Am. J. Hum. Genet. 63:1237–1240, 1998

On the Probability of Neanderthal Ancestry

To the Editor:

The controversial relationship between Neanderthals and modern humans recently received much attention, owing to the recovery of a Neanderthal mtDNA fragment, the analysis of which indicated that the most-recent common ancestor (MRCA) of Neanderthal and modern-human mitochondria was several times more ancient than that of modern humans only (Krings et al. 1997; fig. 1). This finding was considered to be strong evidence that Neanderthals and anatomically modern humans are separate species, the latter having replaced the former without interbreeding ("In our genes?" 1997; Kahn and Gibbons 1997; Lindahl 1997; Wade 1997; Ward and Stringer 1997). Here, I investigate the strength of this evidence by considering the probability of erroneous rejection of interbreeding (i.e., the probability of a type I error). I demonstrate that, although completely random mating clearly can be rejected, more-relevant models of interbreeding cannot.

The question of whether Neanderthals and anatomically modern humans interbred is a question of ancient levels of gene flow. Thus, although the relevant features of the data can be conveniently summarized as in figure 1, this figure is not, a priori, a phylogenetic tree for Neanderthals and humans: indeed, the question is whether such a tree exists. Figure 1 is simply a genealogical tree representing the history of the sampled mtDNA. In the following discussion, I ignore the considerable uncertainty in the estimation of this history and focus on the question of whether, given perfect knowledge of mtDNA genealogy, we would be able to conclude that anatomically modern humans and Neanderthals did not interbreed.

First, I consider whether Neanderthals and anatomically modern humans could have mated randomly. Two features of the data summarized in figure 1 provide evidence against such a scenario: The first is the topology, with the modern sample being monophyletic, and the second is the more than fourfold difference between T_r, the age of the MRCA of the modern humans and the Neaderthal, and T_e, the age of the MRCA of the modern humans only. If anatomically modern humans and Neanderthals mated randomly, the probability of such a result can be calculated as follows. Let $A_n(t) \in \{1,...,n\}$ be the random number of ancestors, at time t, of a sample of n mtDNAs at $t = 0$; its distribution is known under a variety of neutral models (Tavaré 1984). Conditional on $A_{986}(t_s) = k$, the number of ancestors of the modern sample who are contemporary with the sampled Neanderthal, the probability sought can be written as the product of the probability that a compatible topology is observed and the probability that sufficiently extreme coalescence times are observed. The former probability is easily shown to be $P[\text{topology} \mid A_{986}(t_s) = k] = 2/[k(1 + k)]$ (this also may be obtained as a special case of more-general results [Watterson 1982; Saunders et al. 1984]). An exact expression for the latter probability also can be obtained (T. Nagylaki and M. Nordborg, unpublished data) but is cumbersome and in some cases difficult to evaluate numerically. Estimation of the probability through standard Monte Carlo–simulation techniques is more convenient (e.g., Marjoram and Donnelly 1997).

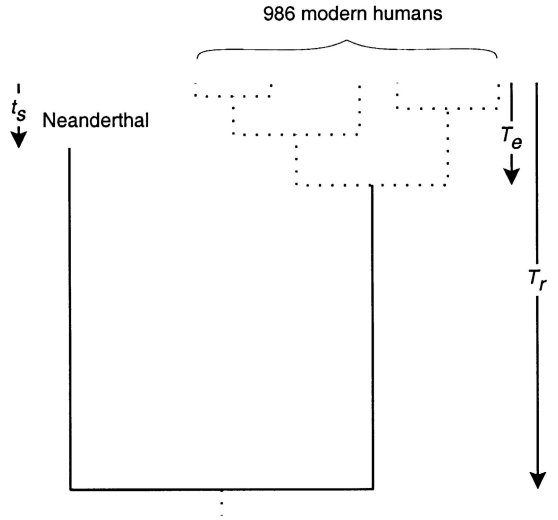

Figure 1 Schematic genealogy of the 986 modern-human mtDNAs and a single Neanderthal mtDNA (the carrier of which lived at time t_s before the present). The MRCA of the entire sample was inferred to be at least four times more ancient than the MRCA of the modern sample—that is, $T_r \geqslant 4T_e$ (Krings et al. 1997).

Two simple scenarios for human demography were used—namely, constant population size and constant ancient-population size followed by exponential growth 50,000 years ago. For both cases, the effective number of females in the constant population was assumed to be 3,400, growing exponentially to 5×10^8 for the latter case. These parameters were chosen so that the probability would be high that T_e lies within the range 100,000–200,000 years, when a generation time of 20 years is assumed. The age of the sampled Neanderthal, t_s, was assumed to be 30,000–100,000 years (the recovery of DNA more ancient than 100,000 years seems highly doubtful [Krings et al. 1997]). I argue below that the absolute values of all these parameters are of considerably lesser importance than their relative values.

