A Human-Specific Gene in Microglia

Toshiyuki Hayakawa,[1,2*] Takashi Angata,[1,2*†] Amanda L. Lewis,[1,3] Tarjei S. Mikkelsen,[6] Nissi M. Varki,[1,4] Ajit Varki[1,2,5‡]

Recent studies have shown differences between humans and chimpanzees in sialic acid (Sia) biology, including Sia-binding receptors called Siglecs (*1*–*3*). Analysis of the ~3-kb region of homology between human *SIGLEC11* (*hSIGLEC11*) and pseudogene *hSIGLECP16* (Fig. 1A) showed that an ~2-kb segment (designated A/A′) including the first five exons is 99.3% identical. However, the rest of this region (designated B/B′) has a much lower identity (94.6%). A phylogenetic tree of B/B′ from the human (h) and chimpanzee (c) genomes shows the topology expected from the gene orthology (Fig. 1B). In contrast, the A/A′ tree shows a within-species clustering of *hSIGLEC11* and *hSIGLECP16* (Fig. 1C), suggesting a recent gene conversion event.

Pairwise genetic distances between human and chimpanzee A/A′ regions were calculated, after excluding intron 2 and exons (*4*) (fig. S1), and compared with the standard human-chimpanzee genetic distance, calculated from downstream intronic regions (0.0141 ± 0.0013). The genetic distance between *hSIGLEC11* and *hSIGLECP16* is much smaller (0.0049 ± 0.0024), and that between corresponding chimpanzee forms is much larger (0.0337 ± 0.0065). Thus, although the presumed gene duplication event yielding ancestral *SIGLEC11* and *SIGLECP16* predated the human-chimpanzee common ancestor, the subsequent gene conversion occurred only in the human lineage. Both *hSIGLEC11* and *hSIGLECP16* are more closely related to *cSIGLECP16* than to *cSIGLEC11*, and the genetic distance between *hSIGLEC11* and *cSIGLECP16* is nearly the same as that between *SIGLECP16* orthologs (*4*). Thus, *hSIGLECP16* converted *hSIGLEC11*. Inclusion of bonobo, gorilla, and orangutan *SIGLEC11* sequences confirmed that this gene conversion occurred only in the human lineage (*4*) (fig. S2). A search of human single-nucleotide polymorphism databases (*4*) suggests that it is universal to modern humans.

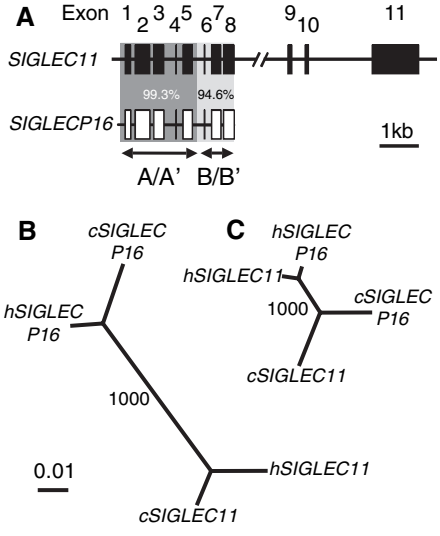

Fig. 1. Gene structure and phylogenetic analyses. (**A**) Comparison of *SIGLEC11* and the similar region of *SIGLECP16*. Exons are represented by solid and open boxes. (**B** and **C**) Phylogenetic relationships of (B) B/B′ and (C) A/A′. The label at the internode represents bootstrap support for 1000 replications.

