

Were neandertal and modern human cranial differences produced by natural selection or genetic drift?

Timothy D. Weaver^{a,b,*}, Charles C. Roseman^c, Chris B. Stringer^d

^a Department of Anthropology, University of California, One Shields Avenue, Davis, CA 95616, USA

^b Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany

^c Department of Anthropology, University of Illinois at Urbana-Champaign, 109 Davenport Hall, 607 South Matthews Avenue, Urbana, IL 61801, USA

^d Department of Palaeontology, The Natural History Museum, London SW7 5BD, UK

Received 10 July 2006; accepted 1 March 2007

Abstract

Most evolutionary explanations for cranial differences between Neandertals and modern humans emphasize adaptation by natural selection. Features of the crania of Neandertals could be adaptations to the glacial climate of Pleistocene Europe or to the high mechanical strains produced by habitually using the front teeth as tools, while those of modern humans could be adaptations for articulate speech production. A few researchers have proposed non-adaptive explanations. These stress that isolation between Neandertal and modern human populations would have led to cranial diversification by genetic drift (chance changes in the frequencies of alleles at genetic loci contributing to variation in cranial morphology). Here we use a variety of statistical tests founded on explicit predictions from quantitative- and population-genetic theory to show that genetic drift can explain cranial differences between Neandertals and modern humans. These tests are based on thirty-seven standard cranial measurements from a sample of 2524 modern humans from 30 populations and 20 Neandertal fossils. As a further test, we compare our results for modern human cranial measurements with those for a genetic dataset consisting of 377 microsatellites typed for a sample of 1056 modern humans from 52 populations. We conclude that rather than requiring special adaptive accounts, Neandertal and modern human crania may simply represent two outcomes from a vast space of random evolutionary possibilities.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Neandertals; Hominin evolution; Evolutionary quantitative genetics; Population genetics

Introduction

For about a century and a half researchers have documented morphological differences between the crania of Neandertals¹ and those of modern humans (Schaafhausen, 1857; Boule, 1911–13). It is true that for most features Neandertal and modern human ranges of variation overlap, but on average, and when multiple features are considered in combination,

most specialists agree that Neandertal and modern human crania can be distinguished morphologically from one another (e.g., Stringer, 1974; Harvati et al., 2004). Many of the metric differences relate to the degree of mid-facial projection and the shape of the neurocranium (for a recent summary see Franciscus, 2002) (Fig. 1). There is no consensus, however, on why these differences exist.

This issue can actually be addressed at both mechanistic and evolutionary levels. Mechanistic explanations focus on the developmental shifts that produced these differences (e.g., Smith, 1991; Lieberman, 1998; Spoor et al., 1999; Ponce de LeÓN and Zollikofer, 2001; Ackermann and Krovitz, 2002; Lieberman et al., 2002; Lieberman et al., 2004). Evolutionary explanations concentrate on which combination of the forces of evolution (i.e., natural selection, mutation, gene flow, and

* Corresponding author. Department of Anthropology, University of California, One Shields Avenue, Davis, CA 95616, USA. Tel.: +1 530 754 4575.

E-mail address: tdweaver@ucdavis.edu (T.D. Weaver).

¹ Because of a spelling change in German, there are two acceptable spellings for the name of the group. Neanderthal rather than Neandertal is the preferred spelling of the third author.



Fig. 1. Neandertal and modern human cranial differences. On the left is a Neandertal from France (cast of La Ferrassie 1) and on the right is a robust recent modern human from Polynesia.

genetic drift) were responsible for producing these differences. Here we investigate the evolutionary level.

Most evolutionary explanations for cranial differences between Neandertals and modern humans emphasize adaptation by natural selection. Given the northern geographic range of Neandertals, one long-standing possibility is that distinctive features of the Neandertal cranium are adaptations to the glacial climate of Pleistocene Europe (e.g., Sergi, 1958; Coon, 1962; Brose and Wolpoff, 1971). Another possibility is that these distinctive features are adaptations to the high mechanical strains produced by habitually using the front teeth as tools (e.g., Brace, 1964; Brose and Wolpoff, 1971; Rak, 1986; Demes, 1987; Trinkaus, 1987; Smith and Paquette, 1989; Spencer and Demes, 1993). Adaptive explanations are not limited to explaining Neandertal cranial morphology. Distinctive aspects of modern human crania may be adaptations for articulate speech production (e.g., Lieberman and Crelin, 1971; Laitman et al., 1979). A few researchers (e.g., Howell, 1952; Antón, 1994, 1996; Hublin, 1998; Franciscus, 2003) have proposed non-adaptive explanations. These stress that isolation between Neandertal and modern human populations would have led to cranial diversification by genetic drift (chance changes in the frequencies of alleles at genetic loci contributing to variation in cranial morphology).

The mechanistic and evolutionary levels of investigation are of course linked. An improved understanding of development is valuable for carefully defining the phenotypes available to natural selection at different points in ontogeny. Indeed, any evolutionary change in morphology must be accompanied by a shift in development. Even individual variation within a population in adult morphology is the result of individual variation in development. Nevertheless, a detailed understanding of development does not allow us to distinguish between the actions of natural selection versus genetic drift because both these evolutionary forces can shape morphology through similar developmental shifts. To tackle the evolutionary level, the specifics of the evolutionary forces themselves must be explicitly modeled using quantitative and population genetics.

We ask the following question: can genetic drift explain cranial differences between Neandertals and modern humans? We choose to focus on genetic drift rather than natural selection

because, as noted by many researchers since Kimura (1968) first proposed the neutral theory of molecular evolution, diversification by genetic drift is a useful starting point, even if incorrect, by giving unambiguous, testable predictions (Kreitman, 1996). In contrast, while it is possible to test specific natural selection models, it is impossible to test a general natural selection model because the predictions depend heavily on the details of how natural selection is acting morphological features (e.g., the strength of natural selection, whether it is constant or randomly fluctuating, etc.) (Lynch, 1989).

Our general approach follows that of Lande (1977). We start with a theoretical model: cranial diversification of Neandertals and modern humans by genetic drift (neutral diversification or divergence). We then build a null hypothesis based on explicit predictions from quantitative- and population-genetic theory about the patterns of morphological variation within and between groups we would expect if the model were true. Next, we use four statistical tests to compare the null hypothesis to the patterns of variation we actually observe within modern human populations and between Neandertals and modern humans. Finally, we conduct further analyses to evaluate the statistical power of our tests. In particular, we perform three of the same statistical tests that we use to make the comparisons between Neandertals and modern humans on all possible pairs of modern human populations, and for two of the tests we compare the results for modern human cranial measurements with those for microsatellites. These additional tests serve both to evaluate statistical power and as a benchmark for understanding any potential differences between how natural selection has acted to produce differences among modern human populations versus between Neandertals and modern humans.

