ORIGINAL INVESTIGATION

Increased constraints on MC4R during primate and human evolution

**David A. Hughes · Anke Hinney · Harald Brumm ·
Anne-Kathrin Wermter · Heike Biebermann ·
Johannes Hebebrand · Mark Stoneking**

Abstract The melanocortin 4 receptor (*MC4R*) is routinely investigated for the role it plays in human obesity, as mutations in *MC4R* are the most common dominantly inherited form of the disease. As little is known about the evolutionary history of this locus, we investigated patterns of variation at *MC4R* in a worldwide sample of 1,015 humans from 51 populations, and in 8 central chimpanzees. There is a significant paucity of diversity at *MC4R* in humans, but not in chimpanzees. The spectrum of mutations in humans, combined with the overall low level of diversity, suggests that most (if not all) of the observed non-synonymous polymorphisms are likely to be transient deleterious mutations. The *MC4R* coding region was resequenced in 12 primate species and sequences from an additional 29 vertebrates were included in molecular evolutionary analyses. *MC4R* is highly conserved throughout vertebrate evolution, and has apparently been subject to high levels of continuous purifying selection that increased approximately threefold during primate evolution. Furthermore, the strong selection extends to codon usage bias, where most silent mutations are expected to be either quickly fixed or removed from the population, which may help explain the unusually low levels of silent polymorphisms in humans. Finally, there is a significant tendency for non-synonymous mutations that impact *MC4R* function to occur preferentially at sites that are identified by evolutionary analyses as being subject to very strong purifying selection. The information from this study should help inform future epidemiological investigations of *MC4R*.

D. A. Hughes (✉) · M. Stoneking
Department of Evolutionary Genetics,
Max Planck Institute for Evolutionary Anthropology,
Deutscher Platz 6, Leipzig, Germany
e-mail: hughes@eva.mpg.de

A. Hinney · J. Hebebrand
Department of Child and Adolescent Psychiatry,
University of Duisburg-Essen, Virchowstr. 174,
45147 Essen, Germany

H. Brumm · H. Biebermann
Institute of Experimental Pediatric Endocrinology,
Charité Universitätsmedizin Berlin,
Augustenburger Platz 1, 13353 Berlin, Germany

A.-K. Wermter
Department of Child and Adolescent Psychiatry,
Philipps-University, Schuetzenstr. 49, 35039 Marburg, Germany

Introduction

With approximately 1.6 billion overweight individuals and at least 400 million clinically obese adults world wide, obesity is one of the world's greatest health concerns. It is a disease which leads to associated health complications including diabetes, cardiovascular disease, musculoskeletal disorders, and cancer (http://www.who.int/topics/obesity). As a result obesity is the focus of much scientific interest. Among several genes that have been associated with obesity, the melanocortin 4 receptor gene (*MC4R*) has been intensively studied, as mutations in *MC4R* account for about 6% of clinical obesity (Farooqi et al. 2003), more than any other single gene. Moreover, homozygous *Mc4r* knockout mice ($-/-$) become obese as a result of hyperphagia, while heterozygous knockout mice ($\pm$) are intermediate in weight and appetite to that of wild-type mice

(Huszar et al. 1997). MC4R is a G-protein coupled receptor (GPCR) that is critical for energy homeostasis. It is involved in the melanocortin system, controlling appetite and energy expenditure, is expressed predominantly in the hypothalamic nucleus paraventricularis of the brain and is activated by melanocyte stimulating hormones, which are processed from proopiomelanocortin by prohormone convertases (Butler 2006). Upon activation, MC4R coupled to the Gs/adenylyl cyclase signaling system leads to increased levels of intracellular cyclic AMP (cAMP), which in turn signals the brain to decrease energy intake, i.e. food consumption (Butler 2006).

Many studies have screened the *MC4R* locus in thousands of obese probands to identify reduced or loss of function mutations associated with dominantly inherited obesity (Biebermann et al. 2003; Branson et al. 2003; Dubern et al. 2001; Farooqi et al. 2003; Gu et al. 1999; Hinney et al. 2003; Hinney et al. 1999; Kobayashi et al. 2002; Larsen et al. 2005; Lubrano-Berthelier et al. 2003; Miraglia Del Giudice et al. 2002; Santini et al. 2004; Srinivasan et al. 2004; Vaisse et al. 2000; Valli-Jaakola et al. 2004; Yeo et al. 2003). To date about 90 such mutations have been identified and functionality determined by an in vitro cAMP accumulation assay. A quantitative effect of *MC4R* mutations on body mass index (BMI) has also been observed (Dempfle et al. 2004); individuals who are homozygotes or compound heterozygotes for *MC4R* mutations (non-synonymous, frameshift or premature stop mutations) are more obese than individuals who are heterozygous for the same allele (Hinney et al. 2006), consistent with observations in *Mc4r* knockout mice. However, it is not uncommon for heterozygous relatives of obese probands to not exhibit early onset obesity (Dempfle et al. 2004). Thus, there are undoubtedly complex interactions between genotype, genetic background and environment which influence the probability and extent of obesity, given a particular *MC4R* mutation.

Mutations at the *MC4R* locus appear to be rare in humans. The mean allele frequency of non-synonymous and synonymous mutations in a sample of 4,068 Germans in the KORA study (Hinney et al. 2006) was 0.021%, exemplifying the rarity of *MC4R* mutations in a random population sample. However, little is known about variation in *MC4R* in worldwide populations of humans, or about the molecular evolution of *MC4R*. Here, we analyze *MC4R* variation in 1,051 individuals from 51 worldwide populations. To determine if *MC4R* variation in humans is unusual, we sequenced *MC4R* from eight chimpanzees. In addition to humans and chimpanzees we also sequenced *MC4R* from 10 other primate species, and analyzed the molecular evolution of *MC4R* in a total of 41 vertebrate species, to determine the mode of evolution (purifying, neutral or adaptive) that has been acting on *MC4R* across

420 million years of vertebrate evolution, as well as to determine if the pattern of evolution has changed on any particular lineage during this time. Given previous observations of human MC4R variation (Hinney et al. 2006) and GPCR conservation (Rompler et al. 2007), we hypothesize that MC4R has been largely subject to purifying selection throughout vertebrate evolution, and that MC4R variants in human populations will be rare, deleterious, de novo mutations.

Materials and methods

Human DNA samples and *MC4R* mutation screening

The *MC4R* coding sequence from the Human Genome Diversity Panel (HGDP-CEPH), which consists of 1,051 individuals from 51 human populations worldwide (Cann et al. 2002), was analyzed by denaturing high performance liquid chromatography (dHPLC) as described previously (Hinney et al. 1999, 2003). Mutation screens of two overlapping fragments were performed using dHPLC analysis (Yu et al. 2006) on a Transgenomic WAVE® system (Transgenomic, Cheshire, UK). PCRs were performed with primers amplifying *MC4R* in two overlapping parts with the primer pairs MC4R-1F/MC4R-1R (F1; 615 bp) and MC4R-2F/MC4R-2R (F2, 622 bp); primer sequences and PCR conditions were as described previously (Hinney et al. 1999). The optimal melting temperatures for separation of homo- and hetero-duplexes were deduced from the melting temperature of the PCR-amplicon using WAVEmaker software, version 4.0 (Transgenomic, Cheshire, UK). Five microliters of unpurified denatured–renatured PCR products were loaded onto a preheated specific column (DNASep column, Transgenomic) and eluted from the column by a linear acetonitrile gradient in a 0.1-m triethylammonium acetate (TEAA) buffer, pH 7, and at a constant flow rate of 0.9 ml/min. The gradient was created by mixing eluents A (0.1 m TEAA) and B (0.1 m TEAA, 25% acetonitrile). For analyses of the *MC4R* F1-fragment, PCR products were eluted from the column (61°C) by increasing buffer B from 54 to 59% within 30 s and from 59 to 68% within 4.5 min. The oven temperature for analyzing the MC4R F2-fragment was set to both 59 and 60°C, to insure detection of variants. For elution of the MC4R F2 products, buffer B was increased from 54 to 59% within 10 s and from 59 to 68% within 4.5 min. All chromatograms were compared with chromatograms of sequenced wild-type samples; PCR amplicons with more than one peak or differences in peak appearance from the wild-type pattern were sequenced (SeqLab, Göttingen, Germany).

MC4R functional assays

Generation of N-terminal HA-tagged mutant receptors

For functional characterization, wild type and mutant MC4 receptors were amplified from genomic DNA using a forward primer containing the Kosac sequence and an N-terminal hemeagglutinine (HA) tag. Wild-type and mutant *MC4R*s were cloned into the pcDps expression vector (Prof. Schöneberg, Leipzig) and sequenced (3100 automated sequencer, Applied Biosystems) to ensure correctness.

