GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Molecular evolution of the melanocortin 1-receptor pigmentation gene in rodents

G.L. Gonçalves, V.R. Paixão-Côrtes and T.R.O. Freitas

Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

Corresponding author: G.L. Gonçalves
E-mail: lopes.goncalves@ufrgs.br

Genet. Mol. Res. 12 (3): 3230-3245 (2013)
Received June 27, 2012
Accepted November 22, 2012
Published February 28, 2013
DOI http://dx.doi.org/10.4238/2013.February.28.24

ABSTRACT. Adaptive variation in the melanocortin 1-receptor gene (*MC1R*), a key locus in melanogenesis, has been identified in some species of rodents. However, in others, *MC1R* has no causative role in pigmentation phenotypes despite their coat color variation. In this study, we characterized the rates and patterns of *MC1R* nucleotide and amino acid sequence evolution and, particularly, selective pressures in the separated domains of the protein using a comparative analysis of 43 species representing three major lineages of rodents with variable coat colors. We found high amino acid variation (44% of sites) throughout the protein. Most substitutions were observed in extracellular and transmembrane domains; the intracellular segment was conserved across species. Pairwise non-synonymous substitutions did not vary significantly in different domains among the rodent lineages - i.e., variation was not associated with phylogenetic distance. Phylogeny-based likelihood analysis suggested that purifying selection has mostly shaped the evolutionary course of *MC1R*. However, a high proportion of sites (27%) were under relaxation of functional constraints (ω = 0.38), and four sites (3, 14, 26, and 251) clearly evolved under positive selection ($\omega \cong 2.9$). Thus, our data indicate a high proportion of sites

evolving under relaxed evolutionary constraints, which might indicate the evolvability of the system in the generation of adaptive changes in specific taxa in rodent lineages.

Key words: *MC1R*; Molecular evolution; Positive selection; Rodentia

INTRODUCTION

Coat color variation in small mammals is evidence of phenotypic adaptation in response to selection in different environments. Several examples of intraspecific color variability have been identified in rodents, including the deer field mouse (*Peromyscus maniculatus*), old-field mouse (*Peromyscus polionotus*), rock pocket mouse (*Chaetodipus intermedius*), mole rat (*Spalax ehrenbergi*), Botta's pocket gopher (*Thomomys bottae*), Rio Negro tuco-tuco (*Ctenomys rionegrensis*), and collared tuco-tuco (*Ctenomys torquatus*) (Gonçalves and Freitas, 2009).

Genes underlying such variation have been investigated in these species (Nachman et al., 2003; Wlasiuk and Nachman, 2007; Gonçalves et al., 2012), as well as in others with conspicuous melanic phenotypes in addition to the most common phenotype [e.g., the gray squirrel (*Sciurus carolinensis*)] (McRobie et al., 2009). The main candidate locus targeted was melanocortin 1-receptor (*MC1R*), a member of the G protein-coupled superfamily that acts as a pigment switch in the production of melanin. When activated by α-melanocyte-stimulating hormone, it signals the production of eumelanin (black/brown pigment) via cyclic adenosine monophosphate; in the absence or inhibition of stimulation, pheomelanin (red/yellow pigment) is synthesized (Jackson et al., 1994). In mice, dominant mutations that disable the binding of α-melanocyte-stimulating hormone lead to constitutive activity (constant signaling of eumelanin synthesis) and predominantly black coat color (Jackson, 1997). Both types of phenotypic change have been linked to missense mutations in the *MC1R* of domestic and wild mammals (Hoekstra, 2006).

Previous studies in natural populations of rodents have suggested that *MC1R* may be a target of positive selection. According to Nachman et al. (2003) and Hoekstra et al. (2006), it underlies adaptive melanism in rock pocket mice and adaptive light coat color in beach mice, respectively. Moreover, McRobie et al. (2009) have observed that the melanic form of gray squirrel is the result of a 15-bp deletion that constitutively activates the receptor. By contrast, species with pigmentation phenotype variation, such as tuco-tucos and Botta's pocket gopher, display no changes in *MC1R* sequences that can be directly linked to the observed differences (Wlasiuk and Nachman, 2007; Gonçalves et al., 2012).

Some studies in rodents have provided evidence of positive selection, particularly in genes involved in immune response (e.g., Toll-like and β-2-microglobulin) and the reproductive system (e.g., Zp-3, Sry, and Tcp-1) (Jansa et al., 2003; Tschirren et al., 2012). In addition, loci related to pigmentation pathways, such as *MC1R*, are also potential candidates to find signatures of positive selection based on rates of synonymous substitution (silent; d_S) and non-synonymous substitution (amino acid replacement; d_N). These estimates are critical to understand the dynamics of molecular sequence evolution (Kimura, 1983). Because synonymous mutations are considered invisible to natural selection, whereas non-synonymous mutations may be under strong selective pressure, comparison of the rates of fixation provides a powerful tool for understanding the mechanisms of DNA sequence evolution. For example, variable d_N/d_S rate ratios among lineages may indicate adaptive evolution or relaxed selective constraints along certain lineages

(Crandall and Hillis, 1997). Similarly, models of variable d_N/d_S rate ratios among sites may provide important insights into functional constraints at various amino acid sites and may be used to detect sites under positive selection (Nielsen and Yang, 1998).

Thus, in this study we investigated patterns of molecular evolution in *MC1R* sequences from rodents given that this gene is often implicated in adaptation in some species but not in others. Consequently, we asked whether it represents a lineage-specific pattern or a particular trait in some taxa. We performed a comparative analysis of *MC1R* evolution rates and patterns in the entire gene and in separate domains (extracellular, transmembrane, and intracellular) to characterize selective pressures, searching in particular for positively selected sites or relaxation in functional constraints in rodents.

MATERIAL AND METHODS

Species surveyed and *MC1R* variation

Nucleotide and amino acid sequences of the coding region of *MC1R* from 43 species of rodents representing 6 families (Muridae, Cricetidae, Heteromyidae, Geomyidae, Sciuridae, and Ctenomyidae) were obtained from GenBank. These families included the three major currently recognized lineages: 1) "mouse-related clade" (Muridae + Cricetidae + Heteromyidae + Geomyidae), 2) "squirrel-related clade" (Sciuridae), and 3) Ctenohystrica (Ctenomyidae) (Table 1 and Figure 1) (Blanga-Kanfi et al., 2009). In addition, *MC1R* sequences of primates [*Lemur* (AY205143), *Pan* (AY205086), and *Gorilla* (AY205088)], and lagomorphs [*Lepus* (HQ005375) and *Oryctolagus* (AM180880)] were incorporated to compare rates and patterns of substitution in relation to sister lineages.