Materials and methods

Data

Cranial measurements. We base our analyses on 37 standard cranial measurements collected on a sample of 2524 modern human crania from 30 globally distributed populations (Howells, 1973, 1989, 1995) and 20 Neandertal specimens (Table 1). We use the modern human data to calculate measurement means as well as measurement variances and covariances. Due to the fragmentary nature of fossil remains the Neandertal sample contains some missing data (Table 1), but we only use this sample to estimate the Neandertal mean for each measurement. The average sample size for estimating the Neandertal means is 11, with a minimum of 6 and a maximum of 19. A much larger and less fragmentary Neandertal sample would be necessary to robustly estimate variance-covariance patterns for Neandertals.

Genetic loci. To further examine the influence of natural selection on patterns of cranial variation among modern human populations we compared our results for cranial measurements with those for a set of presumably neutrally-evolving genetic repeat loci (microsatellites). If cranial differences among

modern human populations were primarily shaped by genetic drift, then we would expect our results for cranial measurements to be similar to those for microsatellites. We used the 377 microsatellites from the study by Rosenberg and colleagues (2002). The sample consists of 1056 modern humans from 52 globally distributed populations.

Null hypothesis

Statistical tests based on theoretical predictions from population genetics are widely applied to genetic data. However, here we are dealing with the phenotype, which necessitates an additional layer describing the inheritance of phenotypic traits in a way consistent with population genetics. Quantitative genetics provides this layer through explicit statements about heredity. Here we assume the classical quantitative genetics model of heredity (more details below), which posits that the genetic basis of each cranial measurement is a large number of genetic loci that contribute equally and additively (i.e., no interactions between them) to the length of the measurement (Falconer, 1981). If future research expands our knowledge about the genetic basis of cranial morphology, it will be possible to refine this model while still remaining within the framework of quantitative genetics (Walsh, 2001).

Importantly, the null hypothesis we consider here is a composite of models for evolutionary process and heredity, so there is more than one interpretation if the null hypothesis is rejected. Rejection either means natural selection played a significant role in producing the cranial divergence among groups or the classical quantitative genetics model of heredity inadequately describes the measurements (or a combination of the two).

Heredity. For heredity, the null hypothesis states that the within modern human phenotypic variance-covariance matrix (\mathbf{P}_w) is proportional to the additive genetic variance-covariance matrix (\mathbf{G}) irrespective of the population of origin, which is the same as setting equal heritabilities for all measurements for all populations. It also states that \mathbf{G} and \mathbf{P}_w for Neandertals are proportional to \mathbf{G} and \mathbf{P}_w for modern humans. These statements are reasonable given that \mathbf{G} and \mathbf{P}_w tend to be proportional to one another for morphological characteristics (Cheverud, 1988; Roff, 1996), and \mathbf{P}_w matrices from different modern human populations are quite similar (González-Jose et al., 2004).

There are two basic models for neutral divergence between populations in the evolutionary quantitative genetics literature: the constant-heritability (CH) and the mutation drift equilibrium (MDE) models (Lande, 1976; Lynch and Hill, 1986; Turelli et al., 1988). The null hypothesis assumes that either the CH or the MDE model adequately describes neutral divergence. Specifically, the null hypothesis states that either 1) new mutations have a minimal effect on \mathbf{G} over the time period considered (CH model) or 2) the matrix of additive genetic variances and covariances introduced by mutation per generation is proportional to \mathbf{G} and the populations are near to mutation drift equilibrium (MDE model). Both of these statements lead to the expectation that the between Neandertal-

modern human variance-covariance matrix (\mathbf{P}_b) will be proportional to \mathbf{P}_w if the divergence was due to genetic drift (Lande, 1979, 1980; Lynch, 1989).

For both models, the divergence rate depends on the strength of random genetic drift, which increases with smaller effective population sizes (N_e) and with larger amounts of heritable variation. The difference is that the CH model does not directly consider mutations, so the amount of heritable variation in the populations is not explicitly related to N_e . The MDE model explicitly links the amount of heritable variation with N_e by assuming equilibrium between the addition of variation due to mutations and the loss of variation due to random genetic drift. The equilibrium amount of heritable variation will be higher in a population with more frequent inputs of mutational variation. Also, because the strength of random genetic drift decreases with increasing N_e , it will be higher in a population with a larger N_e . Therefore, for populations in MDE the opposing effects of N_e on the strength of random genetic drift balance each other, making the divergence rate dependent only on the rate of input of mutational variation. This means that for populations in MDE the divergence rate will not depend on N_e .

The assumptions behind the CH model are appropriate for about $N_e/5$ generations since the founding of new populations (Turelli et al., 1988). The MDE model works best for populations that have had time to reach equilibrium. This depends on initial conditions, but any population will be close to equilibrium after about $4N_e$ generations (Turelli et al., 1988).

Evolutionary process and history. For evolutionary process, the null hypothesis states that the cranial divergence between Neandertals and modern humans was entirely the consequence of genetic drift. Note that over long time periods, mutation is important for maintaining genetic variation in the diverging populations (see above discussion about the MDE model), but the divergence mechanism is genetic drift. For evolutionary history, the null hypothesis states that after diverging from a common ancestor, Neandertals and modern humans evolved as separate evolutionary lineages with negligible gene flow between them. One statistical test described below, however, is still applicable under some models of population structure that include gene flow.

