

The Rise and Fall of the Chemoattractant Receptor GPR33[§]

Received for publication, April 1, 2005, and in revised form, June 29, 2005
Published, JBC Papers in Press, June 29, 2005, DOI 10.1074/jbc.M503586200

Holger Römler[‡], Angela Schulz[‡], Christian Pitras[§], Graham Coop[¶], Molly Przeworski^{||},
Svante Pääbo^{**}, and Torsten Schöneberg[‡] ^{‡‡}

From the [‡]Institute of Biochemistry, Molecular Biochemistry, Medical Faculty, University of Leipzig, 04103 Leipzig, Germany, the [§]Institute for Zoo and Wildlife Research, Department of Evolutionary Genetics, 10315 Berlin, Germany, the [¶]Department of Statistics, University of Oxford, Oxford, OX1 3TG, United Kingdom, the ^{||}Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, and the ^{**}Max-Planck-Institute of Evolutionary Anthropology, 04103 Leipzig, Germany

Chemokine and chemoattractant receptors are members of the large superfamily of G protein-coupled receptors (GPCR), which control leukocyte chemotaxis. In addition to their physiological role, several chemokine and chemoattractant receptors, such as CCR5 and Duffy, have been directly associated with pathogen entry. GPR33 is an orphan chemoattractant GPCR that was previously identified as a pseudogene in humans. GPR33 evolved in mammals about 125–190 million years ago. The cloning and analysis of more than 120 mammalian GPR33 orthologs from 16 of 18 eutherian orders revealed an inactivation of this chemoattractant GPCR not only in humans, but also in several great ape and rodent species. Intriguingly, in all ape and some rodent species where the inactivation occurred, samples harbored both pseudogene and intact gene variants. The analysis of over 1200 human individuals representing all major linguistic groups revealed that the intact allele of *GPR33* is still present in the human population. Estimates of the age of the human alleles suggest inactivation in the past 1 million years. Similarly, analysis of more than 120 wild-caught gray rats (*Rattus norvegicus*), revealed that inactivation of *gpr33* is worldwide fixed and occurred in less than 0.7 million years ago. The coincidental inactivation and its fixation in several species of distantly related mammalian orders suggest a selective pressure on this chemoattractant receptor gene.

Recruitment of cells involved in the cellular defense against pathogens is mediated by an armada of cell surface proteins, chemokines, complement factors, and leukotrienes. The receptors that mediate the action of these factors all belong to the large family of G protein-coupled receptors (GPCR),¹ and con-

stitute within this family structural classes sharing sequence similarities and common transduction pathways (1). In addition to the well known repertoire of CC and CXC chemokine receptors, there are several orphan GPCRs that share structural similarities with chemokine and other chemoattractant receptors and that are currently under intensive investigation. Less attention, however, has been directed to pseudogenes of chemoattractant receptors, among them GPR33, which was identified as a pseudogene in human but as an intact gene in mouse (2). The human *GPR33* gene contains a premature stop codon within the coding sequence of the second intracellular loop but no other obvious structural defects, suggesting a recent inactivation of the receptor. Its close structural relation to ChemR23, a chemokine receptor for chemerin and SIV coreceptor (3, 4), formylmethionylleucylphenylalanine receptors and other chemokine receptors, focused our attention on the potential role of GPR33 in the cellular defense that was lost during human evolution.

By screening all vertebrate classes it turned out that GPR33 arose together with the first mammals. The analysis of GPR33 orthologs revealed an almost simultaneous inactivation in humans, several apes, as well as in several rodent species about 0.5–1 million years ago. Our data provide evidence for a selective pressure to inactivate this chemoattractant GPCR. However, a small fraction of the human population still harbors an intact *GPR33* allele. Our finding may be of great medical relevance because mutational inactivation of other chemokine GPCRs, such as CCR5 and Duffy, are associated with resistance to infections by HIV-1 and by *Plasmodium vivax*, respectively (5–7).

EXPERIMENTAL PROCEDURES

Cloning of *gpr33* Orthologs—To identify the *gpr33* sequence in other vertebrates, genomic DNA samples were prepared from tissue or peripheral mononuclear blood cells of various species or were kindly provided by several other laboratories (supplemental materials Table S1). Tissue samples were digested in lysis buffer (50 mM Tris/HCl, pH 7.5, 100 mM EDTA, 100 mM NaCl, 1% SDS, 0.5 mg/ml proteinase K) and incubated at 55 °C for 18 h. DNA was purified by phenol/chloroform extraction and ethanol precipitation. Based on the human and mouse *GPR33* sequences (2), sets of degenerated primer pairs (supplemental materials Table S2) were applied to amplify *gpr33*-specific sequences. PCR were performed with *Taq* polymerase under variable annealing and elongation conditions. Specific PCR products were directly sequenced and/or subcloned for sequencing into the pCR2.1-TOPO vector (Invitrogen, La Jolla, CA). In case of heterozygosity, allelic separation was performed by subcloning and subsequent sequencing. Sequencing reactions were performed on PCR products with a dye-terminator cycle sequencing kit (Applied Biosystems) on an ABI 3700 automated se-

* This work was supported by the Deutsche Forschungsgemeinschaft, Bundesministerium für Bildung und Forschung, Fonds der Chemischen Industrie, and IZKF Leipzig. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[§] The on-line version of this article (available at <http://www.jbc.org>) contains additional text, Tables S1–S7, and Fig. S1.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY490569–AY490743, AY493989–AY494004, AY502103, AY502104, and AY528865 (supplemental materials Table S1).

^{‡‡} To whom correspondence should be addressed: Institute of Biochemistry, Molecular Biochemistry (Max-Planck-Institute Interim), Medical Faculty, University of Leipzig, Deutscher Platz 6, 04103 Leipzig, Germany. Tel.: 49-341-3550-850; Fax: 49-341-3550-855; E-mail: schoberg@medizin.uni-leipzig.de.

¹ The abbreviations used are: GPCR, G protein-coupled receptor; IP,

inositol phosphate; TRITC, tetramethylrhodamine isothiocyanate; HIV-1, human immunodeficiency virus type 1.

quencer (Applied Biosystems). Based on considerable sequence similarities of the 5'- and 3'-untranslated regions of *gpr33* genes, primers were designed (supplemental materials Table S2) that allowed for the identification of sequences encoding the N and C termini of mammalian GPR33 orthologs.

***gpr33* Sequence Analyses in Human, Chimpanzee, Bonobo, and Rat Samples**—To screen large sample sets for the presence of the TGA allelic variant we established a restriction analysis. Because the TGA codon 140 and the flanking sequence does not contain a suitable site for restriction analysis, we designed a one-base mismatch antisense primer (DdeI-AS 5'-AATGCTGAAGCCCAGCGCGGGcTC-3') that introduces a new DdeI restriction site in the PCR product if TGA codon 140 is present. Together with a sense primer (DdeI-S 5'-CTGGAACCTTGGAAGTGCCTTGTGC-3') PCR resulted in a 166-bp fragment. The reactions were initiated with denaturation at 92 °C for 1 min, followed by 35 cycles of denaturation at 92 °C for 20 s, annealing at 55 °C for 20 s, and elongation at 72 °C for 20 s. A final extension step was performed at 72 °C for 10 min. PCR products were treated with DdeI (37 °C overnight) and separated in a 3% agarose gel. In the case of the TGA allele, the 166-bp PCR product was cleaved into two fragments (139 and 27 bp). Uncut fragments were always sequenced. This DdeI restriction analysis was performed for 1217 human individuals (for details, see supplemental materials Table S3). The DNA sample collection containing mostly samples of the Centre d'Etude du Polymorphisme Human panel, a panel representing all major linguistic groups (DNA panel from Ref. 8), 45 individuals from Papua New Guinea (kindly provided by Mark Stoneking), and 21 Yoruba individuals (International HapMap Project), was used.