Sequences of the *MC1R* coding region were aligned using Codon Code Aligner (Codon Code Corp.). All insertion/deletions (indels) and substitutions were examined by eye. Phylogenetic reconstruction based on nucleotide sequences of 43 species of rodents rooted with a lagomorph was implemented in MRBAYES 3.2 (Huelsenbeck et al., 2001) using all codon positions. The topology and branch lengths were estimated using a Bayesian approach with a TN93+G model of nucleotide evolution (defined by the Akaike information criteria implemented in MRMODELTEST 2) (Nylander et al., 2004), with parameters estimated from the data set. The numbers of d_S and d_N among lineages were estimated for 15 segments of *MC1R* [4 extracellular domains (EDs), 7 transmembrane (TM) domains, and 4 intracellular domains (IDs); Figure 2A] and the entire gene using the Nei and Gojobori (1986) method in the CODEML program of the PAML 4.4 package (Yang, 2007). Additionally, divergence at the amino acid level was estimated among clades using p-distance with the Molecular Evolutionary Genetics Analysis 5 software (Tamura et al., 2011).

Three classes of comparison were used to estimate distance and standard deviation with 1000 bootstrap replications: 1) mouse-related clade *vs* squirrel-related clade, 2) mouse-related clade *vs* Ctenohystrica, and 3) squirrel-related clade *vs* Ctenohystrica, representing different levels of phylogenetic relationships in rodents. In addition, average pairwise d_N rates were characterized in the 3 lineages of rodents for all 15 domains of *MC1R* and the entire gene. Comparisons of the 6 families were performed to characterize differences among lineages during the evolutionary history of rodents. Additionally, pairwise comparisons of d_N were made between rodents and lagomorphs and primates and lagomorphs to investigate patterns of acceleration in the substitution rate in relation to rodent sister clades.

Table 1. Species of rodents used in this study.

Family	Species	Common name	Lineage	GenBank accession No.
Cricetidae				
	Eothenomys melanogaster	Père David's Vole	Mouse-related	GU001573
	Peromyscus maniculatus	Deer mouse	Mouse-related	GQ337977
	Peromyscus polionotus	Beach mouse	Mouse-related	FJ389440
Ctenomyidae				
	Ctenomys australis	Sand-dune tuco-tuco	Ctenohystrica	JF910108
	Ctenomys boliviensis	Bolivian tuco-tuco	Ctenohystrica	JF910109
	Ctenomys azarae	Azara's tuco-tuco	Ctenohystrica	JF910110
	Ctenomys dorbygni	Dorbigny's tuco-tuco	Ctenohystrica	JF910111
	Ctenomys flamarioni	Sand-dune tuco-tuco	Ctenohystrica	JF910112
	Ctenomys haigi	Haig's tuco-tuco	Ctenohystrica	JF910113
	Ctenomys lami	-	Ctenohystrica	JF910114
	Ctenomys leucodon	White-toothed tuco-tuco	Ctenohystrica	JF910115
	Ctenomys maulinus	Maule tuco-tuco	Ctenohystrica	JF910116
	Ctenomys mendocinus	Mendoza tuco-tuco	Ctenohystrica	JF910117
	Ctenomys minutus	-	Ctenohystrica	JF910118
	Ctenomys nattereri	Natterer's tuco-tuco	Ctenohystrica	JF910119
	Ctenomys pearsoni	Pearson's tuco-tuco	Ctenohystrica	JF910120
	Ctenomys perrensi	Goya tuco-tuco	Ctenohystrica	JF910121
	Ctenomys porteousi	Porteous' tuco-tuco	Ctenohystrica	JF910122
	Ctenomys rionegrensis	Rio Negro tuco-tuco	Ctenohystrica	JF910123
	Ctenomys roigi	Roig's tuco-tuco	Ctenohystrica	JF910124
	Ctenomys sociabilis	Social tuco-tuco	Ctenohystrica	JF910125
	Ctenomys steinbachi	Steinbach's tuco-tuco	Ctenohystrica	JF910126
	Ctenomys talarum	Talas's tuco-tuco	Ctenohystrica	JF910127
	Ctenomys torquatus	Collared tuco-tuco	Ctenohystrica	JF910128
Geomyidae				
	Thomomys bottae	Pocket gopher	Mouse-related	EF488834
Heteromyidae				
	Chaetodipus baileyi	Bailey's pocket mouse	Mouse-related	AY258938
	Chaetodipus intermedius	Rock pocket mouse	Mouse-related	AY247634
	Chaetodipus penicillatus	Desert pocket mouse	Mouse-related	AY258934
Muridae				
	Meriones unguiculatus	Mongolian gerbil	Mouse-related	AY800269
	Mus booduga	Little Indian field mouse	Mouse-related	AB306316
	Mus fragilicauda	Sheath-tailed Mouse	Mouse-related	AB306317
	Mus terricolor	Earth-colored mouse	Mouse-related	AB306318
	Mus caroli	Ryukyu mouse	Mouse-related	AB306319
	Mus cervicolor	Fawn-colored mouse	Mouse-related	AB306320
	Mus cookii	Cook's mouse	Mouse-related	AB306321
	Mus musculus	House mouse	Mouse-related	AB306322
	Mus mattheyi	Matthey's mouse	Mouse-related	AB306323
	Mus pahari	Gairdner's shrew-mouse	Mouse-related	AB306324
	Mus platythrix	Flat-haired mouse	Mouse-related	AB306325
	Rattus norvegicus	Brown rat	Mouse-related	AB306978
	Rattus rattus	Black rat	Mouse-related	AB576624
	Rattus tanezumi	Tanezumi rat	Mouse-related	AB576604
Sciuridae				
	Sciurus carolinensis	Gray squirrel	Squirrel-related	EU604831

Tests of selection

Patterns of selection and rates of evolutionary change in *MC1R* were evaluated with Yang and Bielawski (2000) tests. The established structure and inferred function of the *MC1R* protein (Chhajlani et al., 1996) allowed us to make *a priori* predictions about the domains expected to evolve under the relaxation of selective constraints and therefore to discuss the results from a functional perspective. We used the phylogeny-based maximum likelihood analysis of ω (d_N/d_S ratio) as implemented in the CODEML program of the PAML package to test

statistically for positive selection or relaxation of functional constraints acting on *MC1R* codons. Log-likelihood values were generated for models in which ω was allowed to vary among sites within the interval 0-1 [neutral models (M0 and M1a) and a discrete model (M3)] and for models that allowed ω > 1 for some sites [selection models (M2a and M8)] following Yang et al. (2000). First, we tested whether ω differs among sites by comparing model M0, which assumes a constant ω across all sites, to model M3, which allows ω to vary among sites. To test formally for the presence of sites evolving under positive selection, we then compared a nearly neutral model of ω variation (M1a) to a model that allows for positive selection (M2a). Subsequently, we also compared a neutral model, M7, which estimates ω with a beta distribution over the interval 0-1, to a selection model, M8, which additionally allows for positively selected sites (ω > 1) (Yang et al., 2000). We compared the models using the likelihood ratio test (LRT) in the HYPHY 1 program (Pond et al., 2005).

