Front. Agric. China 2010, 4(1): 109–115
DOI 10.1007/s11703-009-0081-3

RESEARCH ARTICLE

Huiqin ZHENG, Xianglong LI, Rongyan ZHOU, Lanhui LI, Xiuli GUO, Jingfen KANG, Dongfeng LI

Bioinformatics analysis of tyrosinase-related protein 1 gene (*TYRP1*) from different species

Abstract As one member of the tyrosinase-related family directly involved in the production of melanin, *TYRP1* is involved in not only melanogenesis but also prevention of melanocyte death, stabilizing tyrosinase and helping determine the shape of melanosomes, etc. Multi-species sequence comparisons showed that there were two evolutionally conserved non-coding regions (from −1306 to −733 and from −642 to −515 according to AL138753) upstream of translational initiation sites, representing putative regulatory regions subject to subsequent experimental tests. Coding sequence length variation and genetic diversity analysis showed that *Felis catus*, *Homo sapiens* and *Canis familiaris* had more genetic diversities than the other species for *TYRP1*, especially *Felis catus* that could be a better choice for studying the *TYRP1*-associated genetic basis underlying the color diversity. As a 75 kDa type-1 transmembrane glycoprotein, mature TYRP1 possesses about 17 kDa modifying components, whose function predominantly depends on the existing glycosyl-groups and the Cu components. In addition, the mutated amino acids within species and the highly conserved amino acids among species were listed in our paper.

Keywords melanin, *TYRP1*, bioinformatics

1 Introduction

Discrete variation in external morphology or appearance that segregates in pedigrees provides a cornerstone of genetics in all organisms, so it is no surprise that coat color variation plays a crucial role in animal genetics. Color in animals depends on the presence of melanin in the skin and hair. The melanin resides in cellular organelles called melanosomes, which is produced in melanocytes. As a class of derivatives of the amino acid tyrosine, melanin is synthesized in two different types of pigment cells: neural-crest-derived melanocytes and optic cup-derived retinal pigment epithelium (RPE). The tyrosinase-related family includes tyrosinase, tyrosinase-related protein 1 (TYRP1) and tyrosinase-related protein 2 (TYRP2) involved in this enzymatic process that converts tyrosine to melanin pigments. Indeed, two types of melanin are produced by melanocytes, pheomelanins, which are red or yellow, and eumelanins, which are brown or black (Bertolotto et al., 1998). The first two steps of both eumelanin and pheomelanin production involve tyrosinase catalysing the conversion of tyrosine to 3, 4-dihydroxy-L-phenylalanine (DOPA) and of DOPA to DOPA quinine (Hearing, 1987; Prota, 1988). Then pheomelanogenesis seems to be the default pathway in the absence of MC1R signaling, with a low tyrosinase activity and a high concentration of thiolic compounds, such as cysteine. In another way, eumelanin synthesis requires α-MSH binding to MC1R (Ito, 2003; Alonso et al., 2008), which transcriptionally activates tyrosinase and upregulates *TYRP1* and *TYRP2* (Jiménez-Cervantes et al., 1994; Kobayashi et al., 1994; Kobayashi et al., 1995).

In addition to its role in melanin synthesis, TYRP1 is involved in stabilizing tyrosinase protein and modulating its catalytic activity. TYRP1 is also involved in maintenance of melanosome structure and affects melanocyte proliferation and melanocyte cell death (Sarangarajan and Boissy, 2001; Rad et al., 2004; Kobayashi and Hearing, 2007). In a word, *TYRP1* affects the final pigmentation and color pattern. Besides putative cis-regulatory region detection, bioinformatics analysis of variation and conservation of coding sequence (CDS) and amino acid sequence of TYRP1 within and among species (Table 1) will all make for efficient experiment design and help us better understand the color genetics basis.

Received May 7, 2009; accepted July 5, 2009

Huiqin ZHENG, Xianglong LI (✉), Rongyan ZHOU, Lanhui LI, Xiuli GUO, Jingfen KANG, Dongfeng LI
College of Animal Science and Technology, Agricultural University of Hebei, Baoding 071001, China
E-mail: lixianglongcn@yahoo.com

