A Recent Evolutionary Change Affects a Regulatory Element in the Human *FOXP2* Gene

Tomislav Maricic,*[1] Viola Günther,[2] Oleg Georgiev,[2] Sabine Gehre,[1] Marija Ćurlin,[3] Christiane Schreiweis,[1] Ronald Naumann,[4] Hernán A. Burbano,‡[,1] Matthias Meyer,[1] Carles Lalueza-Fox,[5] Marco de la Rasilla,[6] Antonio Rosas,[7] Srećko Gajović,[3] Janet Kelso,[1] Wolfgang Enard,[1] Walter Schaffner,[2] and Svante Pääbo[1]

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[2]Institute of Molecular Life Sciences, University of Zürich, Zürich, Switzerland

[3]Croatian Institute for Brain Research, Zagreb University School of Medicine, Zagreb, Croatia

[4]Transgenic Core Facility, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

[5]Institute of Evolutionary Biology, Consejo Superior de Investigaciones Científicas, Universitat Pompeu Fabra, Barcelona, Spain

[6]Área de Prehistoria, Departamento de Historia, Universidad de Oviedo, Oviedo, Spain

[7]Departamento de Paleobiología, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid, Spain

‡Present address: Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen, Germany

***Corresponding author:** E-mail: tomislav_maricic@eva.mpg.de.

Associate editor: Beth Shapiro

Sequences reported in this article have been deposited in the European Bioinformatics Institute European Nucleotide Archive (accession nos. ERP001291 and ERP000837).

Abstract

The *FOXP2* gene is required for normal development of speech and language. By isolating and sequencing *FOXP2* genomic DNA fragments from a 49,000-year-old Iberian Neandertal and 50 present-day humans, we have identified substitutions in the gene shared by all or nearly all present-day humans but absent or polymorphic in Neandertals. One such substitution is localized in intron 8 and affects a binding site for the transcription factor POU3F2, which is highly conserved among vertebrates. We find that the derived allele of this site is less efficient than the ancestral allele in activating transcription from a reporter construct. The derived allele also binds less POU3F2 dimers than POU3F2 monomers compared with the ancestral allele. Because the substitution in the POU3F2 binding site is likely to alter the regulation of *FOXP2* expression, and because it is localized in a region of the gene associated with a previously described signal of positive selection, it is a plausible candidate for having caused a recent selective sweep in the *FOXP2* gene.

***Key words:* FOXP2, Neandertals, ancient DNA, gene regulation, speech and language.**

Article

Introduction

The inactivation of one *FOXP2* allele results in a severe impairments of language and speech development (Vargha-Khadem et al. 1995; Lai et al. 2001; Fisher and Scharff 2009). *FOXP2* encodes a transcription factor of 715 amino acids that is highly conserved among mammals but carries two amino acid substitutions at positions 303 and 325. These substitutions occurred after the divergence of humans from their common ancestor with chimpanzees some 4–7 Ma (Enard et al. 2002; Enard 2011) and are caused by two nucleotide substitutions located in exon 7 of the gene. Mice carrying the two substitutions in their endogenous *Foxp2* gene show alterations in dopamine levels, striatal synaptic plasticity, and neuronal morphology (Enard et al. 2009; Reimers-Kipping et al. 2011), suggesting that these amino acid changes could have contributed to the evolution of human speech and language by adapting cortico-basal ganglia circuits (Enard 2011; Reimers-Kipping et al. 2011).

Among humans, sequence variation around exon 7 shows an excess of derived nucleotide variants at high frequencies and of rare nucleotide variants, indicating that the region has been affected by a selective sweep (Enard et al. 2002; Zhang et al. 2002; Yu et al. 2009). It has been estimated that this happened within the last 200,000 years (Enard et al. 2002) or 55,000 years (Coop et al. 2008). Because it was initially assumed that at least one of the two amino acid substitutions were the cause of the sweep, it was expected that at least one of them would not be present in Neandertals, who shared a common ancestor with modern humans 370–450,000 years ago (Green et al. 2010). However, both nucleotide substitutions were found in two Neandertals from Spain (Krause et al. 2007) as well as in Neandertals from Croatia (Green et al. 2010), and in Denisovans, an extinct Asian hominin group

related to Neandertals (Reich et al. 2010). Furthermore, it was found that linkage disequilibrium extends across exon 7 in present-day humans, which is not expected if one of the two amino acid changes in exon 7 was the target of selection (Ptak et al. 2009). Hence, although at least one of the two amino acid changes is very likely evolutionarily relevant given the functional data and the conservation of FOXP2, they are not likely to be the cause for the selective sweep. Assuming that a sweep did occur, it must therefore be caused by some other variant in the region, possibly affecting the regulation or splicing of *FOXP2*.

To identify such nucleotide changes, we have used a DNA capture approach to isolate the *FOXP2* gene from a Neandertal. We identify a number of nucleotide differences between this ancient individual and the majority of present-day humans. One of these changes modifies a predicted binding site for the transcription factor POU3F2. We show that this nucleotide change affects the way in which the transcription factor POU3F2 is bound to the site in vitro and that this modifies its ability to drive transcription in cells. The change in this transcription factor binding may be the cause of the positive selective sweep in humans.