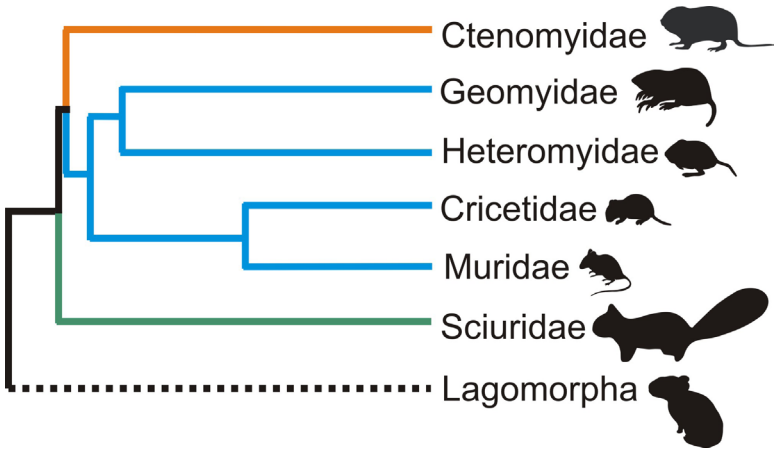

Figure 1. Phylogenetic tree of rodents reconstructed by Bangla-Kanfi et al. (2009) including families from the three major lineages: Ctenohystrica (orange), mouse-related clade (blue), and squirrel-related clade (green).

Twice the log-likelihood difference (2ΔLnL) between models follows a chi-square distribution, with the degrees of freedom equal to the difference in the number of parameters between the models. All models were run multiple times with different starting values for ω to ensure correct estimation of the model parameters. The unrooted tree input file for these analyses was generated using the maximum likelihood method implemented in PAUP (Swofford, 2003) with the data set of 43 rodent species. We used empirical Bayes approaches implemented in CODEML to infer which sites of the *MC1R* sequence may have evolved under positive selection. Two approaches were used to determine sites under selection: the naive-empirical Bayes and the Bayes-empirical Bayes methods. Positive selection was inferred if the posterior probability of ω > 1 for a site was higher than 0.95. Codon usage bias is well known to affect estimation of d_S and d_N rates. Thus, we applied model F61, which uses empirical estimates of individual codon frequencies.

Finally, we tested the heterogeneity of evolutionary rates among lineages by applying the clade models using CODEML in PAML. Branches on the phylogeny were divided *a priori*

into two clades (branch type 0 as the squirrel-related clade, and branch type 1 as the mouse-related clade + Ctenohystrica), and an LRT was used to evaluate divergences in selective pressures between them indicated by ω ratios. We applied clade model type C, which assumes three site classes, and compared it with M1a using an LRT with two degrees of freedom. This model is used primarily to detect positive selection, but our goal was also to evaluate acceleration during the evolutionary history through direct inferences based on d_N/d_S ratio differences.

RESULTS

MC1R: evolutionary patterns in rodents

Amino acid variation across rodents was abundant throughout the *MC1R* gene (Figure 2B and Table 2). However, in general, several completely conserved segments were observed, but usually in short continuous residues (i.e., <10 amino acids). These segments included 178 (56%) sites that were identical across rodents and mostly concentrated in ID1 and ID2 (see Figure 2B and Table 2). Most of the variation (127 codons) was observed in ED1 and ED2 and in ID3 and TM4 (see Figure 2A and Table 2). ED1 was particularly variable, including multiple amino acid replacements and 5 indel events: a 3-amino acid deletion in *Mus* spp and squirrel, 2- and 1-amino acid insertions in tuco-tucos, and a 7-amino acid deletion in pocket gopher (see Figure 2B). The polarity of these indels was inferred based on the rodent phylogeny proposed by Blanga-Kanfi et al. (2009) (see Figure 1). When the 15 domains of the protein were divided into three categories (ED, ID, and TM), measures of variation among rodents were higher in the ED than in the TM or ID, particularly in ED1 and ED3 (see Table 2). Heterogeneity in divergence estimates was observed within the ID and TM categories, with the lowest values in the three classes of comparisons obtained for ID1, ID4, and TM6. Divergences estimated with p-distance were saturated at the various levels of phylogenetic relationships compared (squirrel-related clade *vs* mouse-related clade; squirrel-related clade *vs* Ctenohystrica; mouse-related clade *vs* Ctenohystrica), except TM3 and TM5 (see Table 2).

Variation in the rates and patterns of nucleotide substitution was observed among the three lineages of rodents. The phylogeny reconstructed based on *MC1R* sequences revealed accelerated substitution rates in one lineage of the mouse-related clade (Cricetidae + Muridae) that was not recovered as monophyletic, as verified by the longer branch length (Figure 3). Additionally, a species-specific variation was also observed in Ctenohystrica, indicated by the outer position of *Ctenomys leucodon* (see Figure 3).

MC1R rates of substitution: acceleration in rodents

Pairwise d_N versus d_S calculated among rodent families for the whole gene showed an accelerated rate in all comparisons, but a phylogenetic pattern was not evident - i.e., all lineages showed a similar d_N/d_S ratio (Figure 4). We observed differences in d_N/d_S within rodents for specific domains, in which Ctenohystrica seemed to display an increased rate compared to that of the mouse- and squirrel-related clades, as observed in the ED1 and ID2 segments (Figure 5A). Significant acceleration in the mouse lineage was observed in the ED4, ID1, TM1, and TM4. The squirrel-related clade showed acceleration only in the ID4.

Overall, pairwise d_N differences varied significantly among comparisons of rodents

and primates, particularly in five domains: ED1, ED2, ID3, ID4, and TM2 (Figure 5B). In some domains, the pattern was reversed, and acceleration was observed in primates instead of rodents, mainly in ID4 and TM3 and partially in TM1 (see Figure 5B).

