

No Evidence of Neandertal Admixture in the Mitochondrial Genomes of Early European Modern Humans and Contemporary Europeans

Silvia Ghirotto, Francesca Tassi, Andrea Benazzo, and Guido Barbujani*

Department of Biology and Evolution, University of Ferrara, Italy

KEY WORDS ancient DNA; population genetics; Approximate Bayesian Computations; coalescent simulations; admixture

ABSTRACT Neandertals, the archaic human form documented in Eurasia until 29,000 years ago, share no mitochondrial haplotype with modern Europeans. Whether this means that the two groups were reproductively isolated is controversial, and indeed nuclear data have been interpreted as suggesting that they admixed. We explored the range of demographic parameters that may have generated the observed mitochondrial diversity, simulating 3.0 million genealogies under six models differing as for the relationships among contemporary Europeans, Neandertals, and Upper Palaeolithic European early modern humans (EEMH), who coexisted with Neandertals for millennia. We compared by Approximate Bayesian Computations the simulation results with mito-

chondrial diversity in 7 Neandertals, 3 EEMH, and 150 opportunely chosen modern Europeans. A model of genealogical continuity between EEMH and contemporary Europeans, with no Neandertal contribution, received overwhelming support from the analyses. The maximum degree of Neandertal admixture, under the model of gene flow supported by nuclear data, was estimated at 1.5%, but this model proved 20–32 times less likely than a model without any gene flow. Nuclear and mitochondrial evidence might be reconciled if smaller population sizes led to faster lineage sorting for mitochondrial DNA, and Neandertals shared a longer period of common ancestry with the non-African's than with the African's ancestors. *Am J Phys Anthropol* 146:242–252, 2011. ©2011 Wiley-Liss, Inc.

Two anatomically different human forms, the archaic Neandertals and the European early modern humans of the Upper Palaeolithic (hereafter EEMH, sometimes referred to as Cro-Magnoids), coexisted in Europe for millennia. Fossil and archaeological data document a progressive withdrawal of Neandertal communities as EEMH expanded; the Neandertal anatomy and their artifacts disappeared from the record at a moment in time which is traditionally placed around 29,000 years ago (Mellars, 1992; Mellars, 2006).

The biological relationships between these human forms are controversial. For many years, the debate focused on the relative merits of two classes of models, Regional Continuity, and Replacement. According to the former, anatomically archaic hominids of the Old World formed a subdivided population, within which a transition from archaic to modern morphology occurred; under the Replacement (or “Out of Africa”) model, anatomically modern humans expanded from Africa replacing all archaic groups. More recently, Assimilation models emerged, i.e., the idea that contemporary populations are largely descended from an anatomically modern group expanding from Africa, but Neandertals contributed to the modern European gene pool to a non-negligible extent (Relethford, 2001; Trinkaus, 2007). Thus, the main controversial point became how much genetic exchange, if any, there has been between the two human forms.

Complete replacement is impossible to demonstrate, because the same genetic consequences are expected both without admixture and with extremely low levels of admixture. However, most morphological and genetic evidence seems to agree with the predictions of a model in which anatomically-modern people and archaic

humans did not hybridize. With very few possible exceptions (see e.g., Zilhão, 2006), European fossils are clearly classified as either archaic or modern; the absence of intermediate morphologies suggests that levels of admixture were extremely low or nil (Tattersall and Schwartz, 1999). Neandertal mitochondrial DNA (mtDNA) sequences fall out of the range of current European variation (Krings et al., 1997; Briggs et al., 2009), so that even a small mitochondrial contribution of Neandertals to the modern human gene pool appeared unlikely (Currat and Excoffier, 2004; Belle et al., 2009). By contrast, in the first survey of the whole Neandertal nuclear genome, patterns of allele sharing with modern humans have been interpreted as suggesting gene flow from Neandertals into the ancestors of modern non-Africans, before the Eurasian populations separated (Green et al., 2010).

No clear consensus has yet emerged from studies of modern DNA diversity either; compare e.g., Labuda

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Italian Ministry of Universities (MIUR); Grant numbers: PRIN 2008.

*Correspondence to: Guido Barbujani, Department of Biology and Evolution, University of Ferrara, Via L. Borsari 46, 44121 Ferrara, Italy. E-mail: g.barbujani@unife.it

Received 23 March 2011; accepted 10 May 2011

DOI 10.1002/ajpa.21569

Published online 24 August 2011 in Wiley Online Library (wileyonlinelibrary.com).

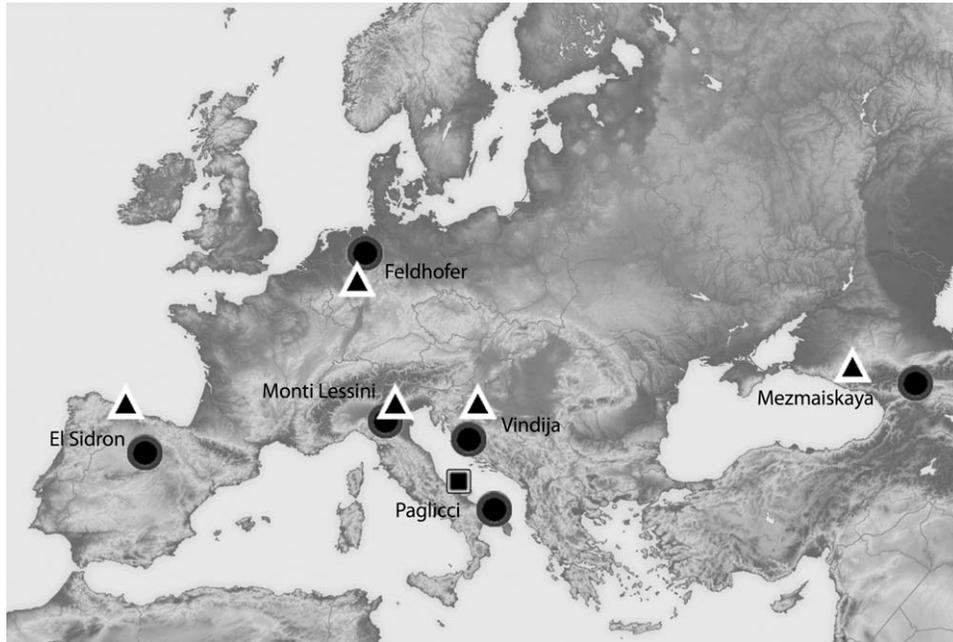


Fig. 1. Geographic localization of the samples considered. Triangles: Neandertals; Squares: Early European Modern Humans, or EEMH; Circles: Modern Europeans.

et al. (2000), Plagnol and Wall (2006) and Templeton (2007), with Excoffier (2002), Hodgson and Disotell (2008) and Jakobsson et al. (2008). In most cases, the evidence suggesting some degree of genetic exchange between archaic and modern human forms is a deep basal node in the gene tree, reflecting high levels of haplotype divergence. Such deep splits, accompanied by high linkage disequilibrium, are expected if there was admixture between groups that have long evolved in isolation (Templeton, 2005), but may also reflect genetic structuring of the common ancestors of modern humans and Neandertals (see e.g., Currat et al., 2006). Accordingly, many regard the current genetic data as insufficient to discriminate between models.

