doi 10.4436/jass. 96009

JASs *Invited Reviews*
Journal of Anthropological Sciences
Vol. 96 (2018), pp. 7-26

Neutral theory and the evolution of human physical form: an introduction to models and applications

Timothy D. Weaver

Department of Anthropology, University of California, Davis, CA 95616, USA
e-mail: tdweaver@ucdavis.edu

Summary - *Anthropologists have long been interested in explaining patterns of variation in human physical form, in both present-day and ancient humans. Starting in the 1950s, their explanations became more firmly rooted in evolutionary theory, but they have typically focused on adaptive accounts. Neutral explanations – those grounded in models of evolution by mutation, genetic drift, and gene flow rather than natural selection – provide an alternative to adaptive explanations, and in recent years, neutral models have become an important tool for researchers investigating the evolution of human physical form. Neutral models have implications for many areas of biological anthropology, including using morphology to reconstruct the histories and migrations of recent human populations, using morphology to infer the evolutionary relationships among hominin taxa, and clarifying how natural selection has acted on physical form throughout human evolution. Their application to anthropological questions has stimulated biological anthropologists to more seriously consider the roles of history and chance in human evolution. In light of the growing importance of neutral explanations in biological anthropology, the goal here is to provide an introduction to neutral models of phenotypic evolution and their application to human physical form.*

Keywords - *Human evolution, Human variation, Genetic drift, Mutation, Gene flow, Coalescence.*

Introduction

Anthropologists have long been interested in explaining patterns of variation in human physical form, in both present-day and ancient humans. Indeed, one major division of anthropology, physical anthropology, initially concerned itself almost exclusively with human physical form, skeletal morphology in particular (Boas, 1899; Hrdlicka, 1908), before this field diversified into other areas, such as genetics, non-human primate behaviour and ecology, and human biology, and began to be commonly called biological anthropology. As von Cramon-Taubadel (2014) recently reviewed in this journal in her article about human cranial morphology, early views of human physical form were typological, with the goals of identifying the characteristics of different human groups and classifying individuals (e.g., Hooton, 1926). Starting in the 1950s, investigations of human physical form became more firmly rooted in evolutionary theory, but they have typically focused on adaptive explanations (e.g., Washburn, 1951).

Neutral models – models of evolution by mutation, genetic drift, and gene flow rather than natural selection – for protein and subsequently DNA-sequence evolution were first proposed in the 1960s and 1970s (Kimura, 1968; King & Jukes, 1969; Kimura & Ohta, 1971), and these foundational neutral models were followed by neutral models of phenotypic evolution in the 1970s and 1980s (Lande, 1976; Lynch & Hill, 1986; Turelli *et al.*, 1988). Although the earliest neutral models considered only mutation and genetic drift, subsequent ones also included gene flow (e.g., Kimura & Maruyama, 1971; Lynch, 1988a). Neutral models forever changed evolutionary biology, perhaps most significantly because they highlighted that evolution and natural selection are not synonyms, which forced researchers to evaluate adaptive explanations

Tab. 1 - Definition of mathematical terms.

t_g	split time in generations
g	generation length
$\bar{x}_i$	mean of a single phenotypic trait in group i
$\bar{x}$	grand mean (mean of the group means) of a single phenotypic trait
d	number of groups
i	number of steps separating two groups in a circular stepping-stone model
d_s	number of sampled groups
V_W	within-group phenotypic variance
V_B	between-group phenotypic variance
h^2	(narrow sense) heritability
V_{W_A}	within-group additive genetic variance (h^2V_W)
N	population size (number of breeding individuals) of each group
N_e	effective population size of each group
σ^2_m	additive genetic variance introduced by mutation
B	between-group phenotypic variance-covariance matrix
P	within-group phenotypic variance-covariance matrix
G	within-group additive-genetic covariance-matrix
U	additive genetic variances and covariances introduced by mutation
τ_W	average coalescence time of pairs of alleles from the same group
τ_B	average coalescence time of pairs of alleles from different groups
τ	average coalescence time of pairs of alleles from the collection of groups sampled
$M/2$	proportion of individuals exchanged each generation by two adjacent groups, in each direction
F_{ST}	measure of genetic differentiation calculated from genetic data
Q_{ST}	measure of genetic differentiation calculated from experimental phenotypic data
P_{ST}	measure of genetic differentiation calculated from observational phenotypic data
c	scale factor for the between-group phenotypic variance
Δ_A	difference between two groups in their additive-genetic-effect means
Δ_O	difference between two groups in their other-effect (dominance genetic, interaction genetic, environmental) means

against an alternative to natural selection that was often difficult to dismiss. Additionally, it became clear that even if one was primarily interested in adaptive changes, neutral models provided a foundation for any evolutionary investigation, because natural selection will act in concert with the neutral evolutionary processes of mutation, genetic drift, and gene flow.

Neutral models began to make inroads into studies of human physical form in the 1980s and 1990s (Rogers & Harpending, 1983; Lynch, 1989; Relethford, 1994), and they became a prevalent tool in these investigations in the 2000s (reviewed by Roseman & Weaver, 2007; von Cramon-Taubadel & Weaver, 2009). Many of these studies of human physical form were based on the global dataset of human cranial variation collected by Howells (1973, 1989, 1995). Neutral models provide a theoretical basis for using morphology to reconstruct the history and structure of recent human populations and the evolutionary relationships among hominin taxa. They can be used as a baseline against which to evaluate adaptive hypotheses about human physical form. Perhaps most fundamentally, studies demonstrating the importance of neutral models to anthropological questions have stimulated biological anthropologists to more seriously consider the roles of history and chance in human evolution.

With this background in mind, the goal here is to provide an introduction to neutral models of phenotypic evolution and their application to human physical form. More detailed descriptions of specific neutral models can be found elsewhere, but there is currently no single source that gathers together the models relevant to biological anthropological investigations and discusses the connections among them. The mathematical terms used in the paper are defined in Table 1 and as they are discussed. Table 2 provides a glossary of terms from quantitative and population genetics and evolutionary biology (see also Relethford, 2007; Konigsberg, 2012; von Cramon-Taubadel, 2014). Table 3 gives brief explanations for how key quantitative and population genetic parameters can be estimated.

Tab. 2 - Glossary.

Adaptive evolution. Typically, adaptive evolution refers to evolution (change) due to natural selection, and this is how this phrase is used here. However, adaptive – and therefore adaptive evolution – is not defined the same way by all evolutionary biologists. See also neutral evolution.

Additive genetic variance. Variation among individuals for a particular trait is typically quantified by variance, which is the average squared deviation of each individual from the mean (average). Additive genetic variance is the fraction of the variance that is due to additive genetic effects, which excludes genetic variance from genetic interactions between alleles at the same locus (dominance effects) or different loci (epistatic effects). Evolutionary quantitative genetic models focus on additive genetic variance because recombination tends to break down genetic interactions across generations. Variance also results from environmental effects. See also environmental variance and phenotypic variance.

Allele. The genome of any particular individual can be divided up into different locations, or loci, on the individual's chromosomes. At each locus there can be different variants (DNA sequences), and these different variants are called alleles.

Arithmetic mean. The typical average, which is simply the mean of the quantity of interest. See also harmonic mean.

Coalescence. Because of genetic drift, a sample of present-day lineages will share fewer and fewer ancestral lineages the further back in time one looks. The sample of present-day lineages will eventually share a single, common ancestral lineage at a particular time in the past, which is called the coalescence time (e.g, 9,600 years ago).

Effective population size. Roughly, the number of breeding individuals in an idealized population that would have as much genetic drift as in the actual population. It is often quite different from both the breeding size (number of breeding individuals) and the census size (total number of individuals) of the actual population.

Environmental variance. Variance from developmental responses to environmental stimuli. See also additive genetic variance and phenotypic variance.

Equilibrium. As used here, an equilibrium is when a balance, or steady state, is reached between different evolutionary processes (forces).

Fitness. Roughly, fitness measures reproductive success, which depends on many components, including survival, mating success, and the production of offspring.

F_{ST}. A measure of the amount of genetic differentiation among groups that varies from zero to one, with zero and one corresponding respectively to the minimum and maximum amounts of differentiation.

Gene flow. Exchange of genes among groups, usually through the migration of individuals from one group to another group.

Generation length. The average age of the parents when their children are born.

Genetic drift. Random changes in the genetic composition of a group (population) because there are only a finite number of individuals in the group.

Harmonic mean. The reciprocal of the mean of the reciprocals of the quantity of interest. The harmonic mean is strongly influenced by small values.

Heritability. The ratio of the additive genetic variance to the phenotypic variance for a trait within a group. Heritability measures the degree to which offspring resemble their parents, and consequently, it affects the rate of evolution. Importantly, heritability can often be overestimated in humans, because offspring can resemble their parents for non-genetic reasons (e.g., common environment, cultural transmission), and heritability is not fixed but can vary across populations and environments. See also Box 1.

Heterozygosity. A measure of genetic variation, which reflects that with more alleles, individuals will have a higher chance of being heterozygous. Heterozygous individuals have two different alleles at a particular locus. See also allele.

Linkage disequilibrium. When alleles at different loci tend to be found together. Linkage disequilibrium occurs because blocks of DNA are inherited together, but over time, recombination will break down these associations. See also allele.