Results

Neandertal DNA Capture and Sequencing

We designed an oligonucleotide array containing 60-bp-long probes tiled by 3 bp along the *FOXP2* gene (excluding repetitive elements) from 36,942 nt upstream of the transcription start site to 127,832 nt downstream of the inferred polyadenylation site (NCBI transcript NM_014491.3). We used the array to capture DNA fragments from a library constructed from an approximately 49,000-year-old Neandertal bone fragment from the El Sidrón cave site in Asturias, Spain (Rosas et al. 2006; de Torres et al. 2009). This bone was excavated using methods designed to prevent contamination with present-day human DNA (Fortea et al. 2008), and the DNA was extracted and the library constructed in our clean room facility using procedures to minimize DNA contamination. The DNA adaptors used to construct the sequencing library carry a DNA sequence tag that allows detection of subsequent contamination of the library from other sources (Briggs et al. 2007). This library has been previously used to capture DNA regions showing accelerated evolution in humans, and its extent of contamination with present-day human DNA is estimated to be below 1% based on mitochondrial DNA and autosomal DNA (Burbano et al. 2012). The library was hybridized to a first array, washed, and captured DNA fragments were eluted. After amplification using primers complementary to the library adaptors, a second hybridization to an array carrying the same probes was performed, followed by washing, elution, and amplification (Burbano et al. 2010).

The isolated DNA fragments were sequenced from both ends on an Illumina GAII instrument. Reads from the same fragments were merged (Green et al. 2010) and mapped to the human genome. Fragments with identical end coordinates were fused as they are likely to represent amplification duplicates. The average coverage of positions targeted by the array and covered with at least one probe was 3.9-fold, and 90% of the positions were covered with at least one Neandertal sequence. To improve the coverage of the *FOXP2* region further, we added published data from the genomes of three Croatian Neandertals (Green et al. 2010) and a Denisovan (Reich et al. 2010). This allowed 416 kb or 94% of the targeted region to be covered by at least one read from an archaic human (table 1).

Table 1. Coverage of the *FOXP2* Region by Archaic Human Sequences.

	No. of Bases	Percent
FOXP2 region	443,550	
Human–chimpanzee alignable sequence	441,534	100.0
Captured Neandertal bases	359,612	81.4
Neandertal genome coverage	242,116	54.8
Denisovan genome coverage	314,124	71.1
Total archaic human coverage	416,426	94.3

Substitutions on the Human Lineage

We extracted 1,669 substitutions that likely occurred in and around the *FOXP2* gene on the human evolutionary lineage by identifying positions in the human genome sequence that are different from the orthologous positions in chimpanzee, orangutan, and rhesus macaque. For these positions, we had Neandertal and Denisovan information for 1,601 positions (95.9%). A total of 1,415 of them carried the derived (human-like) allele, and 143 of them carried the ancestral (ape-like) allele (supplementary table S1, Supplementary Material online). One position carried a third base different from both the human genome and the apes. At 42 positions, at least one copy of both the derived and the ancestral allele was seen in the archaic genomes (supplementary table S1, Supplementary Material online). From the latter 42 positions, we excluded 20 positions at which only one ancestral allele carrying a T or an A was seen, whereas one or more derived alleles carrying a C or a G were seen, respectively: such substitutions are likely to be the result of deamination of cytosine residues in the ancient DNA (Briggs et al. 2007; Brotherton et al. 2007). Thus, 165 positions that carry substitutions that occurred recently on the human evolutionary lineage remained for further analysis.

To identify those substitutions that are fixed or at high frequency among present-day humans, we used the array described above to isolate and sequence the *FOXP2* gene from 50 humans, who each come from a different population in the world-wide Human Genome Diversity Panel (HGDP, supplementary table S2, Supplementary Material online) (Cann et al. 2002; Burbano et al. 2010). For positions for which there were data available from 20 or more humans in the panel, we required the derived allele to be at a frequency of 95% or greater. For the remaining positions, for which there were data available from less than 20 individuals, we added allele counts retrieved from the single nucleotide polymorphism database dbSNP (version 132) to the panel

