Inactivation of *MOXD2* and *S100A15A* by Exon Deletion during Human Evolution

Yoonsoo Hahn,[*,1] *Sangkyun Jeong,*[†,2] *and Byungkook Lee*[*]

*Laboratory of Molecular Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD; and †Laboratory of Mammalian Genes and Development, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

We devised a bioinformatics method for systematic identification of putative human-specific exon-deletion mutations that occurred after the divergence of human and chimpanzee and experimentally verified 2 of the predicted mutations in *MOXD2* and *S100A15A* genes. *MOXD2* gene encodes a monooxygenase that is highly conserved in mammals and is mostly expressed in the olfactory epithelium in mouse. The presence of a deletion of the last 2 exons and a polymorphic nonsense mutation in exon 6 suggests that *MOXD2* gene is inactive in humans. S100A15A is a member of the S100 family of calcium-binding proteins, the mouse ortholog of which is expressed during epidermal maturation. Human *S100A15A* gene is likely to be inactive because the start codon–bearing exon is deleted in human. We propose that modification or inactivation of *MOXD2* and *S100A15A* genes have contributed to the loss of certain smell sense in humans and to the development of human skin.

Introduction

Humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*) diverged from a common ancestor around 6 MYA (Glazko and Nei 2003; Kumar et al. 2005; Patterson et al. 2006). Since then, humans have acquired various phenotypic traits distinct from other great apes (reviewed in Varki and Altheide 2005). Many of these traits represent reduction or loss of great ape features, such as small jaws and teeth, weakened jaw muscles, decreased smell sensitivity, and reduced body hair. Some of these diminished features have been proposed to be associated with a gain of beneficial traits in humans. For example, smooth hairless skin was suggested to be evolved for enhanced thermoregulation (Folk and Semken 1991) and/or for reduction of parasite loads (Pagel and Bodmer 2003). The human-specific 2-bp deletion in the coding region of *MYH16* gene resulting in frameshift and premature termination was suggested to have caused the reduction of jaw muscle and allowed humans to have bigger brains (Stedman et al. 2004). These examples accord with the "less-is-more" model, according to which the gene loss is one of the mechanisms for acquisition of novel phenotypic traits during evolution (Olson 1999; Olson and Varki 2003).

Release of the complete human genome sequence (International Human Genome Sequencing Consortium 2004) and the draft chimpanzee genome sequence (The Chimpanzee Sequencing and Analysis Consortium 2005) enables one to directly compare the 2 genome sequence data for a systematic identification of human-specific genetic changes (Pollard et al. 2006 and reviewed in Li and Saunders 2005; Varki and Altheide 2005; Kehrer-Sawatzki and Cooper 2007). Several strategies were employed to systematically identify genes of which coding region had been disrupted in the human lineage after the divergence of humans and chimpanzees. For examples, human-specific frameshift mutations (Hahn and Lee 2005) and nonsense mutations (Hahn and Lee 2006) were identified, some of which were proposed to lead to a loss or modification of function of the affected genes. Analysis of a nonprocessed human pseudogene data set also yielded a large number of genes that seemed to be specifically pseudogenized in humans (Wang et al. 2006). In this study, we developed a bioinformatics procedure for prediction of exon-deletion mutations in the human genome that might have occurred after the human–chimpanzee split. We then experimentally verified the presence of the deletions in 2 highly plausible candidates, *MOXD2* and *S100A15A* genes. We collected sequence data of orthologs and homologs for each gene for molecular evolutionary analyses. We also discuss possible phenotypic changes that might have been caused by the loss of these genes during human evolution.

Materials and Methods
Data Sets

We downloaded human genome versus chimpanzee genome alignment data, mRNA versus chimpanzee genome alignment data, and other genome annotation data from the University of California Santa Cruz (UCSC) Genome Browser Database (http://genome.ucsc.edu/) (Hinrichs et al. 2006) as of July 2006. Both of the human and the chimpanzee genomes were March 2006 assemblies. For the selection of nonhuman vertebrate mRNA sequences, we used the National Center for Biotechnology Information (NCBI) Taxonomy Browser database (http://www.ncbi.nlm.nih.gov/Taxonomy/).

Bioinformatics Prediction of Human-Specific Exon-Deletion Mutations

First, we collected chimpanzee-only genomic fragments by filtering the database table panTro2.netHg18. The chimpanzee-only regions that fulfilled the following conditions were selected: the region does not align with a syntenic region of the human genome; it is longer than or equal to 100 bp in length; it does not match a nonsyntenic region of the human genome; and the syntenic region of the

[1] Present address: Department of Life Science, College of Natural Sciences, Chung-Ang University, Seoul, South Korea
[2] Present address: Center for Development and Differentiation, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea

Key words: human, chimpanzee, exon deletion, gene inactivation.

E-mail: bk@nih.gov.

human genome does not contain a sequencing gap. We obtained 16,524 chimpanzee-only fragments that meet all of these criteria. These fragments may include chimpanzee-specific insertions as well as human-specific deletions.

Next, we collected only those fragments that possibly contain an exon by inspecting whether they overlap with any nonhuman vertebrate mRNA sequence. For the alignment data of the nonhuman and nonchimpanzee vertebrate RefSeqs and mRNAs to the chimpanzee genome, we used the database tables panTro2.xenoRefSeqAli and panTro2.xenoMrna. There were 819 chimpanzee genome fragments that were absent in the human genome and aligned with a nonhuman vertebrate mRNA by at least 10 bp. By using the coding region data of mRNAs derived from the database table hg18.cds, we observed that 33 of these matched a coding region of a nonhuman gene.

As the final step, we manually scrutinized the 33 candidates to collect highly plausible instances. We excluded the cases showing low sequence quality in genome, short or poor alignment between the genome sequences, uncertain orthology between the 2 species, or chimpanzee-specific insertion. In some cases, which were also excluded, the collected chimpanzee-only fragment matched human mRNAs or expressed sequence tags (ESTs), raising the possibility of a missassembly of the human genome, false alignment of the 2 genomes, or deletion polymorphism among human population.

Genomic DNA Samples and Polymerase Chain Reaction

Human and nonhuman primate genomic DNA samples were purchased from the Coriell Cell Repositories (http://ccr.coriell.org/). The repository numbers are as follows Mbuti Pygmy, NA10492; Basque, NA15884; Icelandic, NA15756; Northern European, NA17003; Southern Chinese, NA11321; Surui, NA10970; chimpanzee, NG06939; bonobo (*Pan paniscus*), NG05253; gorilla (*Gorilla gorilla*), NG05251; Sumatran orangutan (*Pongo pygmaeus abelii*), NG12256; pigtailed macaque (*Macaca nemestrina*), NG08452; and common woolly monkey (*Lagothrix lagotricha*), NG05356. We performed genomic polymerase chain reaction (PCR) by using aforementioned DNA samples in order to detect the predicted deletions of *MOXD2* and *S100A15A* genes. PCR primer sequences are available in supplementary table 1 (Supplementary Material online). The primers used in this study were synthesized by Invitrogen (Carlsbad, CA).

