When Did Neanderthals and Modern Humans Diverge?

PETER BEERLI AND SCOTT V. EDWARDS

Our analysis of the Neanderthal-modern human gene divergence resulted in gene divergence times in a range of 631–789 KY. This gene divergence time is informative, but the biologically more relevant parameter is the time of population divergence, which necessarily occurs after the observed gene divergence when there is any polymorphism in the ancestral population. Here, we provide an estimate and approximate confidence intervals of the population divergence time of Neanderthals and modern humans, using coalescence theory for single-locus data.

The availability of sequences from the mtDNA hypervariable region I and II of a specimen of the Neanderthal,[1,2] *Homo neanderthalensis*, permits estimation of the time of divergence of Neanderthals and modern humans. Based on substitution rates derived from comparison of human and chimpanzee, Krings et al. provided an estimate of 550–690 KY[1] and 317–741 KY[2] before present for the split between the Neanderthal and contemporary human mtDNA. These estimates are dates for the split of the genes, and its confidence intervals are based on the uncertainty of the dating of the human-chimpanzee divergence (assumed to be 4–5 MY ago). The details of the estimation of the gene divergence time are only vaguely described in the seminal presentations of these Neanderthal findings. Vital information, such as values of the substitution rate and its shape parameter of the mutation rate distribution, is missing. Our reanalysis of the Neanderthal-modern human gene divergence resulted in gene divergence times in a range of 631–789 KY. We used a mutation rate μ of 0.0634 per site per million years from the human-chimpanzee divergence time of 5 MY, and of 0.0793 for the 4 MY estimate, and a substitution rate variation shape parameter α of 0.46 (cf. Excoffier and Yang,[3] Wakeley,[4] and our own analysis). We concentrate on the HVR I because we believe that there are some difficulties with the alignment in the HVR II. The interpretation of the three thymidine residues after site 340 as an insertion as presented by Krings et al.[2] needs further investigation.

This gene divergence time is informative, but the biologically more relevant parameter is the time of population divergence, which necessarily occurs after the observed gene divergence when there is any polymorphism in the ancestral population.[5] Furthermore, the gene divergence between modern human and Neanderthal mtDNA is likely to be recent enough so that the difference in time between gene and population divergence may be substantial.

In the light of fossil evidence,[6] the estimates of Krings et al.[1,2] may seem rather old, but morphological analyses[7] suggest an even longer separation between modern human and Neanderthal lineages. Here, we provide an estimate and approximate confidence intervals of the population divergence time of Neanderthals and modern humans, using coalescence theory for single-locus data.

Our analysis (Fig. 1A) and that of Krings et al.[1] of many contemporary human mtDNA HVR I sequences and the single Neanderthal sequence strongly suggest that, going back in time, the modern human lineages coalesced with each other before coalescing with the Neanderthal sequence, and that subsequently this ancestral population coalesced with the chimpanzees. This simple sequence of events facilitates the estimation of population divergence time, making it unnecessary to use elaborate methods, such as those developed by Nielsen[8] that take the uncertainty of the gene genealogy into account, because monophyly of contemporary human mtDNAs is not in question. When descendant lineages have achieved reciprocal monophyly, however, there is no information in a single locus left to provide information on the population size of the ancestors of Neanderthals and modern humans. With only a single locus, there is no information from the variance in coalescence time among loci that can be used to estimate ancestral population

Peter Beerli is at the Department of Genome Sciences, Box 357730, University of Washington, Seattle, WA 98195.
E-mail: beerli@gs.washington.edu
Scott V. Edwards is at the Department of Zoology, Box 351800, University of Washington, Seattle, WA 98195.
Grant sponsor: National Institutes of Health; Grant numbers: GM 51929-02, GM 01989-02; Grant sponsor: National Science Foundation; Grant numbers: BIR 9527687, DEB 9707458, DEB 9815800.

Key words: Neanderthal; human; mtDNA control region; HVR; population divergence; gene divergence; Bayes method; substitution rate variation

Evolutionary Anthropology, Suppl 1:60–63 (2002)
DOI 10.1002/evan.10058
Published online in Wiley InterScience (www.interscience.wiley.com).

