Supplementary material for The timing of selection at the human *FOXP2* gene

Graham Coop[1,3], Kevin Bullaughey,[2], Francesca Luca[1] and Molly Przeworski[1,3]
[1] *Department of Human Genetics, University of Chicago, 920 E. 58th Street, CLSC 5th floor Chicago, IL 60637*
[2] *Dept. of Ecology & Evolution, University of Chicago*
[3] *To whom correspondence should be addressed: gcoop@uchicago.edu, mfp@uchicago.edu*

1 Dating the selected haplotype at *FOXP2* in modern humans

Our aim was to estimate the mean time to the most recent common ancestor (tMRCA) of the selected haplotype, using an approach that does not rely on assumptions about the demographic and selective history of the region but only on the assumption that observed polymorphisms are neutral [1, 2]. We note that, assuming selection acted on a new mutation, the tMRCA of the selected haplotype is always more recent than the time at which the selected mutation occurred, but that the difference between the two is not large. Our approach requires us to identify the selected haplotype and the precise estimate of the tMRCA depends on this step; however, the goal here is mainly illustrative.

To estimate the tMRCA, we analyzed the genotype variation data obtained by Enard et al. [3] for a 14 kilobase region near exon 7 (see Figure 1 (a)). First, we inferred haplotypes using the program PHASE [4]. We ran the program five times; since the inferred haplotypes were identical in four of the five cases, we used the most likely phase assignments in our analysis. We then excluded haplotypes that carry the ancestral background, as defined by the nine sites with high frequency derived alleles considered by Krause et al. [5] (see Figure 1 (b)). The average number of derived mutations per bp, M, on the remaining haplotypes reflects the mean tMRCA [1]. To translate M into time in years, we used the estimate of human-chimpanzee divergence for the region reported by Enard et al. [3] of 0.87% per bp. Assuming a time to the most recent

common ancestor of 6.5 My for the two species, the tMRCA is then given by $(M/0.0087) \times 2 \times 6,500,000 = 42,000$ years. We estimated the 95% confidence intervals (CI) of the tMRCA [2] to be 38,000-45,500 years. Uncertainty in the time to the most recent common ancestor of human and chimpanzee widens the CI, but even a common ancestor 8 million years ago leads to an upper CI of only 55ky. The estimate of the tMRCA would become even smaller if we used genome-wide average divergence levels ($\sim 1.2\%$; [6]) between human and chimpanzee instead of local divergence levels to estimate the mutation rate. Since error in haplotype estimation could affect our conclusion, we also estimated the tMRCA for individuals that are homozygous for the derived haplotypes (i.e., for which phasing is unambiguous). Using this subset of the data, we obtained an estimate of 53,000 years. There is considerable uncertainty associated with the estimates of the tMRCA, and, in particular, our confidence intervals do not include the uncertainty in reconstructing the ancestral haplotype on which the selected mutation arose. However, the estimates are an order of magnitude smaller than the split time of modern human and Neandertal populations.

2 Possible explanations for the *FOXP2* data from modern humans and Neandertals

Assuming that positive selection acted near the intron surveyed for variation in modern humans, and that there was no (or very little) modern human contamination, there are (at least) three possible explanations for the variation data from Neandertals and modern humans.

(i) One possibility is that selection occurred in modern humans subsequent to the population split with Neandertals and acted on an allele already segregating at appreciable frequency in the ancestral population. As pointed out by Krause et al. [5], however, this type of selection is not expected to lead to the patterns of variation observed in modern humans i.e., it does not explain the excess of high frequency and rare alleles [7].

(ii) As also pointed out by the authors, an alternative explanation could be that the evidence for selection at *FOXP2* in modern humans is not due to the amino-acid variants in exon 7 but to subsequent beneficial substitutions nearby. If the high frequency derived haplotype found in humans is due to selection at a linked site, however, how to explain its presence in the two Neandertals? (See main text.)

(iii) An explanation more consistent with the data may be admixture between Neandertals and modern humans since their split. The two Neandertal samples examined by Krause et al. [5] are from Spain and dated to 43 Kya, a time and place where Neandertals and modern humans are not thought to have co-existed. However, the two populations are thought to have overlapped for tens of thousands of years further East; had some interbreeding occurred there, the favored FOXP2 allele could have spread by gene flow within Neandertals to

Spain. Alternatively, the beneficial mutation could have arisen in Neandertals, and passed into the modern human gene pool more recently than 43 Kya. This possibility is especially intriguing given the extensive controversy surrounding the origin of modern linguistic abilities.