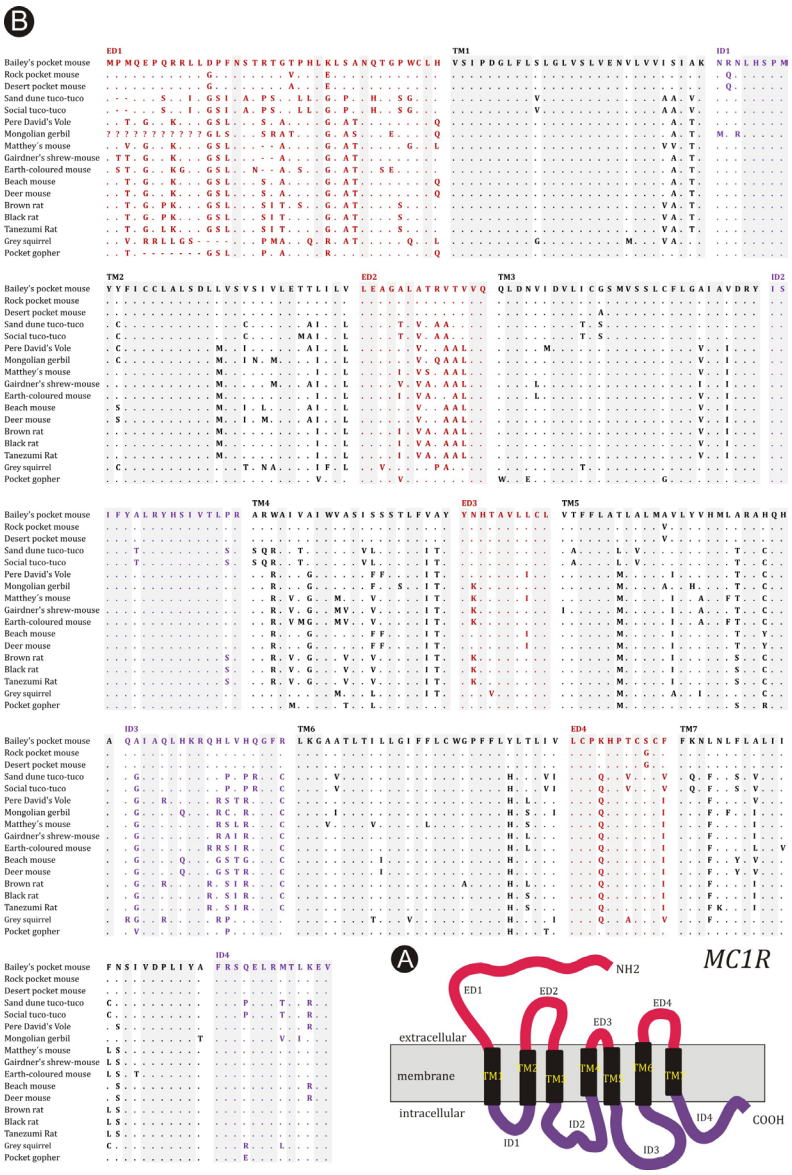

Figure 2. MC1R protein structure and alignment across rodents. **A.** Bi-dimensional protein depiction, highlighting positions of the domains colored as red, black, and blue in the alignment data set. **B.** Amino acid sequence in 17 representative species of all families surveyed. Dots indicate identity to top sequence. The initial position of each of the 15 *MC1R* domains is indicated above the sequences: extracellular (ED) in red, transmembrane (TM) in black, and intracellular (ID) in blue. Domain boundaries are based on Robbins et al. (1993). Amino acid residues completely conserved across the surveyed species are shaded.

Table 2. Amino acid variability in the different domains of the MC1R protein.

Domain	No. of sites	Variable sites	Conserved sites	Mean p-distance between groups ± SE		
				Mouse-related clade *vs* squirrel-related clade	Mouse-related clade *vs* Ctenohystrica	Squirrel-related clade *vs* Ctenohystrica
ED1	41	28 (69%)	13 (31%)	0.25 ± 0.03	0.30 ± 0.03	0.31 ± 0.04
ED2	19	12 (64%)	7 (36%)	0.27 ± 0.05	0.23 ± 0.05	0.23 ± 0.05
ED3	4	2 (50%)	2 (50%)	0.11 ± 0.04	0.21 ± 0.10	0.17 ± 0.04
ED4	14	8 (57%)	6 (43%)	0.12 ± 0.05	0.21 ± 0.07	0.14 ± 0.06
ID1	9	2 (22%)	7 (78%)	0.12 ± 0.05	0.12 ± 0.05	0.07 ± 0.04
ID2	22	6 (28%)	16 (72%)	0.14 ± 0.04	0.18 ± 0.04	0.15 ± 0.04
ID3	27	14 (52%)	13 (48%)	0.23 ± 0.04	0.19 ± 0.04	0.20 ± 0.05
ID4	15	6 (44%)	9 (56%)	0.05 ± 0.02	0.21 ± 0.06	0.22 ± 0.07
TM1	25	6 (24%)	19 (76%)	0.16 ± 0.04	0.20 ± 0.05	0.08 ± 0.03
TM2	27	10 (37%)	17 (63%)	0.18 ± 0.04	0.18 ± 0.04	0.16 ± 0.05
TM3	22	5 (23%)	17 (77%)	0.15 ± 0.04	0.15 ± 0.04	0.05 ± 0.02
TM4	21	12 (57%)	9 (43%)	0.16 ± 0.04	0.19 ± 0.07	0.15 ± 0.08
TM5	25	11 (44%)	14 (56%)	0.19 ± 0.04	0.22 ± 0.04	0.24 ± 0.05
TM6	25	9 (46%)	16 (64%)	0.14 ± 0.03	0.13 ± 0.03	0.12 ± 0.04
TM7	22	9 (41%)	13 (59%)	0.21 ± 0.05	0.21 ± 0.05	0.16 ± 0.05
MC1R total	319	140 (44%)	178 (56%)	0.18 ± 0.01	0.20 ± 0.01	0.17 ± 0.01

ED = extracellular domain; ID = intracellular domain; TM = transmembrane domain; SE = standard error.

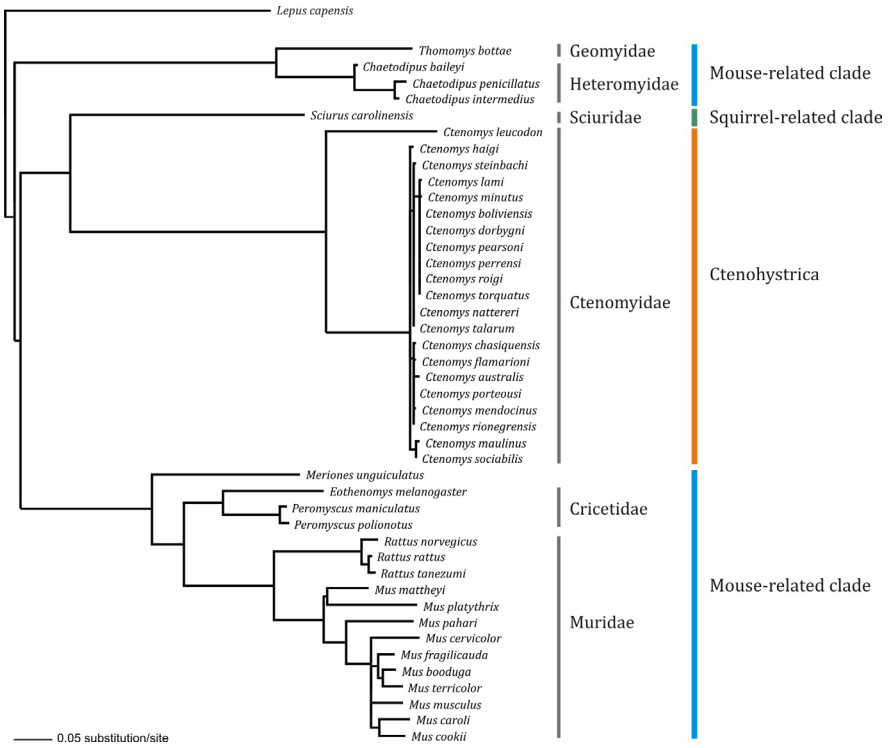

Figure 3. Phylogenetic tree of 43 rodent *MC1R* nucleotide sequences, generated using all codon positions (data set of 945 bp), and rooted using *Lepus capensis*. The topology and branch lengths were estimated using a Bayesian approach, with T93+ G model of nucleotide evolution. Major phylogenetic groups proposed by Bangla-Kanfi et al. (2009) are described on the right side of the tree.