Microsatellite. A rapidly evolving block of DNA in which a simple DNA sequence is repeated multiple times and individuals vary in their number of repeats. Also called a short tandem repeat (STR).

Mutation. A change to the DNA sequence, which can be passed on to offspring if it occurs in a cell on the path to becoming a gamete (egg or sperm).

Neutral evolution. Evolution (change) by neutral evolutionary processes (forces): mutation, genetic drift, and gene flow. See also adaptive evolution.

Natural selection. Changes in the genetic composition of a group (population) and associated phenotypic traits because there is a causal relationship between aspects of the phenotype and fitness. Directional selection shifts the population mean up or down, whereas stabilizing selection maintains the status quo. See also adaptive evolution.

Pedigreed sample. A sample for which the genealogical relationships among individuals are known from breeding records.

Phenotypic variance. A measure of variation among individuals for a particular trait. See also additive genetic variance and environmental variance.

Short tandem repeat. See microsatellite.

Stochastic. Random.

Zygote. The cell that forms when the male and female gametes (egg and sperm) combine.

Tab. 3 - Estimating quantitative and population genetic parameters.

Additive genetic variance, covariance, and correlation. Additive genetic variances, covariances, and correlations have traditionally been estimated by fitting a mixed statistical model to a pedigreed sample (see Lynch & Walsh, 1998; Runcie & Mukherjee, 2013). More recently, methods have been developed that leverage genomic data to obtain estimates from unpedigreed individuals (see Lee *et al.*, 2012). Large sample sizes are needed to accurately estimate these quantities, and depending on how the individuals in the sample are related, the effective sample size can be much lower than the number of individuals (Cheverud, 1988). Consequently, phenotypic variances, covariances, and correlations, which are easier to accurately estimate, are often substituted for their additive-genetic counterparts in anthropological studies (e.g., Ackermann & Cheverud, 2004; Weaver *et al.*, 2007; Grabowski & Roseman, 2015). This substitution, termed "Cheverud's Conjecture", is supported by empirical findings in both humans (Sodini *et al.*, 2018) and non-humans (Cheverud, 1988; Roff, 1996). Mixed models have been used to estimate these quantities for a variety of human traits, including cranial (e.g., Carson, 2006; Martínez-Abadías *et al.*, 2009) and dental (e.g., Stojanowski *et al.*, 2017) measurements. There has also been similar work on non-human primates (e.g., Hlusko & Mahaney, 2009; Roseman *et al.*, 2010).
Effective population size. Effective population size can be estimated in humans by making assumptions about the relationship between census size and effective size (e.g., effective size is a third of the census size, Cavalli-Sforza *et al.*, 1994; Relethford *et al.*, 1997). However, the relationship between effective size and census size may not be the same for different populations or through time, because of variation in, for example, mating practices, age structure, and demographic history. Phenotypic data could be used to estimate effective population size, if one assumes that the traits were evolving neutrally (e.g., Relethford & Harpending, 1994). Effective population size can also be estimated from molecular genetic data. Older methods focused on what could be inferred about long-term effective population size, or effective population size at a few points in the past, from genetic variation, linkage disequilibrium, or coalescence time (e.g., Harpending & Rogers, 2000; Hayes *et al.*, 2003), whereas newer molecular genetic methods combine genomic data with statistical modeling to estimate a roughly continuous record of changes in effective population size (e.g., Li & Durbin, 2011). The long-term effective population size is the average (harmonic mean) effective population size over a number of generations. Evolutionary quantitative genetic studies of human and non-human primate cranial and dental variation have made use of effective population size estimates from molecular genetic data (e.g., Weaver & Stringer, 2015; Rathmann *et al.*, 2017; Reyes-Centeno *et al.*, 2017; Schroeder & von Cramon-Taubadel, 2017).
Generation length. Generation length is estimated in humans (e.g., Fenner, 2005) and non-human primates (e.g., Langergraber *et al.*, 2012), which are slowly reproducing species with overlapping generations, by averaging over all parents in demographic databases.
Heritability. Heritability has traditionally been estimated from comparisons of close relatives (e.g., parents to offspring, full to half siblings, or monozygotic to dizygotic twins) with regression or correlation analyses, or from comparisons across extended pedigrees with mixed statistical models (see Falconer & Mackay, 1996; Lynch & Walsh, 1998; Visscher *et al.*, 2008; Tenesa & Haley, 2013). More recently, methods have been developed that leverage genomic data to estimate heritability from unpedigreed individuals (see Visscher *et al.*, 2008; Tenesa & Haley, 2013). Mixed models have been used to estimate heritability for a variety of human traits, including cranial (e.g., Carson, 2006; Martínez-Abadías *et al.*, 2009) and dental (e.g., Stojanowski *et al.*, 2017) measurements. There has also been similar work on non-human primates (e.g., Hlusko & Mahaney, 2009; Roseman *et al.*, 2010).
Mutational variance. Mutational variance is, more precisely, the additive genetic variance introduced by mutation (per zygote per generation) for a particular phenotypic trait. Mutational variance is usually expressed as a (small) fraction (e.g., 10^{-4}) of the environmental component of the within-group phenotypic variance, and this fraction is called the mutational heritability. Mutational heritability is estimated from experiments conducted on taxa (e.g, fruit flies, mice) that can be readily bred in a controlled manner in the laboratory, so that mating follows a well-defined pattern and natural selection can be minimized (see Lynch, 1988b; McGuigan *et al.*, 2015). Studies of human physical form assume that mutational heritability estimates from other taxa apply to humans.
Split time (population divergence time). Split time can be estimated from the fossil record. For example, if the population represented by the Sima de los Huesos fossils is ancestral to Neandertals, then the split between the Neandertal and modern human evolutionary lineages must predate ≈430,000 years ago, the age of the site based on multiple radiometric dating techniques (Arsuaga *et al.*, 2014). Split time can also be estimated from molecular genetic data with a variety of approaches (see Noonan *et al.*, 2006; Prado-Martinez *et al.*, 2013; Prüfer *et al.*, 2014). Evolutionary quantitative genetic studies of human and non-human primate cranial variation have made use of split time estimates from molecular genetic data (e.g., Weaver & Stringer, 2015; Reyes-Centeno *et al.*, 2017; Schroeder & von Cramon-Taubadel, 2017).

Classic approaches

Genetic drift and mutation

To begin discussing neutral models of phenotypic evolution, imagine that an ancestral human group (population, species) gives rise to two descendant groups, and the descendant groups evolve independently for t_g generations (Fig. 1). The time in the past of the split in years is given by gt_g where g is the generation length (average age of the parents when their children are born). This scenario describes, at least to a first approximation, many situations of interest to biological anthropologists (e.g., the split between the Neandertal and modern human evolutionary lineages from a common ancestor, before Neandertals and modern humans came into contact many generations later in Eurasia), and it can be extended to more complicated situations by positing a series of splits. We are interested in tracking the change through time of a single phenotypic trait (e.g., cranial length), which has an average value $\bar{x}_1$ in one of the descendant groups and $\bar{x}_2$ in the other descendant group. Suppose that the groups are equally variable for the trait – i.e., they have the same within-group phenotypic variance, V_W.

To start, consider the simplified situation in which genetic drift is the only evolutionary process (force) acting. A heritable trait will evolve when the covariance for the trait and fitness is non-zero (e.g., when larger values of the trait are associated with higher fitness than smaller values of the trait). We tend to think about this covariance in terms of natural selection – that is, some values of the trait are more *helpful* than others or are the *cause* of differences in fitness – but covariance for a trait and fitness can also arise by stochastic (random) evolutionary processes, particularly in small populations, even if there is no causal relationship between different values of a trait and fitness (Rice, 2004, 2008). In particular, genetic drift – random changes in the genetic composition of a group (population) because of finite size – can produce such covariances (Fig. 2). On average, we would not expect a random process to result in covariance for a trait and

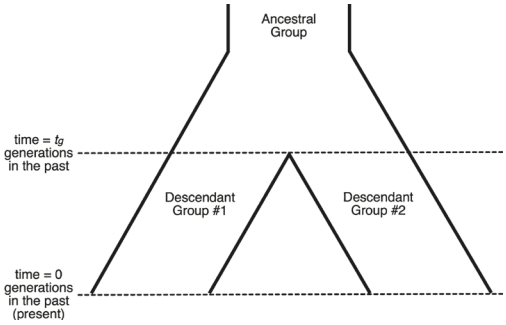

Fig.1 - Model of a split into two groups. An ancestral group split into two descendant groups tg generations in the past.

fitness, because positive covariances will tend to cancel negative covariances. However, it is the realized covariance each generation, rather than the expected covariance, that determines the direction and magnitude of evolutionary change.

