The Timing of Selection at the Human *FOXP2* Gene

Graham Coop, Kevin Bullaughey,† Francesca Luca,* and Molly Przeworski**

*Department of Human Genetics, University of Chicago; and †Department of Ecology and Evolution, University of Chicago

Krause J, Lalueza-Fox C, Orlando L, et al. recently examined patterns of genetic variation at *FOXP2* in 2 Neanderthals. This gene is of particular interest because it is involved in speech and language and was previously shown to harbor the signature of recent positive selection. The authors found the same 2 amino acid substitutions in Neanderthals as in modern humans. Assuming that these sites were the targets of selection and no interbreeding between the 2 groups, they concluded that selection at *FOXP2* occurred before the populations split, over 300 thousand years ago. Here, we show that the data are unlikely under this scenario but may instead be consistent with low rates of gene flow between modern humans and Neanderthals. We also collect additional data and introduce a modeling framework to estimate levels of modern human contamination of the Neanderthal samples. We find that, depending on the assumptions, additional control experiments may be needed to rule out contamination at *FOXP2*.

Recent technological advances allow genomic sequence to be retrieved from extinct species, providing answers to evolutionary questions that were previously inaccessible. The ability to recover hominid autosomal DNA, in particular, promises to resolve controversies surrounding possible interbreeding between modern and archaic humans and the timing of modern human adaptations.

In these respects, the *FOXP2* gene is notable, both because of its involvement in speech and language and because previous studies indicated that the protein was under recent positive selection in modern humans, with (the only) 2 amino acid substitutions as putative targets (Enard et al. 2002; Zhang et al. 2002). Genetic variation from extant humans further suggested that the last beneficial fixation occurred in the past 120 thousand years (Ky) (Enard et al. 2002). Recently, however, Krause et al. (2007) reported that 2 Neanderthals also carried the 2 derived amino acid alleles. The authors assumed that modern humans and Neanderthals did not interbreed and so argued that selection must have occurred in the ancestor to the 2 populations, over 300 thousand years ago (Kya) (Noonan et al. 2006), and that the previous dating was incorrect.

Any model for selection that arises from ancient DNA findings must account for patterns of variation in both archaic and modern human samples. Yet 2 features of the modern human variation data are actually quite unlikely under the scenario proposed by the authors. First, the sample of modern humans shows an excess of high-frequency–derived alleles (i.e., a low frequency of ancestral alleles) in an intron near the 2 amino acid substitutions (see fig. 1). This pattern is consistent with the recent fixation of a linked beneficial allele and recombination between selected and neutral loci during the selective sweep (Fay and Wu 2000; Przeworski 2002). But if selection took place in the ancestral population of humans and Neanderthals, the ancestral alleles in the intron would have had to remain at low frequency in humans for over 300 Kya. Moreover, the authors also report that 1 of the intronic sites is polymorphic in the 2 Neanderthals, indicating that the ancestral allele was not lost from the Neanderthal population either. Given that low-frequency variants will tend to be rapidly lost from the population by genetic drift, the persistence of ancestral alleles for close to 300 Kya in both Neanderthal and human lineages is unlikely.

A second feature of the human data is also puzzling. The authors' scenario predicts that the selective sweep was close to completion 300 Kya, yet the selected haplotype appears to have accumulated few mutations since. To demonstrate this, we estimated the time of the most recent common ancestor (tMRCA) of the selected haplotype (see fig. 1), using an approach sometimes called phylogenetic dating (Thomson et al. 2000; Hudson 2007). This method does not make assumptions about demography and selection, but only requires that the mutations in the intron be neutral or nearly neutral. Taking this approach, we obtained a mean tMRCA of 42 Kya (see Supplementary Material online for details). Although there is considerable uncertainty associated with this estimate, it is surprisingly recent if selection took place over 300 Kya (see Supplementary Material online). In other words, the selective scenario proposed by the authors cannot account readily for patterns of variation in modern humans. Given that we have no power to detect a beneficial substitution that occurred over 250 Kya (cf. Sabeti et al. 2006) yet we see a footprint of positive selection at *FOXP2*, the conclusion of a recent selective sweep at *FOXP2* is not surprising.