Table 1 gives the results for models of random mating. As expected, the probability that both a compatible topology and an extreme difference between T_e and T_r would be observed is low, and, therefore, the hypothesis

Table 1

Results for Models of Random Mating

	Constant Population Size and t_s (in Years) =		Recent Population Growth and t_s (in years) =	
Parameter	30,000	100,000	30,000	100,000
$E[A_{986}(t_s)]$	4.86	1.75	782	2.86
P(topology)	.085	.56	3.3×10^{-6}	.24
P(topology and $T_r \geqslant 4T_e$)	.0063	.035	3.7×10^{-8}	.002

Note.—$E[A_{986}(t_s)]$ is the expected number of ancestors of the modern sample who are contemporary with the sampled Neanderthal. P(topology) is the probability that the topology in figure 1 would be observed, and P(topology and $T_r \geqslant 4T_e$) is the probability that both unlikely features of the data would be observed. All values were estimated through Monte Carlo simulation, as well as by calculation from the analytical results, except for those in the third column, for which the latter approach proved to be computationally too difficult. The 95% confidence intervals for the simulated values do not alter the decimals given. In the constant–population-size model, the expected T_e was ~136,000 years, with an SD of ~70,000 years; for recent exponential growth, the expected T_e was ~180,000 years, with, again, an SD of ~70,000 years.

that modern humans and Neanderthals were a randomly mating population may be rejected. However, closer inspection reveals the more interesting fact that the topology alone may not be unlikely. The reason for this is that, unless the sampled Neanderthal lived long after human populations had started to grow exponentially, most of the modern mtDNA lineages would have coalesced at t_s: if, for example, the modern sample only had two ancestors who were contemporary with the sampled Neanderthal, it would not be surprising if they were monophyletic (probability of 1/3). A large difference between T_e and T_r, on the other hand, is always unlikely under random mating.

Thus, the data constitute considerable evidence against the hypothesis that all sequences were drawn from a single population. This perhaps should not be surprising: the recovered Neanderthal sequence clearly was not sampled from a random individual at time t_s but was sampled specifically from an individual who was morphologically distinct from anatomically modern humans. Furthermore, fossil data strongly suggest that Neanderthals and anatomically modern humans were not a randomly mating population. To ask questions about interbreeding, more-interesting null hypotheses are needed. One pleasingly simple scenario is the following. Assume that Neanderthals were an isolated population for a long time, until they encountered anatomically modern humans at time t_m and *merged* with them to form a single, randomly mating population, with a fraction, c, of the population being Neanderthal. Then, the so-called replacement hypothesis is simply that $c = 0$. The data in figure 1 are perfectly consistent with this scenario; that is, the probability of the data is 1, without interbreeding. However, this provides support for replacement only to the extent that alternative scenarios can be shown to have a much lower probability. Therefore, the probability of the data must be found for different values of $c > 0$.

Under the assumption that the sampled Neanderthal lived before t_m (i.e., a "pure" Neanderthal), the probability sought is simply the probability that none of the ancestors at time t_m came from the Neanderthal fraction of the population. This probability can be written as $\sum_{k=1}^{986}(1-c)^k P[A_{986}(t_m) = k]$, which is the probability-generating function for $A_{986}(t_m)$. Figure 2 shows a plot for the two demographic scenarios described above, with t_m = 30,000 or 100,000 years. Clearly, for the scenarios in which the expected number of ancestors at t_m is low (table 1), the data tell us little about interbreeding, except perhaps that the Neanderthals did not make up the majority. The situation is completely different if the expected number of ancestors at t_m is high. In this case, all but very small values of c may be rejected.

In cases for which we expect few ancestors at t_m, the probability that none of the 986 *sampled* mtDNAs came from the Neanderthal fraction of the population does not differ much from the probability that none of the *currently existing* mtDNAs did so. This latter probability is equal to the well-known probability that an allele starting at frequency c is lost, through drift, by time t_m (Kimura 1955). Under this assumption, another question of interest can be addressed: Given that extant humans do not carry Neanderthal mtDNA, what does this sug-

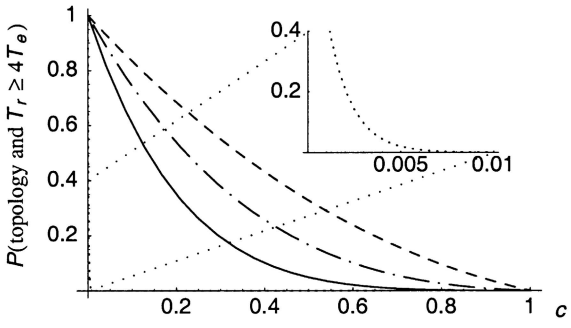

Figure 2 Probability of the data, if Neanderthals and anatomically modern humans merged at time t_m, with Neanderthals composing a fraction, c, of the new population. The four curves are for different demographic assumptions (see text) and values of t_m: constant population size, t_m = 30,000 years (solid line); constant population size, t_m = 100,000 years (dashed line); recent exponential growth, t_m = 30,000 years (dotted line [magnified in insert]); and recent exponential growth, t_m = 100,000 years (dotted-dashed line). The plots were calculated numerically by use of the known probability-generating function (Tavaré 1984), except for the third scenario, for which Monte Carlo simulation was used because of computational difficulties.

