DOI: 10.1534/genetics.108.093518

A Fundamental Relationship Between Genotype Frequencies and Fitnesses

Joseph Lachance[1]

*Graduate Program in Genetics, Department of Ecology and Evolution, State University of New York,
Stony Brook, New York 11794-5222*

Manuscript received July 3, 2008
Accepted for publication August 7, 2008

ABSTRACT

The set of possible postselection genotype frequencies in an infinite, randomly mating population is found. Geometric mean heterozygote frequency divided by geometric mean homozygote frequency equals two times the geometric mean heterozygote fitness divided by geometric mean homozygote fitness. The ratio of genotype frequencies provides a measure of genetic variation that is independent of allele frequencies. When this ratio does not equal two, either selection or population structure is present. Within-population HapMap data show population-specific patterns, while pooled data show an excess of homozygotes.

WHAT patterns of genetic variation are possible within a population, and how does natural selection affect these patterns? R. A. Fisher remarked "it is often convenient to consider a natural population not so much as an aggregate of living individuals but as an aggregate of gene ratios" (Fisher 1953, p. 515). This mathematical abstraction allows key questions in evolutionary genetics to be addressed. A population of diploid individuals can be characterized by a set of genotype frequencies (P_{AA}, P_{AB}, P_{BB}, etc.). This population genetic state is represented by a point in genotype frequency space, where each dimension corresponds to the frequency of a particular genotype. As genotype frequencies change over time, evolving populations explore genotype frequency space (Rice 2004).

However, not every possibility can be realized. Populations are constrained to a restricted set of genotype frequencies. Trivially, genotype frequencies must sum to one. Mendelian segregation and patterns of mating further restrict the set of possible genotype frequencies. For example, in a randomly mating population it is unlikely that every individual will be the same heterozygous genotype. Natural selection also influences patterns of genetic variation, as high-fitness genotypes are found at higher frequencies than neutral expectations. What genotype frequencies can one expect to find, and how does genotype-specific fitness influence this? Any equation summarizing the set of all possible population genetic states must contain frequency and fitness terms for every genotype. Subsequently, genotype frequency data can be used to infer a ratio of genotypic fitnesses. While mathematical descriptions exist for loci with two segregating alleles (Cannings and Edwards 1968), such formulations are lacking for arbitrary numbers of segregating alleles. Here, a general equation describing the set of possible postselection genotype frequencies is derived. Much like how the Hardy–Weinberg principle describes population genetic states in the absence of selection, this novel equation describes population genetic states in the presence of selection. In the context of genotype-frequency space, this is a multidimensional surface, the curvature of which is influenced by natural selection (Figure 1). Evolution involves adaptive walks toward regions of high mean fitness on this surface (Wright 1932; Ewens 1989; Edwards 2000). The set of possible genotype frequencies is analogous to the ecological concept of a fundamental niche (Hutchinson 1957) and the Ramachandran diagrams of biochemistry (Ramachandran *et al.* 1963). The former describes the full range of environmental conditions under which an organism can exist, while the latter describes the possible conformations of dihedral angles for a polypeptide. In each case, valid regions of parameter space are described.

MODEL

A standard single-locus model of theoretical population genetics is considered (diploidy, autosomal inheritance, random mating, and infinite population size). Fitnesses are assumed to be constant and frequency independent. If there are n segregating alleles at a single locus, $n(n+1)/2$ different genotypes are possible, of which n are homozygous and $n(n-1)/2$ are heterozygous. Thus, genotype-frequency space spans $n(n+1)/2$ dimensions. Under random mating, each point in allele-frequency space maps to a single point in genotype-frequency space. Consequently, the surface of possible genotype frequencies is $n-1$ dimensional. The recursion equations of classical population genetics give genotype frequency in the present generation (P_{ij}) as a