If only genetic drift is acting, the expected between-group phenotypic variance for the two descendant groups,

$$V_B = \left(\bar{x}_1 - \tfrac{\bar{x}_1+\bar{x}_2}{2}\right)^2 + \left(\bar{x}_2 - \tfrac{\bar{x}_1+\bar{x}_2}{2}\right)^2 = \tfrac{(\bar{x}_1-\bar{x}_2)^2}{2},$$

after t_g generations is

$$\mathsf{E}\{V_B\} = \frac{h^2 V_W}{N_e} t_g \qquad (1)$$

where E denotes the average evolutionary outcome (mathematical expectation), h^2 is the (narrow sense) heritability, and N_e is the effective population size of each group (Lande, 1976). Heritability measures the degree to which offspring resemble their parents. The effective population size roughly corresponds to the number of breeding individuals in an idealized population that would have as much genetic drift as in the actual population, and it is often quite different from both the breeding size (number of breeding individuals) and the census size (total number of individuals) of the actual population. For simplicity, we assume that h^2 and N_e are the same for both groups, but it is possible to modify Eq. (1) to account for differences in these quantities, and in V_W. The product $h^2 V_W$ is the

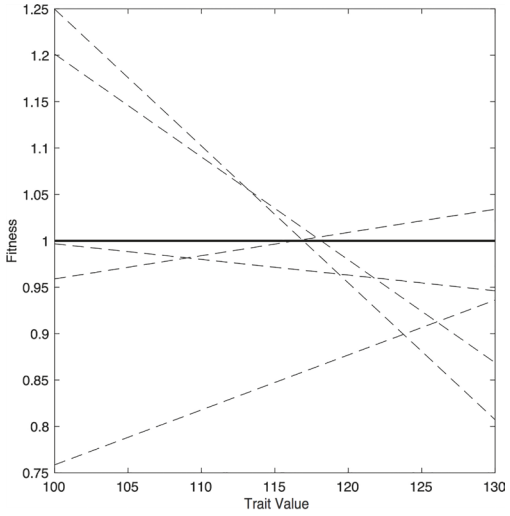

Fig. 2 - Illustration of how genetic drift can produce covariance for a trait and fitness. With just genetic drift acting, because there is no natural selection, the expected fitness for each value of the trait is the same (fitness=1), so the expected covariance for the trait and fitness is zero (solid line). However, because of genetic drift, the realized fitness for each value of the trait is sampled from a distribution, so the realized covariance for the trait and fitness can be non-zero (dashed lines). The plot is based on a population size of ten and a normal fitness distribution (mean=1, standard deviation=0.1).

within-group additive genetic variance, which is the component of variation important for evolutionary change, because it quantifies variation due to genetic effects on the phenotype that can be passed from parents to offspring. According to Eq. (1), the between-group phenotypic variance increases linearly with time with a rate that is faster with more additive genetic variance and smaller effective population size. Note that the expected difference between the group means, $\mathsf{E}\{\bar{x}_1 - \bar{x}_2\}$ remains zero regardless of split time, because positive differences will tend to cancel negative differences.

Schroeder *et al.* (2014) used an extension of Eq. (1) to multiple traits (see section "Multiple traits" below) to investigate the transition in human evolution from *Australopithecus* to *Homo*. Specifically, they assessed whether or not the differences in cranial and mandibular measurements between taxa separated in time were consistent – given assumptions about heritability, effective population size, and generation length – with neutral divergence. They found that many of the possible pairwise comparisons between taxa were consistent with neutrality, although they often detected too much morphological divergence, and sometimes too little morphological divergence, to be consistent with neutrality.

Substituting $\frac{(\bar{x}_1-\bar{x}_2)^2}{2}$ into Eq. (1) and rearranging terms gives

$$\mathsf{E}\left\{\frac{(\bar{x}_1-\bar{x}_2)^2}{V_W}\right\} = \frac{2h^2}{N_e}t_g. \qquad (2)$$

The quantity on the left of Eq. (2) is the (squared) Mahalanobis distance for a single variable. In general, the Mahalanobis distance measures dissimilarity aggregated across multiple traits in a way that accounts for within-group variances and covariances, but in the case of a single trait there are no covariances, only a variance. Although Eq. (2) is for a single trait, it is possible to show that a similar relationship holds for the Mahalanobis distance for multiple traits. The important point for the discussion here is that Eq. (2) shows that the Mahalanobis distance is expected to increase linearly with split time under neutrality (Lynch, 1990). A genetic distance for short tandem repeats (STRs, microsatellites), $(\delta\mu)^2$, is also expected to increase linearly with split time under neutrality (Goldstein *et al.*, 1995b). Harvati & Weaver (2006a,b) made use of this correspondence when they compared Mahalanobis distances calculated from anatomical landmarks collected on recent human crania to a group-matched set of $(\delta\mu)^2$ genetic distances. With this comparison, they assessed how much different cranial regions deviated from neutral expectations, as a way to decide which regions would be most useful for inferring population history or phylogeny.

Although Eq. (1) focuses on how genetic drift will affect between-group variation, it also implicitly assumes that genetic drift will not

affect within-group variation. In fact, with just genetic drift acting, within-group variation will decrease with time at a rate that is faster with smaller effective population size and slows as the within-group variation decreases. More precisely, the expected change in within-group additive genetic variance from one generation to the next is

$$\mathrm{E}\left\{h^2 V_{W(t_g+1)}\right\} = \left(1 - \frac{1}{2N_e}\right) h^2 V_{W(t_g)} \qquad (3)$$

where the t_g subscript indicates the current generation, and $t_g + 1$ the subscript indicates the next generation (Turelli *et al.*, 1988).

Unlike the predictions of Eq. (3), within-group variation often remains fairly constant [as implicitly assumed by Eq. (1)]; this is because as genetic drift removes variation, mutation adds variation. Accordingly, for a more complete neutral model one needs to model the effects of both genetic drift and mutation. Importantly, because the addition of variation by mutation is roughly constant per generation but the subtraction of variation by genetic drift decreases as the within-group variation decreases, eventually, the within-group additive genetic variance will stabilize at an equilibrium value

$$h^2 V_{W(\mathrm{equil})} = 2N_e\sigma_m^2 \qquad (4)$$

where σ_m^2 is the additive genetic variance introduced by mutation (per zygote per generation) (Lande, 1979, 1980; Lynch & Hill, 1986; Turelli *et al.*, 1988). The mutational variance is usually expressed as a (small) fraction (e.g., 10^{-4}) of the environmental component of the within-group phenotypic variance (environmental variance). The environmental variance is equal to the difference between the phenotypic and additive genetic variances in a model for which the phenotypic variance is only due to additive genetic and environmental effects. Figure 3 illustrates how the within-group additive genetic variance approaches and stabilizes at an equilibrium value through the interaction between genetic drift and mutation.

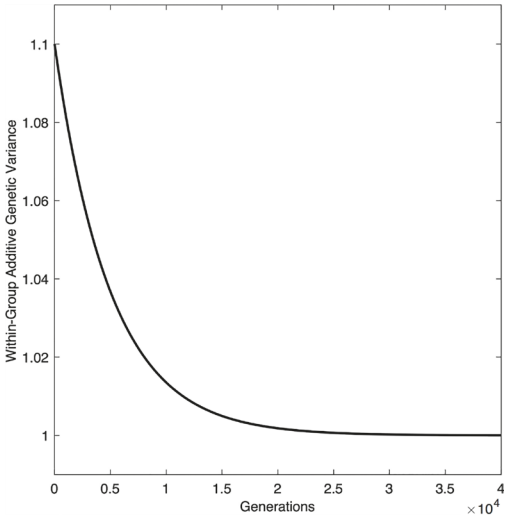

Fig. 3 - Illustration of how the within-group additive genetic variance approaches and stabilizes at an equilibrium value. The plot is based on neutral evolution with $V_{W(initial)}$ = 2.75, h^2 = 0.4, σ_m^2 = 2 × 10⁻⁴, and N_e = 2500.

Substituting the equilibrium within-group additive genetic variance given by Eq. (4) into Eq. (1) gives

$$\mathrm{E}\{V_B\} = 2\sigma_m^2 t_g. \qquad (5)$$

Eq. (5) shows that under a more complete neutral model the rate of increase of between-group variation does not depend on effective population size; it only depends on how much mutational variation is added each generation. This prediction is analogous to how, according to Kimura's (1968, 1983, 1989) neutral model of molecular evolution, the rate of molecular evolution only depends on mutation rate. Effective population size does not appear in Eq. (5) because it cancels when Eqs. (1) and (4) are combined. Intuitively, it is often assumed that neutral divergence between groups will be faster with small population sizes, but Eq. (5) shows that this is not the case at equilibrium.

Lynch (1989) used an extension of Eq. (5) to multiple traits (see section "Multiple traits" below) in one of the first demonstrations that the patterns of recent human cranial variation were

consistent with neutral divergence. This result was somewhat surprising because, at the time, most explanations for variation in human cranial form were adaptive (e.g., Guglielmino-Matessi *et al.*, 1979; Beals *et al.*, 1983). In a subsequent paper, Lynch (1990) examined the rates of evolution in mammals, including primates, concluding that in most lineages stabilizing selection has slowed the rate of evolution relative to what would be expected under neutrality. Because these papers appeared in biology journals, they did not strongly impact the field of anthropology when they were published, but soon afterward, anthropologists independently reached similar conclusions using a somewhat different approach, and these papers and the studies they stimulated shifted many biological anthropologists' views on human cranial variation (see section "P_{ST} to F_{ST} comparisons" below).