To search for individual variations we sequenced the *GPR33* coding region in 85 humans (DNA panel from Ref. 8). PCR fragments were amplified with genomic DNA samples and primers (hu-32, 5'-ACTGTT-TCTCACTCCACAGGTC-3' and hu-33, 5'-CTATTGGTATAATTGAC-CAAGTGC-3'). Genomic DNA (100 ng) was used in the PCR (50 μ l) with primers (10 pmol each), standard buffer (PerkinElmer Life Sciences), dNTP (200 μ M), and *Taq* polymerase (1 unit; PerkinElmer Life Sciences). The reactions were initiated with denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 94 °C for 45 s, annealing at 60 °C for 45 s, and elongation at 72 °C for 2 min. A final extension step was performed at 72 °C for 10 min. PCR products were separated in a 1% agarose gel, purified by using a PCR Product Purification Kit (Qiagen), and sequenced.

For dating purposes several kilobase 5'- and 3'-flanking regions of human, chimpanzee, bonobo, *Rattus rattus*, and *Rattus norvegicus* were amplified by PCR and directly sequenced (supplemental materials Tables S4 and S5). Similarly, the coding region of the human *GPR33* and approximately 4 kb of the 3'-flanking region were sequenced from 21 Yoruba individuals (International HapMap Project) and analyzed for signatures of positive selection using the D-tests of Tajima (9) and Fu and Li (10).

To analyze the allelic variance in rat *gpr33*, genomic DNA was prepared from 114 wild-caught *R. norvegicus* from different parts of Germany (Nordrhein-Westfalen, Mecklenburg-Vorpommern, and Berlin-Brandenburg), 5 wild-caught *R. norvegicus* from Russia (Siberia), 6 wild-caught *R. norvegicus* from the United States (Alaska), 2 wild-caught *R. norvegicus* from Japan, 2 wild-caught *R. rattus* from Germany (Berlin-Brandenburg), and 4 wild-caught *R. rattus* from Paraguay. *gpr33* coding regions were amplified by PCR and sequenced.

Sequence Analyses and Age Estimates—Nucleotide and amino acid sequence alignments were made with the Clustal X program (11). Maximum likelihood trees reconstructed with the PHYLIP software package (12) and the neighbor-joining method (13) gave essentially similar results. Bootstrap replications were conducted to assess the reliability of the trees.

The expected age of a selected allele at a given population frequency is a well studied problem. However, in this case, we only know the sample frequency of the derived allele, not the population frequency. Griffiths (14) investigated various properties of a derived allele given the sample frequency under a range of models. The approach makes use of a diffusion process, which is the large population limit of a wide range of random mating, constant population size models. The derived allele appears in the population at an arbitrarily small frequency, after a single mutational event.

We have implemented the approach of Griffiths (14) to evaluate the expected age of an allele given its sample frequency under a model of genetic selection and under a neutral model. Specifically, Griffiths (14) derives the joint probability density function of the population frequency of the allele and its age given the sample frequency. From this, we computed the expected age of an allele for a given sample frequency

by numerically integrating the conditional expectation of the age of the allele given a population frequency over the marginal distribution of the population frequency conditional on the sample frequency.

This approach yields the expected age in time scaled in units of $2N_e$, generations where N_e is the effective population size of the species. Thus, an estimate of N_e and of the generation time is required to translate the age into years. We used a generation time of 25 for humans and 20 for chimpanzee and orangutan, and estimates of N_e reported in Table II. Because nothing is known about the effective population size of siamang, we omit the species from this analysis.

To obtain a second estimate of the age of the allele, assuming neutrality, we first used the program PHASE version 2.1 to infer haplotypes from genotypes, using default parameters and multiple runs to check the accuracy of the results (15). We then used the GENETREE program to estimate the age of the mutation (16). The program finds the maximum likelihood estimate of the age of the allele for a given population mutation rate, θ , through neutral coalescent simulations conditional on the full polymorphism data set. The age was evaluated for the estimate of Watterson (17), θ , based on the number of segregating sites in the sample.

Construction of Wild-type and Mutant GPR33, Cell Culture, Transfection, and Functional Assays—Full-length *GPR33* sequences were inserted into the mammalian expression vector pcDps. The human and mouse *GPR33* were tagged with an N-terminal hemagglutinin and a C-terminal FLAG epitope. The *GPR33* mutation (A125D) was introduced into the hemagglutinin-tagged version of the murine *Gpr33* using PCR-based site-directed mutagenesis and restriction fragment replacement strategy. The identity of the various constructs and the correctness of all PCR-derived sequences were confirmed by restriction analysis and sequencing. COS-7 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 100 units/ml penicillin, and 100 μ g/ml streptomycin at 37 °C in a humidified 7% CO₂ incubator. Lipofectamine (Invitrogen) was used for transient transfection of COS-7 cells. Thus, cells were split into 12-well plates (1.5 \times 10⁵ cells/well) and transfected with a total amount of 1 μ g of plasmid DNA/well. To measure inositol phosphate (IP) formation, transfected COS-7 cells were incubated with 2 μ Ci/ml of *myo*-[³H]inositol (18.6 Ci/mmol, PerkinElmer Life Sciences) for 18 h. Thereafter, cells were washed once with serum-free Dulbecco's modified Eagle's medium containing 10 mM LiCl followed by incubation for 1 h at 37 °C. Intracellular IP levels were determined by anion-exchange chromatography as described (18).

Enzyme-linked Immunosorbent Assay and Immunofluorescence Studies—To estimate cell surface expression of receptors carrying an N-terminal hemagglutinin tag, we used an indirect cellular enzyme-linked immunosorbent assay (19). Immunofluorescence studies were carried out to examine the subcellular distribution of the human *GPR33*. COS-7 cells were transferred into 6-well plates containing sterilized glass coverslips and transfected. Approximately 48 h later, cells were fixed, permeabilized with 0.1% Triton X-100 in phosphate-buffered saline, and probed with a monoclonal anti-hemagglutinin antibody (10 μ g of 12CA5/ml in phosphate-buffered saline) and an affinity purified rabbit anti-FLAG antibody (10 μ g/ml in phosphate-buffered saline). The primary antibodies were detected using species-specific TRITC- and fluorescein isothiocyanate-labeled secondary antibodies (Sigma). Fluorescence images were obtained with a confocal laser-scanning microscope (LSM 510, Carl Zeiss Jena, Jena, Germany).