All three explanations assume that there was indeed selection in *FOXP2*, based on the evidence presented in Enard et al. [3] and Zhang et al. [8]. Further evidence for recent positive selection at *FOXP2* could potentially be garnered by comparing patterns of polymorphism in *FOXP2* to genome-wide patterns to assess whether they are unusual.

3 The contamination assays

In what follows, we discuss the assumptions of the contamination assays performed by Krause et al. [5] and reanalyze the control data, examining whether the assays are sufficient to determine with high certainty that the two derived amino-acid alleles were carried by the two Neandertals sampled. We note that the purpose of a control assay is not simply to establish that some Neandertal nuclear DNA is recovered, but to help estimate the contamination rate. At each control or target (*FOXP2*) site, there are a number of amplifiable fragments (i.e., fragments that overlap the primers designed to amplify the small region), some of which derive from the Neandertal sample and some of which are modern human contaminants (a set hereafter referred to as the pool of alleles). The contamination assay must allow one to assess whether the ratio of recoverable non-contaminant to contaminant DNA is sufficiently high to ensure that the true Neandertal alleles are very likely to have been observed in the recovered products at the target sites (here, sites in *FOXP2*). The distinction between the two goals of a contamination assay – establishing that some Neandertal DNA was recovered versus that a sufficient fraction of Neandertal DNA was recovered – is somewhat blurred in the authors' conclusion that "we infer that the multiplex PCR products derive from Neandertal nuclear - DNA sequence".

3.1 Assumptions of the assays

The authors focus on two male Neandertal bones for which a mtDNA assay suggested a low level of modern human mtDNA contamination. The relevance of this assay in assessing nuclear DNA contamination is uncertain. For one, because it is not clear from the text whether the mtDNA assays (performed in the Leipzig lab) were performed on the same extracts than the nuclear DNA assays or on different extracts from the same bones. Even if they were performed on the same extracts, differences in the stochiometry and decay patterns of mitochondria and nuclear DNA could lead to differences in the level of amplifiable contamination. For example, recovery of mtDNA through high-throughput sequencing in the Green et al. [9] study indicated correctly that Neandertal mtDNA is an outgroup to modern human mtDNA, suggesting low mtDNA con-

tamination – yet the same sequencing run produced autosomal data that were subsequently shown to suffer from high levels of modern human contamination [10]. Thus, mtDNA assays can help to exclude some cases of obvious modern human contamination, but their relevance in ruling out high contamination for the nuclear DNA remains uncertain.

Given this uncertainty, the authors also perform two-stage multiplex reactions for eight sites (in addition to those in *FOXP2*) to rule out contamination of nuclear DNA in the bone extracts. The goal of these control experiments is to assess the extent to which the recovered fragments are from the Neandertal at the *FOXP2* sites (the target sites) by testing a set of other sites in the genome for which there is a "known answer". Five of the control sites are on the Y chromosomes, one is on the X chromosome and two are on autosomes; the results are summarized in Table 1.

These control assays rely on a number of implicit assumptions. Firstly, that sites typed elsewhere in the genome are informative about contamination rates in the region of interest. This assumption might be undermined by:

1. Differences in the quantity/quality of Neandertal fragments along the genome (e.g., resulting from differences in damage across regions [11]) could affect the amount of amplifiable Neandertal DNA and hence the ratio of contaminant to Neandertal DNA at a given site.

2. The two stages of the amplification procedure [12]; only the first stage of the amplification is multiplex while, in the second stage, the product for each site is amplified independently [12]. As a result, the contamination rate may potentially differ across sites [13].

3. The amplification from low numbers of molecules of DNA, which raises the possibility that contamination is highly stochastic, such that contamination levels have a high variance across control sites (rather than being constant).

Another implicit assumption is that the contamination rate is the same across multiplex reactions, i.e., that successfully recovered control sites in one multiplex reaction (a column of Table 1) are informative about the contamination rate in other multiplex reactions. This assumption may be problematic as:

4. Additional primers are introduced in the final two multiplex reactions (3 and 4).

5. The number of successfully recovered sites differs markedly over multiplex reactions, suggesting that (stochastic) experimental conditions differ across reactions and so might the proportion of contamination recovered.

