Genetic and Environmental Contributions to Variation in Baboon Cranial Morphology

Charles C. Roseman,[1]* Katherine E. Willmore,[2] Jeffrey Rogers,[3] Charles Hildebolt,[4] Brooke E. Sadler,[5] Joan T. Richtsmeier,[2] and James M. Cheverud[6]

[1]Department of Anthropology, University of Illinois, Urbana, IL 61801
[2]Department of Anthropology, Pennsylvania State University, University Park, PA 16802
[3]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78245-0549
[4]Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO 63110
[5]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85287-2402
[6]Department of Anatomy and Neurobiology, Washington University Medical School, St. Louis, MO 63110

ABSTRACT   The development, function, and integration of morphological characteristics are all hypothesized to influence the utility of traits for phylogenetic reconstruction by affecting the way in which morphological characteristics evolve. We use a baboon model to test the hypotheses about phenotypic and quantitative genetic variation of traits in the cranium that bear on a phenotype's propensity to evolve. We test the hypotheses that: 1) individual traits in different functionally and developmentally defined regions of the cranium are differentially environmentally, genetically, and phenotypically variable; 2) genetic covariance with other traits constrains traits in one region of the cranium more than those in others; 3) and regions of the cranium subject to different levels of mechanical strain differ in the magnitude of variation in individual traits. We find that the levels of environmental and genetic variation in individual traits are randomly distributed across regions of the cranium rather than being structured by developmental origin or degree of exposure to strain. Individual traits in the cranial vault tend to be more constrained by covariance with other traits than those in other regions. Traits in regions subject to high degrees of strain during mastication are not any more variable at any level than other traits. If these results are generalizable to other populations, they indicate that there is no reason to suppose that individual traits from any one part of the cranium are intrinsically less useful for reconstructing patterns of evolution than those from any other part. Am J Phys Anthropol 143:1–12, 2010.   © 2010 Wiley-Liss, Inc.

Cranial and dental characteristics play major roles in phylogenetic reconstruction in anthropology (Johanson and White, 1979; Tobias, 1980; Olson, 1981, 1985a,b; White et al., 1981; Kimbel et al., 1984; Skelton et al., 1986; Wood and Chamberlain, 1986; Chamberlain and Wood, 1987; Skelton and McHenry, 1992; McCollum and Ward, 1997; Strait et al., 1997; McCollum, 1999; Strait and Grine, 1999; Nevell and Wood, 2008). Identifying the most efficacious combination of methods and characters for reconstructing phylogenetic relationships enhances our understanding of the relationships between fossil hominin species, a major focus of paleoanthropological investigation (Gonzalez-Jose et al., 2008; Smith, 2009). Questions of character selection and definition are often approached by testing hypotheses about the intrinsic variational properties of characteristics, which are hypothesized to influence the evolutionary potential of characteristics such that they differentially reflect phylogenetic relationships among species (Lieberman, 1995; Lieberman et al., 1996; Collard and Wood, 2001, 2007; Wood and Lieberman, 2001; Lycett and Collard, 2005). The roles of development and constraints, because of integration among traits along with the possibility that different traits are exposed to different environmental stimuli, have also been an important focus of recent research into the variational and evolutionary properties of skeletal morphology (Lieberman, 1995; Lieberman et al., 1996, 2000a,b; Lockwood and Fleagle, 1999; Collard and Wood, 2001, 2007; Ponce de Leon and Zolikoffer, 2001; Wood and Lieberman, 2001; Lockwood et al., 2004;

Scott and Lockwood, 2004; Lycett and Collard, 2005; Lockwood, 2007; Smith et al., 2007). By using a model organism (the baboon) from a population of known pedigree, we are able to partition phenotypic variation into genetic and environmental components to test hypotheses about whether these criteria for selecting characteristics for phylogenetic analysis predict the patterns of phenotypic, genetic, and environmental variation that interact with the forces of evolution to produce differences between species.

*Correspondence to: Charles C. Roseman, Department of Anthropology, University of Illinois, 109 Davenport Hall, 607 S. Matthews Ave., Urbana, IL 61801. E-mail: croseman@uiuc.edu

Present address of Jeffrey Rogers: Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. E-mail: jeffrey.rogers@bcm.tmc.edu

Several reports have suggested that the study of the development and phenotypic integration of traits can identify those traits that most reliably reflect phylogeny and control for the confounding effects of interdependent traits (Lieberman, 1995; Lieberman et al., 1996; Collard and Wood, 2000, 2001; Gibbs et al., 2000; Strait, 2001; Wood and Lieberman, 2001; Gonzalez-Jose et al., 2008). It has been suggested that endochondrally ossifying parts of the cranium, such as the cranial base, reflect phylogeny well because of their early ossification, lack of impact by the strains of mastication, and their hypothesized limited exposure to environmental effects (de Beer, 1937; Scott, 1958; Olson, 1985a,b; Shea, 1985; Skelton and McHenry, 1992; Lieberman et al., 1996, 2000a,b; Strait, 1998; Wood and Lieberman, 2001). The cranial base is also hypothesized to constrain cranial evolution because of its numerous functional and developmental roles as the interface between the cranial vault and face and the remainder of the body, although the results bearing on this question are inconclusive (Lieberman et al., 2000a,b; McCarthy, 2001; Strait, 2001). More recent investigations in mice have shown that cranial base morphology, the cranial base angle in particular, is a function of a complex set of interactions between the cranial base and the effects of spatial packing and brain and face shape and size (Hallgrimsson and Lieberman, 2008; Lieberman and Hallgrimsson, 2008). Whether or not traits in the cranial base display strong integration is particularly important because of the prominent role of the cranial base in distinguishing hypothesized hominin taxa (Nevell and Wood, 2008).

Along similar lines, it has been suggested that more highly heritable characteristics are more useful for reconstructing phylogenetic relationships (Lieberman et al., 1996; Wood and Lieberman, 2001). This criterion posits that highly heritable characteristics have differences among species that appear early in development and are less likely to respond to exposure to environmental effects. It predicts that highly heritable traits should be less variable and less likely to mimic the character states of other species as a result. It is worth pointing out that these prediction do not follow directly from the evolutionary genetic definition of heritability and the justifications for these conjectures remain obscure (see discussion). Nevertheless, the amount of genetic covariance is an important factor for structuring phenotypic responses to evolutionary forces and there is mounting evidence that covariance among traits structures macro-evolutionary diversification among species (Cheetham et al., 1993; Schluter, 1996, 2000; Blows and Higgie, 2003; Marroig and Cheverud, 2004a,b, 2005; Estes and Arnold, 2007; Hunt, 2007; Marroig et al., 2009). This means that measures reflecting the genetic contributions to variation in traits, such as their heritability, should be considered in combination with the ways in which forces of evolution may have acted on phenotypes when asking macroevolutionary questions. A trait that has evolved in a manner such that its contrasting states provide a phylogenetically informative difference must have resulted from the conversion of within-species genetic variation into among-species variation through natural

selection, mutation, and random genetic drift, and as such must have been heritable at some point. This does not mean, however, that the same trait must be heritable within any of the species under study, only that it was heritable at some time.

The potential for environmental effects to cause homoplasy or homoiology is the target of a great deal of scrutiny in biological anthropology (e.g., Lieberman, 1995; Lieberman et al., 1996, 2000a,b; Lockwood and Fleagle, 1999; Wood and Lieberman, 2001; Lycett and Collard, 2005; Collard and Wood, 2007; von Cramon-Taubadel, 2009). This is largely motivated by the observation that traits related to mastication may cause problems when attempting to reconstruct phylogeny because of the functional importance of chewing (Skelton et al., 1986; Walker et al., 1986; Walker and Leakey, 1988; Strait et al., 1997; Skelton and McHenry, 1998; Strait and Grine, 1998). The homoiology hypothesis includes two distinct predictions. The first prediction is that characteristics that are subject to high strains have plastic responses that result in higher levels of variation. The second predictions is that these plastic responses and concomitantly higher levels of variation make them less useful for phylogenetic reconstruction because of their higher probability of approximating the character states of other species for purely environmental reasons.

