Evaluation of the Critical Factors in the Phylogenetic Analysis of Human and Neanderthal mtDNA

Christopher R. Moradi ∗, and Alexander Schuster ∗

∗ San Francisco State University Department of Mathematics, 1600 Holloway Avenue, San Francisco, CA 94132

Submitted to Proceedings of the National Academy of Sciences of the United States of America

We report the results of the first phylogenetic analysis of the relationship of humans and Neanderthals with the inclusion of the El Sidrón and Monte Lessini sequences. By performing multiple analyses on several smaller data sets, we provide evidence of the effects of the inclusion/exclusion of sequences, the substitution model, the tree construction technique, and the branch support method on the separation of these groups. We find no support for the separation of humans and Neanderthals based on the HVRI and HVRII sequences presently available. In addition, we provide evidence that the high support for separation found in previous studies is due to deficiencies in the phylogenetic analyses performed.

Neanderthal mtDNA | human evolution | phylogenetics

Introduction

The Neanderthals are an extinct group of archaic humans from the Middle/Upper Pleistocene that have caused a great deal of controversy, confounding researchers ever since their fossils were discovered in the early nineteenth century. The controversy deals with the ultimate fate of the Neanderthals. Did they make a major contribution to the gene pool of anatomically modern humans, or did they simply fade into oblivion without having any genetic effect on us at all, or does the true answer lie somewhere between these two extremes? Was there evolutionary continuity between the Neanderthals and modern European populations, or were the Neanderthals completely replaced by the modern humans? This question has implications for the more general debate over the origins of modern humans.

The proponents of the Multiregional Continuity Model ([1, 2]) hold that modern human populations evolved from the exodus of Homo erectus from Africa approximately one million years ago. On the other hand, those who favor the Out-of-Africa Model ([3, 4]) claim a much more recent African origin for modern humans. They posit that these anatomically modern Africans spread all over the Old World and replaced the descendants of the earlier African emigrants, which include the Neanderthals.

Researchers on both sides of the argument have claimed that the fossil evidence supports their respective position. Some claim the existence of fossils that provide a bridge between Neanderthals and anatomically modern humans, while others scoff at this position. Many anthropologists agree that the fossil evidence is inconclusive. (See [5] for a review of the fossil data.)

In 1997, a new type of evidence emerged that researchers hoped would shed more light on the situation; Krings et al. ([6]) produced the first hypervariable region one (HVRI) mitochondrial DNA (mtDNA) sequence from Neanderthal remains. Additional sequences have since been produced by Krings et al., as well as other research groups. As of this writing, four unique HVRI-only sequences with more than 300 sites have been extracted from Neanderthal remains, and two sequences with both the hypervariable regions one and two (HVRI+II) data have been extracted (Table 1).

The early sequencing groups analyzed the available Neanderthal sequences in relation to human and chimpanzee sequences and concluded that there was a strongly supported separation between these groups. The papers that present the phylogenetic analysis of the relationship of humans to Neanderthals based on the mtDNA data are summarized in Supplemental Table 1. Comparisons of the average pairwise differences ([6, 7, 9, 8, 10]) and pairwise distances ([14]) between Neanderthal and human sequences were used to support the separation of these groups. The authors of [6, 7, 9] found high support for separation using likelihood-mapping (LM, [16]) of neighbor-joining (NJ, [15]) trees. In [8], the authors found similar support with bootstrapped (17) NJ and maximum parsimony (MP) trees. As a result of these findings, recent papers reporting new Neanderthal sequences ([12, 13, 18]) have limited their analyses to determining the placement of new Neanderthal sequences with respect to the other Neanderthal sequences and estimating the genetic diversity of this population; phylogenetic analyses of these new Neanderthal sequences in relation to human sequences have not been done by these recent sequencing groups. Although the sequencing groups have not continued to analyze these data, [19] used Bayesian techniques to estimate the phylogenetic relationship between these groups. Their analyses showed 100% posterior probability (PP) support for the separation model.