Some of these points are addressed by blank multiplex PCRs performed for each primer mix and each individual to rule out the possibility of contamination from the reagents used. The authors saw no specific product in several mock multiplex reactions for a total of 108 single amplifications (for the sites listed in Table 1). In interpreting these experiments, there are a number of caveats that make the interpretation uncertain. For example, the blank PCRs for Y chromosome primers would obviously not amplify female contamination. Moreover, PCR primers often fail to amplify even when DNA is known to be present: indeed, only 15 out of the 40 amplifications of the 5 Y chromosome primers produced products in the presence of Neandertal DNA (see the supplementary table of Krause et al. [5]). (We also note that if the mock multiplex reactions are used to assess the probability that a multiplex is contaminated due to reagents, then the number of informative trials is the number of multiplexes rather than the number of mock amplifications.) These caveats none withstanding, the blank PCRs suggest that differential probabilities of contamination due to points 2 and 4 above are unlikely. Whether contamination rates differ across sites or multiplex reactions for the other reasons discussed above remains unknown.

Even though this assumption may be invalid, we proceed by assuming, as do the authors, that there is a uniform contamination rate along the genome. We further assume (again, as do the authors) that the reactions start with a single molecule. This assumption is consistent with the observation that only one autosomal allele is ever seen in a given reaction (except for possibly A3 in sample B, see below); it is conservative if the goal is to rule out contamination. We examine the evidence against modern human contamination assuming first that contamination rates are shared across multiplex PCRs, then that they are not.

3.2 Y chromosome control sites

Trivially, Y-linked sites are not informative about potential human female contamination. In turn, the three Y chromosome SNPs (Y1, Y3 and Y5) shaded in gray in Table 1 are not informative with respect to modern European male contamination (the most likely source here), because many European males carry the ancestral allele at those sites and so cannot be distinguished from Neandertal. In particular, the derived alleles at sites Y3 and Y5 (which define haplogroup A_2 and B respectively [5]) are not present in Europe at appreciable frequency (see [14]) and so are not informative about European male contamination (but would be informative about some sources of African male contamination). In turn, the derived allele at site Y1 (which defines haplogroups P, R_1 and R_2 [5]) is at intermediate frequency in European males (see [14]); thus, observing the ancestral allele at this site does not exclude European male contamination (it was only recovered once, so that excluding it from our analysis of European contamination rates makes little difference anyway).

3.3 Non-Y nuclear control sites

The authors chose three autosomal (A1, A2, and A3) and one X chromosome (X4) control sites for which the Green et al. [9] data suggest that *one* Neandertal has the ancestral allele and where the human reference sequence has the derived allele, stating that these control sites are not known to vary in modern humans. The assumption of a fixed derived site in humans is pivotal; if it is not valid and both derived and ancestral alleles are in fact found in humans, then the recovery of the ancestral allele at a control site in the Neandertal sample could simply reflect modern human contamination and the assay becomes uninterpretable. In other words, the X and autosomal controls are only useful if the derived allele is fixed (or at least close to fixed) in humans.

Setting up a control assay that meets these assumptions is not trivial. Given the recent split time of human and Neandertal populations [15], the majority of sites that distinguish the human reference sequence from a putative Neandertal sequence are expected to be polymorphic in humans; moreover, because only a minority of human polymorphic sites are known [16], many will be as yet untyped. Thus, the control sites cannot be assumed to be monomorphic in humans simply because they are not known to vary but should be typed in a world-wide panel.

The site A2 failed in all Neandertal samples, so we considered the three remaining non-Y nuclear control sites. Examining the three control sites (A1, A3 and X4) considered by the authors, we found that Craig Venter [17] is homozygote for the ancestral allele (i.e., carries the "Neandertal allele") for one control site (A3), that all three control sites are polymorphic in a sample of HapMap [16] Yoruba and that both autosomal sites are polymorphic in HapMap Europeans (see Table 2). Given that all three sites are actually polymorphic in humans, human contamination could potentially result in the alleles seen at the three control sites. This said, at the X-linked site, only the derived allele was found in a (small) sample of Hapmap Europeans, so this site can potentially be used to assess European contamination levels. (In what follows, we assume that this is the case.)