The homoiology prediction about the relative phylogenetic utility of traits has been soundly refuted as characters that are subject to higher strain do not appear to be any less useful for phylogenetic reconstruction (Collard and Wood, 2001, 2007; Lycett and Collard, 2005; Collard and Lycett, 2008; von Cramon-Taubadel, 2009). The second prediction about variational properties of different traits, however, has not been falsified, as more highly strained characteristics in the cranium appear to also show higher levels of variation relative to characteristics that are not subject to high levels of strain when the effects of sex are taken into account (Lycett and Collard, 2005; von Cramon-Taubadel, 2009). Removing the effects of sex is important as highly strained characteristics in the cranium show considerable overlap with those showing the most sexual dimorphism (Leigh and Cheverud, 1991; Plavcan, 2003; Willmore et al., 2009).

The variational predictions of the homoiology hypothesis have been general, only including the claim that high strain imparts higher variation on characteristics. An important distinction missing from this discussion is that between the mean level of an environmental effect, such as strain imparted through mastication, and the variance in that environmental effect. Although the variational prediction of the homoiology hypothesis is framed in terms of the mean effect of strain, there is no good *a priori* reason to believe that bony phenotypes may be more variable in a high strain environment as opposed to a low strain environment (see discussion). Should the strength of an environmental effect vary across individuals, however, some may exhibit larger deviations from the population mean than others. This variation in the environmental effect may occur at the lower end of the spectrum of effects, however, which means that there is no reason to suspect that simply being exposed to a stronger mean environmental influence results in higher phenotypic variance.

To illustrate this point, imagine a situation with several populations with identical genotypic frequencies (Represented by state A in Fig. 1). All else held equal, the means and variances of a trait are identical across

---
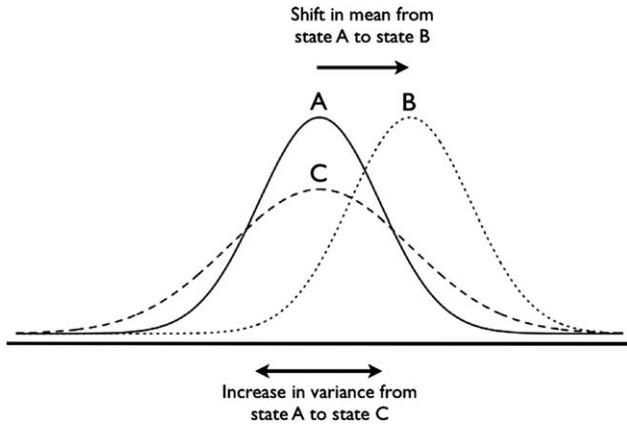
Fig. 1. A hypothetical example of the distribution of phenotypes in three populations with identical genotype frequencies under three different environmental regimes. Curve A represents a baseline population in an arbitrarily normative environment. Curve B represents a population in which every individual has been subject to environmental influence greater than that in A, thus increasing the mean. All individuals are exposed to the exact same dose of the environmental effect. Curve C represents a situation where the mean environmental effect is the same as in population A but the strength of the effect varies across individuals. This changes the variance without affecting the mean.

all of the groups. Likewise, the heritability of the trait is identical in all of the groups. Imagine taking one of the populations and subjecting every individual to an environmental factor such as strain from mechanical forces (Contrast state A with state B in Fig. 1). In the case of bony phenotypes, there is often a tendency for bones to exhibit a plastic response to biomechanical forces (Trinkaus et al., 1994; Lieberman et al., 2003; Ruff et al., 2006; Ravosa et al., 2007; Menegaz et al., 2009). If all individuals are subjected to an identical level of this stronger environmental factor and the degree to which individuals respond to the change in the factor does not vary, then the mean of the affected population may shift, resulting in a difference between the affected group and its unaffected counterparts. This effect has been used to study the behavior in past populations (Trinkaus et al., 1994; Ruff et al., 2006) and could result in a shift in the mean value of a trait in one lineage such that it approximates the character states of another group, thus confounding phylogenetic reconstruction.

Contrast this with a situation where individuals within one of the groups are differentially affected by varying levels of the environmental factor but the mean strength of the factor is the same as its level in the other populations in which it does not vary (State A vs. state C in Fig. 1). This could result in a situation in which the mean of the trait does not shift but the variance in the trait increases. This would result in an increase in a coefficient of variation (CV), as predicted by Wood and Lieberman 2001 and Lycett and Collard 2005, and could result in a situation where some individuals at one of the tails of the distribution in one species might approximate the character states of another species, which, in combination with a sparsely sampled fossil record, could cause difficulties for phylogenetic reconstruction and taxonomic diagnosis, but in a different manner than in the first example.

There are a large number of possible alternatives to consider here. For example, one way in which both the

mean and variance in a trait could change together is a situation in which there is genetic variance in the response of individuals to a change in the environmental factor (Pigliucci, 2001). If there is genetic variance in the extent to which some individuals form more bone than others in response to the change in the mean of the factor through genotype by environment interactions (keeping the variance of the factor the same), then both the mean and the variance of a trait changes as a result of a plastic response in the form of a shift in the mean and the unmasking of *hitherto* hidden genetic variation.

It is important to emphasize that whether or not environmental or genetic variational properties of a trait change is an empirical issue and there is no good theoretical reason to expect one outcome or another independent of the particulars of the traits, populations, and environments (Pigliucci, 2001). Furthermore, the degree of heritability of a trait manifest in one environment is no indicator of how plastic it is when expressed in a population facing a novel environmental condition (Feldman and Lewontin, 1975; Lewontin, 2000; Pigliucci, 2001).

The homoiology hypothesis is usually phrased such that it focuses on the elevated mean level of strain on some traits and not elevated variance in the effect of these forces on these traits. The intended message of the authors (reinforced by comments from an anonymous reviewer), however, often appears to be that it is an increase in variance in strain, made possible by increasing the mean exposure to strain, that drives the differences among traits (Wood and Lieberman, 2001; Lycett and Collard, 2005; Collard and Lycett, 2008; von Cramon-Taubadel, 2009). If negative values of measures of strain are disallowed, this opens up the opportunity for the raw variance to be larger for traits with higher mean values because they have more room to vary, much in the same way that linear traits tend to have variances that positively scale to the mean (Hansen and Houle, 2008). If values of strain are measured on an interval scale where both negative and positive values can occur, the relationship between the mean value and the variance becomes complicated. We entertain both the mean and variance versions of the homoiology hypothesis here.

An important difference between the present study and previous studies looking into the homoiology hypothesis and most studies about the variational properties of primate cranial features is that the animals used in this study come from a captive population. The animals used here do not face the same hazards as animals in the wild and it is probably reasonable to assume that an animal in captivity is not exposed to the same range of environments as wild animals, which may serve to diminish environmental variance caused by variance in exposure to strain.

There are two things that need to be emphasized here. The animals used in this study do chew on hard objects and show varying degrees of tooth wear, even within age groups (personal observation), suggesting that there is some variance in the degree to which mechanical forces impact individuals. Whether this is enough to appreciably inflate phenotypic variance is unknown. The other consideration is the fact that there are several potentially confounding sources of variance that cannot be accounted for in many wild-shot museum collections. For example, extra phenotypic variation may arise in a wild population in response to differences in nutritional status, the effects of age, and the effects of undetected pop-

ulation structure within species. Furthermore, the distributions of exposures to strain in wild populations are not known. As a result, neither the captive sample here nor the samples of wild individuals used in previous studies are ideal to address the homoiology question. The advantage of the pedigreed population used here is that it affords us the opportunity to explore the genetic variational properties of traits.

Whether or not a characteristic is able to evolve differences between species that can be used for phylogenetic reconstruction depends both on the evolvability of a trait and the forces of evolution that act on that trait and traits with which it genetically covaries. There are several meanings of evolvability that have different applications to evolution at different timescales (Houle, 1992; Wagner and Altenberg, 1996; Hansen and Houle, 2004, 2008; Wagner and Laubichler, 2004; Jones et al., 2007; Lynch, 2007a,b). Evolutionary genetic theory suggests several different ways in which researchers may measure the differential propensity for traits to evolve over the short term (Houle, 1992; Hansen and Houle, 2008). Given that many phylogenetic analyses in paleoanthropology use discretized continuous traits (Strait and Grine, 2004; Gonzalez-Jose et al., 2008) and that the differences exhibited among species are usually matters of degree rather than of kind, using an evolutionary quantitative genetic definition of evolvability seems most appropriate in this context rather than an evolutionary developmental definition.