Cell culture and transfection

COS-7 cells were maintained in Dulbecco's modified Eagle's medium (DMEM, Sigma, Deisenhausen) supplemented with 10% fetal calf serum (FCS) and 20 mM Glutamin. Cells were incubated at 37°C in humidified air containing 5% CO_2. Transfections were carried out using Metafectene (Biontex, Munich) according to the manufacturer's protocol. Cyclic AMP assays were performed in 12 well plates (2×10^5 cells/well). Cells were transfected with 0.25 µg DNA/well and 1.25 µl Metafectine/well. Cell surface expression assays were performed in 48 well plates (2×10^5 cells/well) and cells were transfected with 0.25 µg DNA/well and 1 µl Metafectene/well.

Measurement of cAMP accumulation

After 48 h of transfection, cells were labeled with 2 µCi/ml of [^{3}H]adenine (Amersham) and incubated overnight. Cells were washed in serum-free DMEM containing 1 mM 3-iso-butyl-1-methylxanthine (Sigma, Deisenhausen) and incubated for 1 h at 37°C in the presence of α-MSH, β-MSH and NDP-α-MSH respectively, in a ten-step series with the highest concentration of 1,000 nM (α- and β-MSH) or 100 nM (NDP-α-MSH). Aspiration of the medium and addition of 1 ml ice cold 5% trichloroacetic acid terminated receptor stimulation. The measurement of intracellular cAMP in the supernatant was carried out as described elsewhere (Salomon et al. 1974).

Measurement of cell surface expression

After 72 h of transfection, cells were washed two times with Dulbecco's phosphate buffered saline (DPBS, Biochrom, Berlin) and fixed for 30 min in 4% formaldehyde in DPBS followed by two times washing in DPBS. After incubation in blocking buffer (10% FCS supplemented DMEM) for 1 h at 37°C followed by a washing step in DPBS, cells were incubated for 2 h in blocking buffer with 1 µg/ml biotin labeled anti-HA monoclonal antibody (Roche, Mannheim) at 37°C followed by three washes in DPBS and

incubation in blocking buffer with 1 µg/ml streptavidin labeled peroxidase (Dianova, Hamburg) at 37°C for 1 h followed by three times washing. The color reaction was carried out with 0.1% H_2O_2 and 10 µg *o*-phenylendiamine in 0.1 M citric acid and 0.1 M Na_2HPO_4 at pH 5.2. The reaction was stopped after 10 min with 1 M Na_2SO_3 in 1 M HCl. Colorimetry was carried out using an anthos reader 2001 (anthos labtech instruments, Salzburg).

MC4R resequencing in primates

The *MC4R* coding sequence was resequenced in seven Ethiopians, six Germans, eight central chimpanzees (Ptr, *Pan troglodytes*) from Gabon (Fischer et al. 2004), and one individual from each of ten other primate species, including: bonobo (Ppa, *Pan paniscus*), gorilla (Ggo, *Gorilla gorilla*), orangutan (Ppy, *Pongo pygmaeus*), crab-eating macaque (Mfa, *Macaca fascicularis*), baboon (Pha, *Papio hamadryas*), gray langur (Sen, *Semnopithecus entellus*), marmoset (Cja, *Callithrix jacchus*), cotton-top tamarin (Soe, *Saguinus oedipus*), howler monkey (Aca, *Alouatta caraya*), and spider monkey (Abe, *Ateles belzebuth*). The Ethiopian and German samples were described previously (Kayser et al. 2003); the bonobo, gorilla, and orangutan DNA samples were provided by Svante Pääbo of the Max Plank Institute for Evolutionary Anthropology; and the macaque, baboon, langur, marmoset, cotton-top tamarin, howler monkey, and spider monkey DNA samples were donated by Christian Roos of the German Primate Center (DPZ).

PCR reactions were carried out in a RoboCycler 96 Gradient Cycler (Stratagene-Agilent Technologies, La Jolla, CA, USA), consisting of an initial denaturing and *Taq* activation step at 95°C for 15 min, followed by 35 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 2 min, and a final extension step at 72°C for 10 min. Reactions (25 µl) consisted of 7.5 ng of template DNA, 1× PCR buffer (Applied Biosystems, Foster City, CA, USA), 200 µM dNTPs (Amersham Biosciences, Uppsala, Sweden), 400 nM of each primer (*MC4R*_Fa: GGATTGGTCAGAAGGAAGC and *MC4R*_Ra: GATATTCTCAACCAGTACCCT ACAC) (Biotez, Berlin, Germany) and 1.25 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA, USA). PCR products were subsequently sequenced in both directions in a total volume of 10 µl. PCR primers and four other primers (*MC4R*_Fb: ACACTTCTCTGCACCTCTG GAAC, *MC4R*_Rb: CAGCAATCCTCTTAATGTGAAGC, *MC4R*_Fc: AACATTATGACAGTTAAGCGG GT and *MC4R*_Rc: GGAGAGACAAAAAGTTGCTCGTA) were used as sequencing primers in reactions consisting of 1 µl of Applied Biosystems Big Dye Terminator v 1.1, 0.75× sequencing buffer, 320 nM primer (Biotez, Berlin, Germany) and ~12 ng of amplified template DNA. Sequencing

was carried out in an ABI 3730 DNA Analyzer. Sequence data were analyzed manually with Applied Biosystems SeqScape Software V2.5 (Applied Biosystems, Foster City, CA, USA) and with comparison to a reference sequence (NCBI accession number NT_025028, bases: 5829430..5830428).

Additional coding sequence sources

MC4R Coding sequences (cds) from 29 species [bushbaby (0ga, *Otolemur garnettii*), tree shrew (Tbe, *Tupaia belangeri*), mouse (Mmu, *Mus musculus*), mouse (Msp, *Mus spretus*), rat (Rno, *Rattus norvegicus*), guinea pig (Cpo, *Cavia porcellus*), squirrel (Str, *Spermophilus tridecemlineatus*), red fox(Vvu, *Vulpes vulpes*), artic fox (Ala, *Alopex lagopus*), raccoon dog (Npr, *Nyctereutes procyonoides*), dog (Cfa, *Canis familiaris*), cat (Fca, *Felis catus*), horse (Eca, *Equus caballus*), boar (Ssc, *Sus scrofa*), cow (Bta, *Bos Taurus*), microbat (Mlu, *Myotis lucifugus*), hedgehog (Eeu, *Erinaceus europaeus*), elephant (Ema, *Elephas maximus*), opossum (Mdo, *Monodelphis domestica*), platypus (Oan, *Ornithorhynchus anatinus*), chicken (Gga, *Gallus gallus*), goose (Aan, *Anser anser*), clawed frog (Xtr, *Xenopus tropicalis*), fugu (Tru, *Takifugu rubripes*), stickleback (Gac, *Gasterosteus aculeatus*), medaka (Ola, *Oryzias latipes*), zebrafish (Dre, *Danio rerio*), goldfish (Cau, *Carassius auratus*)] were downloaded from ENSEMBL (http://www.ensembl.org) and the NCBI core nucleotide (http://www.ncbi.nlm.nih.gov) databases. Species and accession numbers can also be found in Supplemental Table 1.

Data analysis

Coding sequences were converted into amino acid sequences with MEGA 3.1 (Kumar et al. 2004), aligned with MUSCLE 3.6 (Edgar 2004), exported in CLUSTAL W format, and converted into a cds alignment with a custom PERL script. The alignment is provided in Supplemental Table 3. A species tree (Benton and Donoghue 2007; Bininda-Emonds et al. 2007) was generated and the transition/transversion bias (κ) was estimated using the CODEML program in the PAML 4a package (Yang 2007). The PHYLIP package (Felsenstein 2004) was used to construct nucleotide and amino acid neighbor-joining (NJ) trees. The program dnadist was used to estimate nucleotide distances, using the estimated κ (2.58) and both the F84 and Kimura substitution models. Amino acid distances were estimated with the program protdist using both the JTT and PAM substitution matrixes, and assuming a gamma distribution for the rate variation. Finally, the program neighbor was used to generate NJ trees. Sequences were randomized on input. The species tree and constructed nucleotide and amino acid trees were analyzed by the K–H and S–H test

(Kishino and Hasegawa 1989; Shimodaira and Hasegawa 1999), as implemented in CODEML of the PAML 4a package, to determine if any one tree significantly fit the data better than alternative trees.