[1] *Author e-mail:* joseph.lachance@sunysb.edu

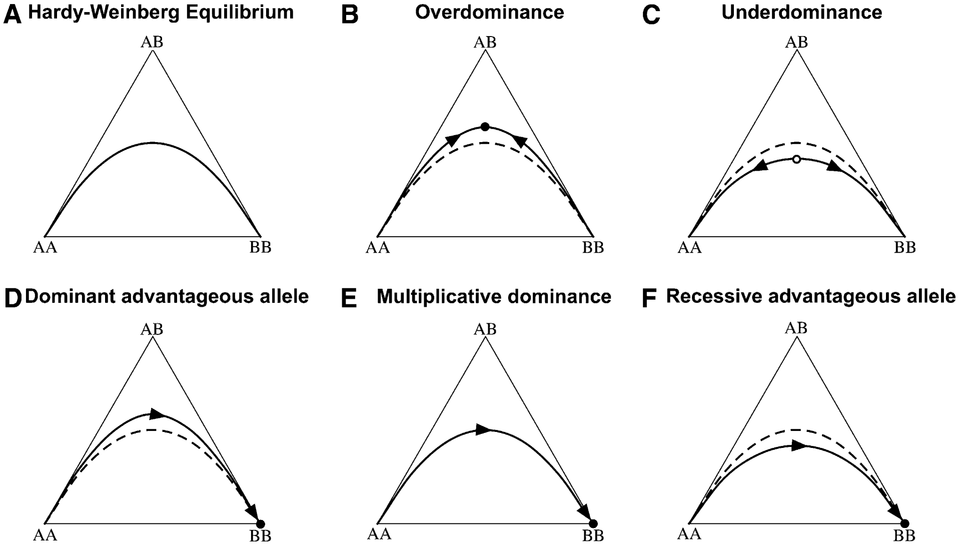
A Hardy-Weinberg Equilibrium
B Overdominance
C Underdominance
D Dominant advantageous allele
E Multiplicative dominance
F Recessive advantageous allele
AB
AA
BB

FIGURE 1.—De Finetti diagrams describing the set of possible genotype frequencies for two segregating alleles. The solid line represents genotype frequencies that satisfy the equation $P^*_{ij}w^*_{ii} = 2P^*_{ii}w^*_{ij}$. Stable equilibria are solid circles, and unstable equilibria are open circles. Hardy–Weinberg proportions are denoted by a dashed line. (A) Neutrality ($\Phi = 2$). (B) Overdominance ($\Phi > 2$). (C) Underdominance ($\Phi < 2$). (D) Directional selection of a dominant advantageous allele ($\Phi > 2$). (E) Directional selection with multiplicative dominance ($\Phi > 2$). (F) Directional selection of a recessive advantageous allele ($\Phi < 2$).

function of genotype fitness (w_{ij}) and allele frequencies in the past generation (p_i).

Derivation of genotypic ratio: Subsequent to mating, but prior to selection, genotype frequencies are found in Hardy–Weinberg proportions. Postselection homozygote frequencies are equal to $P_{ii} = p_i^2 w_{ii}/\bar{w}$ while postselection heterozygote frequencies are equal to $P_{ij} = 2p_i p_j w_{ij}/\bar{w}$ (RICE 2004). Mean fitness ($\bar{w}$) equals the weighted sum of all genotype fitnesses. It is useful to algebraically manipulate these recursion equations so that a ratio of genotype frequency to genotype fitness is on the left-hand side and a ratio of allele frequencies to mean fitness is on the right-hand side. Subsequently, terms for multiple genotypes can be multiplied.

A natural division of genotypes involves homozygotes and heterozygotes. Every allele has a corresponding homozygous genotype, and the product of all homozygote ratios is

$$\frac{\prod_{i=1}^{n} P_{ii}}{\prod_{i=1}^{n} w_{ii}} = \frac{\prod_{i=1}^{n} p_i^2}{\bar{w}^n}. \qquad (1)$$

Since all terms in the above equation are positive, each side of Equation 1 can be raised to the $(n(n-1)/2)$th power:

$$\frac{(\prod_{i=1}^{n} P_{ii})^{n(n-1)/2}}{(\prod_{i=1}^{n} w_{ii})^{n(n-1)/2}} = \frac{\prod_{i=1}^{n} p_i^{n(n-1)}}{\bar{w}^{(n(n(n-1)/2))}}. \qquad (2)$$

Every allele also can be found in heterozygous genotypes, and the product of all heterozygote ratios is

$$\frac{\prod_{i=1,j>i}^{n} P_{ij}}{\prod_{i=1,j>i}^{n} w_{ij}} = 2^{(n(n-1)/2)}\frac{\prod_{i=1}^{n} p_i^{n-1}}{\bar{w}^{n(n-1)/2}}. \qquad (3)$$

Moving the constant term to the left-hand side and raising every term of Equation 3 to the nth power,

$$2^{-(n(n(n-1)/2))}\frac{(\prod_{i=1,j>i}^{n} P_{ij})^n}{(\prod_{i=1,j>i}^{n} w_{ij})^n} = \frac{\prod_{i=1}^{n} p_i^{n(n-1)}}{\bar{w}^{(n(n(n-1)/2))}}. \qquad (4)$$

