THE RISE AND FALL OF THE CHEMOATTRACTANT RECEPTOR GPR33

Holger Römpler¹, Angela Schulz¹, Christian Pitra², Graham Coop³, Molly Przeworski⁴, Svante Pääbo⁵, and Torsten Schöneberg¹

From the ¹Institute of Biochemistry, Molecular Biochemistry, Medical Faculty, University of Leipzig, Leipzig, Germany; ²Institute for Zoo and Wildlife Research, Department of Evolutionary Genetics, Berlin, Germany; ³Department of Statistics, University of Oxford, Oxford, United Kingdom; ⁴Department of Ecology and Evolutionary Biology, Brown University, Providence, USA; ⁵Max-Planck-Institute of Evolutionary Anthropology, Leipzig, Germany

Running title: GPR33 evolution
Address correspondence to: Torsten Schöneberg, 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

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 which was previously identified as a pseudogene in humans. GPR33 evolved in mammals about 125-190 million years (Myr) ago. The cloning and analysis of more than 120 mammalian GPR33 orthologs from 16 out 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 harboured 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 one million years (Myr). 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 Myr. The coincidental inactivation and its fixation in several species of distantly related mammalian orders suggest a selective pressure on this chemoattractant receptor gene.

INTRODUCTION

Recruitment of cells involved in the cellular defense against pathogens is mediated by an armada of cell surface proteins, chemokines, complement factors and leucotriens. The receptors that mediate the action of these factors all belong to the large family of G-protein-coupled receptors (GPCR), and constitute 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 which 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 co-receptor (3, 4), N-formyl-peptide (fMLP) 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 evidences 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 \textit{P. vivax}, respectively (5-7).

**EXPERIMENTAL PROCEDURES**

**Cloning of GPR33 orthologs** - To identify 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 labs (Suppl. 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 hours. DNA was purified by phenol/chloroform extraction and ethanol precipitation. Based on the human and mouse GPR33 sequences (2) sets of degenerated primer pairs (Suppl. Table S2) were applied to amplify GPR33 specific sequences. PCR reactions 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 sequencer (Applied Biosystems). Based on considerable sequence similarities of the 5’- and 3’-untranslated region (UTR) of GPR33 genes primers were designed (Suppl. Table S2) which 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 an one-base mismatch antisense primer (DdeI-AS 5’-AATGCTGGAAGCCAGCGGGGC-3’) which introduces a new DdeI restriction site in the PCR product if the TGA codon 140 is present. Together with a sense primer (DdeI-S 5’-CTTGAACACTTGTGAACCTT-3’) PCR reaction 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 case of the TGA allele the 166-bp PCR product was cleaved into two fragments (139 bp and 27 bp). Uncut fragments were always sequenced. This DdeI restriction analysis was performed for 1217 human individuals (for details see Suppl. Table S3). The DNA sample collection containing most samples of the Centre d’Étude du Polymorphisme Humain (CEPH) 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’-ACTGTGTTTCTCAGTCAGGCAGGC-3’ and hu-33 5’-CTATGGGTATACGGCAAGTGCTG-3’). Genomic DNA (100 ng) was used in PCR reactions (50 µl) with primers (10 pmol each), standard buffer (Perkin Elmer), dNTP (200 µM) and Taq polymerase (1 U; Perkin Elmer). 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 kb 5'- and 3'-flanking regions of human, chimpanzee, bonobo, R. rattus, and R. norvegicus were amplified by PCR and directly sequenced (Suppl. Tables S4 and S5). Similarly, the coding region of the human GPR33 and approx. 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 Tajima’s D-test (9) and the Fu and Li’s D-Test (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, Berlin-Brandenburg), 5 wild-caught R. norvegicus from Russia (Siberia), 6 wild-caught R. norvegicus from 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 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 his approach to evaluate the expected age of an allele given its sample frequency under a model of genic 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 2. Since 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 neutrally, we first used the program PHASE v2.1 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, $\theta$, through neutral coalescent simulations conditional on the full polymorphism data set. The age was evaluated for Watterson's estimate of $\theta$ based on the number of segregating sites in the sample (17).

**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 (HA) and a C-terminal FLAG epitope. GPR33 mutation (A125D) was introduced into the HA-tagged version of the murine GPR33 using a 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 DMEM supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 µg/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 x 10⁵ cells/well) and transfected with a total amount of 1 µg of plasmid DNA/well. To measure inositol phosphate (IP) formation, transfected COS-7 cells were incubated with 2 µCi/ml of myo-[³H]-inositol (18.6 Ci/mmol, NEN) for 18 hours. Thereafter, cells were washed once with serum-free DMEM containing 10 mM LiCl followed by incubation for one hour at 37°C. Intracellular IP levels were determined by anion-exchange chromatography as described (18).

ELISA and immunofluorescence studies - To estimate cell surface expression of receptors carrying an N-terminal HA-tag, we used an indirect cellular ELISA (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 cover slips and transfected. Approximately 48 h later, cells were fixed, permeabilized with 0.1% Triton X-100 in PBS, and probed with a monoclonal anti-HA antibody (10 µg 12CA5/ml in PBS) and an affinity purified rabbit anti-FLAG antibody (10 µg/ml in PBS). The primary antibodies were detected using species specific TRITC- and FITC-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 which leads to a premature termination of translation (see Suppl. 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. 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. Figure 2 shows the geographic distribution of the GPR33 pseudogene and the intact allele which was found with a 2.1% frequency in 2434 alleles screened (Table 1).
GPR33 pseudogenes predominantly occur in rodents and primates but not in other mammals

To find out if GPR33 inactivation is unique to hominoids, we set out to identify GPR33 orthologs from all vertebrate classes (mammals, birds, reptiles, amphibians, fishes) by an extensive PCR-based ortholog cloning strategy (22). GPR33 orthologs were found in eutherians and marsupials (Suppl. Table S1) but not in other vertebrate clades tested. Database search of the available fish genomes (zebrafish, fugu, 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 Myr old (23). All attempts to amplify GPR33 sequence from two monotremes, platypus and echidna, failed. This in turn implies that GPR33 arose 125-190 Myr ago (24, 25) or alternatively, was lost in the monotreme lineage.

Next, we analyzed GPR33 orthologs from 121 species representing 16 out 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 which 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 relative. 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 (Suppl. Table S6). Assuming that half of the mutations occurred in humans and chimpanzees, respectively, and a split of both species 5 Myr 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 Myr old (28). This is consistent with our finding of 1.1-1.3 substitutions per kb between intact and pseudogene alleles in chimpanzees while the pairwise divergence between bonobo and chimpanzee is 2.6 to 3.5 substitutions per kb (Suppl. 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 Myr in humans and a very similar time in other ape species (Table 2). A second approach to estimate the age of a neutral allele led to similar results (Table 2). 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 ten 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 Myr ago (29) GPR33 is intact (five analyzed 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 (Suppl. 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 Myr 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 RT-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 fMLP 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 leucocytes (Fig. 4B and Supplemental Materials).

Most chemoattractant and chemokine-like GPCR mediate their signal transduction via coupling to Goi/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 (TMD3/IL2) transition. The crystal structure of rhodopsin proposes that the acidic 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 TMD3/IL2 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 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-fold and ~3-fold increased, respectively, whereas GFP-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 due to the mutational change in the DRY motif, Ala125 in the mouse ARY motif was reversed to Asp. Replacement of Ala125 by an Asp residue resulted in a decrease of basal receptor activity (Fig. 5B). Plasma membrane expression of Ala125Asp was almost unchanged (110 ± 14 %) when compared with the wild type murine GPR33 (100 %) as determined by an indirect cellular ELISA. These data indicate that the decreased constitutive activity of Ala125Asp 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 due to a loss of constraint or due to positive selection.