Statistical tests

We evaluate the adequacy of the null hypothesis in explaining the observed data using four statistical tests. For all the tests we use the 37 cranial measurements to calculate a set of thirty-seven morphological distances between Neandertals and modern humans (more details below). Each test deals with a different aspect of the distribution of this set of morphological distances. The first test deals with the mean morphological distance, the second and third with the variance and the shape of the distribution, and the fourth with the patterning of the distances relative to within population variation (the proportionality of \mathbf{P}_b to \mathbf{P}_w). We wrote scripts in MATLAB (Mathworks, Natick, MA) to perform all the statistical

Table 1
Neandertal sample and measurement means

Measurement abbreviation	Measurement name	Amud 1	Forbes' Quarry	Guattari 1	Krapina C	Krapina E	La Chapelle-aux-Saints	La Ferrassie 1	La Quina 5
BNL	Basion-nasion length		X	X			X	X	
OCC	Lambda-opisthion chord		X	X			X	X	
AVR	Molar alveolus radius	X	X					X	
SSR	Subspinale radius	X	X	X			X	X	
PRR	Prosthion radius	X	X	X			X	X	
BPL	Basion-prosthion length		X	X			X	X	
FRS	Nasion-bregma subtense	X	X	X		X	X	X	X
NPH	Nasion-prosthion height	X	X	X			X	X	
GOL	Glabella-occipital length	X	X	X			X	X	X
NOL	Nasio-occipital length	X	X	X			X	X	X
BBH	Basion-bregma height		X	X			X	X	
XCB	Maximum cranial breadth	X	X	X	X		X	X	X
XFB	Maximum frontal breadth	X	X	X	X	X	X	X	X
AUB	Biauricular breadth	X	X	X			X	X	X
ASB	Biasterionic breadth	X	X	X			X	X	X
NLH	Nasal height	X	X	X			X	X	
OBH	Orbital height	X	X	X	X		X	X	
OBB	Orbit breadth	X	X	X	X		X	X	
NLB	Nasal breadth		X	X	X		X	X	
MAB	Palate breadth	X	X	X			X	X	X
MDH	Mastoid height	X		X	X		X	X	
ZMB	Bimaxillary breadth	X	X	X	X		X	X	X
SSS	Subspinale radius	X	X	X			X	X	
FMB	Bifrontal breadth	X	X	X	X	X	X	X	X
NAS	Nasio-frontal subtense	X	X	X	X	X	X	X	X
EKB	Biorbital breadth	X	X	X	X		X	X	X
DKB	Interorbital breadth	X	X	X	X		X	X	
FRC	Nasion-bregma chord	X	X	X		X	X	X	X
FRF	Nasion-subtense fraction	X	X	X		X	X	X	X
PAC	Bregma-lambda chord	X	X	X			X	X	X
PAS	Bregma-lambda subtense	X	X	X			X	X	X
PAF	Bregma-subtense fraction	X	X	X			X	X	X
OCS	Lambda-opisthion subtense		X	X			X	X	
OCF	Lambda-subtense fraction		X	X			X	X	
VRR	Vertex radius	X	X	X			X	X	X
NAR	Nasion radius	X	X		X		X	X	X
ZMR	Zygomaxillare radius	X	X	X	X		X	X	

Abbreviations from [Howells \(1973\)](#); an “X” indicates data present; one of us (C.B.S.) collected the Neandertal data with the exception of the Shanidar measurements, which are from E. Trinkaus and casts.

analyses, making use of the statistical toolbox and Richard E. Strauss' MATLAB function package.

Calculation of the morphological distances. To calculate the morphological distances, we first estimate the within modern human variance-covariance matrix (\mathbf{P}_w). We do this by pooling the within population variance-covariance matrices across all the modern human populations in our sample. Next, we decompose \mathbf{P}_w into its principal components (eigenvectors). The resulting 37 principal components comprise a new set of measurements, which are uncorrelated with each other and represent an equal amount of variation as in the original set of measurements. To actually calculate the morphological distances, for each principal component we take the ratio of the variance between Neandertals and modern humans to the variance within modern human populations (eigenvalues). These 37 morphological distances reflect the degree of morphological separation between Neandertals and modern

humans in the directions of each of the principal components (for similar calculations see [Ackermann and Cheverud, 2002](#)).

Mean divergence. Under neutral divergence, each of the morphological distances is expected to be proportional to divergence time because the amount of morphological divergence will be proportional to within group variation and time since divergence ([Lande, 1977](#); [Lynch, 1989](#)). There will be substantial variation around this expectation, but it should hold on average. Each of the morphological distances can be thought of as unscaled divergence time estimates. Calculating an actual divergence time for two groups requires additional assumptions about generation length, heritability, and mutation rate ([Turelli et al., 1988](#)). If two pairs of groups are compared to each other, however, these additional terms cancel, and the ratio of the morphological distance between one pair of groups to the morphological distance between the

Table 1 (continued)

Neandertal 1	Saccopastore 1	Saccopastore 2	Saint-Césaire	Sala 1	Shanidar 1	Shanidar 2	Shanidar 4	Shanidar 5	Spy 1	Spy 2	Tabun C1	Neandertal measurement mean (mm)
	X	X			X							111.1
	X				X							93.4
	X	X	X		X			X				98.9
	X	X	X		X			X				112.5
	X	X	X		X			X				119.4
	X	X			X							116.9
X	X			X	X			X	X	X		19.9
	X	X	X		X			X				85.3
X					X				X	X	X	201.8
X	X				X				X	X	X	197.0
	X				X							126.6
X	X				X				X	X	X	148.7
X	X			X	X			X	X	X	X	121.6
	X	X	X		X			X	X	X	X	134.0
	X				X					X	X	121.1
	X	X	X		X			X				62.1
	X	X	X		X			X			X	36.4
	X	X	X		X			X				44.1
	X	X	X		X			X			X	32.5
	X	X	X		X	X		X			X	74.3
	X	X	X		X	X		X	X		X	22.1
		X	X		X	X	X	X				111.5
		X	X		X	X	X	X				36.4
X		X	X	X	X	X	X	X	X	X	X	112.7
X		X	X	X				X	X		X	22.5
		X	X					X				109.5
	X	X	X		X	X		X				28.1
X	X			X	X			X	X	X	X	111.0
X	X			X	X			X	X	X	X	57.3
X	X				X			X	X	X	X	108.5
X	X				X			X	X	X	X	18.6
X	X				X			X	X	X	X	54.7
	X				X							33.2
	X				X							40.7
	X				X				X	X	X	118.4
	X	X	X		X			X	X		X	107.3
		X	X		X			X				77.2

other pair should be equal to the ratio of their respective divergence times.

We used genetic estimates for the divergence time between Neandertals and modern humans and between modern human populations from sub-Saharan African versus the rest of the world to calculate a range of divergence time ratios to compare with the mean ratio based on morphology. We used the sub-Saharan African–other divergence because it is the deepest one for modern human populations (Zhivotovsky et al., 2003). Analyses of ancient Neandertal and extant human DNA (both mitochondrial and autosomal) suggest that the Neandertal and modern human lineages diverged about 800,000 to 300,000 years ago (Kriings et al., 1997; Ovchinnikov et al., 2000; Green et al., 2006; Noonan et al., 2006). Molecular estimates for the sub-Saharan African–other divergence based on numerous autosomal microsatellites place this divergence time between 50,000 and 150,000 years ago (Zhivotovsky

et al., 2003). The maximum ratio of these divergence times is 16 (800,000/50,000), and the minimum ratio is 2 (300,000/150,000), yielding a range from 16 to 2. A ratio based on morphology outside of this range would lead to rejection of the null hypothesis. This test is not particularly powerful, however, because the considerable uncertainty in the genetic estimates makes the range quite large.