Demographic changes

So far, we have assumed that population size remains constant, but real populations often grow or shrink in size, so this section discusses how we can model neutral phenotypic evolution in the face of demographic changes. Imagine that a subset of the individuals in an ancestral group establishes a descendant group, a subset of individuals in the descendant group establishes a new descendant group, and this process continues numerous times as the collection of groups expands across a geographic area. This process is called a serial founder effect. Because each descendant group is smaller, at least initially, than its ancestral group, we expect the effective population sizes of the groups to decrease with increased geographic distance from the source location. Even if all of the groups subsequently grow in size, the average effective population size of a group will stay close to its initial size for some time. Statistically, this is because the relevant average is the harmonic mean (rather than the arithmetic mean), which is strongly influenced by small values (i.e., reductions in effective population size). Biologically, this is because it takes time for the variation lost in a founder effect to be replenished by mutation. Eq. (4) shows that if a phenotypic trait is evolving neutrally, we expect the within-group phenotypic variance for the trait to decrease with the group's geographic distance from the source location, because within-group variance is expected to be proportional to effective population size.

Present-day human DNA sequences show decreasing heterozygosity with genetic (Eller, 1999; Harpending & Rogers, 2000) and geographic (Prugnolle *et al.*, 2005; Ramachandran *et al.*, 2005) distance from sub-Saharan Africa, which has been interpreted to indicate that a serial founder effect underlies the expansion of modern humans from Africa ≈50,000 years ago (Harpending & Rogers, 2000; Prugnolle *et al.*, 2005; Ramachandran *et al.*, 2005; Liu *et al.*, 2006; DeGiorgio *et al.*, 2009). Accordingly, if neutral processes played an important role in the evolution of recent human physical form, Eq. (4) would predict a decrease in within-group phenotypic variance with geographic distance from Africa (assuming the heritability of traits does not vary with geographic distance from Africa in such a way to obscure decreases in within-group additive genetic variance). This pattern has, in fact, been found for various measurements of the human cranium, dentition, and pelvis (Manica *et al.*, 2007; von Cramon-Taubadel & Lycett, 2008; Hanihara, 2008; Betti *et al.*, 2009, 2013), although given the modest strength of the relationships, evolutionary processes other than neutral ones could still be important.

Building on work on STRs (Goldstein *et al.*, 1995a,b; Zhivotovsky & Feldman, 1995; Zhivotovsky, 2001), Weaver *et al.* (2008) extended Eq. (5) to situations in which the effective population sizes had grown from the ancestral to the descendant groups

$$\mathrm{E}\{V_B\} = 2\sigma_m^2 t_g - h^2 (V_{W_1} + V_{W_2} - 2V_{W_{12}}) \quad (6)$$

where V_{W_1} is the within-group phenotypic variance for one of the descendant groups, V_{W_2} is the within-group phenotypic variance for the other descendant group, and $V_{W_{12}}$ is the within-group

phenotypic variance for the ancestral group. If the effective population sizes of the descendant groups are larger than the effective population size of the ancestral group, the second term on the right-hand side of the equation will be negative, so the expected between-group phenotypic variance will be lower than at mutation-drift equilibrium [Eq. (5)]. The two groups will diverge phenotypically at a slower rate than at equilibrium because they will have too little within-group phenotypic variation for their (larger) effective population sizes [i.e., the effective population size that corresponds to the numerator will be smaller than the effective population size in the denominator of Eq. (1)]. If effective population size, and thus within-group additive genetic variance [see Eq. (4)], has not changed from the ancestral group to the descendant groups, Eq. (6) reduces to Eq. (5). Finally, if the split was a long time ago (≈35,000 generations), the first term on the right-hand side of Eq. (6) will dominate, so Eqs. (5) and (6) will give similar predictions for the between-group variance. In other words, the percentage error from (incorrectly) assuming mutation-drift equilibrium will be small (<5%) if the split time was a long time ago (Fig. 4).

Weaver *et al.* (2008) and Weaver & Stringer (2015) used Eq. (6) in a slightly different form – solved for t_g instead of V_B – to estimate split times from cranial measurements between Neandertals and modern humans and between subspecies of common chimpanzee. Consistent with neutral divergence, the split time estimates for Neandertals and modern humans based on morphology were similar to those based on DNA sequences. In contrast, the split time estimates for common chimpanzees based on morphology were much lower than those based on DNA sequences, possibly because cranial divergence in common chimpanzees has been constrained by stabilizing selection.

Multiple traits

It is possible to extend models of between-group divergence by genetic drift and mutation

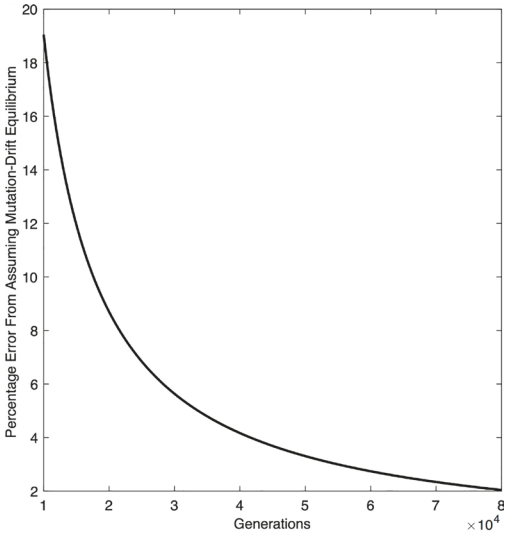

Fig. 4 - Illustration of how the percentage error from assuming mutation-drift equilibrium decreases with split time. The plot is based on neutral evolution with h^2= 0.4, σ_m^2 = 2 × 10^{-4}, V_{w_1} = V_{w_2} = 1.0, and $V_{w_{12}}$ = 0.2.

to multiple traits by replacing the variances with variance-covariance matrices (Lande, 1979, 1980). Matrices are rectangular (row and column) displays of numbers. A variance-covariance matrix is a square matrix (equal number of rows and columns) with variances in the cells along the top-left to bottom-right diagonal and covariances in remaining cells. Importantly, these multiple traits do not have to be traditional linear measurements; they could be the Cartesian coordinates of anatomical landmarks (x, y in 2-D or x, y, z in 3-D for each landmark). With multiple traits, Eq. (1) becomes

$$\mathrm{E}\{\mathbf{B}\} = \frac{h^2}{N_e}\mathbf{P}t_g \qquad (7)$$

where **B** is the between-group phenotypic variance-covariance matrix and **P** is the within-group phenotypic variance-covariance matrix. Eq. (7) assumes that **P** is proportional to the within-group additive-genetic covariance matrix, **G**, where the proportionality constant is h^2, which is often a reasonable assumption for morphological traits (Cheverud, 1988; Roff, 1996).

Similarly, with multiple traits, Eq. (5) becomes

$$\mathrm{E}\{\mathbf{B}\} = 2\mathbf{U}t_g \quad (8)$$

where **U** contains the additive genetic variances and covariances introduced by mutation (per zygote per generation). If we assume, as with Eq. (7), that **P** is proportional to **G**, then **P** is also proportional to **U**, because the equilibrium value of **G** is $2N_e\mathbf{U}$. Therefore, both Eq. (7) and Eq. (8) agree that $\mathrm{E}\{\mathbf{B}\} \propto \mathbf{P}$.

Building on these insights, Ackermann & Cheverud (2002) proposed two statistical tests to detect deviations from neutral divergence (see also Lofsvold, 1988; Ackermann & Cheverud, 2004; Marroig & Cheverud, 2004). For both tests, they decomposed **P** into its principal components (eigenvectors), projected the group means onto the principal components, and calculated the between-group variance along each of the principal components. For the first test, they regressed the log-transformed between-group variances on the log-transformed within-group variances (log-transformed eigenvalues) to evaluate whether or not the slope of the regression differed from one. For the second test, which can only be applied when there are more than two groups (ideally, many more), they evaluated whether the group mean scores along the principal components were correlated. Both of these tests are designed to detect deviations from $\mathrm{E}\{\mathbf{B}\} \propto \mathbf{P}$ with expectations under neutral divergence corresponding to a slope of one and no correlation respectively. Weaver *et al.* (2007) developed additional statistical tests, which considered the distribution expected under neutrality for the ratio of the between-group to the within-group variance along each of the principal components. Using these approaches, Ackermann & Cheverud (2004), Weaver *et al.* (2007), and Schroeder & Ackermann (2017) argued for the importance of neutral evolutionary processes in the evolution and diversification of the skull in the genus *Homo*. The implication of these studies is not that natural selection played no role but that it may have been less important than neutral evolutionary processes. Perhaps analogously, in recent humans, for which we have the most information to evaluate different explanations, there is evidence that natural selection and/or mechanical responses to shifts in diet shaped skull morphology (e.g, Carlson & Van Gerven, 1977), but the effects appear to be small or localized in comparison with neutral evolutionary processes mediated by population history (Katz *et al.*, 2016, 2017; von Cramon-Taubadel, 2017).