The “loss of constraint” scenario may occur due to 1.) loss of the endogenous agonist for GPR33, 2.) loss of an exogenous agonist, and 3.) the development of functional redundancy and mechanisms which 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) which 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 odour 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 which interacts with GPR33 should also release the constraint on this GPCR. Such 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 which share the environment with those species harbouring a GPR33 pseudogene. But we can not rule out the theoretical possibility that a potential exogenous/environmental factor was specific to only humans, great apes, and some rodents. Even if one follow the argument that there was a loss of constraint only in a distinct subset of species it appears very unlikely that drift led to fixation of different nonsense mutations in different GPR33 orthologs in the past one Myr. 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 which 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 out of 5 orthologs contain a stop codon. This also implicates that there are several inactivating missense mutations which 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 Plasmodium 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-∆32) 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 fMLP and chemokine receptors, and its coupling to G₁o 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 hominoidotrope 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 which 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 (Suppl. Table S7). Diversity levels are π = 13.3 x 10^-4, which is not unusually low by comparison to π values (mean π = 9.3 ± 4.2 x 10^-4) estimated in 24 African-Americans for 217 genic regions (http://pga.gs.washington.edu/summary_stats.html) (39). Moreover, commonly used summaries of the allele frequencies, such as Tajima’s D (D = -0.121) and Fu and Li’s D-Test (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 Myr 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 which led to fixation of pseudogenes in several species appear to be very unlikely in less than one Myr. Given the fact that inactivation occurred independently and at similar times in several species of unrelated orders, the GPR33 inactivation is likely due to selection. A likely cause of GPR33 inactivation by selection is its interplay with an exogenous factor, e.g. a rodent-hominoidotopic pathogen. If so, the finding that an appreciable fraction of humans still harbor an intact copy of the gene may have important medical implications.

**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., Goyaerts, C., Mollereau, C., Vassart, G., Doms, R. W., and Parmentier. M. (1998) *Eur. J. Immunol.* 28, 1689-1700
4. Wittamer, V., Fransson, J. D., Vulcano, M., Mirjolet, J. F., Le Poul, E., Migeotte, 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., Cognaux, J., Forceille, C., Muyldermans, 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. Dept. of Genetics,
University of Washington, Seattle

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) (Springer-Verlag, New York), pp 165-182

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. Invest.* **100**, 1547-1556

20. Koessler, 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. Mol. Biol.* **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 D. M. Reeder (eds) (1993) *Mammal Species of the World.* (Smithsonian Institution Press)

27. Schöneberg, T., Schulz, A., Biebermann, H., Hermsdorf, T., Römler, 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. Vaughan, 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. Mommaerts, 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. Basmad, 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., Hrafnkellson, B., Gulcher, J. R., Ward, R., and Stefansson, K. (2003) *Am. J. Hum. Genet.* **72**, 1370-1388

**FOOTNOTES**

*Acknowledgments*

We would like to thank the numerous contributors for the species samples (Suppl. Table S1). We are grateful to Mario Mörl and Michael Hofreiter for many suggestions and critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft, Bundesministerium für Bildung und Forschung, Fonds der Chemischen Industrie and IZKF Leipzig.
Abbreviations
CEPH panel, Centre d’É tude du Polymorphisme Humain panel; fMLP, N-formyl-peptide; GPCR, G-protein-coupled receptor; HA, hemagglutinin; IL, intracellular loop; IP, inositol phosphate; Myr, million years; TMD, transmembrane domain

Data deposition
The sequences reported in this paper have been deposited in the GeneBank database (accession no. AY490569-AY490743, AY493989-AY494004, AY502103, AY502104, AY528865, Suppl. Table S1)

Supporting Information
Suppl. Fig. S1
Suppl. Tables S1-S7
FIGURE LEGENDS

Fig. 1. GPR33 pseudogenization in humans and great apes. The nucleotide sequences (nt position 64 to 997 referred to the sequence of the human GPR33) of 27 primate GPR33 receptors (for accession numbers see Suppl. Table S1) were aligned with Clustal X (11). To reduce the complexity of the tree, positions which differ between alleles within a species were degenerated. The sequence of the tree shrew was used as 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) which statistically support clades are shown. The locations of stop mutations are depicted in schemes of GPR33, the missing receptor parts are shown in white.

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 Dde1 restriction analysis was performed for 1217 individuals (for details see Experimental Procedure). The numbering refers to a list of samples given in Suppl. Table 3 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.

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 which 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 which differ from the sequence of R. rattus are boxed.

Fig. 4. Expression pattern of GPR33 in mouse and human tissues. The expression pattern of GPR33 was determined by RT-PCR. Thus, GPR33 specific primer pairs and the mouse and the human cDNA panels (Clontech) were used (for details see Supplemental Materials). 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 TMD3/IL2 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_ion proteins. It has been demonstrated that replacement of the four C-terminal amino acids of G_ion with the corresponding G_ion residues (referred to as G_ion) confers the ability to stimulate the PLC-β pathway onto G_ion-coupled receptors (41). Therefore, receptor constructs were co-expressed with the chimeric G_ion protein in COS-7 cells and IP assays were performed as described (Experimental Procedures). As controls plasmids encoding GFP, the human P2Y_12 and the human GPR33 pseudogene were co-transfected with G_ion. Basal IP formation is expressed as fold over basal levels of GFP-transfected cells (347 ± 8.6 cpm/well). Data are presented as means ± SEM of three independent experiments, each carried out in triplicate.
Table I
GPR33 allele frequency in Hominoidea

| Species (number of individuals investigated) | Allele 1 – Allele 2 | Number of individuals |
|---------------------------------------------|---------------------|-----------------------|
| Human (1217)                                | Arg140 - Arg140     | 1                     |
|                                             | Arg140 - Stop140    | 50                    |
|                                             | Stop140 - Stop140   | 1166                  |
| Chimpanzee (9)                              | Ser39/Arg140 - Ser39/Arg140 | 5          |
|                                             | Ser39/Arg140 - Stop39/Arg140 | 2          |
|                                             | Ser39/Stop140 - Stop39/Arg140 | 2          |
| Orangutan (4)                               | His171 - His171    | 2                     |
|                                             | His171 - Stop171    | 1                     |
|                                             | Stop171 - Stop171   | 1                     |
| Siamang (3)                                 | Trp94 - Stop94      | 3                     |

Polymorphic GPR33 inactivation (bold) was found in humans (see Fig. 2 and Suppl. 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.
Table II
Age estimates for the pseudogene alleles

| Species      | Nb. of chrom. surveyed | Nb. 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 $^f$ |
|--------------|------------------------|-----------------------------------|------------------------|----------------------------------|--------------------------------------|---------------------------------|----------------------------------------------------------|
| Human        | 170$^c$                | 164                               | 10,000                 | 970,900                          | 48,150                               | 800,000-1,150,000$^e$            | 1,145,000                                               |
| Chimpanzee   | 18                     | 4$^d$                             | 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                          | not determined                                          |

$^a$ 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), while for chimpanzees and orangutans it is 20 years.

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

$^c$ The entire coding region of GPR33 was sequenced from 85 individuals of a human genomic DNA panel used in (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.

$^d$ 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.

$^e$ 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. While 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.

$^f$ Estimate of the allele age on the basis of mutation rates (see text and Suppl. Table S6).
Figure 1

Figure 2
Figure 3

A

- PC
- standard
- heart
- brain
- placenta
- lung
- liver
- skeletal muscle
- pancreas
- spleen
- thymus
- prostate
- testis
- ovary
- small intestine
- colon
- blood/leukocyte

NC

mouse GPR33

mouse GPDH

Figure 4

B

- PC
- standard
- heart
- brain
- placenta
- lung
- liver
- skeletal muscle
- pancreas
- spleen
- thymus
- prostate
- testis
- ovary
- small intestine
- colon
- blood/leukocyte

NC

human GPR33

human GPDH
RESULTS

Nonsense mutation of GPR33 causes truncation of the receptor protein

Because all GPR33 pseudogenes found in humans and apes were due to stop codons we addressed the question whether the stop codons indeed terminates the translation of the receptor proteins. An increasing number of studies showed that not all stop codons are equal and some allow a leaky translation mainly in plants and yeast (44-46). To test whether the stop codon within the human GPR33 pseudogene leads to a truncated protein the receptor was N- and C-terminally epitope tagged with an HA-epitope and a FLAG-epitope, respectively. Immunofluorescence studies revealed the presence of the N-terminal epitope, however, the C terminus could not be detected (Suppl. Fig. S1). Similar results were obtained with the pseudogenes of chimpanzee and orangutan (data not shown). Interestingly, aminoglycoside antibiotics are able to suppress premature stop codons in the cystic fibrosis transmembrane regulator, dystrophin and the V2 vasopressin receptor genes (48-50). Similarly, COS-7 cells transiently transfected with the human GPR33 regained the ability to produce a full length receptor following incubation with the aminoglycoside geneticin (Suppl. Fig. S1).

Aminoglycoside antibiotics can suppress premature stop codons in the cystic fibrosis transmembrane regulator, dystrophin and the V2 vasopressin receptor genes (48-50). Similarly, COS-7 cells transiently transfected with the human GPR33 regained the ability to produce a full length receptor following incubation with the aminoglycoside geneticin (Suppl. Fig. S1).