Variance and shape of the divergence distribution. The thirty-seven morphological distances form a distribution, and the particular values assumed for heritability and mutation rate do not affect the shape of this distribution as long as they are reasonably constant. Under neutral divergence, the mean of this distribution is proportional to divergence time. Furthermore, this distribution is χ^2 distributed with one degree of freedom (Lande, 1979; Turelli et al., 1988; Lynch, 1989). These expectations about the distribution are the basis for the second and third tests. Note that like the previous test,

the second and third tests assume a bifurcating evolutionary history with negligible gene flow between Neandertals and modern humans.

A χ^2 distributed (with one degree of freedom) set of measurements will have a squared coefficient of variation ($[(CV)^2]$, which is a standardized measure of variation) equal to two (Turelli et al., 1988; Lynch, 1989). Thus, the second test uses an F-test of equality of variances to evaluate if the observed $(CV)^2$ is significantly different from two. The third test considers the entire shape of the distribution by using a Kolmogorov-Smirnov (KS) test to assess if the distribution (standardized by the mean of the distribution) differs in shape from a χ^2 distribution with one degree of freedom.

Patterning of the distances. The fourth test can be looked at in two different ways. Described in the framework of the morphological distances used to describe the other three tests, this test evaluates whether the magnitudes of the morphological distances are patterned by within population variation. Specifically, it tests for a relationship between the magnitude of a morphological distance and the amount of variation (eigenvalue) described by the principal component used to calculate it.

Looked at another way, this test evaluates the proportionality of \mathbf{P}_b to \mathbf{P}_w (Lande, 1979; Ackermann and Cheverud, 2002). This is a test of the neutral expectation that the between group variation should be dictated by the within group variation, accounting for integration (covariance structure). Following Ackermann and Cheverud (2002), to test for proportionality of \mathbf{P}_b to \mathbf{P}_w , we use ordinary least-squared regression to calculate a line describing the relationship between the log-transformed between Neandertal-modern human variances along each of the principal components and the corresponding log-transformed within modern human variances along these components (log-transformed eigenvalues). Proportionality, and thus neutral divergence, is rejected if the slope of the regression line (β) is significantly different from one (Ackermann and Cheverud, 2002).

If \mathbf{P}_b were proportional to \mathbf{P}_w , another prediction would be that the scores along the different principal components would be uncorrelated not just within groups (by definition) but also among groups (Ackermann and Cheverud, 2002). Unfortunately, we cannot perform this correlation test of proportionality here because we have only two groups.

Gene flow. Gene flow between Neandertals and modern humans will affect the tests differently. The test for patterning of the morphological distances (β statistic) (Ackermann and Cheverud, 2002), while originally formulated for a bifurcating model with no gene flow, also holds for some models of population history and structure that include gene flow (see Equation 16 in Rogers and Harpending, 1983). The tests based on the $(CV)^2$ and KS statistics (variance and shape) do not hold for migration models because the distribution around the expected value for the between group variance has an unknown shape that is likely to be a complex function of the effective population sizes, the number of populations connected by gene flow, the mutation rate, the migration rates, and whether an equilibrium has been reached (Rogers and Harpending, 1983).

Statistical power and comparisons among modern human populations

A final concern is the power of the statistical tests. It is possible that natural selection could cause small enough deviations from the null hypothesis that we do not have enough statistical power to detect them. There are two relevant questions here. First, how much power do the statistical tests have with large, well-preserved samples? Second, do the small Neandertal sample size, missing data, and potential for deformation further reduce the power?

To address the first question, we wanted to see if we could find situations where it was possible to detect significant deviations from the null hypothesis. We already know that the tests have at least some power because Ackermann and Cheverud (2004) were able to reject a null hypothesis of cranial divergence by genetic drift among some australopith and early *Homo* taxa using the last test discussed above (β statistic). Additionally, Lynch (1990) used tests similar to the first one described here to show that the rates of cranial evolution in many mammalian taxa are too slow to be accounted for by random genetic drift, probably as a result of stabilizing selection.

To further examine the power issue, we performed the same statistical tests that we used to compare Neandertals and modern humans on all possible pairs of populations in our modern human sample. We compared population one to two, one to three, one to four, and so on, producing, in total, 435 population pairs. These comparisons produced distributions of values for the β , $(CV)^2$, and KS statistics (the last three tests described above) for all possible pairwise comparisons among the modern human populations in our sample. Finding some significant results for these pairwise comparisons would show that the tests have some statistical power.

The answer to the second question about power is more complicated and depends on the particulars of the statistical test. The power of the tests based on the $(CV)^2$ and KS statistics is related to the number of measurements, which is the same for both Neandertals and modern humans. For the test based on the β statistic, the strength of the correlation also influences the power (Ackermann and Cheverud, 2002), but the correlation coefficient of 0.61 for the Neandertal-modern human comparison is similar to the average correlation coefficient for the comparisons among modern human populations (0.65).

As a further check, we performed simulations to determine if the pattern of missing data in the Neandertal sample made significant results less likely for the Neandertal-modern human comparisons than for the comparisons among modern human populations. For each modern human population we picked 20 individuals at random to produce samples the same size as the Neandertal sample. Next, from these samples we calculated measurement means for each population in two different ways: first, using the measurements from all 20 individuals (complete means), and second, using a subset of the measurements that mimicked the pattern of missing data in the Neandertal sample as shown in Table 1 (missing means). We then performed all the statistical tests for all pairwise comparisons

among modern human populations twice: once using the complete means and once using missing means. We found that we were no more or less likely to get significant results with the missing means than with the complete means, showing that the pattern of missing data in the Neandertal sample does not bias our results.

Finally, any deformation of the fossils would tend to increase the variance of the between Neandertal-modern human variances along the principal components, making it more likely that the null hypothesis of neutral divergence would be rejected. In sum, the tests should be equally powerful for comparisons between Neandertals and modern humans as for those among modern human populations.