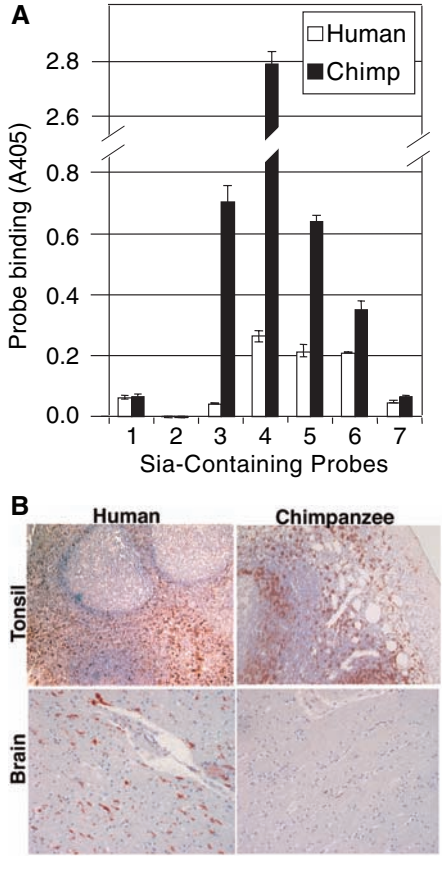

Fig. 2. Sia-recognition and expression patterns (*4*). (**A**) Binding of recombinant human and chimpanzee Siglec-11 to Sia-containing probes: 1, Neu5Acα2-3Lac; 2, Neu5Acα2-6Lac; 3, Neu5Acα2−; 4,Neu5Gcα2−; 5, $(Neu5Ac\alpha2\text{-}8)_2$; 6, $(Neu5Ac\alpha2\text{-}8)_3$; and 7, $(Neu5Ac\alpha2\text{-}8)_{5\text{-}6}$. Error bars indicate standard deviation. A405, absorbance at 405 nm. (**B**) Examples of human and chimpanzee brain and tonsil tissue probed with antibody to human Siglec-11 (brown staining).

The converted region includes a 5′ upstream region and exons encoding the Sia recognition domain. We therefore studied the Sia-binding properties of recombinant, soluble Siglec-11 protein. Chimpanzee Siglec-11 showed more robust binding than hSiglec-11 (Fig. 2A), especially to Neu5Gc, the Sia type missing in humans (*3*). Human Siglec-11 still bound oligosialic acids [$(Neu5Ac\alpha2\text{-}8)_{2\text{-}3}$], which are enriched in the brain.

Human brain cortex microglia showed strong Siglec-11 staining (*5*) in eight individuals. However, despite positive macrophage staining in other tissues, microglia showed only occasional staining in five chimpanzees (Fig. 2B) and none in two orangutans. Thus, brain microglia gained specific and prominent Siglec-11 expression in the human lineage, possibly due to changes of regulatory sequences in the 5′ upstream region.

This work also supports the notion that pseudogenes are better called "potogenes" or potential genes (*6*). We suggest that this human-specific gene conversion event may be related to the evolution of genus *Homo*. Multiple approaches could shed light on its evolutionary significance, including further analysis of the converted region, promoter studies, calculation of the event timing, detection of Siglec-11 ligands in the brain, and a search for humans with *SIGLEC11* mutations.

References and Notes

1. A. Varki, *Nat. Genet.* **36**, 1034 (2004).
2. Chimpanzee Sequencing and Analysis Consortium, *Nature*, in press.
3. A. Varki, T. Angata, *Glycobiology*, in press.
4. Materials and methods are available as supporting material on *Science Online*.
5. T. Angata *et al.*, *J. Biol. Chem.* **277**, 24466 (2002).
6. E. S. Balakirev, F. J. Ayala, *Annu. Rev. Genet.* **37**, 123 (2003).
7. Supported by NIH grant no. R01GM32373 (A.V.), a Japan Society for the Promotion of Science Fellowship for Research Abroad (T.H.), and by the Mathers Foundation (A.V.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5741/1693/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Table S1
References and Notes

2 May 2005; accepted 8 August 2005
10.1126/science.1114321

[1]Glycobiology Research and Training Center, [2]Department of Cellular and Molecular Medicine, [3]Department of Biological Sciences, [4]Department of Pathology, [5]Department of Medicine, University of California at San Diego, La Jolla, CA 92093, USA. [6]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02141, USA.

*These authors contributed equally to this work.
†Present Address: Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan.
‡To whom correspondence should be addressed. E-mail: a1varki@ucsd.edu

Hayakawa et al.