Better data may take long to accumulate, but more refined biostatistical approaches are already available. In this study, we compared for the first time ancient and modern mtDNA sequences under the framework of Approximate Bayesian Computations (ABC). The available mitochondrial data allow one to test hypotheses for which fossil evidence is inconclusive, and which cannot be currently tested at the nuclear level. We could thus evaluate posterior probabilities for a set of models differing as for the genealogical relationships of two ancient (Neandertal and EEMH) and six modern Europeans populations. We also estimated the demographic parameters of the best-fitting model, and demonstrated that the data have enough statistical power to identify the correct model. These results restrict the range of hypotheses potentially accounting for the genetic relationships between Neandertal and modern people, and show how analyses of nuclear and mitochondrial diversity can be reconciled without invoking admixture processes.

MATERIALS AND METHODS

The data

We investigated diversity in the mitochondrial hypervariable region I, spanning 360 bp, in 150 modern and

10 ancient individuals. The latter are seven Neandertal individuals: two from Feldhofer in Germany (Krings et al., 1997; Schmitz et al., 2002), two from the Vindija Cave in Croatia (Krings et al., 2000), one from Mezmaiskaya in the Russian Caucasus (Ovchinnikov et al., 2000), one from Monti Lessini in Italy (Caramelli et al., 2006), and one from El Sidrón cave, Asturias, Spain (Lalueza-Fox et al., 2006), plus 3 EEMH sequences from the Paglicci cave, Italy (Caramelli et al., 2003; Caramelli et al., 2008). No other sequences of the entire mitochondrial hypervariable region I are available for European Neandertals or EEMH (Hodgson and Disotell, 2008).

To have similar effects of geography for ancient and modern populations, we chose modern samples from Germany (Richards et al., 1996), Croatia (Babalini et al., 2005), the Caucasus (Nasidze and Stoneking, 2001), two regions of Italy (Babalini et al., 2005) and Spain (Corte-Real et al., 1996) (see Fig. 1). Sample sizes were very different; to avoid any resulting confounding effect (such as those due to an excessive weight of the largest samples upon estimates of haplotype sharing), we randomly resampled 25 sequences from each modern population. In a preliminary step, we made sure that the summary statistics calculated on the resampled datasets are consistent with those calculated on the complete datasets.

Serial coalescent simulations

Three million mitochondrial genealogies were generated by the serial coalescent algorithm implemented in the Bayesian version of SERIALSIMCOAL (Anderson et al., 2005). With this program one can generate multiple gene genealogies according to any demographic model. Suppose that one has samples of sizes $n_0, n_1, n_2, \dots, n_k$ from populations studied $t_0, t_1, t_2, \dots, t_k$ generations ago. The program generates genealogies proceeding backwards in time, starting with n_0 samples in the present (t_0) and adding n_1, n_2, \dots, n_k samples at the appropriate moments in the past. The genealogies were extended

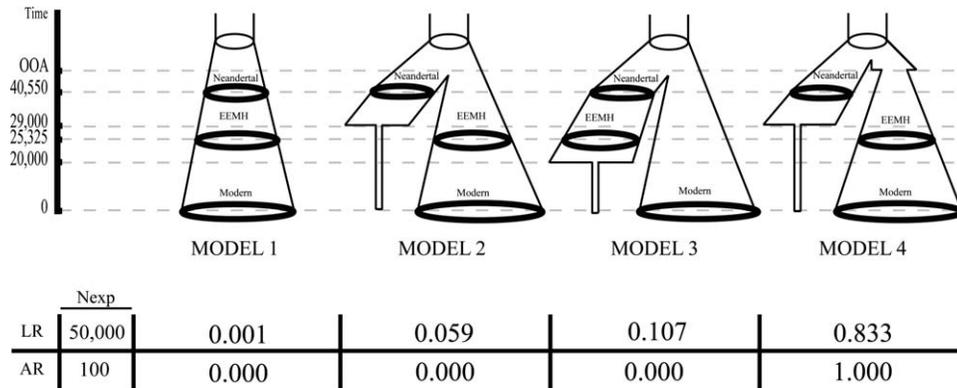


Fig. 2. Schematic presentation of the six models tested. Numbers on the Y-axis refer to years from the present. OOA (Out of Africa) is the time of the dispersal from Africa. The posterior probability is given under each model, according to two criteria: AR, or acceptance-rejection, and LR, or weighted multinomial logistic regression. *Nexp* is the number of experiments considered.

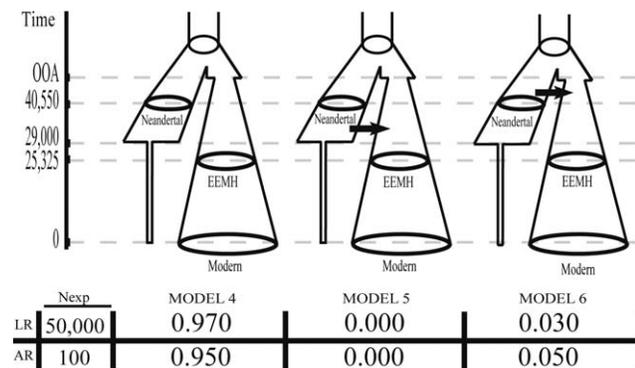


Fig. 3. Comparison between Models 4, 5 and 6. The arrows in Models 5 and 6 represent gene flow from the Neandertal to the EEMH population (Model 5) and gene flow from Neandertal to the ancestors of modern European and Asian populations (Model 6).

backwards until, through a series of coalescence events, they reached their most recent common ancestor, or MRCA. Mutations were then randomly distributed onto the tree, under a finite-site model with two potential allelic states for each site, a transition bias = 0.9375 and a rate-heterogeneity parameter = 0.26 (see Belle et al., 2009 and reference therein).