How then can the data from modern humans and Neanderthals be reconciled? If we assume, as do Krause et al. (2007), that one or both of the 2 amino acid substitutions were the targets of positive selection, then the data may reflect gene flow between modern human and Neanderthal populations. Low levels of admixture, which currently cannot be excluded (Serre et al. 2004; Noonan et al. 2006), could readily explain how the 2 populations share derived *FOXP2* alleles, even though modern human variation data show the footprint of very recent selection. Moreover, assuming that an allele is globally beneficial, low gene flow levels—potentially just a few viable hybrids—could be sufficient for it to spread to another population (Barton and Hewitt 1985). Thus, loci such as *FOXP2* are precisely the regions of the genome where we might expect to see evidence for allele sharing, given low but nonzero admixture between modern humans and Neanderthals.

Alternatively, the recent episode of positive selection in modern humans may not have involved the amino acid sites, in contrast to what has been assumed in previous studies of *FOXP2*. If so, the nonsynonymous substitutions could have occurred at any point since the common ancestor of humans and chimpanzees, either due to positive

Key words: selective sweep, ancient DNA, Neanderthal, *FOXP2*.

E-mail: mfp@uchicago.edu.

Mol. Biol. Evol. 25(7):1257–1259. 2008
doi:10.1093/molbev/msn091
Advance Access publication April 15, 2008

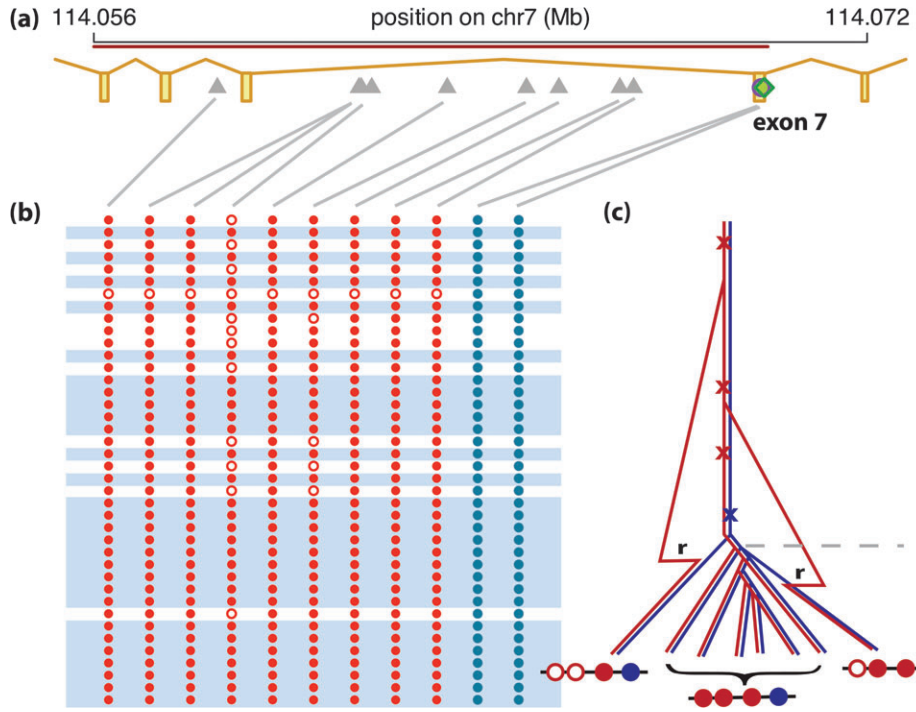

FIG. 1.—The gene model for *FOXP2*, variation data for a modern human sample and genealogical cartoon. (*a*) A subset of the gene model (based on Ensembl 48) for *FOXP2* showing exons 4–8 (tan boxes) with spliced-out introns in between. The region surveyed in Enard et al. (2002) is indicated by the red bar, below which are the polymorphic intronic sites (gray triangles) and the 2 human-specific amino acid substitutions in exon 7 (overlapping purple circle and green diamond). (*b*) Phased haplotypes from Enard et al. (2002), for the 9 sites with high-frequency–derived alleles in the extant human sample. Open red circles denote the ancestral variant, and closed red circles denote the derived variant. Ancestral haplotypes (white) are defined as those carrying the ancestral allele at one or more of the sites with high-frequency–derived alleles. Blue circles indicate the fixed amino acid substitutions. (*c*) A cartoon of the genealogical relationship between individuals in this region of *FOXP2*. A hypothetical genealogy (blue) at the selected site is superimposed on a genealogy (red) of a region linked to the selected site. At the selected site, all lineages coalesce rapidly because of the selective sweep. Recombination events (*r*) have occurred on some lineages of the red genealogy, bringing the selected allele onto additional ancestral backgrounds. The crosses on the genealogy indicate the mutation to the selected allele (blue) and 3 mutations leading to high-frequency–derived alleles (red). The resulting haplotypes are shown at the bottom of the figure. We estimate the tMRCA (*t*) of the selected haplotype by counting the average number of mutations on haplotypes carrying the high-frequency–derived alleles (this variation is not shown in the figure).