RESULTS AND DISCUSSION

GPR33 Pseudogenes Are Polymorphic in Human and Several Ape Species—Here, we followed the evolutionary trail of an orphan chemoattractant GPCR, *GPR33*, which was originally identified as a pseudogene in humans (2). The human *GPR33* gene contains a stop codon within the coding sequence of the second intracellular loop that leads to a premature termination of translation (see supplemental materials Fig. S1). To investigate if the inactivation of *GPR33* is recent in primate evolution, we amplified and sequenced the *gpr33* orthologs in the great apes and gibbons. Surprisingly, different inactivating nonsense mutations were found together with intact variants in our samples of chimpanzees, orangutans, and siamangs (Fig. 1, Table I). The positions of most stop codons within the *gpr33* genes were different from the one (Stop140) truncating the human *GPR33*. However, the same Stop140 mutation as in humans was also identified in two chimpanzee individuals

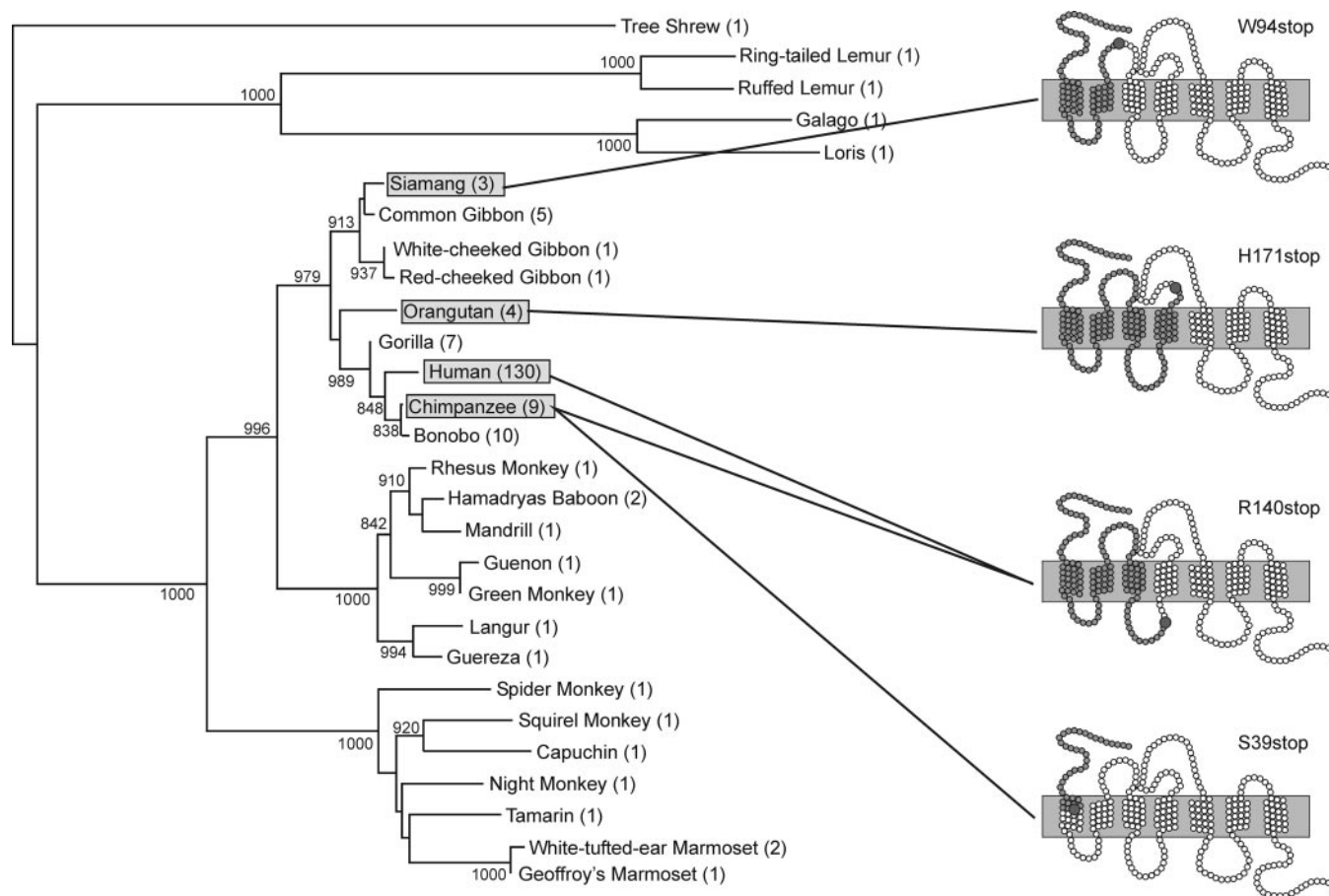


FIG. 1. **GPR33 pseudogenization in humans and great apes.** The nucleotide sequences (nucleotide positions 64 to 997 referred to the sequence of the human *GPR33*) of 27 primate *GPR33* receptors (for accession numbers see supplemental materials Table S1) were aligned with Clustal X (11). To reduce the complexity of the tree, positions that differ between alleles within a species were degenerated. The sequence of the tree shrew was used as the outgroup. The number of individuals investigated is shown in parentheses. Based on the multiple alignment a neighbor-joining tree was constructed and bootstrap values are the result of 1,000 neighbor-joining replicates. Only relevant bootstrap values (>750) that statistically support clades are shown. The locations of stop mutations are depicted in schemes of *GPR33*, the missing receptor parts are shown in white.

TABLE I
GPR33 allele frequency in Hominoidea

Polymorphic *GPR33* inactivation (bold) was found in humans (see Fig. 2 and supplemental materials Table S3 for sample origins), chimpanzees, orangutans, and siamangs. The allele combinations found and the numbers of individuals carrying them were determined by direct sequencing the *GPR33* coding region from species and populations indicated.

| Species (number of individuals investigated) | Allele 1-allele 2 | Number of individuals |
|--|--|--------------------------|
| Human (1217) | Arg ¹⁴⁰ -Arg ¹⁴⁰ | 1 |
| | Arg ¹⁴⁰ - Stop140 Stop140-Stop140 | 50 1166 |
| Chimpanzee (9) | Ser ³⁹ /Arg ¹⁴⁰ -Ser ³⁹ /Arg ¹⁴⁰ | 5 |
| | Ser ³⁹ /Arg ¹⁴⁰ - Stop39 /Arg ¹⁴⁰ | 2 |
| | Ser ³⁹ / Stop140-Stop39 /Arg ¹⁴⁰ | 2 |
| Orangutan (4) | His ¹⁷¹ -His ¹⁷¹ | 2 |
| | His ¹⁷¹ - Stop171 | 1 |
| | Stop171-Stop171 | 1 |
| Siamang (3) | Trp ⁹⁴ - Stop94 | 3 |

(Table I). Because a polymorphism in the ancestor of humans and chimpanzees is highly unlikely to persist for over 5 million years (20), this is likely to be the result of two mutations at the same site. Two observations strengthen this interpretation: first, the site is at a CpG, a dinucleotide known to experience an elevated mutation rate (21). Second, an allelic variant at another site (silent C to T transition codon 305) is in complete linkage disequilibrium with the Stop140 mutation in the chim-

panzee sample, but is absent from the human sample. Thus, it appears that the Stop140 mutation was introduced independently in the human and chimpanzee lineages in the last 5 million years after the human-chimpanzee split (20). An extensive sequence analysis of 17 additional primates did not reveal any other *GPR33* orthologs with an obvious gene inactivation (Fig. 1), suggesting that *gpr33* pseudogenization is restricted to hominoids among primates.