Human and rat GPR33 pseudogenes are expressed

To gain more information beyond structural relations, we set out to clarify the expression pattern of GPR33 in mouse tissue by RT-PCR. The mouse GPR33 is a single copy gene at chromosome 12 (12B3) and the structure of the mRNA was unknown. To avoid false positive results in the RT-PCR analysis with primers derived from the coding region because of genomic contamination of the cDNA we first analyzed the structure of the mouse GPR33 transcript. Thus, 5’ rapid amplification of cDNA ends (RACE) PCR was performed with a cDNA library from mouse spleen (Clontech, Palo Alto, CA). PCR fragments were cloned and sequenced. In several clones the coding region of the mouse GPR33 was 5’ joined with noncoding DNA sequence which matched with the genomic mouse database sequence (Acc.
number NT_039551) about 3.6 kb upstream the start ATG of GPR33. The intron of the 5'-UTR is flanked by a classical consensus sequence. The sequence of the GPR33 5'-UTR was verified by direct genomic sequencing. For RT-PCR a primer pair was chosen which flanks the intron and a 413-bp PCR product is obtained only when the exons are properly spliced. As shown in Fig. 4A, GPR33 is mainly expressed in mouse spleen, lung and testis. Then, several murine cell lines were analyzed for GPR33 expression by RT-PCR. In consensus with a predominant expression in immunologically relevant tissues GPR33 transcripts were identified in RAW 264.7 cells, a murine macrophage cell line.

Next, we asked whether GPR33 pseudogenes are also transcribed. Comparison of the rat genomic sequence with the mouse GPR33 locus revealed high sequence similarities in the coding region and the 5'-UTR. Similarly, primers were designed to amplify a 284-bp fragment from spleen cDNA of R. norvegicus. As found for the mouse GPR33 a specific product was amplified and its identity was verified by sequencing.

Except of the coding sequence and the splice acceptor site (5' of the translation start ATG), no similarities are obvious when the mouse and the human GPR33 locus were compared in the 5' region. Therefore, primers derived from the coding region had to be used in the human RT-PCR. As shown in Fig. 4B, the human GPR33 transcript was highly expressed in spleen and lung but also in heart, liver, kidney, pancreas, thymus, gonads, and leukocytes. The tissue panel was controlled for genomic contamination by using the identical antisense primer together with a sense primer directly upstream of the putative splice acceptor site. Although a specific PCR product was amplified with human genomic DNA as control, all samples of the cDNA panel revealed no such product indicating that there is no contamination. Sequence analysis of rat and human pseudogene cDNA revealed no evidence for exon skipping or intergenic splicing which may produce fusion proteins with neighbored genes as found e.g. for P2Y_{11} and SSF1 (51).

REFERENCES (Suppl.)

44. Skuzeski, J. M., Nichols, L. M., Gesteland, R. F. and Atkins, J. F. (1991) J. Mol. Biol. 218, 365-373
45. Namy, O., Hatin, I. and Rousset, J. P. (2001) Impact of the six nucleotides downstream of the stop codon on translation termination. EMBO Rep. 2, 787-793
46. Keeling, K. M., Lanier, J., Du, M., Salas-Marco, J., Gao, L., Kaenjak-Angeletti, A. and Bedwell, D. M. (2004) RNA 10, 691-703
47. Mankin, A. S. and Liebman, S. W. (1999) Nat. Genet. 23, 8-10
48. Barton-Davis, E. R., Cordier, L., Shoturma, D. I., Leland, S. E. and Sweeney, H. L. (1999) J. Clin. Invest. 104, 375–381
49. Wilschanski, M., Yahav, Y., Yaacov, Y., Blau, H., Bentur, L., Rivlin, J., Aviram, M., Bdolah-Abram, T., Bebok, Z., Shushi, L., Kerem, B. and Kerem, E. (2003) N. Engl. J. Med. 349, 1433-1441
50. Sangkuhl, K., Schulz, A., Römpler, H., Yun, J., Wess, J. and Schöneberg, T. (2004) Hum. Mol. Genet. 13, 893-903
51. Communi, D., Suarez-Huerta, N., Dussossoy, D., Savi, P., Boeynaems, J. M. (2001) J. Biol. Chem. 276, 16561-16566
Suppl. Figure S1. Reconstitution of the polypeptide chain of the human GPR33 pseudogene. Aminoglycosides are able to suppress premature stop codons, thereby permitting protein translation to continue to the normal end of the gene. To examine the ability of geneticin to suppress the stop codon 140, COS-7 cells were transfected with expression plasmids coding for the human GPR33 Stop140. Transfected COS-7 cells were incubated with or without 75 µg geneticin/ml one day prior to immunostaining. Permeabilized cells were incubated with a monoclonal anti-HA-antibody (red) and a polyclonal antibody against the FLAG epitope (green). The primary antibodies were detected using species specific TRITC- and FITC-labeled secondary antibodies. Nuclei were stained with DAPI (blue). Confocal immunofluorescence pictures are representative of three independent experiments.
Suppl. Table S1
Accession numbers, description and sources of genomic DNA samples
Abbreviations: NT, N terminus; CT, C terminus; TMD, transmembrane domain, IL, intracellular loop