At one autosomal site (A3), Krause et al. [5] also recover an allele not seen in chimpanzee or humans in one reaction (labeled 'N' in table 1). Because this variant does not match a known type of ancient DNA damage, the authors argue that this may be a novel Neandertal allele (potentially providing further evidence for low human contamination levels). However, the vast majority of variable nucleotide positions in humans are biallelic, with few triallelic positions. Given that human and Neandertal populations only split approximately 300kyrs [15], it is highly unlikely that in assaying three SNPs, a triallelic site was found. Instead, the novel allele could reflect some sort of experimental artifact.

3.4 Genotyping the non-Y nuclear control sites in the human Hapmap sample

We initially investigated the three control sites by using BLAT to locate the primers (and hence the locations) for the X and the two autosomal control sites in the human reference genome and then located these sites in Craig Venter's genome [17]. We found that his genome had only the derived allele at sites A1 and X4, but was homozygote for the ancestral allele at site A3.

We then performed additional genotyping. Specifically, a total of 56 samples from the three HapMap populations were amplified in three independent PCR reactions targeting the X and the two autosomal control sites considered by Krause et al. [5]. The number of samples successfully analyzed is reported in Table 2. The PCRs were performed in a total volume of 25 μl, using 0.4 μM of each primer, 200 μM of dNTPs, 1.5 mM of $MgCl_2$ and 1U of Bioline Taq DNA Polimerase. Following Exo-SAP purification, the PCR products were sequenced with ABI BigDye Terminator v. 3.1 Cycle Sequencing kit and the products were analyzed on an ABI 3730 automated sequencer (Applied Biosystems). A list of the PCR and Sequencing primers used in the assay is provided in Table 3. The Phred-Phrap-Consed package (Polyphred version 6.11) was used to assemble and analyze all sequences [18]. For each sample, the genotype at the control site was further confirmed by visual inspection.

3.5 An approach to assessing contamination levels

We present an approach to quantify the level of contamination in the Neandertal male samples, and the probability that we have observed the true Neandertal allele at a target site. We used a Bayesian approach to incorporate the uncertainty in the estimates of the contamination rate. We note that this framework rests on a number of assumptions, and so should be treated with caution. These assumptions are implicit in the work of Krause et al. [5], and represent a best case scenario.

The parameter of interest is the contamination rate c. We assume that the parameter c is shared across all sites (and initially across all multiplexes from the same individual). The only exception is the X-linked site in the presence of modern human female contamination of a male Neandertal sample. If c is the contamination rate on the autosomes, then $c' = \frac{2c}{2c+1-c} = \frac{2c}{c+1}$ is the contamination rate on the X for female contamination. On the X chromosome, female human contamination of male Neandertal nuclear DNA results in a higher (i.e., more detectable) contamination rate than on the autosomes.

The idea behind the framework used here is a simple one. Given a constant contamination rate along the genome, each recovery (draw) of an allele not found in European humans adds to the information about the level (rate) of European human contamination. Consider a locus with alleles 1 and 2, where allele 2

is fixed in Europeans and allele 1 is polymorphic (or fixed) in Neandertals. The unknown frequency of allele 1 in Neandertals is x. Our task is somewhat simplified by the fact that, in the Krause et al. [5] study, the only informative nuclear markers are hemizygotic (the Y and X in males), and so a Neandertal sample can have only one allele at each of the control sites of interest. The probability that we see allele 1 at a locus in a multiplex reaction is:

$$x(1-c) \tag{1}$$

Note that this depends linearly on the Neandertal frequency, and so the Neandertal frequency cancels out in the subsequent formulas. The posterior probability (i.e., the probability of the parameter given the data) of the contamination rate, $P(c|\text{Control Data})$, can be calculated assuming an uniform prior on the contamination rate:

$$P(c|\text{Control Data}) = \frac{P(\text{Control Data}|c)}{\int_0^1 P(\text{Control data}|c)dc} \tag{2}$$

If across multiplex reactions, we saw no evidence of contamination (no derived allele) at k control sites, then:

$$P(c|\text{Control Data}) = \frac{(1-c)^k}{\int_0^1 (1-c)^k dc} \tag{3}$$

We can use this probability to calculate the key probability of interest, namely the probability that, because of human contamination, the true Neandertal allele is not seen at a target site:

$$P(\text{Neandertal allele not seen}) = \int_0^1 c^l P(c|\text{Control Data})dc \tag{4}$$

where l is the number of successful recovered alleles at the target site. This represents the probability that we drew contaminated fragments all l times at a site, and thus that the Neandertal allele has never been observed.