Because the Stop140 was introduced in the human *GPR33* after speciation, we determined the Stop140 allele frequency in 1217 individuals representing all major linguistic groups by DdeI restriction analysis (see "Experimental Procedures"). The inactive allele (TGA, Stop140) was present in all parts of the world at high frequency (Fig. 2, Table I). Thus, it is likely that *GPR33* inactivation occurred before the last migration of Africa about 50,000–100,000 years ago (20). Strikingly, we identified individuals that were heterozygous ($n = 50$) or homozygous ($n = 1$) for an allele resistant to DdeI cleavage. In all latter individuals the entire *GPR33* coding region was sequenced. Sequencing revealed the Stop140 allele and the ancestral sequence of codon 140 (CGA, Arg¹⁴⁰, intact receptor protein) in heterozygous individuals. One individual was homozygous for the ancestral allele. Fig. 2 shows the geographic distribution of the *GPR33* pseudogene and the intact allele that was found with a 2.1% frequency in 2434 alleles screened (Table I).

GPR33 Pseudogenes Predominantly Occur in Rodents and Primates but Not in Other Mammals—To find out if *GPR33*

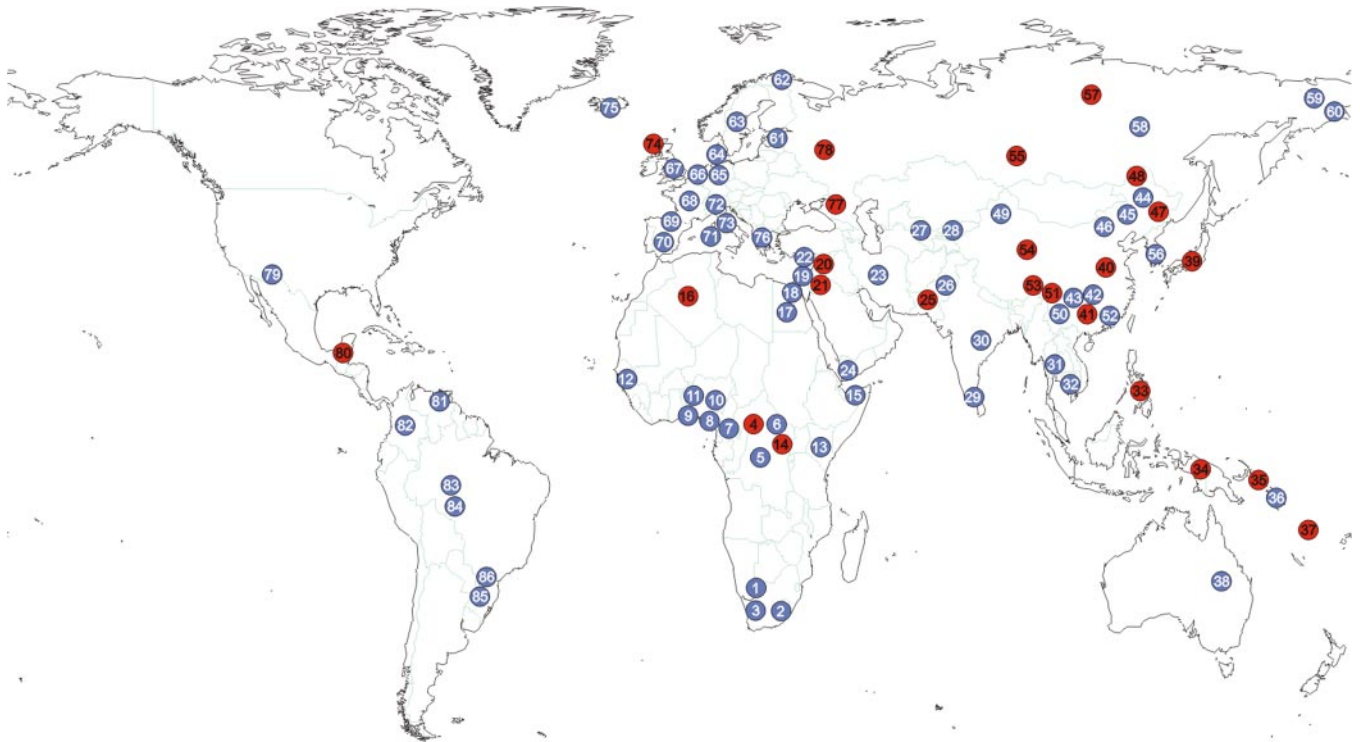


FIG. 2. **Geographic distribution of human *GPR33* alleles.** To analyze the geographic distribution of the wild-type and the Stop140 alleles of *GPR33* in human populations, DdeI restriction analysis was performed for 1217 individuals (for details see “Experimental Procedures”). The numbering refers to a list of samples given in supplemental materials Table III and the approximate geographic origin of each sample group is indicated. Sample groups in which only the Stop140 allele was found are marked in *blue*. The presence of both alleles, the wild-type and the Stop140, are marked in *red*.

inactivation is unique to hominoids, we set out to identify *GPR33* orthologs from all vertebrate classes (mammals, birds, reptiles, amphibians, and fishes) by an extensive PCR-based ortholog cloning strategy (22). *GPR33* orthologs were found in eutherians and marsupials (supplemental materials Table S1) but not in other vertebrate clades tested. Data base search of the available fish genomes (zebrafish, fugu, and tetraodon) and the chicken genome revealed no putative *GPR33* homologues. The absence of *GPR33* in amphibians, birds, and reptiles implies that the receptor is less than ~310 million years old (23). All attempts to amplify the *gpr33* sequence from two monotremes, platypus and echidna, failed. This in turn implies that *GPR33* arose 125–190 million years ago (24, 25) or alternatively, was lost in the monotreme lineage.

Next, we analyzed *GPR33* orthologs from 121 species representing 16 of a total of 18 existing eutherian orders (26). *Gpr33* pseudogenes generated by frame shifting deletions or insertions were found in two gerbil species, the laboratory rat, the Syrian hamster, the European mole and, as a polymorphic stop mutation, in the yellow-toothed cavy. In other ordinal clades no obvious *GPR33* inactivation was found. It has been estimated that about 65% of all inactivating mutations naturally found in GPCRs are missense mutations that can only be identified by functional testing (27). The other third is obvious because of stop codons, deletions, or insertions. Taking in account that 5 of 25 rodent species and 4 of 27 primate species present an apparent inactivation of *GPR33*, we would have expected to see at least one obvious inactivation in Carnivora (19 species) and Cetartiodactyla (20 species) if *gpr33* pseudogenes are as frequent in these orders as in rodents and primates ($p < 0.01$, Fisher’s Exact Test, comparing Carnivora and Cetartiodactyla to primates and rodents).