Table 1 *TYRP1* sequences from GenBank and related characteristics

species	CDS sequence		protein sequence		gene sequence	stop codon	CDS length
Felis catus					—		
	AY965743.1,	AY965744.1,	AAY87040.1,	AAY87041.1,		TAA(7)	1614(5)
	AY965745.1,	AY965746.1,	AAY87042.1,	AAY87043.1,		TGA(2)	1663(1)
	AY956310.1,	AH014499.1,	AAY29636.1,	AAV65117.1,			1666(1)
	AH014500.1,	AH014501.1,	AAV65118.1,	AAV65119.1,			300(2)
	NM_001042560.2		NP001036025.2				
Homo sapiens	NM_000550.2,	AF001295.1,	NP_000541.1,	AAC15468.1,	AL138753.8	TAA(5)	1614(3)
	X51420.1,	BC052608.1,	CAA35785.1,	AAH52608.1,		TGA(2)	1611(1)
	CR407683.1,		CAG28611.1,				1584(2)
	AL138753.8 (2 spliced variants)		CAD13328.1,				741(1)
			CAH73438.1				
Mus musculus	AL670884.7,	X03687.1,	CAM22450.1,	CAA27323.1,	AL670884.7	TGA(7)	1614(7)
	NM_031202.2,	BC076598.1,	NP_112479.1,	AAH76598.1,			
	AK148332.1,	AK148370.1,	BAE28487.1,	BAE28511.1,			
	AK148441.1		BAE28554.1				
Ovis aries	EU760771.1,	EF102110.1,	ACF21681.1,	ABK97637.1,	—	TAA(3)	1614(3)
	EF102111.1,	EF102112.1,	ABK97638.1,	ABK97639.1,		?(4)	?(4)
	EU760770.1,	EF102109.1,	ACF21680.1,	ABK97636.1,			
	NM_001130023.1		NP_001123495.1				
Bos taurus	AF400250.1,	AF445638.2,	AAK85404.1,	AAL38167.2,	NW_001495437.1	TAG(3)	1614(2)
	NM_174480.3		NP_776905.2				1614(revised)
Canis familiaris	XM859450.1,	XM859433.1,	XP_864543.1,	XP_864526.1,	—	TAA(3)	1614(1)
	XM531934.2		XP_531934.1				1617(2)
Coturnix japonica	EU046599.1,	EU046600.1,	ABW05105.1,	ABW05106.1,	—	TGA(2)	1611(2)
	AB005228.1		BAA89535.1			TAA(1)	1611(revised)
Danio rerio	BC076406.1,	BX901913.8,	AAH76406.1,	CAM13049.1,	—	TGA(3)	1581(2)
	NM001002749.1		NP001002749.1				1572(1)
Sus scrofa	NM_001025226.1,	AB207240.1	NP_001020397.1,	BAD99581.1	—	TGA(2)	1614(2)
Equus caballus	BK000021.1,	NM001081840.1	DAA00083.1,	NP001075309.1	NW_001799690.1	TAA(2)	1626(2)
Xenopus laevis	NM001087023.1,	BC043815.1	NP001080492.1,	AAH43815.1	—	TGA(2)	1605(2)
Rattus norvegicus	XM00110664.1		NP001100134.1		—	TGA(1)	1614(1)
Pan troglodytes	XM520488.2		XP520488.2		—	TAA(1)	1614(1)
Macaca mulatta	XM001111475.1		XP001111475.1		—	TAA(1)	1614(1)
Capra hircus	AF136926.2		AAD34802.2		—	TAA(1)	1614(1)
Gallus gallus	NM205045.1		NP_990376.1		—	TGA(1)	1608(1)
Xenopus tropicalis	CR760901.2		CAJ82935.1		—	TGA(1)	1605(1)
Takifugu rubripes	AF397401.1		AAL84110.1		—	TGA(1)	1566(1)

2 Materials and methods

All the complete gene sequences and CDSs were obtained from NCBI (Table 1). Multi-species sequence comparisons were carried out using the zPicture software (Hardison, 2000; Pennacchio and Rubin, 2001; Ovcharenko et al., 2004), which compared pairwise the reference sequence, as the mouse *TYRP1* locus (AL670884), with corresponding sequences from other species, that is, *Homo sapiens* (AL138753), *Bos taurus* (NW001495437) and *Equus caballus* (NW001799690).

All of the 55 CDSs were aligned using the Clustal W program implemented in BioEdit (version7.0.5). DnaSP (version 4.0) software was used to analyze the number of haplotypes (h), the haplotype diversity (*Hd*), the average number of nucleotide differences (*K*) (Tajima, 1983), the nucleotide diversity (π), synonymous nucleotide diversity (π_s), nonsynonymous nucleotide diversity (π_a) with the

Jukes and Cantor correction, the polymorphic site (S), the singleton variable sites (SP), the parsimony informative sites (PIP), the number of singleton variable sites with 2 variants (NS2), and the number of parsimony-informative sites with 2 variants (NP2) for each species.

We positioned the conserved amino acids including the sorting and targeting signals among species after alignment of the 55 amino acid sequences using the same Clustal W program and then scanned for motifs (pattern, HMMs) in the sequence CAM22450 on MyHits web site (Pagni et al., 2007; Hulo et al., 2008; http://myhits.isb-sib.ch). Hydrophobicity analysis was carried out with the algorithm of Kyte and Doolittle in BioEdit followed by SMART (Simple Modular Architecture Research Tool, SMART) analysis online (http://smart.embl-heidelberg.de).