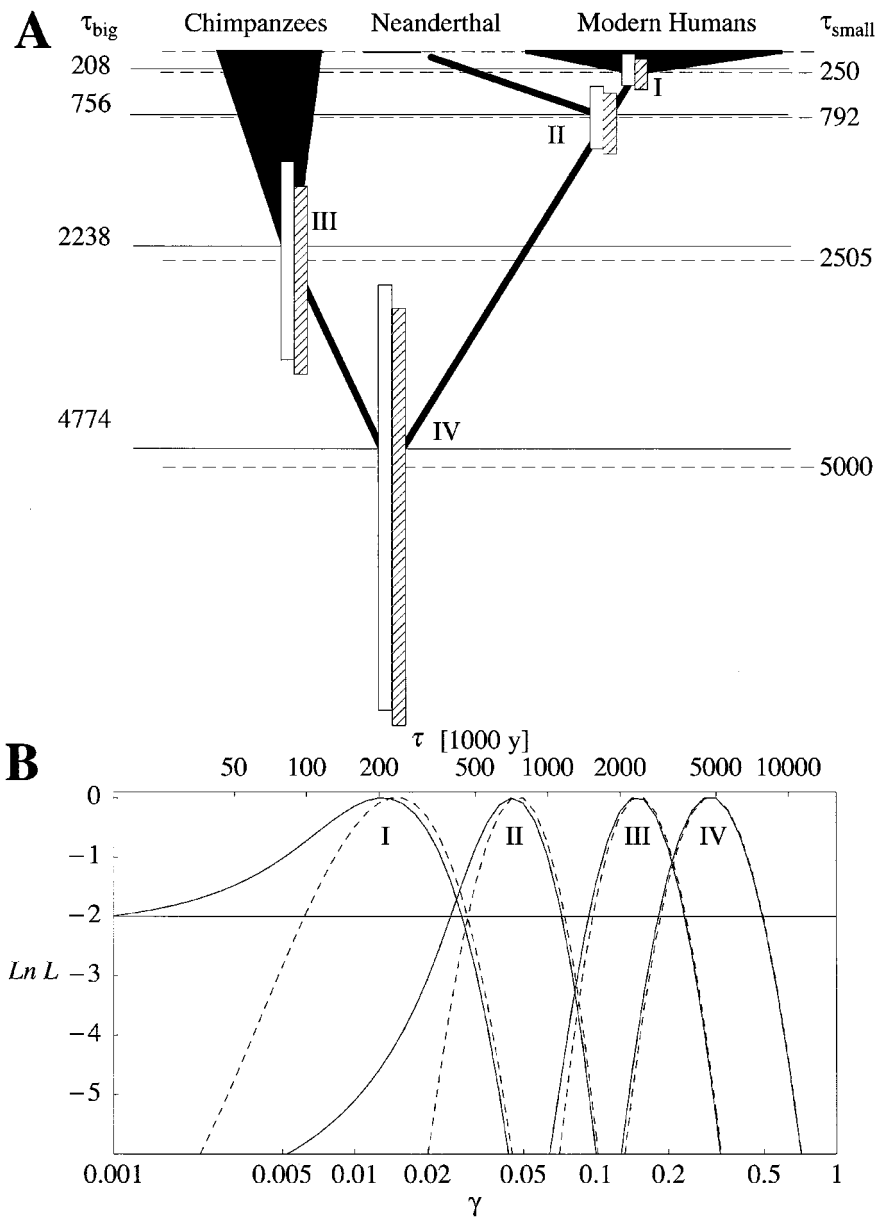

Figure 1. **A:** Genealogy of a Neanderthal mtDNA HVR I sequence and of sequences of modern humans and chimpanzees. A bootstrap of neighbor-joining trees using maximum likelihood distances and sequences of 100 contemporary humans, 1 Neanderthal, and 21 chimpanzees corroborates monophyly of chimpanzees (100% support) and nests the Neanderthals within modern humans only twice in 100 trials. Various divergence times were estimated with a Bayesian approach, using sequences from the Anderson reference, a !Kung (Genebank M76244), the Neanderthal (A01222), and two chimpanzees (*Pan troglodytes,* L35400; *Pan paniscus,* L35443). Setting the gene split of humans and chimpanzees at 5My ago (dashed curve IV in B) yields a mutation rate estimate of 0.0634. Hatched bars are approximate confidence intervals using our Bayesian estimator with a small population size; open bars are approximate confidence intervals, using a large population size (see B). Divergence times are also dependent on these population sizes. **B:** log likelihood of population divergence time. I. a deep split among modern humans; II, split between Neanderthals and modern humans; III, a deep split in chimpanzees; and IV, divergence time of humans and chimpanzees. γ is divergence time scaled by mutation rate. For population divergence time, same mutation rate as in A was used. We used a rectangular prior distribution for ancestral population size, with ranges from 6×10^{-8} to 0.001208 (dashed curve) and 6×10^{-8} to 0.01208, respectively. Using the mutation rate from A, these values translate into 1–20,000, and 1–200,000 individuals, respectively. The maximum of each solid and dashed curve marks the most likely divergence. γ values with ln L below −2 are outside the approximate 95% confidence boundaries, using the likelihood ration test criterion.