From an evolutionary genetic perspective, heritability is not very useful for comparing the relative evolvability of traits or the degree to which a trait may be affected by environmental factors. A trait could be highly heritable because it: 1) has a large amount of genetic variance or 2) has a very small amount environmental variance (Houle, 1992). Because h^2 is a dimensionless ratio of additive genetic (V_A) to phenotypic (V_P) variances ($h^2 = V_A/V_P$. Falconer and Mackay, 1996), h^2 gives us no sense of the expected magnitude of variation at the phenotypic level and does not tell us whether the genetic or environmental components of variation are elevated. These sources of variation are important to tease apart when evaluating the criteria for trait selection in phylogenetic analysis, as models of constraint predict a dearth of genetic variation when a trait is constrained (Hansen and Houle, 2004) and the homoiology hypothesis predicts that the environment differentially affects the traits. As such, separate estimates of phenotypic, genetic, and environmental variation are desirable when confronting these problems.

Furthermore, h^2 is not the best measure of the evolutionary potential for a trait. Results from evolutionary quantitative genetic theory suggest that the coefficient of additive genetic variation ($CV_A = \sqrt{V_A}/\mu$) or its square ($e = V_A/\mu^2$), where μ is the trait mean, would be a better measure of evolvability than h^2 (Houle, 1992; Hansen and Houle, 2004, 2008). The latter measure of evolvability, e, is the expected proportional change in a trait per unit of a mean-standardized selection gradient. The operation of developmental constraint could cause low e by canalizing a trait such that mutation is unlikely to generate variation, or internal stabilizing selection may rid the trait of most new variation. Likewise low e could be the result of stabilizing selection resulting from variation in the external environment that may limit the amount of genetic variation in the population. In the case of morphological characteristics where V_P is often a

good predictor of V_A (Cheverud, 1988; Roff, 1996), e has a straightforward relationship with the square of the phenotypic coefficient of variation (CV^2), allowing for studies of the relative evolvability of traits within a species to be conducted using phenotypic data alone. We can test whether or not this is the case here by comparing additive genetic ($\mathbf{G}$) and phenotypic ($\mathbf{P}$) covariance matrices and their correlation counterparts ($\rho\mathbf{G}$ and $\rho\mathbf{P}$) to see if $\mathbf{P}$ is proportional to $\mathbf{G}$ and, thus, if tests at the phenotypic level would be adequate to test hypotheses about patterns of genetic covariance. This is important for testing hypotheses in comparative and fossil contexts where quantitative genetic parameters cannot be estimated.

Phenotypic integration among traits may result in a situation where some of the V_A of any given phenotype may be shared with other traits through genetic covariance (Hansen et al., 2003; Hansen and Houle, 2008). Those traits with higher levels of intrinsic conditional additive genetic variation or conditional evolvability (c), after accounting for the additive genetic covariance with other traits, are more conditionally evolvable. Traits that share much of their variance in common with other traits (low c relative to e) may be constrained by evolutionary forces acting on other traits, and evolution of that characteristic is likely to cause correlated responses in other characteristics. Individual characteristics that have low c are also more likely to violate assumptions about character independence in phylogeny reconstruction than those with higher c values. If traits in one region of the cranium have lower c values relative to their e values than traits in other areas, those traits are subject to stronger pleiotropic constraints than the others and each one contributes less independent information for phylogenetic reconstruction. The parameterizations of e and c suggest a measure of integration—the tendency for traits to covary and correlate with one another—in the form $i = 1 - (c/e)$, which expresses the proportion of all evolvability of a trait that is accounted for by covariance with other traits. Traits with i values close to 1 have less potential to evolve independently of other traits.

If environmental forces act to make some traits more variable than others then we expect these traits to differ in terms of their intrinsic environmental variation. This can be measured in a manner similar to the way in which evolvability is measured by calculating the mean-square standardized environmental variance (e_e). Should traits that ossify earlier, for example, be subject to fewer sources of environmental variation than later ossifying traits, we expect the earlier ossifying traits to have lower e_e values than the late ossifying traits.

These different ways of modeling and measuring genetic and environmental sources of phenotypic variation allow us to construct more precise hypotheses about how developmental and pleiotropic constraints and the effects of environmental factors affect the propensity of individual traits to evolve and, thus, their utility in evolutionary investigations. Using crania from a pedigreed sample of baboons, we seek to answer a series of questions about the genetic and environmental bases of phenotypic variation for a variety of traits in the cranium (Fig. 2, Table 1). Do traits in different functionally and developmentally defined regions of the cranium vary in their levels of e_e, h^2, and e? Do traits in some functionally and developmentally defined regions tend to have higher levels of c or i than traits in other regions, thus
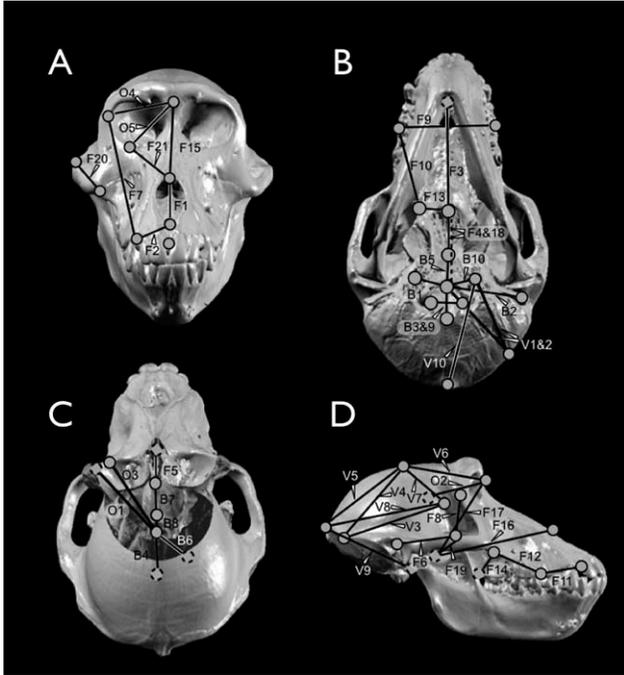
Fig. 2. Anterior (**A**), inferior (**B**), superior cutaway (**C**), and right lateral (**D**) views of the baboon skull with interlandmark distances used in this analysis marked using lines, and landmarks with open circles. Circles with solid boundaries represent landmarks in their natural position and circles with dashed boundaries represent landmarks obscured by bone indicating that the trait passes behind the visible bone to the hidden landmark. The letter that begins each trait label corresponds to the groupings listed in Table 1 and the number corresponds to the number of that trait within that grouping. Landmarks indicated on the skull (in alphabetical order with their abbreviations) are: Anterior nasal spine (ans), Asterion (as), Basion (bas), Bregma (breg), Carotid (car), Crista galli anterior (cga), Crista galli posterior (cgp), External auditory meatus (eam), Foramen magnum lateral (fm), Frontal-zygomatic juncture (fzj), Jugular process (jug), Lambda (lam), Midpoint between the first molar and last premolar on the alveolar margin (mda), Maxillary tuberocity (mxt), Nasale (nal), Nasion (nas), Opisthion (opi), Premaxillary maxillary boundary (pmm), Posterior nasal spine (pns), Pterion (pt), Sella turcica (sella), Vomer-sphenoid juncture (vsj), Zygomatic arch superior (zas), Zygomaxillary inferior (zmi), and Zygomaxillary superior (zms). Interdentale superior (ids) is also marked as it is referred to in the text even though no traits include this landmark.

indicating that some traits are more constrained than others? Are phenotypic and genetic correlations and covariances similar to one another so that studies of phenotypic variation can be used as surrogates for studies of quantitative genetic variation? Are characteristics that are subject to higher levels of masticatory strain more variable, less heritable, more environmentally variable, and less evolvable than traits subject to less strain? The answers to these questions have clear implications for testing hypotheses about evolution and estimating phylogenies in fossil humans and in other primates.