3 Results and discussion

3.1 Analysis of the regulatory elements

In order to find the evolutionally conserved non-coding regions as putative regulatory regions, multi-species sequence comparisons were carried out using the zPicture software, which showed that two regions (from −1306 to −733 and from −642 to −515) were highly conserved among the species *Mus musculus*, *Homo sapiens*, *Bos taurus* and *Equus caballus*. The most common method of testing the cis-regulatory function of such a region was to link it to a reporter gene encoding either b-galactosidase (LacZ) or a fluorescent protein, make a transgenic animal or melanin-associated cell line containing the reporter construct, and then look for a reporter activity *in vivo*, while the correlation analysis between *TYRP1* polymorphism and color variation was another powerful way. The base number was set according to AL138753, as per numbering of Shibahara et al. (1986), in which the base before the translational initiation codon was −1.

The reported cis-regulatory activities of binding sites for MITF (M-box: AA/GTCATGTGCT) as a determining factor in melanocyte development, OCT-1 or related (POU domain) proteins (octamer motif: ATTTGAAT), Pax3 (two melanocyte-specific elements termed the MSEu and MSEi: GTGTGA), and Otx2 (bicoid/Otx consensus element: GGATTA) found in the first region (−1306 to −733) not only testified the regulatory importance of this region, but also supported the approaches from Elnitski et al. (2003) to distinguish the regulatory regions from neutrally evolving DNA (Lowings et al., 1992; Yavuzer and Goding, 1994; Martínez-Morales et al., 2003; Levy et al., 2006).

3.2 Variation of stop codon

In this paper, we got three stop codon mutations within *Felis catus*, *Homo sapiens* and *Coturnix japonica*, each in one species. In *Felis catus*, seven CDSs in Table 1 have TAA as stop codon with two variants using premature TGA stop codon due to C to T nonsense mutation at position 298, which is associated with the cinnamon phenotype in domestic cats (Lyons et al., 2005; Schmidt-Küntzel et al., 2005). In *Homo sapiens*, most CDSs have stop codon TAA, except for two variants using TGA due to the tetranucleotide (AGAA) insertion after position 1575 and before TATGA. And in *Coturnix japonica*, there are two forms of stop codon (TGA and TAA) at the corresponding position. The stop codon usage bias for *TYRP1* among species from our analysis do not conform to Sun et al. (2005) well, who showed that TAG was the least used in all eukaryotes, TAA was overrepresented in less complex eukaryotes such as yeast and invertebrate, and TGA was overrepresented in more complex eukaryotes, especially in vertebrates, while for *TYRP1*, although TAG was only found in *Bos taurus*, TAA was mainly found in higher vertebrates like *Homo sapiens*, except for *Mus musculus*, *Sus scrofa* and *Rattus norvegicus* using TGA and only *Bos taurus* using TAG, and TGA was mainly found in lower vertebrates such as *Daniorerio* and *Xenopus laevis* (Table 1). Therefore, the stop codon usage bias have some specificity for different species and different genes.

Since the three stop codons had different stopping efficiency, with TAA the highest and TGA the lowest (Tate et al., 1999), stop codon variation could be implicated in phenotype variation. Compared to the other two stop codons, TGA tends to contribute to phenotype variety to a higher degree via affecting the expression level of functional proteins. However, direct correlation between stop codon variation and color variation was not concluded here and further investigation is needed.

3.3 CDS length variation within and among species

The overall sequence length profile can be seen in Table 1. The complete CDS length variation of *TYRP1* among species showed that in the higher vertebrates, such as *Homo sapiens* and *Mus musculus*, most of the analyzed sequences had a coding region of 1614 bp, except for that from *Equus caballus* (1626 bp) and some variants especially in *Felis catus* and *Homo sapiens*; however, in the lower vertebrates, such as *Danio rerio* and *Xenopus laevis*, the length of complete CDS was all shorter than 1614 bp, and with more heterogeneity. The interspecific complete CDS length variations were found in *Canis familiaris*, *Danio rerio* and especially in *Felis catus* and *Homo sapiens*.