While consensus has rapidly formed that Neanderthals did not make any genetic contribution to modern humans, supporting the Out-of-Africa model, several studies have sharply criticized this conclusion. Questions have been raised about the authenticity of the sequences—primarily of FE1 ([6])—by several papers ([10, 19, 20, 21, 22, 23]). In addition, opponents argue that the HVRI and HVRII regions are inadequate to resolve these phylogenetic questions ([20, 24]). Gutiérrez et al. ([22]) criticized the use of pairwise differences because of biases in the human sequence pool and the use of LM because it would overestimate branch support levels. As part of their own analysis of the sequences, they found no support for separating these groups based on bootstrapping NJ trees or using the interior branch test ([25]) on NJ trees.

This research is directed towards reanalyzing the human and Neanderthal data with the inclusion of the new Neanderthal sequences (SI2 and MLS). In addition, we examine the critical factors for the phylogenetics analysis of these data to determine the stability of the best-fit model selection, and the support for separating these groups with neighbor-joining and Bayesian inference. We also evaluate the separation by comparing the likelihood scores of maximum likelihood (ML) trees containing a human monophyly (HM) against those that do not. Finally, we provide evidence that the high support values

Reserved for Publication Footnotes
Results and Discussion

Model Selection. The model selection process implemented with PAUP* [26] and MODELTEST [27] was unable to find a consistent best-fit substitution model for these data across all data sets. For all selection criteria used, the selected model varied in complexity from HKY+Γ (28), to the most complex model available in PAUP*, GTR+Γ+i (29). The results of the model selection tests are shown in Supplemental Table 2.

The differences in human sequences included in each data set led to the selection of different models. While the use of a Jukes-Cantor (JC, [48]) NJ base tree is typically considered adequate for estimating parameter values and the likelihood scores for evaluation in MODELTEST ([30]), this substitution model may be far too simplistic to estimate an adequate base tree for these sequences. Alternatively, the lack of data (short sequence lengths) likely contributes to the inability of this method to find a consistent best-fit model. We found that the Bayesian information criterion (BIC, [31]) chose the simplest model in nearly all tests, which matches the findings of [32]; the Akaike information criterion (AIC, [33]) generally selected the most complex model for the small data sets, while the hierarchical likelihood ratio test (hLRT) generally selected the most complex model for the large data sets.

While we found no single best-fit model for the HVRI and HVRII regions of human, Neanderthal, and chimpanzee mtDNA sequences, we did find parameters that should be present in the model. Firstly, all of the models selected over all criteria included the three parameters to specify unequal base frequencies. Secondly, all of the selected models included the gamma shape parameter (with α < 1), and nearly all (152 of 160) included the proportion of invariable sites parameter. This indicates that the sequences show a significant amount of rate heterogeneity by site that must be taken into account for an accurate analysis of the data. Finally, all models selected included at least two rate categories (e.g., transitions occur at a different rate from transversions).

Although the authors of [6, 7, 9] did not explicitly state the model that they used for tree construction or branch support, Gutiérrez et al. ([22]) stated that the model was F84. This model would be insufficient to determine an accurate tree as well as to find accurate branch support values, because it neglects the high levels of rate heterogeneity exhibited in the HVRI and HVRII regions of human mtDNA.

Bootstrapped Neighbor-joining Analysis. The bootstrapped NJ trees constructed show that there is a lack of support for the separation of humans and Neanderthals based on the HVRI and HVRII sequence data presently available. We did find 100% support for a chimpanzee monophyly (CM) with all data sets. In addition, we found high (above 90%) bootstrap proportion (BP) values for a monophyly of the Neanderthals; however, the placement of this clade with respect to the human sequences was not well supported. Figure 1 shows the BP values of the human monophyly (HM) when all sites were included in the analysis (full), as well as for the data sets from which sites with a majority of missing and ambiguous data were removed (trimmed). The support fell below 20% for nearly all data sets containing all sites once rate heterogeneity was included in the substitution model. Analysis of the trimmed data showed an increase in the BP support over the BP values of the full data set. However, when an adequate substitution model was used (e.g., F84+Γ+I [34, 35]), the support values for nearly all data sets were still below 30% and 60% for the HVRI and HVRII-II sets, respectively.