We now consider the case in which the contamination rate is not necessarily the same over each multiplex reaction. In this case, we need to estimate the posterior of the contamination rate for each multiplex reaction. The probabilities can be constructed in a similar way as previously, but now considering the evidence from each multiplex reaction independently. For a multiplex reaction m, we can calculate $P(c|\text{Control Data in multiplex m})$ as in equation 3 where k is now the number of control sites in the particular multiplex reaction where no evidence of contamination (i.e., no derived allele) was seen. For each multiplex reaction m and each target site, we calculate P(Neandertal allele not seen in m) using $P(c|\text{Control Data in multiplex m})$, where l in equation 4 is now the number of successful recoveries (in these data, usually 1) of the target site. Information can then be combined over multiplexes as

$$P(\text{Neandertal allele not seen}) = \prod_m P(\text{Neandertal allele not seen in m}) \tag{5}$$

where $\prod_m$ is the product over multiplexes where at least one informative control allele has been typed.

As in all Bayesian approaches, the choice of prior distribution influences the outcome of the analysis, more so in situations when we have little information about the parameter of interest, as here regarding the contamination rate. The prior distribution captures our a priori beliefs about the probability that a given allele is Neandertal or modern human contaminant. Thus, the prior can be chosen to favor low or high contamination rates. In particular, the choice could be made to reflect additional information about the contamination rate, for example from the mtDNA assays. However, as discussed above, the relevance of the mtDNA is unclear. Since our goal is to assess the information provided by the *nuclear* control assays, we instead chose to use a uniform prior on the contamination level (this makes our analysis equivalent to a likelihood-based approach).

In our view, the dependence of the analysis on the prior does not reflect a shortcoming of the method, but instead stems from the fact that when there is insufficient information about the contamination rate, the subjective aspect of the analysis (the prior) becomes important. In designing nuclear contamination assays, the goal can be thought of as performing a sufficient number of informative assays that the choice of prior becomes relatively unimportant to the outcome – in this case, has little influence on the estimate of the contamination rate.

We note further that the information about the contamination rate could be combined across samples and sexes in a number of ways. We chose to present the analysis separately for male and female contamination of the different Neandertal extracts. This choice makes it easier to understand the source of information about contamination for each extract; the evidence can then be readily combined. Moreover, treating each sample separately highlights that the experiments performed in different labs on distinct samples yield varying levels of information about contamination.

3.6 Estimates of contamination levels

We consider each of the Neandertal male samples in turn and estimate the probability that the true Neandertal allele has been retrieved at least once, i.e. that the true Neanderthal allele has been seen and not just human contamination. In doing so, we assume that the Neandertal is homozygote for the allele observed at the target site (if he is in fact heterozygote, the probability that we have seen both alleles would be lower). To assess the possibility of contamination of the genomic DNA by European males, we include information from both sites Y2 and Y4 and the X chromosome site; when assessing contamination of the genomic DNA by a European female, we use the only informative control site, which is the X chromosome site. In discussing the evidence about contamination

we ignore the mtDNA assay, and restrict ourselves to discussing the nuclear control assays. Obviously if the mtDNA assay provides reliable information about the contamination level of the nuclear DNA, this increases the confidence in the results.

Neandertal 1:
Assuming the contamination rate is shared over multiplex reactions, then if the source of contamination is European male ($k = 5$ draws of controls), there is a 99.5% probability that we have seen the Neandertal allele at F911 at least once, 96% for target sites recovered twice and 86% for sites recovered only once in the individual (e.g., for F977 and a number of the intron sites). European female contamination of the nuclear DNA cannot be ruled out *a priori*, however, as there are no informative nuclear control loci. As a result, we cannot exclude high levels of contamination and hence the possibility that the target *FOXP2* sites have the derived alleles purely due to human contamination.

If multiplex reactions do not share a contamination rate then, for European male contamination, we can be 97% certain that we have seen the Neandertal F911 allele at least once, 75% sure for F977 allele, and 67% sure for intronic sites. Once again, if European female contamination may have occurred, we cannot exclude the possibility that the derived alleles at *FOXP2* result entirely from human contamination.

The results from this Neandertal individual were also confirmed by two independent laboratories, using a subset of the control markers (recovered a small number of times). Assuming a constant rate of contamination along the genome, their findings greatly strengthen our confidence that the results are not due entirely to modern human European male contamination. However, neither of the other two groups successfully recovered the X chromosome control. (The laboratory in Lyon did initially, but in an extract that showed substantial contamination; a second extract was then analyzed, from which the X4 site was not recovered successfully.) Since no nuclear control assays were performed on this individual that are informative about female contamination, high levels of female nuclear contamination of the genomic DNA of this individual could have gone undetected by all three groups.