There were three variation forms in *Homo sapiens* and two variation forms in *Felis catus*. In *Homo sapiens*, the first variation form (only 741 bp) (one variant from AL138753) was the combined result of one alternative transcriptional initiation, whose site was set in the otherwise exon3, and one alternative splicing that cut out the otherwise exon4, which together depleted the first

larger part of the wild-type protein sequence and then held back the transportation of this alternative product into the endoplasmic reticulum (ER), let alone into the melanosome. The second one (1611 bp) was made up of only one deletion (TTA) after position 1082 (another variant from AL138753). The third one (1584 bp) existing in two sequences (X51420 and CR407683) was caused by tetranucleotide (AGAA) insertion and the subsequent frameshift mutation that lead to pre-termination. In *Felis catus*, the first one (only 300 bp) found in two sequences (AH014500 and AY965746) was caused by C to T nonsense mutation at position 298, which also affected the transportation of the protein products into the melanosome due to the missing of the targeting signals in C terminal. The second one (1663 bp and 1666 bp) was made up of one large insertion (AAACATTCTCCTATGCTTTTTTGCGGGCTCAGCAAGGAGACCGTTCAGGTGATG) encoding ETFSYAFLRAQQGDRSGD after position 1262 in AY965745, and the same large insertion just missing GTG in AY965744, associated with the chocolate phenotype in domestic cats (Lyons et al., 2005; Schmidt-Küntzel et al., 2005).

There was only one CDS length variation in both *Canis familiaris* and *Danio rerio*, respectively. The length variation in *Canis familiaris* came from alternative splicing that only gave little base variation such as trinucleotide (AAA) insertion after position 382 in XM859433 and dinucleotide (TG) insertion after position 1265 in XM859450. The length variation in *Danio rerio* was just due to one 9-bp (ACATGCTTG) deletion encoding DML after position 679 in BX901913.

The CDS length variation caused by insertions, deletions or alternative splicing as a primary aspect gave protein products polymorphism. The species with more CDS length variations might have more phenotype varieties and could be used as a choice of target to study the genetic basis underlying the corresponding phenotype; the result in our paper brought forward *Felis catus* as a good target for color study.

3.4 Genetic diversity within and among species

The alignment of the 52 CDSs (excluding the three excessively short ones) listed in Table 1 was carried out using BioEdit. The subsequent DnaSP analysis indicated that the selected region (1–1693) of 52 sequences from different species had 1511 sites, excluding sites with gaps. There were 595 invariable (monomorphic) sites and 916 variable (polymorphic) sites including 52 singleton variable sites and 864 parsimony informative sites. Among the 52 singleton variable sites, there were 48 sites with 2 variants and 4 sites with 3 variants. There were 435 sites with 2 variants, 301 sites with 3 variants, and 128 sites with 4 variants in the 864 parsimony informative sites. The nucleotide diversity and average number of nucleotide differences for all sequences were 0.18157 and 274.359, respectively.

Table 2 shows the polymorphic information of *TYRP1* for each species. Different species have different genetic diversity estimates for the *TYRP1*. Usually, more genetic diversity is more useful for nature selection. Because of the limited sequences from each species, we could only compare the species with the same number of sequences. Therefore, for the species with seven sequences and *Homo sapiens* with six sequences, *Mus musculus* had the smallest average number of nucleotide differences (0.952) and nucleotide diversity (0.00059), which was in conformity with the somewhat unitary color phenotypes in mouse, while *Felis catus* had the largest average number of nucleotide differences (8.190) and nucleotide diversity (0.00510), which was consistent with the diversiform color phenotypes in cat. For the species with three sequences, *Canis familiaris* had the largest average number of nucleotide differences (5.333) and nucleotide diversity

Table 2 The polymorphic information of *TYRP1* for each species

species	N	h	*H*d	*K*	π	π_s	π_α	S	SP	PIP	NS2	NP2
Felis catus	7	5	0.905	8.190	0.00510	0.01514	0.00202	16	0	16	0	16
Mus musculus	7	2	0.476	0.952	0.00059	0.00000	0.00078	2	0	2	0	2
Ovis aries	7	6	0.952	7.619	0.00483	0.01505	0.00166	18	5	13	5	13
Homo sapiens	6	4	0.800	4.733	0.00300	0.00000	0.00392	10	3	7	3	7
Bos taurus	3	2	0.667	3.333	0.00207	0.00349	0.00163	5	5	0	5	0
Canis familiaris	3	3	1.000	5.333	0.00332	0.00179	0.00217	8	8	0	8	0
Coturnix japonica	3	2	0.667	4.667	0.00290	0.00878	0.00109	7	7	0	7	0
Danio rerio	3	2	0.667	2.667	0.00170	0.00707	0.00000	4	4	0	4	0
Sus scrofa	2	1	0.000	0.000	0.00000	0.00000	0.00000	0	0	0	0	0
Equus caballus	2	1	0.000	0.000	0.00000	0.00000	0.00000	0	0	0	0	0
Xenopus laevis	2	1	0.000	0.000	0.00000	0.00000	0.00000	0	0	0	0	0

Note: N, h, *H*d and *K* represent No. of sequence, No. of haplotypes, haplotype diversity and average number of nucleotide differences respectively. π, π (s), and π (α) represent nucleotide diversity, synonymous nucleotide diversity and nonsynonymous nucleotide diversity respectively. S, SP, PIP, NS2 and NP2 represent No. of polymorphic sites, singleton variable sites, parsimony informative sites, No. of singleton variable sites with 2 variants and No. of parsimony-informative sites with 2 variants respectively.