Demographic models and priors

The six models are outlined in Figures 2 and 3. Model 1 assumes genealogical continuity between Neandertal, EEMH, and modern samples; under Model 2, the Neandertal lineage separates from the lineage leading to EEMH and modern Europeans; under Model 3 the EEMH population is descended from Neandertal ancestors, whereas the modern populations are part of another lineage. Model 4 resembles Model 2, but the lineage that gave rise to EEMH and modern Europeans undergoes a founder effect associated with the dispersal from Africa. Models 5 and 6 add gene flow from the Neandertals, either (Model 5) during the maximum span of the possible coexistence of Neandertals and EEMH, 42,000 to 30,000 years ago, or (Model 6) starting 80,000 years ago, as suggested by Green et al. (2010). In all simulations, the modern samples were placed at genera-

tion 0. The EEMH and Neandertal samples were at generations 1,013 and 1,622, corresponding to the average age of the respective specimens, 25,325 and 40,550 years, assuming a generation time = 25 years (Currat and Excoffier, 2004; Fenner, 2005; Noonan et al., 2006; Fagundes et al., 2007). All population sizes increased exponentially through time, at constant rate, starting >1,622 generations ago.

Under Model 2 the Neandertal lineage got extinct 1,160 generations (29,000 years) ago (Mellars, 1992); other authors (Walker et al., 2008; Zilhão et al., 2010) proposed that this has happened earlier, around 37,000 years ago in the Iberian peninsula, but a well-dated specimen shows that Neandertals were still present, at least in the Caucasus, 29,000 years ago (Ovchinnikov et al., 2000). At any rate, choosing a late date of Neandertal disappearance increases the time interval through which there might have been contact with EEMH, thus favoring the admixture model. Under Model 3, in which EEMH and Neandertals were in the same lineage, both became extinct at an arbitrary time, 20,000 years ago. In both cases, the date of extinction has no effect on the results of the tests, in so far as it is more recent than the age of the youngest specimen.

Under Models 4–6, we added to Model 2 a founder effect in the Cro-Magnoid and Modern lineage, bringing the population size to 500 at a moment between 50,000 and 80,000 years ago (Liu et al., 2006; Fagundes et al., 2007). A complete description of the prior information considered is in Supporting Information Tables 1 and 2.

Summary statistics

Internal genetic diversity was summarized by the number of different haplotypes, the average pairwise sequence difference, and the haplotype diversity, calculated with Arlequin ver. 3.11 (Excoffier et al., 2005). We compared pairs of samples in two ways: (a) by estimating F_{ST} (Hudson et al., 1992); (b) by classifying the segregating sites into four categories, namely (1, 2) those that are polymorphic in one population and monomorphic in the other (i.e., exclusive sites for population 1 or 2); (3) polymorphic sites shared between populations 1 and 2 (shared differences); (4) fixed differences between populations 1 and 2 (Wakeley and Hey, 1997; Leman et al., 2005; Becquet and Przeworski, 2007). In this way, we

TABLE 1. Observed summary statistics describing genetic variation within and between samples: Neandertal, NE; Early European modern humans, EEMH; Modern Europeans, ME

	NE	EEMH	ME		
Number of sequences	7	3	150		
Number of distinct haplotypes	6	2	115		
Segregating site	17	1	92		
Mean pairwise difference	5.238	0.667	4.549		
Haplotype diversity	0.952	0.667	0.990		
Fst Hudson					
NE/EEMH 0.8600	EEMH/NE 0.0966	ME/NE 0.7839			
Class of Segregating Sites					
SHARED SITES			FIXED SITES		
ME_EEMH 1	ME_NE 9	EEMH_NE 0	ME_EEMH 0	ME_NE 7	EEMH_NE 15
EXCLUSIVE SITES					
ME_EEMH 91	EEMH_NE 0	ME_NE 83	NE_ME 8	EEMH_NE 1	NE_EEMH 17

obtained 12 statistics (four counts of segregating sites, times three pairwise comparisons) (Table 1).

These statistics were chosen from a larger set, in order to capture the whole information contained in the data. Indeed, the larger the number of summary statistics, the larger the statistical noise included in the posterior estimation (known as “curse of dimensionality”) (Joyce and Marjoram, 2008). To this aim, we assessed by principal component analysis the correlation between each descriptor of genetic diversity and the genetic diversity generated by our simulations, to choose only those statistics having a substantial impact on the inference.

Approximate Bayesian Computations

All the following procedures were developed in the R environment (R Development Core Team, 2008) using scripts available at <http://www.rubic.rdg.ac.uk/~mab/stuff/>. The ABC procedure included three main steps (Beaumont et al., 2002). First, for each model, 500,000 genealogies were simulated (for a total of 3,000,000 experiments), considering as random variables the demographic and evolutionary parameters of the model. Therefore, for every simulation experiment, these values were chosen at random from the corresponding prior distributions. Next, we summarized the genetic variation of the samples calculating the same set of statistics in the observed data and in each simulated dataset. Finally, we calculated for each experiment a Euclidean distance between observed and simulated statistics, thus ordering the experiments according to their distance from the observed dataset. The choice of the best model and the parameter estimation were based on the subset of simulation experiments producing the shortest Euclidean distances.

Model selection

We compared the posterior probabilities of the models in two ways, using the calmod function, also available at <http://www.rubic.rdg.ac.uk/~mab/stuff/>. The first criterion is a simple acceptance-rejection procedure (AR) (Pritchard et al., 1999). For each model, we initially counted the number of simulations (n_i) which were found among the N simulations with the shortest Euclidean distance. The posterior probability for the i -th model was

simply n_i/N . This method is considered reliable only when based on a small set of simulations showing an excellent fit with the observed data (Beaumont, 2008); in this case, we chose to retain the 100 simulations producing statistics closest to the observed statistics. We also resorted to a second method, estimating posterior probabilities by weighted multinomial logistic regression (LR) (Beaumont, 2008). Under this procedure, each model represents a categorical dependent variable Y_i (where i is again the identity number of the model), and the summary statistics are the predictive variables. The probability of the model is evaluated at the point corresponding to the observed vector of summary statistics. For this calculation we retained the 50,000 simulation experiments associated with the shortest Euclidean distances.

Parameter estimation

Within models, the parameters of the 1,000 experiments showing the shortest Euclidean distances between simulated and observed statistics were logtan transformed (Hamilton et al., 2005); we then calculated a weighted local regression, using summary statistics as predictors to adjust the parameter values towards the values expected in correspondence of the observed summary statistics (Beaumont et al., 2002). We thus obtained the posterior distributions of four classes of parameters, namely effective population sizes, separation times, mutation rates and migration rates.