gest about the rest of the genome? For the constant–population-size model, for example, assume that Neanderthals and anatomically modern humans merged 1 coalescent-time unit ago (equivalent to $t_m = 68{,}000$ years, for the population size used above) and that Neanderthals composed 25% of the new population. Then, the probability that all Neanderthal mtDNA was lost through drift is .52 (the probability that Neanderthal mtDNA was not in the sample [calculated as above] is the same, to two decimal places). At the same time, each nuclear locus, for which the coalescence-time scale is four times slower, would have lost all Neanderthal alleles with probability .10 and would have become fixed for them with probability 9.8×10^{-5}. Thus, 90% would still be segregating for Neanderthal alleles.

In conclusion, data such as those shown in figure 1 shed little light on the issue of replacement versus interbreeding, unless the number of ancestors of the sample was large throughout the periods of interest. This is part of a general problem: in order to estimate gene flow, a large sample is needed, and, in order to estimate ancient-gene flow, a large ancient sample is needed. According to coalescent theory, large ancient samples usually cannot be obtained by the sampling of modern populations. The rate of coalescence is quadratic in the number of ancestors and linear in the inverse of the population size. Thus, the expected number of ancestors of a sample usually decreases rapidly as earlier time periods are studied. Exceptions include exponentially growing populations, in which the number of ancestors may be large shortly after the onset of growth (reviewed in Donnelly and Tavaré 1995; Marjoram and Donnelly 1997). In the present case, it seems clear that the statistical power to detect interbreeding that took place before the human population started to grow exponentially is close to zero.

I also have considered the mtDNA genealogy as known. The extreme uncertainty of the reconstruction of ancient DNA and the genealogy shown in figure 1 presumably suggests that conclusions from the data should be made with even more caution. Additional Neanderthal mtDNA sequence data would reduce these sources of uncertainty, but the main problem discussed above can be alleviated only by the study of data from several unlinked loci. The fact remains that an inference about population properties that is based on a single locus (or a nonrecombining genome) is an inference from a single data point. This does not mean that single loci contain no information: I have shown that random mating can be rejected, and the existence of a single Neanderthal mtDNA that differed little from modern mtDNA would allow rejection of the hypothesis that there was no interbreeding. Such an observation probably could never be made, however, since contamination would be impossible to rule out.

Finally, the above analysis depends on the selective neutrality of mtDNA variation. It is well known that human mtDNA variation suggests a genealogy that is "star shaped": this has been interpreted as the result of a historical population expansion (Di Rienzo and Wilson 1991; Merriwether et al. 1991; Vigilant et al. 1991; Rogers and Harpending 1992). However, data from several nuclear loci do not show this pattern (Harding et al. 1997; Hey 1997). Together, these observations may constitute evidence against neutrality, with a plausible alternative being a recent selective sweep in human mtDNA (Hey 1997). The conclusions in this paper clearly are not robust to this type of violation of assumptions: if there has been a recent selective sweep in human mtDNA, even random mating cannot be rejected.

Acknowledgments

I thank B. Bengtsson, A. Di Rienzo, P. Donnelly, R. Harding, the reviewers, and especially T. Nagylaki, for their comments on the manuscript. This work was supported by the Erik Philip-Sörensen Foundation.

MAGNUS NORDBORG

Department of Genetics
Lund University
Lund
Sweden

References

Di Rienzo A, Wilson AC (1991) The pattern of mitochondrial DNA variation is consistent with an early expansion of the human population. Proc Natl Acad Sci USA 88:1597–1601

Donnelly P, Tavaré S (1995) Coalescents and genealogical structure under neutrality. Annu Rev Genet 29:401–421

Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, et al (1997) Archaic African *and* Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Hey J (1997) Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol Biol Evol 14: 166–172

In our genes? (1997) The Economist 344(8025), July 12th, pp 71–72

Kahn P, Gibbons A (1997) DNA from an extinct human. Science 277:176–178

Kimura M (1955) Solution of a process of random genetic drift with a continuous model. Proc Natl Acad Sci USA 41: 144–150

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neanderthal DNA sequences and the origin of modern humans. Cell 90:19–30