Our results seem to match those found by Gutiérrez et al. ([22]), but are much lower than those found by Ovchinikov et al. ([8]). Gutiérrez et al. reported BP support for an HM of 29% for the HVRI data and less than 32% for the HVRII+II data set. Ovchin-
Topology Tests. GARLI found trees with similar likelihood scores when both a positively and a negatively human monophyletic constraint was used. If there existed a strong phylogenetic signal that showed the separation of the Neanderthals and humans, we would expect a significant improvement in the likelihood score of the tree with the positive HM constraint over the negatively constrained tree for all of our data sets. For all data sets analyzed, the Shimodaira-Hasegawa (SH) test as implemented in PAUP* showed that neither tree was significantly better than the other. While there may be a concern that two runs we performed for each data set is not enough to adequately examine the tree space, we estimate that the ratio of positively constrained HM trees to the set of all trees is $2.6 \times 10^{-19}$ (see Supplemental data). Therefore, the search is more likely to miss an optimal tree without an HM than one showing an HM. Because there were no data sets that showed a significant difference in the likelihood score of the two constrained trees, this reinforces the findings of our other analyses—showing that the data available does not support the separation of these groups.

Additional Comments on Prior Work

Pairwise Differences. Several studies analyzing the relationship of human and Neanderthal sequences ([6, 7, 9, 8, 10]) used the number of pairwise differences (PWD) between the human and Neanderthal sequences to show the separation of these groups. In addition to the concerns over bias in the sequence database ([22]), we would like to point out additional concerns with the use of pairwise differences comparisons. Firstly, this analysis technique fails to acknowledge that substitutions are not created equal; there are high levels of rate heterogeneity in the HVRI and HVRRII portions of the human D-loop. Some changes will be much more common than others, and therefore add less to the divergence of these sequences. However, by simply counting the number of differences between two sequences, we are assigning each change an equal weight. Secondly, we believe the comparison of PWD is much more susceptible to errors in the sequence than phylogenetic methods. Several papers have questioned the authenticity of the FE1 sequence ([22, 20, 10, 21, 19]). In particular, [10] raised concerns over the four substitutions at positions 16107-8 and 16111-2, that are unique to the FE1 sequence and not present in the other Neanderthal sequences. Additionally, post-mortem artifacts are expected to occur more frequently at sites with higher evolutionary rates ([38]). Hence, these errors will have less effect when distance-based or ML trees are constructed than changes in regions with low substitution rates. However, both of these types of substitutions will have equal effect on the divergence calculations. With phylogenetic analysis, sequence errors will affect the branch lengths, but may not be enough to affect the placement of the sequence in the tree.

Likelihood Mapping. We believe that in [6, 7, 9], Krings et al. used likelihood mapping (LM, [16]) in an incorrect manner to establish branch support for their NJ tree. Gutierrez et al. ([22]) criticized their use of LM by showing that it was susceptible to long-branch attraction (LBA). We agree with this critique since Krings et al. reported their results using three basins of attraction for the mapping triangle. However, TREE-PUZZLE ([39]) also provides results using a mapping triangle divided into seven basins. This additional partitioning of the triangle helps to mitigate the LBA effects noted in [22]; the center triangle shows the percentage of sequence comparisons with a star signal (all quartets have similar likelihood scores), while the three rectangular side sections show the sequence comparisons with a network-like signal (two quartets show similar likelihood scores). Figure 3 shows an LM analysis we performed on a data set selected because the results approximate the results presented in [6] (only the FE1 sequence was included in the analysis). When we consider the mapping triangle with three basins, there appears to be 88.4% support for grouping the Neanderthal sequence with the chimpanzee sequences. However, when we examine the seven basins output triangle, 15% of the quartets show an inconclusive signal from the sequences (at least two quartets show similar likelihood scores). In addition, 4% of the comparisons yield quartets that show a strong pairing of the Neanderthal sequence with one human sequence.