As an additional step, we compared our results for modern human cranial measurements with those for a large set of microsatellite loci (Rosenberg et al., 2002). Both microsatellites and morphological measurements evolve by lengthening and shortening. Consequently, like the morphological distances based on cranial measurements, the distribution of genetic distances for a set of neutral microsatellite loci will be χ^2 distributed with one degree of freedom (Zhivotovsky and Feldman, 1995). This makes it possible to directly compare cranial measurements to microsatellites for the $(CV)^2$ and KS statistics for all possible pairwise comparisons among modern human populations. This last test evaluates the importance of natural selection in shaping cranial diversification among modern human populations. If cranial differences among modern human populations were primarily shaped by genetic drift, then we would expect our results for cranial measurements to be similar to those for microsatellites. Combined with the results of the other tests, this final test provides a benchmark for understanding any potential differences between how natural selection has acted to produce differences among modern human populations versus between Neandertals and modern humans.

Results and discussion

Statistical tests

The mean ratio for all the morphological distances of the between Neandertal–modern human distance to the between sub-Saharan African–other distance is 6.03 (± 1.98). This result is well within the range (16–2) expected under neutral divergence based on genetic estimates of divergence time. This ratio is consistent with a Neandertal modern human divergence time of about 900,000 to 300,000 years ago, when compared with the deepest within-modern human divergence range of 150,000–50,000 years ago that we discussed earlier. For the test that deals with the variance of the distribution $[(CV)^2]$, the null hypothesis is not rejected (two-tailed F-test, numerator d.f. = 36, denominator d.f. = ∞ , $(CV)^2 = 2.42$, $F = 1.21$, $P = 0.36$). Similarly, the Kolmogorov-Smirnov (KS) test, which evaluates the shape of the distribution, yields a non-significant result (KS statistic = 0.14, $P = 0.23$, corrected using a simulation for the fact that standardizing by the mean forces the observed distribution to have a mean of one). For the last test, which deals with the patterning of the

morphological distances (β statistic), the null hypothesis is also not rejected (t-test, d.f. = 35, $\beta = 1.27$, $t = 0.97$, $P = 0.34$). In sum, using four statistical tests we are unable to reject the null hypothesis of neutral divergence between Neandertals and modern humans.

Statistical power and comparisons among modern human populations

The results of the pairwise comparisons among modern human populations that we use to evaluate statistical power are striking. Using the same significance level ($\alpha = 0.05$) as for the Neandertal-modern human tests, 40% (175/435) of the pairwise comparisons among modern human populations are significant for at least one test. This demonstrates that the statistical tests are actually quite powerful, as they are able to detect deviations from the null hypothesis in the cranial variation found among modern human populations. This makes it even more noteworthy that we were unable to detect deviations for the Neandertal–modern human comparisons. The Neandertal–modern human comparisons are, in fact, quite ordinary relative to those among modern human populations (Fig. 2).

If anything, it appears as if natural selection has been more important in producing cranial differences between some pairs of modern human populations than between Neandertals and modern humans. But does modern human cranial diversity appear to have been strongly affected by natural selection? From the results of previous studies, a case can definitely be made that natural selection has shaped particular cranial measurements (Roseman, 2004; Roseman and Weaver, 2004; Harvati and Weaver, 2006a,b), but overall, cranial differences among modern human populations appear to be predominantly the result of genetic drift (Lynch, 1989; Relethford, 1994; Roseman, 2004; Roseman and Weaver, 2004). Additionally, distance matrices based on cranial measurements and presumably neutral genetic loci tend to be significantly associated with each other (González-Jose et al., 2004; Roseman, 2004; Harvati and Weaver, 2006a,b). Here we present further evidence for the importance of genetic drift in shaping modern human cranial diversity. The mean values for cranial measurements and microsatellites are markedly similar to each other for both the $(CV)^2$ and KS statistics (Table 2). This is the case even though we do not expect the results to be identical, because the cranial and microsatellite datasets are not from the same individuals or populations.

On the other hand, we did detect many significant results for the pairwise tests among the modern human populations in our sample. Simply by chance we expect 5% of the comparisons for each test to be significant just by setting α equal to 0.05. The number of significant comparisons are 25%, 5%, 19% for the tests based on the $(CV)^2$, KS, β and statistics respectively. For one of the tests (KS statistic) the number of significant results is the same as expected by chance, but for the other two tests ($[(CV)^2]$ and β statistics) there are between four and five times more significant results. Perhaps this indicates that natural selection has played a more important role in modern human cranial diversification than previously detected, but