MATERIALS AND METHODS

Crania from baboons (*Papio hamadryas*) from the pedigreed population housed at the Southwest National Primate Research Center (SNPRC), San Antonio, TX were

TABLE 1. Trait definitions and their hypothetical modular groupings

Base (B)	Face (F)	Orbit (O)	Vault (V)	High strain
1: bas_car	1: ans_nal	1: cgp_fzj	1: ast_bas	ans_pmm
2: bas_eam	2: ans_pmm	2: cgp_nas	2: ast_jug	ans_pns
3: bas_opis	3: ans_pns	3: cgp_zms	3: ast_pt	fzj_pmm
4: bas_sella	4: bas_pns	4: fzj_nas	4: breg_as	fzj_zas
5: bas_vsj	5: cga_nas	5: zms_nas	5: breg_lam	mda_mda
6: car_sella	6: eam_zas		6: breg_nas	mda_mxt
7: cga_cgp	7: fzj_pmm		7: breg_pt	mda_pmm
8: cgp_sella	8: fzj_zas		8: lam_pt	mda_zmi
9: fm_fm	9: mda_mda		9: lam_bas	mxt_pns
10: jug_bas	10: mda_mxt		10: lam_jug	mxt_zmi
	11: mda_pmm			zas_zmi
	12: mda_zmi			
	13: mxt_pns			
	14: mxt_zmi			
	15: nal_nas			
	16: nal_vsj			
	17: nas_vsj			
	18: pns_vsj			
	19: zas_vsj			
	20: zas_zmi			
	21: zms_nal			

The traits listed in the base, face, vault, and orbit categories together contain all traits used in the analysis and each trait is accompanied by a number for use in conjunction with Figure 2. The high strain category contains a subset of these, and those not included in it are low to moderate strain characters.

used to estimate quantitative genetic parameters. The number of pedigreed animals ranged from 364 to 410, depending on the trait, with a mean of 382. The large majority of the baboons in this population were purebred olive (*P.h. anubis*), although there were a few yellow baboons (*P. h. cynocephalus*) and hybrids between the two groups. The founding population was composed of roughly 300 individuals and there is substantial genetic variation in the population (Rogers et al., 2000). These animals have been used to successfully study the quantitative genetics of craniodental characteristics (Hlusko and Mahaney, 2003; Hlusko et al., 2004, 2006, 2007; Sherwood et al., 2008). Any animal in the collection that was under six years of age was excluded from the analysis.

Heads were collected at necropsy, macerated in a water-bath, and air dried in a fume hood. Three-dimensional CT scans were acquired of each cranium (Siemens Medical Systems, Erlanger, Germany). Pixel size varied from 0.2 to 0.61 mm, as a function of the size of the individual specimen, and slice thickness was 0.75 mm. Two observers scored a total of 43 landmarks divided into different sets on the face, cranial base, and cranial vault (Fig. 2). A set of 46 interlandmark linear distances were calculated from the coordinates (Table 1), although no distances were used that included interdentale superior (ids) as it was regularly obscured because of the resorption of bone associated with incisor loss. Landmarks located on the base and face were digitized using the CT scans and the eTDIPS software package (National Institutes of Health and The National University of Singapore). Because some points on the cranial vault could not be reliably identified from the scans, the locations of landmarks on the vault and a subset of those on the facial skeleton and cranial base were recorded directly from the crania using a Microscribe digitizer (Immersion, San Jose, CA; Fig. 2). To acquire

the landmarks on the cranial vault, modeling clay was used to reattach the calotte to the cranium. All individuals were landmarked twice and repeatabilities (Falconer and Mackay, 1996) of the Euclidean distances among landmarks were very high (mean repeatability = 0.98). More individuals were analyzed for the vault set than other sets because of differences in data acquisition schedules.

Quantitative genetic parameter estimation was conducted using maximum likelihood in the SOLAR quantitative genetics computer package (Almasy and Blangero, 1998). We estimated h^2; the proportion of environmental variance, e^2, and their associated standard errors and estimates of their statistical significance. The covariates sex, age, and age^2 were screened and corrected for when significant at the $P = 0.10$ level. The effective sample sizes (N_{eff}) of the traits were estimated using the procedure in Cheverud (1995). Before estimating ρP, ρG, and ρE, we reduced the set of traits by removing those with a N_{eff} smaller than 10. This was done to avoid producing biased estimates of squared genetic correlations for traits carrying little genetic information (Cheverud, 1995). All traits were included in the univariate analyses, whereas multivariate analyses dependent on an estimate of G were conducted on the reduced trait set.

We estimated ρP and P to determine whether they could be reliably used as surrogates for ρG and G. We calculated estimates of e, e_e, and CV to compare different kinds of variation across regions of the cranium. We use the CV rather than its square to evaluate intrinsic phenotypic variation because it is the standard statistic used in other tests of the homoiology hypothesis and similar studies dealing with the variational properties. All mean standardized statistics reported here were calculated using covariate adjusted means. In spite of substantial sexual dimorphism in this population, the cross-sex genetic correlations for most traits in this sample are high and indistinguishable from unity, in most cases, rendering results in one sex directly proportional to the other (Willmore et al., 2009). Furthermore, results using the single sex means were highly correlated with each other and with the results from the pooled mean.

To address the hypothesis of constraint and interdependence among traits, we calculated the squared-mean standardized conditional additive genetic variance, c (Houle, 1992; Hansen et al., 2003; Hansen and Houle, 2008). The c statistic quantifies the free or independent variation of a trait that is not shared with any other traits through covariance. To compare conditional genetic variation at the phenotypic and genetic levels, we calculated a phenotypic equivalent to c (c_P, using P as a substitute for G). Because G could not be estimated using all traits owing to the small N_{eff} of a subset of the characteristics, c_P was also estimated for the reduced trait set (c_{PR}) substituting the P for the reduced set of traits for G to make it strictly comparable with c. Furthermore, because G is singular in this sample, a generalized inverse was used in the estimation of the c values in different traits. This introduces some degree of error into the estimation of c values, making the comparison with their phenotypic counterparts, c_P and c_{PR}, all the more important. Likewise, we estimate i and its phenotypic counterpart i_p, to test if the relative magnitude of integration is different across regions of the cranium.

To estimate the repeatability of each of the covariance matrices, we employed a parametric bootstrap coupled with a random skewers procedure (Cheverud et al., 1983; Cheverud and Marroig, 2007). Matrix repeatability for the correlation matrices was assessed by Mantel tests between the ρG estimated from the real data and one thousand ρG estimates simulated following the procedures for estimating the repeatability of G and P (Cheverud et al., 1983). To estimate matrix repeatability for ρG and G, we used a matrix bending procedure that uses the N_{eff} of the traits as weights to make the matrices positive definite so that they could be resampled (Jorjani et al., 2003). To test the similarity of P, G, and E, we used the random skewers procedure as outlined above. For the correlation counterparts of the three matrices, we used a Mantel test with significance assessed by 1,000 random permutations of the rows and columns of the matrices. To test the hypothesis that h^2 was distributed in a manner one would expect if it represented random deviations around a single expectation, we tested the fit of the distribution of h^2 estimates to a normal distribution using a Shapiro-Wilk test (Sokal and Rohlf, 1995). We tested the equality of the variance of the variances of the h^2 estimates and the mean sampling variance (the square of the standard error) to see if the dispersion of the distribution was what one expected, given the uncertainty of estimating h^2 in this population using an F-test (Sokal and Rohlf, 1995).

To test the hypothesis that functional and developmental classifications of traits predict the variational properties of their constituent traits, we divided the traits into regions defined by their functional and developmental roles in forming the cranium (Cheverud et al., 1983) (Table 1). Traits were assigned to the face, orbit, vault and base, and we evaluated groupings contrasting traits in the face/orbit and base/vault, the face/orbit, base, and vault, and the face, orbit, base and vault to test hypotheses about the distribution of genetic and environmental variation across the cranium. We used ANOVA (Sokal and Rohlf, 1995) to test the hypothesis of no difference between trait groups for levels of phenotypic, genetic, and environmental variation. Mean or mean square standardized parameters tend to be log-normally distributed in continuous biological traits (Lycett and Collard, 2005). We log (base e) transformed the values of these estimates before applying the test to make the data better fit the distributional assumptions of the ANOVA.

To test the homoiology hypothesis, an additional hypothesis was done that contrasted a set of traits shown to be subject to high levels of strain in experimental strain-gage studies (i.e., traits describing the palate, alveolus surrounding the dentition, and the zygoma) with a set of traits from the remainder of the cranium that experience low to moderate strain (Currey, 1984; Martin et al., 1998; Lycett and Collard, 2005; von Cramon-Taubadel, 2009). We used the same ANOVA procedure outlined above for testing hypotheses of differential variation across these regions. All statistical analyses were performed in R, JMP 5.01 (SAS Institute, Inc., Cary, NC), or by hand.