We also note that the first extract examined in Lyon suffered from high levels of male contamination (2 out of the 9 products recovered at the nuclear DNA sites where European male contamination would be detectable i.e. Y2, Y4 and X4) yet contamination was not reported for either the mtDNA assay or the blank PCRs performed by Lyon. These findings may cast doubt on the informativeness of these assays in estimating nuclear contamination rates.

Neandertal 2
If the contamination rate is shared over multiplexes then, for European male contamination ($k = 4$), there is a 95% probability that we have seen the true

Neandertal allele at target sites where we have recovered the allele twice (F911 and F977), and an 83% chance if the allele was recovered once. If the contamination source was a European female ($k = 2$ from the X-linked control), then we are 93% certain that we have seen the Neandertal allele at sites recovered twice and 81% certain for an allele recovered once.

If multiplexes cannot be assumed to share a single contamination rate and there is potentially European male contamination, we can be only 75% certain that we have seen the Neandertal allele for sites recovered in multiplex 4. If there is potentially European female contamination, we can only be 71% certain that we have seen the Neandertal allele for sites recovered in multiplex 4.

More generally, our inferences about female contamination are strongly influenced by our assumption that female contamination of a male sample has a greater impact on the X than on the autosomes (as discussed above and quantified in c'). While the average level of contamination for X chromosome is likely to meet this assumption, it is unclear whether the level of contamination at a particular site (i.e., X4) does, as contamination may be highly stochastic. If this assumption is not met for the one site examined by the authors, then the uncertainty about the contamination rate increases, making it even harder to rule out high levels of female nuclear contamination.

3.7 Conclusions regarding the contamination assay

We present a framework for assessing levels of human contamination in the *FOXP2* data. This concern is not a hypothetical one, as the two other laboratories reported European male contamination in initial extracts (detected from Y chromosome controls), suggesting that contamination can occur (even undetected by mtDNA or blank PCR assays, see Section 3.6).

Our conclusion is that if the authors' assumptions (outlined above) are met, then the Y and X chromosome sites are sufficient to rule out with high probability the possibility that the derived alleles seen at the *FOXP2* sites are entirely due to contamination by a European male, for both Neandertal samples. Even if the male contamination rate cannot be assumed to be the same over multiplex reactions (but the contamination rate is shared over sites), it is unlikely that European male contamination can explain the presence of the derived alleles at the two amino acid replacement sites in *FOXP2* in Neandertal, or the derived alleles in Neanderthals at the sites that form the high frequency derived haplotype in humans (as these alleles are seen in both Neandertals).

Whether the high frequency derived haplotype was polymorphic in Neandertals (as stated by the authors) is uncertain however, as the ancestral allele at site S6 is seen only once in individual 2. Establishing that this haplotype is polymorphic in both humans and Neandertals is important, as it strongly influ-

ences which models, and in particular which timings of selection, are consistent with the data (see text). Specifically, if the haplotype is polymorphic in both modern humans and Neandertals, this would provide strong evidence against the most recent selection having occurred 300 Kya, the scenario proposed by the authors.

Only the non-Y assays allow one to rule out female contamination of the nuclear DNA. However, as we have shown, the two autosomal sites are polymorphic in Europeans. Thus, results from the two autosomal sites are uninterpretable and only the site on the X chromosome can be used for this purpose. Since the X assay was unsuccessful in Neandertal 1 (in all three labs), we cannot exclude the possibility that European female nuclear DNA contamination is responsible for the presence of the derived *FOXP2* mutations in this sample. The X-linked site was successfully recovered twice in Neandertal 2, but only in one multiplex reaction in which *FOXP2* alleles were also seen. If contamination rates are shared over multiplexes and sites, we can be reasonably confident (93% per site) that the second Neandertal had the amino-acid replacements, but our certainty about a given high frequency derived allele (usually seen only once) is lower (83% per site). If contamination rates are not shared across multiplexes (but are across sites) then in the presence of potential female contamination, we may only be 75% certain that they have seen the true Neandertal allele at least once in Neandertal 2.