(0.00332), while *Bos taurus* had the smallest average number of nucleotide differences (3.333) and nucleotide diversity (0.00207).

Compared with other species in this paper, *Felis catus* and *Canis familiaris* had more genetic diversity for *TYRP1*. Hence, relational analysis of *TYRP1* polymorphism should be carried out on these two species.

3.5 Amino acid variation and conservation

Table 3 shows the mutations of Ala3Gly, Asn459Lys, Arg100Stop, Ser166Stop, Gly254Arg, 362delTyr, Ser363Asn, codon368delA, Cys86Tyr, Arg302Cys, Thr352Ile, Glu451Gly, Asn102Ser, Cys290Phe, Phe94Ser, Phe342Tyr, Ser41Cys, Gln331Stop, 345delPro, Phe282Ser, and 227-229delAsp-Met-Leu, all take place at conserved positions and thus may have important genetic effects. The effects of Ala3Gly, Arg100Stop, Ser166Stop, codon368delA, Cys86Tyr, Arg302Cys, Cys290Phe, Ser41Cys, Gln331Stop, 345delPro, and Phe282Ser had already been testified (Shibahara et al., 1992; Manga et al., 1997; Schmutz et al., 2002; Lyons et al., 2005; Schmidt-Küntzel et al., 2005; Gratten et al., 2007; Nadeau et al., 2007).

In addition, we also searched for the signal motifs responsible for correctly intracellular sorting and targeting of TYRP1 protein product from ER to cis-Golgi (di-acidic code DXE) and from trans-Golgi to melanosome (QPLLTD and GY) (Williams and Fukuda, 1990; Mathews et al., 1992; Vijayasaradhi et al., 1995; Nishimura and Balch, 1997; Xu et al., 1998). It was shown that the motifs GY at 461 and QPLLTD at 512 (except for *Xenopus* with QPLLGE, *Takifugu rubripes* with QPLLID and *Danio rerio* with EPLLGE) were more conserved than the di-acidic code DXE at 525, in which the D and the E had some variation.

Theoretically, the amino acids conserving among species should play a crucial role in protein function. Site-directed mutagenesis followed by phenotype observation could be an efficient way for association analysis between these conserved amino acids in Table 3 and TYRP1 function, while in the presence of enough color statistics, screening of SNP at these conserved sites within or among populations with more color diversity might be another good method.

Table 3 Mutated and conserved amino acids

species	mutated amino acids	conserved amino acids
Felis catus	Ala3Gly[1], Ala31Val, Met155Ile, Asn459Lys*, Arg100Stop[2], 421ins17/18[1]	C(30,65,99,113,122,261,290), D(54,71,234,332), E(140,428,477), F(105,220,244,400,411), G(39,107,116,119,202,254,267,367), I(311), L(36), N(449), P(172,330,426,365), Q(124,236), S(281,278,284,349,351), T(154,407), Y(79,185,369,464,522), CG(56-57,110-111), CN(101-102,303-304) PH(74-75), DG(82-83), WP(90-91), AL(147-148), AK(151-152), HP(156-157), VI(160-161), EN(180-181), HY(192-193), SV(195-196), KT(198-199), GQ(204-205), HL(227-228), LE(231-132), ML(238-239), CD(258-259), RS(269-270), FD(272-273), NS(280-281), RP(321-322), CL(336-337), TG(387-388), AP(456-457), WP(468-469), VA(490-491), PL(513-514), CCP(41-43), GRG(61-63), RRN(130-132), TRR(163-165), FVW(188-190), DDL(263-265), LGT(299-301), LHT(403-405), DDRE(85-88), RFFN(93-96), TWHR(222-225), NLGY(459-462), EGYS(360-363), AQFPR(24-28), VQRLP(324-328), FDTPP(342-346), SFRNT(354-358), TGHNR(432-436), FSQWRV(283-288), QTHLSP(390-395), DFSHEGP(212-218), LPYWNFA(246-252), RNPAGNV(313-319), NMVPFWPP(439-446), EMFVT/SAPD/ENLGY(451-462), RSLHNLAHLFLN(374-385)
Homo sapiens	Val48Ala, (Ser166Stop[3]), Gly254Arg*, 362delTyr*, Ser363Asn*, (codon368delA[3]), an 1575insAGAA induced frameshift mutation, a transcriptional variant (see in length variation)	
Mus musculus	(Cys86Tyr[4], Arg302Cys[4]),Thr352Ile*, Glu451Gly*	
Ovis aries	Ala68Val, Arg100His, Asn102Ser*, Cys290Phe[5], Ile447Val	
Bos taurus	Phe94Ser*, Phe342Tyr*, Asn531Ser	
Canis familiaris	alternative splicing, (Ser41Cys[6], Gln331Stop[6], 345delPro[6])	
Coturnix japonica	Val21Ala, Phe282Ser[7]	
Danio rerio	227-229delAsp-Met-Leu*	
Sus scrofa	—	
Equus caballus	—	
Xenopus laevis	—	
Rattus norvegicus	—	
Pan troglodytes	—	
Macaca mulatta	—	
Capra hircus	—	
Gallus gallus	—	
Xenopus tropicalis	—	
Takifugu rubripes	—	

