doi: 10.1111/j.1469-1809.2006.00261.x

Distribution of the F374 Allele of the *SLC45A2* (*MATP*) Gene and Founder-Haplotype Analysis

I. Yuasa[1,*], K. Umetsu[2], S. Harihara[3], A. Kido[4], A. Miyoshi[5], N. Saitou[6], B. Dashnyam[7], F. Jin[8], G. Lucotte[9], P. K. Chattopadhyay[10], L. Henke[11] and J. Henke[11]

[1]*Division of Legal Medicine, Faculty of Medicine, Tottori University, Yonago, Japan*
[2]*Department of Experimental and Forensic Pathology, Faculty of Medicine, Yamagata University, Yamagata, Japan*
[3]*Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan*
[4]*Department of Legal Medicine, Faculty of Medicine, University of Yamanashi, Nakakoma, Japan*
[5]*Department of Forensic Medicine, Fukuoka University School of Medicine, Fukuoka, Japan*
[6]*Division of Population Genetics, National Institute of Genetics, Mishima, Japan*
[7]*Institute of Biological Sciences, Mongolian Academy of Sciences, Ulaan Baator, Mongolia*
[8]*Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China*
[9]*Center of Molecular Neurogenetics, Paris, France*
[10]*Amity Institute of Forensic Sciences, Defence Colony, New Delhi, India*
[11]*Institut für Blutgruppenforschung, Cologne, Germany*

Summary

The membrane-associated transporter protein (MATP) plays an important role in melanin synthesis. The L374F mutation in the *SLC45A2* gene encoding MATP has been suggested to be associated with skin colour in major human populations. In this study more detailed distribution of the F374 allele was investigated in 1649 unrelated subjects from 13 Eurasian populations and one African population. The highest allele frequency was observed in Germans (0.965); French and Italians showed somewhat lower frequencies; and Turks had an intermediate value (0.615). Indians and Bangladeshis from South Asia were characterized by low frequencies (0.147 and 0.059, respectively). We also found the F374 allele in some East and Southeast Asian populations, and explained this by admixture. Haplotype analysis revealed that the haplotype diversity was much lower in Germans than in Japanese, and suggest that the L374F mutation occurred only once in the ancestry of Caucasians. The large differences in distribution of the F374 allele and its haplotypes suggest that this allele may be an important factor in hypopigmentation in Caucasian populations.

Keywords: haplotype, linkage disequilibrium, membrane-associated transporter protein, polymorphism, population study, SLC45A2

Introduction

Variation in skin colour is one of the most conspicuous and polymorphic traits of humans. Skin colour is primarily due to a pigment called melanin, located in the epidermis, and is lighter in more northerly latitudes. The clinal distribution of skin colour observed among indigenous peoples is correlated with ultraviolet radiation levels. This variation in skin colour has been suggested to reflect biological adaptation to such environments and is affected by natural selection. However, this view has not been universally accepted (Jablonski & Chaplin, 2000; Relethford, 2002; Jobling *et al*. 2004; Diamond, 2005).

Currently 127 loci associated with mouse coat colour have been identified, and 59 of the actual genes have been cloned and sequenced. Of these the mouse

*Address for correspondence: Isao Yuasa, Division of Legal Medicine, Faculty of Medicine, Tottori University, Yonago 683-8503, Japan. Tel. +81-859-38-6123. Fax +81-859-38-6120. E-mail: yuasai@grape.med.tottori-u.ac.jp

underwhite (*uw*) locus is a major determinant of pigmentation, and mutations at the *uw* locus also lead to hypopigmentation (Lehman *et al*. 2000; Bennett & Lamoreux, 2003). The *SLC45A2* gene, originally called the "antigen in melanoma (AIM-1)" gene, is expressed in most human melanoma cell lines and melanocytes, but not in normal tissue, and has been cloned in medaka (a small fresh-water teleost), mouse and human. Mutations in this gene reduce the melanin content in medaka. "AIM-1" was assigned to the *uw* locus and the encoded protein referred to as membrane-associated transporter protein (MATP). Thus, MATP takes part in melanin synthesis in melanosomes and is suspected to function as a membrane-transporter by directing the traffic of melanosomal proteins to the melanosome (Fukamachi *et al*. 2001; Harada *et al*. 2001; Newton *et al*. 2001; Kushimoto *et al*. 2003).