Sequence Analysis of *MOXD2* and *S100A15A* Genes

We determined full coding sequences of *MOXD2* and *S100A15A* genes from gorilla and orangutan genomes by direct sequencing of amplified exons. Primer sequences used for amplification of exons are presented in supplementary table 1 (Supplementary Material online). The PCR products were sequenced by Macrogen USA (Rockville, MD). The sequences have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/GenBank/) under accession numbers EF044235 (gorilla *MOXD2*), EF044236 (Sumatran orangutan *MOXD2*), EF044237 (gorilla *S100A15A*), and EF044238 (Sumatran orangutan *S100A15A*). We also collected mammalian orthologs of *MOXD2* and *S100A15A* by deducing coding sequences from the genome assemblies available in the UCSC Genome Browser Database, mRNAs or genomic sequences in GenBank, and the whole-genome shotgun (WGS) clone sequences from the NCBI Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/). We carried out Blast searches of the nonredundant protein database at NCBI web site (http://www.ncbi.nlm.nih.gov/BLAST/) to gather homologous proteins of MOXD2 and S100A15A, respectively. Presence of a putative signal peptide, a glycosylphosphatidylinositol (GPI) anchor signal, and functional domains of the proteins were inferred by using SignalP (http://www.cbs.dtu.dk/services/SignalP/) (Bendtsen et al. 2004), DGPI (http://129.194.185.165/dgpi/), and Pfam (http://pfam.janelia.org/) (Finn et al. 2006), respectively. A multiple alignment of orthologous and homologous protein sequences was prepared using ClustalW (Thompson et al. 1994) and CHROMA (Goodstadt and Ponting 2001). A phylogenetic analysis was performed by using MEGA3 (Kumar et al. 2004). All the accession numbers analyzed in this study are given in supplementary table 2 (Supplementary Material online).

Molecular Evolutionary Analyses

The estimated nonsynonymous/synonymous mutation ratios ($\omega = d_N/d_S$) were computed using the maximum likelihood method for codons (CODEML program in the PAML package, v3.15 obtained from http://abacus.gene.ucl.ac.uk/software/paml.html) (Yang 1997). We used the χ^2 test on the log likelihood difference to decide if the differences in ω values between different lineages and different models were significant (Yang 1998; Stedman et al. 2004). The probabilities were computed by using the "Chi Square Calculator" (available at http://www.stat.tamu.edu/~west/applets/chisqdemo.html). Coding sequences of *MOXD2* and *S100A15A* genes from human, chimpanzee, gorilla, orangutan, rhesus macaque, and marmoset were aligned by using ClustalW (Thompson et al. 1994). The region that is deleted in human and the stop codon positions have been removed from the alignment prior to analyses.

Results and Discussion
Identification of Putative Human-Specific Exon-Deletion Mutations

We developed a procedure for identification of putative human-specific exon-deletion mutation candidates as described in the Materials and Methods. We manually inspected 33 candidates to narrow them to 6 highly plausible cases, which we named after the mouse orthologs as *ADAM5*, *CMAH*, *MOXD2*, *MSR2*, *NLRP10*, and *S100A15A*. The exon deletion and subsequent inactivation of the *CMAH* gene has been previously reported (Chou et al. 1998). In the cases of *ADAM5* and *MSR2*, the chimpanzee orthologs also seem to be defective due to frameshift and/or nonsense mutations. The human NLRP10 (also known as NALP10) protein which is involved in inflammation (Tschopp et al. 2003) exhibits altered C-terminus when compared with putative chimpanzee ortholog due to
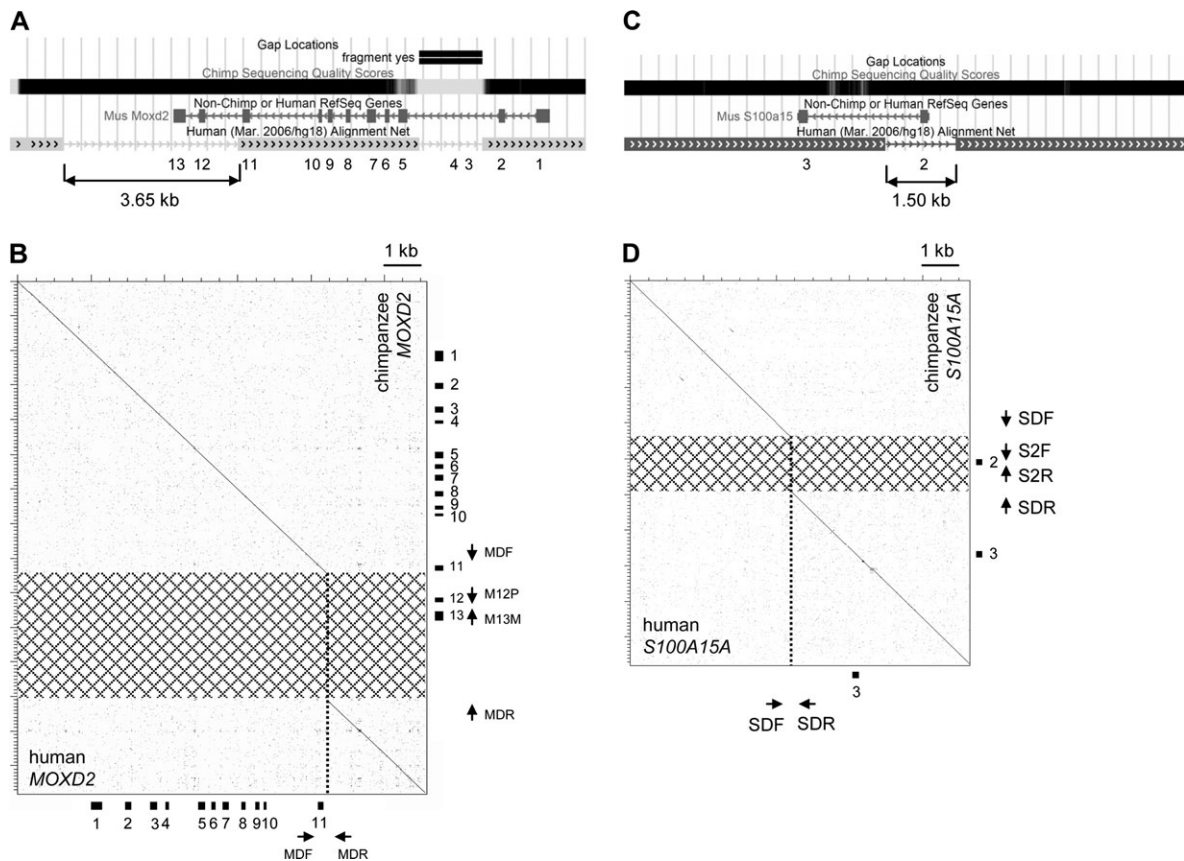
Fig. 1.—(A) The human-specific exon deletion in the *MOXD2* gene. A part of the annotated chimpanzee genome available at the UCSC Genome Browser Database is shown. The chimpanzee chromosome track is depicted in shades of black, representing the quality scores of bases. The light gray region is a sequencing gap, which spans exons 3 and 4 of chimpanzee *MOXD2*. The bottom track indicates the human chromosome 7 that corresponds to the orthologous chimpanzee chromosome: thick, aligned; thin, unaligned due to either deletion in the human chromosome or sequencing gap in 1 of the 2 sequences. The smaller boxes between the human and the chimpanzee chromosomes indicate exons of mouse *Moxd2* mRNA that match the chimpanzee ortholog. The exons of *MOXD2*/*Moxd2* are numbered below the human chromosome track: 1–13 in reverse direction. The predicted 3.65-kb deletion in the human chromosome, which includes exons 12 and 13, is indicated at the bottom (double-ended arrow). (B) A dot-plot comparison of human *MOXD2* and chimpanzee *MOXD2* genomic sequences. The sequencing gap spanning exons 3 and 4 in the current chimpanzee genome assembly was filled using WGS clone sequences obtained at the NCBI Trace Archive. The deleted segment in the human genome is crosshatched. Exons identified based on mouse *Moxd2* gene are marked at the right and bottom. Locations of the PCR primers used for experimental verification are indicated (arrows). (C) The human-specific exon deletion in the *S100A15A* gene. The human *S100A15A* gene in the chromosome 1 (the bottom track) has a 1.50-kb deletion that removes the start codon–containing exon 2 in the orthologous chimpanzee gene. The location of the putative noncoding exon 1 was not determined. (D) A dot-plot comparison of human *S100A15A* and chimpanzee *S100A15A* genomic sequences. The crosshatch marks the deleted region in human *S100A15A*. The locations of the PCR primers for experimental verification are indicated (arrows).