| accession number | common name    | species            | cloned sequence | Source                                                                 |
|------------------|----------------|--------------------|-----------------|------------------------------------------------------------------------|
| **Primates**     |                |                    |                 |                                                                        |
| AY490588         | human          | *Homo sapiens*     | NT-CT           | Dr M. Kayser, MPI, Leipzig, Germany                                     |
| AY490589         |                |                    |                 | Dr M. Stoneking, MPI, Leipzig, Germany                                  |
| AY493998         |                |                    |                 | D. Sere, MPI, Leipzig, Germany                                          |
| AY493999         |                |                    |                 | own source                                                             |
| AY494000         | gorilla        | *Gorilla gorilla*  | NT-CT           | Dr J. Bullerdiek, University of Bremen                                  |
|                  |                |                    |                 | Dr T. Haaf, MPI Molecular Genetics Berlin, Germany                     |
|                  |                |                    |                 | Dr J. Rhyne, University of Maryland, Baltimore, USA                     |
|                  |                |                    |                 | Dr M. Rocchi, University of Bari, Italy                                |
|                  |                |                    |                 | Dr L. Vigilant, MPI, Leipzig, Germany                                   |
| AY490569         | siamang        | *Hylobates syndactylus* | NT-CT         | Dipl. Biol. C. Roos, Primate Center Goettingen, Germany               |
| AY490570         |                |                    |                 | Dr T. Haaf, MPI Molecular Genetics Berlin, Germany                     |
|                  | common gibbon  | *Hylobates lar*    | NT-CT           | Dipl. Biol. C. Roos, Primate Center Goettingen, Germany               |
|                  |                |                    |                 | Dr W. Enard, MPI, Leipzig, Germany                                     |
|                  |                |                    |                 | Dr J. Bullerdiek, University of Bremen                                  |
|                  |                |                    |                 | Dr M. Rocchi, University of Bari, Italy                                |
| AY490573         | red-cheeked gibbon | *Hylobates gabriellae* | NT-CT       | Dipl. Biol. C. Roos, Primate Center Goettingen, Germany               |
| AY490574         | white-cheeked gibbon | *Hylobates leucogenys* | NT-CT       | Dipl. Biol. C. Roos, Primate Center Goettingen, Germany               |
| AY490575         | white-tufted-ear marmoset | *Callithrix jacchus* | NT-CT       | Dr T. Haaf, MPI Molecular Genetics Berlin, Germany                     |
| AY490576         | Geoffroy's marmoset | *Callithrix geoffroyi* | NT-CT       | Dr J. Rhyne, University of Maryland, Baltimore, USA                     |
| AY490577         | silvered langur | *Presbytis cristata* | NT-CT          | Dr T. Haaf, MPI Molecular Genetics Berlin, Germany                     |
| Accession | Species             | Genus   | NT/IL | Institute of Collection |
|-----------|---------------------|---------|-------|-------------------------|
| AY490580  | chimpanzee          | Pan     | NT-CT | Dr T. Haaf, MIPI Molecular Genetics Berlin, Germany |
| AY490581  |                     |         |       | Dr M. Rocchi, University of Bari, Italy |
| AY490582  |                     |         |       | Dr W. Enard, MPI, Leipzig, Germany |
| AY490583  |                     |         |       | Dr M. Rocchi, University of Bari, Italy |
| AY490584  |                     |         |       | Dr W. Enard, MPI, Leipzig, Germany |
| AY490585  |                     |         |       | Dr W. Enard, MPI, Leipzig, Germany |
| AY490586  |                     |         |       | Dr W. Enard, MPI, Leipzig, Germany |
| AY490587  | pygmy chimpanzee, bonobo | Pan     | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| AY490590  | hamadryas baboon    | Papio   | NT-CT | Dr M. Rocchi, University of Bari, Italy |
| AY490591  | rhesus monkey       | Macaca  | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| AY490592  | orangutan           | Pongo   | NT-CT | Dr J. Rhyne, University of Maryland, Baltimore, USA |
| AY490595  | potto               | Perodicticus | NT-IL3 | Dr T. Haaf, MIPI Molecular Genetics Berlin, Germany |
| AY490596  | emperor tamarin     | Saguinus | NT-CT | Dr J. Rhyne, University of Maryland, Baltimore, USA |
| AY490597  | ruffed lemur        | Varecia  | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| AY490598  | ring-tailed lemur   | Lemur   | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| AY490599  | mandrill            | Mandrillus | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| AY490600  | common squirrel monkey | Saimiri | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| AY490603  | red guenon          | Erythrocebus | NT-CT | Dr J. Rhyne, University of Maryland, Baltimore, USA |
| AY490604  | Azara's night monkey | Aotus   | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| AY490605  |                     |         |       | Dr W. Enard, MPI, Leipzig, Germany |
| AY490606  | brown capuchin      | Cebus   | NT-CT | Dr W. Enard, MPI, Leipzig, Germany |
| Accession | Species | Genus | Scientific Name | Collaborator | Institution |
|-----------|---------|-------|-----------------|--------------|-------------|
| AY490607  | guereza | Colobus | guereza         | Dr W. Enard | MPI, Leipzig, Germany |
| AY490608  | slow loris | Nycticebus | bengalensis | Dr W. Enard | MPI, Leipzig, Germany |
| AY490609  | South African galago | Galago | moholi         | Dr W. Enard | MPI, Leipzig, Germany |
| AY490611  | brown-headed spider monkey | Ateles |fusciceps       | Dr W. Enard | MPI, Leipzig, Germany |
| AY490613  | African green monkey | Cercopithecus | aethiops | COS-7, ATCC – cell collection |
| AY490614  | European woodmouse | Apodemus | sylvaticus      | Dr G. Gerlach | University of Konstanz, Germany |
| AY490617  | wood mice | Apodemus | sp.            | Dr G. Gerlach | University of Konstanz, Germany |
| AY490615  | yellow-necked field mouse | Apodemus | flavicollis    | Dr H. Meisel | Humboldt University Berlin, Germany |
| AY490621  | sibling vole | Microtus | epiroticus     | Dr G. Gerlach | University of Konstanz, Germany |
| AY490622  | common vole | Microtus | arvalis        | Dr H. Meisel | Humboldt University Berlin, Germany |
| AY490624  | bank vole | Clethrionomys | glareolus   | Dr G. Gerlach | University of Konstanz, Germany |
| AY490626  | eastern European house mouse | Mus | musculus       | Dr A. Orth | University of Montpellier, France |
| AY490627  | Macedonian mouse | Mus | macedonicus     | Dr A. Orth | University of Montpellier, France |
| AY490628  | southeastern Asian house mouse | Mus | musculus castaneus | Dr A. Orth | University of Montpellier, France |
| AY490629  | house mouse | Mus | musculus sp.   | Dr A. Orth | University of Montpellier, France |
| AY490630  | Japanese mice | Mus | musculus molossinus | Dr A. Orth | University of Montpellier, France |
| AY490631  | western wild mouse | Mus | spreus         | Dr A. Orth | University of Montpellier, France |
| AY490632  | steppe mouse | Mus | spicilegus     | Dr A. Orth | University of Montpellier, France |
| AY490633  | Ryukyu mouse | Mus | caroli         | Dr A. Orth | University of Montpellier, France |
| AY490634  | servant mouse | Mus | famulus        | Dr A. Orth | University of Montpellier, France |
| AY490635  | western European house mouse | Mus | musculus domesticus | Dr A. Orth | University of Montpellier, France |
| AY490636  | house mice | Mus | musculus 129/SV | Research Facility for Experimental Medicine, Berlin, Germany |
| AY490637  | house mice | Mus | musculus Bal 6/3 | Charles River Laboratories, Wilmington, USA |
| AY490638  | house mice | Mus | musculus OBA / 2 | Charles River Laboratories, Wilmington, USA |
| AY490639  | house mice | Mus | musculus CBA / J | Charles River Laboratories, Wilmington, USA |
| AY490640  | house mice | Mus | musculus FvB   | Charles River Laboratories, Wilmington, USA |
| AY490641  | midday gerbil | Meriones | meridianus    | Dr C. Pitra, IZW, Berlin, Germany |
| AY490642  | Mongolian gerbil | Meriones | unguiculatus | ECL2-CT | Research Facility for Experimental Medicine, Berlin, Germany |
| Accession Number | Species | Strain | Source |
|------------------|---------|--------|--------|
| AY490643-44      | Norway rat | Rattus norvegicus | NT-CT |
| AY490602         | Norway rat | Rattus norvegicus | Dipl. Biol. M. Buske, Federal Environmental Agency, Berlin, Germany |
| AY490645-47      | black rat | Rattus rattus | NT-CT |
| AY490648-49      | black rat | Rattus rattus | Dipl. Biol. M. Buske, Federal Environmental Agency, Berlin, Germany |
| AY490646         | Dassie rat | Petromus typicus | Dr R. Adkins, University of TN, USA |
| AY490647         | naked mole-rat | Heterocephalus glaber | Dr R. Adkins, University of TN, USA |
| AY490650-52      | muskrat | Ondatra zibethicus | NT-TMD7 |
| AY490651         | spiny mouse | Acomys sp. | NT-TMD4 |
| AY502103-14      | Buffy Cuis | Galea monastieri | NT-CT |
| AY490653-54      | European hare | Lepus europaeus | NT-CT |
| AY490655         | rabbit | Oryctolagus cuniculus | NT-CT |
| AY490656-57      | minke whale | Balaenoptera acutorostrata | NT-CT |
| AY490658-61      | domestic cow | Bos taurus | NT-CT |
| AY490662-64      | red deer | Cervus elaphus | NT-CT |
| AY490663         | reindeer | Rangifer tarandus | NT-CT |
| AY490664         | roe deer | Capreolus capreolus | NT-CT |
| Accession | Species | Genus | Author          | Location                      |
|-----------|---------|-------|-----------------|-------------------------------|
| AY490665  | moose   | Alces | K. Zierau, Berlin, Germany |
| AY490666  | water buffalo | Bubalus | own source |
| AY490667  | European bison | Bison | Dr C. Pitra, IZW, Berlin, Germany |
| AY490668  | domestic pig | Sus | own source |
| AY490669  | wild pig | Sus | Dipl. Med. J. Moeller, Erfurt, Germany |
| AY490670  | alpaca | Lama | Dr C. Pitra, IZW, Berlin, Germany |
| AY490671  | lama | Lama | Dr C. Pitra, IZW, Berlin, Germany |
| AY490672  | Arabian camel | Camelus | Dr C. Pitra, IZW, Berlin, Germany |
| AY490673  | Bactrian camel | Camelus | Dr C. Pitra, IZW, Berlin, Germany |
| AY490674  | pygmy hippopotamus | Hexaprotodon | Dr C. Pitra, IZW, Berlin, Germany |
| AY490675  | Mhorr gazelle | Gazella | Dr C. Pitra, IZW, Berlin, Germany |
| AY490676  | Sabel antelope | Hippotragus | Dr C. Pitra, IZW, Berlin, Germany |
| AY490677  | giraffe | Giraffa | Dr C. Pitra, IZW, Berlin, Germany |
| AY490678  | muskox | Ovis | Dr C. Pitra, IZW, Berlin, Germany |
| AY490679  | sheep | Ovis | Dr C. Pitra, IZW, Berlin, Germany |
| AY490680  | mouflon | Ovis | Dr C. Pitra, IZW, Berlin, Germany |

**Perissodactyla**

| Accession | Species | Genus | Author          | Location                      |
|-----------|---------|-------|-----------------|-------------------------------|
| AY490690  | horse | Equus | own source |
| AY490691  | Przewalski horse | Equus przewalskii | Dr C. Pitra, IZW, Berlin, Germany |
| AY490692  | kulan | Equus | Dr C. Pitra, IZW, Berlin, Germany |
| AY490693  | tapirs | Tapirus | Dr C. Pitra, IZW, Berlin, Germany |
| AY490694  | white rhinoceros | Ceratotherium | Dr C. Pitra, IZW, Berlin, Germany |