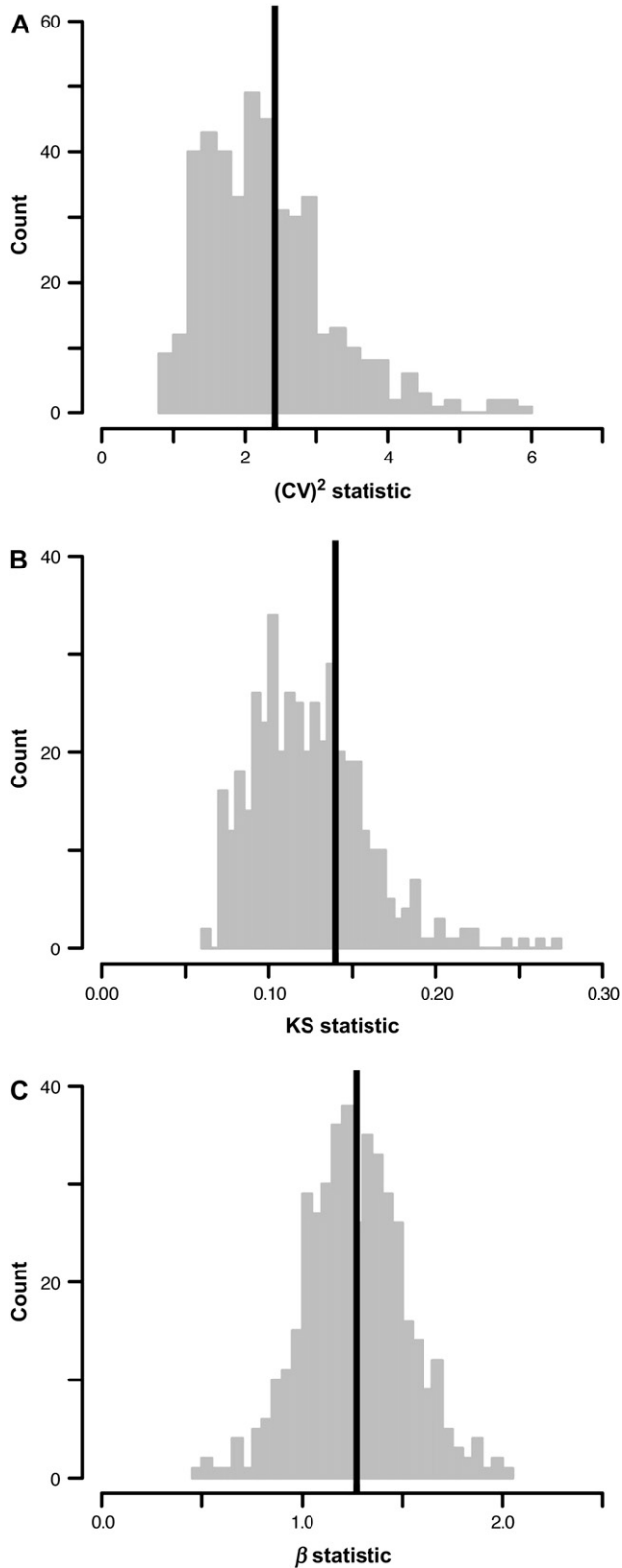


Fig. 2. Neandertal-modern human versus among modern human comparisons. **A**, $(CV)^2$ statistic. **B**, KS statistic. **C**, β statistic. Gray histograms show the distributions of these statistics for all possible pairwise comparisons among the 30 modern human populations in our sample. Vertical black lines indicate the value for each statistic for the Neandertal-modern human comparison.

Table 2
Microsatellites versus cranial measurements

	$(CV)^2$	KS
Microsatellites		
Dinucleotide	2.15 (0.79)	0.11 (0.03)
Trinucleotide	2.03 (0.77)	0.10 (0.03)
Tetranucleotide	2.82 (1.48)	0.05 (0.02)
Morphology		
Cranial measurements	2.29 (0.88)	0.12 (0.03)

Means for the pairwise among modern human population comparisons, with standard deviations in parentheses (note that these are not appropriate for statistical testing because the comparisons are not independent).

It is important to remember that deviations from the model of heredity or gene flow between populations could also produce significant results. Additionally, the pairwise comparisons among modern human populations are not independent, so the significance percentages are not as robust as they would appear from the large numbers of pairwise comparisons. At minimum these results serve to emphasize the sensitivity of the tests used in our study and make it even more noteworthy that we did not detect deviations from the null hypothesis for the Neandertal-modern human comparisons.

Cranial integration in Neandertals and modern humans

Our analyses assume that the phenotypic (\mathbf{P}_w) and genetic variance-covariance matrices (\mathbf{G}) for Neandertal crania are proportional to those for modern human crania. Proportionality of variance-covariance matrices results from similar cranial integration, because differences in integration will manifest themselves in different variances and covariances. Multivariate comparisons that assume that variance-covariance patterns in modern humans adequately describe those in Neandertals have been criticized because variance-covariance patterns in Neandertals may, in fact, have been quite different (e.g., Wolpoff, 1993; Ackermann, 2005). We agree that this criticism is, in principle, valid for many multivariate comparisons of fossil and extant taxa, but it does not apply to our study. In our case, the assumption of similar variance-covariance patterns is part of the null hypothesis, so we actually test this assumption, and we are unable to reject the null hypothesis using multiple statistical tests that appear to be quite powerful. The implication is that variance-covariance patterns in Neandertals and modern humans may not be so different, and cranial integration in Neandertals and modern humans may actually be quite similar, at least relative to differences among modern humans.

Distribution of morphological differences between Neandertals and modern humans

Investigations of Neandertal taxonomy tend to either focus on the morphological features for which Neandertals and modern humans are extremely different or on the features for which they are quite similar. If Neandertal and modern human cranial differences were shaped by genetic

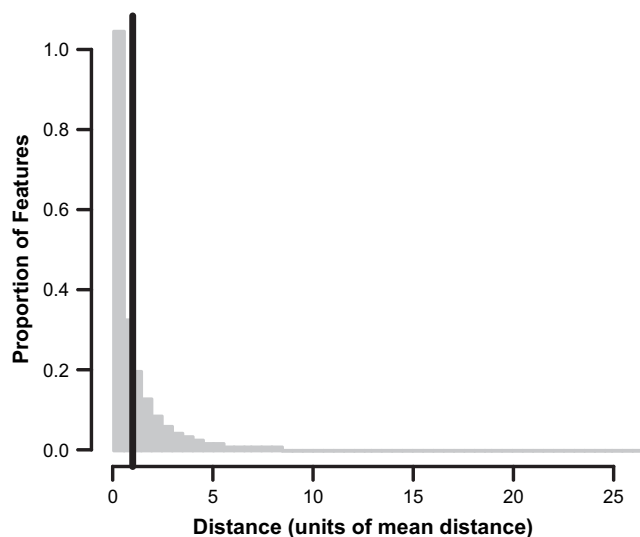


Fig. 3. Distribution of morphological distances between Neandertals and modern humans expected for divergence by genetic drift (χ^2 distribution with one degree of freedom). Vertical black line indicates the mean of the distribution. Morphological distances are given in multiples (units) of the mean morphological distance.

drift, then when we consider all the morphological features together, we expect a particular distribution of morphological distances (χ^2 distribution with one degree of freedom; Fig. 3). (As a practical aside, all the features must be independent within groups for the predictions about the distribution to hold, so the data must usually be transformed first, such as by principal components analysis based on the pooled within group variance-covariance matrix, before it can be analyzed.) For this distribution, the mean distance is expected to be proportional to the divergence time between the Neandertal and modern human lineages (the basis for statistical test one). However, because there is substantial variation around this expectation and the distribution is highly asymmetrical, there will be a large number of features for which Neandertals and modern humans are quite similar (distances close to zero, left side of Fig. 3). There will also be some features for which Neandertals and modern humans are extremely different (large distances, tail of the distribution on the right side of Fig. 3). While concentrating on certain subsets of features that may be useful for taxonomy, if we want to understand the evolution of Neandertals and modern humans, then we need to examine the whole distribution rather than focusing on either one end or the other.

Conclusions

Our results show that diversifying natural selection has not left an obvious signature on differences between Neandertals and modern humans in cranial morphology. Detailed functional analyses also fail to support several candidate adaptive explanations for Neandertal cranial form (Antón, 1994; Franciscus, 2003; O'Connor et al., 2005). In contrast, Ackermann

and Cheverud (2004) did find evidence that natural selection shaped facial differences between australopiths and early *Homo* and among australopith taxa, although not among taxa of early *Homo*. It may be that with the advent of the genus *Homo* natural selection played a much more limited role in hominin cranial diversification (Ackermann and Cheverud, 2004).