Coalescence-based approaches

Reformulating classic models

A major theoretical advance in population genetics, starting in the 1980s, was to reformulate classic problems in terms of coalescence (reviewed by Rosenberg & Nordborg, 2002). Instead of thinking about time in the typical way as advancing forward, coalescence-based approaches start at the present and work backwards in time (Fig. 5). It turns out that all of the classic models discussed here can be reformulated in terms of the average coalescence times of pairs of alleles (average pairwise coalescence times quantify how far back in time, on average, two alleles shared a common ancestral allele). For example, the expected within-group additive genetic variance under neutrality is

$$\mathrm{E}\left\{h^2 V_W\right\} = \tau_W \sigma_m^2 \quad (9)$$

and the expected between-group variance under neutrality is

$$\mathrm{E}\{V_B\} = 2\sigma_m^2(\tau_B - \tau_W) \quad (10)$$

where τ_W is the average coalescence time of pairs of alleles from the same group and τ_B is the average coalescence time of pairs of alleles from different groups (Whitlock, 1999; Weaver, 2016).

These coalescence-based expressions can be connected with classic results. For example, Eq. (4) can be derived from Eq. (9) by recognizing that $\tau_W = 2N_e$ at mutation-drift equilibrium, and Eq. (5) can be derived from Eq. (10) by recognizing that $\tau_B = t_g + \tau_W$ at mutation-drift equilibrium. To make another link with classic results, Eq. (6) can be derived from Eqs. (9) and (10) by recognizing that, when an ancestral group gives rise

to two descendant groups, $\tau_B = t_g + \tau_{W_{12}}$ where $\tau_{W_{12}}$ is the average coalescence time of pairs of alleles in the ancestral group (for more details see supplementary material for Weaver & Stringer, 2015). The key point is that with a coalescence-based approach, it is possible to predict expected amounts of within-group and between-group variation for a neutrally-evolving phenotypic trait by sampling – or simulating – neutrally evolving alleles from a collection of groups. All one needs is average coalescence times within and between groups, which increasingly will be possible to estimate as more DNA sequence data become available.

Gene flow

Until now, we have assumed that different groups evolve independently; that is, they never exchange migrants, so we can ignore gene flow. However, real groups often exchange migrants, particularly with adjacent groups. Recent genomic evidence for admixture between hominin lineages (reviewed by Wolf & Akey, 2018) has increased the relevance of models that consider gene flow to the evolution of human physical form. These models provide a theoretical basis for interpreting potential morphological evidence of admixture in the fossil record (e.g., Duarte *et al.*, 1999) as well as the results of hybridization experiments (e.g., Warren *et al.*, 2018). While it is possible to discuss gene flow using classic approaches (e.g., Rogers & Harpending, 1983; Relethford & Blangero, 1990; Lynch, 1988a; Lande, 1992), the coalescence-based approaches we have just discussed can be readily extended to incorporate gene flow.

Imagine four groups arranged in a square on the landscape. These groups have been exchanging migrants for many generations (i.e., an equilibrium state has been reached). The migration between groups is structured so that only adjacent groups exchange migrants, and the proportion of individuals exchanged each generation is $M/2$ (Fig. 6). Each group's population size (number of breeding individuals), N, is constant. This

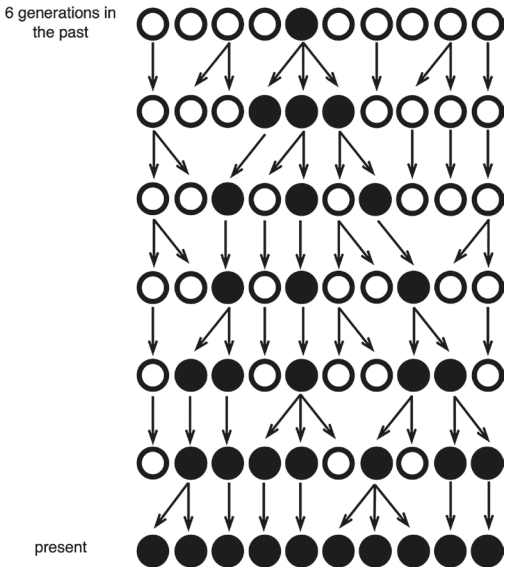

Fig. 5 - Coalescence example. Because of genetic drift, not every lineage has descendants in the next generation, so looking backwards in time, the present-day lineages will eventually share a single, common ancestral lineage (coalesce). In this example of coalescence in a population that is constant in size the ten present-day lineages coalesce six generations in the past. The filled circles indicate which lineages in each generation have at least one descendant in the present.

four-group model of gene flow is a simple, circular version of the stepping-stone model, which has a long history of investigation in population genetics (Kimura & Weiss, 1964; Slatkin, 1991). Under this model

$$\tau_{W(\text{step4})} = 8N \tag{11}$$

$$\tau_{B_{\text{adj}}(\text{step4})} = 8N + \frac{3}{2M} \tag{12}$$

$$\tau_{B_{\text{opp}}(\text{step4})} = 8N + \frac{2}{M} \tag{13}$$

where $\tau_{W\ (step4)}$ is the average coalescence time of pairs of alleles from the same group, $\tau_{B_{adj}\ (step4)}$ is the average coalescence time of pairs of alleles from adjacent groups, and $\tau_{B_{opp}\ (step4)}$ is the average coalescence time of pairs of alleles from opposite (diagonally across from each other) groups

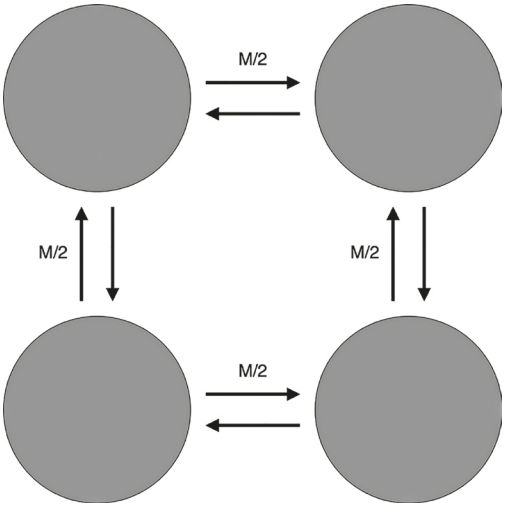

Fig. 6 - Model of gene flow among four groups. Four groups are arranged in a square on the landscape. The proportion of individuals exchanged each generation by adjacent groups, in each direction, is M/2.

(Slatkin, 1991). According to Eq. (11), the average coalescence time within each group increases linearly with group size. Eqs. (12) and (13) show that alleles from opposite groups (which indirectly exchange migrants) tend to coalesce deeper in the past than alleles from adjacent groups (which directly exchange migrants), and that the differences between within-group and between-group average coalescence times decrease as migration rate increases.

Eqs. (11)-(13) can be substituted into Eqs. (9) and (10), giving $8N\sigma_m^2$ for the within-group additive genetic variance, $\frac{3}{M}\sigma_m^2$ for the between-adjacent-group variance, and $\frac{4}{M}\sigma_m^2$ for the between opposite-group variance.

To generalize this four-group model, imagine d groups arranged in a circle, with d an even number, and $M/2$ the migration rate between adjacent groups. Under this more general model

$$\tau_{W(\text{step})} = 2Nd \qquad (14)$$

$$\tau_{B_i(\text{step})} = 2Nd + \frac{(d-i)i}{2M} \qquad (15)$$

where $\tau_{W(\text{step})}$ is the average coalescence time of pairs of alleles from the same group and $\tau_{B_i(\text{step})}$ is the average coalescence time of pairs of alleles from groups i steps apart (Slatkin, 1991).
Similarly to above, Eqs. (14) and (15) can be substituted into Eqs. (9) and (10), giving $2Nd\sigma_m^2$ for the within-group additive genetic variance and $\frac{(d-i)i}{M}\sigma_m^2$ for the between-group variance.
The circular stepping-stone model describes just one specific case of gene flow, but this example demonstrates how a coalescence-based approach can be used to find the expected within-group and between-group variances for a phenotypic trait when there is gene flow.

P_{ST} to F_{ST} comparisons

In population genetics, F_{ST} is a classic measuree of genetic differentiation. A paper by Relethford & Blangero (1990) – along with Relethford's freely-available RMET software – prompted anthropologists to estimate F_{ST} for many morphological traits collected on humans and non-human primates (e.g., Relethford, 1994; Relethford *et al.*, 1997; Schillaci & Froehlich, 2001; Leigh *et al.*, 2003; Roseman & Weaver, 2004; Roseman, 2004; Hanihara, 2008; Hubbe *et al.*, 2009; Smith, 2009; Weaver, 2014; Reyes Centeno *et al.*, 2017). In the biology literature (Prout & Barker, 1993; Spitze, 1993; Leinonen *et al.*, 2006), a morphological estimate of F_{ST} is designated Q_{ST} or P_{ST}, to distinguish it from a typical estimate, which comes from molecular data (here I will use P_{ST} for a morphological estimate; see Box 1 for further details). The molecular quantity measures genetic differentiation at the genetic loci themselves, and the morphological quantities measure genetic differentiation at the genetic loci underlying the phenotypic trait. One common approach is to compare P_{ST} for the phenotypic traits of interest to F_{ST} for presumably neutral DNA markers as a way to identify potential cases of directional selection ($P_{ST}>F_{ST}$)) or stabilizing selection ($P_{ST}<F_{ST}$) (e.g., Rogers & Harpending, 1983; Roseman & Weaver, 2004; Relethford, 2002; Weaver, 2014).