a deletion encompassing the 3′ part of the last exon and the following flanking region. Comparison of mouse and chimpanzee NLRP10 orthologs showed relatively high divergence at their C-termini. This suggests that a variation of the C-terminus of this protein is common among mammals and not necessarily specific to humans. Although these 3 genes could prove to be interesting, we concentrated on *MOXD2* and *S100A15A* for this study because their respective orthologs from mouse and chimpanzee encode proteins that are highly conserved.

Human *MOXD2* and *S100A15A* Genes Are Predicted to Have a Deletion

MOXD2 is the unannotated human ortholog of mouse *Moxd2*, which encodes a protein named "monooxygenase, DBH-like 2" (GenBank accession number NM_139296). Analysis of the sequence alignment between human and

chimpanzee genome assemblies showed a chimpanzee-only segment of 3658 bp in length (from 142,904,107 to 142,907,764 of the chimpanzee chromosome 7), which matched 2 exons of mouse *Moxd2* mRNA (fig. 1A). The corresponding genomic segments of human and chimpanzee were retrieved. The exons of human and chimpanzee *MOXD2* genes were identified based on mouse *Moxd2* gene. A dot matrix comparison (Sonnhammer and Durbin 1995) of the human and chimpanzee *MOXD2* genes revealed a deletion in the human chromosome 7 (fig. 1B). The deletion removed the last 2 of the 13 coding exons including 3′ untranslated region (UTR).

The *S100A15A* is an unannotated gene in the human chromosome 1. Initial analysis of the sequence alignment data showed a 1501-bp-long segment in the chimpanzee genome (from 132,521,559 to 132,523,059 of the chimpanzee chromosome 1), which overlaps with an exon of mouse *S100a15* mRNA but is missing in the human genome
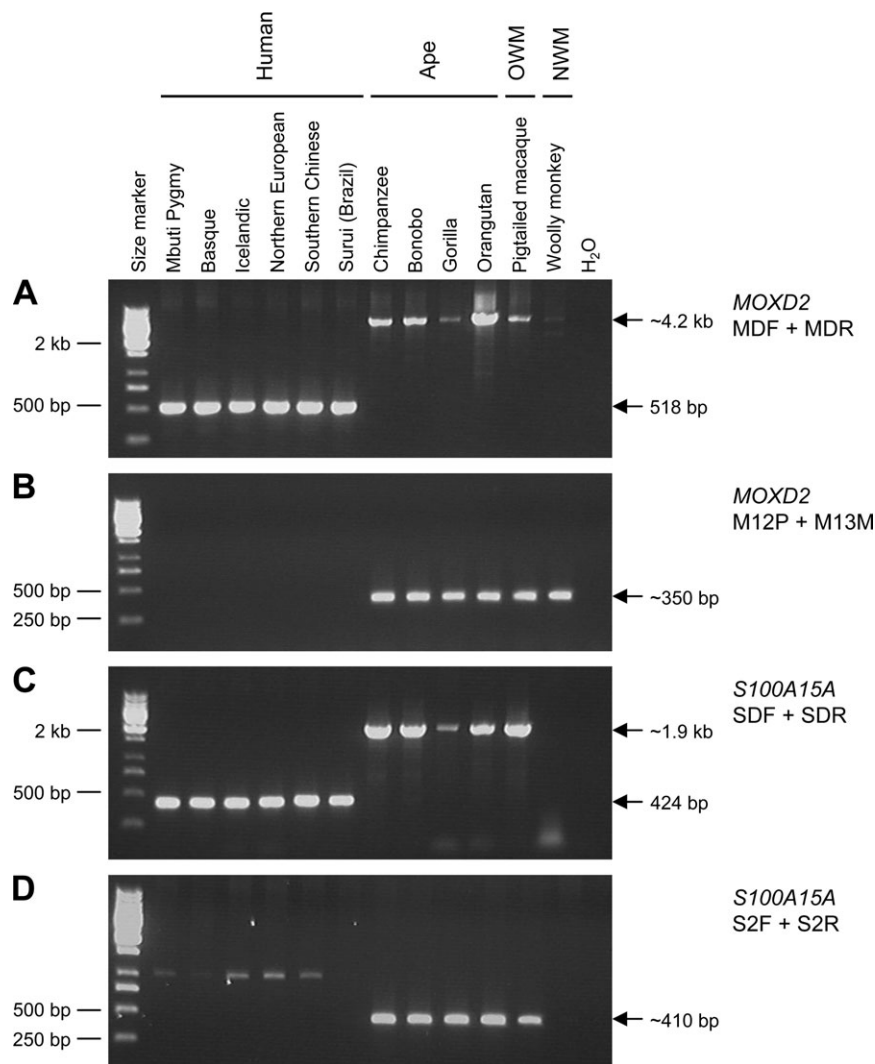
Fig. 2.—PCR analysis of the human-specific genomic deletions in *MOXD2* (*A*, *B*) and *S100A15A* (*C*, *D*) genes. PCR primer pairs designed to encompass the deleted region produced smaller bands from human DNA samples collected from various geographical locations but larger bands in nonhuman primate species (*A*, *C*). The amplified products containing the exonic segment were successfully obtained from the nonhuman primates but not from the human (*B*, *D*). Some faint bands in the human genomic DNA samples, which lack authentic targets, are probably due to nonspecific primer annealing. PCR products, target genes, and primer pairs are indicated at the right. OWM, Old World monkey; NWM, New World monkey.