The inferred coding sequence for each individual from the HGDP-CEPH panel, as well as for each resequenced Ethiopian, German, chimpanzee and other primate, were analyzed in DNAsp 4.0 (Rozas et al. 2003). Summary statistics for 319 genes, including sequence length, number of segregating sites, nucleotide diversity, and Tajima's D, were downloaded from the SeattleSNPs database (http://pga.mbt.washington.edu/) in May 2008, and used as a comparative empirical dataset. The estimated number of non-synonymous and synonymous sites for the genes in the SeattleSNPs database was taken from previously published estimates (Bustamante et al. 2005); such estimates were available for 268 genes. Neutral simulations were performed with Hudson's ms (Hudson 2002). For the human data, simulations were carried out using demographic parameters found to provide the best fit, based on empirical data from three human populations: Africans, Europeans, and East Asians (Schaffner et al. 2005). The model used allows for migration between each population, includes a bottleneck for non-African populations, growth in Africans, and a post-agricultural expansion in each population. For chimpanzee data, simulations included a current effective population size of 118,000 (N_2), an ancestral population size of 16,000 (N_1), and growth from 500,000 years ago to today from N_1 to N_2 (Caswell et al. 2008). Simulations were based on the observed number of segregating sites (S), and for humans on a recombination rate of 0.00103/base (Ptak et al. 2004) across the entire sequence length of 999 bp. The significance of observed summary statistics was then determined by comparing empirical observations to 10,000 simulated loci.

Amino acid sequence evolution

The role of selection across the *MC4R* phylogeny was investigated by analyzing patterns of evolution at the codon level. The non-synonymous to synonymous substitution rate ratio (ω) was estimated by maximum likelihood methods (Goldman and Yang 1994) using the CODEML programs of the PAML4a package (Yang 2007). Three different codon substitution models were used in each analysis to insure the robustness of the results: the F3X4 codon substitution model (Goldman and Yang 1994) derives equilibrium codon frequencies from each of the three codon positions using three nucleotides, and thus has 9 free parameters; the F61 substitution model (Goldman and Yang 1994) uses each codon as a free parameter, estimating codon frequencies from the empirical distribution; and the recently developed FMutSel codon substitution model

(Yang and Nielsen 2008) specifically models mutational biases and selection at silent sites by introducing a parameter of codon fitness. This latter model was developed to address possible selection at silent sites resulting in codon usage bias. Furthermore, several models were fit to the data which allow ω to vary across sites, across branches (Yang 1998; Yang and Nielsen 1998), and across clades (Bielawski and Yang 2004). These models and analyses allow inferences concerning the type of selection (neutral, negative, or positive) which has acted across a phylogeny, as well on specific branches and/or sites. A likelihood ratio test (LRT = $2(l_1 - l_0)$, where l_1 and l_0 are the log likelihoods of the two models) was used to determine if one model fit the data significantly better than another model.

Results

HGDP-CEPH screening

The HGDP-CEPH panel was screened for polymorphisms at the *MC4R* locus by dHPLC (WAVE) melting curve technology, and aberrant PCR products were re-sequenced to identify polymorphisms. Fourteen segregating sites (S) were found across 1,074 bps in 2,030 chromosomes, with an average variant allele frequency of only 0.17% (Table 1). The most frequent polymorphism (V103I) had a variant allele frequency of just 1.33%. Eleven of the segregating sites are non-synonymous mutations, two are synonymous mutations and one is found in the 5′ region. The mean nucleotide diversity estimate (π) for *MC4R* is 0.00005, which is lower than the π value for any of the 319 genes in the SeattleSNPs database (mean $\pi = 0.0009$, minimum $\pi = 0.00016$). The π value for *MC4R* is also significantly lower than expected (Pvalue = 0.001) based on simulations that incorporate the best-fitting demographic parameters (Schaffner et al. 2005) for three human populations (Africans, Europeans, and Asians). Since the HGDP-CEPH includes more population structure than would be observed in just these three human populations (Li et al. 2008; Rosenberg et al. 2002, 2005), the comparison of π values from the HGDP-CEPH to the empirical SeattleSNPs data and to the simulations based on three populations is conservative, making the extremely low level of diversity at *MC4R* all the more unusual. In addition to the dHPLC screening of the HGDP-CEPH panel, the *MC4R* locus was amplified and directly sequenced in six Germans and seven Ethiopians. No mutations were identified in these 26 chromosomes, consistent with the low levels of diversity found in the HGDP-CEPH screening. Thus the low level of diversity found in this worldwide sample is not an artifact of the dHPLC technology.

This extremely low level of diversity could reflect either very recent directional selection in all human populations, or continuous purifying selection. Under neutral expectations, the rate of mutations at non-synonymous and synonymous sites are expected to be equal, i.e. the intra-species ratio of the non-synonymous to synonymous polymorphism rate per codon site should equal 1 (pN/pS = ~1). The observed pN/pS value for *MC4R* is 1.98, but this is not significantly greater than 1 ($\chi^2 = 0.826$, $P = 0.362$). However, the pN/pS ratio for *MC4R* is significantly larger ($P = 0.008$) than pN/pS estimates for 268 genes from the SeattleSNPs database (SeattleSNPs), and moreover it is in the top 1% of estimated values from the Celera exon re-sequencing dataset (Bustamante et al. 2005), when considering genes with informative polymorphic sites at both non-synonymous and synonymous positions. The excess of mutations at non-synonymous sites would seem to suggest that directional selection (or relaxed constraints) may be acting on *MC4R*, resulting in the accumulation of non-synonymous mutations that may alter protein function. However, a caveat to this observation is that no polymorphism in the CEPH data is above 1.4% in frequency. It is common for low-frequency alleles to be removed from analyses of polymorphic synonymous and non-synonymous mutations, as slightly deleterious mutations segregate at such frequencies and bias estimates of intra-population variation and tests of selection, including the MK test (Charlesworth 1994; Eyre-Walker 2006; Fay et al. 2001). However, removing low-frequency alleles in the MC4R dataset would remove all of the variants. Eight of the observed polymorphisms are singletons, three are doubletons, two are observed in four chromosomes, and V103I, which has been dubbed a protective allele (Heid et al. 2008), is the most common non-synonymous allele observed in 27 out of 2,030 chromosomes (frequency = 1.33%). Thus all non-synonymous mutations are rare and are likely to be transient, deleterious mutations which will ultimately be removed from the population (Boyko et al. 2008). In regard to synonymous polymorphisms, a possible explanation for their rarity in *MC4R*, beyond the persistent problem of background selection, is codon usage bias resulting in selection on silent sites; we explore this possibility in more detail below.

Functional characterization of non-synonymous mutations

Eight of the eleven observed non-synonymous mutations in the HGDP-CEPH screening were tested in cAMP accumulation and cell surface accumulation assays to measure the effect of each mutation on MC4R function as compared to the wild-type (Table 2); the remaining three sites (R7C, I251L and V103I) have been previously analyzed (Srinivasan et al. 2004; Stutzmann et al. 2007; Xiang et al. 2006). Only one of the 11 non-synonymous mutations (N62S) exhibited a reduced function, as indicated by a strongly reduced cell surface expression, and by signal transduction

Table 1 CEPH panel mutations

Variable site (bp)	Amino acid	Mutation	Freq (%)	Population(s)	Individual(s)-HGDP#
−59	Promoter region	G/A	0.05	Sardinian	1071
19	Arg-7-Cys	C/T	0.05	Russia	882
106	Ser-36-Thr	T/A	0.05	Bantu	993
110	Asp-37-Gly	A/G	0.05	Palestine	679
119	Cys-40-Tyr	G/A	0.05	Russia	884
185	Asn-62-Ser	A/G	0.05	Yizu	1181
307	Val-103-Ile	A/G	1.33	Biaka, Yoruba, Bantu, Mozabite, Druze, Palestine, Brahui, Balochi, Makrani, Pathan, Burusho, Tujia, Yizu, Japanese, French, North Italian, Orcadian	01086, 00926, 01411, 01413, 01419, 01265, 01280, 01281, 00572, 00695, 00011, 00017, 00043, 00078, 00086, 00136, 00143, 00262, 00346, 00364, 01096, 01183, 01185, 00756, 00539, 01171, 00797
429	Ile-143-Ile*	T/A	0.1	NAN Melanesian	00788, 00824
496	Val-166-Ile	G/A	0.2	Mongola, Hezhen, Xibo	01226, 01229, 01238, 01243
508	Ile-170-Val	A/G	0.1	Bantu, Han	00994, 01290
572	Ser-191-Thr	G/C		Adygei	1402
594	Ile-198-Ile*	C/T	0.3	Mandenka, Yoruba, Pathan, Bedouin, Sindhi	00910, 00929, 00936, 00226, 00619, 00175
606	Phe-202-Leu	C/A	0.1	Bedouin, Sindhi	00619, 00175
751	Ile-251-Leu	A/C	0.05	French Basque	1376