There are, however, two important caveats to keep in mind. First, we are referring to natural selection's role in producing differences between Neandertal and modern human crania, not its role in shaping Neandertal and modern human cranial morphology *per se*. Natural selection, in the form of stabilizing selection to maintain a properly functioning cranium, may have been important in maintaining similarities between Neandertals and modern humans. Second, we cannot rule out a role for natural selection in the diversification of behavioral, physiological, or other morphological features. For instance, body proportions and other aspects of Neandertal postcranial anatomy may be adaptations to the glacial climate of Pleistocene Europe (Trinkaus, 1981; Ruff, 1994; Holliday, 1997a,b; Pearson, 2000; Weaver, 2003), and our analyses are limited to external features that can be quantified using standard osteometric tools. Other aspects of cranial morphology such as internal structures or features that are difficult to measure may prove to have been shaped by natural selection. Nevertheless, the results of this investigation, combined with those from functional analyses, imply that adaptive accounts are not necessary to explain Neandertal and modern human cranial differences and that genetic drift is the most strongly supported explanation for these differences.

Acknowledgments

We thank S. Leigh, J. Polk, J.-J. Hublin, T. Steele, M. Stoneking, and members of the University of Illinois Biological Anthropology Journal Club for feedback; the late W. Howells and N. Rosenberg for making their raw data publicly available; the institutes that made their fossil material available for study; The Natural History Museum (London) for the picture for Fig. 1; and the Max Planck Society, the University of Illinois, the Medical Research Council, and the Natural History Museum (London) for financial support.

References

- Ackermann, R.R., 2005. Variation in Neandertals: a response to Harvati (2003). *J. Hum. Evol.* 48, 643–646.
- Ackermann, R.R., Cheverud, J.M., 2002. Discerning evolutionary processes in patterns of Tamarin (*Genus Saguinus*) craniofacial variation. *Am. J. Phys. Anthropol.* 117, 260–271.
- Ackermann, R.R., Cheverud, J.M., 2004. Detecting genetic drift versus selection in human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101 (52), 17946–17951.
- Ackermann, R.R., Krovitz, G.E., 2002. Common patterns of facial ontogeny in the hominid lineage. *Anat. Rec.* 269, 142–147.
- Antón, S.C., 1994. Mechanical and other perspectives on Neandertal craniofacial morphology. In: Corruccini, R.S., Ciochon, R.L. (Eds.), *Integrative Paths to the Past: Paleoanthropological Advances in Honor of F. Clark Howell*. Prentice Hall, Englewood Cliffs, pp. 677–695.