For a single phenotypic trait, according to Relethford & Blangero (1990), P_{ST} is estimated as

$$\frac{1}{d_s}\sum_{i=1}^{d}\frac{(\bar{x}_i-\bar{x})^2}{h^2V_W}\Bigg/\left(2+\frac{1}{d_s}\sum_{i=1}^{d}\frac{(\bar{x}_i-\bar{x})^2}{h^2V_W}\right) \quad (16)$$

where d_s is the number of groups that have been sampled, $\bar{x}_i$ is the trait mean for the i^{th} group, and $\bar{x}$ is the trait grand mean (mean of the group means). Relethford and colleagues (Relethford & Blangero, 1990; Relethford *et al.*, 1997; Relethford, 1994) present a version for multiple traits that depends on matrices rather than scalars and allows for unequal weighting of groups based on differences in effective population size. (Note that these authors refer to this quantity as F_{ST} in their papers.)

If the phenotypic trait is evolving neutrality, P_{ST} as given by Eq. (16) is expected to equal

$$\frac{\tau-\tau_W}{\tau} \quad (17)$$

where τ is the average coalescence time of pairs of alleles from the collection of groups sampled (Slatkin, 1995; Whitlock, 1999; Weaver, 2016). Eq. (17) connects P_{ST} to coalescence times, which allows expected P_{ST} under neutrality to be calculated for models of genetic drift, mutation, and gene flow. One simply needs to work out – analytically or by simulation – τ and τ_W, which allows for quite a bit of flexibility in the specifics of the models. This connection also makes it possible to show that, regardless of population structure or effective population size, morphological (P_{ST}) and molecular (F_{ST}) estimates of genetic differentiation are expected to be equal under neutrality (Whitlock, 1999; Weaver, 2016).

Relethford (1994, 2002) used a comparison of P_{ST} to F_{ST} at a global scale to argue for the importance of neutral evolutionary processes in shaping human cranial variation. Specifically, he found that when P_{ST} was calculated assuming an average heritability for cranial measurements based on clinical data (Devor, 1987), it corresponded closely with F_{ST} estimates. As discussed earlier in the context of Lynch (1989), this conclusion was somewhat surprising when it was published, given the general preference at the time for adaptive explanations for human cranial variation. Roseman (2004), Roseman & Weaver (2004), Smith (2009), and Hubbe *et al.* (2009) used similar approaches to identify the cranial traits or regions for which P_{ST} corresponded best with F_{ST}, and which traits were potentially affected by directional natural selection (e.g., cranial breadth, facial shape). These examples illustrate that even studies whose focus is on adaptation can make use of neutral models by using neutral expectations as a null hypothesis, which when it is rejected potentially indicates the action of natural selection. Although not precisely a comparison of P_{ST} to F_{ST}, comparing morphological to genetic distances to detect the action of natural selection (e.g., studies of human cranial variation by Harvati & Weaver, 2006a,b; von Cramon-Taubadel, 2009, 2011) is based on the same principle as a comparison of P_{ST} to F_{ST} – that under neutrality the amount of morphological and genetic differentiation should match.

Conclusions

Neutral models of phenotypic evolution have now become an important tool for researchers investigating the evolution of human physical form. These models have implications for many areas of biological anthropology, including using morphology to reconstruct the histories and migrations of recent human populations, using morphology to infer the evolutionary relationships among hominin taxa, and clarifying how natural selection has acted on physical form throughout human evolution. Their application to anthropological questions has stimulated biological anthropologists to more seriously consider the roles of history and chance in human evolution. These formal mathematical models have the added benefit of forcing researchers to explicitly incorporate their assumptions about the process of evolution into the analysis phase of research. To date, most insights from applying neutral theory to the evolution of human physical form have come from applications of classic

Box 1: Connections between F_{ST}, Q_{ST}, P_{ST}, and heritability.

Initially in the anthropology literature, both molecular and morphological estimates of genetic differentiation were designated F_{ST} (Relethford & Blangero, 1990; Relethford, 1994). In the biology literature, F_{ST} was always reserved for a molecular estimate, and a morphological estimate was designated Q_{ST} (Prout & Barker, 1993; Spitze, 1993). More recently, Leinonen *et al.* (2006) proposed that Q_{ST} should only be used for morphological estimates based on experimental data, for which additive genetic and other (dominance genetic, interaction genetic, environmental) effects can be separated by controlled breeding, and P_{ST} should be used for morphological estimates based on observational data, for which breeding is not controlled. Following this classification, the proper term for human morphological estimates would be P_{ST} (Roseman & Weaver, 2007; Reyes-Centeno *et al.*, 2014). Even so, currently, there is no consensus in the anthropological literature about which term to use for morphological estimates of genetic differentiation, but here I follow Leinonen *et al.* (2006) in using P_{ST} (see also Leinonen *et al.*, 2013).

A concern with morphological estimates derived from observational data (P_{ST}) is that variance that is not additive genetic could bias estimates, particularly in the direction of too much differentiation. If a large number of groups are being compared, Relethford & Blangero (1990)'s equation for estimating P_{ST} [Eq. (16)] can be rewritten as $V_B/(V_B + 2h^2V_W)$ (see Weaver, 2016). With the rewritten equation it is apparent that the within-group phenotypic variance is, in principle, adjusted for variance that is not additive genetic by scaling it by h^2, but the between-group phenotypic variance is not similarly adjusted. Therefore, if the average phenotypic trait values of two groups differ for reasons other than additive genetic effects (e.g., environmental effects from different diets), P_{ST} will be too large; it will be an inflated estimate of the amount of differentiation at the genetic loci underlying the trait.

To address this issue, some researchers have argued that V_B should be scaled by a factor c, giving $cV_B/(cV_B + 2h^2V_W)$ as a revised equation for estimating P_{ST}, and further that, as a first approximation, c should be set equal to h^2 (Brommer, 2011; Zaidi *et al.*, 2016). A similar assumption was implicitly made by Manica *et al.* (2007) when they compared coefficients of determination for regressions of within-group phenotypic (cranial) variance on group distance from sub-Saharan Africa to h^2. The justification for setting c equal to h^2 is that, without information to the contrary, it is reasonable to assume that the trait is equally heritable within and between groups (Brommer, 2011; Zaidi *et al.*, 2016). While this justification makes sense intuitively - notwithstanding the pitfalls of equating between-group with within-group sources of variation (Feldman & Lewontin, 1975) - it overlooks that V_W and V_B are fundamentally different in that V_W measures variation across individuals whereas V_B measures variation across group means. Considering the situation of two groups in which each individual's phenotype is the sum of additive genetic and other effects (i.e., no interaction terms), $c = \Delta_A^2/(\Delta_A + \Delta_O)^2$ where Δ_A is the difference between the groups in their additive-genetic-effect means and Δ_O is the difference between the groups in their other effect means. In contrast, for each group $h^2 = V_{W_A}/V_W$ where V_{W_A} is the within- group additive genetic variance. Unlike h^2, which will be less than one when any of the individual-level variation within a group is not additive genetic ($V_{W_A} < V_W$), c will only deviate from one if the groups differ in their other-effect means ($\Delta_O \neq 0$). In other words, individual-level variation that is not additive genetic will result in $h^2<1$, but only group-level differences for reasons other than additive genetic effects will result in $c<1$. Consequently, h^2 will not be a good proxy for c. Accordingly, while the presence of between-group variance that is not additive genetic can bias P_{ST} estimates, and this possibility should be evaluated whenever possible, there is no reason to assume that less biased estimates will be obtained by setting c equal to h^2.

approaches, but as more DNA sequence data become available, coalescence-based approaches promise to provide further insights by being able to readily and flexibly leverage these data.

Acknowledgments

I would like to thank D. Katz, N. von Cramon-Taubadel, and three anonymous reviewers for helpful comments on a version of this manuscript, and N. von Cramon-Taubadel for inviting me to write this paper.

References

Ackermann R.R. & Cheverud J.M. 2002. Discerning evolutionary processes in patterns of Tamarin (Genus *Saguinus*) craniofacial variation. *Am. J. Phys. Anthrop.,* 117: 260– 271.

Ackermann R.R. & Cheverud J.M. 2004. Detecting genetic drift versus selection in human evolution. *Proc. Natl. Acad. Sci. USA,* 101: 17946–17951.

Arsuaga J.L., Martínez I., Arnold L.J., Aranburu A., Gracia-Téllez A., Sharp W.D. *et al.* 2014. Neandertal roots: cranial and chronological evidence from Sima de los Huesos. *Science,* 344: 1358– 1363.

Beals K.L., Smith C.L. & Dodd S.M. 1983. Climate and the evolution of brachycephalization. *Am. J. Phys. Anthrop.,* 62: 425–437.

Betti L., Balloux F., Amos W., Hanihara T. & Manica A. 2009. Distance from Africa, not climate, explains within-population phenotypic diversity in humans. *Proc. R. Soc. Lond. B, Biol. Sci.,* 276: 809–814.

Betti L., von Cramon-Taubadel N., Manica A. & Lycett S.J. 2013. Global geometric morphometric analyses of the human pelvis reveal substantial neutral population history effects, even across sexes. *PLoS One,* 8: e55909.