The human *SLC45A2* gene (Entrez Gene ID: 51151; GenBank accession numbers: NM_016180 and NT_006576) spans 40 kb on chromosome 5p and consists of 7 exons, encoding the 530-amino acid MATP polypeptide. The human *SLC45A2* gene underlies oculocutaneous albinism type 4 (OCA4), for which several mutations have been identified in Germans, Japanese, and an individual of Turkish descent, (Newton *et al*. 2001; Inagaki *et al*. 2004; Rundshagen *et al*. 2004). Only a few polymorphisms have been identified in exons of this gene. Of these the L374F mutation, resulting from a G-to-C transversion in exon 5, has been found to have a unique distribution: the F374 allele was observed at high frequencies of 0.89 and 0.96 in white South Africans and Germans, respectively, while it has not been observed in Ghanaians, Japanese and New Guinea Islanders. The F374 allele is likely to be associated with skin pigmentation in the major human populations. The striking difference in distribution of the F374 allele may be a consequence of natural selection, i.e. adaptation to lower amounts of ultraviolet radiation (Nakayama *et al*. 2002; Yuasa *et al*. 2004). However, knowledge of the allele frequency and haplotype diversity in the gene encoding MATP is very limited. Elucidation of its genetic polymorphisms and haplotypes would serve as a starting point for molecular approaches to its anthropological genetics. In this study, we investigated i) the distribution of the F374 allele in several populations, ii) 13 additional SNPs in *SLC45A2* and its immediately adjacent genes, *AMACR* (alpha-methylacyl-CoA racemase; Gene ID 23600) and *RLN3R1* (relaxin 3 receptor 1, also known as SALPR; Gene ID 51289) in Africans, Japanese and Germans, and iii) founder haplotype(s) with the F374 allele in a German population.

Materials and Methods

DNA Samples

DNA samples were obtained from a total of 1649 unrelated individuals, mostly living in various regions of Eurasia (Fig. 1 & Table 1). The donors were selected at random, irrespective of skin colour, and their phenotypic data were unavailable. German samples were obtained from Germans living in Northrhine-Westphalia. French and Italians samples were from Rheims and Genoa, respectively. Turkish samples were collected from Turks living in West Germany as immigrants, and all subjects were born in various regions of Turkey. The DNA samples from Bangladeshis and Indonesians were described previously (Dobashi *et al*. 2003); Bangladeshis and Indians investigated in this study were Indo-European-speaking. Khalhas and Buryats are major and minor ethnic groups in Mongolia, respectively. Chinese samples came from three Han populations from Shenyang, Wuxi and Huizhou. Japanese from Okinawa were also studied. African samples were obtained from various Sub-Saharan Africans immigrating into Germany or Japan: 10 Africans were from Ghana and the others were from the Congo, Nigeria, Zambia, Ivory Coast, Uganda, Kenya and Zimbabwe. The samples from Munich and Tottori used for haplotype analysis are the same set as used in our previous study (Yuasa *et al*. 2004). The data on average pigmentation for the selected populations have been summarized in the references (Jablonski & Chaplin, 2000; Jobling *et al*. 2004). This study was approved by the Ethical Committee at the Faculty of Medicine, Tottori University.

Population Study of the F374 Allele

DNA samples were typed for the L374F polymorphism by the amplified product length polymorphism (APLP) method as described previously (Yuasa *et al*. 2004).