Supporting Online Materials

Materials and Methods

Extraction of chimpanzee *SIGLEC11* and *SIGLECP16* genomic sequences. Reciprocal best BLASTZ alignments with human *SIGLEC11* and *SIGLECP16* genomic sequences were performed to extract chimpanzee orthologous regions from the draft sequence (NCBI Build 1 Version 1) of the chimpanzee genome. The sequence qualities of extracted chimpanzee sequences were examined using Phred quality scores for each site. All unique sites and all sites clustering chimpanzee sequences in the homologous regions (see Fig. 1A) were checked, and of these sites (125 sites), 98, 13, 8, 4, 2 and 0 sites showed 50, 40, 30, 20, 10 and 0 Phred quality score, respectively. Most of sites (123 of 125) showed high-quality Phred score (>20). It was therefore concluded that extracted chimpanzee sequences are reliable. The chimpanzee *SIGLECP16* sequence has stop codons in the exons of B/B' region, indicating that chimpanzee *SIGLECP16* is a pseudogene. Since these stop codons are conserved in human *SIGLECP16*, the ancestral *SIGLECP16* was considered to be a pseudogene.

Analyses of proposed gene conversion. Using nucleotide sequence alignments (details to be published elsewhere), genetic distances were calculated with multiple-hit corrections (1). We included the genomic region upstream of exon 1 in the distance calculation. The genomic region upstream of exon 1 includes potential regulatory regions and its evolutionary rate is not clear. However, it is reasonable to think that most of these sites should be under evolutionary neutrality, as only a few positions in regulatory sequence should be essential for its function and under functional constraint. Phylogenetic trees were constructed using the neighbor-joining method (2) and reliability was assessed by bootstrap values with 1,000 replications. Relative rate

tests were performed by using MEGA2 (3).

Analysis of ligand binding specificity of human and chimpanzee Siglec-11. Expression and purification of human Siglec-11-Fc protein (recombinant fusion protein of the first 3 Ig-like domains of human Siglec-11 and human IgG Fc domain) was previously described (4). Chimpanzee Siglec-11 cDNA was cloned by RT-PCR from chimpanzee spleen total RNA using the same primer set used to clone human Siglec-11 cDNA (4). The cDNA segment coding for the first 3 Ig-like domains was amplified by PCR from sequence-verified chimpanzee Siglec-11 cDNA as template using the following primer set: cSL-11 5'Expr (5'-CCCTCTAGAGCCACCATGCTGCTGTTACCCCTGTTG-3'; Xba I site underlined) and SL-11 Chi 3'(3D) (5'-ATCCATCACTCTCAGGTTCTCTGGAGGA-3'). The PCR product was digested with Xba I and subcloned into Xba I – Eco RV site of EK-Fc/pcDNA3.1 (4). Chimpanzee Siglec-11-Fc protein was expressed and purified in the same manner as human Siglec-11-Fc (4). The Sia-binding assay using human and chimpanzee Siglec-11-Fc protein was performed as described (5, 6), using biotinylated polyacrylamide probes multiply substituted with sialylated glycans (Glycotech, Rockville, MD).

Siglec-11 immunostaining. Paraffin sections were deparaffinized, rehydrated, treated to remove endogenous peroxidases, endogenous biotin and non-specific binding to extracellular matrix, and stained with anti-Siglec11 antibody 4C4 and mouse IgG (X63 supernatant from mouse myeloma cell line; negative control) following established protocols (4). All sections showed good histological quality in parallel hematoxylin & eosin stains. Anti-CD68 and anti-vimentin were used as positive controls to further ensure the quality of the sections. Nuclei were counterstained with Mayer's hematoxylin (Sigma, St.Louis, MO) and the slides were mounted for viewing and analysis (Fisher).

Sequencing of great ape *SIGLEC11* sequences. Genomic regions encompassing A/A' and B/B' of *SIGLEC11* were amplified by PCR from genomic DNAs of bonobo, gorilla and orangutan. Purification of PCR products was performed using a QIAquick PCR purification kit (Qiagen) and QIAquick gel extraction kit (Qiagen). The purified PCR products were sequenced after cloning using TOPO TA Cloning Kits (Invitrogen), or were directly sequenced. The PCR primers and sequencing primers are listed in table S1. The obtained nucleotide sequences of bonobo, gorilla and orangutan *SIGLEC11* will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB211392-AB211394.