Note that the right-hand sides of Equations 2 and 4 are identical. Further algebraic manipulation and the transitive property of equality (where $A = B$ and $B = C$ imply $A = C$) allow a single equation containing every genotypic term to be derived:

$$\frac{(\prod_{i=1,j>i}^{n} P_{ij})^n}{(\prod_{i=1}^{n} P_{ii})^{n(n-1)/2}} = 2^{(n(n(n-1)/2))}\frac{(\prod_{i=1,j>i}^{n} w_{ij})^n}{(\prod_{i=1}^{n} w_{ii})^{n(n-1)/2}}. \qquad (5)$$

Since every term in the above equation is positive, Equation 5 can be simplified by taking the $n(n(n-1)/2)$th root of both sides of the equation. This root is the product of the number of homozygote and heterozygote states:

$$\frac{\sqrt[(n(n-1)/2)]{\prod_{i=1,j>i}^{n} P_{ij}}}{\sqrt[n]{\prod_{i=1}^{n} P_{ii}}} = 2\frac{\sqrt[(n(n-1)/2)]{\prod_{i=1,j>i}^{n} w_{ij}}}{\sqrt[n]{\prod_{i=1}^{n} w_{ii}}}. \qquad (6)$$

Note that the geometric mean of n numbers is the nth root of their product. In the absence of assortative mating, patterns of genetic variation reduce to a surprisingly elementary equation. The geometric mean heterozygote frequency divided by the geometric mean homozygote frequency equals two times the geometric mean heterozygote fitness divided by the geometric mean homozygote fitness. Denoting geometric means with asterisks,

$$\frac{P^*_{ij}}{P^*_{ii}} = 2\frac{w^*_{ij}}{w^*_{ii}}. \qquad (7)$$

Description of the genotypic ratio: The above genotypic ratio equation is marked by multiple axes of

TABLE 1

MATLAB simulations confirm analytic theory

Selection	Alleles	Population size	Expected Φ	Observed Φ	Observed f
Overdominant	2	100,000	2.2	2.1981 (2.1×10^{-4})	−0.0472 (1.1×10^{-5})
Underdominant	2	100,000	1.8	1.8011 (1.5×10^{-4})	0.0508 (2.1×10^{-5})
Neutral	2	100,000	2	2.0005 (1.4×10^{-4})	−0.0001 (8.9×10^{-6})
Stochastic fitness	2	100,000	>2	2.0318 (6.9×10^{-2})	−0.0037 (4.3×10^{-3})
Stochastic fitness	3	100,000	>2	2.0010 (2.8×10^{-2})	0.0013 (7.8×10^{-4})
Stochastic fitness	4	100,000	<2	1.9898 (1.5×10^{-2})	0.0016 (3.0×10^{-4})
Directional	2	1,000	2.0976	2.1639 (6.8×10^{-2})	−0.0195 (8.9×10^{-4})
Directional	2	10,000	2.0976	2.0989 (6.3×10^{-3})	−0.0149 (1.3×10^{-4})
Directional	2	100,000	2.0976	2.1013 (6.2×10^{-4})	−0.0156 (2.7×10^{-5})
Directional	3	100,000	2.0646	2.0681 (1.3×10^{-3})	−0.0115 (1.2×10^{-5})
Directional	4	100,000	2.0482	2.0467 (2.6×10^{-3})	−0.0097 (9.4×10^{-6})

Simulations were run for 100 generations and mean and variance of Φ were computed (with variance in observed Φ within parentheses). All alleles were equally frequent at the start of each simulation run. Fitnesses are as follows: overdominant selection ($w_{ij} = 1.1$, $w_{ii} = 1.0$), underdominant selection ($w_{ij} = 0.9$, $w_{ii} = 1.0$), stochastic fitness (fitnesses for each genotype were generated each generation from a Gaussian distribution with a mean of 1.0 and a standard deviation of 0.1), and directional selection (homozygotes and heterozygotes containing a favored dominant allele have a fitness of 1.1, while all other genotypes have a fitness of 1.0).