In addition to the problems with using the three basin mapping triangle, we believe that Krings et al. ([6, 7, 9]) incorrectly grouped the taxon to determine branch support levels. From the data provided it appears that the support for the NM in Figure 7b of [6] was determined by selecting sequences from three groups: chimpanzees, the Neanderthal, and all humans. We believe that support for this branch should have been determined using sequences from four groups: chimpanzees, the Neanderthal, the four Africans nearest the Neanderthal, and the remaining humans. The authors did not elaborate on the method used for obtaining the other interior branch support values. It is unclear what the LM branch support values presented in [6, 7, 9] indicate. They seem to show that the Neanderthal sequences are not deep within the human clade; however, this does not seem to be an argument that anyone is purporting. Supplemental Figure 1 shows the results of an analysis in which we use the methodology of Krings et al. to compare a divergent African sequence (GenBank accession no. AF346992) to chimpanzee and other randomly selected human sequences. There is 94.6% support for a strong grouping of these sequences with chimpanzee sequences; however, in the seven basin map we can see that 2.8% of the comparisons reflect a strong grouping of this sequence with another human. While the support for separating this sequence from other human sequences is strong using this methodology, this does not indicate that the African sequence is from a separate species.

In contrast, we used one of our small HVRI data sets to find the the LM support for a branch separating the chimpanzee sequences and African sequence AF346992 in this set from the Neanderthals and 199 other human sequences. In order to determine the support for the tested branch in Figure 4, we used four groups for our LM analysis: chimpanzees, Neanderthals, the basal African, and the 199 other human sequences. This analysis shows (Figure 4) that 84.2% of the quartets show a strong signal pairing the African sequence with a chimpanzee sequence.

Divergence Time of Humans and Neanderthals. Many of the proponents of the separation of humans and Neanderthals have used the estimation of the most recent common ancestor (MRCA) as evidence against intermixing ([6, 7, 9, 8, 10]). However, we believe that inadequate phylogenetic analyses invalidate these estimates. Firstly, a key to the MRCA calculation is that the correct phylogeny has been estimated; our results as well as those of others ([22]) show that the separation of these groups is unsupported. Secondly, the studies have failed to use either a relative-rate test ([41]) or the likelihood ratio test ([32]) to support the assumption that the data fits a molecular clock ([42]). Thirdly, these studies ignore a main source of error in the estimation of the confidence interval for the divergence times which is the stochastic variation in the clock rate ([43]). Finally, in [6] the authors used a simpler substitution model (F84) for the estimation of the phylogeny and a much more complex model (TN+Γ+I, [50]) to estimate the distances used to calculate the MRCA.

Data and Methods

Data Set Construction. Sequences were obtained through HVR-Base++ ([44]; http://www.hvrbase.de), which is an updated version of the same repository used in many of the previous studies of human and Neanderthal mtDNA. From this database, all available manually aligned HVRI and HVRRI sequences were retrieved and sorted into human, Neanderthal, and primate sequences. Additional Neanderthal and chimpanzee sequences were retrieved from GenBank and manually aligned to the other sequences. Sequences were compared (ignoring ambiguous and missing sites) and duplicates were
discarded, keeping the sequence with the fewest missing/ambiguous sites. HVRI sequences with matching HVRII data were joined to form the HVRI+II sequences. The HVRI and HVRI+II pools contained 11,265 and 3,791 human sequences, respectively; after removing the duplicate sequences these pools contained 4,471 and 2,597 human sequences, respectively.

From each sequence pool, large and small data sets were created for use in our analyses. Thirty large data sets were created from the HVRI and HVRI+II sequence pools. For the HVRI pool, these consisted of 500 randomly selected human, 6 chimpanzee (three Pan paniscus and three Pan troglodytes), and 6 Neanderthal sequences each (FE1, FE2, MEZ, V75, MLS, and SI2); for the HVRI+II pool, these consisted of 500 human, 6 chimpanzee, and 2 Neanderthals sequences each (FE1 and V75). In addition, 50 small data sets were created using the same Neanderthal and chimpanzee sequences, except that these contained only 200 human sequences each.