Contamination from nuclear DNA by non-European individuals would be even harder to exclude on the basis of the control assays. Very high levels of contamination by African female nuclear DNA might be impossible to detect, as the individual could potentially contribute the ancestral allele seen at site X4. In turn, high levels of contamination by a non-European (e.g., African) male would also be harder to rule out, as there are fewer draws at the Y chromosome control sites that pertain to African individuals (and the X chromosome site cannot be used, as it is polymorphic).

All of the results presented here rely on the assumption that contamination rates are equal across sites, an assumption the validity of which is not clear. The contamination rate for a given amplification is the proportion of *amplifiable* DNA due to modern human contamination. Even if the amount of contaminant DNA in an extract is the same across genomic regions, the amount of amplifiable Neandertal DNA may well differ. Indeed, it seems plausible that the preservation of amplifiable Neandertal DNA fragments varies along the genome, leading to variation in the contamination rate. In addition, given that the amplification appears to usually start from a single molecule, the process is likely to be highly stochastic, casting doubt on whether the contamination rate is well modeled by a single contamination parameter. If contamination is not uniform amongst sites or is stochastic, a few nuclear control sites (one site, in the case of female contamination) may provide insufficient information.

In summary, if contamination rates are shared across sites, there seems to be sufficient evidence to rule out the possibility that the results stem from European male contamination. In turn, the evidence against female contamination relies on the mtDNA assay of the bone and the X chromosome assay, which was successful in only a single extract. So the overall conclusion rests on at least three assumptions, namely that:

1. The contamination rate is constant across sites.

2. The mtDNA assay (which shows very low contamination of the two bones) is a good proxy for the contamination rate of the nuclear DNA in the extracts.

3. European female contamination of the nuclear DNA of *all* of the extracts is viewed as *a priori* very unlikely. This assumption depends strongly on whether female contamination is independent across extracts, or may have been present from the start and so is shared across extracts.

If these assumptions are not valid, then high levels of contamination cannot be excluded as an explanation for the observation that two Neandertals carry the derived amino-acid alleles at *FOXP2*. We emphasize that the two Neandertals may very well carry the derived amino-acid alleles and that these results simply show that further discussion of the assays and experiments are needed to increase the confidence in the results reported by Krause et al. [5].

The framework presented here to assess contamination rates is undoubtedly too simple and should be extended, but the assumptions match those implicitly made by the authors. Moreover, a more detailed analysis, which factored in more sources of uncertainty, would likely result in less confidence about the contamination rate in the samples. Multiplex methods can now amplify a large number of primers (i.e., markers), and so numerous control markers can be typed simultaneously. Since the uncertainty about contamination rates decreases rapidly with the number of control markers (under the assumption of a constant rate of contamination along the genome), it may be helpful to perform additional experiments. For assessing contamination in male Neandertals, markers for the X chromosome would be of particular use, as heterozygosity of these markers is automatically diagnostic of contamination. We further suggest that statistical statements about confidence in the results should be presented in ancient DNA papers, although even these measures only represent the level of uncertainty under a particular model. Models in no way replace additional controls and independent confirmation, but they provide a framework within which to explore the sensitivity to assumptions of the contamination assays.

Supplementary Table 1 Alleles recovered in Krause et al. [5]. Alleles in gray indicate control alleles that are not informative with regard to the presence or absence of European human contamination. Each line represents a set of primers used to amplify a site. The identifier for each site is shown in the left-most column. Sites are grouped into categories depending on their location. Alleles are specified by either 'A' for ancestral (i.e, matches the chimp allele), 'D' for derived (matches the human allele) or 'N' for the case when the allele recovered corresponded to neither human nor chimpanzee.

	Sidrón 1253 (Neandertal 1)				**Sidrón 1351c (Neandertal 2)**			
Multiplex	1	2	3	4	1	2	3	4
Y nuclear controls:								
Y1	-	A	-	-	-	-	-	-
Y2	A	A	-	-	-	-	-	-
Y3	A	-	-	-	A	-	-	-
Y4	A	A	-	A	A	-	-	A
Y5	A	-	-	A	-	A	A	A
Non-Y nuclear controls:								
X4	-	-	-	-	-	A	-	A
A1 (Chr1)	-	A	A	D	A	-	-	D
A3 (Chr8)	A	A	-	A	A/N	-	-	-
FOXP2 **amino acid substitutions:**								
F911 1	D	-	D	D	-	-	D	D
F911 2	-	D	-	-	-	-	-	-
F977 1	D	-	-	-	-	-	D	D
FOXP2 **intron sites:**								
S1	-	-	D	D	-	-	-	-
S2	-	-	-	D	-	-	D	-
S5	-	-	D	D	-	-	D	D
S6	-	-	-	D	-	-	-	A
S7	-	-	D	D	-	-	-	D
S8	-	-	-	D	-	-	D	-
S9	-	-	-	D	-	-	-	-