RESULTS

All but two traits had h^2 values that were significantly different from zero at the 0.05 level. Heritability estimates fit a normal distribution well (Shapiro-Wilk $W = 0.983$, $P = 0.855$) and were moderate (mean $h^2 = 0.44$, Supporting Information Table S1). The variance of

TABLE 2. A comparison of genetic, environmental, and phenotypic associations among traits

	ρ**G/G**	ρ**E/E**	ρ**P/P**	Covariance repeatability
ρ**G/G**	–	0.309 ($P = 0.029$)	0.833 ($P = 1.49\mathrm{e}^{-7}$)	0.719
ρ**E/E**	-0.184 ($P < 0.001$)	–	0.741 ($P = 2.58\mathrm{e}^{-6}$)	0.720
ρ**P/P**	0.669 ($P < 0.001$)	0.524 ($P < 0.001$)	–	0.968
Correlation repeatability	0.604	0.531	0.918	

Scores for comparisons of covariance matrices are above the diagonal whereas scores for comparisons of correlation matrices are below the diagonal. Scores of matrix association are expressed as raw correlation coefficients whereas repeatabilities are expressed as coefficients of codetermination.

the distribution of h^2 estimates was not different from the mean of the distribution of the sampling variances for the individual h^2 estimates (F-test. $F_{43\mathrm{d.f.}/44\mathrm{d.f.}} = 1.404$, $P = 0.131$). Seven traits did not make the minimum cutoff of $N_{\mathrm{eff}} > 10$ and were excluded from multivariate analyses (Supporting Information Table S1). Matrix repeatabilities for **P** and ρ**P** were high, in excess of 90% for both variance/covariance and correlation matrices (Table 2). Genetic and environmental matrices had lower repeatabilities owing to the greater error involved in their estimation. The repeatabilities of matrices give the maximum correlation that can be realized in a matrix comparison so that the magnitude of observed associations should be considered relative to their repeatabilities, not 1.0. There were significantly positive correlations between ρ**P** and ρ**G**, and ρ**P** and ρ**E** but a significant negative correlation between ρ**G** and ρ**E** (Table 2), which is likely the result of strong negative covariance between estimates of genetic and environmental correlations (Cheverud, 1995). Mean random skewers scores among **E**, **G**, and **P** were all significantly positive. The values of the mean random skewers statistics are all only slightly lower than their repeatability values indicating that the correlations of the responses to selection of the **G** vs. **P** and **E** vs. **P** are very nearly as high as could be expected if they were proportional matrices given the estimation error of the elements of **G** and **E**.

None of the univariate measures of genetic, environmental, and phenotypic variation (h^2, e, e_{e}, and CV) differed significantly among any of the trait groupings (Table 3). This contrasts with previous investigations that accounted for sex differences when estimating the level of intrinsic variation where traits that were in regions subjected to large amounts of strain tended to be more intrinsically variable (Lycett and Collard, 2005; von Cramon-Taubadel, 2009). Most importantly, the environmental component of variation was not elevated in the high-strain region relative to the rest of the cranium. Measures of the average degree to which individual characteristics were independent of or integrated with other characteristics (c, c_{B} c_{PR}, i, and i_{p}) were different across anatomically defined regions when controlling for multiple comparisons (four tests for each parameter yielding a Bonferroni-adjusted significance threshold of 0.0125) in the case of c, c_{P}, c_{PR}, i, and i_{p}, although c_{PR} in the case of the model comparing the face, base, vault, and orbit did not pass the adjusted threshold (Table 2). *Post-hoc* Tukey–Kramer HSD tests indicate that these results are largely driven by traits in the vault being less conditionally evolvable (lower c, c_{P} and c_{PR}) and more highly integrated (i and i_{p}) in the vault as compared to the remainder of the regions.

One of the measures of c_{p} and i_{p} revealed a significant difference among trait groups at the 5% level in the high

vs. low strain grouping model (Table 3). The difference was not significant, however, when multiple comparisons were considered, and the effect completely disappears when the traits of the vault are removed from the analysis, indicating that it is more likely the contrast of a small number of traits on the face with the more tightly integrated vault rather than a specific effect on traits in more highly strained areas.

DISCUSSION

We found no evidence suggesting that h^2, e, e_{e}, or CV for individual traits vary across different functionally/developmentally defined regions of the cranium or across those regions subject to different levels of masticatory strain. Indeed, these results indicate that the distribution of heritability across the cranium is not different from that expected under a model where all cranial traits have heritability values of $\sim$44%, and the differences in heritability estimates observed is because of a sampling error inherent in estimating quantitative genetic parameters in this population. The same result has been observed in several quantitative genetic studies of primate cranial morphology (Cheverud, 1982, 1995, 1996).

Phenotypic and genetic variation followed similar patterns across the cranium as indicated by the strong similarity of genetic and phenotypic variance/covariance and correlation matrices. In addition, the r^2 values of the comparisons of the measures of intrinsic phenotypic variation with their genetic counterparts range from 0.59 to 0.65 indicating that phenotypic values are giving us some insight into the genetic variances of the traits. Critical to the understanding of these correlation values is that these coefficients are about as strong a correlation as would be allowed given the repeatability of the genetic estimates due to the elevated sampling variance associated with estimating genetic parameters. We predict that the relationships between estimates of phenotypic and genetic variation will strengthen as the size of the collection increases with the ongoing addition of more crania. This indicates that studies of phenotypic variation are fairly reliable guides to patterns of genetic variation in populations. This is not a new conclusion but remains controversial (Cheverud, 1988; Roff, 1996). The upshot of this finding is that estimates of variation gleaned from comparative material while appropriately taking covariates such as sex into consideration add to our understanding of the propensity of traits to evolve.

Our findings do not support the variational prediction of the homoiology hypothesis (Wood and Lieberman, 2001; Lycett and Collard, 2005; Collard and Wood, 2007), which states that cranial traits subject to high strain tend to be more intrinsically variable. We propose that this finding is largely a consequence of the fact that we,

TABLE 3. *Results of ANOVAs (P-values) for each of the statistics summarizing different aspects variation for each of the hypothetical groupings*

	Face/orbit and base/vault	Face/orbit, base, and vault	Face, base, vault, and orbit	Low–moderate strain and high strain
h^2	0.939	0.767	0.899	0.355
CV	0.882	0.242	0.252	0.276
e	0.932	0.583	0.573	0.751
e_e	0.814	0.230	0.322	0.145
c	0.453	0.000	0.001	0.297
c_p	0.254	0.000	0.000	0.029
c_{pr}	0.612	0.011	0.018	0.229
i	0.929	0.001	0.002	0.319
i_p	0.360	0.004	0.009	0.034

and some others (Lycett and Collard, 2005; von Cramon-Taubadel, 2009), corrected estimates of the degree of variation for differences in means between the sexes and for any age-graded variation among adults. Sexual dimorphism in the baboon face is substantially larger than it is for the cranial base and vault (Leigh and Cheverud, 1991). The brain and associated cranial base and vault grow relatively early, approximating their final sizes at a young age before much overall sexual dimorphism has been established. In contrast, the face continues growing until relatively late, when overall body-size growth is quite different between the sexes. Lycett and Collard 2005 and von Cramon-Taubadel 2009 presented the only similar studies to correct for sex and this tended to attenuate the estimates of intrinsic variation in more highly strained areas, although the mean levels of variation were still significantly different across the differently-strained regions. In both cases, this was attributed to higher mean levels of strain and not to variance in the level of strain, but the effects of age, population structure, and nutritional status were not accounted for in these studies. Although age tended to affect traits thought to be exposed to high levels of strain more often than other traits, its effect was small and did not change the mean differences in the CV values of traits in the different regions. We might consider that the effects of age, however, may be elevated in a wild population relative to a captive population.

The second reason why this study may have failed to support the variational prediction of the homoiology hypothesis where other studies do support it is that all the animals in this population are fed a uniform diet of monkey chow and were sheltered from many of the threats to which a wild animal might be exposed. Although monkey chow is reasonably hard, the animals used in this study have obvious signs of tooth wear and the animals are known to chew on hard objects in their environment (which all suggest that the animals are being exposed to strains associated with mastication), the environment in which the animals were raised may have been fairly uniform. If the strain-inducing behaviors in which they engaged did not vary much across individuals, there may not have been much opportunity for plastic responses to differ across individuals and, thus, extra variance would not be introduced into the population. A group of animals with the same genotypes as those in this study but raised in another uniform environment and exposed to different degrees or frequencies of loading during chewing might develop a different mean state without changing the within-population variances. None of the previous studies have demonstrated that nutritional status, population structure, and age are not driv-

ing the difference in the variational properties of traits in different regions of the skull in wild populations (this is discussed more thoroughly later). The variational properties of relevant environmental conditions in wild populations, like in the captive group used here, is also unknown.