Note: The variations with superscript 1 and 2 are associated with chocolate and cinnamon phenotype in domestic cats, respectively (Lyons et al., 2005; Sun et al., 2005). The superscript 3–7 represent mutations that are associated with rufous oculocutaneous albinism (ROCA) (Shibahara et al., 1992), light color in mice (Gratten et al., 2007), light color in soay sheep (Schmutz et al., 2002), brown (including brown and white) phenotype in dogs (Nadeau et al., 2007) and plumage phenotype in Japanese quail (Nishimura and Balch, 1997) respectively. The symbol* represents mutations that happened at conserved positions. The capital letters underlined represent nucleotides and the mutations in round brackets were not found in our paper. The positions of conserved amino acids were set according to CAM22450, as a reference sequence when aligning.

3.6 Motif analysis of amino acid sequence

38 motifs as potential function sites were found in sequence CAM22450 of 537 amino acids, including ten casein kinase II phosphorylation sites, nine protein kinase C phosphorylation sites, eight N-glycosylation sites, five N-myristoylation sites, one cAMP- and cGMP-dependent protein kinase phosphorylation site, one Tyrosine kinase phosphorylation site, one EGF-like domain, one laminin-type EGF-like (LE) domain, one tyrosinase CuA-binding region, and one tyrosinase CuB-binding region, of which five N-glycosylation sites (96–99, 181–184, 304–307, 385–388, and 350–353), three casein kinase II phosphorylation sites (270–273, 405–408, and 455–458), two N-myristoylation sites (267–272 and 300–305), two protein kinase C phosphorylation sites (163–165 and 354–356), the EGF-like domain (99–110), the laminin-type EGF-like (LE) domain (99–122), the tyrosinase CuA-binding region (215–232), and the tyrosinase CuB-binding region (397–408) were highly conserved among the 55 protein sequences and hence probably true positive results.

It was reported that the five N-glycosylation sites were directly involved in sorting, stability and activity of TYRP1 (Jackson, 1994). Also, it was highly predicted that the EGF-like domain may be involved in forming multi-enzyme complex within the tyrosinase-related protein family along with other proteins, and TYRP1 was a copper-containing enzyme with two copper-binding sites (three conserved histidines residues) (Jackman et al., 1991; Xu et al., 2001). The other motifs referred above need further investigation.

3.7 Hydrophobicity profile and SMART analysis

Hydrophobicity analysis of sequence CAM22450 showed that TYRP1 possessed two hydrophobic regions (1–20, 480–500). Based on the available knowledge that mature TYRP1 is a 75 kDa type-1 transmembrane glycoprotein synthesized mainly on ER and then transported to premelanosome via Golgi (Jackman et al., 1991; Chen et al., 2001), we conclude that the N-terminal hydrophobilic peak represents the signal peptide that directs the working ribosome onto ER and the C-terminal one represents the transmembrane domain, which is consistent with the SMART analysis result that showed three functional segments (1–24: signal peptide, 121–471: Pfam.Tyrosinase, 479–501: transmembrane domain). In addition, the results showed that general TYRP1 sequences had a molecular weight of about 58 kDa minus the N-terminal signal sequence of about 20 amino acids, inferring that the modifying components like glycosyl- groups made up the other 17 kDa. It is worthy to note that except for one N-glycosylation site (96–99) implicated in transport and stability of TYRP1 and the EGF-like domain involved in the formation of the multienzyme complex, all the other conserved motifs referred above exist in segment 121–471, which implies their possible important roles in TYRP1 or in the tyrosinase-related protein family.