Supporting Text

Direction of gene conversion between *SIGLEC11* and *SIGLECP16* in human lineage. There are two possible directions of gene conversion: either *SIGLEC11* converted *SIGLECP16* (*SIGLEC11->SIGLECP16*); or, *SIGLECP16* converted *SIGLEC11* (*SIGLECP16->SIGLEC11*). In the former case, the sequence of *SIGLEC11* should not be changed. In the latter case, *SIGLECP16* would introduce new sequences into *SIGLEC11*. We determined the direction of gene conversion by examining the pair-wise genetic distances of converted regions. Since the gene conversion between *SIGLEC11* and *SIGLECP16* occurred only in the human lineage, knowing which chimpanzee sequence is more closely related to human *SIGLEC11* and *SIGLECP16* should indicate the direction of gene conversion. One chimpanzee sequence should provide the same or nearly same distance as the human-chimpanzee standard distance in every interspecies comparison, and another one should provide the same or nearly same distance as the distance between chimpanzee sequences in every interspecies comparison (see fig. S1). Thus, the human ortholog of the former chimpanzee sequence should be a donor of sequences in the gene conversion.

The A/A' region can be divided into two even more highly similar parts separated by a small divergent part including intron 2, and containing 9 of the 15 differences between human *SIGLEC11* and *SIGLECP16* (details to be reported elsewhere). The upstream part of this divergent region is designated A/A'1 and shows 99.6% identity; and the downstream part is designated A/A'2, showing a 99.8% identity. In contrast, the small part separating A/A'1 and A/A'2 shows 94.7% identity, similar to that in B/B', between human *SIGLEC11* and *SIGLECP16* (details to be reported elsewhere). Thus, if gene conversion occurred, the small divergent portion was likely not involved. To examine the gene conversion direction, the pair-wise genetic

distances were calculated using region A/A' after excluding intron 2 and exons. The genetic distances between human *SIGLEC11*/*SIGLECP16* and chimpanzee *SIGLEC11* are 0.0312+/-0.0063 and 0.0337+/-0.0065, nearly the same as the distance (0.0337+/-0.0065) between chimpanzee *SIGLEC11* and *SIGLECP16*. The genetic distance between human *SIGLEC11* and chimpanzee *SIGLECP16* and the distance between human and chimpanzee *SIGLECP16* are 0.0173+/-0.0046 and 0.0198+/-0.0050, similar to the human-chimpanzee standard distance (0.0141+/-0.0013) (details to be published elsewhere). This relationship of genetic distances supports a *SIGLECP16->SIGLEC11* gene conversion in the region A/A'. As mentioned above, region A/A' actually has two parts (A/A'1 and A/A'2) that underwent two independent (likely related) gene conversions. Separate genetic distance calculations for A/A'1 and A/A'2 show the same distance relationship as obtained using the combined A/A' (details to be reported elsewhere).

We also sequenced *SIGLEC11* from other great apes (bonobo, gorilla and orangutan), and used non-coding portions of the region A/A' to determine whether *SIGLECP16->SIGLEC11* gene conversions occurred in these lineages. Phylogenetic analysis of A/A' (A/A'1 + A/A'2) showed that bonobo, gorilla and orangutan *SIGLEC11* are more closely related to chimpanzee *SIGLEC11* than to chimpanzee *SIGLECP16* (see fig. S2). Detailed analyses of A/A'1 + A/A'2 will be reported elsewhere. Overall, the obtained results indicate that the *SIGLECP16->SIGLEC11* gene conversion in region A/A' occurred only in the human lineage.

Gene Conversion Event Appears Universal to Modern Humans. Analysis of informative sites in this region of the current Human SNP database showed no evidence for polymorphisms of gene conversion (data not shown). Taken together with the finding of Siglec-11 expression in all human brains studied, the gene conversion event is likely universal to modern humans.

Perspectives and Future Directions. Based on all the above data, and on timing of the conversion event (details to be published elsewhere), we suggest that the human-specific gene conversion was related to the human genetic loss of expression of the sialic acid Neu5Gc (7), and to the evolution of genus *Homo*, in which dramatic changes in brain size and function occurred (8). Unfortunately, mice do not have a Siglec-11 ortholog, and there is no practically feasible and ethically acceptable way to test this hypothesis in great apes. Thus, a search for humans with *SIGLEC11* mutations (assuming no effect on postnatal viability) is most likely to shed light on the evolutionary significance of this human-specific gene conversion event. We are also currently analyzing further details regarding the gene conversion event, determining the timing of the event during human evolution, studying the putative promoter region of *SIGLEC11* that was altered during the gene conversion process, and exploring the presence of specific ligands for Siglec-11 in the human brain.