(fig. 1*C*). The mouse *S100a15* gene has 3 exons, the coding region spanning from exon 2 to exon 3. A dot-plot analysis revealed that a deletion in the human *S100A15A* gene eliminated the start codon-containing exon 2 (fig. 1*D*). We named this *S100A15A* in order to distinguish it from the known human gene *S100A7a*, which is also known as *S100A15* because it is commonly thought as orthologous to mouse *S100a15* (Wolf et al. 2003, 2006). However, our sequence comparison shows that true human ortholog of mouse *S100a15* is the newly found *S100A15A* and that mouse has lost the ortholog of *S100A7* group of human genes (see below).

Experimental Validation of Exon Deletion in *MOXD2* and *S100A15A* Genes

In order to experimentally verify the predicted deletions in human *MOXD2* and *S100A15A* genes, we per-

formed genomic PCR analysis using human DNA samples and nonhuman primate DNA samples (fig. 2). Six human DNA samples with diverse geographical origin and nonhuman DNA samples isolated from 6 primate species were obtained from the Coriell Cell Repositories. Two pairs of PCR primers were designed to amplify the breakpoint junction and a part of the deleted segment, respectively, for each of *MOXD2* and *S100A15A* genes (see fig. 1*B* and *D* for primer locations). PCRs using the primer pair designed to encompass the breakpoint yielded smaller products from genomic DNA of all of the human individuals but larger products from the nonhuman primates for both *MOXD2* (fig. 2*A*) and *S100A15A* (fig. 2*C*) genes. The size reduction of the amplified segment clearly demonstrates that the predicted deletion is present in the human genomes. We also tried to amplify the exons that are missing from the human genome. We successfully observed amplified bands from the nonhuman primate species but not from the humans (fig. 2*B* and *D*). All amplified products
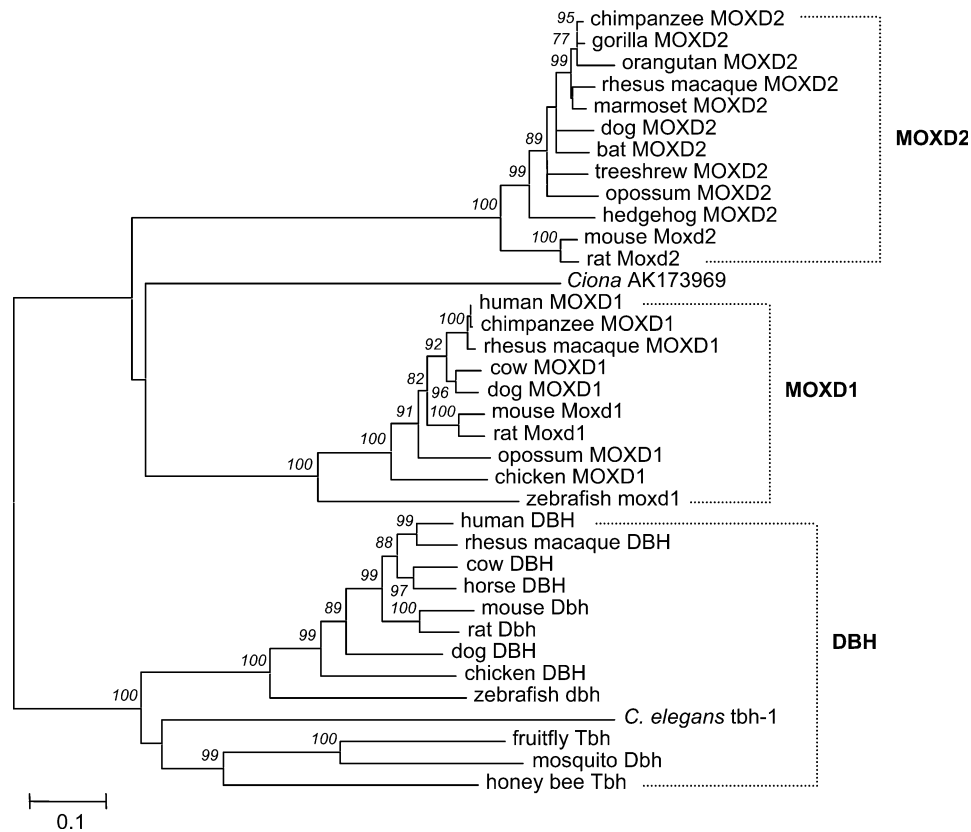
FIG. 3.—Phylogenetic tree of MOXD2, MOXD1, and DBH proteins. A Neighbor-Joining phylogenetic tree was constructed using MEGA3 with 1,000 bootstrap replicates from a multiple alignment of full-length proteins (supplementary fig. 1, Supplementary Material online). Bootstrap values greater than or equal to 70% are indicated on the nodes. The branching patterns of some proteins within orthologous groups do not exactly coincide with the universal species tree, for instance, rodent MOXD2 or MOXD1, possibly due to lineage-dependent difference of the substitution rate. Note that insects have the orthologs of vertebrate DBH.

were in the expected size range depending on the presence or absence of the deletion. It is not clear whether the absence of amplified bands in the case of the woolly monkey *S100A15A* gene is due to an independent deletion or to mismatches in primers which were designed based on intron sequences of the chimpanzee and the rhesus macaque (*Macaca mulatta*) *S100A15A*.