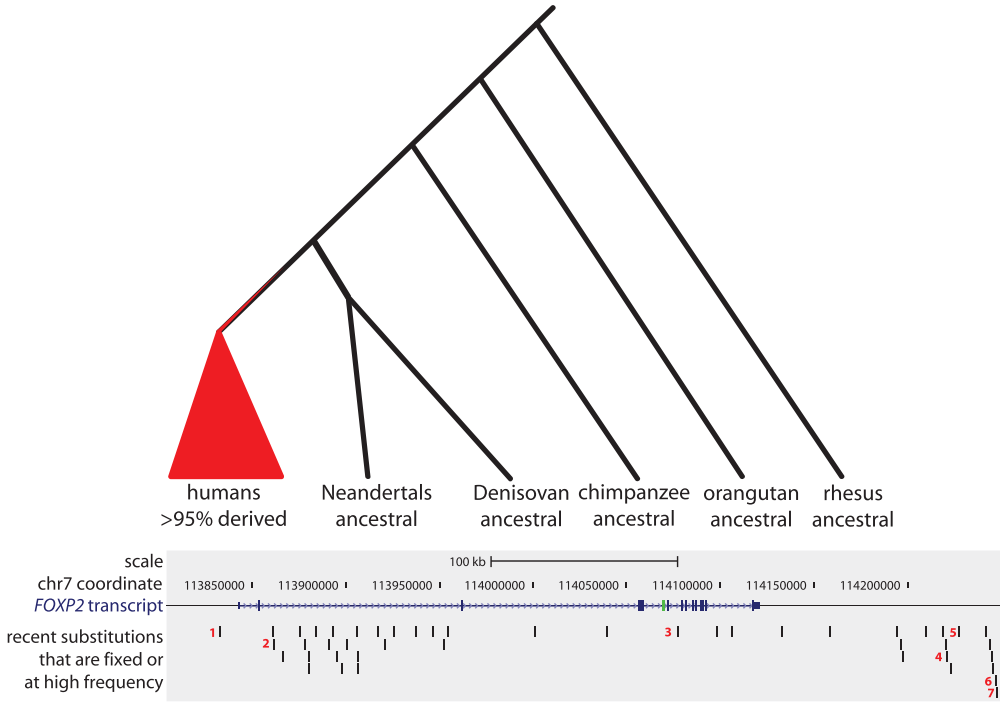

Fig. 1. Recent substitutions that are fixed or at high frequency in present-day humans. Upper panel: schematic illustration of substitutions at high frequency or fixed in humans (red) but ancestral in archaic humans and apes. Lower panel: the *FOXP2* region captured encompassing the *FOXP2* transcript (blue); vertical lines in the transcript indicate exons; exon 7 is green. Among the 46 recent substitutions, those that overlap with highly conserved regions and/or annotated functional regions are indicated by red numbers: 1) PhastCons score: 0.98, DNaseI hypersensitive, ChIP-seq TFs: CTCF, CEBPB, and Rad21; 2) DNaseI hypersensitive; 3) PhastCons score: 1.00; TFBS conserved: POU3F2; 4) TFBS conserved: CDP; 5) DNaseI hypersensitive, ChIP-seq TFs: NFKB, EBF, and BCL11A; 6) ChIP-seq TFs: c-Jun, STAT1, and BAF155; and 7) ChIP-seq TFs: STAT1, BAF155, c-Fos, Max, and BAF170.

data and required the derived allele frequency in the combined data to be 95% or more. One position for which there were no panel data and no record in dbSNP (indicating that the position is not variable in humans) was retained in the further analysis. In total, 46 derived positions (fig. 1, supplementary table S3, Supplementary Material online), which are fixed or at high frequency in present-day humans, were analyzed further.

Exploration of Substitutions

To investigate whether any of the 46 positions are in a functional element, we looked for overlap with conserved regions and/or functional annotation tracks in UCSC genome browser (supplementary tables S3 and S4, Supplementary Material online). Two positions lie within regions that are significantly conserved among vertebrates (Siepel et al. 2005) (fig. 1, lower panel). One of them is located 9,401 bp upstream of the *FOXP2* transcription start site, whereas the other is located 6,963 bp downstream of exon 7 at position 114076877 in intron 8. Both of these positions, as well as five further positions, fall within DNAse hypersensensitive sites and/or binding sites for transcription factors (fig. 1, lower panel). Of the additional five sites, one is located in intron 2, whereas the remaining four are located downstream of the polyadenylation sites.

Of these seven positions, position 114076877 in intron 8 is located in the region where a putative selective sweep was detected (Enard et al. 2002). Specifically, the derived base appears on the putatively selected haplotype, whereas the ancestral base appears on the nonselected haplotype (data in [Ptak et al. 2009]). As this makes the position a likely cause of the sweep, we investigated it further.

Position 114076877 falls at the third position of a 14-bp-long inferred binding motif of the transcription factor POU3F2 (supplementary fig. S1, Supplementary Material online) (Rhee et al. 1998). A 1.4-kb-region around this site is conserved among tetrapods, for example, it is 90% identical in a frog (fig. 2, upper panel). Position 114076877 carries an adenine residue in all tetrapod animals and in all three DNA fragments retrieved from the El Sidrón Neandertal, both chromosomes from a Denisovan (Meyer et al. 2012) and another Neandertal individual from Siberia (unpublished data). However, in the majority of present-day humans (98%), this is replaced by a thymine residue (fig. 2, lower panel). Thus, this position was conserved among animals during more than 700 My (Kumar and Hedges 1998) but has changed recently in humans.

POU3F2 Binding

To test whether, and how, the POU3F2 protein binds to the two allelic variants of the putative binding site, we synthesized two pairs of complementary oligonucleotides that carry the ancestral and derived nucleotide, respectively. Nuclear extracts were isolated from HeLa cells transfected with a