**Carnivora**

| Accession | Species | Genus | Author          | Location                      |
|-----------|---------|-------|-----------------|-------------------------------|
| AY490695  | cat | Felis | Dipl. Med. J. Moeller, Erfurt, Germany |
| AY490696  | dog | Canis | own source |
| AY490697  | beach marten | Martes | own source |
| AY490698  | Eurasian river otter | Lutra | Dr C. Pitra, IZW, Berlin, Germany |
| Accession | Species Name       | Scientific Name            | Code   | Institution                                      |
|-----------|--------------------|----------------------------|--------|-------------------------------------------------|
| AY490701  | cheetah            | Acinonyx jubatus           | ECL1-CT| Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490702  | Baikal seal        | Phoca sibirica             | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490703  | California sealion | Zalophus californianus     | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490704  | brown bear         | Ursus arctos              | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490705  | Malayan sun bear   | Helarctos malayanus       | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490706  | polar bear         | Ursus maritimus           | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490707  | American black bear| Ursus americanus          | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490708  | ring-tailed coati  | Nasua nasua               | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490709  | gray wolf          | Canis lupus               | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490710  | raccoon dog        | Nyctereutes procyonoides  | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490711  | red fox            | Vulpes vulpes             | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490712  | striped hyena      | Hyaena hyaena             | NT-IL3 | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490713  | jaguar             | Panthera onca             | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490714  | lion               | Panthera leo              | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490715  | puma               | Puma concolor             | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490716  | snow leopard       | Uncia uncia               | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490717  | Siberian tiger     | Panthera tigris altaica   | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490718  | Caribbean manatee  | Trichechus manatus        | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490719  | Asiatic elephant   | Elephas maximus           | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490720  | African elephant   | Loxodonta africana        | NT-CT  | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490721  | European shrew     | Sorex araneus             | NT-CT  | own source                                      |
| AY490722  | European mole      | Talpa europaea            | NT-TMD7| own source                                      |
| AY490723  | western European hedgehog | Erinaceus europaeus | IL1-CT | own source                                      |
| AY490724  | flying foxes       | Pteropus sp.              | NT-IL3 | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490725  | Central American yellow bat | Rhogeessa io | NT-IL3 | Dr. C. Pitra, IZW, Berlin, Germany               |
| AY490726  | Silver-haired Myotis | Myotis albenscens      | NT-IL3 | Dr. C. Pitra, IZW, Berlin, Germany               |
Non-mammalian species analyzed in this study for presence of GPR33 orthologs are listed in the Supplementary Section of ref. (22).
| Primer number | Sequence | Orientation | Domain |
|---------------|----------|-------------|--------|
| 1             | 5'-TGYTTGCCAKTWRAAAYATTC-3' | s | 5'-UTR rodent |
| 2             | 5'-GGTTTGAGCCACATGCAWG-3' | s | 5'-UTR rat |
| 3             | 5'-CATGCCATTGGAATTATTAAT-3' | s | 5'-UTR cattle |
| 4             | 5'-TGGCTTCGAGATTTACTATAAAG-3' | s | 5'-UTR Apodemus |
| 5             | 5'-AGATSSRTCAGTCTTTAATTCCTCA-3' | s | 5'-UTR non primate |
| 6             | 5'-GCTCTTTTTTTTTKTCTYATTYYCAGTTTGCATCAAAT-3' | s | 5'-UTR even-toed ungulate |
| 7             | 5'-ATTTTTTTTTTTTCCTTYNTTTTYCCTCCATAGTAAYT-3' | s | 5'-UTR rodents |
| 8             | 5'-CTCCATAGTMYATGAGTYKGRCTCAACT-3' | s | N-term mammalian |
| 9             | 5'-CTCCATAGSAMYATATGGAAYKRGYCARCTC-3' | s | N-term mammalian |
| 10            | 5'-ATGGAYCTGACACCTACTGARYKRWYCTGCTCAAT-3' | s | N-term even-toed ungulate |
| 11            | 5'-ATGGAYCTGACACCTACTGARYKRWYCTGCT-3' | s | N-term even-toed ungulate |
| 12            | 5'-ATGGATTTGGCTCAACTTASTGATT-3' | s | N-term carnivore |
| 13            | 5'-ATGGATCTGATCACTCTACTGATTAC-3' | s | N-term primate |
| 14            | 5'-CCCACCATGATTTCGATACACTCCATC-3' | s | N-term mouse |
| 15            | 5'-CCCACCATGATGCTAGTCAACTCTACTG-3' | s | N-term human |
| 16            | 5'-TCAAGTAAAAATAGACACTATTTTCTAGCTCMYGC-3' | s | N-term odd-toed ungulate |
| 17            | 5'-GATGCAAGGCTAGAAGCAGTGTC-3' | as | N-term primate |
| 18            | 5'-CAACACGCACTGCGTCTCAACTC-3' | s | N-term murinae |
| 19            | 5'-AGCAGCTCTACCAAAAATRATKATTCG-3' | s | TMD1 primate |
| 20            | 5'-ATCASCAYAGGCCCTCTAYCRTG-3' | s | TMD1 mammalian |
| 21            | 5'-ATCASCAYAGGCCCTNTAYCTGTG-3' | s | TMD1 mammalian |
| 22            | 5'-CAATGGTCTCTCTAATCTATG-3' | s | TMD1 even-toed ungulate |
| 23            | 5'-TATCTATGGGCTTNWTTYAARATG-3' | s | TMD1&IL1 |
| 24            | 5'-TTTTCAYCTTTATYTTYTCTHTAYTAT-3' | s | TMD2 |
| 25            | 5'-AAATACNYTTYTTTWTYCTGACTCTATT-3' | s | TMD2 |
| 26            | 5'-AAATAAAATAGGAGAGATGAGATG-3' | as | TMD2 mammalian |
| 27            | 5'-TCMACATTGATCTGCCMTTYATG-3' | s | TMD2 mammalian |
| 28            | 5'-TAACTATGCTGTGTCYASRACTCCG-3' | s | EL2 |
| 29            | 5'-CCTGCTTCTCTCCCAATAGTGMSMC-3' | as | EL2 |
| 30            | 5'-CTGTCTTCTCTCTTCTGNGCCATMRG-3' | s | TMD3 |
| 31            | 5'-CACGTAGCTTGGGAATCATCACTG-3' | s | TMD3 |
| 32            | 5'-CCTTGCACATGGCAGTCCAAAAGCTCC-3' | as | TMD3 |
| 33            | 5'-GAAGGTAACAGATCAAKACYKATRGC-3' | as | IL2 |
| 34            | 5'-CCYTAYCTGATTCCCGGGAGACC-3' | s | TMD4 |
| 35            | 5'-ATGACATAAAGAGAGAGTGACCTG-3' | s | EL2 |
| 36            | 5'-GTCACTTAAATCACCTAAGGGTTGTCG-3' | s | IL3 |
| 37            | 5'-GAGTAGCCCAAAAGATGAAAGAGAG-3' | s | IL3 |
| 38            | 5'-GATGACTTTGAAAGGTTGTGTAAG-3' | as | IL3 |
| 39            | 5'-GAAAGTACGCGAAAGATGAAAG-3' | s | IL3 |
| 40            | 5'-TCYTTCCTTSTGTYGATGCC-3' | s | TMD6 |
|   | Sequence                                      | Type     | Species         |
|---|----------------------------------------------|----------|-----------------|
| 41 | 5’-TCTTTTTGTGTGTTGGMTC-3’                   | s        | TMD6 mammalian  |
| 42 | 5’-TARGYCATCCACACASRAAGAARGA-3’             | as       | TMD6           |
| 43 | 5’-GGTTRAYRTGRTANGGCATCCARCA-3’             | s        | TMD6 mammalian  |
| 44 | 5’-GGTAGGGCATCCAACACASRAAGA-3’             | as       | TMD6 mammalian  |
| 45 | 5’-GTGTAYRTGRTAGGGCATCCARCA-3’             | as       | TMD6 mammalian  |
| 46 | 5’-TCTTCTGAGACCAAGCCACTGCTG-3’             | as       | TMD6 mammalian  |
| 47 | 5’-GGGCTTAATGCTCCTAGAAGACCGG-3’             | s        | TMD6 lemur      |
| 48 | 5’-GACTCTGATACCTAGTGCTAAC-3’               | s        | TMD7           |
| 49 | 5’-GAARTTCTCCCARYAAYAARTARARG-3’           | as       | TMD7 mammalian  |
| 50 | 5’-TCCCCARYAAATAARTARAGWRYNGGAGA-3’         | as       | TMD7 mammalian  |
| 51 | 5’-CCATTCTTTGCTTGTTGAGTC-3’                | s        | C-term primate  |
| 52 | 5’-GAGCAAGAATRGAYTTTTRAAAC-3’              | as       | C-term mammalian |
| 53 | 5’-TTCTGRRRTYAGTTTTGNGTYCTTTC-3’           | as       | C-term mammalian |
| 54 | 5’-GAATTAAAAAYTTMGRTTCTGARTTNAG-3’         | as       | C-term mammalian |
| 55 | 5’-TTACATGTCCGGCTTCTGAATTTAGG-3’           | as       | C-term cattle   |
| 56 | 5’-TTAAATGCTCTGAGTGTCAGTTTGG-3’            | as       | C-term Apodemus |
| 57 | 5’-TCYAGAATTIAATTTMRGRTTCTG-3’             | as       | C-term mammalian |
| 58 | 5’-TCTGAGTTAGGGTYGYYCYTKTC-3’              | as       | C-term mammalian |
| 59 | 5’-GAGACTTCTTGAGACGTTTTCGARTTC-3’          | as       | C-term primate  |
| 60 | 5’-TCCCCACCAAGAGGATCCACAAT-3’              | as       | C-term mouse    |
| 61 | 5’-CCAAATGCTGGTTGCTTAAG-3’                 | as       | C-term human    |
| 62 | 5’-GTGCRTGTTTCTAGTCTAGATCT-3’              | as       | 3’-UTR primitive |
| 63 | 5’-TTCAGTGGTCTGGTTGCTTAAG-3’               | as       | 3’-UTR primitive |
| 64 | 5’-CTAGTTATCTCTCTATATAGGG-3’               | as       | 3’-UTR          |
| 65 | 5’-GAATCATTTTCCTAGATCTTG-3’                | as       | 3’-UTR          |
| 66 | 5’-GGTTCTAGACATATGCTGTCGATGATTTG-3’        | as       | 3’-UTR murinae  |
| 67 | 5’-TCCCTAGGGRCTGGGGTTATAGAC-3’             | as       | 3’-UTR murinae  |
Suppl. Table S3  