MOXD2 Is a Membrane-Bound Monooxygenase in Olfactory Epithelium

MOXD2 proteins are highly conserved among mammals including opossum (*Monodelphis domestica*), a marsupial (Supplementary fig. 1, Supplementary Material online). Protein sequence identity between the chimpanzee and the opossum orthologs is 82%. The exceptional level of conservation implies that this protein may play an important role in mammalian biology. MOXD2 contains a signal peptide, a GPI anchor signal, and a catalytic core of the copper monooxygenase, indicating that the protein is a membrane-associated enzyme. MOXD2 shows sequence similarity to dopamine β-hydroxylase (DBH) and MOXD1 (monooxygenase, DBH-like 1) (see supplementary fig. 1, Supplementary Material online). Sequence comparison, exon organization, and phylogenetic analysis indicate that *MOXD2/MOXD1/DBH* genes originated from a common ancestor by gene

duplication and divergence (fig. 3). The DBH protein, which does not have a GPI anchor signal in its C-terminus, is also attached to the membrane by using the uncleaved signal peptide as the anchor (Feng et al. 1992; Houhou et al. 1995). Cleavage of the signal peptide also generates soluble form of the enzyme. DBH converts dopamine to norepinephrine in the synaptic vesicles of postganglionic sympathetic neurons, and its mutation causes several psychiatric disorders (Kim et al. 2002; Cubells and Zabetian 2004; Timmers et al. 2004). The MOXD1 protein is composed of the similar domains when compared with MOXD2. The divergence of MOXD1 and MOXD2 seems to have occurred in an ancestral chordate. MOXD1 proteins form tightly membrane-associated oligomeric enzymes that are predicted to hydroxylate a hydrophobic substrate in the endoplasmic reticulum (Xin et al. 2004). The mouse *Moxd2* gene is highly expressed in the medial olfactory epithelium (the GNF Expression Atlas 2 [Su et al. 2004] track at the UCSC Genome Browser Database). The specialized expression pattern and the molecular function of homologs suggest that Moxd2 protein may play a role in olfactory sense by converting one neurotransmitter to another or by modifying incoming odorant molecules. Loss of functional MOXD2 might be implicated in the alteration of smell sensitivity of the human.

The deletion in the human *MOXD2* gene removes 3′ UTR and the poly(A) signal as well as coding region in the

last 2 terminal exons. Loss of the 3′ UTR regulatory elements may interfere with proper mRNA processing and/or translation. In addition to the deletion, the gene suffers a nonsense mutation (codon 305, CGA/TGA) within exon 6, which is polymorphic in human population (http://www.ncbi.nlm.nih.gov/projects/SNP/; the Single Nucleotide Polymorphism database [dbSNP] accession number rs4376428). The dbSNP record indicates that the nonsense allele is rather frequent in Asians but scarce in Africans. The presence of the secondary mutation strengthens the notion that the gene is inactive in humans. However, because the promoter and the preceding exons may be intact in alleles without a nonsense mutation in exon 6, we cannot rule out the possibility that the gene is still transcribed and even produces the truncated proteins. The expression of the gene in the pooled human tissue has been observed as ESTs, for examples, DY654834 and DY655575.

MOXD2 Genes Are Mutated in Higher Primates

When we analyzed *MOXD2*/*Moxd2* orthologs of mammalian species derived from various sources including sequences of amplified genomic fragments and the WGS clones at the NCBI Trace Archive, we found more cases of disruptive mutations in higher primates. Besides the deletion and the nonsense mutations in the human, we observed 2 nonsense mutations and a splice site mutation in the orangutan *MOXD2* and a frameshift mutation in the rhesus macaque ortholog (supplementary fig. 1, Supplementary Material online). The 2 nonsense mutations, one in exon 8 (codon 402, TAA) and the other in exon 13 (codon 570, TAG), were detected in both Sumatran orangutans and Bornean orangutans (*Pongo pygmaeus pygmaeus*). The second nonsense mutation exhibits polymorphism in Sumatran orangutans: CAG/TAG. The splice donor consensus sequence GT of intron 3 was absent in Sumatran orangutans due to a 2-bp deletion but present in their Bornean cousins. The frequent incidence of such deleterious mutations in the orangutan *MOXD2* gene indicates that the gene also became not in use in the early stage of the orangutan evolution and is now decaying. It is uncertain whether the independent inactivation of the gene in orangutan features the same phenotypic change as in human.

The rhesus macaque genome assembly (January 2006 freeze) and the WGS clone sequences examined reveal a 13-nt deletion in the last exon of the *MOXD2* gene, resulting in frameshift and premature termination. The altered open reading frame would encode a protein with 586 residues. Other primate genes without a mutation would produce proteins with 618 amino acids. The mutation should not modify the enzymatic core of the protein but abolishes the GPI anchor sequence at its C-terminus. Therefore, the enzyme would be soluble, if not attached to the membrane by using a signal anchor as in the case of DBH. It is not clear whether the altered form of MOXD2 protein in the rhesus macaque confers similar enzymatic activity and biological role as in other species.

S100A15A Is Involved in Skin Differentiation

We compared orthologous S100A15A proteins deduced from various sequence data of mammalian species (supplementary fig. 2, Supplementary Material online). The orthologous protein sequences were well conserved even in a marsupial species. The chimpanzee and the opossum orthologs share 68% sequence identity. Unlike the case of *MOXD2* genes, we do not find any mutation causing coding disruption in nonhuman *S100A15A* genes examined. The S100A15A protein contains 2 calcium-binding domains and belongs to the subfamily A of the family of S100 proteins (supplementary fig. 2, Supplementary Material online). S100 proteins are involved in cell growth and differentiation, cell cycle regulation, and metabolic control (Donato 2003). Many members are implicated in human diseases, including skin disease, as well as cancer pathogenesis (Heizmann et al. 2002; Eckert et al. 2004; Emberley et al. 2004).

More than 20 genes encoding S100 proteins exist as a cluster in a genomic locus of each mammalian genome (e.g., on human chromosome 1q21.3 and on mouse chromosome 3qF1, according to the UCSC Genome Browser Database) (Marenholz et al. 2004; Ravasi et al. 2004). The multigenic nature and lineage-dependent expansion or pseudogenization of some members make nomenclature of the S100 protein genes and correct assignment of orthologs complicated (Marenholz et al. 2004, 2006). For example, the mouse S100a15 protein was considered as the ortholog of human S100A7a and S100A7c (also known as Psoriasin) in recent studies (Webb et al. 2005; Eckert and Lee 2006; Wolf et al. 2006) and was renamed as S100a7a (Marenholz et al. 2006). However, a phylogenetic analysis of mammalian S100A proteins including newly identified S100A15A orthologs indicates that S100A7 and S100A15A are distinct from each other even though they are the closest members among S100A proteins (fig. 4). Human and chimpanzee S100A7a and S100A7c show 32–33% amino acid sequence identity, respectively, with mouse S100a15, whereas the full-length chimpanzee S100A15A and mouse S100a15 share 84% amino acid identity. The *S100A15A* and *S100A7* genes, which are tandemly located in the genome, seem to have been formed by a segmental duplication before the mammal emerged because the opossum, a marsupial species, also possesses *S100A15A*. The S100A7 proteins are found in horse and cow but not in mouse, suggesting that it was lost during rodent evolution. As a result, the mouse *S100a15* and the human *S100A7* are the most similar members between the 2 species, despite their low level of sequence identity, that are still functioning in the cells of each species. However, they must be considered evolutionarily paralogous on which disparate selection pressure might have been exerted. In order to build a more accurate evolutionary history and a better orthology relationship of S100A proteins, more number of genomes need to be inspected, including rhesus macaque, mouse, rat, cow, and dog.