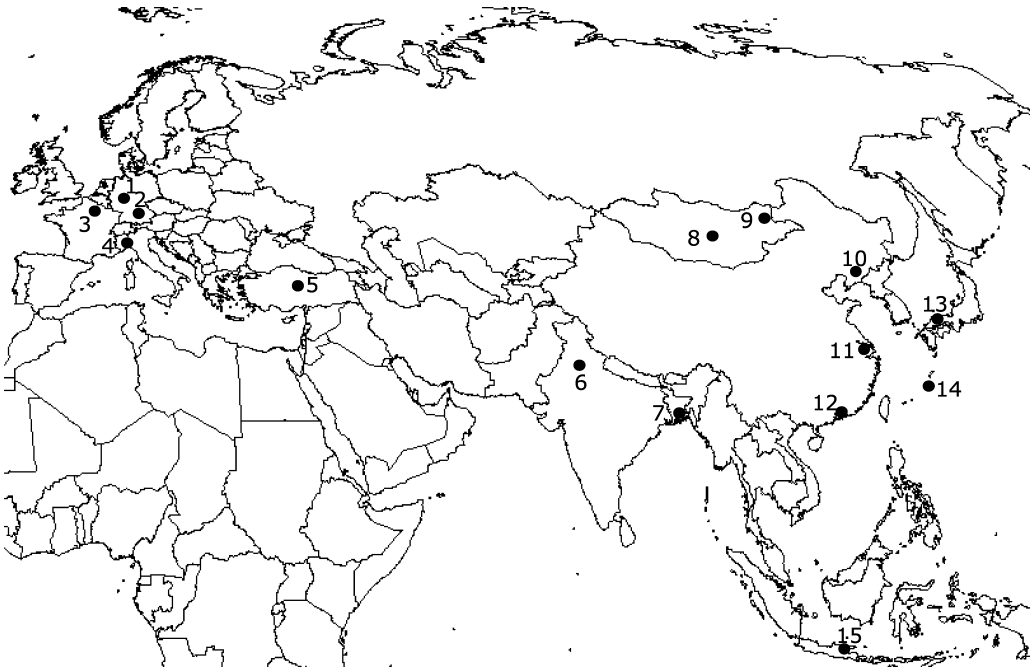

Figure 1 Location of Eurasian populations analyzed in the present study.

Table 1 Distribution of L374F genotypes and the F374 allele in various populations

No	Population	Collection place	n	Genotypes of L374F			Frequency of F374	References
				L	L/F	F		
1	German	Northrhine-Westphalia	241	0	17	224	0.965	This study
2	German	Munich	93	0	7	86	0.962	Yuasa *et al.* (2004)
3	French	Rheims	98	1	19	78	0.893	This study
4	Italian	Genoa	97	5	19	73	0.851	This study
5	Turk	West Germany	200	41	72	87	0.615	This study
6	Indian	New Delhi	51	37	13	1	0.147	This study
7	Bangladeshi	Dhaka	118	104	14	0	0.059	This study
8	Khalha	Ulaan Baator	173	135	37	1	0.113	This study
9	Buryat	Dashbalbar	143	111	31	1	0.115	This study
10	Han	Shenyang	89	84	5	0	0.028	This study
11	Han	Wuxi	119	119	0	0	0.000	This study
12	Han	Huizhou	111	110	1	0	0.005	This study
13	Japanese	Tottori	103	103	0	0	0.000	Yuasa *et al.* (2004)
14	Japanese	Okinawa	87	87	0	0	0.000	This study
15	Indonesian	Surabaya	105	104	1	0	0.005	This study
16	African	Germany/Japan	17	17	0	0	0.000	This study
	White South African		54				0.89	Nakayama *et al.* (2002)
	Ghanaian		50				0.00	Nakayama *et al.* (2002)
	New Guinea Islander		52				0.00	Nakayama *et al.* (2002)
	Japanese		49				0.00	Nakayama *et al.* (2002)