Age Estimation of *GPR33* Pseudogenization in Hominoids and Rats—*GPR33* pseudogenization events in hominoids and rats took place after divergence from their closest living rela-

tive. To further delimit the time point of *GPR33* inactivation in great apes and humans, we sequenced the 5’ and 3’ non-coding genomic regions flanking intact and pseudogene *GPR33* alleles of human, chimpanzee, and bonobo. Comparison of ~5.4-kb homologous genomic sequence revealed an average of 9.6 mutations/1,000 bp between chimpanzee and human alleles (supplemental materials Table S6). Assuming that half of the mutations occurred in humans and chimpanzees, respectively, and a split of both species 5 million years ago, there is an average mutation rate of 10^{-9} per bp per year similar to previous studies (20). The human intact and pseudogene haplotypes differ from each other by 2.2 substitutions/1,000 bp. Thus, by comparison to human-chimpanzee divergence, we would estimate the inactivation of the *GPR33* in human to be less than 1.1 million years old (Table II). Similarly, the fact that the bonobo *gpr33* allele does not carry one of these stop mutations implicates an inactivation in chimpanzees <800,000 years ago (28). This is consistent with our finding of 1.1–1.3 substitutions per kb between intact and pseudogene alleles in chimpanzees, whereas the pairwise divergence between bonobo and chimpanzee is 2.6 to 3.5 substitutions per kb (supplemental materials Table S6).

We can also estimate the age of the disruption mutation from its frequency in the sample. If we assume the pseudogene allele is neutral, we estimate an age of ~1 million years in humans and a very similar time in other ape species (Table II). A second approach to estimate the age of a neutral allele led to similar results (Table II). If, instead, we assume that it is weakly favored, the age could be as recent as ~50,000 years in humans. In general, the estimate of the age decreases with increasing strength of selection.

Sequence analysis of *gpr33* in 10 laboratory rats (*R. norvegicus*) revealed an inactivation caused by a 14-bp deletion. In the black rat (*R. rattus*), which diverged from *R. norvegicus* ~5.5 million years ago (29), *GPR33* is intact (five analyzed

TABLE II
Age estimates for the pseudogene alleles

| Species | No. of chromosomes surveyed | No. of copies of pseudogene allele | Estimate of N_e^a | Expected age in years if neutral | Expected age in years if favored ^b | Estimate of age using GENETREE | Estimate of age from divergence between allelic classes ^c |
|------------|-----------------------------|------------------------------------|---------------------|----------------------------------|---|--------------------------------|--|
| Human | 170 ^d | 164 | 10,000 | 970,900 | 48,150 | 800,000–1,150,000 ^e | 1,145,000 |
| Chimpanzee | 18 | 4 ^f | 30,000 | 968,040 | 22,080 | 530,000–840,000 ^e | 572,000–677,000 |
| Orangutan | 8 | 3 | 30,000 | 1,273,680 | 24,396 | 720,000 | ND ^g |

^a The expected age of a selected allele at a given population frequency was evaluated under a model of genic selection and under a neutral model (see “Experimental Procedures”) based on N_e values for humans (42) and for chimpanzee and orangutan (20). We assume that the generation time for humans is 25 years (43), whereas for chimpanzees and orangutans it is 20 years.

^b Assuming that the selective advantage of the favored allele is 0.005.

^c Estimate of the allele age on the basis of mutation rates (see text and supplemental materials Table S6).

^d The entire coding region of *GPR33* was sequenced from 85 individuals of a human genomic DNA panel used in Ref. 8. Except of the wild-type and the Stop140 alleles no further sequence variations or evidences of recombination were observed within the coding region of *GPR33*. Age estimates were performed on the basis of these sequencing results.

^e There is also a distinct gene disruption mutation present twice in the sample. We estimated the expected age of a selected allele (**Stop39/Arg¹⁴⁰**) ignoring interference between them.

^f The two estimates are for the two positions at which the mutation could occur in the genealogy. Please note that the GENETREE approach assumes no recombination. Whereas there is no evidence for recombination in the coding region of *GPR33*, this does not preclude that recombination occurred in the history of the sample. Thus, the results of this analysis should be interpreted with care.

^g ND, not determined.

```

398                               438
CACTGTTGGTCCCAGCAGC-----ACCGAACCCACACTGGGCTTCC  M. musculus
CGGTTTGGTCCCAGCAGC-----ACCGAACCCACAGCTGGGCTTCC  R. rattus
CGGTTGTTGGTCCCAGCAGC-----TGGGCTTCC              R. norvegicus A
CGGTTGTTGGTCCCAGCAGCACCGAACCCACGACCGAACCCACAGCTGGGCTTCC  R. norvegicus B

```

FIG. 3. An unequal meiotic crossover event caused inactivation of *gpr33* in *R. norvegicus*. Sequence analysis of the *gpr33* ortholog from *R. norvegicus* revealed two alleles. Allele A contains a 14-bp deletion and allele B a 14-bp insertion that represents the deleted portion of allele A. Sequences were compared with *GPR33* orthologs from *Mus musculus* and *R. rattus*. The numbering is referred to the *gpr33* coding sequence from *R. rattus*. Positions that differ from the sequence of *R. rattus* are boxed.

individuals from Germany and Paraguay). Interestingly, we identified a second *gpr33* allelic variant with a 14-bp insertion in wild-caught *R. norvegicus*, which exactly matches the deleted portion of the first allele discovered (Fig. 3). This unique finding suggests a simultaneous inactivation of both alleles in *R. norvegicus* by an unequal meiotic crossing over event leading to a reciprocal 14-bp deletion and duplication. We took advantage of this finding to date the pseudogenization event by sequencing 4 kb of the 5' and 3' non-coding genomic regions flanking both *R. norvegicus* alleles and the *R. rattus* allele (supplemental materials Table S6). Based on a comparison of the sequence differences between both *R. norvegicus* alleles and the divergence from *R. rattus*, we estimate that the inactivation occurred less than 0.7 million years ago. Similar to the human *GPR33* pseudogene, *GPR33* inactivation is probably fixed worldwide in *R. norvegicus* populations as indicated by the presence of both pseudogene alleles in 127 wild-caught *R. norvegicus* analyzed from Germany, Alaska, Russia, and Japan.

***GPR33* Has Structural and Functional Features of an Chemokine/Chemoattractant GPCR**—To gain more functional information, we set out to clarify the tissue expression pattern and signal transduction of *GPR33*. First, we analyzed the expression pattern of the intact *GPR33* in mouse tissue by reverse transcriptase-PCR (for details see supplemental materials). The mouse *Gpr33* mRNA is predominantly expressed in lung, spleen, and testis (Fig. 4A) similar to those found for several formylmethionylleucylphenylalanine and chemokine receptors (30, 31). In consensus with a predominant expression in immunologically relevant tissues *Gpr33* transcripts were identified in RAW 264.7 cells, a murine macrophage cell line (data not shown). The persistence of expression in humans and rats may also be indicative to a recent gene inactivation. To date, there are only a few pseudogenes found to be expressed (32). As found for the mouse *Gpr33*, transcripts of the *gpr33* pseudogene of *R. norvegicus* were amplified from spleen cDNA and its identity was verified by sequencing (for details see

supplemental materials). Similarly, the human *GPR33* pseudogene is still transcribed in spleen and lung but also in heart, liver, kidney, pancreas, thymus, gonads, and leukocytes (Fig. 4B and supplemental materials).