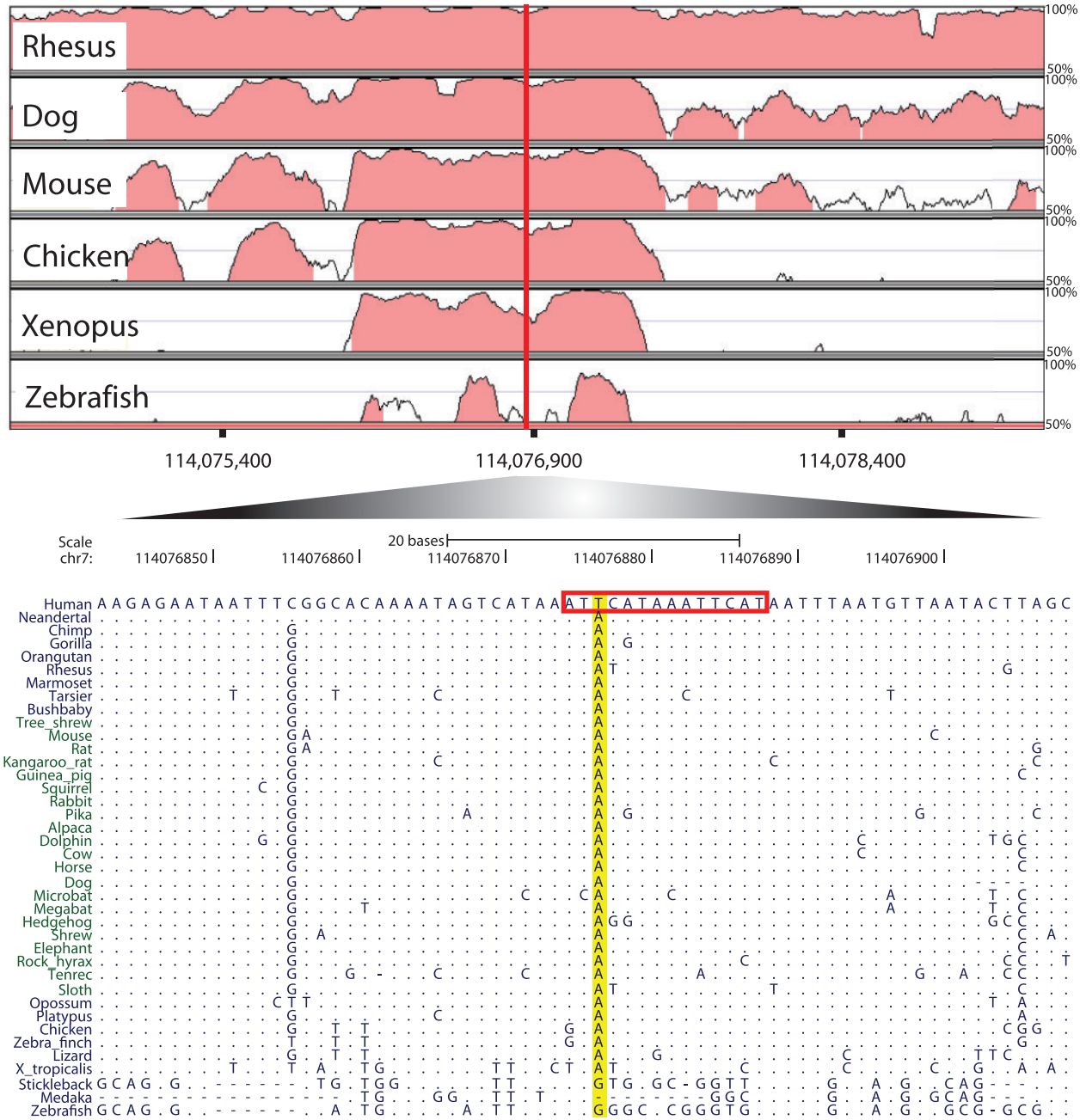

Fig. 2. Strong evolutionary conservation of the POU3F2 binding site. Upper panel: VISTA browser (Frazer et al. 2004) window where the *Y*-axis represents percent sequence similarity among human and other species. The red line indicates the position of the POU3F2 binding site. Lower panel: UCSC *multiz* alignment (Blanchette et al. 2004). The red rectangle highlights the POU3F2 binding site, and the position derived in humans is outlined in yellow. Dots indicate bases identical to the human, dashes absence of bases in the alignment.

POU3F2 expression plasmid (Schreiber et al. 1989) and incubated with one of the two double-stranded radioactively labeled oligonucleotides and subsequently subjected to acrylamide gel electrophoresis under nonreducing conditions (Radtke et al. 1993). As described (Rhee et al. 1998), we find that POU3F2 binds both as a monomer and dimer to the binding site (fig. 3). However, across seven experiments where the amount of POU3F2 was varied, the ratio of dimer to monomer bound by the derived allele was less (59%, range: 50–66%) than the ratio bound by the ancestral allele (table 2). Thus, in each experiment performed, the derived allele binds less of the dimeric form relative to the monomeric form of POU3F2 than does the ancestral allele.

Transcriptional Activation

To test to what extent the two allelic variants of the POU3F2 binding sites are able to drive transcription, we generated three pairs of reporter plasmids that each carried either the derived or ancestral allele of the POU3F2 binding site in various configurations relative to a β-globin reporter gene (fig. 4, left panel): 1) two copies of the POU3F2 binding site variant upstream of the reporter gene, 2) one copy of the binding site

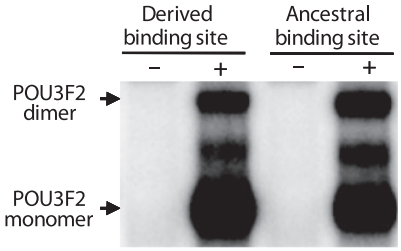

Fig. 3. POU3F2 binding to the allelic variants of the POU3F2 binding site. Migration positions of POU3F2 monomers and dimers are indicated, +/— refers to presence and absence of transfection of the *POU3F2* expression vector, respectively. The weak band between the dimer and monomer bands might be partly misfolded POU3F2 bound to the oligonucleotide.

Table 2. Ratios of Monomer and Dimer Band Photo Stimulated Luminescence Obtained by the EMSAs.

	POU3F2 Nuclear Extract (µg)	Derived Dimer/ Monomer	Ancestral Dimer/ Monomer	(Derived Dimer/Monomer)/ (Ancestral Dimer/Monomer)
EXP 1	10	0.15	0.28	0.53
EXP 2	10	0.26	0.51	0.50
EXP 3	8	0.10	0.17	0.63
EXP 4	4	0.06	0.09	0.66
EXP 5	8	0.15	0.23	0.64
EXP 6	16	0.25	0.39	0.63
EXP 7	30	0.44	0.78	0.56
			Average	0.59

upstream and four copies downstream of the reporter gene, and 3) one copy upstream and an SV40 enhancer downstream of the reporter gene. These plasmids were separately transfected into HeLa cells together with a reference plasmid and a POU3F2 expression plasmid if indicated.

Two days after transfection, RNA was extracted from the cells, and reporter and reference gene expression was assayed by an S1 nuclease protection assay (Weaver and Weissmann 1979). The first two plasmid pairs failed to show any appreciable transcriptional activity (fig. 4, right panel). In contrast, both the ancestral and derived alleles caused reporter gene expression in the plasmid that contained the SV40 enhancer downstream of the reporter gene. As this expression was dependent on coexpression of POU3F2 in the cells (fig. 4, upper-right panel), it is most likely mediated by POU3F2 binding to the binding sites. Across three experiments, cells transfected with the derived allele accumulated less (66%, range: 63–69%, t-test $P = 0.003$) of the reporter mRNA than cells transfected with the ancestral allele (fig. 4, lower-right panel), indicating that the derived allele is less efficient than the ancestral allele in driving transcription from the reporter gene.