Because of the open questions concerning the influence of ambiguous data in phylogenetic analyses ([22, 45, 46, 47]), we repeated experiments with the ambiguous sites excluded and highly-ambiguous sites removed. For the “trimmed” data sets, the ends of the full data sets were removed from HVRI and HVRII regions so that only the regions that contained data for the Neanderthal sequences were included in the analysis. In addition, columns were removed where the majority of sequences showed either a gap or an ambiguous site. Since this process was performed separately on each data set, some of the resulting sets included more sites than others. The minimum number of sites in the trimmed sets were 360 for all of the HVRI data sets, and 645 and 659 for the small and large HVRI+II data sets, respectively. Finally, we created a third group of data sets, which we will refer to as the “deleted-sequence” data sets, by excluding all sequences with a large number of ambiguous sites from the first 10 data sets of each of the small and large HVRI and HVRI+II full sets.

Model Selection. The model selection process described in [32] was applied to each full data set (all sequences and sites), using PAUP* 4.0b10 ([26]) and MODELTEST 3.7 ([27]). The best-fit model for each data set was determined using three selection criteria: hLRT, AIC, and BIC. For BIC, 438 and 875 were used for the number of characters parameter for the full HVRI and HVRI+II data sets, respectively. The defaults were used for all other MODELLTEST settings.

Bootstrapped Neighbor-joining Analysis. In order to resolve the conflicting results of [8] and [22], we performed neighbor-joining (NJ) bootstrap analysis of our data sets for seven different substitution models: JC ([48]), JC+Γ, F81 ([49]), F84 ([34, 35]), F84+Γ+I, TN+Γ+I ([50]), GTR+Γ+I ([29]). By using the multiple small and large data sets, we sought to determine the impact of the human sequences included on the resulting bootstrap proportion (BP) support values. While most phylogenetic analyses are done with a single substitution model, we used seven different models to quantify the influence that specific model parameters have on the support of a CM, NM, and HM. For each data set, an NJ tree was constructed based on ML distances with the model parameters fixed to pre-specified values (see Supplemental Table 2). Using this tree, the model parameters were re-estimated and used to construct a final NJ tree. A bootstrap analysis was performed on the final tree with 500 replicates, and the group frequencies were used to find the bootstrap proportion (BP) for the desired branches. In all cases, ties were broken randomly in the neighbor-joining construction. For models with a gamma distribution specifying rate heterogeneity, four rate categories were used. Identical analyses were done on the trimmed and the deleted-sequence data sets. For all bootstrap analyses, a seed value of 12345 was used so that comparisons could be made between tests while removing the variability due to replicate construction.

In addition to collecting the BP support values for the three monophyletic groups of interest, the group frequencies from the bootstrap analyses were used to identify divergent human sequences. In the first case, we searched for the highest grouping of human sequences with the CM to the exclusion of Neanderthal sequences. In the second case, we looked for the highest grouping of human sequences with the NM to the exclusion of the chimpanzee sequences. The human sequences that showed one of these pairings most frequently across all data sets and substitution models were considered divergent sequences.

Bayesian Tree Inference. We used MrBayes 3.1.2 ([51, 52]) to investigate the support for the separation of humans and Neanderthals under the Bayesian paradigm. We limited our analyses to one of the fifty small HVRI data sets that contained all sites and sequences. We chose to analyze the HVRI set over an HVRI+II set because it allowed us to include all 6 Neanderthal sequences. In addition, while [19] concluded that their HVRI-only analysis showed strong support for the separation of humans and Neanderthals, the 63% PP support that they found for an NM is too low to support this conclusion. Since [19] found high support for a human monophyly using a larger set of human sequences for their Bayesian inference, we selected a set that contained a divergent African sequence (GenBank no. AF346992). Similar to [19], we used the defaults of MrBayes and GTR+Γ+I as the substitution model. The analysis was run for 25 million generations, sampling every 1000 generations and discarding 25% for the burn-in phase. We verified that the two runs converged based on the average standard deviations of the split frequencies and Potential Scale Reduction Factors.