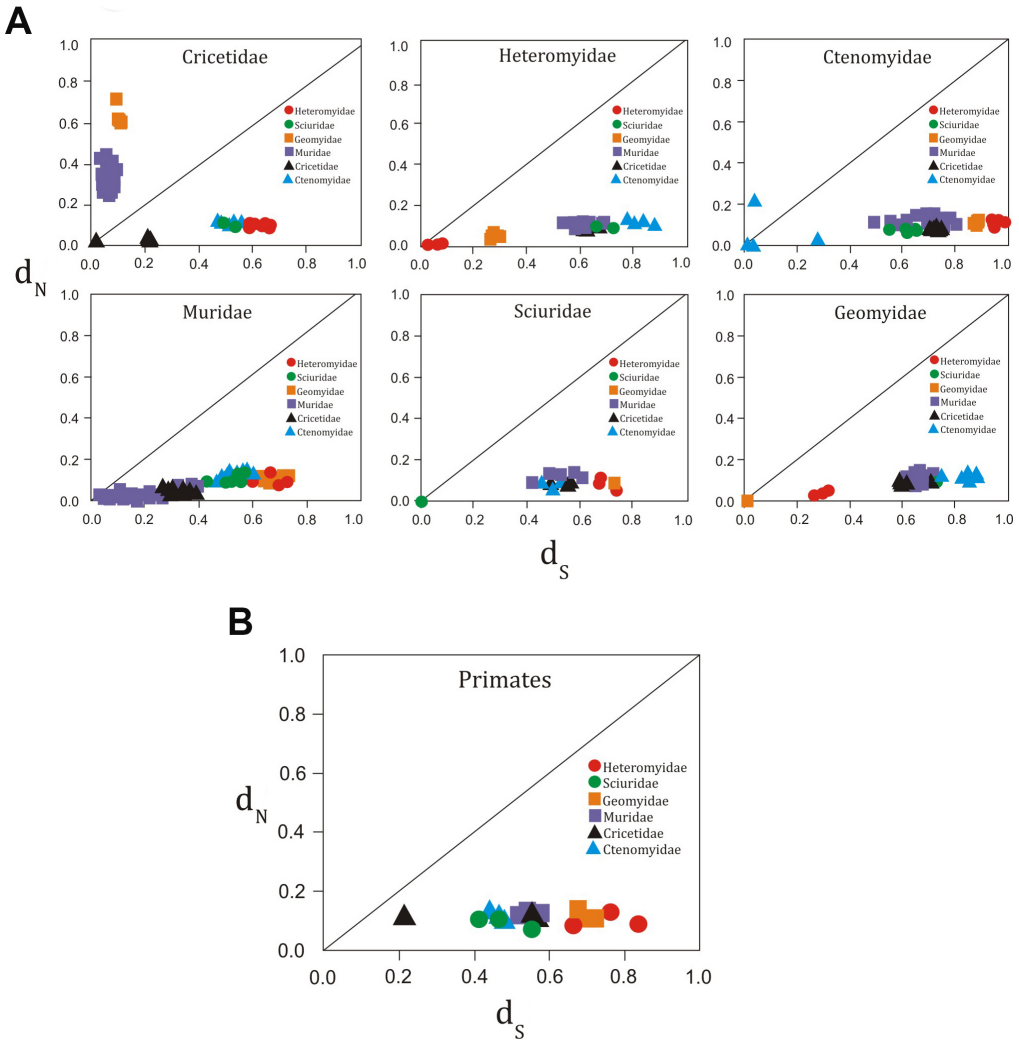

Figure 4. Scatterplots of non-synonymous (d_N) per synonymous (d_S) estimates for the *MC1R*. **A.** Pairwise comparison between families. **B.** Pairwise comparison among primates and rodent families.

Positive selection in *MC1R*

Phylogeny-based maximum likelihood approaches provided evidence that positive selection and relaxation in functional constraints acted on the codons of *MC1R* during the evolutionary history of rodents (Figure 6). We detected significant ω heterogeneity across the entire *MC1R* sequence. A comparison of neutral models M0, M1a, and M7 with variation models M2a, M3, and M8 revealed that M3 and M8 performed significantly better than M0 and M7 (Table 3). M3 and M8 suggested that a small proportion of sites (1%) are under positive selection (ω ≅ 2.9). Four codons (3 M, 14 K, 26 H, and 251 T) have likely evolved under posi-

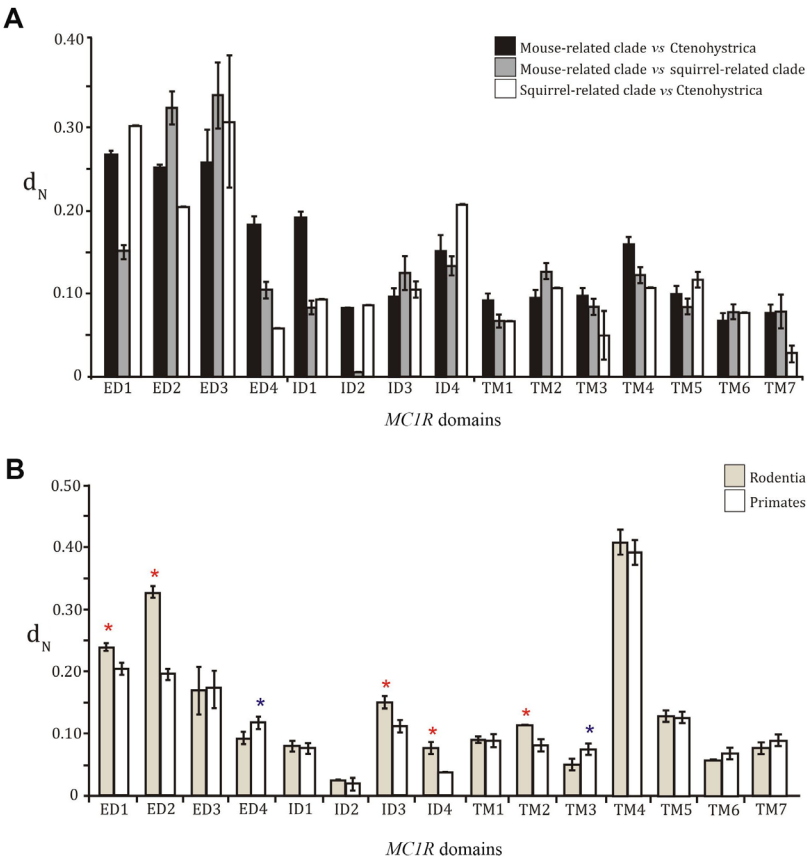

Figure 5. Graph showing estimated number of non-synonymous (d_N) substitution rate for each *MC1R* domain. **A.** Substitutions were estimated as the mean of pairwise nucleotide sequence comparisons in the three major lineages of rodents: mouse-related clade, Ctenohystrica, and squirrel-related clade. **B.** Substitutions estimated by comparing lagomorphs with both rodents and primates. Asterisks in red indicate acceleration in rodents, and in blue, in primates. For domain abbreviations, see legend to Table 2.