Transgenic Mouse Enhancer Assay

To test whether the 1.8-kb-long conserved region surrounding the POU3F2 binding site can drive transcription during mouse development, we produced transgenic mice carrying either of the two variants of the conserved region in front of a minimal promoter and a lacZ reporter gene (supplementary fig. S2, Supplementary Material online) (Visel et al. 2007). We analyzed embryos and animals for expression of the reporter gene in brain, heart, esophagus, and lungs at embryonic day E11.5 and at postpartal days P4, P28, and P60. At neither time point did we detect any reproducible staining although a positive control vector produced expected results (supplementary table S5, Supplementary Material online). Thus, in the construct tested, the conserved region does not act as an enhancer in the mouse at the developmental stages tested.

Discussion

Several lines of evidence suggest that the substitution at position 114076877 in intron 8 of the *FOXP2* gene, which differs between the majority of present-day humans and Neandertals and Denisovans, affects a regulatory DNA element. First, it is located in a sequence that is conserved over 700 My of vertebrate evolution. Second, it falls at the third position of a 14-bp-sequence motif identified as a binding site for the transcription factor POU3F2 by in vitro selection experiments (Rhee et al. 1998). Third, there is another putative POU3F2 binding site 78 bp downstream of position 114076877 that is conserved among mammals (Kent et al. 2002), consistent with the observation that transcription factor binding sites that serve as transcriptional enhancers often occur in close proximity to other homotypic transcription factor binding sites (Gotea et al. 2010). Fourth, we show that the DNA sequence where the substitution occurs binds POU3F2 in vitro (fig. 3). Fifth, we show that this sequence in conjunction with an SV40 enhancer can drive transcription in cells in tissue culture (fig. 4).

Given the evidence that the substitution falls within a regulatory element, the fact that we could not detect any enhancer activity using a reporter construct in transgenic mice is likely to reflect that the human DNA sequences tested do not work in the mouse background or at the developmental stages tested. It is also possible that the element requires cooperation with sequences outside the fragment tested as it requires an associated SV40 enhancer in the transcriptional activation assay. Further work is therefore required to address where and when this element is physiologically active in humans.

It is likely that the putative regulatory activity of the POU3F2 site will affect the *FOXP2* gene rather than other genes on chromosome 7 because the protein CTCF has been shown to bind to DNA regions upstream and downstream of the *FOXP2* gene (supplementary fig. S3, Supplementary Material online) (Bao et al. 2008). CTCF is a hallmark of "insulators," cis-acting DNA elements that limit the effects of enhancers and silencers of transcription along a chromosome (Raab and Kamakaka 2010). Thus, any regulatory activity at the POU3F2 site is unlikely to have effects on genes other than *FOXP2*.

The POU family of transcription factors (to which POU3F2 belongs) tends to bind to their target sequences in the form of homodimers or monomers (Rhee et al. 1998). It is therefore

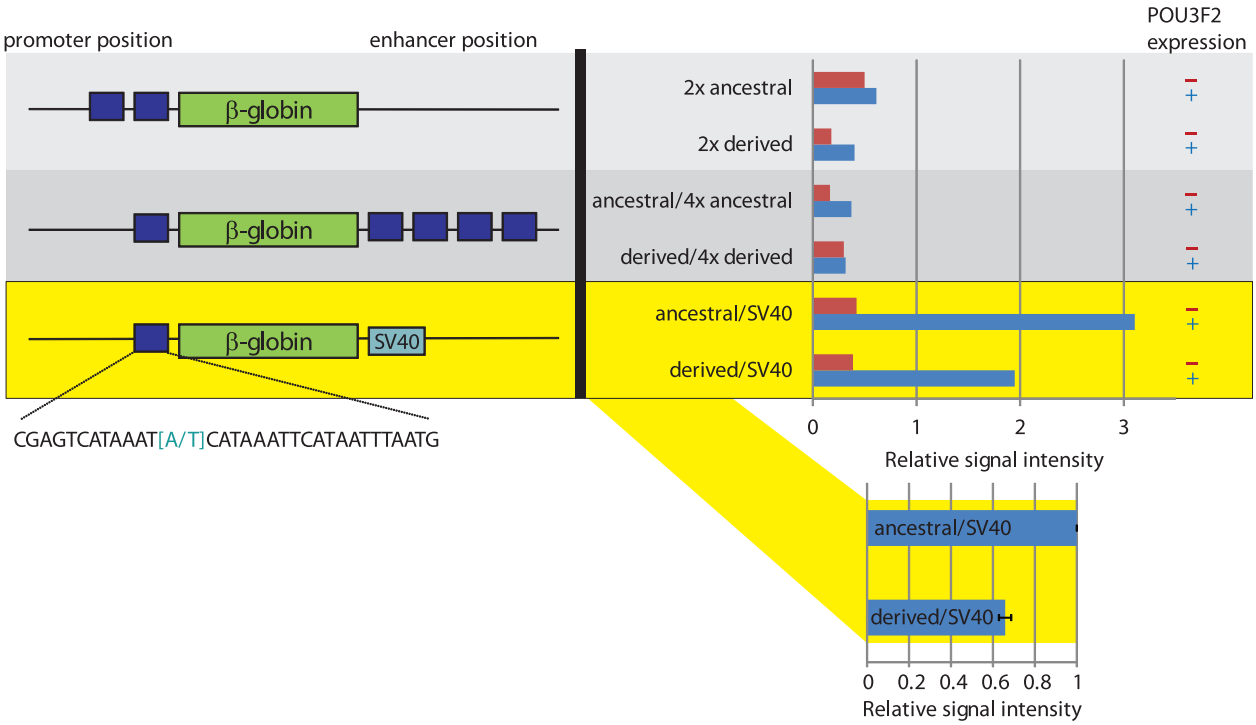

Fig. 4. Reporter gene transcription driven by the POU3F2 binding sites. Left panel: three reporter constructs used for the S1 nuclease protection assay. Blue boxes represent POU3F2 binding sites. Upper-right panel: Relative signal intensities of β-globin transcripts for each construct given in the left panel. Red bars are without (−) and blue bars with (+) POU3F2 expression plasmid cotransfection. Lower-right panel: Relative reporter transcript amounts generated by the two allelic variants of the construct containing the SV40 enhancer are shown. The signal of the ancestral binding site in each independent experiment was set to one, and the standard deviation of three experiments was calculated for the derived binding site. In the three experiments, different cell culture dishes were transfected on different days to isolate RNA.