size. This situation necessitates indirect methods of inference. Our method is an extension of earlier work[9] and of methods devised by Takahata and Satta[5] and Yang,[10] who used a maximum-likelihood method based on coalescence theory to estimate jointly the ancestral population size and the divergence of two populations or species, using one sequence from each population for each of multiple loci. For single-locus data sets under reciprocal monophyly, one needs to know either the ancestral population size Θ (for mtDNA this is $N_e\mu$, where N_e is the effective population size and μ is the mutation rate per site per generation) or the population divergence time γ (= $\mu\tau$, where τ is the divergence time in generations). One cannot estimate both parameters τ and μ at once, because they are confounded. For the ancestral population of Neanderthals and modern humans, we do not know, but we can guess, a range of population sizes a priori. This approximation can be incorporated into a Bayesian approach that integrates over that range of possible values of Θ to achieve a likelihood curve for γ. Our estimator of the likelihood of γ, $L(\gamma)$ (see formula 1), is based on the coalescent and includes an arbitrary prior distribution for Θ; we use a finite-sites mutation model allowing for unequal base frequencies and rate heterogeneity among sites.[11]

$$L(\gamma) = \int_{\Theta=0}^{\infty} f(\Theta) \int_{t=0}^{\infty} \mathrm{Prob}\left(D \middle| \gamma + t - \frac{\gamma_0}{2}, \alpha\right) e^{-2t/\Theta} \frac{2}{\Theta} dt d\Theta, \quad (1)$$

where γ_0 is the age of the Neanderthal specimen (we set this to zero because the age is unknown and errors here will not change the outcome). Prob(D|. . .) is the likelihood of the data over the lengths of the genealogy, and α is the shape parameter of the substitution rate distribution. Our program approximates the integral over all rates with a discretization of the Gamma distribution.[10]

Essentially, formula 1 is a distance measure that takes into account the facts that the ancestral population was not zero, that there are unequal nucleotide frequencies, and that there

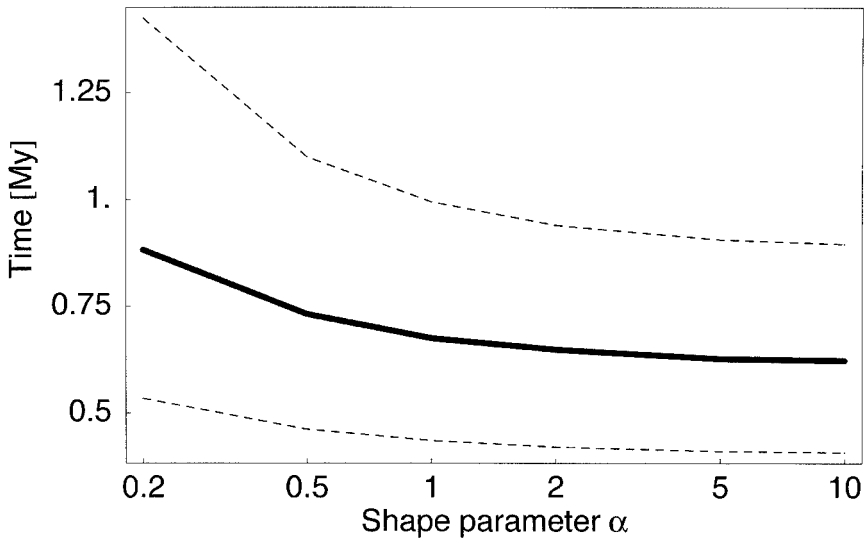

Figure 2. Relationship of estimated population divergence time with shape parameter α of gamma distribution assumed for variation of substitution rate among sites. Data were HVR I sequences of the Neanderthal and a !Kung (individual 10). Distance between top and bottom line is 95% confidence interval of time; center line is maximum likelihood estimate.

is substition rate variation among sites. For simplicity, we assume here a prior distribution for Θ that is rectangular, i.e., in which the probability of Θ being outside some minimum and maximum values is zero, and within that range is constant and greater than zero. We used two ranges (Fig. 1B), assuming that the mean ancestral population size was small (1–20,000 individuals with a mean of 10,000) and somewhat larger (1–200,000 individuals, mean 100,000), respectively. The sequence of divergence events using mtDNA of chimpanzees, Neanderthals, and contemporary humans is easily recovered using the Bayesian approach: chimpanzees diverge from *Homo*, after which humans split into Neanderthals and modern humans (Fig. 1A). As expected, the confidence limits on population divergences at each split exceed those of gene divergences, with a smaller mean ancestral Θ, and the relative difference between the most likely gene divergence time and the most likely population divergence time increases at more recent divergence events (data not shown). The estimation of the splits of humans and chimpanzees or between Neanderthal and contemporary humans is virtually independent of the human sequence used (we tried several individuals from the human mtDNA database *hvrBase*[13]).