Lindahl T (1997) Facts and artifacts of ancient DNA. Cell 90: 1–3

Marjoram P, Donnelly P (1997) Human demography and the time since mitochondrial Eve. In: Donnelly P, Tavaré S (eds)

Progress in population genetics and human evolution. Springer-Verlag, New York, pp 107–131
Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, et al (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33: 543–555
Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569
Saunders IW, Tavaré S, Watterson GA (1984) On the genealogy of nested subsamples from a haploid population. Adv Appl Prob 16:471–491
Tavaré S (1984) Line-of-descent and genealogical processes, and their applications in population genetic models. Theor Popul Biol 26:119–164
Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507
Wade N (1997) Neanderthal DNA sheds new light on human origins. New York Times, July 11, sec A
Ward R, Stringer C (1997) A molecular handle on the Neanderthals. Nature 388:225–226
Watterson GA (1982) Mutant substitutions at linked nucleotide sites. Adv Appl Prob 14:206–224

Address for correspondence and reprints: Dr. Magnus Nordborg, Department of Genetics, Lund University, Sölvegatan 29, 223 62 Lund, Sweden. E-mail: magnus.nordborg@gen.lu.se

0002-9297/98/6304-0043$02.00

Am. J. Hum. Genet. 63:1240–1242, 1998

Do Human Chromosomal Bands 16p13 and 22q11-13 Share Ancestral Origins?

To the Editor:

Ancient duplications and rearrangements within a genome are believed to be important mechanisms of evolution. Although most duplications are of gene segments, single genes, or chromosomal segments, molecular evidence has been gathered suggesting that whole-genome duplication has facilitated evolution in yeast (Wolfe and Shields 1997). Identifying these duplicated genomic areas can be valuable not only for understanding the timing and nature of evolutionary events; additionally, this information can greatly facilitate the pinpointing of novel (disease-related) genes by positional cloning techniques.

While mapping and cloning the human gene encoding the CREB-binding protein (CBP, encoded by the *CREBBP* gene) on chromosome band 16p13.3 (Giles et al. 1997*b*), we noticed an emerging pattern concerning the genomic relationship between this chromosome band and a region of chromosome 22q. CBP exhibits extensive homology to the adenovirus E1A–associated protein p300, whose gene has been mapped to human chromosome band 22q13 (Eckner et al. 1994; Lundblad et al. 1995). At that time we noted with interest that the heme oxygenase-1 (*HMOX1*) gene, just centromeric of *CREBBP* on 16p13.3, has a paralogue mapping to chromosome band 22q12, heme oxygenase-2 (*HMOX2*; Kutty et al. 1993). Our interest was further piqued when the molecular defect in families with carbohydrate-deficient glycoprotein type I syndrome (CDG1) was determined to be caused by mutations in the phosphomannomutase 2 gene (*PMM2*) on 16p13 (Matthijs et al. 1997*a*); the same investigators had previously mapped the first phosphomannomutase gene (*PMM1*) to 22q13 (Matthijs et al. 1997*b*). Sequence comparison at the amino acid level revealed that homologies between these paralogous proteins are high: homology between CBP and p300 is 63% (Arany et al. 1995), that between PMM1 and PMM2 is 66% (Matthijs et al. 1997*a*), and that between HMOX1 and HMOX2 is 74% (authors' observation). Subsequent examination of genome databases (e.g., OMIM) resulted in six additional sets of paralogues mapping to chromosomes 16p13 and 22q11-13, although the extent of homology between these paralogue sets is not known (table 1). YAC contigs connecting outlying genes of each paralogous cluster, *CREBBP* to *MYH11* on chromosome 16 and the *CRYB* genes to *PMM1* on chromosome 22, suggest that the extent of the redundant area presented here is ~12–14 Mb. Furthermore, *CREBBP* and *MYH11* are also thought to be near the borders for the conserved synteny group in mouse chromosome 16 (Doggett et al. 1996).

We propose that the existence of these paralogous sets suggests that chromosome bands 16p13 and 22q11-13 share ancestral origins and that at some point a large-scale duplication gave rise to this second set of genes. It is well established that such duplicated regions exist (Lundin 1993; Holland et al. 1994), and a catalogue of putative paralogous regions can be found on-line (Database of Duplicated Human Chromosomal Regions). This database suggests two duplicated regions for areas of 16p: a well-documented gene cluster on chromosome band 16p11.1, which shares high homology with a locus on Xq28 (Eichler et al. 1996), and a region of 16p13, which resembles 19p13, although no specific genes are named.

A hypothesis set forth by Ohno (1993) suggests that at the stage of fish, the mammalian ancestral genome underwent tetraploid duplication. Although certain aspects of this hypothesis are not universally accepted, most scientists agree that the fourfold increase, in the number of genes, between invertebrates and vertebrates implies at least two rounds of genome duplication (Aparicio 1998). Paralogues such as the HOX-