Boas F. 1899. Anthropology. *Science,* 9: 93–96.

Brommer J.E. 2011. Whither P_{ST}? The approximation of Q_{ST} by P_{ST} in evolutionary and conservation biology. *J. Evol. Biol.,* 24: 1160–1168.

Carlson D.S. & Van Gerven D.P. 1977. Masticatory function and post-Pleistocene evolution in Nubia. *Am. J. Phys. Anthrop.,* 46: 495–506.

Carson E.A. 2006. Maximum likelihood estimation of human craniometric heritabilities. *Am. J. Phys. Anthrop.,* 131: 169–180.

Cavalli-Sforza L.L., Menozzi P. & Piazza A. 1994. *The history and geography of human genes*. Princeton University Press, Princeton.

Cheverud J.M. 1988. A comparison of genetic and phenotypic correlations. *Evolution,* 42: 958– 968.

DeGiorgio M., Jakobsson M. & Rosenberg N.A. 2009. Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proc. Natl. Acad. Sci. USA,* 106: 16057– 16062.

Devor E.J. 1987. Transmission of human craniofacial dimensions. *J. Craniofac. Genet. Dev. Biol.,* 7: 95–106.

Duarte C., Maurício J., Pettitt P.B., Souto P., Trinkaus E., van der Plicht H. *et al.* 1999. The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. *Proc. Natl. Acad. Sci. USA,* 96: 7604–7609.

Eller E. 1999. Population structure and isolation by distance in three continental regions. *Am. J. Phys. Anthrop.,* 108: 147–159.

Falconer D.S. & Mackay T.F.C. 1996. *Introduction to quantitative genetics*. Pearson, Harlow.

Feldman M. & Lewontin R.C. 1975. The heritability hang-up. *Science,* 190: 1163–1168.

Fenner J.N. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthrop.,* 128: 415– 423.

Goldstein D.B., Linares A.R., Cavalli-Sforza L.L. & Feldman M.W. 1995a. An evaluation of genetic distances for use with microsatellite loci. *Genetics,* 139: 463–471.

Goldstein D.B., Ruiz Linares A., Cavalli-Sforza L.L. & Feldman M.W. 1995b. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA,* 92: 6723–6727.

Grabowski M. & Roseman C.C. 2015. Complex and changing patterns of natural selection explain the evolution of the human hip. *J. Hum. Evol.,* 85: 94–110.

Guglielmino-Matessi C.R., Gluckman P. & Cavalli-Sforza L.L. 1979. Climate and the evolution of skull metrics in man. *Am. J. Phys. Anthrop.,* 50: 549–564.

Hanihara T. 2008. Morphological variation of major human populations based on nonmetric dental traits. *Am. J. Phys. Anthrop.,* 136: 169–182.

Harpending H.C. & Rogers A.R. 2000. Genetic perspectives on human origins and differentiation. *Annu. Rev. Genomics Hum. Genet.,* 1: 361–385.

Harvati K. & Weaver T.D. 2006a. Human cranial anatomy and the differential preservation of population history and climate signatures. *Anat. Rec. A Discov. Mol. Cell Evol. Biol.,* 288A: 1225–1233.

Harvati K. & Weaver T.D. 2006b. Reliability of cranial morphology in reconstructing Neanderthal phylogeny. In K. Harvati & T. Harrison (eds): *Neanderthals revisited: new approaches and perspectives*, pp. 239–254. Springer, New York.

Hayes B.J., Visscher P.M., McParthlan H.C. & Goddard M.E. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.,* 13: 635– 643.

Hlusko L.J. & Mahaney M. 2009. Quantitative genetics, pleiotrophy, and morphological integration in the dentition of *Papio hamadryas*. *Evol. Biol.,* 36: 5–18.

Hooton E.A. 1926. Methods of racial analysis. *Science,* 63: 75–81.

Howells W.W. 1973. *Cranial variation in man: a study by multivariate analysis of patterns of difference among recent human populations*. Harvard University, Cambridge.

Howells W.W. 1989. *Skull shapes and the map: craniometric analyses in the dispersion of modern* Homo. Harvard University, Cambridge.

Howells W.W. 1995. *Who's who in skulls: ethnic identification of crania from measurements*. Harvard University, Cambridge.

Hrdlicka A. 1908. Physical anthropology and its aims. *Science,* 28: 33–43.

Hubbe M., Hanihara T. & Harvati K. 2009. Climatic signatures in the morphological differentiation of worldwide human populations. *Anat. Rec. A Discov. Mol. Cell Evol. Biol.,* 292: 1720–1733.

Katz D.C., Grote M.N. & Weaver T.D. 2016. A mixed model for the relationship between climate and human cranial form. *Am. J. Phys. Anthrop.,* 160: 593–603.

Katz D.C., Grote M.N. & Weaver T.D. 2017. Changes in human skull morphology across the agricultural transition are consistent with softer diets in preindustrial farming groups. *Proc. Natl. Acad. Sci. USA,* 114: 9050–9055.

Kimura M. 1968. Evolutionary rate at the molecular level. *Nature,* 217: 624–626.

Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.

Kimura M. 1989. The neutral theory of molecular evolution and the world view of the neutralists. *Genome,* 31: 24–31.

Kimura M. & Maruyama T. 1971. Pattern of neutral polymorphism in a geographically structured population. *Genet. Res.,* 18: 125–131.

Kimura M. & Ohta T. 1971. Protein polymorphism as a phase of molecular evolution. *Nature,* 229: 467–469.

Kimura M. & Weiss G.H. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics,* 49: 561–576.

King J.L. & Jukes T.H. 1969. Non-Darwinian evolution. *Science,* 164: 788–798.

Konigsberg L.W. 2012. Quantitative variation and genetics. In S. Stinson, B. Bogin & D. O'Rourke (eds): *Human biology: an evolutionary and biocultural perspective*, pp. 143–173. John Wiley & Sons, Hoboken.

Lande R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution,* 30: 314–334.

Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution,* 33: 402–416.

Lande R. 1980. Genetic variation and phenotypic evolution during allopatric speciation. *Am. Nat.,* 116: 463–479.

Lande R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution,* 46: 381–389.

Langergraber K.E., Prüfer K., Rowney C., Boesch C., Crockford C., Fawcett K. *et al.* 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. USA,* 109: 15716–15721.

Lee S.H., Yang J., Goddard M.E., Visscher P.M. & Wray N.R. 2012. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics,* 28: 2540–2542.

Leigh S.R., Relethford J.H., Park P.B. & Konigsberg L.W. 2003. Morphological differentiation of gorilla subspecies. In A.B. Taylor & M.L. Goldsmith (eds): *Gorilla biology: a multidisciplinary perspective*, pp. 104–131. Cambridge University Press, Cambridge.

Leinonen T., Cano J.M., Mäkinen H. & Merilä J. 2006. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. *J. Evol. Biol.,* 19: 1803–1812.

Leinonen T., McCairns R.J.S., O'Hara R.B. & Merilä J. 2013. Q_{ST}-F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.,* 14: 179–190.

Li H. & Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature,* 475: 403–407.

Liu H., Prugnolle F., Manica A. & Balloux F. 2006. A geographically explicit genetic model of worldwide human-settlement history. *Am. J. Hum. Genet.,* 79: 230–237.

Lofsvold D. 1988. Quantitative genetics of morphological differentiation in *Peromyscus*. II. Analysis of selection and drift. *Evolution,* 42: 54–67.

Lynch M. 1988a. The divergence of neutral quantitative characters among partially isolated populations. *Evolution,* 42: 455–466.

Lynch M. 1988b. The rate of polygenic mutation. *Genet. Res.,* 51: 137–148.

Lynch M. 1989. Phylogenetic hypotheses under the assumption of neutral quantitative-genetic variation. *Evolution,* 43: 1–17.

Lynch M. 1990. The rate of morphological evolution in mammals from the standpoint of neutral expectation. *Am. Nat.,* 136: 727–741.

Lynch M. & Hill W.G. 1986. Phenotypic evolution by neutral mutation. *Evolution,* 40: 915–935.

Lynch M. & Walsh B. 1998. *Genetics and analysis of quantitative traits*. Sinauer Associates, Sutherland.

Manica A., Amos W., Balloux F. & Hanihara T. 2007. The effect of ancient population bottlenecks on human phenotypic variation. *Nature,* 448: 346–349.

Marroig G. & Cheverud J.M. 2004. Did natural selection or genetic drift produce the cranial diversification of neotropical monkeys? *Am. Nat.,* 163: 417–428.

Martínez-Abadías N., Esparza M., Sjøvold T., Gonzáles-José R., Santos M. & Hernández M. 2009. Heritability of human cranial dimensions: comparing the evolvability of different cranial regions. *J. Anat.,* 214: 19–35.

McGuigan K., Aguirre J.D. & Blows M.W. 2015. Simultaneous estimation of additive and mutational genetic variance in an outbred population of *Drosophila serrata*. *Genetics,* 201: 1239–1251.

Noonan J.P., Coop G., Kudaravalli S., Smith D., Krause J., Alessi J. *et al.* 2006. Sequencing and analysis of Neanderthal genomic DNA. *Science,* 314: 1113–1118.