It is also notable that the cytosolic tail of Siglec-11 can recruit the protein-tyrosine phosphatase SHP-1 (Src homology domain 2-containing phosphatase 1), an intracellular regulator of many signaling pathways (4). Notably, SHP-1 deficient mice are known to show a marked decrease in the number of microglial cells (9), suggesting that SHP-1 is essential to the maintenance of microglia in the central nervous system. Thus, the human-specific gain of Siglec-11 expression could have influenced microglial numbers and/or trophic functions, and contributed to human brain evolution. Interestingly, SHP-1deficient mice also have a slightly smaller brain size than littermate controls (9). Finally, microglia play prominent roles in human brain diseases such as Alzheimer disease, HIV-1-associated dementia and multiple sclerosis (10). Characterization of this human-specific microglial change may help explain why the pathology of such diseases is common among humans, but appears to be rare in great apes (11).

Supporting References and Notes

1. T. H. Jukes, C. R. Cantor, H. N. Munro, Ed. (Academic Press, New York, 1969), pp. 21-132.
2. N. Saitou, M. Nei, *Mol Biol Evol* **4**, 406 (1987).
3. S. Kumar, K. Tamura, I. B. Jakobsen, M. Nei, *Bioinformatics* **17**, 1244 (2001).
4. T. Angata *et al.*, *J Biol Chem* **277**, 24466 (2002).
5. N. Patel *et al.*, *J Biol Chem* **274**, 22729 (1999).
6. T. Angata, A. Varki, *J Biol Chem* **275**, 22127 (2000).
7. H. H. Chou et al., *Proc Natl Acad Sci U S A* **99**, 11736 (2002).
8. B. Wood, M. Collard, *Science* **284**, 65 (1999).
9. C. A. Wishcamper, J. D. Coffin, D. I. Lurie, *J Comp Neurol* **441**, 118 (2001).
10. A. Minagar *et al.*, *J Neurol Sci* **202**, 13 (2002).
11. M. V. Olson, A. Varki, *Nat Rev Genet* 4, 20 (2003).

Supporting Figure Legends

Fig. S1. Hypothetical phylogenetic trees related to gene conversion events. The gene duplication event yielding ancestral *SIGLEC11* and *SIGLECP16* predated the human-chimpanzee split. T1 and T2 correspond to the human-chimpanzee divergence time and the *SIGLEC11-SIGLECP16* divergence time, respectively. S11, *SIGLEC11*; P16, *SIGLECP16*; Hsa, *Homo sapiens*; Ptr, *Pan troglodytes*. **A.** No gene conversion. **B.** Gene conversion of *SIGLECP16* by *SIGLEC11* in the human lineage. **C.** Gene conversion of *SIGLEC11* by *SIGLECP16* in the human lineage.

Fig. S2. Phylogenetic relationships of regions A/A' of human, chimpanzee, bonobo, gorilla and orangutan *SIGLEC11* and human and chimpanzee *SIGLECP16*. The non-coding parts were used for tree construction after elimination of intron 2 (rationale for eliminating intron 2 is that it was apparently excluded from the gene conversion event, details to be published elsewhere). Hsa, *Homo sapiens*; Ptr, *Pan troglodytes*; Ppa, *Pan paniscus*; Ggo, *Gorilla gorilla*; Ppy, *Pongo pygmaeus*.

Table S1. Sequences of PCR and sequencing primers

Name	Sequence
SL-XI5'UTR	5'-GGGACAGGCCCAGCCCCAGAGCCC-3'
Sig11ExtrDn	5'-TTCTCCCTGAACCTCAGCCC-3'
3'RACE-1	5'-GGTGATCAGAGACGCGCAGAGGG-3'
3'RACE-2	5'-CGGTCATCTGTGTGTTTAACTGGG-3'
SeqA1	5'-CCTCCCTGAAGCCAGCTCAC-3'
SeqA2	5'-GGGCTGGAGCTGCGTGGG-3'
SeqA2c	5'-ACCCTGGGGCTGGAGCTGCG-3'
SeqA3	5'-AGAGCCGGTCCCCTGTCTCC-3'
SeqA4	5'-CTGCACAGAGAGGTCCAGGG-3'
XIChi(2D)3'	5'-CATCACTCTCAGGTTCTCTGGAGGA-3'
CS115'-2	5'-TGTGTTCTTGTAGACTCTTCCAGTG-3'
CS115'-3	5'-CCAGTTGTCCTTTCCCAGTGAAGTC-3'
S11Seq-2	5'-AGCAGCATCTCTGGGCTCTG-3'

Sig11ExtrDn and CS115'-3 were used to amplify specifically *SIGLEC11* sequence in genomic PCR.