Mouse S100a15 mRNAs and/or proteins were detected during skin maturation, especially in differentiating cells of the hair follicles and the cornified layer (Webb et al. 2005; Wolf et al. 2006). It is also expressed in mouse mammary gland and is upregulated during mammary tumorigenesis (Webb et al. 2005) Mouse S100a15 may have similar function to human S100A7 because they are the closest members in the S100A protein family. However, the 2
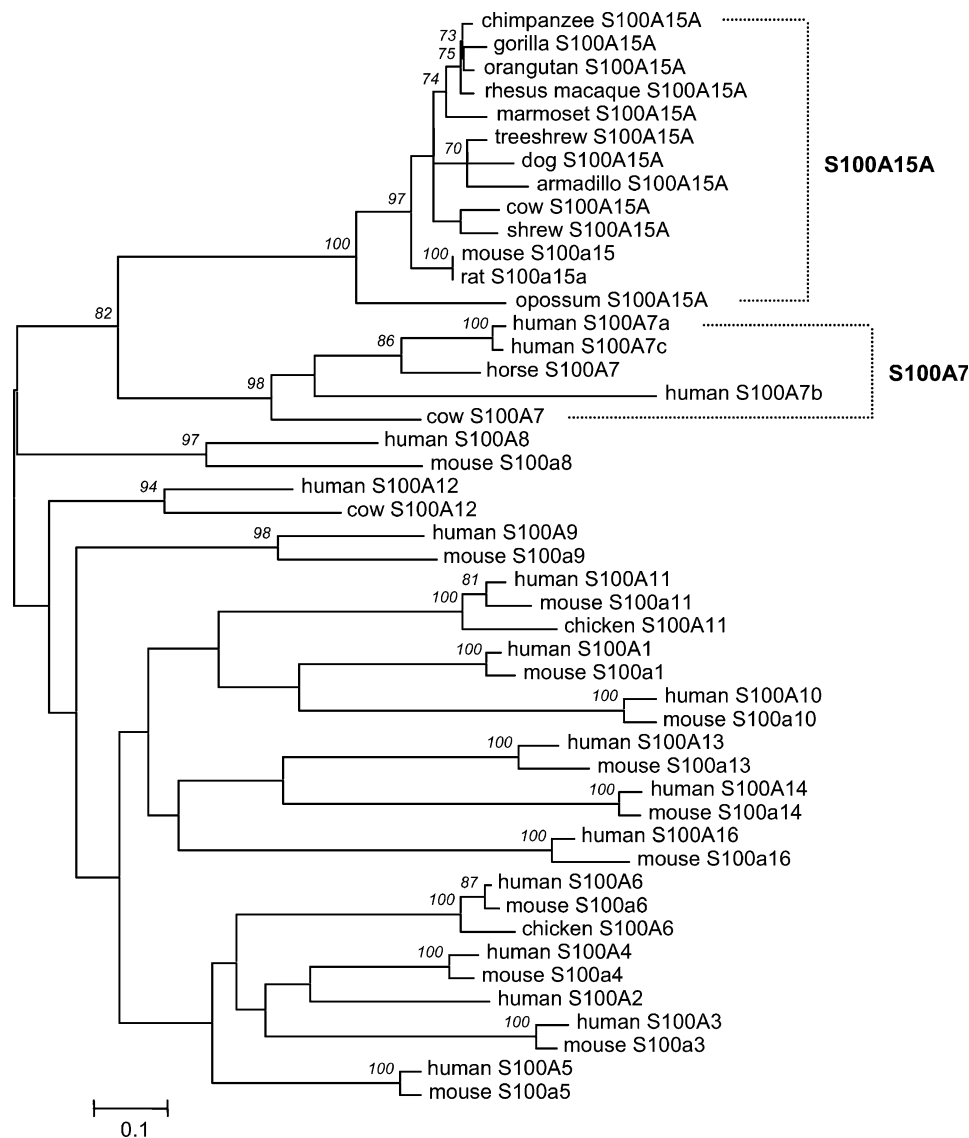
FIG. 4.—Phylogenetic tree of S100A15A and homologous proteins. A tree was generated based on a multiple alignment of S100A proteins (supplementary fig. 2, Supplementary Material online) as described in figure 3. S100A7 is the closet paralogous group of S100A15A. The orthologous S100a7 seems to be lost in rodent lineage, making S100a15 the closest functional mouse homolog of human S100A7.

proteins have low sequence similarity even in the same species, for example, the sequence identity between S100A7a and S100A15A in chimpanzee is 33%. Chimpanzee S100A7a and S100A7c share 94% sequence identity. The conservation levels of the S100A7 orthologs and S100A15A orthologs among mammals are quite different: the sequence identity between the S100A15A orthologs of chimpanzee and cow is 87% but that between chimpanzee S100A7a and cow S100A7 is 62%. Accelerated divergence among S100A7 proteins and duplicative expansion of *S100A7* genes in primates suggests that S100A7 may confer specialized function in some mammalian lineages such as primates, whereas S100A15A retains more basal function for all mammalian species. Thus, although S100A7 may have substituted some of the functions of S100A15A, the substitution may not have been complete, in which case the loss of *S100A15A* gene function might effect human-

specific features in the structure or physiology of human skin.