Posterior predictive test and quality of the estimates

Next, we tested whether the model we chose could indeed generate patterns of genetic diversity resembling the observed ones. For that posterior predictive test (Gelman et al., 2004) we simulated 10,000 datasets according to the model with the highest probability, using the estimated posterior parameter distribution. We then estimated 9 additional descriptors of genetic diversity, namely the number of segregating sites within each population and the haplotype sharing between samples, which had not been considered during the inferential step, and compared them with the observed ones. If the model is realistic and our posterior distributions esti-

TABLE 2. Demographic parameters estimated under Model 4: prior distribution (U: uniform in all cases), median, mode, lower (0.05) and upper (0.95) limits of the 90% credible interval and coefficient of determination (R^2)

Model 4	Prior distribution	Median	Mode	0.05	0.95	R^2
Time MRCA	*	16,433	16,079	9,586	25,086	0.73
Ne Modern	{U: 1,000,000–10,000,000}	4,173,069	1,027,003	1,000,000	7,945,095	0.02
Ne Neandertal (before extinction)	{U: 5,000–100,000}	32,263	15,676	4,375	90,080	0.29
Time Out of Africa	{U: 50,000–80,000}	2,725	3,037	2,106	3,200	0.07
Separation time	{U: 80,025–900,000}	11,785	11,031	3,877	25,848	0.62
Mutation Rate	{U: 0.0002–0.008}	0.0008	0.0007	0.0006	0.0011	0.90
Ancestral Ne EEMH	{U: 5–5,000}	2,071	5	5	4,025	0.03
Ancestral Ne Neandertal	{U: 5–5,000}	766	5	5	4,105	0.10

The time to the most recent common ancestor, Time MRCA was estimated from the simulated data and not extracted from a prior distribution.

mates are plausible, summary statistics inferred from the simulated and observed datasets should not significantly differ. A posterior predictive P -value for each summary statistics was thus estimated, and these probabilities were combined into a global P -value (see Ghirotto et al., 2010), taking into account nonindependence of the individual statistics (Voight et al., 2005).

Finally, we asked whether models could actually be discriminated on the basis of the available data. To answer, we simulated 1,000 datasets from the prior distribution of each model and we treated them as observed datasets in an ABC analysis using previously simulated models. For each dataset we calculated Type I error as the number of experiments in which the simulated model was not recognized by the model selection procedure (AR), or Type I Error.

To assess the reliability of the inferred parameters we calculated the coefficient of determination (R^2 , indicating the percentage of the parameter's variance explained by the summary statistics we used), the relative bias and the relative root mean square error (RMSE), that allow to evaluate biases toward overestimation and underestimation of parameters point estimate, and the 95% coverage for each estimated parameter (see Neuschwander et al., 2008 and, for an application, Ghirotto et al., 2010).

RESULTS

Model selection

Table 1 reports the statistics summarizing mtDNA sequence variation in 7 Neandertal (NE), 3 EEMH and 150 Modern Europeans (ME). Of the demographic models initially considered (see Fig. 2) Model 4 showed by far the highest ability to reproduce the observed variation, reaching probabilities $>83\%$ under the LR approach and $= 100\%$ under the AR approach. This means that almost all simulated datasets showing good agreement with the observed data had been generated assuming independent genealogies for the NE versus the EEMH and ME mtDNAs. Moreover, none of the best-fitting experiments was simulated assuming genetic continuity between NE and ME or EEMH, since Model 1 has $P = 0$ under both methods used for comparison.

We then added the possibility of migration from the NE into the EEMH between 42,000 and 30,000 years ago (Model 5), or during the early stage of the dispersal from Africa (Model 6), as suggested by Green et al. (2010). In the comparison of Models 4, 5, and 6, under both estimation procedures, Model 4 reached probabilities between 95% and 97%, appearing 20–32 times as likely as Model 6 (see Fig. 3).

Parameter estimation

Table 2 shows the posterior parameter estimates for Model 4. The age of the Most Recent Common Ancestor (MRCA), the separation time between the NE and EEMH-ME lineages and the mutation rate were well estimated, as shown by the high value of their R^2 , respectively 73%, 62%, and 90% (posterior distributions in Fig. 4). The likely age of the Most Recent Common Ancestor (TMRCA) is around 411,000 years (median value), in agreement with previous estimates (Kriings et al., 1999; Ovchinnikov et al., 2000; Briggs et al., 2009), but lower than the 660,000 years ago inferred from the survey of the entire mitochondrial genome (Green et al., 2008). The separation time between populations is estimated at about 295,000 years ago, close to the value inferred by Noonan et al. (2006) from 65,000 nuclear bp, and within the range estimated considering the whole genome (270,000–440,000) (Green et al., 2010). The mutation rate (0.0008 per generation for the 360-bp region, hence about 0.1 mutational events per million year per nucleotide) is lower than recently estimated (Henn et al., 2009; Soares et al., 2009) but appears reliable, considering that across our long evolutionary times multiple mutational events on the same site may occur, reducing the apparent mutation rate.

Our median estimate for the time of dispersal from Africa is about 69,000 years ago, and its 90% credible interval is between 52,650 and 80,000 years ago. These values suggest that the dispersal might have occurred earlier than inferred from studies of modern DNA diversity, i.e. 51,000 (Fagundes et al., 2007) or 56,000 (Liu et al., 2006) years ago, although the R^2 value is admittedly low (7%). We could not substantially improve the prior estimates of other parameters, including the effective sizes of the ME and EEMH populations, both associated with broad posterior distributions and low R^2 . Changing the prior distribution from uniform to log-uniform had basically no effect upon the estimate and on its accuracy (Supporting Information Table 3).

Even though Models 5 and 6 appeared to be highly unlikely, we also estimated their parameters and found they have largely overlapping distributions. The gene flow rate from Neandertals to EEMH, under Model 5, has a 90% upper bound of the posterior density at 0.0057, with a mode = 0.0008 and a median = 0.001 (Supporting Information Table 4a). This means on average 3 events of migration per generation, considering the NE effective female population size estimated during the period of coexistence with the EEMH. Simulating migration right after early modern human dispersal