4 Conclusions

Of the two evolutionally conserved non-coding regions detected in this paper, the first one (from −1306 to −733 according to AL138753) very probably represents the pivotal cis-regulatory constituent in *TYRP1* transcription, while the same attention should be given to the other region (−642 to −515). The relevance between stop codon variation and color variety was very subtle and thus needs to be combined with CDS length variation or SNP for more precise color genetic correlation analysis. CDS length variation and genetic diversity analysis showed that *Felis catus*, *Homo sapiens* and *Canis familiaris* had more genetic diversity than other species for *TYRP1*, and that *Felis catus* could be used as a better choice of target to study the *TYRP1*-associated genetic basis underlying the color diversity. Both amino acid variations within species with genetic effects and the conserved amino acids or motifs among species can be good stuff for function association analysis. The site-directed mutagenesis techniques and SNP screening in specific species may give more confirmation. As a 75 kDa type-1 transmembrane glycoprotein, mature TYRP1 possesses modifying components of about 17 kDa, whose function predominantly depends on the existing glycosyl- groups and the Cu components.

Acknowledgements This work was supported by the Fund for the Doctoral Program of Higher Education (RFDP) of China (No. 20070086003).

References

Alonso S, Izagirre N, Smith-Zubiaga I, Gardeazabal J, Díaz-Ramén J L, Díaz-Pórez J L, Zelenika D, Boyano M D, Smit N, de la Rúa C (2008). Complex signatures of selection for the melanogenic loci TYR, TYRP1 and DCT in humans. BMC Evol Biol, 8: 74

Bertolotto C, Buscà R, Abbe P, Bille K, Aberdam E, Ortonne J P, Ballotti R (1998). Different cis-acting elements are involved in the regulation of TRP1 and TRP2 promoter activities by cyclic AMP: pivotal role of M boxes (GTCATGTGCT) and of microphthalmia. Mol Cell Biol, 18(2): 694–702

Chen H, Salopek T G, Jimbow K (2001). The role of phosphoinositide 3-kinase in the sorting and transport of newly synthesized tyrosinase-related protein-1 (TRP-1). J Investig Dermatol Symp Proc, 6(1): 105–114

Elnitski L, Hardison R C, Li J, Yang S, Kolbe D, Eswara P, O'Connor M J, Schwartz S, Miller W, Chiaromonte F (2003). Distinguishing regulatory DNA from neutral sites. Genome Res, 13(1): 64–72

Gratten J, Beraldi D, Lowder B V, McRae A F, Visscher P M, Pemberton J M, Slate J (2007). Compelling evidence that a single nucleotide substitution in TYRP1 is responsible for coat-colour polymorphism

in a free-living population of Soay sheep. Proc Biol Sci, 274(1610): 619–626

Hardison R C (2000). Conserved noncoding sequences are reliable guides to regulatory elements. Trends Genet, 16: 369–372

Hearing V J (1987). Mammalian monophenol monooxygenase (tyrosinase): purification, properties and reactions catalyzed. Methods Enzymol, 142: 154–165

Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche B A, de Castro E, Lachaize C, Langendijk-Genevaux P S, Sigrist C J (2008). The 20 years of PROSITE. Nucleic Acids Res, 36(Database issue), D245–249

Ito S (2003). A chemist's view of melanogenesis. Pigment Cell Res, 16: 230–236

Jackman M P, Hajnal A, Lerch K (1991). Albino mutants of *Streptomyces glaucescens* tyrosinase. Biochem J, 274(Pt 3): 707–713

Jackson I J (1994). Molecular and developmental genetics of mouse coat color. Annu Rev Genet, 28: 189–217

Jiménez-Cervantes C, Solano F, Kobayashi T, Urabe K, Hearing V J, Lozano J A, García-Borrón J C (1994). A new enzymatic function in the melanogenic pathway. J Biol Chem, 269(27): 17993–18001

Kobayashi T, Hearing V J (2007). Direct interaction of tyrosinase with Tyrp1 to form heterodimeric complexes *in vivo*. J Cell Sci, 120: 4261–4268

Kobayashi T, Urabe K, Winder A, Jiménez-Cervantes C, Imokawa G, Brewington T, Solano F, García-Borrón J C, Hearing V J (1994). Tyrosinase related protein 1 (TRP1) functions as a DHICA oxidase in melanin biosynthesis. EMBO J, 13(24): 5818–5825

Kobayashi T, Vieira W D, Potterf B, Sakai C, Imokawa G, Hearing V J (1995). Modulation of melanogenic protein expression during the switch from eu-topheomelanogenesis. J Cell Sci, 108: 2301–2309

Levy C, Khaled M, Fisher D E (2006). MITF: master regulator of melanocyte development and melanoma oncogene. Trends Mol Med, 12(9): 406–414

Lowings P, Yavuzer U, Goding C R (1992). Positive and negative elements regulate a melanocyte-specific promoter OCT-1. Mol Cell Biol, 12(8): 3653–3662