Topology Tests. Bootstrapped neighbor-joining has produced conflicting results with regards to analysis of human and Neanderthal mtDNA ([8, 22]), which have led some to speculate ([19]) that this technique is inadequate for these data. In addition, Bayesian inference is often criticized because of the priors chosen for parameter values ([53]). Because of the concerns regarding these methods, we decided to add a ML tree search to our test methods. Instead of using bootstrap to determine branch support, we sought to examine the difference in the likelihood scores of the best tree under the constraint that the human sequences form a monophyletic group to the exclusion of Neanderthal sequences against the best tree that did not contain a human monophyly. GARLI 0.95 ([54]) was used for these tests because of its ability to rapidly find good ML trees under topology constraints. For each data set, we searched for the best ML tree under an HM constraint, and then for the best tree under the negative of this constraint (i.e., a tree that does not contain an HM). This was done for only the small HVRI and HVRI+II data sets because of the time required to find each tree. Two runs were performed for each data set, with the best tree kept for each type of constraint. F84+Γ+I was used as the substitution model for all tests. A random starting tree was used. The default values were used for all other settings. The final topologies were then reoptimized in PAUP*, as per the instructions of GARLI’s author, and a final likelihood score was obtained.

The two trees found under the positive and negative constraint of an HM were compared using the SH test. These likelihood scores were compared using the resampling of estimated log-likelihoods (RELL) SH test in PAUP* to see if the likelihood score of the tree with the human monophyly was significantly better than the tree that did not contain this monophyly.

Although this comparison may be considered a misuse of the SH test because the trees being compared were selected by ML search-violating the test’s assumptions–this is acceptable to show that the trees are not significantly different. If however, we intended to show that the one tree was optimal, our use of the SH test would not be acceptable. In that scenario, we would be required to compare the
ML tree to all other possible optimal trees, and to show that our ML tree was significantly better than the set of other trees.

ACKNOWLEDGMENTS.

This paper benefited from the careful review and comments provided by Nicoletta Caldararo, Eric Routman, and Greg Spicer. The authors thank the Dragutin Petkovic and Mike Wong of the Center for Computing in the Life Sciences (CCLS) for providing the computing power for this research. C.M. was partially supported by a CCLS Mini-grant. A.S. was partially supported by NSF grant DMS 0601060.


Table 1. List of Neanderthal mtDNA sequences with more than 300 sites.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Location of remains</th>
<th>Abbrev.</th>
<th>Type</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>[6]</td>
<td>Feldhofer, Germany</td>
<td>FE1</td>
<td>HVRI</td>
<td>379</td>
</tr>
<tr>
<td>[7]</td>
<td>Feldhofer, Germany</td>
<td>FE1</td>
<td>HVRII</td>
<td>345</td>
</tr>
<tr>
<td>[8]</td>
<td>Mezmaiskaya Cave</td>
<td>MEZ</td>
<td>HVRI</td>
<td>345</td>
</tr>
<tr>
<td>[9]</td>
<td>Vindija, Croatia</td>
<td>V75</td>
<td>HVRI</td>
<td>357</td>
</tr>
<tr>
<td>[9]</td>
<td>Vindija, Croatia</td>
<td>V75</td>
<td>HVRII</td>
<td>288</td>
</tr>
<tr>
<td>[10]</td>
<td>Feldhofer, Germany</td>
<td>FE2</td>
<td>HVRI</td>
<td>357</td>
</tr>
<tr>
<td>[12]</td>
<td>El Sidrón, Spain</td>
<td>SI2</td>
<td>HVRI</td>
<td>303</td>
</tr>
</tbody>
</table>

*Identical to V75

Fig. 1. NJ bootstrap support for an H M for full (all sites) and trimmed (excluding missing/ambiguous sites) data sets

Fig. 2. Consensus tree from Bayesian inference on HVRI data set containing: 6 chimpanzee, 6 Neanderthal, and 200 human sequences.
Fig. 3. LM support with six chimpanzee (Group A), FE1 Neanderthal (Group B), 200 Human sequences (Group C). Human data set chosen to approximate results in [6].

Fig. 4. LM analysis of HVRI region with six chimpanzee (Group A), six Neanderthal (Group B), a divergent African human (Group C), and 199 randomly selected human sequences (Group D). Based on TN+Γ-I with parameter values from [22].