interesting that the derived transcription factor binding site binds less of the POU3F2 dimer than the ancestral allele (fig. 3), as this is likely to affect its ability to drive transcription. This is borne out by the tissue culture experiments where the derived allele drives less reporter gene expression than the ancestral allele to an extent that quantitatively matches the binding of POU3F2 dimers relative to POU3F2 monomers to the two alleles. Interestingly, in another experimental system, the first two nucleotides in the consensus binding site have been shown to affect binding of the POU-domain protein dimers (Rhee et al. 1998). A fuller understanding of how POU3F2 dimers and monomers affect transcription awaits further studies.

The transcription factor POU3F2 is expressed exclusively in the central nervous system (Schreiber et al. 1993), more specifically in postmitotic neurons and glia (Hagino-Yamagishi et al. 1997). Within the central nervous system, *FOXP2* is expressed in postmitotic neurons (Ferland et al. 2003). Thus, it is reasonable to assume that POU3F2 regulates expression of *FOXP2* in neurons. It is furthermore interesting that position 114076877 is located at the point in intron 8 of the *FOXP2* gene where the pattern of allele frequencies among humans indicates that a functional change occurred that could be responsible for a positive selective sweep affecting the *FOXP2* gene during the last 50,000 years (Coop et al. 2008). It is noteworthy that this is the only nucleotide variant in that region where the majority of present-day people carry a derived variant that is not present in Neandertals and Denisovans. Thus, it is possible that this change was positively selected recently during the evolution of fully modern humans.

Because the loss of one functional copy of *FOXP2* in humans affects language and speech (Vargha-Khadem et al. 1995) and because POU3F2 is expressed only in the central nervous system, it is tempting to speculate that the substitution at position 114076877 in intron 8 of the *FOXP2* gene was involved in the evolution of modern language. However, other work (Enard et al. 2009; Reimers-Kipping et al. 2011) has shown that two amino acid substitutions in the FOXP2 protein that are shared with archaic humans (Krause et al. 2007; Green et al. 2010; Reich et al. 2010) affect synaptic plasticity and dendritic trees in cortico-basal ganglia circuits when introduced into mice. Thus, one may speculate that these earlier coding changes that occurred before the divergence of archaic and modern humans were followed by later regulatory changes that were unique to modern humans.

The ancestral allele occurs at frequencies of ~10% in some African populations (supplementary table S6, Supplementary Material online). Therefore, individuals homozygous for the ancestral allele can be expected to occur at a frequency of approximately 1% in the population. In such individuals, the phenotype of the ancestral allele should be observable even if is recessive to the derived allele. Further work will explore the phenotypes of such homozygous carriers of the ancestral allele and the consequences of the substitution at position 114076877 on *FOXP2* transcription in model systems.

Materials and Methods

HGDP Library Capture, Sequencing, Processing, and Mapping

Probes were designed (Burbano et al. 2010) on an Agilent 244 k array to cover a 444 kb long region (chr7:113,805,346-114,248,895, fig. 1, lower panel) centered on exon 7 and encompassing the entire *FOXP2*. Indexed DNA libraries were made as described (Burbano et al. 2010) from 50 humans, each from a different population in the HGDP (supplementary table S2, Supplementary Material online). One array was used to capture the *FOXP2* region, and captured products were sequenced on the Illumina GAII platform (v4 chemistry) as a multiplexed, paired-end run of 2×76 cycles with additional 7 cycles for reading the index sequence. Base calling was done using *Ibis* (Kircher et al. 2009) and reads assigned to individuals based on their index sequence. Sequences are deposited in European Bioinformatics Institute European Nucleotide Archive (accession number ERP001291; http://trace.ddbj.nig.ac.jp/DRASearch/study?acc=ERP001291, last accessed December 6, 2012). If possible, reads from the same DNA fragments were merged (Burbano et al. 2010) and treated as single-end reads, whereas nonmerged reads were treated as pair-end reads when mapped to *hg18* with *bwa* (Li and Durbin 2009). Duplicate reads were removed with *SAMtools* (Li et al. 2009). Only sequences with a *Phred*-scaled mapping quality 30 or higher were kept for analysis. The average read depth of captured positions varied from 4.0- to 7.8-fold (supplementary fig. S4, Supplementary Material online).

Neandertal Sequencing

Neandertal DNA was extracted from a total of 1.4 g of bone (SD1253) as described previously (Rohland and Hofreiter 2007). Uracil residues were enzymatically removed, and 454 sequencing libraries were prepared (Briggs et al. 2010), amplified (Burbano et al. 2010), pooled, amplified, and 16.8 µg used for two consecutive captures on identical Agilent 244 k arrays. Captured products were converted to an Illumina library by amplification with tailed primers (Burbano et al. 2010) and sequenced as above except that sequencing primers that allow the clean room "key" to be read were used (Burbano et al. 2010). Sequences are deposited in European Bioinformatics Institute European Nucleotide Archive (accession number ERP000837; http://trace.ddbj.nig.ac.jp/DRASearch/study?acc=ERP000837. last accessed December 6, 2012). Reads were merged (Green et al. 2010) and mapped as above. Sequences with the same start and end coordinates (amplification duplicates) were collapsed into a single consensus sequence. The base with the highest sum of quality scores was called at each position, and when more than one base had the highest score, one of them was chosen randomly. Only sequences with a mapping quality of 30 or higher were kept for further analysis.