Figure S1. Hayakawa et al.

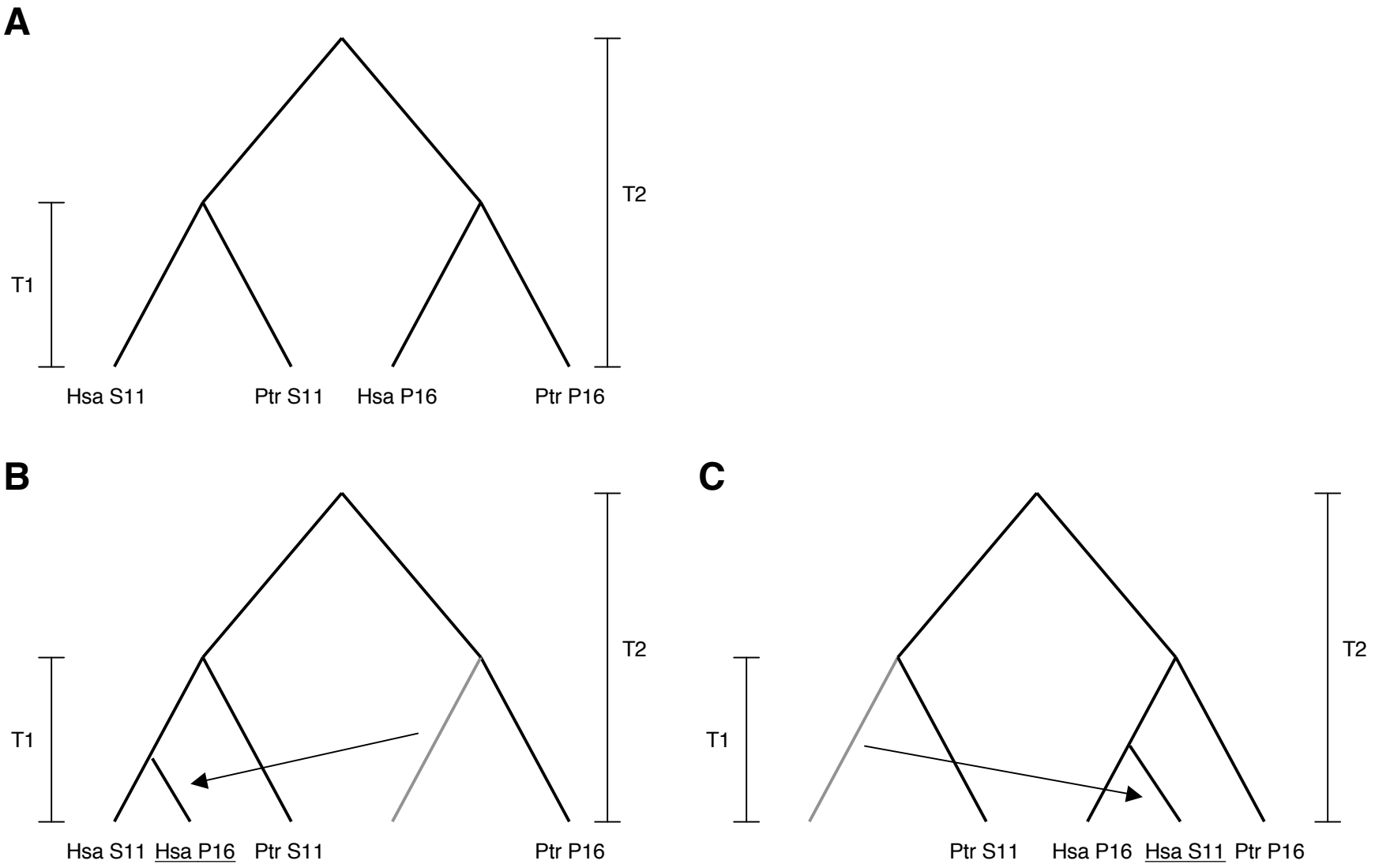

Figure S2. Hayakawa et al.