB19 loci in 26 human ethnic groups. *Tissue Antigens* **54**, 533–542.

Chen, J. J., Duan, T., Single, R., Mather, K. & Thomson, G. (2005). Hardy–Weinberg testing of a single homozygous genotype. *Genetics* **170**, 1439–1442.

Curie-Cohen, M. (1982). Estimates of inbreeding in a natural population: a comparison of sampling properties. *Genetics* **100**, 339–358.

Dobzhansky, T. & Levene, H. (1948). Genetics of natural populations. XVII. Proof of operation of natural selection in wild populations of *Drosophila pseudoobscura*. *Genetics* **33**, 537–547.

Elandt-Johnson, R. C. (1971). *Probability Models and Statistical Methods in Genetics*. New York: Wiley.

Ginzburg, L. R. (1979). Why are heterozygotes often superior in fitness? *Theoretical Population Biology* **15**, 264–267.

Gomes, I., Collins, A., Lonjou, C., Thomas, N. S., Wilkinson, J., Watson, M. & Morton, N. (1999). Hardy–Weinberg quality control. *Annals of Human Genetics* **63**, 535–538.

Greenwood, B. M., Bradley, A. K., Greenwood, A. M., Byass, P., Jammeh, K., Marsh, K., Tulloch, S., Oldfield, F. S. J. & Hayes, R. (1987). Mortality and morbidity from malaria among children in a rural area of The Gambia, West Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **81**, 478–486.

Guo, S. W. & Thompson, E. A. (1992). Performing the exact test of Hardy–Weinberg proportions for multiple alleles. *Biometrics* **48**, 361–372.

Hare, M. P., Karl, S. A. & Avise, J. C. (1996). Anonymous nuclear DNA markers in the American oyster and their implications for the heterozygote deficiency phenomenon in marine bivalves. *Molecular Biology & Evolution* **13**, 334–345.

Hedrick, P. (1990). Evolution at HLA: possible explanations for the deficiency of homozygotes in two populations. *Human Heredity* **40**, 213–220.

Hedrick, P. (2004). Estimation of relative fitnesses from relative risk data and the predicted future of haemoglobin alleles S and C. *Journal of Evolutionary Biology* **17**, 221–224.

Hedrick, P. (2005). *Genetics of Populations*, 3rd edn. Sudbury, MA: Jones and Bartlett.

Hill, W. G., Babiker, H. A., Ranford-Cartwright, L. C. & Walliker, D. (1995). Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. *Genetical Research* **65**, 53–61.

Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., Riley, J., Purvis, I. & Xu, C. F. (2004). Detection of genotyping errors by Hardy–Weinberg equilibrium testing. *European Journal of Human Genetics* **12**, 395–399.

Karlin, S. (1981). Some natural viability systems for a multiallelic locus: a theoretical study. *Genetics* **97**, 457–473.

Karlin, S. & Feldman, M. W. (1981). A theoretical and numerical assessment of genetic variability. *Genetics* **97**, 475–493.

Lewontin, R. C. & Cockerham, C. C. (1959). The goodness-of-fit test for detecting natural selection in random mating populations. *Evolution* **13**, 561–564.

Lewontin, R. C., Ginzburg, L. R. & Tuljapurkar, S. D. (1978). Heterosis as an explanation for large amounts of genic polymorphism. *Genetics* **88**, 149–170.

Li, C. C. (1959). Notes on relative fitness of genotypes that forms a geometric progression. *Evolution* **13**, 564–567.

Li, C. C. (1969). Population subdivision with respect to multiple alleles. *Annals of Human Genetics* **33**, 23–29.

Li, C. C. & Horvitz, D. G. (1953). Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics* **5**, 107–117.

Louis, E. J. & Dempster, E. R. (1987). An exact test for Hardy–Weinberg and multiple alleles. *Biometrics* **43**, 805–811.

Mandel, S. P. H. (1959). The stability of a multiple allelic system. *Heredity* **13**, 289–302.

Mandel, S. P. H. (1970). The equivalence of different sets of stability conditions for multiple allelic systems. *Biometrics* **26**, 840–845.