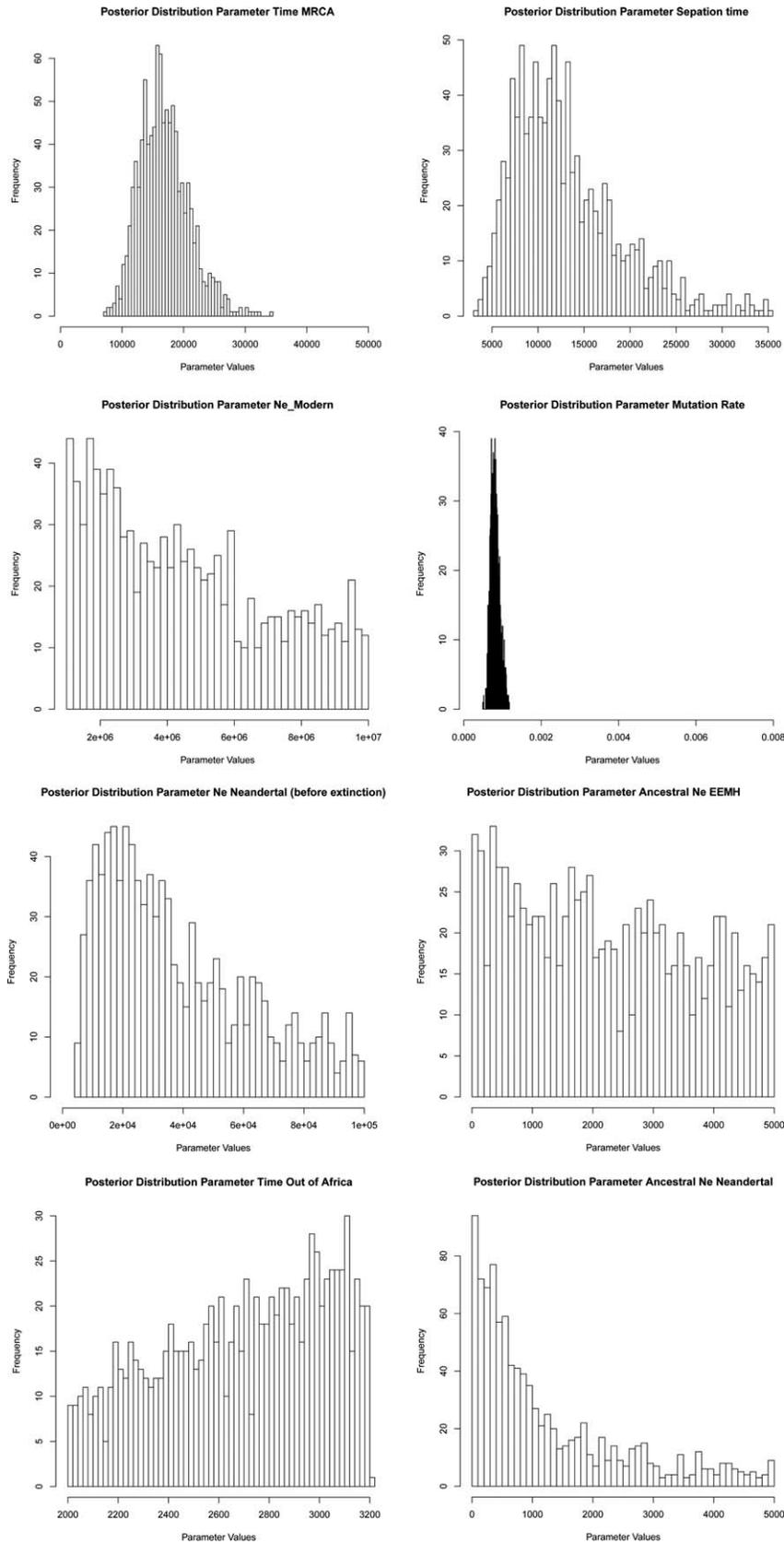


Fig. 4. Posterior distributions of parameters under Model 4, based on the best 1,000 simulations. The X axis covers the range of the (uniform) prior distributions.

TABLE 3. Power of the AR procedure to recover the true model, estimated as the proportion of cases in which data generated by the models listed on the Y-axis were attributed to the models listed on the X-axis

		% of model attribution				Type I error
		MOD1	MOD2	MOD3	MOD4	
SIMULATED MODEL	MOD1	0.992	0.005	0.003	0	0.008
	MOD2	0.031	0.947	0.009	0.012	0.052
	MOD3	0.015	0.022	0.962	0.001	0.038
	MOD4	0.002	0.031	0.002	0.963	0.035

Type I error, in the last column, is the fraction of cases in which the true model was not identified.

from Africa as suggested by Green et al. (2010) (Model 6, which proved at least 20 times less likely than a model without gene flow) the upper bound of the 90% posterior density distribution reaches 0.015, but the modal estimate is 0 (Supporting Information Table 4b). In other words, gene flow between NE and EEMH appears extremely unlikely (see Fig. 3), and if it was not zero it was nearly so. This parameter, with its narrow posterior distribution and good R^2 (45%, Supporting Information Tables 4a and 4b), seems well estimated.

Posterior predictive test and quality assessment of the estimates

Additional tests support the reliability of our estimates. For Model 4, the P -values representing the discrepancy between the observed data and the datasets generated drawing parameter values from the estimated posterior distributions were insignificant for all statistics (Supporting Information Table 5), and the global P -value was 0.478. Therefore, this model, besides having a probability close to 100% with respect to alternative models, is also capable to generate patterns of diversity fully consistent with all observed statistics.

We then asked whether the method used for the model selection (AR) is powerful enough to identify the model under which the data were generated (Table 3). Comparing Models 1 to 4, all the datasets were correctly identified, with probabilities of recovery from 94% to 99% and hence a Type I Error never $>5.2\%$. In other words, had mtDNA diversity evolved according to one of the other simulated models, the present analysis would have shown it. As could be expected, slightly less power was shown in the comparison between Model 5 and Model 6 (Type I Error: 38%, Supporting Information Table 6). This is due to the fact that these models are very similar, differing just for the time of the gene flow events from Neandertal into EEMH, and hence generate similar patterns in the data.

Finally, we ran several tests to assess the quality of the parameters estimated under Model 4 (Supporting Information Table 7). Both relative Bias and RMSE for the modes and medians of each parameter are generally low. Only the modes of the two ancestral N_e have high values of both Bias and RMSE, suggesting, as said before, that these parameters could not be well estimated. Most of the parameters show high values of the 95% coverage, indicating that their posterior distributions are in general well estimated (Supporting Information Table 6).

DISCUSSION

Nordborg (1998) first remarked that the nonoverlap between Neandertal and modern mtDNA variation does not imply that there was no admixture, because, at the low Palaeolithic population sizes, drift could have eliminated rare, and even not-so-rare, haplotypes. The ques-

tion, then, became how rare a haplotype should be, and how small the population, to produce the observed absence of Neandertal haplotypes in modern subjects, despite admixture having actually occurred.

Currat and Excoffier (2004) demonstrated by simulation that the absence of Neandertal mtDNAs in the modern gene pool is compatible with a maximum interbreeding rate = 0.1%, which translates into 1 admixture event every 100 years, during the coexistence of the two human forms in Europe. Belle et al. (2009) incorporated EEMH sequences in their analyses, but still failed to find evidence for any appreciable degree of Neandertal admixture in the European mtDNA pool. For methodological reasons, in both studies mutation rates and population sizes had to be fixed at the start of the simulation. Conversely, the ABC methods we employed in this study allowed us to explore for each model a broad and continuous range of population sizes, mutation rates and, when applicable, separation times and gene flow rates. In this way, the models were compared in a statistically rigorous way, and their final performance is independent of any specific value of the simulation parameters. We found that the best estimate by far of mitochondrial admixture between Neandertals and the ancestors of modern Europeans is zero. Even at very low population sizes and with high mutation rates, the patterns of diversity observed in ancient and modern samples appear incompatible with a Neandertal contribution to the mitochondrial genealogy of EEMH and modern Europeans.