Translation of Θ and γ into absolute measures of population size N_e and time τ, respectively, requires an estimate of the substitution rate of the mtDNA control region domain I. This rate is currently hotly debated.[14] Allowing for substitution rate heterogeneity among sites with a shape parameter of the mutation rate distribution of 0.4 (cf. Excoffier and Yang[3]), and setting the gene split of human and chimpanzees at 5.0 My ago, suggest a substitution rate for HVR I of 0.0634 per site per milllion years. Using these parameters and the two different rectangular priors for the ancestral population size, we estimate a population divergence time between modern humans and Neanderthals of 792 KY (500–1,227 KY; dashed line II, Fig. 1B) or 756 KY (425–1,172 KY, solid line II, Fig. 1B), respectively.

Despite our taking ancestral Θ into account, these divergence times are greater than those previously reported.[1,2] The rate heterogeneity correction which Krings et al.[1,2] applied seems to be overly biased by contemporary human sequences and too high, rendering their estimates of the gene divergence too low.

A 20-fold higher mutation rate per site per year has been derived from studies of human families.[14] Divergence dates of Neanderthal and contemporary humans based on this rate (41 KY with a range of 26–63 KY, and 39 KY with a range of 22–61 KY, respectively) are too small, given that the Neanderthals probably died out around 30 KY ago and were distinct from other human groups by at least 200 KY ago.[6] All these absolute times, however, are highly dependent on the correct estimation of the human-chimpanzee divergence, mutation rates, and particularly the value of specifying rate heterogeneity among sites (Fig. 2).[10] A firm dating will require multiple unlinked loci,[15] the variance among which can be used to estimate the population size of the ancestor of modern humans and Neanderthals. Even though we can calculate an approximate population divergence time between modern humans and Neanderthals, the mtDNA data cannot rule out a period of gene flow between Neanderthals and modern humans. But, as recognized by Krings et al.,[1] one sequence that happens to coalesce deeply before the most common recent ancestor of modern human mtDNA probably was not part of the gene pool of modern humans. Our divergence time estimates are reconcilable with a scenario in which the ancestor of modern humans and Neanderthals split into African and non-African lineages, with Neanderthals developing in situ, and contemporary humans arising from African lineages. An alternative scenario,[6] with contemporary humans and Neanderthals diverging about 250 KY ago, is unlikely given our confidence limits, and would require an ancestral population size in excess of 200,000 individuals, a population size vastly greater than the estimated 10,000 individuals of recent paleobiological studies.[6] Alternatively, one needs to invoke a highly substructured large ancestral population. Coalescent simulations using two subpopulations with equal subpopulation size show that one would need a very low migration rate m ($N_e m < 0.01$) to accommodate a 250 KY divergence.

The C source code for the population divergence estimation program *diverge* will be available from http://

evolution.gs.washington.edu/beerl;/vicariance.html.

ACKNOWLEDGMENTS

We thank Thomas Uzzell and Robert D. Martin for their critical comments. P.B. was supported by National Institutes of Health grants GM 51929-02 and GM 01989-02, and by National Science Foundation grant BIR 9527687, all to Joseph Felsenstein. S.V.E. was supported by National Science Foundation grants DEB 9707458 and DEB 9815800.

REFERENCES

1. 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.

2. Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S. 1999. DNA Sequence of the mitochondrial hypervariable region II from the Neanderthal type specimen. Proc Natl Acad Sci USA 96:5581–5585.

3. Excoffier L, Yang Z. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16:1357–1368.

4. Wakeley J. 1994. Substitution-rate variation among sites and the estimation of transition bias. Mol Biol Evol 11:436–442.

5. Takahata N, Satta Y. 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc Natl Acad Sci USA 94:4811–4815.

6. Lahr MM, Foley RA. 1998. Towards a theory of modern human origins: geography, demography and diversity in recent human evolution. Yrbk Phys Anthropol 41:135–176.

7. Zollikofer CPE, Ponce de León MS, Martin RD. 1998. Computer-assisted paleoanthropology. Evol Anthropol 6:41–54.

8. Nielsen R. 1998. Maximum likelihood estimation of population divergence times and population phylogenies under the infinite sites model. J Theor Popul Biol 53:143–151.

9. Edwards SV, Beerli P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839–1854.

10. Yang Z. 1997. On the estimation of ancestral population sizes of modern humans. Genet Res 69:111–116.

11. Felsenstein J, Churchill GA. 1996. A hidden Markov model approach to variation among sites in rate of evolution. Mol Biol Evol 13:93–104.

12. Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximative methods. J Mol Evol 39:306–314.

13. Handt O, Meyer S, von Haeseler A. 1998. Compilation of human mtDNA control region sequences. Nucleic Acids Res 26:126–129.

14. Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllensten U. 1998. Mitochondrial mutation revisited: hot spots and polymorphisms. Nat Genet 18:109–110.

15. Nordborg M. 1998. On the probability of Neanderthal ancestry. Am J Hum Genet 63:1237–1240.