These results strongly suggest that the region that a trait occupies in the cranium and exposure to high mean levels of strain, as opposed to variance in strains across traits, do not cause traits to be more phenotypically variable nor do they affect the evolutionary potential of individual traits. This uniformity in evolvability that we have demonstrated in cranial traits in this population of baboons supports the thesis that the particular location of a trait on the cranium does not impart an intrinsic limitation on the evolution of traits in any one region of the skull or any differential utility for reconstructing phylogeny, an observation that accords with studies of the relative efficacy of different traits for phylogenetic reconstruction (Collard and Wood, 2001, 2007; Lycett and Collard, 2005; Collard and Lycett, 2008; von Cramon-Taubadel, 2009). The results of this investigation, in combination with the results of these phylogenetic analyses, indicate that traits from one region of the cranium do not appear to be better for reconstructing phylogeny than any other. Because traits in one region are no more or less evolvable than any other, at least in this population, there is not any reason to expect the intrinsic variational properties of traits to affect their evolution in different ways.

That bone can display plastic responses to different types of environmental effects is indisputable (Trinkaus et al., 1994; Lieberman et al., 2003; Ruff et al., 2006; Ravosa et al., 2007; Menegaz et al., 2009). There is no indication that a change in the level of strain, however, leads to an increase in the intrinsic variation of a trait. Those traits that show higher strain under mastication in model organisms may also be more liable to display variance as a result of a lifting of a ceiling on the degree to which strain affects traits, thus providing room for strain to vary. But it is still important to distinguish between the mean and variance of an effect.

It is also important to point out that strain may not be the only cause of the additional variation seen in parts of the face in some studies. As has been emphasized before, facial characteristics tend to finish developing later in life than those on the remainder of the cranium, and this leaves open the opportunity for various insults to growth in the form of disease, injury, or malnutrition, among others, to affect those traits (Lieberman, 1995, 1997; Wood and Lieberman, 2001). In this context, the elevated levels of intrinsic variation observed in these

traits in samples of wild populations may be because of any number of factors, and the role of variance in masticatory strain may not be easily separated from other potential causes. If traits in the face vary more among groups within a species than traits in other parts of the cranium and the museum specimens were sampled from multiple locations within a species' range, this would cause elevated levels of variance that could be easily mistaken for the effects of a variable environment. Without a sense of how variable exposure to strains associated with mastication is within and among populations, there is no good way to decide if it is solely or partly responsible for the observed pattern.

In contrast to the results of the investigation into the variational properties of traits irrespective of their covariance with other traits, the average level of integration and conditional evolvability of traits does vary across regions. As a result, characteristics in the cranial vault are more likely to be constrained in their evolution by stabilizing selection on other traits and more likely to show stronger correlated responses to selection on other traits. Using multiple highly correlated characteristics in a phylogenetic analysis can result in a false sense of security in the strength of support of a particular phylogenetic hypothesis as correlation among traits effectively reduces the number of traits available for analysis (Cheverud, 2001). This also implies that traits in the cranial vault may not be able to evolve independently of the rest of the cranium as easily as other traits in the cranium. Patterns of integration, however, evolve in their own right (Cheverud, 1982; Steppan et al., 2002; Whitlock et al., 2002; Wagner et al., 2007; Arnold et al., 2008) and there are some substantial differences between integration in baboons and those of other primates (Marroig et al., 2009; Porto et al., 2009), so the extent to which these results can be generalized is unknown.

Several claims have been made about the implications of the heritability of traits for reconstructing phylogeny. Some discussions of character selection include claims that highly heritable characteristics should be less variable than less heritable characteristics. For example, Wood and Lieberman 2001 hypothesized that "Features that have higher coefficients of narrow sense heritability (h^2) are predicted to show less intraspecific variability and are thus be more taxonomically valent than features with lower heritabilities" (p. 16) and that "dental features that are highly heritable are also either unresponsive, or are only weakly responsive, to epigenetic stimuli such as strain" (p. 16). Lieberman 1996 recommended that "Characters used in any (phylogenetic) analysis must be heritable, with as clear a link to the genome as possible" (p. 99).

These statements suggest that we may expect highly heritable characteristics tend to be highly canalized and less sensitive to environmental effects in general. As mentioned previously, heritability is a dimensionless ratio that gives no sense of the absolute magnitude of any component of variance so there is no reason to expect traits with different heritability values to display different magnitudes of variation. This is one of the rationales for using measures of evolvability in evolutionary quantitative genetics instead of h^2 (Houle, 1992; Hansen and Houle, 2008). Differences between individuals can be assigned to genetic and environmental sources of variation and to gene-by-environment interactions (Falconer and Mackay, 1996), but the developmental trajectory of any one individual cannot be decomposed into distinct genetic and environmental components. The construction of an individual's phenotype is the result of environment-gene interplay and is not reducible to component parts (Lewontin, 2000). Some traits may be more canalized than others in that the outcomes of their development may be more restricted for a given set of genotypes across a range of environments, but this is unrelated to the heritability of the trait. It is also true that we would really rather have characteristics that are not heritable for use in our phylogenetic analyses in that we would prefer them to not vary at all within species and, instead, to display variation exclusively among groups (e.g. the presence or absence of an amnion is not heritable within species because it does not vary within species).

Another way to approach this problem would be to assess the phylogenetic equivalent of heritability, phylogenetic h^2 (Lynch, 1991; Housworth et al., 2004) for various characteristics. The degree of phylogenetic h^2 for a trait is an indicator of how closely interspecific differences are associated with degrees of phylogenetic similarity, much in the same way that h^2 predicts how individual differences are associated with degrees of relatedness in a pedigree. The methods for estimating phylogenetic h^2 (phylogenetic mixed models. Houseworth et al., 2004) are very similar to the methods used in this investigation to estimate h^2 in a pedigree. Phylogenetic h^2 was originally developed to address the problem of phylogenetic effects in comparative studies. It could be used, however, to see which traits track phylogeny with high fidelity (given a known phylogeny) and, thus, may be useful for deciding which characters to choose in a paleoanthropological investigation. We caution, however, that the phylogenetic h^2 needs to be estimated using large numbers of species (just as h^2 requires large numbers of individuals with known relationships) and a characteristic with a high phylogenetic h^2 may not reflect relationships well in small subsets of a phylogeny or in groups that were not included in the initial estimate of phylogenetic h^2.

It is worth emphasizing that phylogenetic h^2 is based on a loose analogy to h^2 in the narrow sense. It is estimated in a similar way but it does not have the same solid base in Mendelian rules of inheritance that supports the use of h^2 in evolutionary quantitative genetics (Lynch, 1991; Houseworth et al., 2004). A nonzero phylogenetic heritability could be motivated by the response of phenotypes differential distribution of environments across several clades of organisms (a phylogeny-environment covariance). This could mean that traits with differences among groups that are caused by environmental effects rather than genetic effects might actually be useful in phylogenetic reconstruction, such as the bicondylar angle, which only develops the characteristic human condition in a particular mechanical environment (Tardieu, 1995; Shefelbine et al., 2002). A difference like this would not be the result of the accumulation of additive genetic variance among species via mutation, drift, and selection, but it is not clear why an investigator should care so long as a trait was capable of making useful distinctions among groups. We do not think that this will happen very often, but it is far from certain that cases like these would never arise.

In conclusion, we find that traits from any region of the cranium are about as well suited for phylogenetic analysis as those from any other region based on their intrinsic variational properties, at least as far as the

results from this population are concerned. In this one sample there is no distinction among traits in cranial regions in levels of intrinsic variation, heritability, evolvability, or intrinsic environmental variation. Characteristics on the cranial vault tend to have less independence than traits from other regions. We caution that the results from this analysis may not be generalizable, particularly because they are based on a single, captive group and groups in other contexts may be exposed to environments that could serve as sources of additional variation that are not detectable here. One reason to think that this study has some generalizability when it comes to statistics reflecting genetic variation, is that the large majority of like studies show very similar results (Roff, 1986; Cheverud, 1988). Importantly, the meaning of "heritable" with regard to the issue at hand needs clarification and other measures of genetic variation are more useful for modeling evolution. Overall, we think that at the present state of knowledge, combined with the results presented here, there are not compelling reasons to exclude characteristics from phylogenetic analyses based on their intrinsic genetic or environmental variational properties. Other criteria, however, may be useful, but only in combination with an account of how the forces of evolution have acted to diversify a group.