Most chemoattractant and chemokine-like GPCR mediate their signal transduction via coupling to $G_{i/o}$ proteins, which inhibit adenylyl cyclases and, therefore, decrease intracellular cAMP levels (1). According to the current model of GPCR function, receptor overexpression can result in constitutive activation of signaling pathways. Thus, coupling abilities of several receptors, including “orphan” receptors, have been characterized by overexpression in the absence of agonist (22).

As a member of the rhodopsin-like GPCR family, *GPR33* presents common scaffold residues such as the Asp-Arg-Tyr (DRY) motif located at the transmembrane domain 3/intracellular loop 2 transition. The crystal structure of rhodopsin proposes that the acetic residue in the DRY motif forms a salt bridge to Arg and probably acts as a proton acceptor during receptor activation (33). Interestingly, all *GPR33* orthologs cloned from genus *Mus* display an ARY motif, whereas all other *GPR33* orthologs contain the classical DRY motif at the transmembrane domain 3/intracellular loop 2 transition (Fig. 5A). There are numerous reports showing that Ala mutation of the acetic residue in the DRY motif results in constitutive activity of many receptors (34). To test the functional consequence of the Asp to Ala mutation in murine *Gpr33* orthologs we utilized second messenger assays. As shown in Fig. 5B, basal activity of the human ADP receptor (P2Y₁₂), which served as positive control (22), and the mouse *Gpr33* were ~5- and ~3-fold increased, respectively, whereas green fluorescent protein-transfected or human *GPR33*-transfected cells displayed no significant changes in basal IP levels. To control whether the high basal activity of the murine *Gpr33* is indeed because of the mutational change in the DRY motif, Ala¹²⁵ in the mouse ARY motif was reversed to Asp. Replacement of Ala¹²⁵ by the Asp residue resulted in a decrease of basal receptor activity (Fig. 5B). Plasma membrane expression of A125D was almost un-

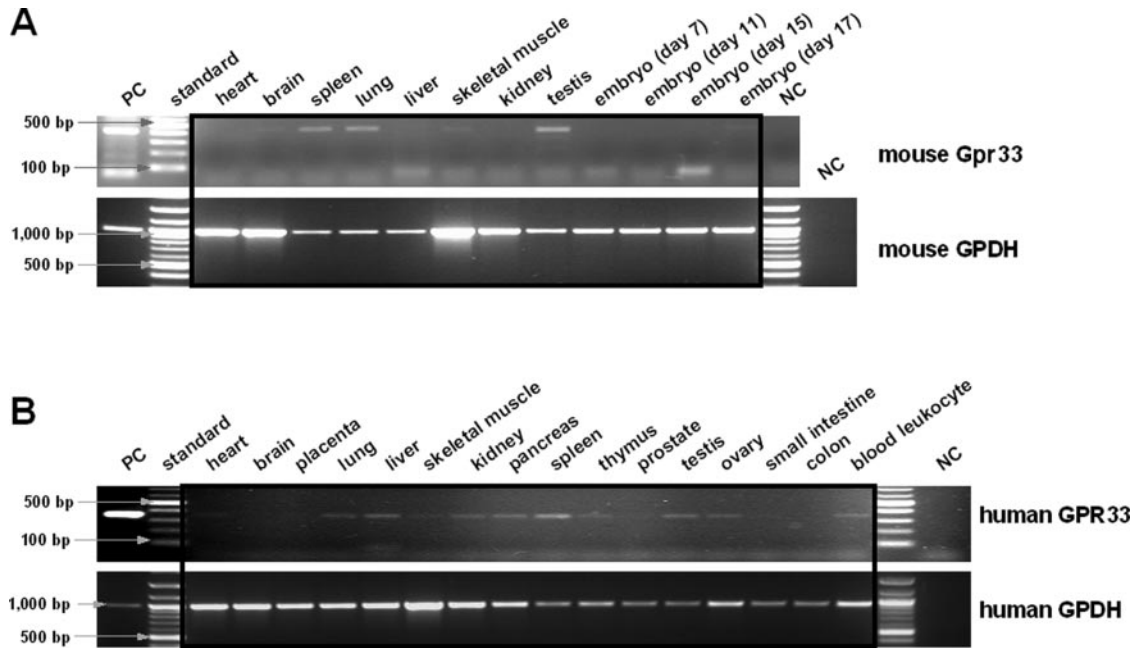


FIG. 4. **Expression pattern of GPR33 in mouse and human tissues.** The expression pattern of GPR33 was determined by reverse transcriptase-PCR. Thus, GPR33-specific primer pairs and the mouse and human cDNA panels (Clontech) were used (for details, see supplemental materials). Glyceraldehyde phosphate dehydrogenase (*GPDH*) expression was monitored in parallel. *NC*, negative control; *PC* positive control.