symmetry: frequencies are on the left-hand side while fitnesses are on the right-hand side, and heterozygous terms are found in numerators while homozygous terms are found in denominators. Genotype frequencies satisfy the above equation after a single generation of random mating and viability selection. As expected, postselection genotype frequencies show increased heterozygosity when heterozygote fitnesses are large relative to homozygote fitnesses. The right-hand side of Equation 7 involves a ratio of fitnesses, indicating that relative, rather than absolute, fitnesses determine genotype frequencies. Under conditions of neutrality Equation 7 reduces to Hardy–Weinberg proportions. However, these proportions also arise when fitnesses are multiplicative (Lewontin and Cockerham 1959). By extension, one can expect to find the same ratio of genotypic frequencies as Hardy–Weinberg when the assumptions of this model are met and $w^*_{ij} = w^*_{ii}$. Regardless of selection coefficients, heterozygote frequencies are maximized at intermediate allele frequencies. The constant 2 on the right-hand side of Equation 7 is due to diploidy and equivalence between *ij* and *ji* heterozygotes. Singularities in the above equation are nonproblematic, as any genotype with zero fitness must also have a postselection frequency of zero. Equation 7 holds for both equilibrium and nonequilibrium populations. Genotype frequencies of natural populations are much easier to obtain than genotype-specific fitnesses. Consequently, Equation 7 allows one to infer the ratio of genotype fitnesses from genotype-frequency data (so long as population size is large and mating is random).

Fitness dominance influences the relative proportions of heterozygotes and homozygotes. The ratio of geometric mean heterozygote frequency to geometric mean homozygote frequency (*i.e.*, the left-hand side of Equation 7) is denoted by Φ:

$$\Phi = \frac{P^*_{ij}}{P^*_{ii}}. \qquad (8)$$

$\Phi < 2$ indicates an excess of homozygotes relative to neutral expectations, and $\Phi > 2$ indicates an excess of heterozygotes. When fitnesses are multiplicative (*i.e.*, fitness dominance is absent), $\Phi = 2$. Geometric means are always less than or equal to the arithmetic mean. Therefore, additive fitnesses (*i.e.*, the fitnesses of heterozygotes are equal to the mean of the relevant homozygotes) result in $\Phi > 2$. Concave fitness functions (where fitnesses of heterozygotes are greater than the arithmetic mean of the relevant homozygote fitnesses) yield $\Phi > 2$. Depending on heterozygote fitnesses, convex fitness functions yield $\Phi < 2$, $\Phi = 2$, or $\Phi > 2$. Note that enzyme kinetics of metabolic pathways are associated with concave fitness functions (Hartl *et al.* 1985; Gillespie 1991), and overdominance and underdominance are exaggerated forms of concave and convex fitness functions, respectively. MATLAB simulations (Mathworks 2005) verify the effects of fitness dominance and also show that Φ is independent of allele frequency (see Table 1).

The genotypic ratio equation (Equation 7) also holds for subsets of alleles. In principle, this allows genotype-specific fitness effects to be detected. The ratio of geometric mean heterozygote frequency to geometric mean homozygote frequency for a subset of alleles is denoted $\Phi_{i,j,k\ldots}$ (*i.e.*, for the alleles *A*, *B*, and *C* the genotypic frequency ratio is equal to Φ_{ABC}). For example, if there are three segregating alleles and the genotype *AA* is deleterious relative to all other genotypes, one would expect Φ_{AB}, Φ_{AC}, and Φ_{ABC} to be >2 and Φ_{BC} to be 2. This application can identify nonneutral genotypes of highly polymorphic loci, such as microsatellites or genes encoding blood group anti-

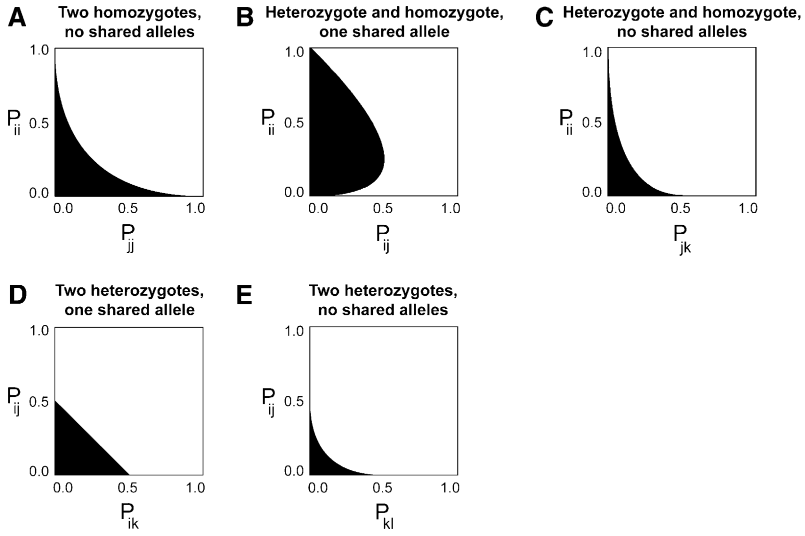

FIGURE 2.—Two-dimensional slices through genotype frequency space. Solid regions indicate possible genotype frequencies for $n \geq 2$ and $w^*_{ij} = w^*_{ii}$. (A) Two homozygotes sharing zero alleles (*e.g.*, *AA* and *BB*). (B) One homozygote and one heterozygote sharing one allele (*e.g.*, *AA* and *AB*). (C) One homozygote and one heterozygote sharing zero alleles (*e.g.*, *AA* and *BC*). (D) Two heterozygotes sharing one allele (*e.g.*, *AB* and *AC*). (E) Two heterozygotes sharing zero alleles (*e.g.*, *AB* and *CD*).