Prado-Martinez J., Sudmant P.H., Kidd J.M., Li H., Kelley J.L., Lorente-Galdos B. *et al.* 2013. Great ape genetic diversity and population history. *Nature,* 499: 471–475.

Prout T. & Barker J.S.F. 1993. *F* statistics in *Drosophilia buzzatii*: selection, population size and inbreeding. *Genetics,* 134: 369–375.

Prüfer K., Racimo F., Patterson N., Jay F., Sankararaman S., Sawyer S. *et al.* 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature,* 505: 43–49.

Prugnolle F., Manica A. & Balloux F. 2005. Geography predicts neutral genetic diversity of human populations. *Curr. Biol.,* 15: R159–R160.

Ramachandran S., Deshpande O., Roseman C.C., Rosenberg N.A., Feldman M.W. & Cavalli-Sforza L.L. 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. USA,* 102: 15942–15947.

Rathmann H., Reyes-Centeno H., Ghirotto S., Creanza N., Hanihara T. & Harvati K. 2017. Reconstructing human population history from dental phenotypes. *Sci. Rep.,* 7: 12495.

Relethford J.H. 1994. Craniometric variation among modern human populations. *Am. J. Phys. Anthrop.,* 95: 53–62.

Relethford J.H. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. *Am. J. Phys. Anthrop.,* 118: 393–398.

Relethford J.H. 2007. The use of quantitative traits in anthropological genetic studies of population structure and history. In M.H. Crawford (ed): *Anthropological genetics: theory, methods and applications*, pp. 187–209. Cambridge University Press, Cambridge.

Relethford J.H. & Blangero J. 1990. Detection of differential gene flow from patterns of quantitative variation. *Hum. Biol.,* 62: 5–25.

Relethford J.H., Crawford M.H. & Blangero J. 1997. Genetic drift and gene flow in post-famine Ireland. *Hum. Biol.,* 69: 443–465.

Relethford J.H. & Harpending H.C. 1994. Craniometric variation, genetic theory, and modern human origins. *Am. J. Phys. Anthrop.,* 95: 249–270.

Reyes-Centeno H., Ghirotto S., Détroit F., Grimaud-Hervé D., Barbujani G. & Harvati K. 2014. Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proc. Natl. Acad. Sci. USA, 111:7248-7253.*

Reyes-Centeno H., Ghirotto S. & Harvati K. 2017. Genomic validation of the differential preservation of population history in modern human cranial anatomy. *Am. J. Phys. Anthrop.,* 162: 170–179.

Rice S.H. 2004. *Evolutionary theory: mathematical and conceptual foundations*. Sinauer Associates, Sunderland.

Rice S.H. 2008. A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evol. Biol.,* 8: 262.

Roff D.A. 1996. The evolution of genetic correlations: an analysis of patterns. *Evolution,* 50: 1392– 1403.

Rogers A.R. & Harpending H.C. 1983. Population structure and quantitative characters. *Genetics,* 105: 985–1002.

Roseman C.C. 2004. Detection of interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proc. Natl. Acad. Sci. USA,* 101: 12824–12829.

Roseman C.C. & Weaver T.D. 2004. Multivariate apportionment of global human craniometric diversity. *Am. J. Phys. Anthrop.,* 125: 257–263.

Roseman C.C. & Weaver T.D. 2007. Molecules versus morphology? Not for the human cranium. *Bioessays,* 29: 1185–1188.

Roseman C.C., Willmore K.E., Rogers J., Hildebolt C., Sadler B.E., Richtsmeier J.T. *et al.* 2010. Genetic and environmental contributions to variation in baboon cranial morphology. *Am. J. Phys. Anthrop.,* 143: 1–12.

Rosenberg N.A. & Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.,* 3: 380–390.

Runcie D.E. & Mukherjee S. 2013. Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices. *Genetics,* 194: 753–767.

Schillaci M.A. & Froehlich J.W. 2001. Nonhuman primate hybridization and the taxonomic status of Neanderthals. *Am. J. Phys. Anthrop.,* 115: 157–166.

Schroeder L. & Ackermann R.R. 2017. Evolutionary processes shaping diversity across the *Homo* lineage. *J. Hum. Evol.,* 111: 1–17.

Schroeder L., Roseman C.C., Cheverud J.M. & Ackermann R.R. 2014. Characterizing the evolutionary path(s) to early *Homo*. *PLoS One,* 9: e114307.

Schroeder L. & von Cramon-Taubadel N. 2017. The evolution of hominoid cranial diversity: a quantitative genetic approach. *Evolution,* 71: 2634–2649.

Slatkin M. 1991. Inbreeding coefficients and coalescence times. *Genet. Res.,* 58: 167–175.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics,* 139: 457–462.

Smith H.F. 2009. Which cranial regions reflect molecular distances reliably in humans? Evidence from three-dimensional morphology. *Am. J. Hum. Biol.,* 21: 36–47.

Sodini S.M., Kemper K.E., Wray N.R. & Trzaskowski M. 2018. Comparison of genotypic and phenotypic correlations: Cheverud's Conjecture in humans. *Genetics,* 209: 941–948.

Spitze K. 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics,* 135: 367–374.

Stojanowski C.M., Paul K.S., Seidel A.C., Duncan W.N. & Guatelli-Steinberg D. 2017. Heritability and genetic integration of tooth size in the South Carolina Gullah. *Am. J. Phys. Anthrop.,* 164: 505–521.

Tenesa A. & Haley C.S. 2013. The heritability of human disease: estimation, uses and abuses. *Nat. Rev. Genet.,* 14: 139–149.

Turelli M., Gillespie J.H. & Lande R. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. *Evolution,* 42: 1085–1089.

Visscher P.M., Hill W.G. & Wray N.R. 2008. Heritability in the genomics era – concepts and misconceptions. *Nat. Rev. Genet.,* 9: 255–266.

von Cramon-Taubadel N. 2009. Congruence of individual cranial bone morphology and neutral molecular affinity patterns in modern humans. *Am. J. Phys. Anthrop.,* 140: 205–215.

von Cramon-Taubadel N. 2011. The relative efficacy of functional and developmental cranial modules for reconstructing global human population history. *Am. J. Phys. Anthrop.,* 146: 83–93.

von Cramon-Taubadel N. 2014. Evolutionary insights into global patterns of human cranial diversity: population history, climatic and dietary effects. *J. Anthropol. Sci.,* 92: 43–77.

von Cramon-Taubadel N. & Lycett S.J. 2008. Brief communication: human cranial variation fits an iterative founder effect model with African origin. *Am. J. Phys. Anthrop.,* 136: 108–113.

von Cramon-Taubadel N. & Weaver T.D. 2009. Insights from a quantitative genetic approach to human morphological evolution. *Evol. Anthropol.,* 18: 237–240.

von Cramon-Taubadel N. 2017. Measuring the effects of farming on human skull morphology. *Proc. Natl. Acad. Sci. USA,* 114: 8917– 8919.

Warren K.A., Ritzman T.B., Humphreys R.A., Percival C.J., Hallgrímsson B. & Ackermann R.R. 2018. Craniomandibular form and body size variation of first generation mouse hybrids: a model for hominin hybridization. *J. Hum. Evol.,* 116: 57–74.

Washburn S.L. 1951. The new physical anthropology. *Trans. N. Y. Acad. Sci.,* 13: 298–304.

Weaver T.D. 2014. Brief communication: quantitative- and molecular-genetic differentiation in humans and chimpanzees: implications for the evolutionary processes underlying cranial diversification. *Am. J. Phys. Anthrop.,* 154: 615–620.

Weaver T.D. 2016. Estimators for Q_{ST} and coalescence times. *Ecol. Evol.,* 6: 7783– 7786.

Weaver T.D., Roseman C.C. & Stringer C.B. 2007. Were Neandertal and modern human cranial differences produced by natural selection or genetic drift? *J. Hum. Evol.,* 53: 135–145.

Weaver T.D., Roseman C.C. & Stringer C.B. 2008. Close correspondence between quantitative- and molecular-genetic divergence times for Neandertals and modern humans. *Proc. Natl. Acad. Sci. USA,* 105: 4645–4649.

Weaver T.D. & Stringer C.B. 2015. Unconstrained cranial evolution in Neandertals and modern humans compared to common chimpanzees. *Proc. R. Soc. Lond., B, Biol. Sci.,* 282: 20151519.

Whitlock M.C. 1999. Neutral additive genetic variance in a metapopulation. *Genet. Res.,* 74: 215–221.

Wolf A.B. & Akey J.M. 2018. Outstanding questions in the study of archaic hominin admixture. *PLoS Genetics,* 14: e1007349.

Zaidi A.A., Mattern B.C., Claes P., McEcoy B., Hughes C. & Shriver M.D. 2016. Investigating the case of human nose shape and climate adaptation. *PLoS Genetics,* 13: e1006616.

Zhivotovsky L.A. 2001. Estimating divergence time with the use of microsatellite genetic distances: impacts of population growth and gene flow. *Mol. Biol. Evol.,* 18: 700–709.

Zhivotovsky L.A. & Feldman M.W. 1995. Microsatellite variability and genetic distances. *Proc. Natl. Acad. Sci. USA,* 92: 11549–11552.

Editor, Noreen von Cramon-Taubadel