Complex Evolutionary History of *MOXD2* and *S100A15A* in Great Apes

To perform molecular evolutionary analyses, we prepared multiple sequence alignment of human, chimpanzee, orangutan, gorilla, rhesus macaque, and marmoset coding sequences of *MOXD2* and *S100A15A* genes. The alignments include 1492 bp for *MOXD2* and 174 bp for *S100A15A*. The nonsynonymous/synonymous substitution ratios ($\omega = d_N/d_S$) were computed based on the codon maximum likelihood method (Yang 1998; Stedman et al. 2004) by using CODEML program in the PAML package (Yang 1997). The estimated numbers of nonsynonymous (N) and
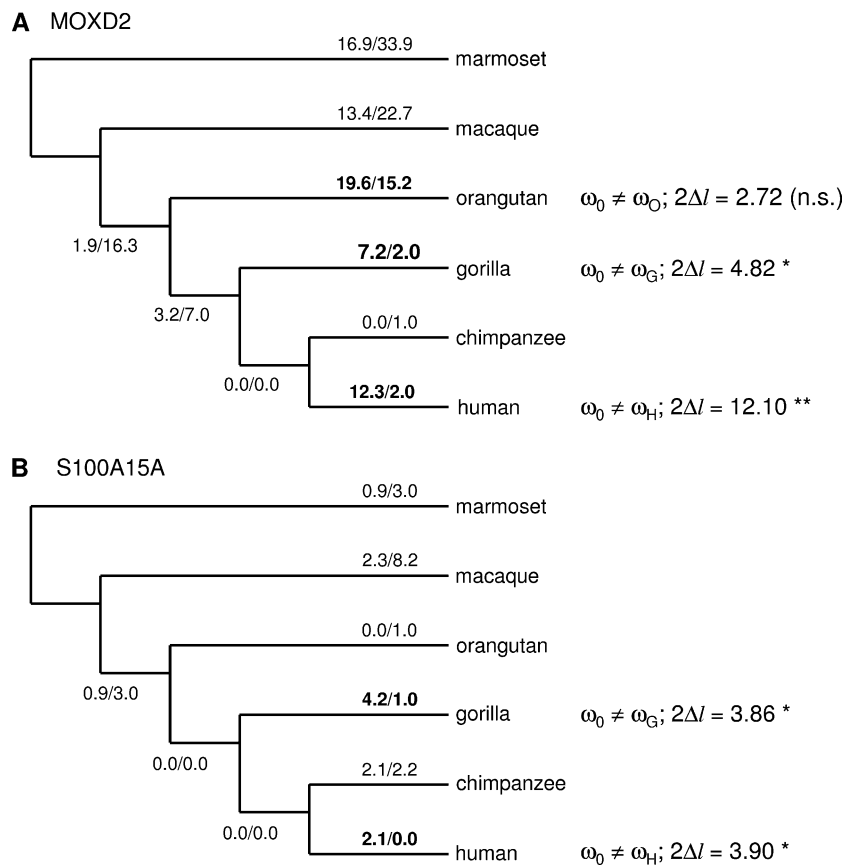
A MOXD2

B S100A15A

FIG. 5.—Molecular evolutionary analyses of primate *MOXD2* (*A*) and *S100A15A* (*B*) orthologs. The nonsynonymous/synonymous ratios ($\omega = d_N/d_S$) along the branches were computed by using the maximum likelihood method for codons available in PAML. In all branches, the ω ratio was set free to vary (the free-ratio model). The maximum likelihood estimates of the number of nonsynonymous (N) and synonymous (S) substitutions are presented along each branch. It is notable that a higher nonsynonymous change (N > S) is observed in some lineages (bold). Selected tested hypotheses are listed at the right. ω_H, ω_C, ω_G, ω_O, ω_R, and ω_M stand for the ω value in the human, the chimpanzee, the gorilla, the orangutan, the rhesus macaque, and the marmoset lineage, respectively, and ω_0 in all other branches not free to vary. Single asterisk indicates significant ($0.01 < P < 0.05$; $\chi_1^2 = 3.84$). Double asterisk indicates extremely significant ($P < 0.01$; $\chi_1^2 = 6.63$). n.s., not significant. See supplementary tables 3 and 4 (Supplementary Material online) for further details.

synonymous (S) substitutions were calculated along each branch.

Analysis of *MOXD2* orthologs revealed an elevated nonsynonymous change (N > S) in each of the branches leading to human, gorilla, and orangutan genes (fig. 5*A*). The estimated ω values for these branches are 2.4709, 1.4315, and 0.5157, respectively (see supplementary table 3A [Supplementary Material online] for all ω values). According to the log likelihood ratio test (Yang 1998), ω values for the human (ω_H) and the gorilla (ω_G) are significantly different from that of background value ω_0 (fig. 5*A* and test numbers 1 and 3 in supplementary table 3B; Supplementary Material online). They are larger than 1, but we cannot reject the null hypothesis that they are equal to 1 (test numbers 7 and 8 in supplementary table 3B, Supplementary Material online), indicating that these genes are not under strongly positive or negative (purifying) selection. It is also notable that there is no nonsynonymous substitution along the branch either for the human–chimpanzee ancestor or for the chimpanzee, indicating a strong purifying selection in these branches.

Comparative analysis of *S100A15A* orthologs also reveals accelerated nonsynonymous substitutions in the

human and the gorilla lineages that are statistically significant compared with the background level (fig. 5*B* and supplementary tables 4A and 4B; Supplementary Material online). The ω value of human *S100A15A* (ω_H) is infinite because the number of synonymous substitution is 0. However, again we cannot reject the null hypothesis that it is equal to 1 (test number 7 in supplementary table 4B, Supplementary Material online). Strong purifying selection is hinted at in the orangutan, the human–chimpanzee ancestor, and the human–chimpanzee–gorilla ancestor branches. It is important to note that the small size of *S100A15A* sequence analyzed (174 bp) may restrict computation of reliable ω values.

Possible Mechanism for Exon-Deletion Mutation

A mechanism known as *Alu* recombination–mediated deletion has been reported to play a role in the human-specific inactivation of the *CMAH* (Hayakawa et al. 2001) and other genes (Sen et al. 2006). However, no *Alu* repeats or other repetitive elements are present near the deletion boundaries of the *MOXD2* and the *S100A15A* genes in

the human and the chimpanzee genome. The cleaved and rejoined regions do not show any other special sequence features such as inverted repeats in either of the cases. Sequence alignment between the human and the chimpanzee genomic sequences surrounding the breakpoint junction shows no mismatched or additional sequences in each of the cases, indicating that the genomic fragment was cleanly removed. We assume that the deletion in the *MOXD2* and the *S100A15A* genes were mediated by DNA double-strand break and nonhomologous end joining (Cahill et al. 2006).

Conclusion

In summary, we developed a semiautomatic procedure for systematic identification of human-specific exon-deletion mutations and experimentally verified those in the *MOXD2* and the *S100A15A* genes. The *MOXD2* gene encodes a monooxygenase of which mouse ortholog is highly expressed in the olfactory sensory organ. Inactivation of this gene in humans may have resulted in an alteration of the olfactory sense. A preferential expression of mouse *S100a15* gene in the skin suggests that the loss of this protein in humans may contribute to the structural or physiological difference of the human skin from other mammals.

Supplementary Material

Supplementary tables 1–4 and figures 1 and 2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 340:783–795.

Cahill D, Connor B, Carney JP. 2006. Mechanisms of eukaryotic DNA double strand break repair. Front Biosci. 11:1958–1976.

Chou HH, Takematsu H, Diaz S, Iber J, Nickerson E, Wright KL, Muchmore EA, Nelson DL, Warren ST, Varki A. 1998. A mutation in human CMP-sialic acid hydroxylase occurred after the Homo-Pan divergence. Proc Natl Acad Sci USA. 95:11751–11756.

Cubells JF, Zabetian CP. 2004. Human genetics of plasma dopamine beta-hydroxylase activity: applications to research in psychiatry and neurology. Psychopharmacology (Berl). 174:463–476.