gens. A similar approach has been developed that uses genotype-specific fixation indexes (ALVAREZ 2008). Note, however, that the absolute magnitude of selection-induced departures from Hardy–Weinberg proportions is expected to be small for most sets of genotypic fitnesses (PEREIRA and ROGATKO 1984; HERNÁNDEZ and WEIR 1989). Consequently, sample sizes needed to detect selection would need to be quite large (WEIR 1996).

If the assumption of constant genotypic fitness is relaxed, the magnitude of the genotype-frequency ratio depends on the number of segregating alleles. Consider a stochastic fitness scenario where genotype-specific fitnesses vary from generation to generation and are drawn from the same arbitrary distribution (*i.e.*, no genotype is more fit "on average" than any other genotype). When there is temporal variation in fitness, the geometric mean fitness of genotypes applies (HALDANE and JAYAKAR 1963). The stochastic fitness expectation of Φ is greater than the constant fitness expectation when a small number of alleles are segregating and less than the constant fitness expectation when a large number of alleles are segregating. The geometric mean of a number of independent random variables decreases as the number of variables increases (F. J. ROHLF, personal communication). This is because the geometric mean is sensitive to low values, and each random variable has a chance of resulting in a low value. Random variables in this case refer to genotypic fitnesses. Consequently, the magnitude of Φ is contingent on the relative numbers of heterozygous and homozygous genotypes (which are a function of the number of segregating alleles). Stochastic fitness also influences the genotype-frequency ratio independent of the number of segregating alleles. This is because Φ in a stochastic fitness scenario involves the ratio of two random variables. The geometric mean of a ratio of two identical random variables has an expectation of one. Due to the arithmetic mean–geometric mean inequality, the arithmetic mean of a stochastic fitness ratio is greater than or equal to one, resulting in $\Phi > 2$. Allele-dependent and independent effects of stochastic fitness combine in a complex manner, and MATLAB simulations indicate that when three or fewer alleles are segregating, $\Phi > 2$ (see Table 1).

Visualization of genotype frequencies: The high dimensionality of genotype-frequency space makes visualization difficult. However, it is possible to take two-dimensional slices through genotype-frequency space and view possible frequencies for pairs of genotypes (Figure 2). Five different curves are possible, depending on the number of shared alleles and whether the genotypes in question are homozygous or heterozygous. For example, if one genotype in question involves a homozygote (*ii*) and the other genotype involves a heterozygote that shares zero alleles with the homozygote (*jk*), then Figure 2C applies. Given the assumptions of this model, populations can exist only within the solid regions of Figure 2. Areas and shapes of solid regions are contingent on the ratio of the geometric mean heterozygote fitness to the geometric mean homozygote fitness. The exact position of a population genetic state depends on allele frequencies. For example, one will not find *AA* homozygotes and *AB* heterozygotes at high frequencies if a third allele, *C*, happens to be common. Note that heterozygote advantage in a multiallelic system is unlikely to result in the maintenance of many segregating alleles (LEWONTIN *et al.* 1978), although spatial heterogeneity in selection pressures relaxes these constraints (STAR *et al.* 2007).

Comparison of heterozygosity and Φ: Heterozygosity and the genotypic ratio, Φ, are complementary measures of genetic variation. Both measures exhibit an excess of heterozygotes when there is overdominance and an excess of homozygotes when there is underdominance. However, heterozygosity is maximized at intermediate allele frequencies, while Φ is independent of

TABLE 2

Population-specific patterns of Φ emerge from HapMap data

Population	CEU	CHB	JPT	YRI	Pooled
No. of individuals	60	45	45	60	210
No. of SNPs	400	400	400	400	400
Median Φ	2.0389	2.0207	2.0000	1.9158	1.6500
Mean Φ	2.1543	2.2099	2.0975	1.9782	1.6871
SD Φ	0.7686	0.8909	0.8243	0.6685	0.3561
SE Φ	0.0019	0.0022	0.0021	0.0017	0.0009
P-value (mean $\Phi \neq 2$)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Mean *f*	−0.0034	−0.0039	0.0103	0.0277	0.0895
SD *f*	0.1289	0.1557	0.1463	0.1278	0.1002

f is Wright's inbreeding coefficient. *P*-values were computed using a one-sample *t*-test with 399 d.f.