- Antón, S.C., 1996. Tendon-associated bone features of the masticatory system in Neandertals. *J. Hum. Evol.* 31, 391–408.
- Boule, M., 1911–13. L'Homme fossile de La Chapelle-aux-Saints. *Annales de Paléontologie*, 6–8.
- Brace, C.L., 1964. The fate of the “classic” Neanderthals: a consideration of hominid catastrophism. *Curr. Anthropol.* 5 (1), 3–43.
- Brose, D.S., Wolpoff, M.H., 1971. Early Upper Paleolithic man and late Middle Paleolithic tools. *Am. Anthropol.* 73, 1156–1194.
- Cheverud, J.M., 1988. A comparison of genetic and phenotypic correlations. *Evolution* 42, 958–968.
- Coon, C.S., 1962. *The Origin of Races*. Alfred A. Knopf, New York.
- Demes, B., 1987. Another look at an old face: biomechanics of the Neandertal facial skeleton reconsidered. *J. Hum. Evol.* 16, 297–303.
- Falconer, D.S., 1981. *Introduction to Quantitative Genetics*, second ed. Longman Group Limited, Essex.
- Franciscus, R.G., 2002. Neanderthals. In: Pagel, M. (Ed.), *Encyclopedia of Evolution*. Oxford University Press, Oxford, pp. 493–497.
- Franciscus, R.G., 2003. Internal nasal floor configuration in *Homo* with special reference to the evolution of Neandertal facial form. *J. Hum. Evol.* 44, 701–729.
- González-Jose, R., Van der Molen, S., Gonzales-Perez, E., Hernandez, M., 2004. Patterns of phenotypic covariation and correlation in modern humans as viewed from morphological integration. *Am. J. Phys. Anthropol.* 123, 69–77.
- Green, R.E., Krause, J., Ptak, S.E., Briggs, A.W., Ronan, M.T., Simons, J.F., Du, L., Egholm, M., Rothberg, J.M., Paunovic, M., Pääbo, S., 2006. Analysis of one million base pairs of Neandertal DNA. *Nature* 444, 330–336.
- Harvati, K., Frost, S.R., McNulty, K.P., 2004. Neandertal taxonomy reconsidered: implications of 3D primate models of intra- and interspecific differences. *Proc. Natl. Acad. Sci. U.S.A.* 101 (5), 1147–1152.
- Harvati, K., Weaver, T.D., 2006a. Human cranial anatomy and the differential preservation of population history and climate signatures. *Anat. Rec.* 288A, 1225–1233.
- Harvati, K., Weaver, T.D., 2006b. Reliability of cranial morphology in reconstructing Neandertal phylogeny. In: Harvati, K., Harrison, T. (Eds.), *Neandertals Revisited: New Approaches and Perspectives*. Springer, New York, pp. 239–254.
- Holliday, T.W., 1997a. Body proportions in Late Pleistocene Europe and modern human origins. *J. Hum. Evol.* 32, 423–447.
- Holliday, T.W., 1997b. Postcranial evidence of cold adaptation in European Neandertals. *Am. J. Phys. Anthropol.* 104, 245–258.
- Howell, F.C., 1952. Pleistocene glacial ecology and the evolution of “classic” Neandertal man. *Southwest. J. Anthropol.* 8 (4), 377–410.
- 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.
- Hublin, J.-J., 1998. Climatic changes, paleogeography, and the evolution of the Neandertals. In: Akazawa, T., Aoki, K., Bar-Yosef, O. (Eds.), *Neandertals and Modern Humans in Western Asia*. Plenum Press, New York, pp. 295–310.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Kreitman, M., 1996. The neutral theory is dead. Long live the neutral theory. *Bioessays* 18, 678–682.
- Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., Pääbo, S., 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90, 19–30.
- Laitman, J.T., Heimbuch, R.C., Crelin, E.S., 1979. The basicranium of fossil hominids as an indicator of their upper respiratory systems. *Am. J. Phys. Anthropol.* 51, 15–34.
- Lande, R., 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30 (2), 314–334.
- Lande, R., 1977. Statistical tests for natural selection on quantitative characters. *Evolution* 31 (2), 442–444.
- Lande, R., 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33 (1), 402–416.
- Lande, R., 1980. Genetic variation and phenotypic evolution during allopatric speciation. *Am. Nat.* 116 (4), 463–479.
- Lieberman, D.E., 1998. Sphenoid shortening and the evolution of modern human cranial shape. *Nature* 393, 158–162.
- Lieberman, D.E., Krovitz, G.E., McBratney-Owen, B., 2004. Testing hypotheses about tinkering in the fossil record: the case of the human skull. *J. Exp. Zool. (Mol. Dev. Evol.)* 302B, 284–301.
- Lieberman, D.E., McBratney, B.M., Krovitz, G., 2002. The evolution and development of cranial form in *Homo sapiens*. *Proc. Natl. Acad. Sci. U.S.A.* 99 (3), 1134–1139.
- Lieberman, P., Crelin, E.S., 1971. On the speech of Neandertal man. *Linguistic Inquiry* 2 (2), 203–222.
- Lynch, M., 1989. Phylogenetic hypotheses under the assumption of neutral quantitative-genetic variation. *Evolution* 43 (1), 1–17.
- Lynch, M., 1990. The rate of morphological evolution in mammals from the standpoint of neutral expectation. *Am. Nat.* 136 (6), 727–741.
- Lynch, M., Hill, W.G., 1986. Phenotypic evolution by neutral mutation. *Evolution* 40 (5), 915–935.
- Noonan, J.P., Coop, G., Kudaravalli, S., Smith, D., Krause, J., Alessi, J., Chen, F., Platt, D., Pääbo, S., Pritchard, J.K., Rubin, E.M., 2006. Sequencing and analysis of Neandertal genomic DNA. *Science* 314, 1113–1118.
- O'Connor, C.F., Franciscus, R.G., Holton, N.E., 2005. Bite force production capability and efficiency in Neandertals and modern humans. *Am. J. Phys. Anthropol.* 127, 129–151.
- Ovchinnikov, I.V., Götherström, A., Romanova, G.P., Kharitonov, V.M., Lidén, K., Goodwin, W., 2000. Molecular analysis of Neandertal DNA from the northern Caucasus. *Nature* 404, 490–493.
- Pearson, O.M., 2000. Activity, climate, and postcranial robusticity. *Curr. Anthropol.* 41 (4), 569–607.
- Ponce de León, M.S., Zollikofer, C.P.E., 2001. Neandertal cranial ontogeny and its implications for late hominid diversity. *Nature* 412, 534–538.
- Rak, Y., 1986. The Neandertal: A new look at an old face. *J. Hum. Evol.* 15, 151–164.
- Relethford, J.H., 1994. Craniometric variation among modern human populations. *Am. J. Phys. Anthropol.* 95, 53–62.
- Roff, D.A., 1996. The evolution of genetic correlations: an analysis of patterns. *Evolution* 50 (4), 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. U.S.A.* 101 (35), 12824–12829.
- Roseman, C.C., Weaver, T.D., 2004. Multivariate apportionment of global human craniometric diversity. *Am. J. Phys. Anthropol.* 125, 257–263.
- Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., Feldman, M.W., 2002. Genetic structure of human populations. *Science* 298, 2381–2385.
- Ruff, C.B., 1994. Morphological adaptation to climate in modern and fossil hominids. *Yearb. Phys. Anthropol.* 37, 65–107.
- Schaaffhausen, H., 1857. *Correspondenzblatt. Verhandlungen des Naturhistorischen Vereins der Preussischen Rheinlande und Westfalens* 14, 50–52.
- Sergi, S., 1958. Die Neandertalischen Palaeanthropen in Italien. In: von Koenigswald, G.H.R. (Ed.), *Hundert Jahr Neandertaler*. Kemink en Zoon, Utrecht, pp. 38–51.
- Smith, F.H., 1991. The Neandertals: Evolutionary dead ends or ancestors of modern people? *J. Anthropol. Res.* 47, 219–238.
- Smith, F.H., Paquette, S.P., 1989. The adaptive basis of Neandertal facial form, with some thoughts on the nature of modern human origins. In: Trinkaus, E. (Ed.), *The Emergence of Modern Humans: Biocultural Adaptations in the Later Pleistocene*. Cambridge University Press, Cambridge, pp. 181–210.
- Spencer, M.A., Demes, B., 1993. Biomechanical analysis of masticatory system configuration in Neandertals and Inuits. *Am. J. Phys. Anthropol.* 91, 1–20.
- Spoor, F., O'Higgins, P., Dean, C., Lieberman, D.E., 1999. Anterior sphenoid in modern humans. *Nature* 397, 572.
- Stringer, C.B., 1974. Population relationships of later Pleistocene hominids: A multivariate study of available crania. *J. Archaeol. Sci.* 1, 317–342.

- Trinkaus, E., 1981. Neanderthal limb proportions and cold adaptation. In: Stringer, C.B. (Ed.), *Aspects of Human Evolution*. Taylor and Francis, London, pp. 187–224.
- Trinkaus, E., 1987. The Neanderthal face: evolutionary and functional perspectives on a recent hominid face. *J. Hum. Evol.* 16, 429–443.
- Turelli, M., Gillespie, J.H., Lande, R., 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. *Evolution* 42 (5), 1085–1089.
- Walsh, B., 2001. Quantitative genetics in the age of genomics. *Theor. Popul. Biol.* 59, 175–184.
- Weaver, T.D., 2003. The shape of the Neanderthal femur is primarily the consequence of a hyperpolar body form. *Proc. Natl. Acad. Sci. U.S.A.* 100 (12), 6926–6929.
- Wolpoff, M.H., 1993. Reply to Dr. Foote. *Am. J. Phys. Anthropol.* 90, 381–384.
- Zhivotovsky, L.A., Feldman, M.W., 1995. Microsatellite variability and genetic distances. *Proc. Natl. Acad. Sci. U.S.A.* 92, 11549–11552.
- Zhivotovsky, L.A., Rosenberg, N.A., Feldman, M.W., 2003. Features of evolution and expansion of modern humans, inferred from genomewide microsatellite markers. *Am. J. Hum. Genet.* 72, 1171–1186.