selection or to genetic drift. This hypothesis could account for the presence of the derived amino acid alleles in Neanderthals. However, if the high-frequency–derived haplotype found in modern humans is due to recent selection at an (unknown) linked site, its presence in 2 Neanderthals becomes difficult to explain. Although one might invoke a set of selective sweeps at different time points, such a scenario would have to account for all aspects of the data, including the recent tMRCA in modern humans as well as the persistence of the ancestral haplotype in both populations for over 300 Ky.

As in all studies of ancient human DNA, an additional concern in interpreting the data is modern human contamination, which could produce the observed results. In the *FOXP2* study, the authors performed a number of assays to try to ensure that the level of genomic DNA contamination was low (see Supplementary Material online for discussion). They chose 2 bones for which a mitochondrial DNA (mtDNA) assay revealed low levels of modern human mtDNA contamination in their extracts. They then considered a number of control sites in their 2-stage multiplex PCRs, including 2 Y-linked sites to assess contamination by European males (trivially, these are not informative about female contamination). In addition, they assayed 3 non-Y nuclear sites where a previously obtained Neanderthal sequence carried the ancestral allele (see Supplementary Material online). Assuming that modern humans were fixed for derived alleles at these 3 sites, they interpreted the recovery of ancestral alleles from Neanderthal samples as evidence against contamination.

We genotyped these sites and found all 3 to be polymorphic in a human sample (see Supplementary Material online). This finding strongly limits the utility of these sites as controls because the recovery of the ancestral allele no longer establishes that Neanderthal DNA was recovered (see table in Supplementary Material online). This, in turn, reduces the number of interpretable controls performed by Krause et al. (2007): for example, only their mtDNA assay and a single marker on the X chromosome (that only worked in 1 reliable extract) can now be relied on to rule out modern European, female contamination.

The interpretation of the remaining controls depends strongly on a set of implicit assumptions, notably on the relevance of mtDNA contamination assays for nuclear DNA, and whether the amplification of fragments of Neanderthal DNA is uniform among sites and multiplex reactions. We discuss these assumptions in detail in Supplementary Material online, where we present a framework

within which to analyze the control experiments. For the data reported in Krause et al. (2007), we find that, depending on which of the assumptions are met, the controls performed may not be sufficient to rule out contamination. By this, we are not suggesting that Neanderthals do not carry the derived *FOXP2* variants, but simply that additional experiments appear to be needed in order to gain greater confidence in the findings of Krause et al. (2007) and hence in any interpretation.

Acknowledgments

We thank C. Becquet for help with phasing. K.B. was supported by a National Science Foundation graduate research fellowship. This work was supported by National Institutes of Health grants GM79558, GM72861, and HG002772.

Literature Cited

Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. Annu Rev Ecol Syst. 16:113–148.

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S. 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. Nature. 418:869–872.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics. 155:1405–1413.

Hudson RR. 2007. The variance of coalescent time estimates from DNA sequences. J Mol Evol. 64:702–705.

Krause J, Lalueza-Fox C, Orlando L, et al. (13 co-authors). 2007. The derived *FOXP2* variant of modern humans was shared with Neandertals. Curr Biol. 17:1908–1912.

Noonan JP, Coop G, Kudaravalli S, et al. (11 co-authors). 2006. Sequencing and analysis of Neanderthal genomic DNA. Science. 17:1113–1118.

Przeworski M. 2002. The signature of positive selection at randomly chosen loci. Genetics. 160:1179–1189.

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, Palma A, Mikkelsen TS, Altshuler D, Lander ES. 2006. Positive natural selection in the human lineage. Science. 312:1614–1620.

Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Possnert G, Pääbo S. 2004. No evidence of Neandertal mtDNA contribution to early modern humans. PloS Biol. 2:E57.

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA. 97:7360–7365.

Zhang J, Webb DM, Podlaha O. 2002. Accelerated protein evolution and origins of human-specific features: *FOXP2* as an example. Genetics. 162:1825–1835.

Dan Graur, Associate Editor

Accepted April 7, 2008