allele frequency. This is because both homozygous and heterozygous genotypes containing rare alleles will be found at low frequencies (canceling out in Equation 7). Heterozygosity varies as allele frequencies change due to selection. By contrast, the genotypic ratio does not change during an adaptive walk. In addition, expected heterozygosity is greater when more alleles are segregating, while Φ is independent of the number of segregating alleles. Equilibrium heterozygosity of neutral loci depends on population size and mutation rate, while $\Phi = 2$ for neutral loci regardless of population size and mutation rate. Both measures of variation decrease over time when there is inbreeding. Positive assortative mating results in an excess of homozygotes, and negative assortative mating results in an excess of heterozygotes. Population structure also affects both measures of variation. When subpopulations differ in allele frequencies, the frequencies of homozygotes in a pooled population are larger than the mean homozygote frequency of unmixed subpopulations (Hedrick 2005). This reduction in heterozygosity due to population structure is known as the Wahlund effect. The above properties of heterozygosity and Φ hint at the ability to distinguish between alternative evolutionary hypotheses. For example, genotypic ratio data can be combined with other information (such as linkage disequilibrium, allele frequency spectra, and reduced heterozygosity) to provide integrated evidence of selection.

One common measure of genetic variation is Wright's inbreeding coefficient, *f*. This is equal to one minus observed heterozygosity over expected heterozygosity, $f = 1 - (H_{\text{obs}}/H_{\text{exp}})$. An *f* greater than zero corresponds to an excess of homozygotes and an *f* less than zero corresponds to an excess of heterozygotes. While this measure is directly related to the concept of heterozygosity, its relationship to selection coefficients is more convoluted. This is because the magnitude of *f* depends on allele frequencies and does not significantly differ from zero when one allele is rare (see supplemental information). Consider a recessive deleterious allele in mutation–selection balance ($w_{AA} = 0.9$, $w_{AB} = 1.0$, $w_{BB} = 1.0$, $\hat{p}_A = 0.0032$). This scenario results in $f = -0.0003$ and $\Phi = 2.1082$. Each measure of genetic variation is sensitive to a different range of genotype frequencies. If values from different generations and/or loci are averaged, it is possible to have an excess of heterozygotes from one measure and an excess of homozygotes from the other measure. When genotype-frequency data are condensed into a single summary statistic like *f* or Φ, information is unavoidably lost. Thus, a more complete picture of genetic variation arises when both *f* and Φ are calculated (see Tables 1 and 2).

Genomic analysis of Φ: The signature of selection tends to be local within the genome, while population structure often results in genomewide patterns. Ideally, one could calculate Φ across all loci and look for outliers (with the reasoning that large departures from $\Phi = 2$ are indicative of selection). A Bayesian formulation of Φ exists for two segregating alleles (Pereira and Rogatko 1984), allowing the estimation of type I and type II error rates. In practice, however, sample sizes are rarely large enough to detect significant departures from Hardy–Weinberg proportions. This is confounded when genomic data are used because multiple-testing issues arise. An alternative is to calculate the genomewide mean of Φ for different populations. This allows departures from random mating to be detected, as putatively neutral markers are expected to have a mean $\Phi = 2$. Data from the International HapMap Project are well suited for this type of analysis and were used here (International HapMap Consortium 2003). Here, 60 individuals from northern and western Europe (CEU), 45 Han Chinese individuals from Beijing (CHB), 45 Japanese individuals from Tokyo (JPT), and 60 Yoruban individuals from Ibadan, Nigeria (YRI) were sequenced at ~800,000 SNP markers. HapMap Data Release 23a was used (phase II, March 2008, NCBI B36 assembly). Φ was calculated for 400 randomly selected SNPs covering the short arm of the third chromosome (see supplemental information for a list of SNPs and genotype frequencies). Linkage disequilibrium in human populations decays substantially over 200 kb (Ke *et al.* 2004). To ensure independence of data points SNPs were chosen that were at