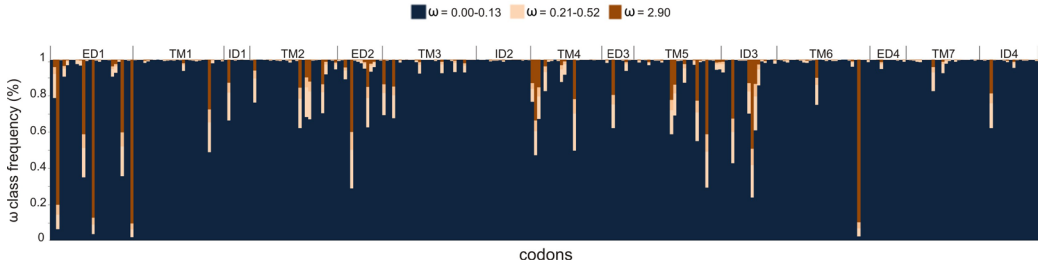

Figure 6. Posterior probabilities that each site is from the 11 site classes (here grouped in only three classes, defined as: 1: ω = 0.00-0.013; 2: ω = 0.21-0.52; 3: ω = 2.9) under the M8 (selection) model, calculated using the Bayes empirical Bayes procedure from each codon of *MC1R* in rodents. For domain abbreviations, see legend to Table 2.

tive selection, 2 (26 H, 251 T) with a posterior probability of >0.95 (see Table 3 and Figure 6). Interestingly, the positively selected sites identified by the empirical Bayes criterion are located in both the ED and TM. Selection models did not perform significantly better than the discrete model M3 when analyzing domains separately for most segments except for TM4 and TM6 (see Table 3). In particular, ID4 was the only segment in which we observed no differences in ω across all samples in which the neutral model M0 was most likely.

Table 3. Parameters estimated under different models of substitution codons for the entire *MC1R* gene in rodents.

Model	Parameters estimates	Log likelihood	P*	Positively selected
M0	$\omega = 0.1170$	-4276.48		None
M3	$p_0 = 0.7150$, $\omega_0 = 0.0365$ $p_1 = 0.2714$, $\omega_1 =$ **0.3836** $p_2 = 0.0135$, $\omega_2 = 2.8530$	-4218.61	<0.001	*3 M*, *14 K*, **26 H**, **251 T**
M1a	$p_0 = 0.8826$, $\omega_0 = 0.0754$ $p_1 = 0.1173$, $\omega_1 = 1$	-4228.67		Not allowed
M2a	$p_0 = 0.8826$, $\omega_0 = 0.0754$ $p_1 = 0.0869$, $\omega_1 = 0$ $p_2 = 0.0303$, $\omega_2 = 1$	-4228.67	>0.999	*14 K*, **26 H**, **251 T**
M7	$p = 0.3223$, $q = 1.7418$	-4222.98		Not allowed
M8	$p_0 =$ **0.9871**, $p = 0.3994$, $q = 2.5055$ $p_1 =$ **0.0128**, $\omega = 2.9009$	-4218.47	0.01	*3 M*, *14 K*, **26 H**, **251 T**

p_0 = proportion of sites where $\omega < 1$; p_1 = proportion of sites where $\omega = 1$; p_2 = proportion of sites where $\omega > 1$ (selection models only). For models M7 and M8, p and q represent parameters of the beta distribution. PAML site models and positively selected sites were identified by empirical Bayes approaches. Positive selection was inferred if the posterior probability of $\omega > 1$ for a site was 0.95 or higher (bold). Sites with a posterior probability of $\omega > 1$ between 0.50 and 0.949 are also shown (italic). Amino acids correspond to the *Mus musculus* sequence. *Degrees of freedom: M0-M3 = 4; M1-M2 = 2; M7-M8 = 2; LRT: $2\Delta l = 2(l_1 - l_0)$.

All other segments showed differences in d_N/d_S ratio, and some also showed relaxation in a small percentage of sites (Table 4). The clade model C used to infer differences in d_N/d_S ratio among rodent lineages was not significantly better (P = 1) than the neutral model M1a. In addition, no significant variation was detected between branch types 0 and 1 among the three site classes (Table 5).

DISCUSSION

Patterns and rates of *MC1R* variation

The alignment of *MC1R* across rodent species indicated the occurrence of several short conserved motifs intercalated with variable segments. In total, 178 amino acids were completely conserved among lineages, particularly in the IDs, showing that these sites are likely the ones under the strongest functional constraints. Therefore, variants at these positions can be expected to have more significant effects on *MC1R* activity than those identified in other domains of the gene. The IDs provide the binding interfaces for heterotrimeric G proteins and contain phosphorylation targets involved in the regulation of signaling, internalization, and cycling (Strader et al., 1994). In mice, the natural tobacco mutation S69L, located within ID1, leads to *MC1R* hyperactivity (Robbins et al., 1993), suggesting that this segment is important for normal receptor activity (Garcia-Borrón et al., 2005).

Table 4. Parameter estimates under models of variable ω ratios (d_N/d_S) among sites for separated domains of *MC1R*.

Domain	Model fit	d_N/d_S [a]	Estimates of parameters	Positively selected sites	P*
ED1	M3: discrete	0.4249	p0 = 0.1928, p1 = 0.7288, p2 = 0.0783 ω0 = 0.0000, ω1 = 0.4004, ω2 = 1.6989	3 M, 14 K	0.009
ED2	M3: discrete	0.0257	p0 = 0.4915, p1 = 0.2744, p2 = 0.2340 ω0 = 0.0000, ω1 = 0.0000, ω2 = 0.1099	None	0.008
ED3	M2: positive selection	0.3653	p0 = 0.6347, p = 0.3622, q = 0.0030 ω0 = 0.0000, ω1 = 1.0000, ω2 = 1.0000	None	0.018
ED4	M3: discrete	0.0665	p0 = 0.1785, p1 = 0.3200, p2 = 0.5013 ω0 = 0.0036, ω1 = 0.0036, ω2 = 0.1290	None	0.001
ID1	M3: discrete	0.3492	p0 = 0.5058, p0 = 0.3072, p2 = 0.1868 ω0 = 0.0210, ω1 = 0.3796, ω2 = 1.1874	17 R	0.007
ID2	M0: one-ratio	0.0244	Not allowed	Not allowed	0.001
ID3	M3: discrete	0.1590	p0 = 0.8308, p1 = 0.0941, p2 = 0.0750 ω0 = 0.1103, ω1 = 0.7062, ω2 = 0.7062	None	0.001
ID4	M3: discrete	0.0663	p0 = 0.5376, p1 = 0.1623, p2 = 0.3000 ω0 = 0.0000, ω1 = 0.0832, ω2 = 0.0832	None	0.001
TM1	M3: discrete	0.1990	p0 = 0.5804, p1 = 0.3789, p2 = 0.0406 ω0 = 0.0000, ω1 = 0.2160, ω2 = 2.7816	**26 H**	0.001
TM2	M3: discrete	0.2369	p0 = 0.5889, p1 = 0.2892, p2 = 0.1217 ω0 = 0.0099, ω1 = 0.3821, ω2 = 1.0606	None	0.001
TM3	M3: discrete	0.0845	p0 = 0.5851, p1 = 0.0056, p2 = 0.4092 ω0 = 0.0000, ω1 = 0.2037, ω2 = 0.2037	None	0.001
TM4	M3: discrete	0.1712	p0 = 0.2123, p1 = 0.5811, p2 = 0.2065 ω0 = 0.0000, ω1 = 0.1020, ω2 = 0.5416	None	0.001
TM5	M3: discrete	0.2851	p0 = 0.4481, p1 = 0.5087, p2 = 0.0431 ω0 = 0.0143, ω1 = 0.3900, ω2 = 1.8059	None	0.001
TM6	M2: positive selection	0.3072	p0 = 0.9123, p1 = 0.0476, p2 = 0.0400 ω = 0.0558, ω1 = 1.0000, ω2 = 5.2135	**251 T**	0.006
TM7	M7: discrete	0.0760	p0 = 8.18354, q = 99.0000	None	1