Markov, T., Hedrick, P. W., Zuerlein, K., Danilovs, J., Martin, J., Vyvial, T. & Armstrong, C. (1993). HLA polymorphism in the Havasupai: evidence for balancing selection. *American Journal of Human Genetics* **53**, 943–952.

Marks, R. W. & Spencer, H. G. (1991). The maintenance of single-locus polymorphism. II. The evolution of fitnesses and allele frequencies. *American Naturalist* **138**, 1354–1371.

Modiano, D., Luoni, G., Sirima, B. S., Simporé, J., Verra, F., Konaté, A., Rastrelli, E., Olivieri, A., Calissano, C., Paganotti, G. M., D'Urbano, L., Sanou, I., Sawadogo, A., Modiano, G. & Coluzzi, M. (2001). Haemoglobin C protects against clinical *Plasmodium falciparum* malaria. *Nature* **414**, 305–308.

Nei, M. (1965). Variation and covariation of gene frequencies in subdivided populations. *Evolution* **19**, 256–258.

Purser, A. F. (1966). Increase in heterozygote frequency with differential fertility. *Heredity* **21**, 322–327.

Roberts, D. F. & Boyo, A. E. (1960). On the stability of haemoglobin gene frequencies in West Africa. *Annals of Human Genetics* **24**, 375–387.

Roberts, D. F. & Boyo, A. E. (1962). Abnormal haemoglobins in childhood among the Yoruba. *Human Biology* **34**, 20–37.

Robertson, A. (1965). The interpretation of genotypic ratios in domestic animal populations. *Animal Production* **7**, 319–324.

Robertson, A. & Hill, W. G. (1984). Deviations from Hardy–Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics* **107**, 703–718.

Rousset, F. & Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics* **140**, 1413–1419.

Ruiz, A., Fontdevila, A., Santos, M., Seoane, M. & Torroja, E. (1986). The evolutionary history of *Drosophila buzzatii*. VIII. Evidence for endocyclic selection acting on the inversion polymorphism in a natural population. *Evolution* **40**, 740–755.

Spencer, H. G. & Marks, R. W. (1988). The maintenance of single-locus polymorphism. I. Numerical studies of a viability selection model. *Genetics* **120**, 605–613.

Spencer, H. G. & Marks, R. W. (1992). The maintenance of single-locus polymorphism. IV. Models with mutation from existing alleles. *Genetics* **130**, 211–221.

Teo, Y. Y., Fry, A. E., Clark, T. G., Tai, E. S. & Seielstad, M. (2007). On the usage of HWE for identifying genotyping errors. *Annals of Human Genetics* **71**, 701–703.

Vogel, F. & Motulsky, A. G. (1997). *Human Genetics*. Berlin: Springer.
Wittke-Thompson, J. K., Pluzhnikov, A. & Cox, N. J. (2005). Rational inferences about departures from Hardy–Weinberg equilibrium. *American Journal of Human Genetics* **76**, 967–986.
Weir, B. S. (1970). Equilibria under inbreeding and selection. *Genetics* **65**, 371–378.
Weir, B. S. (1996). *Genetic Data Analysis II*. Massachusetts: Sinauer.
Workman, P. L. (1969). The analysis of simple genetic polymorphisms. *Human Biology* **41**, 97–114.
Xu, J., Turner, A., Little, J., Bleecker, E. R. & Meyers, D. (2002). Positive results in association studies are associated with departure from Hardy–Weinberg equilibrium: hint for genotyping error? *Human Genetics* **111**, 573–574.
Yasuda, N. (1968). Estimation of the inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring. *Biometrics* **24**, 915–935.
Ziehe, M. & Gregorius, H. (1981). Deviations of genotypic structures from Hardy–Weinberg proportions under random mating and differential selection between the sexes. *Genetics* **98**, 215–230.
Zou, G. Y. & Donner, A. (2006). The merits of testing Hardy–Weinberg equilibrium in the analysis of unmatched case-control data: a cautionary note. *Annals of Human Genetics* **70**, 923–933.