Human DNA samples

The DNA sample collection contained most samples of the CEPH 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. "#" indicates a homozygote individual.

| Number | sample group                  | Number of individuals | TGA allele | CGA allele |
|--------|-------------------------------|-----------------------|------------|------------|
| 1      | San                           | 8                     | 16         | 0          |
| 2      | Bantu (south east)            | 5                     | 10         | 0          |
| 3      | Bantu (south west)            | 3                     | 6          | 0          |
| 4      | Biaka Pygmies                 | 37                    | 69         | 5          |
| 5      | Mbuti Pygmies                 | 19                    | 38         | 0          |
| 6      | Mbenzele Pygmies              | 2                     | 4          | 0          |
| 7      | Bamileke                      | 1                     | 2          | 0          |
| 8      | Igbo                          | 1                     | 2          | 0          |
| 9      | Yoruba                        | 47                    | 94         | 0          |
| 10     | Effik                         | 1                     | 2          | 0          |
| 11     | Hausa                         | 1                     | 2          | 0          |
| 12     | Mandenka                      | 24                    | 48         | 0          |
| 13     | Bantu (Kenya)                 | 12                    | 24         | 0          |
| 14     | Lisombo                       | 1                     | 1          | 1          |
| 15     | Ethiopian jew                 | 2                     | 4          | 0          |
| 16     | Mozabite                      | 30                    | 55         | 5          |
| 17     | Nubian                        | 2                     | 4          | 0          |
| 18     | Kopte                         | 1                     | 2          | 0          |
| 19     | Bedouin                       | 49                    | 98         | 0          |
| 20     | Druze                         | 50                    | 98         | 2          |
| 21     | Palestinian                   | 53                    | 105        | 1          |
| 22     | Habani Syrian                 | 2                     | 4          | 0          |
| 23     | Iranian                       | 3                     | 6          | 0          |
| 24     | Yemenite jew                  | 2                     | 4          | 0          |
| 25     | Brahui, Balochi, Hazara, Sindhi | 100        | 192 8     |            |
| 26     | Burusho, Makrani, Pathan, Kalash | 100   | 200 0     |            |
| 27     | Uzbek                         | 2                     | 4          | 0          |
| 28     | Kirghiz                       | 1                     | 2          | 0          |
| 29     | Tumile                        | 1                     | 2          | 0          |
| 30     | Asian Indian                  | 1                     | 2          | 0          |
| 31     | Thai                          | 2                     | 4          | 0          |
| 32     | Cambodian                     | 11                    | 22         | 0          |
| 33     | Filipino                      | 2                     | 2          | 2#         |
| 34     | Papuan                        | 66                    | 125        | 7          |
| 35     | NAN Melanesian                | 22                    | 43         | 1          |
| 36     | Nasioi                        | 2                     | 4          | 0          |
| 37     | Samoan                        | 2                     | 3          | 1          |
| 38     | Aborigine Australian          | 4                     | 8          | 0          |
| 39     | Japanese                      | 35                    | 69         | 1          |
| 40     | Han                           | 45                    | 87         | 3          |
| 41     | Tujia                         | 10                    | 19         | 1          |
| 42     | Yizu                          | 10                    | 20         | 0          |
| 43     | Miao zu                       | 10                    | 20         | 0          |
| 44     | Oroqen                        | 10                    | 20         | 0          |
| 45     | Daur                          | 10                    | 20         | 0          |
| 46     | Mongola                       | 10                    | 20         | 0          |
| 47     | Hezhen                        | 10                    | 19         | 1          |
|   |     |   |   |
|---|-----|---|---|
| 48 | Xibo | 9 | 17 |
| 49 | Uygur | 10 | 20 |
| 50 | Dai | 10 | 20 |
| 51 | Lahu | 10 | 16 |
| 52 | She | 10 | 20 |
| 53 | Naxi | 10 | 19 |
| 54 | Tu | 10 | 19 |
| 55 | Buryat | 1 | 1 |
| 56 | Korean | 1 | 2 |
| 57 | Yakut | 26 | 51 |
| 58 | Evenki | 1 | 2 |
| 59 | Chukchi | 1 | 2 |
| 60 | Siberian Inuit | 1 | 2 |
| 61 | Estonia | 1 | 2 |
| 62 | Saami | 2 | 4 |
| 63 | Swedish | 1 | 2 |
| 64 | Danish | 1 | 2 |
| 65 | German | 10 | 20 |
| 66 | Dutch | 1 | 2 |
| 67 | English | 1 | 2 |
| 68 | French | 30 | 60 |
| 69 | French Basque | 24 | 48 |
| 70 | Spain | 1 | 2 |
| 71 | Sardinian | 28 | 56 |
| 72 | North Italian | 15 | 30 |
| 73 | Tuscan | 8 | 16 |
| 74 | Orcadian | 16 | 31 |
| 75 | Islandic | 1 | 2 |
| 76 | Greek | 1 | 2 |
| 77 | Adygei | 17 | 33 |
| 78 | Russian | 25 | 49 |
| 79 | Pima | 25 | 50 |
| 80 | Maya | 25 | 49 |
| 81 | Warao Indian | 2 | 4 |
| 82 | Colombian | 13 | 26 |
| 83 | Karitiana | 24 | 48 |
| 84 | Surui | 21 | 42 |
| 85 | Kaingang | 1 | 2 |
| 86 | Guarani | 1 | 2 |
### Suppl. Table S4