Cell Culture and Transfections

HeLa cells were maintained in DMEM (Gibco, Life Technologies) supplemented with 8% fetal bovine serum (Biochrom AG), 100 U/ml penicillin, 100 µg/ml streptomycin, and 2 mM L-glutamine (Gibco, Life Technologies). Cells were transfected with DNA using the calcium phosphate method, washed 14–16 h after the addition of precipitate, and harvested for further experiments the following day.

Preparation of Nuclear Extracts and Electrophoretic Mobility Shift Assay

Nuclear extracts from transiently transfected HeLa cells were prepared as described in Schreiber et al. (1989). Binding reactions with ^{32}P end-labeled double-stranded oligonucleotides and gel conditions are described in Schreiber et al. (1988). Oligonucleotides used were the following (complementary bases are in capital letters):

Derived binding site:
5′-CGAGTCATAAATTCATAAATTCATAATTTAATG-3′
and 5′-tcgaCATTAAATTATGAATTTATGAATTTATGACT
CGagct-3′;

Ancestral binding site:
5′-CGAGTCATAAATACATAAATTCATAATTTAATG-3′
and 5′-tcgaCATTAAATTATGAATTTATGTATTTATGACT
CGagct-3′.

Plasmid Constructions and S1 Nuclease Protection Assay

The POU3F2 expression vector (Schreiber et al. 1993) and the S1-nuclease protection assay reference plasmid (Westin et al. 1987) were described previously. The reporter plasmids were obtained by fusing respective copies of the annealed oligonucleotides containing ancestral or derived POU3F2 binding sites (as described in "Preparation of nuclear extracts and electrophoretic mobility shift assay") upstream or downstream of the β-globin gene of the OVEC-vector (Westin et al. 1987). Reporter constructs with the SV40 enhancer inserted downstream of the transcription unit were described previously (Seipel et al. 1992). Detailed cloning strategies are available on request.

HeLa cells were transfected with 10 µg of the respective reporter and 3 µg of reference plasmid as an internal control and 1 µg of POU3F2 expression plasmid if indicated. Two days after transfection, RNA was isolated and analyzed by the S1 nuclease protection assay as described previously (Weaver and Weissmann 1979; Westin et al. 1987). Signals were visualized and quantified using the FLA-700 analyzer and ImageGauge software (Fujifilm Life Science). Reporter signals were normalized to reference signals.

Transgenic Mouse Enhancer Assay

Polymerase chain reaction (PCR) primers were designed to amplify a DNA region of 1,829 bp that is conserved from humans to *Xenopus* and surrounds the POU3F2 binding site (supplementary fig. S2, Supplementary Material online): F: 5′-TTTGTCTTCCTGAAGTGCTTGC, R: 5′-GCTTCATATTGC

TGTGTTTCCG. PCR was performed on DNA from two Nigerian individuals carrying the derived and the ancestral as well as the derived versions of the POU3F2 binding site, respectively. PCR products were purified and inserted into pENTR/D-TOPO vectors (Invitrogen) and sequenced. Two inserts that differed only at the POU3F2 binding site were transferred into the HsP68-LacZ vector (supplementary fig. S2, Supplementary Material online) via the LR recombination reaction (Invitrogen). Vectors were purified (Qiagen EndoFree plasmid kit) and injected into pronuclei of C57Bl/6/J mouse embryos, which were then transferred to pseudopregnant females. Embryo harvesting at E11.5, genotyping, and LacZ staining was performed as described (Visel et al. 2007). Transgenic mice number, developmental stages, and stained tissues are reported in supplementary table S5, Supplementary Material online. As a positive control, transgenic mouse embryos were produced that carry a previously described DNA construct (Prabhakar et al. 2008); in those, expression patterns similar to those reported (Prabhakar et al. 2008) were detected (data not shown).

Supplementary Material

Supplementary tables S1–S6 and figures S1–S4 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org).

Acknowledgments

The authors are grateful to Nadav Ahituv for providing the Hsp68-lacZ-Gateway vector. They thank Anja Heinze, Jesse Dabney, Johannes Krause, and Adrian Briggs for help with producing DNA libraries, Martin Kircher and Udo Stenzel for data processing, the Sequencing group for DNA sequencing, Lee-Ann Wood for help with UCSC custom tracks analysis, and Aida Andrés, Jeff Jensen, Susan Ptak, Daniel Falush, Sven-Holger Puppel, and the entire Department of Evolutionary Genetics for helpful discussion. This work was supported by the Max Planck Society, the Swiss National Science Foundation, and the Kanton Zürich. El Sidrón research was supported by several Principado de Asturias-Universidad de Oviedo Grants. C.L.-F. was supported by the Ministerio de Ciencia e Innovacion, Spain (grant number BFU2009-06974), and A.R. was supported by the Spanish Ministry of Science (grant number CGL2009-09013).

References

Bao L, Zhou M, Cui Y. 2008. CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators. *Nucleic Acids Res.* 36:D83–D87.

Blanchette M, Kent WJ, Riemer C, et al. (12 co-authors). 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* 14:708–715.

Briggs AW, Stenzel U, Johnson PL, et al. (11 co-authors). 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci U S A.* 104:14616–14621.

Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Paabo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* 38:e87.

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, Cooper A. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res.* 35:5717–5728.

Burbano HA, Green RE, Maricic T, et al. (10 co-authors). 2012. Analysis of human accelerated DNA regions using archaic hominin genomes. *PloS One* 7:e32877.