All mutations found in the CEPH panel. *Variable site* is the position of the observed mutation in accordance with coding sequence location, *Amino Acid* describes the type of mutation observed, *Mutation* is the specific base change observed, *Population(s)* lists the populations for which the mutation was observed, *Individual(s)-HGDP#* are the CEPH individuals in which the mutation was found identified by their HGDP number

Table 2 MC4R functional assays

Construct	Basal (Basal/BasalWT)	cAMP accumulation						Cell surface expression (% of WT-MC4R)	Function	Site class
		NDP-α-MSH		α-MSH		β-MSH				
		E_{max}/E_{max} WT (%)	EC_{50} (nM)	E_{max}/E_{max} WT (%)	EC_{50} (nM)	E_{max}/E_{max} WT (%)	EC_{50} (nM)			
MC4R-WT	1	100	0.36 ± 0.31	100	2.2 ± 1.0	100	10 ± 5	100	NA	NA
R7C	NA	NA	NA	NA	NA	NA	NA	NA	LWT*	2
S36T	0.83 ± 0.23	88 ± 20	0.46 ± 0.18	81 ± 15	6.4 ± 4.9	82 ± 18	32 ± 11	107 ± 8	LWT	2
D37G	0.83 ± 0.28	83 ± 18	0.31 ± 0.18	78 ± 16	1.4 ± 0.5	75 ± 13	15 ± 12	100 ± 19	LWT	2
C40Y	1.08 ± 0.30	103 ± 10	0.49 ± 0.30	99 ± 11	3.0 ± 0.8	98 ± 8	19 ± 12	49 ± 8	LWT	0
N62S	0.85 ± 0.35	22 ± 8	Nd	28 ± 6	Nd	20 ± 7	Nd	51 ± 4	RF	0
V103I	NA	NA	NA	NA	NA	NA	NA	NA	LWT*	0
V166I	1.20 ± 0.29	122 ± 9	0.31 ± 0.07	87 ± 8	3.8 ± 1.1	105 ± 20	29 ± 14	49 ± 14	LWT	0
I170 V	1.01 ± 0.29	98 ± 21	0.16 ± 0.13	75 ± 16	6.4 ± 2.7	97 ± 13	38 ± 24	72 ± 21	LWT	0
S191T	0.96 ± 0.31	99 ± 7	0.98 ± 0.43	101 ± 27	3.2 ± 1.2	97 ± 40	10 ± 3	69 ± 18	LWT	1
F202L	0.76 ± 0.22	83 ± 23	0.30 ± 0.12	79 ± 17	4.1 ± 3.1	79 ± 23	16 ± 10	81 ± 6	LWT	0
I251L	NA	NA	NA	NA	NA	NA	NA	NA	LWT*	0

MC4R cAMP accumulation and cell surface expression assays. Each amino acid replacement observed in the HGDP-CEPH panel is found in Table. Three sites (R7C, V103I, I251L) were not functionally tested in this study as these sites have been identified and assayed previously (Srinivasan et al. 2004; Xiang et al. 2006; Stutzmann et al. 2007; Xiang et al. 2006). *Construct* identifies the derived amino acid residue assayed. *cAMP accumulation* denotes the amount of intracellular cAMP measured after stimulation of α-MSH, β-MSH, and NDP-α-MSH. EC_{50} and E_{max} values were obtained from concentration–response curves from 0.01 to 100 nM NDP-α-MSH and from 10 pM to 10 μM α-MSH and β-MSH using the computer program GraphPad Prism. *Cell surface expression* denotes the percentage of MC4R cell surface expression as compared to the wild type. *Function* is a summarization of each mutation's effect on function based on cAMP assay and cell surface expression (*NA* not applicable, *LWT* like wild type, *RF* reduced function); those with an asterisk (*) were not functionally tested in this study. *Site Class* denotes the site classification of each codon based on the M3(discrete) PAML sites model; 0 = very strong purifying selection ($\omega_0 = 0.009$), 1 = strong purifying selection ($\omega_1 = 0.094$), 2 = purifying selection ($\omega_2 = 0.483$). *Nd* Not determined

properties that did not result in an enhancement of intracellular cAMP levels (Table 2). Some mutants (C40Y, I170V and S191T) displayed a reduced cell surface expression; however, maximal stimulation and other parameters with all three tested ligands were comparable to the wild-type receptor, and therefore we define these mutants as having wild-type function.

Pan troglodytes re-sequencing

To determine if the unusually low π and unusually high pN/pS values are restricted to humans, we sequenced the *MC4R* coding region in eight central chimpanzees. Four segregating sites were observed (2 synonymous and 2 non-synonymous mutations), from which π was estimated to be 0.00111. This observed π value is not unusual ($P = 0.23$) when compared to 26 non-coding regions which were sequenced in the same chimpanzees (Fischer et al. 2006), nor is it unusual ($P = 0.185$) when compared to simulated data based on a demographic model appropriate for chimpanzees (see "Materials and methods"). Thus, unlike humans, the amount of diversity at the *MC4R* locus is not unusual in central chimpanzees. Further, the estimated pN/pS value for *MC4R* in chimpanzees is 0.36, which is smaller than but not significantly different from that of humans (Table 3).

Primate resequencing

The coding sequence of *MC4R* was determined from a total of 13 primate species, including humans (Fig. 1, Supplemental Table 1). Published data were added for another primate species, the galago (*Otolemur garnettii*, Oga) as well as for the tree shrew (*Tupaia belangeri*, Tbe) a species basal to the primate phylogeny. The conservation of the coding sequence is high across primates. The largest p-distance (0.088; proportion of nucleotide differences) is found between *Papio hamadryas* and *Otolemur garnettii*. Only a few amino acid sequence differences (23 in 334 amino acid sites) are found among primates (Fig. 1), across approximately 60 million years of evolution.

Table 3 Polymorphic non-synonymous versus synonymous mutations in homo and pan

	Human	Chimpanzee
Poly_Syn	2	2
Poly_Nonsyn	11	2
		$P = 0.219$

The number of observed polymorphic synonymous (Poly_Syn) and non-synonymous (Poly_Nonsyn) mutations in humans and chimpanzees. The P value is based on a Fisher's exact test

Phylogenetic analysis by maximum likelihood (PAML)

The *MC4R* coding sequences for 29 species, covering all of vertebrate evolution, were retrieved from available databases (accession numbers can be found in Supplemental Table 1). Nucleotide and protein sequence distance matrixes were computed using the F84 and Kimura nucleotide substitution models and the JTT and PAM protein substitution models, from which neighbor-joining trees were generated. A "species tree" was created according to the accepted taxonomy (Benton and Donoghue 2007; Bininda-Emonds et al. 2007). To determine if any one tree statistically fits the data better than any other, we used the likelihood-based S–H and K–H tests (Goldman et al. 2000; Kishino and Hasegawa 1989; Shimodaira and Hasegawa 1999), as implemented in the PAML package. The species tree had the largest log likelihood, but no tree was statistically better than any other. Given this result we chose to conduct all further analyses with the species tree (Fig. 2).

The goal of these analyses is to investigate various ways of estimating ω (the ratio of the rates of non-synonymous to synonymous substitutions per codon site) across the phylogeny to determine if the *MC4R* locus has been evolving neutrally or rather has been subject to functional constraints, and if the latter is the case, to then identify essential residues which are important to protein function. Each analysis was performed using one of three different codon substitution models: F3X4, F61 and FMutSel (see "Materials and methods" for details). The results are largely concordant between substitution models and thus results will only be presented for the F3X4 model, unless otherwise specified. Further, because alignment gaps are ignored in the analyses, each analysis was conducted with both the full data set of 41 species and with a reduced data set which includes the primate species and the basal tree shrew (Tbe) lineage. The full data set provides more power to detect changes in evolutionary rates, whereas the reduced data set permits the evaluation of evolutionary rates at sites that are present in all primate species but missing in one or more of the other vertebrate species. As the results from both the full dataset and the reduced dataset of only primate sequences were largely concordant, we present and discuss the results based on the full dataset, and present those for the reduced dataset only when they differ.

First, a global ω value was estimated for the entire tree, using the M0 model, which constrains each lineage in the tree to evolve with the same ω value. The global ω value for the *MC4R* phylogeny is 0.043, and is a significantly better fit to the data ($P < 0.001$) than when the ω value is set to 1. This ω estimate suggests strong negative selection acting throughout the *MC4R* phylogeny.