There is reason to believe that the estimates we obtained can be trusted. The shapes of the posterior probability distributions, the posterior predictive tests, and several statistics estimated from the simulated data strongly suggest that the information available was sufficient to discriminate among models, and that most parameters are well estimated. The main area of uncertainty concerns the modern population sizes, which appear extremely large and distributed across the whole range of the prior distribution. This finding is not unusual in studies of this kind (Fagundes et al., 2007; Belle et al., 2009; Wegmann et al., 2009; Laval et al., 2010), and does not seem to reflect the choice of priors. Rather, it is probably a consequence of a simplistic, yet unavoidable, assumption, namely that populations evolved in isolation. In reality, people with different mitochondrial features must have migrated for millennia from other regions. This process resulted in an increase of genetic diversity, which the model accommodated by inflating the population size estimates. However, it is hard to imagine that the Neandertal contribution to the modern gene pool would be more likely in a (much more complicated) model also considering successive gene flow from multiple modern sources. On the other hand, such a complicated model would require the estimation of a very large number of parameters (i.e., migration rates between all possible pairs of populations), resulting in a

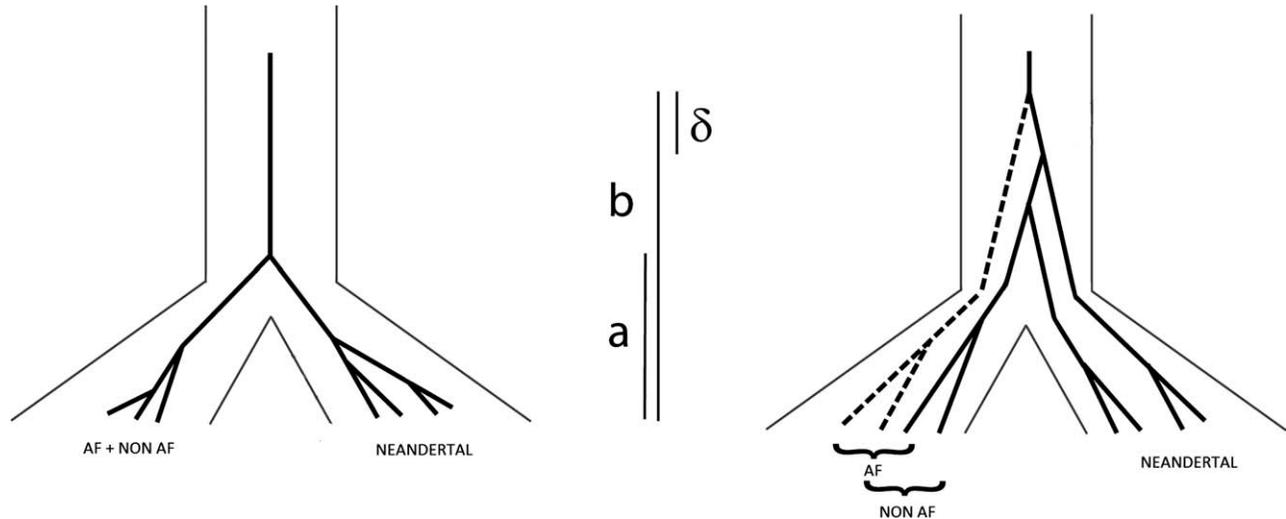


Fig. 5. Schematic view of the gene genealogies for markers transmitted by one (left) and two (right) parents, in Neandertals and modern humans. a: time since the mitochondrial most recent common ancestors; b: time since the autosomal most recent common ancestor; δ : excess evolutionary time shared with Neandertals only by the lineage leading to non-African modern people; AF, NON-AF: African and Non-African modern populations.

loss of accuracy in the estimation of the parameters that really matter, i.e., admixture rates between Neandertals and anatomically modern people.

In the recent genome survey, Neandertals appeared genetically closer to non-Africans than to Africans. This observation was interpreted as evidence of admixture between Neandertals and the common ancestors of Asians and Europeans, in the Levant, resulting in a Neandertal contribution to the modern genomes estimated between 1% and 4%. Alternative explanations are possible, but were considered less likely (Green et al., 2010). However, the poor performance of Model 6 in this study shows that the hypothesis of early admixture in the Levant has some problems too. Unless we have made serious errors in the interpretation of mitochondrial data, the model favored by the analysis of nuclear diversity seems to account very poorly, if at all, for the observed patterns of mitochondrial diversity in archaic and contemporary populations of Europe.

Only one complete Neandertal genome has been studied so far, and, given the rigid standards established to guarantee the quality of the data, sample size is not going to increase any time soon. A second problem is that the admixture model between Neandertal and anatomically modern populations proposed by Green et al. (2010) implies that the ancestors of all modern humans who left Africa had contacts with Neandertals, including those from Papua New Guinea. On the contrary, it is possible that ancestral modern humans also dispersed from Africa via a Southern route, through the Arab peninsula, the Indian subcontinent and Melanesia. This hypothesis was proposed to account for temporal and spatial patterns of cranial diversity (Lahr and Foley, 1994), has been supported by analyses of mtDNA variation (Quintana-Murci et al., 1999; Maca-Meyer et al., 2001; Macaulay et al., 2005) and, recently, by the analysis of >100,000 nuclear single-nucleotide polymorphisms (Ghirotto et al., 2011). If some modern populations of Southern Asia and Papua New Guinea are descended from people who left Africa without crossing Palestine, we see no way that their ancestors could have met, and hybridized with, Neandertals. Therefore, their genetic affinities with Neandertals must have a different origin.

It is thus necessary to find another explanation for the discrepancy between the apparent implications of the mitochondrial and nuclear analyses. In principle, two possibilities, neither simple to support empirically, would be sex-biased gene flow and hybrid selection. The former means that maybe Neandertal males, but not females, admixed with early anatomically modern Europeans. This is in contrast with studies of sex-biased admixture in modern communities, suggesting that the invading population tends to incorporate females more than males (Abe-Sandes et al., 2004; Goncalves et al., 2008; Gonzalez-Andrade et al., 2007; Stefflova et al., 2009; Quintana-Murci et al., 2010); to what extent this might also apply to prehistoric populations, nobody knows. Hybrid selection could account for the observed differences between admixture estimates if Neandertal mtDNAs had lower fitness in combination with a hybrid nuclear genome. Once again, we see no way to test empirically whether that was actually the case.