ACKNOWLEDGMENTS

LITERATURE CITED

Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211.

Arnold SJ, Burger R, Hohenlohe PA, Ajie BC, Jones AG. 2008. Understanding the evolution and stability of the G-matrix. Evolution 62:2451–2461.

Blows MW, Higgie M. 2003. Genetic constraints on the evolution of mate recognition under natural selection. Am Nat 161:240–253.

Chamberlain A, Wood BA. 1987. Early hominid phylogeny. J Hum Evol 16:119–133.

Cheetham AH, Jackson JBC, Hayek LAC. 1993. Quantitative genetics of bryozoan phenotypic evolution. I. Rate tests for random change versus selection in differentiation of living species. Evolution 47:1526–1538.

Cheverud JM. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution 36:499–516.

Cheverud JM. 1988. A comparison of genetic and phenotypic correlations. Evolution 42:958–968.

Cheverud JM. 1995. Morphological integration in the saddleback tamarin (Saguinus fuscicollis) cranium. Am Nat 145:63–89.

Cheverud JM. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (Saguinus oedipus) and saddleback (S. fuscicollis) tamarins. J Evol Biol 9:5–42.

Cheverud JM. 2001. A simple correction for multiple comparisons in interval mapping genome scans. Heredity 87:52–58.

Cheverud JM, Marroig G. 2007. Comparing covariance matrices: random skewers method compared to the common principal components model. Genet Mol Biol 30:461–469.

Cheverud JM, Rutledge JJ, Atchley WR. 1983. Quantitative genetics of development—genetic correlations among age-specific trait values and the evolution of ontogeny. Evolution 37:895–905.

Collard M, Lycett SJ. 2008. Does phenotypic plasticity confound attempts to identify hominin fossil species? Folia Primatol (Basel) 79:111–122.

Collard M, Wood B. 2000. How reliable are human phylogenetic hypotheses? Proc Natl Acad Sci USA 97:5003–5006.

Collard M, Wood B. 2001. Homoplasy and the early hominid masticatory system: inferences from analyses of extant hominoids and papionins. J Hum Evol 41:167–194.

Collard M, Wood B. 2007. Hominin homoiology: an assessment of the impact of phenotypic plasticity on phylogenetic analyses of humans and their fossil relatives. J Hum Evol 52:573–584.

Currey JD. 1984. The mechanical adaptation of bones. Princeton, NJ: Princeton University Press.

de Beer GR. 1937. The development of the vertebrate skull. Oxford: Oxford University Press.

Estes S, Arnold SJ. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am Nat 169:227–244.

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics, 4th ed. Harlow, England: Pearson Education Limited.

Feldman MW, Lewontin RC. 1975. Heritability hang-up. Science 190:1163–1168.

Gibbs S, Collard M, Wood B. 2000. Soft-tissue characters in higher primate phylogenetics. Proc Natl Acad Sci USA 97:11130–11132.

Gonzalez-Jose R, Escapa I, Neves WA, Cuneo R, Pucciarelli HM. 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in homo evolution. Nature 453:775–778.

Hallgrimsson B, Lieberman DE. 2008. Mouse models and the evolutionary developmental biology of the skull. Integr Comp Biol 48:373–384.

Hansen TF, Armbruster WS, Carlson ML, Pelabon C. 2003. Evolvability and genetic constraint in Dalechampia blossoms: genetic correlations and conditional evolvability. J Exp Zool B Mol Dev Evol 16:23–39.

Hansen TF, Houle D. 2004. Evolvability, stabilizing selection, and the problem of stasis. In: Pigliucci M, Preston K, editors. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Oxford, England: Oxford University Press. p 130–150.

Hansen TF, Houle D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J Evol Biol 21:1201–1219.

Hlusko LJ, Do N, Mahaney MC. 2007. Genetic correlations between mandibular molar cusp areas in baboons. Am J Phys Anthropol 132:445–454.

Hlusko LJ, Lease LR, Mahaney MC. 2006. Evolution of genetically correlated traits: tooth size and body size in baboons. Am J Phys Anthropol 131:420–427.

Hlusko LJ, Mahaney MC. 2003. Genetic contributions to expression of the baboon cingular remnant. Arch Oral Biol 48:663–672.

Hlusko LJ, Suwa G, Kono RT, Mahaney MC. 2004. Genetics and the evolution of primate enamel thickness: a baboon model. Am J Phys Anthropol 124:223–233.

Houle D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 13:195–204.

Housworth EA, Martins EP, Lynch M. 2004. The phylogenetic mixed model. Am Nat 163:84–96.

Hunt G. 2007. Evolutionary divergence in directions of high phenotypic variance in the ostracode genus Poseidonamicus. Evolution 61:1560–1576.

Johanson DC, White TD. 1979. A systematic assessment of early African hominoids. Science 203:321–330.

Jones AG, Arnold SJ, Burger R. 2007. The mutation matrix and the evolution of evolvability. Evolution 61:727–745.

Jorjani H, Klei L, Emanuelson U. 2003. A simple method for weighted bending of genetic (co)variance matrices. J Dairy Sci 86:677–679.

Kimbel WH, White TD, Johanson DC. 1984. Cranial morphology of *Australopithecus afarensis*: a comparative study based on a composite reconstruction of the adult skull. Am J Phys Anthropol 64:337–338.

Leigh SR, Cheverud JM. 1991. Sexual dimorphism in the baboon facial skeleton. Am J Phys Anthropol 84:193–208.

Lewontin R. 2000. The triple helix: gene, organism, and environment. Cambridge, MA: Harvard University Press.

Lieberman DE. 1995. Testing hypotheses about recent human evolution from skulls—integrating morphology, function, development, and phylogeny. Curr Anthropol 36:159–197.

Lieberman DE. 1997. Making behavioral and phylogenetic inferences from hominid fossils: considering the developmental influence of mechanical forces. Annu Rev Anthropol 26:185–210.

Lieberman DE, Hallgrimsson B. 2008. Spatial packing, cranial base angulation, and craniofacial shape variation in the mammalian skull: testing a new model using mice. J Anat 212:720–735.

Lieberman DE, Pearson OM, Mowbray KM. 2000a. Basicranial influence on overall cranial shape. J Hum Evol 38:291–315.

Lieberman DE, Pearson OM, Polk JD, Demes B, Crompton AW. 2003. Optimization of bone growth and remodeling in response to loading in tapered mammalian limbs. J Exp Biol 206:3125.

Lieberman DE, Ross CF, Ravosa MJ. 2000b. The primate cranial base: ontogeny, function, and integration. Am J Phys Anthropol 31:117–169 (Suppl).

Lieberman DE, Wood BA, Pilbeam DR. 1996. Homoplasy and early homo: an analysis of the evolutionary relationships of *H. habilis sensu stricto* and *H. rudolfensis*. J Hum Evol 30:97–120.

Lockwood CA. 2007. Adaptation and functional integration in primate phylogeny. J Hum Evol 52:490–503.

Lockwood CA, Fleagle JG. 1999. The recognition and evaluation of homoplasy in primate and human evolution. Am J Phys Anthropol 29:189–232.

Lockwood CA, Kimbel WH, Lynch JM. 2004. Morphometrics and hominin phylogeny: support for a chimpanzee-human clade and differentiation among great ape subspecies. Proc Natl Acad Sci USA 101:4356–4360.

Lycett SJ, Collard M. 2005. Do homoiologies impede phylogenetic analyses of the fossil hominids? an assessment based on extant papionin craniodental morphology. J Hum Evol 49:618–642.

Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. Evolution 45:1065–1080.

Lynch M. 2007a. Colloquium papers: the frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA 104:8597–8604.

Lynch M. 2007b. The origins of genome architecture. Sunderland, MA: Sinauer and Associates, Inc.

Marroig G, Cheverud JM. 2004a. Cranial evolution in sakis (*Pithecia*, platyrrhini). I. Interspecific differentiation and allometric patterns. Am J Phys Anthropol 125:266–278.

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

Marroig G, Cheverud JM. 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in new world monkeys. Evolution 59:1128–1142.