We next allowed the ω value to vary from lineage to lineage by fitting the free-ratio model, which estimates the

```
Hsp   MVNSTH-RGM HTSLHLWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Ptr   MVNSTH-RGM HTSLHLWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNG  [110]
Ppa   MVNSTH-RGM HTSLHLWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Ggo   MVNSTH-RGM HTSLHLWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Ppy   MVNSTH-RGM HTSLHLWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Mfa   MVNSTH-RGM HASLHLWNRS SHRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Mmul  MVNSTH-RGM HASLHLWNRS SHRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Pha   MVNSTH-RGI HASLHLWNRS SHRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Sen   MVNSTH-RGK HASLHLWNRS SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Cja   MVNSTH-RGM HTSLHFWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Soe   MVNSTH-RGM HTSLHFWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Abe   MVNSTH-RGM HTSLPFWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Aca   MVNSTH-RGM HTSLHFWNRS SYRLHSNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Oga   M-NATHQRGM HTSLHFWNRS SLGPHRNASE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]
Tbe   MMNATHHRGM HTSLHFWNRS SYGLHSNTSE SLGKGYSDGG CYEQLFVSPE VFVTLGVISL LENILVIVAI AKNKNLHSPM YFFICSLAVA DMLVSVSNGS ETIVITLLNS  [110]

Hsp   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Ptr   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Ppa   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Ggo   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Ppy   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGVLFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Mfa   TDTDTQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Mmul  TDTDTQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Pha   TDTDTQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Sen   TDTDTQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Cja   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGIT ISCIWAACTV SGILFIIYSE SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Soe   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGIT ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Abe   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGIT ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Aca   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGIT ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]
Oga   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGVLFIIYSD SSAVIICLIT MFLTMLALMA SLYVHMFLMA  [220]
Tbe   TDTDAQSFTV NIDNVIDSVI CSSLLASICS LLSIAVDRYF TIFYALQYHN IMTVKRVGII ISCIWAACTV SGILFIIYSD SSAVIICLIT MFFTMLALMA SLYVHMFLMA  [220]

Hsp   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Ptr   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Ppa   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Ggo   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Ppy   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Mfa   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSVIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Mmul  RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSVIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Pha   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSVIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDV  [330]
Sen   RLHIKRIAVL PGTGAIRQGT NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNFYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC SPLGG-LCEL  [330]
Cja   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSVIDP LIYALRSQEL RKTFKEIICC YPLGS-LCDL  [330]
Soe   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNFYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGS-LCDL  [330]
Abe   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Aca   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSIIDP LIYALRSQEL RKTFKEIICC YPLGGGLCDL  [330]
Oga   RLHIKRIAVL PGTGAIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSVIDP LIYALRSQEL RKTFKEIICC YPLGG-LCDL  [330]
Tbe   RLHIKRIAVL PGTGTIRQGA NMKGAITLTI LIGVFVVCWA PFFLHLIFYI SCPQNPYCVC FMSHFNLYLI LIMCNSVIDP LIYALRSFEL RKTFKEIICR YPLGG-LCDL  [330]

Hsp   SSRY [334]
Ptr   SSRY [334]
Ppa   SSRY [334]
Ggo   SSRY [334]
Ppy   SSRY [334]
Mfa   SSRY [334]
Mmul  SSRY [334]
Pha   SSRY [334]
Sen   SSRY [334]
Cja   SSRY [334]
Soe   SSRY [334]
Abe   SSRY [334]
Aca   SSRY [334]
Oga   SSRY [334]
Tbe   SSRY [334]
```

Fig. 1 Amino acid sequence alignment of all primates used in this study as well as the primate outgroup (Tbe, tree shrew). Variable positions are *shaded in gray*. Species names corresponding to each three letter code can be found in Supplemental Table 1

ω value for each lineage. A likelihood ratio test (LRT, two times the difference in log likelihoods of the two models) indicates that this model provides a significantly better fit to the data, compared to the global M0 model (LRT = 162, $P < 0.001$, $df = 78$), although this is not true for the primate phylogeny alone (LRT = 28.5, $P = 0.39$, $df = 27$). This difference between the full and reduced dataset may reflect a change in the evolutionary rate of MC4R between primates and other vertebrates, a possibility we investigate below. Overall, even as negative selection predominates across the phylogeny there is significant variation in ω values (i.e. levels of constraint) across lineages; the lineage-specific ω values are presented in Fig. 2.

We next investigated variation in ω values across sites, by fitting the M3(discrete) sites model to the *MC4R* phylogeny. This model defines three categories of sites (0, 1 and 2) each with a free and independent estimate of ω. Site class 0 has an ω value (ω_0) of 0.009 with 71% (p_0) of the sites designated to this category; site class 1 has $p_1 = 0.24$ and $\omega_1 = 0.094$; and site class 2 has $p_2 = 0.05$ and $\omega_2 = 0.483$.
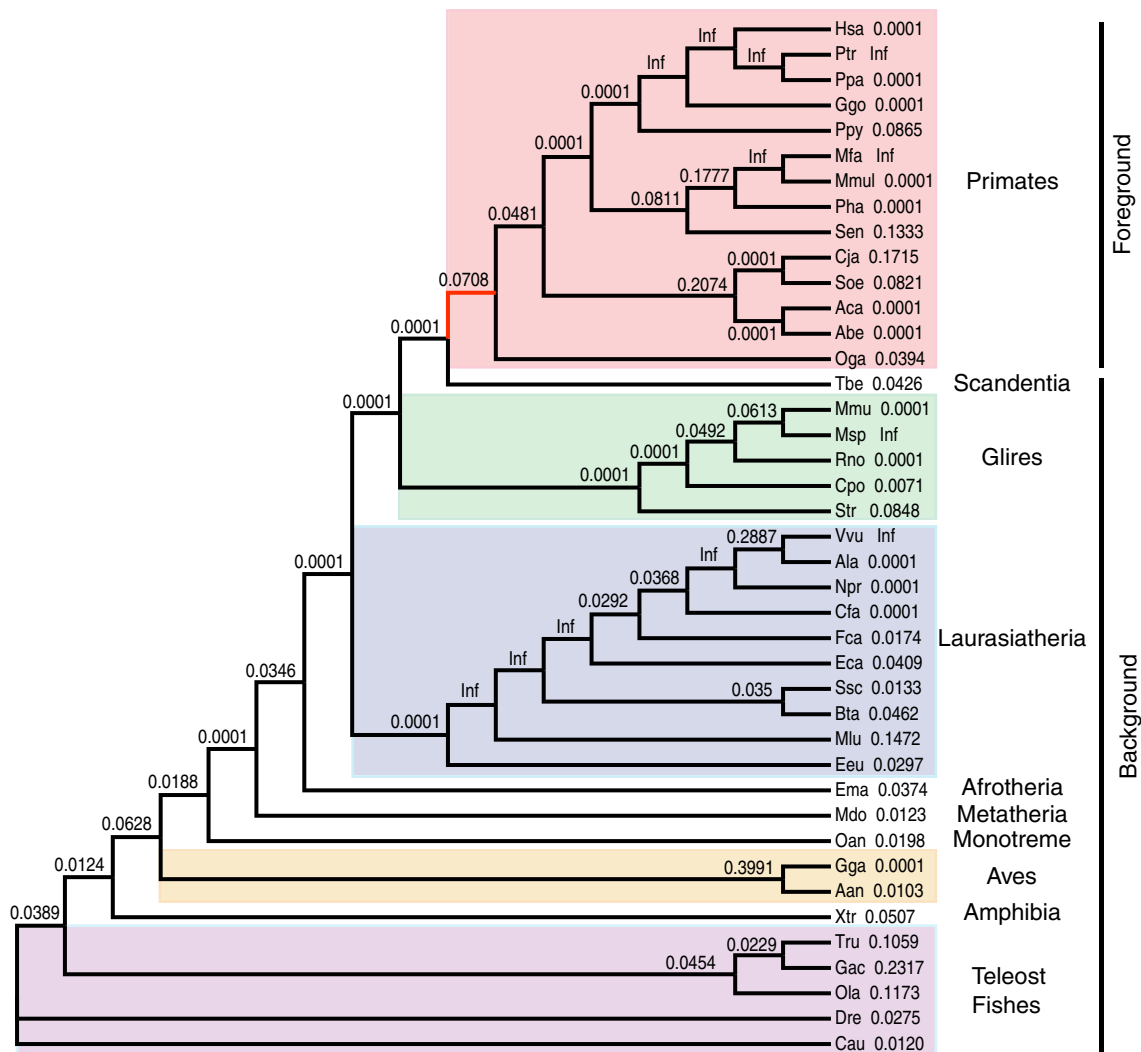
Fig. 2 MC4R species tree. Omega estimates for each lineage, from the free-ratio (M1) model, across the unrooted *MC4R* phylogeny. Estimates equal to infinity (*Inf*) are designated as such when the estimation of dS was equal to 0.0, i.e. no silent substitutions were observed.