Moving on to testable hypotheses, a simple process of genetic drift after admixture is not the explanation we seek (see Fig. 3). In addition, in a simple admixture model, alleles passed from a resident to an invading population are expected to often surf to high frequencies if the invading populations also undergoes demographic growth (Currat et al., 2008). Because the incoming EEMH doubtless increased in numbers, even small Neandertal contributions should be detectable in the gene pool of their descendants, which is not the case for the European mtDNAs (Currat and Excoffier, 2004; this study).

To reconcile findings based on nuclear and mitochondrial variation we thus need a more articulate model, of which genetic drift is only a component. Many studies of modern DNA data have suggested that the common ancestors of Neandertals and modern humans might have been geographically structured (Falush et al., 2003; Harding and McVean, 2004; Lahr and Foley, 1994). A few simple calculations show that this possibility, also mentioned by Green et al. (2010), should be taken seriously (see Fig. 5). The expected time since the MRCA is $2N$ generations for mtDNA, where N is the female population size; if the sex ratio among Neandertals was 1 female : 1 male, the age of the nuclear DNA MRCA

should be 4 times as large. Briggs et al. (2009) quantified the size of the Neandertal female population around 3,500 or less. This means that the mitochondrial and nuclear MRCAs of Neandertals can be placed respectively 7,000 generations (or 175,000 years) and 28,000 generations (or 700,000 years) ago. These figures come with a large standard error, but imply that if the lineages leading to Neandertals and modern humans separated between 175,000 and 700,000 years ago, one would expect exactly what has been observed, namely independent mtDNA genealogies, and a certain degree of allele sharing at the autosomal level (see Fig. 5). On the basis of cranial measurements, anatomically archaic and modern humans separated between 311,000 and 435,000 years ago, with an upper limit of 592,000 (Weaver et al. 2008, and references therein). In this paper, we estimated that the same event occurred about 295,000 years ago (median value), with an upper 95% limit of 646,200 years. Therefore, the Replacement model with structured ancestral population is in reasonable agreement with fossil, nuclear DNA and mtDNA evidence, whereas the model of admixture fails to account for the observed relationships between ancient and modern mtDNAs.

Under a model in which the ancestral population was structured, the greater nuclear similarity between Neandertals and non-Africans would not necessarily require admixture between them. Indeed, if the non-Africans shared with Neandertals a longer section of their genealogy (represented by the interval labeled as δ in Fig. 5), they would also share more alleles than Africans and Neandertals, including the derived alleles upon which Green et al. (2010) based their estimates. This view is also supported by data on the DNA of the human gastric parasite *Helicobacter pylori*, in which ancestral genetic clusters seem to have given rise to two distinct populations, one exclusively African and the other cosmopolitan (Falush et al., 2003), and by the extreme levels of DNA variation still present in Africa (Schuster et al., 2010; Henn et al., 2011). The only additional assumption one has to make to account for the observed results is that the latter population was also ancestral to the European Neandertals typed by Green et al. (2010). Therefore, the hypothesis of genetic drift in a structured ancestral population, in which Neandertals shared a longer period of common ancestry with the non-African's than with the African's ancestors, seems to reconcile most findings about DNA diversity in Neandertal and modern people. This hypothesis predicts that the nuclear alleles preferentially shared by Neandertals and non-African will have MRCAs falling in the upper part of the genealogy (δ interval in Fig. 5), and we are preparing to test this hypothesis.

ACKNOWLEDGMENTS

The authors thank Giorgio Bertorelle and Arnaud Estoup for their useful suggestions.

LITERATURE CITED

- Abe-Sandes K, Silva WA Jr, Zago MA. 2004. Heterogeneity of the Y chromosome in Afro-Brazilian populations. *Hum Biol* 76:77–86.
- Anderson CN, Ramakrishnan U, Chan YL, Hadly EA. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21:1733–1734.
- Babalini C, Martinez-Labarga C, Tolk HV, Kivisild T, Giampaolo R, Tarsi T, Contini I, Barac L, Janicijevic B, Martinovic Klaric I, et al. 2005. The population history of the Croatian linguistic minority of Molise (southern Italy): a maternal view. *Eur J Hum Genet* 13:902–912.
- Beaumont M. 2008. Joint determination of topology, divergence time and immigration in population trees. Simulations, genetics and human prehistory. Cambridge: McDonald Institute for Archaeological Research. p 135–154.
- Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.
- Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res* 17:1505–1519.
- Belle EM, Benazzo A, Ghirotto S, Colonna V, Barbujani G. 2009. Comparing models on the genealogical relationships among Neandertal, Cro-Magnoid and modern Europeans by serial coalescent simulations. *Heredity* 102:218–225.
- Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajkovic D, Kucan Z, et al. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* 325:318–321.
- Caramelli D, Lalueza-Fox C, Condemi S, Longo L, Milani L, Manfredini A, de Saint Pierre M, Adoni F, Lari M, Giunti P, et al. 2006. A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Curr Biol* 16:R630–R632.
- Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, et al. 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc Natl Acad Sci USA* 100:6593–6597.
- Caramelli D, Milani L, Vai S, Modi A, Pecchioli E, Girardi M, Pilli E, Lari M, Lippi B, Ronchitelli A, et al. 2008. A 28,000 years old Cro-Magnon mtDNA sequence differs from all potentially contaminating modern sequences. *PloS one* 3:e2700.
- Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC. 1996. Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 60:331–350.
- Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol* 2:e421.
- Currat M, Excoffier L, Maddison W, Otto SP, Ray N, Whitlock MC, Yeaman S. 2006. Comment on “Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*” and “Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans”. *Science* 313:172; author reply 172.
- Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* 62:1908–1920.
- Excoffier L. 2002. Human demographic history: refining the recent African origin model. *Curr Opin Genet Dev* 12:675–682.
- Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50.
- Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA* 104:17614–17619.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, et al. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299:1582–1585.
- Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am J Phys Anthropol* 128:415–423.
- Gelman A, Carlin J, Stern H, Rubin D. 2004. Bayesian data analysis. Boca Raton, Florida: CRC Press.
- Ghirotto S, Mona S, Benazzo A, Pappalardo F, Caramelli D, Barbujani G. 2010. Inferring genealogical processes from patterns