**Primers used for dating experiments (human, bonobo, chimpanzee)**

| primer number | sequence | orientation | domain |
|---------------|----------|-------------|--------|
| hu-1          | 5'-TGAGGGGCTTAGCAGGAGA-3' | s          | 5'-UTR |
| hu-2          | 5'-GGATAAAAACACCAAGGAGAC-3' | s          | 5'-UTR |
| hu-3          | 5'-TAGAGGCGAAGTATACCTTAGG-3' | s          | 5'-UTR |
| hu-4          | 5'-ACAGAGTCATATGCAAGAAATTG-3' | as        | 5'-UTR |
| hu-5          | 5'-CACCATGTTGGCCAGGATG-3' | s          | 5'-UTR |
| hu-6          | 5'-CATGCTCTGAATTCAGCAGC-3' | as        | 5'-UTR |
| hu-7          | 5'-ACTTCTAACAGCGACCTGGGAGGC-3' | as        | 5'-UTR |
| hu-8          | 5'-CTGCGGCTGGCAAAATGTAG-3' | as        | 5'-UTR |
| hu-9          | 5'-CTGAAAGTCAGGAGTGAGACACC-3' | s          | 5'-UTR |
| hu-10         | 5'-CCATCACTCTTACCTTGCAAGAC-3' | as        | 5'-UTR |
| hu-11         | 5'-GTGACAAAGGCAAGACCTCTGAC-3' | s          | 5'-UTR |
| hu-12         | 5'-AAATCTCCTCAGAATTCATCAAC-3' | as        | 5'-UTR |
| hu-13         | 5'-TATTTATCTGCTAGGGCAAAATGTAGC-3' | s          | 5'-UTR |
| hu-14         | 5'-ACTATGTTGGCCAGGCTCGTC-3' | as        | 5'-UTR |
| hu-15         | 5'-CATATATATTCTATTCTATGTGCTAG-3' | s          | 5'-UTR |
| hu-16         | 5'-CCATACCAATGGCAAGATCATGTGTC-3' | as        | cod reg |
| hu-17         | 5'-GGAACTCCTGGTGAAGGTCGTC-3' | s          | cod reg |
| hu-18         | 5'-CTGACGCCTACCTCTCAGTC-3' | s          | cod reg |
| hu-19         | 5'-CCAGTTAGACACACAGCATAG-3' | as        | cod reg |
| hu-20         | 5'-TCAGAAGGTCATCCATCTGTGCTC-3' | s          | cod reg |
| hu-21         | 5'-ACCTGAGTAGCTGAGAATCTAC-3' | s          | 3'-UTR |
| hu-22         | 5'-GTGATCCGCTGCCCTCAACC-3' | s          | 3'-UTR |
| hu-23         | 5'-AACAAGTCTGTGCCCATAAGAG-3' | s          | 3'-UTR |
| hu-24         | 5'-GAGTTACTAATTTATGGGCAC-3' | s          | 3'-UTR |
| hu-25         | 5'-CATGTACATTCCACCACAGC-3' | s          | 3'-UTR |
| hu-26         | 5'-GTGGAGATGTAACATGTGTC-3' | as        | 3'-UTR |
| hu-27         | 5'-TACTCGAGAGGCTGAGGCA-3' | s          | 3'-UTR |
| hu-28         | 5'-GTGAAACCACATAATATACACTCAC-3' | s          | 3'-UTR |
| hu-29         | 5'-ATTATGAGTAGACAGGACTACC-3' | s          | 3'-UTR |
| hu-30         | 5'-TCAGTGGGCTTGTGCAAGGCACC-3' | as        | 3'-UTR |
| hu-31         | 5'-CAAAGGCCACTCTCTGTAGCTGC-3' | as        | 3'-UTR |
| hu-32         | 5'-CAATGTTGGCTCAGCACCCTCT-3' | s          | 3'-UTR |
| hu-33         | 5'-CACCCCAACTATCCAGGCTT-3' | s          | 3'-UTR |
| hu-34         | 5'-GGCCTGTGCTCTCTCTGATGAT-3' | s          | 3'-UTR |
| hu-35         | 5'-TGTTGCTGAGAGAATCTGCTAT-3' | s          | 3'-UTR |
| hu-36         | 5'-TATAGCCATGGCGAGAAATGCCAA-3' | s          | 3'-UTR |
| hu-37         | 5'-ATGTTTAAGCAATCTGTGTC-3' | s          | 3'-UTR |
| hu-38         | 5'-GAGCTCCCTCTTATTTAGTAACATC-3' | s          | 3'-UTR |
| hu-39         | 5'-ATAAGCTTTTCAATTTCGCAACC-3' | s          | 3'-UTR |
| hu-40         | 5'-TCTACATTCCAGGCAAAACTCTC-3' | s          | 3'-UTR |
|    | Sequence                        | Type | Location |
|----|---------------------------------|------|----------|
| hu-41 | 5’-CCAAAAATTTTTCATCATTCAAGGAC-3’ | s    | 3’-UTR   |
| hu-42 | 5’-GATTCTTAAGCAACACGTCATT-3’    | s    | 3’-UTR   |
| hu-43 | 5’-GGAAAGCAAGGAGATGCAAG-3’      | s    | 3’-UTR   |
| hu-44 | 5’-AAGACAATCAGGAAACTTTGGA-3’    | s    | 3’-UTR   |
| hu-45 | 5’-GCACTCAGTTTCTAGCTCTGATC-3’   | s    | 3’-UTR   |
| hu-46 | 5’-TTCAGTTTCTTACTCCCAGAGTC-3’   | s    | 3’-UTR   |
| hu-47 | 5’-GGTGATTTCATTTGCTATTGAGG-3’   | as   | 3’-UTR   |
| hu-48 | 5’-AACCACATGACTAGAGCTTCAAGC-3’  | as   | 3’-UTR   |
| hu-49 | 5’-TCCCCCTATAATTTCTCACGGATATGGA-3’ | as | 3’-UTR   |
| hu-50 | 5’-TGAGGAAAGACTTACACTGGCATT-3’  | as   | 3’-UTR   |
| hu-51 | 5’-AAATACTTCAAAAAAGCCTGACC-3’   | as   | 3’-UTR   |
| hu-53 | 5’-AGCTTTCAAAAAACCTTGAAATAGGG-3’ | as | 3’-UTR   |
| hu-54 | 5’-GCTGGTGATATTCAAGGACTGATGT-3’ | as   | 3’-UTR   |
| hu-55 | 5’-AACAAGATGAGTGCTTGTGG-3’      | as   | 3’-UTR   |
| hu-56 | 5’-GTTGGTGATGTTGTTAATTTTCGT-3’  | as   | 3’-UTR   |
| hu-57 | 5’-TGGCAGAAAATCCCATCTCTA-3’     | as   | 3’-UTR   |
### Suppl. Table S5

**Primers used for dating experiments (rats)**

| primer number | sequence | orientation | domain       |
|---------------|----------|-------------|--------------|
| rat-1         | 5'-TCATCAGGCACACATGTGGTG-3’ | s            | 5’-UTR       |
| rat-2         | 5’-CAGGGAATCCCAATCAAAGAGC-3’ | s            | 5’-UTR       |
| rat-3         | 5’-GTTGAAATGCTGTCATAACAGGCTC-3’ | as       | 5’-UTR       |
| rat-4         | 5’-AACCGAGGTACAGATCTCTTCTTG-3’ | as       | 5’-UTR       |
| rat-5         | 5’-TCCAACCACCTGGGCTGTCAC-3’ | s            | 5’-UTR       |
| rat-6         | 5’-GACACAAATCTTTTGCGCTGTCAC-3’ | s            | 5’-UTR       |
| rat-7         | 5’-ATTTCATTGGCTTATGGGCCCATC-3’ | as       | 5’-UTR       |
| rat-8         | 5’-CTAGCTGCAGGTGTTCTAGAGG-3’ | s            | 5’-UTR       |
| rat-9         | 5’-CCCTACAGCTACAGACTCTATTC-3’ | as       | 5’-UTR       |
| rat-10        | 5’-TCTAAGGAGCTAAGTCCACTG-3’ | s            | 5’-UTR       |
| rat-11        | 5’-GCTGCATGACAAGTTTTGAG-3’ | s            | 5’-UTR       |
| rat-12        | 5’-CTAAAATCTTTTGATGTGTAATTCG-3’ | s            | 5’-UTR       |
| rat-13        | 5’-GTAAGGACTGACCTTCAAGAG-3’ | as       | 5’-UTR       |
| rat-14        | 5’-TGAAACAACAATAAAGATTAATATG-3’ | as       | 5’-UTR       |
| rat-15        | 5’-GGAGTATTGGAAACACACCAGTGCTG-3’ | as       | cod reg      |
| rat-16        | 5’-CCATGAAAGGCGACAGAATCAATGTGG-3’ | as       | cod reg      |
| rat-17        | 5’-CCCTCTCTCCTGACAGAACACTC-3’ | s            | cod reg      |
| rat-18        | 5’-CGAGAACAGGACTGCTCAGAATCC-3’ | s            | cod reg      |
| rat-19        | 5’-CAGCAAGCCCTTCAAGATCATG-3’ | s            | cod reg      |
| rat-20        | 5’-AATACAAAAATCAGAGGGATTAAACT-3’ | as       | 3’-UTR       |
| rat-21        | 5’-AGCTAAAATCTGGCATTCTGTC-3’ | s            | 3’-UTR       |
| rat-22        | 5’-CTGCAAGAGAGAACACTCTGTATTC-3’ | s            | 3’-UTR       |
| rat-23        | 5’-GCTGGAATATTTATATGTAGATG-3’ | as       | 3’-UTR       |
| rat-24        | 5’-CACATCTCTACACACTTTGTCG-3’ | s            | 3’-UTR       |
| rat-25        | 5’-CCAAGCGCAACTCTTATAAAAGGAC-3’ | as       | 3’-UTR       |
| rat-26        | 5’-TATTCACAAGAATTTAACCTAAAAGAC-3’ | s            | 3’-UTR       |
| rat-27        | 5’-GCCCTCCTAGGCCTGAGAGG-3’ | s            | 3’-UTR       |
| rat-28        | 5’-CACAGAGTGCAGTCTCTCTG-3’ | s            | 3’-UTR       |
| rat-29        | 5’-ACTCAAGACACACGCCCACCAG-3’ | s            | 3’-UTR       |
| rat-30        | 5’-GTTATGCGGGCAAGAGTGTCTTG-3’ | as       | 3’-UTR       |
| rat-31        | 5’-GTGTTTGGGATATTTCCGTTGTC-3’ | as       | 3’-UTR       |
**Suppl. Table S6**