Foreground and background labels are for the MC clade model analysis of divergent selection pressures; all branches descended from the branch highlighted in *red* are foreground branches

This model is a better fit to the data than the M0(one-ratio) model (LRT = 536, $P < 0.001$, $df = 4$). Thus, there is variation in the rate of evolution at different sites across the *MC4R* phylogeny, with most sites (71%) experiencing very low rates of non-synonymous substitutions ($\omega = 0.009$). There is also significant evidence of site variation in evolutionary constraint when using the primate phylogeny alone (LRT = 15, $P < 0.001$, $df = 2$). However, here the M3(discrete) model with three categories of sites is not a better fit than the same model with two categories of sites. As such, when analyzing the primate phylogeny alone 93% of sites (p_0) have a ω_0 estimate of 0.026, while the remaining 7% of sites (p_1) have a ω_1 estimate of 0.63.

Moreover, the three site classes (0, 1, and 2, corresponding to sites under strong, moderate, and weak purifying selection) are not evenly distributed over the topology

of the *MC4R* protein (Fig. 3). The occurrence of each site class in the three topological categories (extracellular, transmembrane, and intracellular; Table 4) shows a significant deviation from random expectations ($\chi^2 = 71.22$, $P < 0.001$). The extracellular region of *MC4R* has a deficiency of class 0 sites (very strong purifying selection), and an excess of class 2 sites (those with the largest ω value). This observation, along with the fact that the extracellular N-terminus has many alignment gaps (Fig. 3) and displays the lowest degree of evolutionary conservation, suggests that this section of the protein can tolerate more mutations, relative to other regions of *MC4R*, while still maintaining protein function. The analogous analysis with the primate phylogeny alone shows a similar, although non-significant ($\chi^2 = 4.14$, $P = 0.11$) pattern (Supplemental Fig. 1).

Fig. 3 Bar plot. Site class probability bar plot for each codon across the *MC4R* sequence. *White spaces* are alignment gaps
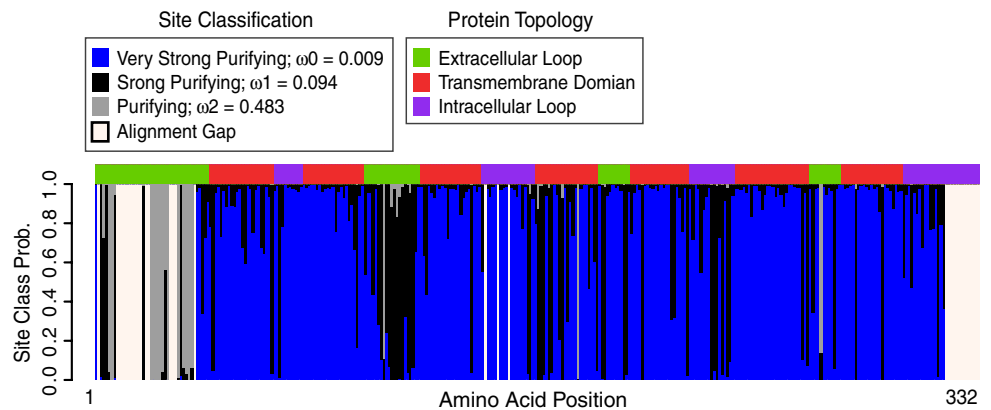

Table 4 Site and protein topology matrices

	Class$_0$	Class$_1$	Class$_2$
Extra	28 (51.68)	28 (14.80)	14 (3.52)
Trans	143 (123.29)	23 (35.31)	1 (8.41)
Intra	49 (45.03)	12 (12.90)	0 (3.07)

The number of class 0, 1 and 2 sites based on the M3 PAML model observed in extracellular (Extra), transmembrane (Trans) and intracellular (Intra) regions of the *MC4R* protein. The first number in each cell is the observed number of sites, the number in parenthesis is the expected number

To determine if these site classifications are informative in identifying functionally critical residues, we compared the association of *MC4R* non-synonymous mutations found in obese cohorts as well as in population based samples, including the CEPH panel, to the site classifications of the M3 analysis. If the level of evolutionary constraint is informative in predicting the deleterious effect of a mutation then we would predict that mutations occurring at class 0 sites (i.e. those under very strong purifying selection) would be more likely to reduce *MC4R* function in cAMP assays. Correspondingly, we would predict that mutations that function like the wild type (LWT) *MC4R* in cAMP assays should occur more frequently at class 1 and 2 sites (i.e. those under less evolutionary constraint). Using our own data (Table 2) and published data we identified 42 codons (Supplemental Table 2) for which a non-synonymous mutation was observed, functional tests were performed, and the site classification could be determined (Farooqi et al. 2000; Gu et al. 1999; Hinney et al. 1999, 2003, 2006; Kobayashi et al. 2002; Larsen et al. 2005; Nijenhuis et al. 2003; Santini et al. 2004; Srinivasan et al. 2004; Stutzmann et al. 2007; Vaisse et al. 2000; Valli-Jaakola et al. 2004; Xiang et al. 2006; Yeo et al. 2003). Of these 42 non-synonymous substitutions, 23 caused a reduction in function (RF), and 22 (96%) of them occur at a class 0 site, under strong evolutionary constraint (Supplemental Table 2). Of the 19 non-synonymous substitutions which

retain LWT function, nine (47%) occurred at a class 0 site. Overall, there is a significant association of class 0 sites with mutations which cause a reduction in function (Fisher's exact test, $P < 0.001$). Hence, non-synonymous mutations that have reduced function are significantly more likely to occur in codons under "very strong" purifying selection, while non-synonymous mutations that retain LWT functionality are over-represented in classes experiencing less evolutionary constraint (site class 1 and 2).

Finally, we investigated the possibility that certain codons in the *MC4R* gene have experienced recurrent positive selection across the *MC4R* phylogeny, possibly providing an adaptive function for particular lineages. We did this by fitting two neutral models of codon evolution (M1a and M7) which allow ω values to range from 0 to 1 across codon sites, and two positive selection models (M2a and M8) which allow a proportion of the sites to have ω values larger than 1 (Nielsen and Yang 1998; Wong et al. 2004; Yang et al. 2005). The M1a and its corresponding M2a model are discrete models, where ω estimates take a discrete value, while the M7 and M8 models estimate ω based on a continuous beta distribution. No selection model was significantly better than the corresponding neutral model, indicating that there is no evidence for recurrent positive selection at particular amino acid residues across the *MC4R* phylogeny.

Divergent selection pressures in primates

The above analyses indicate that overall, the *MC4R* protein has been under strong negative selection throughout vertebrate evolution. Yet, is there any evidence for a change in selection pressure on *MC4R* during primate evolution? To investigate this, a clade model of codon evolution (MC) was fit to the data to determine if there are divergent selection pressures between the primate clade and all other lineages in the *MC4R* phylogeny (Bielawski and Yang 2004). This model has 3 site categories: ω at class 0 sites can range from 0 to 1 (purifying selection); ω at class 1 sites are set to

1 (neutral evolution); and class 2 sites are under divergent selection pressure where ω is freely estimated. Further, class 2 sites have two estimations of ω: one estimate (ω_2) is the ω estimate for the background branches, i.e. all branches not included in the primate clade; the other estimate (ω_3) is the ω estimate for the primate clade. Thus, it is the class 2 sites, which freely estimate ω for the foreground branch (primate lineage) and the background branch (all other lineages) independently, that allow the determination of possible divergent selection pressures acting on these two opposing branches of the MC4R phylogeny (Fig. 2). To construct an LR-test this clade model is compared to the neutral sites model (M1a) with 3 degrees of freedom. The results of this model indicate that 22% of the sites are designated to site class 0 (purifying selection) where ω_0 is equal to 0.134, 3% of the sites belong to site class 1 (neutrally evolving), and 75% of the sites are under divergent selection pressure where ω_2 (background) equals 0.011 and ω_3 (primate clade) equals 0.003. This model is significantly better than the M1a model (LRT = 350, $P < 0.001$), and thus estimates that 75% of the sites across the MC4R locus are evolving in the primate clade under a level of purifying selection that is three times stronger than in the rest of the phylogeny. While MC4R has been subject to strong purifying selection throughout vertebrate evolution, it appears that the level of evolutionary constraint has increased significantly during primate evolution.