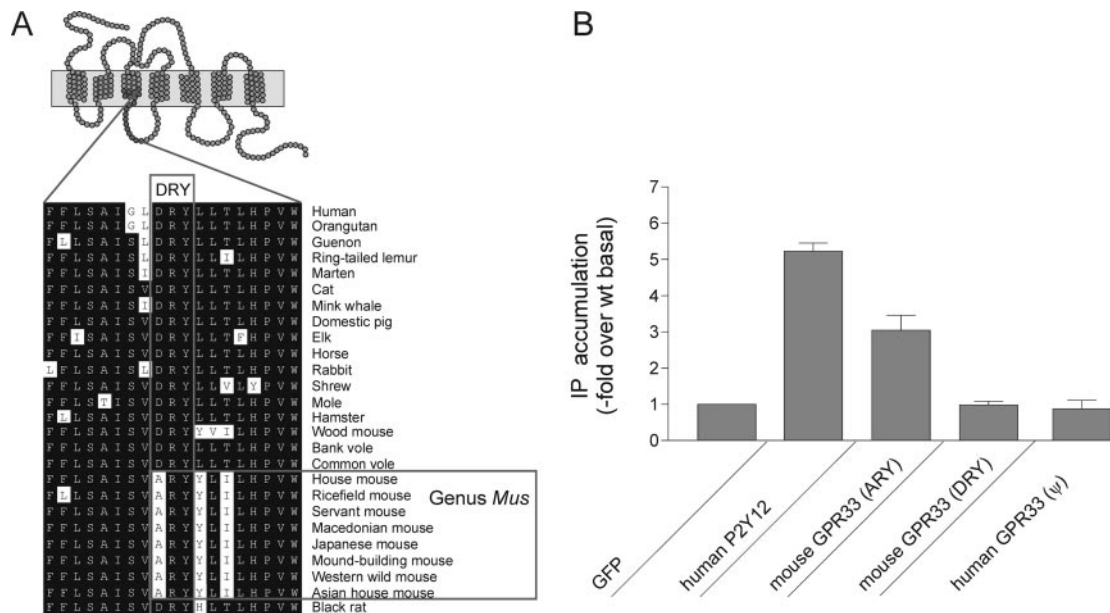


FIG. 5. **The murine *Gpr33* ortholog constitutively activates the G_i proteins.** *A*, the transmembrane domain 3/intracellular loop 2 transition of various GPR33 orthologs is shown, which contains the functionally important DRY motif. Interestingly, GPR33 orthologs from different murine species display an exchange of Asp¹²⁵ to Ala. *B*, most chemoattractant and chemokine-like GPCR mediate their signal transduction via coupling to $G_{i/o}$ proteins. It has been demonstrated that replacement of the four C-terminal amino acids of $G\alpha_q$ with the corresponding $G\alpha_i$ residues (referred to as $G\alpha_{\Delta 6qi4myr}$) confers the ability to stimulate the phospholipase C- β pathway onto G_i -coupled receptors (41). Therefore, receptor constructs were co-expressed with the chimeric $G\alpha_{\Delta 6qi4myr}$ protein in COS-7 cells and IP assays were performed as described (“Experimental Procedures”). As controls, plasmids encoding green fluorescent protein (*GFP*), the human P2Y₁₂, and the human GPR33 pseudogene were co-transfected with $G\alpha_{\Delta 6qi4myr}$. Basal IP formation is expressed as -fold over basal levels of green fluorescent protein-transfected cells (347 ± 8.6 cpm/well). Data are presented as mean \pm S.E. of three independent experiments, each carried out in triplicate.

changed ($110 \pm 14\%$) when compared with the wild-type murine *Gpr33* (100%) as determined by an indirect cellular enzyme-linked immunosorbent assay. These data indicate that the decreased constitutive activity of A125D is indeed caused by stabilizing the inactive receptor conformation and not by a reduction of the receptor number at the cell surface.

Potential Evolutionary Modes of GPR33 Inactivation—Although the existence of GPR33 provides an advantage as reflected by its presence in all mammalian orders, it was recently

inactivated in humans, as well as in a subset of ape and rodent species. Two scenarios have to be discussed that may explain GPR33 inactivation: GPR33 inactivation occurred because of a loss of constraint or because of positive selection.

The “loss of constraint” scenario may occur because: 1) loss of the endogenous agonist for GPR33; 2) loss of an exogenous agonist; and 3) the development of functional redundancy and mechanisms that compensate for loss of GPR33 function. Following gene inactivation by one of these reasons, neutral drift

can lead to fixation of the inactive allele. However, scenarios 1 and 3 are very unlikely because they are not compatible with an almost simultaneous inactivation of GPR33 in several unrelated species but not all mammalian orders. The loss of a potential exogenous/environmental factor (point 2) that somehow participates in GPR33 function is more suitable to explain the timely synchronized pseudogenization in several unrelated species. For example, most GPCR pseudogenes are found in mammalian odorant receptors (35). It is assumed that the repertoire of functional and non-functional olfactory receptors is triggered by the odor composition of the environment and the development of compensatory mechanisms (36). The disappearance of an odorant releases the specific olfactory GPCR from its constraint. Similarly, loss of a potential exogenous factor that interacts with GPR33 should also release the constraint on this GPCR. Such a mechanism would at least account for the simultaneous inactivation in several unrelated mammals. However, GPR33 inactivation is restricted to only a small subset of apes and rodents and is kept intact in most mammalian orders that share the environment with those species harboring a *gpr33* pseudogene. But we cannot rule out the theoretical possibility that a potential exogenous/environmental factor was specific to only humans, great apes, and some rodents. Even if one follows the argument that there was a loss of constraint only in a distinct subset of species it appears very unlikely that the drift led to fixation of different nonsense mutations in different GPR33 orthologs in the past one million years. This is because stop mutations are a rare event. This is reflected by the fact that only 8% of all inactivating mutations in GPCRs are stop codons but almost 65% are missense mutations (27). This calculation does not include all missense mutations that do not interfere with proper receptor function. So, one would expect at least 8 missense mutations per one stop mutation. By contrast, only 1–3 missense mutations became fixed after speciation in GPR33 orthologs of great apes but 3 of 5 orthologs contain a stop codon. This also implicates that there are several inactivating missense mutations that escape detection unless orthologs have been functionally tested.

In a second scenario, *gpr33* pseudogenization provides some advantage and, therefore, is positively selected. A likely cause of chemokine receptor inactivation by selection is its interplay with an exogenous factor. For example, the chemokine receptor CCR5 and the Duffy antigen, both members of the superfamily of GPCRs, act as co-receptors for cell entry of pathogens: CCR5 for the internalization of HIV-1 and Duffy for entry of *P. vivax*. Mutational inactivation of CCR5 and Duffy are associated with resistance to infections by HIV-1 (5, 6) and by *P. vivax* (7), respectively. However, HIV-1 has not infected humans long enough to account for the 10% frequency of inactive allele (CCR5-Δ32) in the European population. Other reasons have been proposed, such as plague and small pox that selected for the null allele (37). The expression of GPR33 in macrophages, lung, and spleen (see Fig. 4), its relation to formylmethionylleucylphenylalanine and chemokine receptors, and its coupling to G_{i/o} protein (see Fig. 5B) all suggest that GPR33 most likely functions as a chemoattractant receptor. Unfortunately, all our attempts to identify the endogenous agonist by functional screening of potential ligands, fractionated tissue, and bacterial extracts failed so far. Despite its close structural relation to other co-receptors for retroviruses, the introduction of HIV into the human population occurred too recently to explain the frequency of GPR33 inactivation. Additionally, GPR33 inactivation was found also in non-primate species, which are usually not the host of hominoid-trope retroviruses. In contrast to GPR33, the inactive CCR5 and Duffy allelic variants display a strong geographic and ethnic restriction to Eurasia and Central Africa, respectively. Therefore,

it is more likely that the inactivation of GPR33 conferred a selective advantage to hosts exposed to another probably older and more widespread pathogen.

Assuming a positive selection of *gpr33* pseudogenes one has to search for signatures that may support this scenario. The recent fixation of a beneficial substitution of an allele at one site leads to a reduction in variation and a high proportion of rare alleles at linked, neutrally evolving sites (38). To address this point we sequenced the *GPR33* coding region (~1 kb) and 3.8 kb of the 3'-untranslated region from 21 Yoruba individuals. For reference purposes the same region was analyzed in a Philippine individual (homozygous for the ancient CGA allele) and in chimpanzee. Sequence data analysis revealed neither of the expected signatures of a recent selection (supplemental materials Table S7). Diversity levels are $\pi = 13.3 \times 10^{-4}$, which is not unusually low by comparison to π values (mean $\pi = 9.3 \pm 4.2 \times 10^{-4}$) estimated in 24 African-Americans for 217 genic regions (pga.gs.washington.edu/summary_stats.html) (39). Moreover, commonly used summaries of the allele frequencies, such as the D-tests from Tajima (9) ($D = -0.121$) and Fu and Li (10) ($D = 1.01$) show no marked departures from the expectations of the standard neutral model, nor were the D values unusual by comparison to values obtained in the same survey of African-Americans. Taken together, the polymorphism data do not suggest a recent selective sweep. However, theoretical investigations suggest that if the fixation event occurred prior to 100,000 years ago, the signature will have weakened considerably and may no longer be detectable (40).