Donato R. 2003. Intracellular and extracellular roles of S100 proteins. Microsc Res Tech. 60:540–551.

Eckert RL, Broome AM, Ruse M, Robinson N, Ryan D, Lee K. 2004. S100 proteins in the epidermis. J Invest Dermatol. 123:23–33.

Eckert RL, Lee KC. 2006. S100A7 (Psoriasin): a story of mice and men. J Invest Dermatol. 126:1442–1444.

Emberley ED, Murphy LC, Watson PH. 2004. S100 proteins and their influence on pro-survival pathways in cancer. Biochem Cell Biol. 82:508–515.

Feng Z, Angeletti RH, Levin BE, Sabban EL. 1992. Glycosylation and membrane insertion of newly synthesized rat dopamine beta-hydroxylase in a cell-free system without signal cleavage. J Biol Chem. 267:21808–21815.

Finn RD, Mistry J, Schuster-Bockler B, et al. (13 co-authors). 2006. Pfam: clans, web tools and services. Nucleic Acids Res. 34:D247–D251.

Folk GE Jr, Semken HA Jr. 1991. The evolution of sweat glands. Int J Biometeorol. 35:180–186.

Glazko GV, Nei M. 2003. Estimation of divergence times for major lineages of primate species. Mol Biol Evol. 20:424–434.

Goodstadt L, Ponting CP. 2001. CHROMA: consensus-based colouring of multiple alignments for publication. Bioinformatics. 17:845–846.

Hahn Y, Lee B. 2005. Identification of nine human-specific frameshift mutations by comparative analysis of the human and the chimpanzee genome sequences. Bioinformatics. 21:i186–i194.

Hahn Y, Lee B. 2006. Human-specific nonsense mutations identified by genome sequence comparisons. Hum Genet. 119:169–178.

Hayakawa T, Satta Y, Gagneux P, Varki A, Takahata N. 2001. Alu-mediated inactivation of the human CMP-*N*-acetylneuraminic acid hydroxylase gene. Proc Natl Acad Sci USA. 98:11399–11404.

Heizmann CW, Fritz G, Schafer BW. 2002. S100 proteins: structure, functions and pathology. Front Biosci. 7:d1356–d1368.

Hinrichs AS, Karolchik D, Baertsch R, et al. (27 co-authors). 2006. The UCSC genome browser database: update 2006. Nucleic Acids Res. 34:D590–D598.

Houhou L, Lamouroux A, Biguet NF, Mallet J. 1995. Expression of human dopamine β-hydroxylase in mammalian cells infected by recombinant vaccinia virus. Mechanisms for membrane attachment. J Biol Chem. 270:12601–12606.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. Nature. 431:931–945.

Kehrer-Sawatzki H, Cooper DN. 2007. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. Hum Mutat. 28:99–130.

Kim CH, Zabetian CP, Cubells JF, Cho S, Biaggioni I, Cohen BM, Robertson D, Kim KS. 2002. Mutations in the dopamine β-hydroxylase gene are associated with human norepinephrine deficiency. Am J Med Genet. 108:140–147.

Kumar S, Filipski A, Swarna V, Walker A, Hedges SB. 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. Proc Natl Acad Sci USA. 102:18842–18847.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.

Li WH, Saunders MA. 2005. The chimpanzee and us. Nature. 437:50–51.

Marenholz I, Heizmann CW, Fritz G. 2004. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). Biochem Biophys Res Commun. 322:1111–1122.

Marenholz I, Lovering RC, Heizmann CW. 2006. An update of the S100 nomenclature. Biochim Biophys Acta. 1763:1282–1283.

Olson MV. 1999. When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet. 64:18–23.

Olson MV, Varki A. 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. Nat Rev Genet. 4:20–28.

Pagel M, Bodmer W. 2003. A naked ape would have fewer parasites. Proc Biol Sci. 270:S117–S119.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature. 441:1103–1108.

Pollard KS, Salama SR, Lambert N, et al. (16 co-authors). 2006. An RNA gene expressed during cortical development evolved rapidly in humans. Nature. 443:167–172.

Ravasi T, Hsu K, Goyette J, Schroder K, Yang Z, Rahimi F, Miranda LP, Alewood PF, Hume DA, Geczy C. 2004. Probing the S100 protein family through genomic and functional analysis. Genomics. 84:10–22.

Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. 2006. Human genomic deletions mediated by recombination between Alu elements. Am J Hum Genet. 79:41–53.

Sonnhammer EL, Durbin R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 167:GC1–GC10.

Stedman HH, Kozyak BW, Nelson A, Thesier DM, Su LT, Low DW, Bridges CR, Shrager JB, Minugh-Purvis N, Mitchell MA. 2004. Myosin gene mutation correlates with anatomical changes in the human lineage. Nature. 428:415–418.

Su AI, Wiltshire T, Batalov S, et al. (13 co-authors). 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA. 101:6062–6067.

The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 437:69–87.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22: 4673–4680.

Timmers HJ, Deinum J, Wevers RA, Lenders JW. 2004. Congenital dopamine-β-hydroxylase deficiency in humans. Ann N Y Acad Sci. 1018:520–523.

Tschopp J, Martinon F, Burns K. 2003. NALPs: a novel protein family involved in inflammation. Nat Rev Mol Cell Biol. 4:95–104.

Varki A, Altheide TK. 2005. Comparing the human and chimpanzee genomes: searching for needles in a haystack. Genome Res. 15:1746–1758.

Wang X, Grus WE, Zhang J. 2006. Gene losses during human origins. PLoS Biol. 4:e52.

Webb M, Emberley ED, Lizardo M, et al. (13 co-authors). 2005. Expression analysis of the mouse S100A7/psoriasin gene in skin inflammation and mammary tumorigenesis. BMC Cancer. 5:17.

Wolf R, Mirmohammadsadegh A, Walz M, Lysa B, Tartler U, Remus R, Hengge U, Michel G, Ruzicka T. 2003. Molecular cloning and characterization of alternatively spliced mRNA isoforms from psoriatic skin encoding a novel member of the S100 family. FASEB J. 17:1969–1971.

Wolf R, Voscopoulos CJ, FitzGerald PC, Goldsmith P, Cataisson C, Gunsior M, Walz M, Ruzicka T, Yuspa SH. 2006. The mouse S100A15 ortholog parallels genomic organization, structure, gene expression, and protein-processing pattern of the human S100A7/A15 subfamily during epidermal maturation. J Invest Dermatol. 126:1600–1608.

Xin X, Mains RE, Eipper BA. 2004. Monooxygenase X, a member of the copper-dependent monooxygenase family localized to the endoplasmic reticulum. J Biol Chem. 279:48159–48167.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13:555–556.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 15:568–573.