**Determination of mutation rates in the GPR33 locus of Hominoidea and rats**

The number of nucleotide differences in the genomic non-coding and coding regions between the indicated GPR33 alleles of human, chimpanzee and bonobo (upper panel; 5,447 bp total) and *R. norvegicus* and *R. rattus* (lower panel; 5,293 bp total) was counted. Data are given as nucleotide differences per 1,000 bp.

| 5,447 bp total | Human Stop140 | Human Arg140 | Bonobo  | Chimpanzee Ser39/Arg140 | Chimpanzee Stop39/Arg140 |
|----------------|---------------|--------------|---------|------------------------|-------------------------|
| Human Arg140   | 2.2           | -            | 9.4     | 9.4                    | -                       |
| Bonobo         |               |              | 9.6     | 9.4                    | 2.8                     |
| Chimpanzee Ser39/Arg140 | 9.4     | 9.2           | 2.6     | 1.3                    | -                       |
| Chimpanzee Stop39/Arg140 | 9.9     | 10.1          | 3.5     | 1.1                    | 2.0                     |

| 5,293 bp total | *R. norvegicus* A | *R. norvegicus* B |
|----------------|-------------------|-------------------|
| *R. norvegicus* B | 3.4               | -                 |
| *R. rattus*      | 30.2              | 29.5              |
Putative Yoruba haplotypes

The GPR33 coding region (~1 kb) and 3.8 kb of the-3’ untranslated region from 21 Yoruba individuals were directly sequenced. The most likely haplotypes for each individual were calculated with PHASE.

### Nucleotide position in Chr14 (29938165-29943210)

| Nucleotide position starting with Start-ATG in GPR33 | 418 | 428 | 915 | 1215 | 1623 | 1842 | 2025 | 2195 | 2277 | 2681 | 2720 | 2829 | 3044 | 3047 | 3063 | 3093 | 3107 | 3121 | 3132 | 3651 | 3726 | 3787 | 3901 | 4143 | 4198 | 4470 |
|-----------------------------------------------------|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|

### ANCESTRAL STATE

| Chimpanzee                           | C     | G     | C     | G     | C     | A     | T     | A     | T     | A     | A     | A     | T     | T     | A     | T     | A     | A     | C     | A     | C     | A     | G     | G     | T     | G     | T     | A     |
|--------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Human CGA Filipino                   | C     | G     | C     | G     | C     | A     | T     | A     | A     | T     | A     | A     | A     | T     | T     | A     | A     | A     | C     | A     | G     | G     | C     | T     | G     | C     | T     | W     |
| Yoruba-21                            | 1     | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-20                            | 2     | T     | T     | -     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-19                            | 3     | T     | T     | -     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-18                            | 4     | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-17                            | 5     | T     | T     | -     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-16                            | 6     | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-14                            | 7     | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-13                            | 8     | T     | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-12                            | 9     | T     | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-11                            | 10    | T     | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-10                            | 11    | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-9                             | 12    | T     | T     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     | -     |
| Yoruba-15 | 13 T - T - T T - - - - - - - - - - - - - - - - - A - - T - - |
|-----------|----------------------------------|
| Yoruba-14 | 11 T - T - - - - - C G - - C - - - - - - - - - - - - A - - T - |
| Yoruba-13 | 14 T - T T - - - - - G - - C - - - - C - - - - A - - T C - |
| Yoruba-12 | 15 T - T - T - - G - - - - - - - - - - - - - - - - - - - - T |
| Yoruba-11 | 16 T - T - - - - - G C - - - - - - - - - - - - - - - - - - T |
| Yoruba-10 | 17 T A T - - - T - - - - - - - - - - - - - - - - A - - T - |
| Yoruba-9  | 18 T - T - T - - - - - - - - G - - - - - - - - - - A - - T - |
| Yoruba-8  | 19 T A T - - - - - - - - - - - - - - - - - - - - - - - - T |
| Yoruba-7  | 20 T - T - T - - - - - - - C - - - - - - - - - - A - - T - |
| Yoruba-6  | 21 T - T - T - - - - - - - - - - - - - - - - - - A - - T - T |
| Yoruba-5  | 22 T - T - - - - - - C G - - - - - - - - - - - - - - A - - T - |
| Yoruba-4  | 23 T - T - T - - - - - - - - - - - - - - - - - - A - - T - |
| Yoruba-3  | 24 T A T - T - T - - - - - - - - - - - - - - - - A - - T C - |
| Yoruba-2  | 25 T - T - - - - - - - - - - - - - - - - - - A - - T C - T |
| Yoruba-1  | 26 T - T - - - - - - C G - - - - - - - - - - - - - - - - - A C - T - T |
| Yoruba-10 | 27 T - T - T - - - - - - - - - - - - - - - - - - C - - T - T |
| Yoruba-9  | 28 T - T - - - - - - - - - - - - - - - - - - A - - T - - |
| Yoruba-8  | 29 T - T - - - - - - - - - - - - - - - - - - - - - C - A - T |
| Yoruba-7  | 30 T A T T - - - G - - - - - - - - - - - - - - - - A - - T - |
| Yoruba-6  | 31 T - T - T - - - - - - - - - - - - - - - - - - C - - T A - T - |
| Yoruba-5  | 32 T - T - - - - - - - - - - - - - - - - - - A - - T - - |

1. Yoruba-15: 13 T - T - T T - - - - - - - - - - - - - - - - - A - - T - -
2. Yoruba-14: 11 T - T - - - - - C G - - C - - - - - - - - - - - - A - - T -
3. Yoruba-13: 14 T - T T - - - - - G - - C - - - - C - - - - A - - T C -
4. Yoruba-12: 15 T - T - T - - G - - - - - - - - - - - - - - - - - - - - T
5. Yoruba-11: 16 T - T - - - - - G C - - - - - - - - - - - - - - - - - - T
6. Yoruba-10: 17 T A T - - - T - - - - - - - - - - - - - - - - A - - T -
7. Yoruba-9: 18 T - T - T - - - - - - - - G - - - - - - - - - - A - - T -
8. Yoruba-8: 19 T A T - - - - - - - - - - - - - - - - - - - - - - - - T
9. Yoruba-7: 20 T - T - T - - - - - - - C - - - - - - - - - - A - - T -
10. Yoruba-6: 21 T - T - T - - - - - - - - - - - - - - - - - - A - - T - T
11. Yoruba-5: 22 T - T - - - - - - C G - - - - - - - - - - - - - - - - - A C - T - T
12. Yoruba-4: 23 T - T - T - - - - - - - - - - - - - - - - - - A - - T -
13. Yoruba-3: 24 T A T - T - T - - - - - - - - - - - - - - - - A - - T C -
14. Yoruba-2: 25 T - T - - - - - - - - - - - - - - - - - - A - - T C - T
15. Yoruba-1: 26 T - T - - - - - - C G - - - - - - - - - - - - - - - - - A C - T - T
16. Yoruba-10: 27 T - T - T - - - - - - - - - - - - - - - - - - C - - T - T
17. Yoruba-9: 28 T - T - - - - - - - - - - - - - - - - - - A - - T - -
18. Yoruba-8: 29 T - T - - - - - - - - - - - - - - - - - - - - - C - A - T
19. Yoruba-7: 30 T A T T - - - G - - - - - - - - - - - - - - - - A - - T -
20. Yoruba-6: 31 T - T - T - - - - - - - - - - - - - - - - - - C - - T A - T -
21. Yoruba-5: 32 T - T - - - - - - - - - - - - - - - - - - A - - T - -
The rise and fall of the chemoattractant receptor GPR33
Holger Rompler, Angela Schulz, Christian Pitra, Graham Coop, Molly Przeworski, Svante Paabo and Torsten Schoneberg

J. Biol. Chem. published online June 29, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M503586200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2005/07/08/M503586200.DC1