Conclusion—After introduction of the chemokine receptor GPR33 into the mammalian genome more than 125–190 million years ago this receptor underwent pseudogenization in humans, other hominoids, and some rodents. This process is still ongoing, as reflected by a polymorphic existence of intact alleles and pseudogenes, e.g. in a worldwide sample of humans. Simultaneous pseudogenization because of loss of constraint and neutral drift that leads to fixation of pseudogenes in several species appear to be very unlikely in less than one million years. Given the fact that inactivation occurred independently and at similar times in several species of unrelated orders, the GPR33 inactivation is likely because of selection. A likely cause of GPR33 inactivation by selection is its interplay with an exogenous factor, e.g. a rodent-hominoid-tropic pathogen. If so, the finding that an appreciable fraction of humans still harbor an intact copy of the gene may have important medical implications.

Acknowledgments—We thank the numerous contributors for the species samples (supplemental materials Table S1). We are grateful to Mario Mörl and Michael Hofreiter for many suggestions and critical reading of the manuscript.

REFERENCES

1. Thelen, M. (2001) *Nat. Immunol.* **2**, 129–134
2. Marchese, A., Nguyen, T., Malik, P., Xu, S., Cheng, R., Xie, Z., Heng, H. H., George, S. R., Kolakowski, L. F., Jr., and O'Dowd, B. F. (1998) *Genomics* **50**, 281–286
3. Samson, M., Edinger, A. L., Stordeur, P., Rucker, J., Verhasselt, V., Sharron, M., Govaerts, C., Mollereau, C., Vassart, G., Doms, R. W., and Parmentier, M. (1998) *Eur. J. Immunol.* **28**, 1689–1700
4. Wittamer, V., Franssen, J. D., Vulcano, M., Mirjolet, J. F., Le Poul, E., Migotte, I., Brezillon, S., Tyldesley, R., Blanpain, C., Detheux, M., Mantovani, A., Sozzani, S., Vassart, G., Parmentier, M., and Communi, D. (2003) *J. Exp. Med.* **198**, 977–985
5. Samson, M., Libert, F., Doranz, B. J., Rucker, J., Liesnard, C., Farber, C. M., Saragosti, S., Lapoumeroulie, C., Cogniaux, J., Forceille, C., Muyltermans, G., Verhofstede, C., Burtonboy, G., Georges, M., Imai, T., Rana, S., Yi, Y., Smyth, R. J., Collman, R. G., Doms, R. W., Vassart, G., and Parmentier, M. (1996) *Nature* **382**, 722–725
6. Liu, R., Paxton, W. A., Choe, S., Ceradini, D., Martin, S. R., Horuk, R., MacDonald, M. E., Stuhlmann, H., Koup, R. A., and Landau, N. R. (1996) *Cell* **86**, 367–377
7. Hadley, T. J., and Peiper, S. C. (1997) *Blood* **89**, 3077–3091
8. Kaessmann, H., Heissig, F., von Haeseler, A., and Pääbo, S. (1999) *Nat. Genet.* **22**, 78–81
9. Tajima, F. (1989) *Genetics* **123**, 585–595
10. Fu, Y. X., and Li, W. H. (1993) *Genetics* **133**, 693–709

11. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25**, 4876–4882
12. Felsenstein, J. (2001) *PHYLIP* (Phylogeny Inference Package), version 3.6a2.1., University of Washington, Seattle, WA
13. Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425
14. Griffiths, R. C. (2003) *Theor. Popul. Biol.* **64**, 241–251
15. Stephens, M., Smith, N. J., and Donnelly, P. (2001) *Am. J. Hum. Genet.* **68**, 978–989
16. Griffiths, R. C., and Tavaré, S. (1997) in *Progress in Population Genetics and Human Evolution* (Donnelly, P., and Tavaré, S., eds) pp. 165–182, Springer-Verlag, New York
17. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7**, 256–276
18. Berridge, M. J. (1983) *Biochem. J.* **212**, 849–858
19. Schöneberg, T., Sandig, V., Wess, J., Gudermann, T., and Schultz, G. (1997) *J. Clin. Investig.* **100**, 1547–1556
20. Kaessmann, H., and Pääbo, S. (2002) *J. Intern. Med.* **251**, 1–18
21. Cooper, D. N., and Krawczak, M. (1989) *Hum. Genet.* **83**, 181–188
22. Schulz, A., and Schöneberg, T. (2003) *J. Biol. Chem.* **278**, 35531–35541
23. Kumar, S., and Hedges, S. B. (1998) *Nature* **392**, 917–920
24. Ji, Q., Luo, Z. X., Yuan, C. X., Wible, J. R., Zhang, J. P., and Georgi, J. A. (2002) *Nature* **416**, 816–822
25. Woodburne, M. O., Rich, T. H., and Springer, M. S. (2003) *Mol. Phylogenet. Evol.* **28**, 360–385
26. Wilson, D. E., and Reeder, D. M. (eds) (1993) *Mammal Species of the World*, Smithsonian Institution Press, Washington, D. C.
27. Schöneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Römpler, H., and Sangkuhl, K. (2004) *Pharmacol. Ther.* **104**, 173–206
28. Fischer, A., Wiebe, V., Pääbo, S., and Przeworski, M. (2004) *Mol. Biol. Evol.* **21**, 799–808
29. Dubois, J. Y., Jekel, P. A., Mulder, P. P., Bussink, A. P., Catzeflis, F. M., Carsana, A., and Beintema, J. J. (2002) *J. Mol. Evol.* **55**, 522–533
30. Vaughn, M. W., Proske, R. J., and Haviland, D. L. (2002) *J. Immunol.* **169**, 3363–3369
31. Migeotte, I., Franssen, J. D., Goriely, S., Willems, F., and Parmentier, M. (2002) *Eur. J. Immunol.* **32**, 494–501
32. Mighell, A. J., Smith, N. R., Robinson, P. A., and Markham, A. F. (2000) *FEBS Lett.* **468**, 109–114
33. Okada, T., Ernst, O. P., Palczewski, K., and Hofmann, K. P. (2001) *Trends Biochem. Sci.* **26**, 318–324
34. Schöneberg, T., Schulz, A., and Gudermann, T. (2002) *Rev. Physiol. Biochem. Pharmacol.* **144**, 143–227
35. Mombaerts, P. (2001) *Annu. Rev. Genomics Hum. Genet.* **2**, 493–510
36. Gilad, Y., Man, O., Pääbo, S., and Lancet, D. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3324–3327
37. Galvani, A. P., and Slatkin, M. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15276–15279
38. Bamshad, M., and Wooding, S. P. (2003) *Nat. Rev. Genet.* **4**, 99–111
39. Akey, J. M., Eberle, M. A., Rieder, M. J., Carlson, C. S., Shriver, M. D., Nickerson, D. A., and Kruglyak, L. (2004) *PLoS Biol.* **2**, e286
40. Przeworski, M. (2002) *Genetics* **160**, 1179–1189
41. Kostenis, E. (2001) *Trends Pharmacol. Sci.* **22**, 560–564
42. Przeworski, M., Hudson, R. R., and Di Rienzo, A. (2000) *Trends Genet.* **16**, 296–302
43. Helgason, A., Hrafnkelsson, B., Gulcher, J. R., Ward, R., and Stefansson, K. (2003) *Am. J. Hum. Genet.* **72**, 1370–1388