Selection on codon usage bias

Recently, a method was developed to address selection on codon usage bias, which would influence variation at silent sites (Yang and Nielsen 2008). This model, FMutSel, explicitly models mutational biases and selection at silent sites in comparison to a null model (FMutSel0) which only models mutational biases. The model not only determines if there is selection at silent sites creating codon usage biases, but also estimates the proportion of all possible mutations that would be advantageous and the mean selection coefficient for advantageous and disadvantageous mutations. With respect to the MC4R phylogeny, the FMutSel model is a significantly better fit to the data than the null model (LRT = 111, $P < 0.001$), suggesting codon usage biases across the MC4R phylogeny. Furthermore, the model estimates that 80% of all observed substitutions at silent sites are advantageous. The mean selection coefficient estimates across the MC4R locus are 1.98 for positively selected sites, and −2.33 for negatively selected sites; these values are approximately five and four times larger, respectively, than the average values estimated for 5,600 genes (Yang and Nielsen 2008). Thus, mutations which occur at silent sites at the MC4R locus are either largely advantageous or largely disadvantageous, and hence would be expected to be quickly fixed or lost via selection.

Discussion

The MC4R locus is highly conserved and exhibits evidence of strong purifying selection across the entire vertebrate phylogeny, as evident by the very low global ω estimate of 0.043. Furthermore, there is no evidence of any one lineage experiencing an acceleration of protein evolution that may be adaptive (Fig. 2). The level of functional constraint appears high across the entire MC4R phylogeny, which is consistent with the MC4R locus, and GPCRs in general, being vital to the organisms carrying them (Rompler et al. 2007; Schoneberg et al. 2007).

Despite the high level of purifying selection across the entire MC4R phylogeny, the amount of functional constraint has apparently increased even more during primate evolution. Approximately 75% of all codons exhibit three times more evolutionary constraint in the primate lineage than all other lineages, which would suggest that the physiological role of MC4R is more constrained in primates than in other vertebrates. This may be a consequence of changes in MC4R regulation and its effect on animal behavior. For example, fish deprived of food increase MC4R expression in the liver (Kobayashi et al. 2008); only mammals express MC4R exclusively in the brain where there is an apparent gene–environment interaction influencing eating behavior (Butler and Cone 2003). It would, however, be necessary to determine if MC4R regulation and function vary between primates and other mammals to fully understand the observation of increased functional constraint in primates.

Against this background of even higher purifying selection on MC4R in primates, observations from human polymorphism data are consistent with a continuing increase in evolutionary constraints on the human lineage. The MC4R locus has a significantly low level of diversity in humans, when compared to other empirical data (Bustamante et al. 2005; SeattleSNPs) and to simulations. This observation appears to be restricted to humans, as MC4R diversity in chimpanzees is not unusually low as compared either to simulations or to empirical data from non-coding regions (Fischer et al. 2006).

Humans do exhibit an unusually large pN/pS ratio, as compared to empirical data, which would suggest relaxation of functional constraints. This observation, at first, seemed at odds with the overall low level of diversity at MC4R. However, the observed level of variation at synonymous and non-synonymous sites is not statistically different from that observed in chimpanzees, although this conclusion may very well be the result of so few variable sites, and hence a corresponding lack of power in the statistical

test. More importantly, all observed human mutations are rare (mean = 0.17%) suggesting that most of these mutations are transient deleterious mutations on their way to being removed from the population, consistent with strong purifying selection on *MC4R* and the functional data of the observed non-synonymous mutations. This, in turn, predicts that the polymorphisms observed in the human data are all new mutations, and hence should reflect the presumably random mutational spectrum at non-synonymous and synonymous positions, driving the unusually large pN/pS. To test this prediction, we performed a very simple simulation: we took the full length of the *MC4R* coding sequence (996 bp, with an estimated 732 non-synonymous positions and 264 synonymous positions), randomly mutated 13 positions, computed the pN/pS value, and then repeated this process 2,000 times. The probability, based on this empirical distribution, of observing a pN/pS value equal to or greater than the value of 1.98 is 0.271. Therefore, the observed pN/pS value for *MC4R* is not significantly different from expectations based on a purely random distribution of mutations.

Although the observed pN/pS value for *MC4R* is not different from purely random expectations, the occurrence of only two synonymous substitutions, and at such low frequency, remains puzzling. Certainly, continuous negative selection at linked non-synonymous sites will maintain a low level of diversity at synonymous sites. However, negative selection alone may not be sufficient to explain the observed data; another factor may be codon usage bias. As demonstrated by the FMutSel analyses, the majority of mutations at silent sites are expected to be either highly advantageous or highly disadvantageous, with average estimated selection coefficients for *MC4R* much larger than those for typical genes. Thus, selection will either quickly remove or fix a silent mutation, which may help explain the lack of silent polymorphisms.

Finally, non-synonymous mutations which have been identified as loss of function mutations tend to occur preferentially at class 0 sites (under very strong purifying selection) of the M3(discrete) model (Supplemental Table 2). Moreover, non-synonymous mutations that do not have an effect on *MC4R* function in the in vitro assay tend to occur preferentially at sites that are under less stringent constraints (i.e., weaker purifying selection). This correlation may be useful in assessing the potential functional consequences of newly identified mutations, and should assist further studies of the epidemiology of the *MC4R* locus.

In conclusion, *MC4R* has experienced strong negative selection throughout vertebrate evolution and an apparent increase in selective constraint along the primate phylogeny. Moreover, humans have a significant paucity of diversity at *MC4R*, which is not observed in chimpanzees. This suggests that the magnitude of negative selection has

continued to increase along the human lineage. Further, many of the observed non-synonymous mutations reduce *MC4R* protein function, suggesting a detriment to an individual's fitness. Given these combined observations we conclude that mutations at *MC4R*, across vertebrates and especially in primates and humans, are overwhelmingly deleterious to protein function, resulting in prominent sequence conservation and significantly low levels of diversity in humans.

References

Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life. Mol Biol Evol 24:26–53

Biebermann H, Krude H, Elsner A, Chubanov V, Gudermann T, Gruters A (2003) Autosomal-dominant mode of inheritance of a melanocortin-4 receptor mutation in a patient with severe early-onset obesity is due to a dominant-negative effect caused by receptor dimerization. Diabetes 52:2984–2988

Bielawski JP, Yang Z (2004) A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. J Mol Evol 59:121–132

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. Nature 446:507–512

Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller KE, Adams MD, Schmidt S, Sninsky JJ, Sunyaev SR, White TJ, Nielsen R, Clark AG, Bustamante CD (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4:e1000083

Branson R, Potoczna N, Kral JG, Lentes KU, Hoehe MR, Horber FF (2003) Binge eating as a major phenotype of melanocortin 4 receptor gene mutations. N Engl J Med 348:1096–1103

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Todd Hubisz M, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG (2005) Natural selection on protein-coding genes in the human genome. Nature 437:1153

Butler AA (2006) The melanocortin system and energy balance. Peptides 27:281–290

Butler AA, Cone RD (2003) Knockout studies defining different roles for melanocortin receptors in energy homeostasis. Ann N Y Acad Sci 994:240–245

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL (2002) A human genome diversity cell line panel. Science 296:261–262

Caswell JL, Mallick S, Richter DJ, Neubauer J, Schirmer C, Gnerre S, Reich D (2008) Analysis of chimpanzee history based on genome sequence alignments. PLoS Genet 4:e1000057

Charlesworth B (1994) The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet Res 63:213–227

Dempfle A, Hinney A, Heinzel-Gutenbrunner M, Raab M, Geller F, Gudermann T, Schafer H, Hebebrand J (2004) Large quantitative effect of melanocortin-4 receptor gene mutations on body mass index. J Med Genet 41:795–800

Dubern B, Clement K, Pelloux V, Froguel P, Girardet JP, Guy-Grand B, Tounian P (2001) Mutational analysis of melanocortin-4 receptor, agouti-related protein, and alpha-melanocyte-stimulating hormone genes in severely obese children. J Pediatr 139:204–219

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113

Eyre-Walker A (2006) The genomic rate of adaptive evolution. Trends Ecol Evol 21:569–575

Farooqi IS, Yeo GS, Keogh JM, Aminian S, Jebb SA, Butler G, Cheetham T, O'Rahilly S (2000) Dominant and recessive inheritance of morbid obesity associated with melanocortin 4 receptor deficiency. J Clin Invest 106:271–279