Supplementary Table 2 At site A1, the ancestral allele is C and the human reference allele is A. At site A3, the ancestral allele is C and the human reference allele is T. At site X4, the ancestral allele is T and the human reference allele is G. See Section 3.3 for methods.

Chr 1 (A1):

	Num. of Indiv.	AA	AC	CC
YRI	13	11	2	0
CEU	16	13	3	0
ASN	6	5	1	0

Chr 8 (A3):

	Num. of Indiv.	CC	CT	TT
YRI	24	13	11	0
CEU	22	16	6	0
ASN	10	5	5	0

Chr X (X4):

	Num. of Chrom.	G	T
YRI	37	32	5
CEU	36	36	0
ASN	15	15	0

Supplementary Table 3 PCR and Sequencing primers used in the genotyping assay of the Non-Y control sites.

Locus	PCR Primers	
Chr 1 (A1)	N1U1:	CTTAAAGCAGCCCATTTTAATGTT
	N1L1:	TTCATGTTGAAAATTATGGAGGTG
Chr 8 (A3)	N8U1:	CTTTTACCCAAGTCCAGCAACTAT
	N8L2:	AGCTTCCTAGCAATGGAGAC
Chr X (X4)	NXU1:	AAATGACAGAATTTCCTTTTGCAT
	NXL1:	TGGAAGCAGAGAAAGACTAATGTG
Locus	**Sequencing Primers**	
Chr 1 (A1)	N1U3:	CCTTGTAGTGTGTGAAAATC
Chr 8 (A3)	N8U1:	CTTTTACCCAAGTCCAGCAACTAT
	N8L2:	AGCTTCCTAGCAATGGAGAC
Chr X (X4)	NXU2:	GTACACATTAAATAGATTTG

References

[1] Thomson R, Pritchard J, Shen P, Oefner P, Feldman M (2000). Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA 97: 7360–7365.

[2] Hudson R (2007). The variance of coalescent time estimates from DNA sequences. J Mol Evol 64: 702–705.

[3] Enard W, Przeworski M, Fisher S, Lai C, Wiebe V, et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418: 869–872.

[4] Stephens M, Smith N, Donnelly P (2001). A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68: 978–989.

[5] Krause J, Lalueza-Fox C, Orlando L, Enard W, Green R, et al. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. Curr Biol 17: 1908–1912.

[6] Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437: 69–87.

[7] Przeworski M, Coop G, Wall J (2005). The signature of positive selection on standing genetic variation. Evolution 59: 2312–2323.

[8] Zhang J, Webb D, Podlaha O (2002). Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics 162: 1825–1835.

[9] Green R, Krause J, Ptak S, Briggs A, Ronan M, et al. (2006). Analysis of one million base pairs of Neanderthal DNA. Nature 444: 330–336.

[10] Wall J, Kim S (2007). Inconsistencies in Neanderthal genomic DNA sequences. PLoS Genet 3: 1862–1866.

[11] Briggs A, Stenzel U, Johnson P, Green R, Kelso J, et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci USA 104: 14616–14621.

[12] Krause J, Dear P, Pollack J, Slatkin M, Spriggs H, et al. (2006). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature 439: 724–727.

[13] Sanchez J, Endicott P (2006). Developing multiplexed SNP assays with special reference to degraded DNA templates. Nat Protoc 1: 1370–1378.

[14] Jobling M, Tyler-Smith C (2003). The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4: 598–612.

[15] Noonan J, Coop G, Kudaravalli S, Smith D, Krause J, et al. (2006). Sequencing and analysis of Neanderthal genomic DNA. Science 314: 1113–1118.

[16] International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–861.

[17] Levy S, Sutton G, Ng P, Feuk L, Halpern A, et al. (2007). The diploid genome sequence of an individual human. PLoS Biol 5: e254.

[18] Bhangale T, Stephens M, Nickerson D (2006). Automating resequencing-based detection of insertion-deletion polymorphisms. Nat Genet 38: 1457–1462.