Marroig G, Shirai LT, Porto A, de Oliveira FB, De Conto V. 2009. The evolution of modularity in the mammalian skull. II. Evolutionary consequences. Evol Biol 36:136–148.

Martin RB, Burr D, Sharkey NA. 1998. Skeletal tissue mechanics. New York: Springer.

McCarthy RC. 2001. Anthropoid cranial base architecture and scaling relationships. J Hum Evol 40:41–66.

McCollum MA. 1999. The robust australopithecine face: a morphogenetic perspective. Science 284:301–305.

McCollum MA, Ward SC. 1997. Subnasoalveolar anatomy and hominoid phylogeny: evidence from comparative ontogeny. Am J Phys Anthropol 102:377–405.

Menegaz RA, Sublett SV, Figueroa SD, Hoffman TJ, Ravosa MJ. 2009. Phenotypic plasticity and function of the hard palate in growing rabbits. Anat Rec 292:277–284.

Nevell L, Wood B. 2008. Cranial base evolution within the hominin clade. J Anat 212:455–468.

Olson TR. 1981. Basicranial morphology of the extant hominoids and Pliocene hominids: the new material from the Hadar Formation, Ethiopia, and its significance in early human evolution and taxonomy. In: Stringer CB, editor. Aspects of human evolution. London: Taylor and Francis. p 99–128.

Olson TR. 1985a. Cranial morphology and systematics of the Hadar formation hominids and *Australopithecus africanus*. In: Delson E, editor. Ancestors: the hard evidence. New York: Alan R Liss. p 102–119.

Olson TR. 1985b. Taxonomic affinities of the immature hominid crania from Hadar and Taung. Nature 316:539–540.

Pigliucci M. 2001. Phenotypic plasticity: beyond nature and nurture. Baltimore, MD: The Johns Hopkins University Press.

Plavcan JM. 2003. Scaling relationships between craniofacial sexual dimorphism and body mass dimorphism in primates: implications for the fossil record. Am J Phys Anthropol 120:38–60.

Ponce de Leon MS, Zollikofer CPE. 2001. Neandertal cranial ontogeny and its implications for late hominid diversity. Nature 412:534–538.

Porto A, de Oliveira FB, Shirai LT, De Conto V, Marroig G. 2009. The evolution of modularity in the mammalian skull. I. Morphological integration patterns and magnitudes. Evol Biol 36:118–135.

Ravosa MJ, Kunwar R, Stock SR, Stack MS. 2007. Pushing the limit: Masticatory stress and adaptive plasticity in mammalian craniomandibular joints. J Exp Biol 210:628–641.

Roff DA. 1996. The evolution of genetic correlations: an analysis of patterns. Evolution 50:1392–1403.

Rogers J, Mahaney MC, Witte SM, Nair S, Newman D, Wedel S, Rodriguez LA, Rice KS, Slifer SH, Perelygin A, Slifer M, Palladino-Negro P, Newman T, Chambers K, Joslyn G, Parry P, Morin PA. 2000. A genetic linkage map of the baboon (*Papio hamadryas*) genome based on human microsatellite polymorphisms. Genomics 67:237–247.

Ruff C, Holt B, Trinkaus E. 2006. Who's afraid of the big bad Wolff? "Wolff's law" and bone functional adaptation. Am J Phys Anthropol 129:484–498.

Schluter D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–1774.

Schluter D. 2000. The ecology of adaptive radiation. Oxford: Oxford University Press.

Scott JE, Lockwood CA. 2004. Patterns of tooth crown size and shape variation in great apes and humans and species recognition in the hominid fossil record. Am J Phys Anthropol 125:303–319.

Scott JH. 1958. The cranial base. Am J Phys Anthropol 16:319–348.

Shea BT. 1985. On aspects of skull form in african apes and orangutans, with implications for hominoid evolution. Am J Phys Anthropol 68:329–342.

Shefelbine SJ, Tardieu C, Carter DR. 2002. Development of the femoral bicondylar angle in hominid bipedalism. Bone 30:765–770.

Sherwood RJ, Duren DL, Havill LM, Rogers J, Cox LA, Towne B, Mahaney MC. 2008. A genome-wide linkage scan for quantitative trait loci influencing the craniofacial complex in baboons (*Papio hamadryas* spp.). Genetics 180:619–628.

Skelton RR, McHenry HM. 1992. Evolutionary relationships among early hominids. J Hum Evol 23:309–349.

Skelton RR, McHenry HM. 1998. Trait list bias and a reappraisal of early hominid phylogeny. J Hum Evol 34:109–113.

Skelton RR, McHenry HM, Drawhorn GM. 1986. Phylogenetic analysis of early hominids. Curr Anthropol 27:21–43.

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

Smith HF, Terhune CE, Lockwood CA. 2007. Genetic, geographic and environmental correlates of human temporal bone variation. Am J Phys Anthropol 134:312–322.

Sokal RR, Rohlf FJ. 1995. Biometry: the principles and practices of biological research. New York, NY: W.H. Freeman and Co.

Steppan SJ, Phillips PC, Houle D. 2002. Comparative quantitative genetics: evolution of the G matrix. Trends Ecol Evol 17:320–327.

Strait DS. 1998. Evolutionary integration in the hominid cranial base. Ph.D. Dissertation, State University of New York, Stony Brook.

Strait DS. 2001. Integration, phylogeny, and the hominid cranial base. Am J Phys Anthropol 114:273–297.

Strait DS, Grine FE. 1998. Trait list bias? A reply to Skelton and McHenry. J Hum Evol 34:115–118.

Strait DS, Grine FE. 1999. Cladistics and early hominid phylogeny. Science 285:1210.

Strait DS, Grine FE. 2004. Inferring hominoid and early hominid phylogeny using craniodental characters: the role of fossil taxa. J Hum Evol 47:399–452.

Strait DS, Grine FE, Moniz MA. 1997. A reappraisal of early hominid phylogeny. J Hum Evol 32:17–82.

Tardieu C. 1995. Homoplasie de l'angle d'obliquite femorale chez l'homme et les grands singes: demonstration morphogentique. Anthropologie 33:69–78.

Tobias PV. 1980. Australopithecus afarensis and *A. africanus*: critique and an alternative hypothesis. Paleont Afr 23:1–17.

Trinkaus E, Churchill SE, Ruff CB. 1994. Postcranial robusticity in *Homo*. II. Humeral bilateral asymmetry and bone plasticity. Am J Phys Anthropol 93:1–34.

von Cramon-Taubadel N. 2009. Revisiting the homoiology hypothesis: the impact of phenotypic plasticity on the reconstruction of human population history from craniometric data. J Hum Evol 57:179–190.

Wagner GP, Altenberg L. 1996. Perspective: complex adaptations and the evolution of evolvability. Evolution 50:967–976.

Wagner GP, Laubichler MD. 2004. Rupert Riedl and the re-synthesis of evolutionary and developmental biology: body plans and evolvability. J Exp Zool B Mol Dev Evol 302:92–102.

Wagner GP, Pavlicev M, Cheverud JM. 2007. The road to modularity. Nature Rev Genet 8:921–931.

Walker A, Leakey RE, Harris JM, Brown FH. 1986. 2.5-MYR Australopithecus boisei from west of Lake Turkana, Kenya. Nature 322:517–522.

Walker AC, Leakey RE. 1988. The evolution of *Australopithecus boisei*. In: Grine FE, editor. Evolutionary history of the 'robust' australopithecines. New York: Aldine de Gruyter. p 247–258.

White TD, Johanson DC, Kimbel WH. 1981. *Australopithecus africanus*: its phyletic position reconsidered. S Afr J Sci 77:445–470.

Whitlock MC, Phillips PC, Fowler K. 2002. Persistence of changes in the genetic covariance matrix after a bottleneck. Evolution 50:1968–1975.

Willmore KE, Roseman CC, Rogers J, Richtsmeier JT, Cheverud JM. 2009. Genetic variation in baboon craniofacial sexual dimorphism. Evolution 63:799–806.

Wood B, Lieberman DE. 2001. Craniodental variation in *Paranthropus boisei*: a developmental and functional perspective. Am J Phys Anthropol 116:13–25.

Wood BA, Chamberlain AT. 1986. *Australopithecus*: grade or clade? In: Wood BA, Martin L, Andrews P, editors. Major topics in primate and human evolution. Cambridge: Cambridge University Press. p 220–248.