- of Bronze-Age and modern DNA variation in Sardinia. *Mol Biol Evol* 27:875–886.
- Ghirotto S, Penso-Dolfin L, Barbujani G. 2011. Genomic evidence for an African expansion of anatomically-modern humans by a southern route. *Hum Biol* 00:00–00.
- Goncalves VF, Carvalho CM, Bortolini MC, Bydlowski SP, Pena SD. 2008. The phylogeography of African Brazilians. *Hum Hered* 65:23–32.
- Gonzalez-Andrade F, Sanchez D, Gonzalez-Solorzano J, Gascon S, Martinez-Jarreta B. 2007. Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador. *Hum Biol* 79:51–77.
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.
- Green RE, Malaspina AS, Krause J, Briggs AW, Johnson PL, Uhler C, Meyer M, Good JM, Maricic T, Stenzel U, et al. 2008. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* 134:416–426.
- Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilineal populations than in matrilineal populations. *Proc Natl Acad Sci USA* 102:7476–7480.
- Harding RM, McVean G. 2004. A structured ancestral population for the evolution of modern humans. *Curr Opin Genet Dev* 14:667–674.
- Henn BM, Gignoux CR, Feldman MW, Mountain JL. 2009. Characterizing the time dependency of human mitochondrial DNA mutation rate estimates. *Mol Biol Evol* 26:217–230.
- Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodríguez-Botigúé L, Ramachandran S, Hon L, Brisbin A, et al. 2011. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci USA* 108:5154–5162.
- Hodgson JA, Disotell TR. 2008. No evidence of a Neandertal contribution to modern human diversity. *Genome Biol* 9:206.
- Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–589.
- Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, et al. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451:998–1003.
- Joyce P, Marjoram P. 2008. Approximately sufficient statistics and Bayesian computation. *Stat Appl Genet Mol Biol* 7:Article26.
- Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M, Paabo S. 2000. A view of Neandertal genetic diversity. *Nat Genet* 26:144–146.
- Krings M, Geisert H, Schmitz RW, Krainitzki H, Paabo S. 1999. DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc Natl Acad Sci USA* 96:5581–5585.
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19–30.
- Labuda D, Zietkiewicz E, Yotova V. 2000. Archaic lineages in the history of modern humans. *Genetics* 156:799–808.
- Lahr MM, Foley RA. 1994. Multiple dispersals and modern human origins. *Evol Anthropol* 3:48–60.
- Lalueza-Fox C, Krause J, Caramelli D, Catalano G, Milani L, Sampietro ML, Calafell F, Martinez-Maza C, Bastir M, Garcia-Taberner A, et al. 2006. Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals. *Curr Biol* 16:R629–R630.
- Laval G, Patin E, Barreiro LB, Quintana-Murci L. 2010. Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS one* 5:e10284.
- Leman SC, Chen Y, Stajich JE, Noor MA, Uyenoyama MK. 2005. Likelihoods from summary statistics: recent divergence between species. *Genetics* 171:1419–1436.
- Liu H, Prugnolle F, Manica A, Balloux F. 2006. A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 79:230–237.
- Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2:13.
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, et al. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–1036.
- Mellars P. 2006. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc Natl Acad Sci USA* 103:9381–9386.
- Mellars PA. 1992. Archaeology and the population-dispersal hypothesis of modern human origins in Europe. *Philos Trans R Soc Lond B Biol Sci* 337:225–234.
- Nasidze I, Stoneking M. 2001. Mitochondrial DNA variation and language replacements in the Caucasus. *Proc Biol Sci* 268:1197–1206.
- Neuenschwander S, Lurgiader CR, Ray N, Currat M, Vonlanthen P, Excoffier L. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol Ecol* 17:757–772.
- Noonan JP, Coop G, Kudaravalli S, Smith D, Krause J, Alessi J, Chen F, Platt D, Paabo S, Pritchard JK, et al. 2006. Sequencing and analysis of Neandertal genomic DNA. *Science* 314:1113–1118.
- Nordborg M. 1998. On the probability of Neandertal ancestry. *Am J Hum Genet* 63:1237–1240.
- Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W. 2000. Molecular analysis of Neandertal DNA from the northern Caucasus. *Nature* 404:490–493.
- Plagnol V, Wall JD. 2006. Possible ancestral structure in human populations. *PLoS Genet* 2:e105.
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol* 16:1791–1798.
- Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat Genet* 23:437–441.
- Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG, Behar DM. 2010. Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. *Am J Hum Genet* 86:611–620.
- R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. Available at: <http://www.R-project.org>. Vienna, Austria: Foundation for Statistical Computing.
- Relethford JH. 2001. Absence of regional affinities of Neandertal DNA with living humans does not reject multiregional evolution. *Am J Phys Anthropol* 115:95–98.
- Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59:185–203.
- Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, Krainitzki H, Paabo S, Smith FH. 2002. The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc Natl Acad Sci USA* 99:13342–13347.
- Schuster SC, Miller W, Ratan A, Tomsho LP, Giardine B, Kasson LR, Harris RS, Petersen DC, Zhao F, Qi J, et al. 2010. Complete Khoisan and Bantu genomes from southern Africa. *Nature* 463:943–947.
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84:740–759.
- Stefflova K, Dulik MC, Pai AA, Walker AH, Zeigler-Johnson CM, Gueye SM, Schurr TG, Rebbeck TR. 2009. Evaluation of group genetic ancestry of populations from Philadelphia and Dakar in the context of sex-biased admixture in the Americas. *PLoS one* 4:e7842.

- Tattersall I, Schwartz JH. 1999. Hominids and hybrids: the place of Neanderthals in human evolution. *Proc Natl Acad Sci USA* 96:7117–7119.
- Templeton AR. 2005. Haplotype trees and modern human origins. *Am J Phys Anthropol Suppl* 41:33–59.
- Templeton AR. 2007. Genetics and recent human evolution. *Evolution* 61:1507–1519.
- Trinkaus E. 2007. European early modern humans and the fate of the Neandertals. *Proc Natl Acad Sci USA* 104:7367–7372.
- Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. 2005. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci USA* 102:18508–18513.
- Wakeley J, Hey J. 1997. Estimating ancestral population parameters. *Genetics* 145:847–855.
- Walker MJ, Gibert J, Lòpez MV, Vincent Lombardi A, Pérez A, Zapata J, Ortega J, Higham T, Pike A, Schwenninger JL, et al. 2008. Late Neandertals in Southeastern Iberia: Sima de las Palomas del Cabezo Gordo, Murcia, Spain. *Proc Natl Acad Sci USA* 105:20631–20636.
- Wegmann D, Leuenberger C, Excoffier L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182:1207–1218.
- Zilhão J. 2006. Neandertals and moderns mixed, and it matters. *Evol Anthropol* 15:183–195.
- Zilhão J, Davis SJM, Duarte C, Soares AMM, Steier P, Wild E. 2010. Pego do Diabo (Loures, Portugal): dating the emergence of anatomical modernity in westernmost Eurasia. *PLoS ONE* 5:e8880.