Burbano HA, Hodges E, Green RE, et al. (20 co-authors). 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 328:723–725.

Cann HM, de Toma C, Cazes L, et al. (41 co-authors). 2002. A human genome diversity cell line panel. *Science* 296:261–262.

Coop G, Bullaughey K, Luca F, Przeworski M. 2008. The timing of selection at the human FOXP2 gene. *Mol Biol Evol.* 25:1257–1259.

de Torres T, Ortiz JE, Grün R, et al. (19 co-authors). 2009. Dating of the hominid (Homo Neanderthalensis) remains accumulation from El Sidrón cave (Piloña, Asturias, North Spain): an example of a multimethodological approach to the dating of upper Pleistocene sites. *Archaeometry* 52:680–705.

Enard W. 2011. FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr Opin Neurobiol.* 21:415–424.

Enard W, Gehre S, Hammerschmidt K, et al. (56 co-authors). 2009. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137:961–971.

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.

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA. 2003. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *J Comp Neurol.* 460:266–279.

Fisher SE, Scharff C. 2009. FOXP2 as a molecular window into speech and language. *Trends Genet.* 25:166–177.

Fortea J, de la Rasilla M, Garcia-Tabernero A, Gigli E, Rosas A, Lalueza-Fox C. 2008. Excavation protocol of bone remains for Neandertal DNA analysis in El Sidron Cave (Asturias, Spain). *J Hum Evol.* 55:353–357.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32:W273–W279.

Gotea V, Visel A, Westlund JM, Nobrega MA, Pennacchio LA, Ovcharenko I. 2010. Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.* 20:565–577.

Green RE, Krause J, Briggs AW, et al. (56 co-authors). 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.

Hagino-Yamagishi K, Saijoh Y, Ikeda M, Ichikawa M, Minamikawa-Tachino R, Hamada H. 1997. Predominant expression of Brn-2 in the postmitotic neurons of the developing mouse neocortex. *Brain Res.* 752:261–268.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res.* 12:996–1006.

Kircher M, Stenzel U, Kelso J. 2009. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* 10:R83.

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.

Kumar S, Hedges SB. 1998. A molecular timescale for vertebrate evolution. *Nature* 392:917–920.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–523.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, et al. (10 co-authors). 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.

Meyer M, Kircher M, Gansauge MT, et al. (34 co-authors). 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–226.

Prabhakar S, Visel A, Akiyama JA, et al. (13 co-authors). 2008. Human-specific gain of function in a developmental enhancer. *Science* 321:1346–1350.

Ptak SE, Enard W, Wiebe V, Hellmann I, Krause J, Lachmann M, Paabo S. 2009. Linkage disequilibrium extends across putative selected sites in FOXP2. *Mol Biol Evol.* 26:2181–2184.

Raab JR, Kamakaka RT. 2010. Insulators and promoters: closer than we think. *Nat Rev Genet.* 11:439–446.

Radtke F, Heuchel R, Georgiev O, Hergersberg M, Gariglio M, Dembic Z, Schaffner W. 1993. Cloned transcription factor MTF-1 activates the mouse metallothionein I promoter. *EMBO J.* 12:1355–1362.

Reich D, Green RE, Kircher M, et al. (28 co-authors). 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468:1053–1060.

Reimers-Kipping S, Hevers W, Paabo S, Enard W. 2011. Humanized Foxp2 specifically affects cortico-basal ganglia circuits. *Neuroscience* 175:75–84.

Rhee JM, Gruber CA, Brodie TB, Trieu M, Turner EE. 1998. Highly cooperative homodimerization is a conserved property of neural POU proteins. *J Biol Chem.* 273:34196–34205.

Rohland N, Hofreiter M. 2007. Comparison and optimization of ancient DNA extraction. *Biotechniques* 42:343–352.

Rosas A, Martinez-Maza C, Bastir M, et al. (18 co-authors). 2006. Paleobiology and comparative morphology of a late Neandertal sample from El Sidron, Asturias, Spain. *Proc Natl Acad Sci U S A.* 103:19266–19271.

Schreiber E, Matthias P, Muller MM, Schaffner W. 1988. Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). *EMBO J.* 7:4221–4229.

Schreiber E, Matthias P, Muller MM, Schaffner W. 1989. Rapid detection of octamer binding proteins with "mini-extracts", prepared from a small number of cells. *Nucleic Acids Res.* 17:6419.

Schreiber E, Tobler A, Malipiero U, Schaffner W, Fontana A. 1993. cDNA cloning of human N-Oct3, a nervous-system specific POU domain transcription factor binding to the octamer DNA motif. *Nucleic Acids Res.* 21:253–258.

Seipel K, Georgiev O, Schaffner W. 1992. Different activation domains stimulate transcription from remote ("enhancer") and proximal ('promoter") positions. *EMBO J.* 11:4961–4968.

Siepel A, Bejerano G, Pedersen JS, et al. (16 co-authors). 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15:1034–1050.

Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R. 1995. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc Natl Acad Sci U S A.* 92:930–933.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA. 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 35:D88–D92.

Weaver RF, Weissmann C. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15 S beta-globin mRNA precursor and mature 10 s beta-globin mRNA have identical map coordinates. *Nucleic Acids Res.* 7: 1175–1193.

Westin G, Gerster T, Muller MM, Schaffner G, Schaffner W. 1987. OVEC, a versatile system to study transcription in mammalian cells and cell-free extracts. *Nucleic Acids Res.* 15:6787–6798.

Yu F, Keinan A, Chen H, Ferland RJ, Hill RS, Mignault AA, Walsh CA, Reich D. 2009. Detecting natural selection by empirical comparison to random regions of the genome. *Hum Mol Genet.* 18: 4853–4867.

Zhang J, Webb DM, Podlaha O. 2002. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* 162:1825–1835.