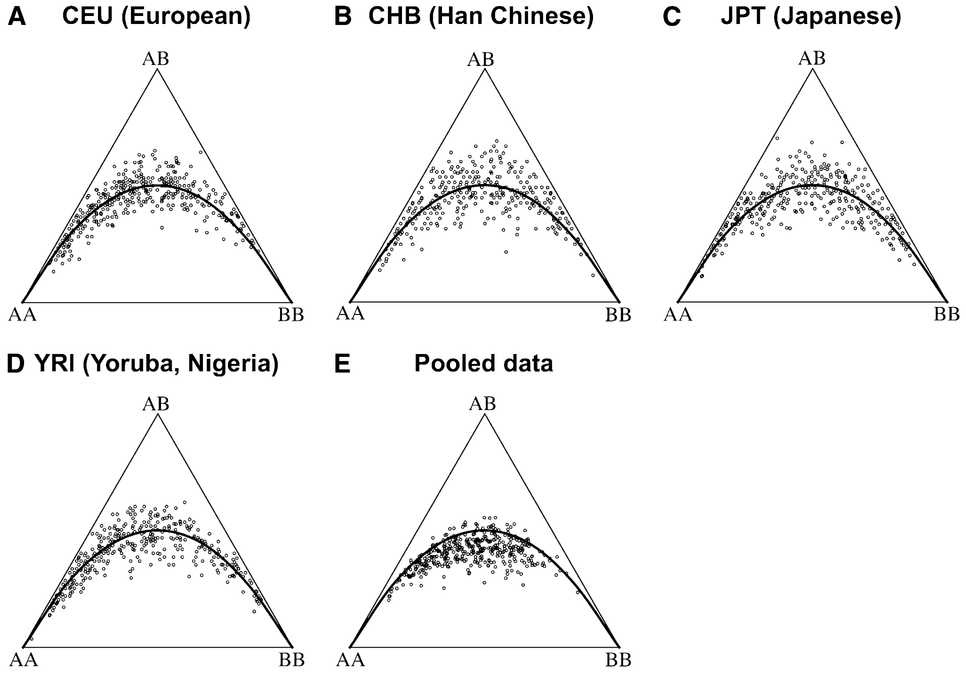

FIGURE 3.—De Finetti diagrams of HapMap genotype frequencies. Each data point corresponds to one of 400 independent SNPs. The reference allele is denoted *A*, and the nonreference allele is denoted *B*. Solid curved lines signify $\Phi = 2$. (A) Sixty individuals from northern Europe. (B) Forty-five individuals from Beijing. (C) Forty-five individuals from Tokyo. (D) Sixty Yoruban individuals from Ibadan, Nigeria. (E) Pooled data from all four HapMap populations.

least 200 kb apart. SNPs were required to be polymorphic in all four populations, and an additional criterion was that heterozygote and both homozygote genotypes were observed. Results are summarized in Figure 3 and Table 2. European and Chinese populations exhibited an excess of heterozygotes while the Yoruban population exhibited an excess of homozygotes. There is a large spread in values of Φ, owing to the relatively small number of individuals in each sample population. When population data are pooled, mean $\Phi < 2$ ($P < 0.0001$, one-sample *t*-test with 399 d.f.). This is indicative of a Wahlund effect. For each population mean Φ significantly differed from 2 ($P < 0.0001$, one-sample *t*-test with 399 d.f.).

Numerous selective and demographic causes can explain these patterns. An excess of heterozygotes is consistent with overdominance, associative overdominance, stochastic fitness of diallelic loci, negative selection against deleterious recessive alleles, and positive selection of dominant advantageous alleles. Conversely, an excess of homozygotes is consistent with underdominance, negative selection against deleterious dominant alleles, and positive selection of advantageous recessive alleles. However, values of Φ seen in the HapMap data set would require very large selection coefficients (on the order of 10%). Also, it is unlikely that all loci in question are under selection (KIMURA 1983, but see HAHN 2008), and there are no *a priori* reasons why the four HapMap populations would have such different signatures of selection. In contrast to the local footprint of selection, demography yields genome-wide patterns. Negative assortative mating, where individuals preferentially mate with individuals with different genotypes, results in an excess of heterozygotes over panmictic expectations. Inbreeding avoidance also results in $\Phi > 2$ (PUSEY and WOLF 1996). Both positive assortative mating and the pooling of subdivided populations result in an excess of homozygotes. Each of the four HapMap populations has a different demographic history, potentially explaining why they differ in mean Φ. Alternatively, ascertainment bias could be responsible for the differences between populations. Individuals were selected via different methods for each population, particularly with respect to the presence of couples (INTERNATIONAL HAPMAP CONSORTIUM 2003). In addition, criteria for ethnic identity ranged from self-identification (Japanese) to all four grandparents sharing the same culture (Yoruban). While it is possible for the effects of selection and population structure to cancel out (resulting in $\Phi = 2$), this is unlikely to occur on a genomic scale. At present, the above causes cannot be distinguished by genotypic ratio data. Indeed, they are not mutually exclusive and pluralistic explanations are possible.