Lyons L A, Foe I T, Rah H C, Grahn R A (2005). Chocolate coated cats: TYRP1 mutations for brown color in domestic cats. Mamm Genome, 16(5): 356–366

Manga P, Kromberg J G, Box N F, Sturm R A, Jenkins T, Ramsay M (1997). *Rufous oculocutaneous* albinism in southern African Blacks is caused by mutations in the TYRP1 gene. Am J Hum Genet, 61(5): 1095–1101

Martínez-Morales J R, Dolez V, Rodrigo I, Zaccarini R, Leconte L, Bovolenta P, Saule S (2003). OTX2 activates the molecular network underlying retina pigment epithelium differentiation. J Biol Chem, 278(24): 21721–21731

Mathews P M, Martinie J B, Fambrough D M (1992). The pathway and targeting signal for delivery of the integral membrane glycoprotein LEP100 to lysosomes. J Cell Biol, 118(5): 1027–1040

Nadeau N J, Mundy N I, Gourichon D, Minvielle F (2007). Association of a single-nucleotide substitution in TYRP1 with roux in Japanese quail (*Coturnix japonica*). Anim Genet, 38(6): 609–613

Nishimura N, Balch W E (1997). A di-acidic signal required for selective export from the endoplasmic reticulum. Science, 277(5325): 556–558

Ovcharenko I, Loots G G, Hardison R C, Miller W, Stubbs L (2004). zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. Genome Res, 14: 472–477

Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel C V, Hau J, Martin O, Kuznetsov D, Falquet L (2007). MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res, 35(Web Server issue): W433–437

Pennacchio L A, Rubin E M (2001). Genomic strategies to identify mammalian regulatory sequences. Nat Rev Genet, 2: 100–109

Prota G (1988). Some new aspects of eumelanin chemistry. Prog Clin Biol Res, 256: 101–124

Rad H H, Yamashita T, Jin H Y, Hirosaki K, Wakamatsu K, Ito S, Jimbow K (2004). Tyrosinase-related proteins suppress tyrosinase-mediated cell death of melanocytes and melanoma cells. Exp Cell Res, 298(2): 317–328

Sarangarajan R, Boissy R E (2001). Tyrp1 and oculocutaneous albinism type 3. Pigment Cell Res, 14(6): 437–444

Schmidt-Küntzel A, Eizirik E, O'Brien S J, Menotti-Raymond M (2005). Tyrosinase and tyrosinase related protein 1 alleles specify domestic cat coat color phenotypes of the albino and brown loci. J Hered, 96 (4): 289–301

Schmutz S M, Berryere T G, Goldfinch A D (2002). TYRP1 and MC1R genotypes and their effects on coat color in dogs. Mamm Genome, 13 (7): 380–387

Shibahara S, Tomita Y, Sakakura T, Nager C, Chaudhuri B, Müller R (1986). Cloning and expression of cDNA encoding mouse tyrosinase. Nucleic Acids Res, 14(6): 2413–2427

Shibahara S, Tomita Y, Yoshizawa M, Shibata K, Tagami H (1992). Identification of mutations in the pigment cell-specific gene located at the brown locus in mouse. Pigment Cell Res, Suppl 2: 90–95

Sun J, Chen M, Xu J, Luo J (2005). Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. J Mol Evol, 61(4): 437–444

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics, 105: 437–460

Tate W P, Mansell J B, Mannering S A, Irvine J H, Major L L, Wilson D N (1999). UGA: a dual signal for 'stop' and for recoding in protein synthesis. Biochemistry (Mosc), 64(12): 1342–1353

Vijayasaradhi S, Xu Y, Bouchard B, Houghton A N (1995). Intracellular sorting and targeting of melanosomal membrane proteins: identification of signals for sorting of the human brown locus protein, gp75. J Cell Biol, 130(4): 807–820

Williams M A, Fukuda M (1990). Accumulation of membrane glycoproteins in lysosomes requires a tyrosine residue at a particular position in the cytoplasmic tail. J Cell Biol, 111(3): 955–966

Xu Y, Bartido S, Setaluri V, Qin J, Yang G, Houghton A N (2001). Diverse roles of conserved asparagine-linked glycan sites on tyrosinase family glycoproteins. Exp Cell Res, 267(1): 115–125

Xu Y, Vijayasaradhi S, Houghton A N (1998). The cytoplasmic tail of the mouse brown locus product determines intracellular stability and export from the endoplasmic reticulum. J Invest Dermatol, 110(4): 324–331

Yavuzer U, Goding C R (1994). Melanocyte-specific gene expression: Role of repression and identification of a melanocyte-specific factor, MSF. Mol Cell Biol, 14(5): 3494–3503