Farooqi IS, Keogh JM, Yeo GS, Lank EJ, Cheetham T, O'Rahilly S (2003) Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. N Engl J Med 348:1085–1095

Fay JC, Wyckoff GJ, Wu CI (2001) Positive and negative selection on the human genome. Genetics 158:1227–1234

Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S (2006) Demographic history and genetic differentiation in apes. Curr Biol 16:1133–1148

Fischer A, Wiebe V, Paabo S, Przeworski M (2004) Evidence for a complex demographic history of chimpanzees. Mol Biol Evol 21:799–808

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Syst Biol 49:652–670

Gu W, Tu Z, Kleyn PW, Kissebah A, Duprat L, Lee J, Chin W, Maruti S, Deng N, Fisher SL, Franco LS, Burn P, Yagaloff KA, Nathan J, Heymsfield S, Albu J, Pi-Sunyer FX, Allison DB (1999) Identification and functional analysis of novel human melanocortin-4 receptor variants. Diabetes 48:635–639

Heid IM, Vollmert C, Kronenberg F, Huth C, Ankerst DP, Luchner A, Hinney A, Bronner G, Wichmann HE, Illig T, Doring A, Hebebrand J (2008) Association of the MC4R V103I polymorphism with the metabolic syndrome: the KORA Study. Obesity (Silver Spring) 16:369–376

Hinney A, Schmidt A, Nottebom K, Heibult O, Becker I, Ziegler A, Gerber G, Sina M, Gorg T, Mayer H, Siegfried W, Fichter M, Remschmidt H, Hebebrand J (1999) Several mutations in the melanocortin-4 receptor gene including a nonsense and a frameshift mutation associated with dominantly inherited obesity in humans. J Clin Endocrinol Metab 84:1483–1486

Hinney A, Hohmann S, Geller F, Vogel C, Hess C, Wermter AK, Brokamp B, Goldschmidt H, Siegfried W, Remschmidt H, Schafer H, Gudermann T, Hebebrand J (2003) Melanocortin-4 receptor gene: case-control study and transmission disequilibrium test confirm that functionally relevant mutations are compatible with a major gene effect for extreme obesity. J Clin Endocrinol Metab 88:4258–4267

Hinney A, Bettecken T, Tarnow P, Brumm H, Reichwald K, Lichtner P, Scherag A, Nguyen TT, Schlumberger P, Rief W, Vollmert C, Illig T, Wichmann HE, Schafer H, Platzer M, Biebermann H, Meitinger T, Hebebrand J (2006) Prevalence, spectrum, and functional characterization of melanocortin-4 receptor gene mutations in a representative population-based sample and obese adults from Germany. J Clin Endocrinol Metab 91:1761–1769

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–348

Huszar D, Lynch CA, Fairchild-Huntress V, Dunmore JH, Fang Q, Berkemeier LR, Gu W, Kesterson RA, Boston BA, Cone RD, Smith FJ, Campfield LA, Burn P, Lee F (1997) Targeted disruption of the melanocortin-4 receptor results in obesity in mice. Cell 88:131–141

Kayser M, Brauer S, Stoneking M (2003) A genome scan to detect candidate regions influenced by local natural selection in human populations. Mol Biol Evol 20:893–900

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170–179

Kobayashi H, Ogawa Y, Shintani M, Ebihara K, Shimodahira M, Iwakura T, Hino M, Ishihara T, Ikekubo K, Kurahachi H, Nakao K (2002) A Novel homozygous missense mutation of melanocortin-4 receptor (MC4R) in a Japanese woman with severe obesity. Diabetes 51:243–246

Kobayashi Y, Tsuchiya K, Yamanome T, Schioth HB, Kawauchi H, Takahashi A (2008) Food deprivation increases the expression of melanocortin-4 receptor in the liver of barfin flounder, *Verasper moseri*. Gen Comp Endocrinol 155:280–287

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

Larsen LH, Echwald SM, Sorensen TI, Andersen T, Wulff BS, Pedersen O (2005) Prevalence of mutations and functional analyses of melanocortin 4 receptor variants identified among 750 men with juvenile-onset obesity. J Clin Endocrinol Metab 90:219–224

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104

Lubrano-Berthelier C, Cavazos M, Le Stunff C, Haas K, Shapiro A, Zhang S, Bougneres P, Vaisse C (2003) The human MC4R promoter: characterization and role in obesity. Diabetes 52:2996–3000

Miraglia Del Giudice E, Cirillo G, Nigro V, Santoro N, D'Urso L, Raimondo P, Cozzolino D, Scafato D, Perrone L (2002) Low frequency of melanocortin-4 receptor (MC4R) mutations in a Mediterranean population with early-onset obesity. Int J Obes Relat Metab Disord 26:647–651

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Nijenhuis WA, Garner KM, van Rozen RJ, Adan RA (2003) Poor cell surface expression of human melanocortin-4 receptor mutations associated with obesity. J Biol Chem 278:22939–22945

Ptak SE, Voelpel K, Przeworski M (2004) Insights into recombination from patterns of linkage disequilibrium in humans. Genetics 167:387–397

Rompler H, Staubert C, Thor D, Schulz A, Hofreiter M, Schoneberg T (2007) G protein-coupled time travel: evolutionary aspects of GPCR research. Mol Interv 7:17–25

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW (2005) Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet 1:e70

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2507

Salomon Y, Londos C, Rodbell M (1974) A highly sensitive adenylate cyclase assay. Anal Biochem 58:541–548

Santini F, Maffei M, Ceccarini G, Pelosini C, Scartabelli G, Rosellini V, Chiellini C, Marsili A, Lisi S, Tonacchera M, Agretti P, Chiovato L, Mammoli C, Vitti P, Pinchera A (2004) Genetic screening for melanocortin-4 receptor mutations in a cohort of Italian obese patients: description and functional characterization of a novel mutation. J Clin Endocrinol Metab 89:904–908

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D (2005) Calibrating a coalescent simulation of human genome sequence variation. Genome Res 15:1576–1583

Schoneberg T, Hofreiter M, Schulz A, Rompler H (2007) Learning from the past: evolution of GPCR functions. Trends Pharmacol Sci 28:117–121

SeattleSNPs NHLBI Program for Genomic Applications. SeattleSNPs, Seattle, WA, vol 5-2008. http://pga.gs.washington.edu

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116

Srinivasan S, Lubrano-Berthelier C, Govaerts C, Picard F, Santiago P, Conklin BR, Vaisse C (2004) Constitutive activity of the melanocortin-4 receptor is maintained by its N-terminal domain and plays a role in energy homeostasis in humans. J Clin Invest 114:1158–1164

Stutzmann F, Vatin V, Cauchi S, Morandi A, Jouret B, Landt O, Tounian P, Levy-Marchal C, Buzzetti R, Pinelli L, Balkau B, Horber F, Bougneres P, Froguel P, Meyre D (2007) Non-synonymous polymorphisms in melanocortin-4 receptor protect against obesity: the two facets of a Janus obesity gene. Hum Mol Genet 16:1837–1844

Vaisse C, Clement K, Durand E, Hercberg S, Guy-Grand B, Froguel P (2000) Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity. J Clin Invest 106:253–262

Valli-Jaakola K, Lipsanen-Nyman M, Oksanen L, Hollenberg AN, Kontula K, Bjorbaek C, Schalin-Jantti C (2004) Identification and characterization of melanocortin-4 receptor gene mutations in morbidly obese Finnish children and adults. J Clin Endocrinol Metab 89:940–945

Wong WS, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–1051

Xiang Z, Litherland SA, Sorensen NB, Proneth B, Wood MS, Shaw AM, Millard WJ, Haskell-Luevano C (2006) Pharmacological characterization of 40 human melanocortin-4 receptor polymorphisms with the endogenous proopiomelanocortin-derived agonists and the agouti-related protein (AGRP) antagonist. Biochemistry 45:7277–7288

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

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Yang Z, Nielsen R (2008) Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol Biol Evol 25(3):568–579

Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118

Yeo GS, Lank EJ, Farooqi IS, Keogh J, Challis BG, O'Rahilly S (2003) Mutations in the human melanocortin-4 receptor gene associated with severe familial obesity disrupts receptor function through multiple molecular mechanisms. Hum Mol Genet 12:561–574

Yu B, Sawyer NA, Chiu C, Oefner PJ, Underhill PA (2006) DNA mutation detection using denaturing high-performance liquid chromatography (DHPLC). Curr Protoc Hum Genet Chapter 7: Unit7 10