p0 = proportion of sites where ω < 1; p1 = proportion of sites where ω = 1; p2 = proportion of sites where ω > 1 (selection models only). For models M7 and M8, p and q represent parameters of the beta distribution. PAML site models and positively selected sites were identified by empirical Bayes approaches. Positive selection was inferred if the posterior probability of ω > 1 was 0.95 or higher (bold) for a given site. Amino acids correspond to position in the *Mus musculus* complete *MC1R* sequence. [a]d_N/d_S ratio is the average across all codons. *Degrees of freedom: M0-M3 = 4; M1-M2 = 2; M7-M8 = 2; LRT: $2\Delta l = 2(l_1 - l_0)$.

Table 5. Branch-site clade model C values of ω (d_N/d_S ratio) estimated for three site classes.

Site class	Proportion	Branch type 0 (Squirrel-related clade)	Branch type 1 (Mouse-related clade + Ctenohystrica)
0	0.5518	ω = 0.0250	ω = 0.0250
1	0.0431	ω = 1.0000	ω = 1.0000
2	0.4050	ω = 0.2132	ω = 0.2289

Heterogeneity in evolutionary rates was evident among the lineages of rodents and *MC1R* domains. Taxon-specific acceleration of the amino acid substitution rate was observed in ED4 and TM7, suggesting that the mouse-related clade has evolved faster than the squirrel-related and Ctenohystrica clades have at these *MC1R* segments. At the nucleotide level, we observed differences in rates across lineages of rodents, but most of them were not significant, as within-lineage variation was evident in several comparisons. We suggest that this high variability might have masked the expected increased differences associated with phylogenetic depth, probably indicating that some lineages have a species-specific pattern. In particular, a contrasting pattern was identified in the mouse-related clade *vs* the squirrel-related clade and in the mouse-related clade *vs* Ctenohystrica for the TM4 domain, in which higher rates in the deepest comparisons were observed in the first group. These observations, interpreted in the

context of the overall pattern of *MC1R* variation, might indicate potential lineage-specific changes in functional constraints or episodic adaptive evolution in this gene.

A faster rate of replacement substitution in half of the *MC1R* domains was observed in rodents relative to comparisons that included primates. Particular segments, including ED1, ED2, ID3, ID4, and TM2, showed a significant difference between these groups. Interestingly, other segments, such as ED4 and TM3, showed the reverse pattern, in which evidence of acceleration in primates was observed. Therefore, in view of this variation, a conspicuous pattern of increased substitution rate in rodents could not be ruled out for *MC1R*.

MC1R substitution rate scenario

Among the 3 rodent clades, the Cricetidae + Muridae lineage within the mouse-related clade seems to have an increased substitution rate, primarily evidenced by longer branch lengths in the phylogenetic tree and higher values of pairwise d_N. In contrast, a potential lineage-specific acceleration could be ruled out in the case of Ctenohystrica, in which one taxon (*C. leucodon*) showed significant variation in relation to other *Ctenomys* species.

Substitution rates can be systematically affected by certain species characteristics, including aspects of evolutionary history, such as population size (Smith and Donoghue, 2008), and life history traits, such as body size (Welch et al., 2008). One recurrent factor for the higher rate of substitution in the rodent lineage is the generation-time effect - i.e., the higher rate occurs because rodents have a comparatively shorter generation time. However, this argument fails to explain the similarity of the *MC1R* substitution rate in rodents, as the generation time is relatively short in all taxa in this order. In this case, the variation might indicate fine-scale differences in species correlated with individual life history traits (Bromham, 2009).

Overall, several amino acid changes in *MC1R* are associated with coat-color polymorphism in mammals. This relationship indicates that the gene has a conspicuous phenotypic effect, which in some lineages might be subject to strong selective pressure, particularly in rodents. This group exhibits marked color polymorphisms, and its members are often predated by visually oriented birds of prey such that maintenance of crypsis is likely under strong ecological pressure. Thus, selection on this morphological trait might also be underlying the high substitution rate of non-synonymous change in *MC1R* for some taxa-specific lineages.

Selection on *MC1R*

Molecular evolution analyses have indicated that purifying selection has acted during the majority of *MC1R* evolutionary history in some groups of mammals, such as cetartiodactyls, mustelids, and primates (Mundy and Kelly, 2003; Hosoda et al., 2005; Ayoub et al., 2009; Shimada et al., 2009). However, our survey of rodents indicated a higher rate of sites that have evolved under the relaxation of functional constraints as well as two sites with evidence of positive selection signatures. Sites with a ω value of approximately 0.38, which indicates relaxation of functional constraints, are located in the ED and TM. The positively selected sites (26 and 251) are also located in both the ED and the TM.

MC1R is unusually polymorphic, and many of the natural variants are functionally relevant (Wong and Rees, 2005). EDs are generally small, particularly ED3, which shows high constitutive activity. Mutation of residues Glu269 and Thr272 to Ala in ED4 of human *MC1R*

reportedly lowers the binding affinity for agonists, suggesting that these residues are involved in ligand recognition (Chhajlani et al., 1996). Moreover, in TMs, the ligand-binding site is a pocket located below the plasma membrane-extracellular medium interface, formed through the contributions of several fragments (Garcia-Borrón et al., 2005). Three-dimensional models of ligand-receptor complexes have been developed (Haskell-Luevano et al., 1996) that suggest that a highly charged region containing Glu94 (TM2), Asp-117, and Asp-121 (TM3) interacts with an Arg residue. A network of aromatic residues located near the extracellular side of TM4, TM5, and TM6 also potentially contributes to agonist binding by interacting with aromatic residues. Interestingly, 11 mutations described in vertebrates are located in TM2 of *MC1R*, and several of them have important functional consequences, particularly involvement in pelage and feather melanism (Hoekstra, 2006).

Our results indicated a high proportion of sites with ω values of 0.38 for TM2, suggesting relaxation of constraints for this segment probably due to a functional implication. In addition, TM4 and TM6 are likely to have evolved under selection rather than neutral or discrete models, evidenced by significantly higher likelihood values.

In particular, one positively selected site (251) observed in *MC1R* with high Bayes-empirical Bayes posterior probability is located in TM6. The other (26) is located in ED1, formally characterized as the N-terminal domain. Although we have no evidence of functional effects of amino acid changes in these sites from our data or from that of previous studies, the results of positive selection indicate that they might have morphological significance. Alternatively, these sites could be involved in other physiological effects of *MC1R* variants, such as a kappa-opioid receptor that mediates analgesia in mice and humans (Mogil et al., 2003).