CONCLUSION

The ratio of geometric mean heterozygote frequency to geometric mean homozygote frequency is coupled to the effects of natural selection. It provides a measure of genetic variation that is complementary to heterozygosity and can be used to detect the signature of evolutionary processes. As larger numbers of individuals are sequenced (as in MACDONALD *et al.* 2005), the utility of the genotypic ratio will increase. Genotype frequencies bear the footprint of differential fitnesses, and elegant mathematical patterns arise from the natural phenomena of Mendelian segregation and Darwinian selection.

I thank A. Onstine, H. Spencer, J. True, S. Yeh, R. Yukilevich, and the Eanes lab for constructive criticism during the preparation of this manuscript. This work was supported by a National Institutes of Health predoctoral training grant (5 T32 GM007964-24).

LITERATURE CITED

Alvarez, G., 2008 Deviations from Hardy-Weinberg proportions for multiple alleles under viability selection. Genet. Res. **90:** 209–216.

Cannings, C., and A. W. F. Edwards, 1968 Natural selection and the de Finetti diagram. Ann. Hum. Genet. **31:** 421–428.

Edwards, A. W. F., 2000 *Foundations of Mathematical Genetics*. Cambridge University Press, Cambridge, UK/London/New York.

Ewens, W. J., 1989 An interpretation and proof of the fundamental theorem of natural selection. Theor. Popul. Biol. **36:** 167–180.

Fisher, R. A., 1953 Population genetics. The Croonian lecture. Proc. R. Soc. Lond. Ser. B **141:** 510–523.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

Hahn, M. W., 2008 Toward a selection theory of molecular evolution. Evolution **62:** 255–265.

Haldane, J. B. S., and S. D. Jayakar, 1963 Polymorphism due to selection of varying direction. J. Genet. **58:** 237–242.

Hartl, D. L., D. E. Dykhuizen and A. M. Dean, 1985 Limits of adaptation: the evolution of selective neutrality. Genetics **111:** 655–674.

Hedrick, P. W., 2005 *Genetics of Populations*. Jones & Bartlett, Sudbury, MA.

Hernández, J. L., and B. S. Weir, 1989 A disequilibrium approach to Hardy-Weinberg testing. Biometrics **45:** 53–70.

Hutchinson, G. E., 1957 Concluding remarks. Cold Spring Harbor Symp. Quant. Biol. **22:** 415–427.

International HapMap Consortium, 2003 The International HapMap Project. Nature **426:** 789–796.

Ke, X., S. Hunt, W. Tapper, R. Lawrence, G. Stavrides *et al.*, 2004 The impact of SNP density on fine-scale patterns of linkage disequilibrium. Hum. Mol. Genet. **13:** 577–588.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

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

Lewontin, R. C., L. R. Ginzberg and S. D. Tuljapurkar, 1978 Heterosis as an explanation for large amounts of genetic polymorphism. Genetics **88:** 149–170.

Macdonald, S. J., T. Pastinen and A. D. Long, 2005 The effect of polymorphisms in the enhancer of split gene complex on bristle number variation in a large wild-caught cohort of *Drosophila melanogaster.* Genetics **171:** 1741–1756.

Mathworks, 2005 *MATLAB 7*. Mathworks, Natick, MA.

Pereira, C., and A. Rogatko, 1984 The Hardy-Weinberg equilibrium under a Bayesian perspective. Rev. Bras. Genet. **4:** 689–707.

Pusey, A., and M. Wolf, 1996 Inbreeding avoidance in animals. Trends Ecol. Evol. **11:** 201–206.

Ramachandran, G. N., C. Ramakrishnan and V. Sasisekharan, 1963 Stereochemistry of polypeptide chain configurations. J. Mol. Biol. **7:** 95–99.

Rice, S. H., 2004 *Evolutionary Theory*. Sinauer Associates, Sunderland, MA.

Star, B., R. J. Stoffels and H. G. Spencer, 2007 Single-locus polymorphism in a heterogeneous two-deme model. Genetics **176:** 1625–1633.

Weir, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Wright, S., 1932 The roles of mutation, crossbreeding, and selection in evolution. Proc. 6th Int. Congr. Genet. **1:** 356–366.

Communicating editor: H. G. Spencer