In summary, our results indicate that purifying selection is not the only evolutionary force that has shaped *MC1R* in rodents, because a significant number of codons have evolved under relaxation of functional constraints, and a few have been positively selected. Although our estimates indicated only a small proportion of sites at which ω is greater than 1, the possibility remains that positive directional selection is the driving force behind the increase in ω estimates. Finding amino acid replacements in an excess of d_N, globally or in specific regions, provides unequivocal evidence of positive selection at the molecular level. Nevertheless, this criterion may be excessively stringent, and we suggest that statistically significant increments in ω might indicate positive selection.

Several examples of single-nucleotide polymorphisms having pronounced evolutionary consequences can be cited - e.g., in genes involved in pathogen virulence (Brault et al., 2007) or pigmentation (Nachman et al., 2003; Manceau et al., 2010). Because the positively selected sites in the *MC1R* identified in this study are located in critical regions of ligand binding of the protein, functional changes in these regions might have direct consequences for some phenotypes. Consequently, the selection and relaxation of functional constraint that we have observed may indicate the evolvability of the system for the generation of adaptive changes in specific taxa in the rodent lineage.

ACKNOWLEDGMENTS

Research supported by Coordenadoria de Aperfeiçoamento Pessoal, Conselho Nacional de Desenvolvimento Científico e Tecnológico, and Fundação de Amparo à Pesquisa do Rio Grande do Sul. G.L. Gonçalves received a doctoral fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (#141604/2007-7).

REFERENCES

Ayoub NA, McGowen MR, Clark C, Springer MS, et al. (2009). Evolution and phylogenetic utility of the melanocortin-1 receptor gene (*MC1R*) in Cetartiodactyla. *Mol. Phylogenet. Evol.* 52: 550-557.

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, et al. (2009). Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evol. Biol.* 9: 71.

Brault AC, Huang CY, Langevin SA, Kinney RM, et al. (2007). A single positively selected West Nile viral mutation confers increased virogenesis in American crows. *Nat. Genet.* 39: 1162-1166.

Bromham L (2009). Why do species vary in their rate of molecular evolution? *Biol. Lett.* 5: 401-404.

Chhajlani V, Xu X, Blauw J and Sudarshi S (1996). Identification of ligand binding residues in extracellular loops of the melanocortin 1 receptor. *Biochem. Biophys. Res. Commun.* 219: 521-525.

Crandall KA and Hillis DM (1997). Rhodopsin evolution in the dark. *Nature* 387: 667-668.

Garcia-Borrón JC, Sanchez-Laorden BL and Jimenez-Cervantes C (2005). Melanocortin-1 receptor structure and functional regulation. *Pigment Cell Res.* 18: 393-410.

Gonçalves GL and Freitas TRO (2009). Intraspecific variation and genetic differentiation of the collared tuco-tuco (*Ctenomys torquatus*) in southern Brazil. *J. Mamm.* 90: 1020-1031.

Gonçalves GL, Hoekstra HE and Freitas TRO (2012). Striking coat colour variation in tuco-tucos (Rodentia: Ctenomyidae): a role for the melanocortin-1 receptor? *Biol. J. Linn. Soc.* 105: 665-680.

Haskell-Luevano C, Sawyer TK, Trumpp-Kallmeyer S, Bikker JA, et al. (1996). Three-dimensional molecular models of the hMC1R melanocortin receptor: complexes with melanotropin peptide agonists. *Drug Des. Discov.* 14: 197-211.

Hoekstra HE (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97: 222-234.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, et al. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313: 101-104.

Hosoda T, Sato JJ, Shimada T, Campbell KL, et al. (2005). Independent nonframeshift deletions in the MC1R gene are not associated with melanistic coat coloration in three mustelid lineages. *J. Hered.* 96: 607-613.

Huelsenbeck JP, Ronquist F, Nielsen R and Bollback JP (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294: 2310-2314.

Jackson IJ (1997). Homologous pigmentation mutations in human, mouse and other model organisms. *Hum. Mol. Genet.* 6: 1613-1624.

Jackson IJ, Budd P, Horn JM, Johnson R, et al. (1994). Genetics and molecular biology of mouse pigmentation. *Pigment Cell Res.* 7: 73-80.

Jansa SA, Lundrigan BL and Tucker PK (2003). Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *Mus*. *J. Mol. Evol.* 56: 294-307.

Kimura M (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Manceau M, Domingues VS, Linnen CR, Rosenblum EB, et al. (2010). Convergence in pigmentation at multiple levels: mutations, genes and function. *Phil. Trans. R. Soc. B* 365: 2439-2450.

McRobie H, Thomas A and Kelly J (2009). The genetic basis of melanism in the gray squirrel (*Sciurus carolinensis*). *J. Hered.* 100: 709-714.

Mogil JS, Wilson SG, Chesler EJ, Rankin AL, et al. (2003). The melanocortin-1 receptor gene mediates female-specific mechanisms of analgesia in mice and humans. *Proc. Natl. Acad. Sci. U. S. A.* 100: 4867-4872.

Mundy NI and Kelly J (2003). Evolution of a pigmentation gene, the melanocortin-1 receptor, in primates. *Am. J. Phys. Anthropol.* 121: 67-80.

Nachman MW, Hoekstra HE and D'Agostino SL (2003). The genetic basis of adaptive melanism in pocket mice. *Proc. Natl. Acad. Sci. U. S. A.* 100: 5268-5273.

Nei M and Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3: 418-426.

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

Nylander JA, Ronquist F, Huelsenbeck JP and Nieves-Aldrey JL (2004). Bayesian phylogenetic analysis of combined data. *Syst. Biol.* 53: 47-67.

Pond SL, Frost SD and Muse SV (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676-679.

Robbins LS, Nadeau JH, Johnson KR, Kelly MA, et al. (1993). Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* 72: 827-834.

Shimada T, Sato JJ, Aplin KP and Suzuki H (2009). Comparative analysis of evolutionary modes in Mc1r coat color gene in wild mice and mustelids. *Genes Genet. Syst.* 84: 225-231.

Smith SA and Donoghue MJ (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science* 322: 86-89.

Strader CD, Fong TM, Tota MR, Underwood D, et al. (1994). Structure and function of G protein-coupled receptors. *Annu. Rev. Biochem.* 63: 101-132.

Swofford DL (2003). PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b 10. Sinauer Associates, Sunderland.

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.

Tschirren B, Andersson M, Scherman K, Westerdahl H, et al. (2012). Contrasting patterns of diversity and population differentiation at the innate immunity gene toll-like receptor 2 (TLR2) in two sympatric rodent species. *Evolution* 66: 720-731.

Welch JJ, Bininda-Emonds OR and Bromham L (2008). Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol. Biol.* 8: 53.

Wlasiuk G and Nachman MW (2007). The genetics of adaptive coat color in gophers: coding variation at Mc1r is not responsible for dorsal color differences. *J. Hered.* 98: 567-574.

Wong TH and Rees JL (2005). The relation between melanocortin 1 receptor (MC1R) variation and the generation of phenotypic diversity in the cutaneous response to ultraviolet radiation. *Peptides* 26: 1965-1971.

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

Yang Z and Bielawski JP (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15: 496-503.

Yang Z, Nielsen R, Goldman N and Pedersen AM (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431-449.