Genotyping of SNPs

Many SNPs at the loci *AMACR*, *SLC45A2* and *RLN3R1* have been reported in the GenBank SNP database. In this study several SNPs which could be genotyped with restriction enzymes were arbitrarily selected (Fig. 2). DNA from 17 Africans, 103 Japanese from Tottori and 92 Germans from Munich were subjected to polymerase chain reaction (PCR) followed by restriction fragment length polymorphism (RFLP)

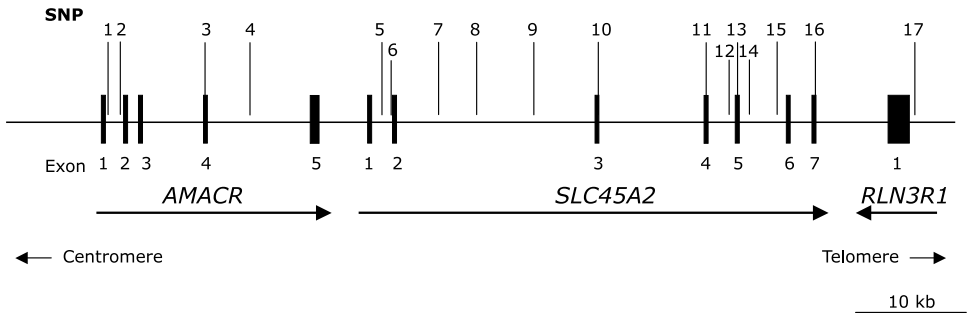

Figure 2 Schematic diagram of the *AMACR*, *SLC45A2* and *RLN3R1* genes, showing locations of exons and 17 SNPs. Arrows below each gene name show the direction of transcription.

Table 2 Oligonucleotide sequence for SNP genotyping in human *AMACR*, *SLC45A2* and *RLN3R1*

SNP	Marker Position on NT_006576	Primer	Sequence (5′→3′) enzyme	Restriction
1	16459765	AMACR-Y1F1	CTTGGGTTCTGAGATACTGCTGTT	*Msp*I
		AMACR-Y1R1	CAGGAAGGGAATCCTATGGCTTT	
2	16458616	AMACR-Y1F2	TGGTTACCTGGGGTTTGTGCT	*Rsa*I
		AMACR-Y1R2	GCCTAGCATCAGTACCTTCACGA	
3	16450989	AMACR-E4F2	ATGCTTTGACACTTGAGTTATCTGG	*Tsp*509I
		AMACR-E4R2	CATCTGCTGTCCTGTAAGTCGT	
4	16447049	AMACR-Y4F1	CATATCCACACTTCACCTCA	*Taq*I
		AMACR-Y4R1	AACTCAGCAGAAGAATACCC	
5	16435295	MATP-Y1F2	GTTCAAAGATGTGGCCTCTGAC	*Msp*I
		MATP-Y1R2	GTAGTCAAACGCAGTGCTTCTC	
6	16434674	MATP-Y1F	TTGTGTTTTCTGGACATTCGCT	*Bsh*1236I
		MATP-Y1R	CTTCCTCCTTGGGTTAGCAATC	
7	16430294	MATP-Y2F6	TATGTCCAACACCTCCCTTCTC	*Eco*130I
		MATP-Y2R6	TCTAAGGCCTCCAACTGACTG	
8	16426848	MATP-Y2F2	ACTGTGGTATCCTCACTCTG	*Hsp*92II
		MATP-Y2R2	TGCTTCAAGTGGAATGCTG	
9	16422112	MATP-Y2F3	AAACGCTTCACGGTGTGCTGA	*Vsp*I
		MATP-Y2R3	ATTTCCTGGTCAGACCCCATGGA	
12	16404484	MATP-Y4F	AGCAATCATGGCAGGCTTCA	*Hsp*92II
		MATP-Y4R	AGTCATAGGTCACCGCATCCT	
14	16402809	MATP-Y5F	ACCATTCTCAGGAGGACTCA	*Mph*1103I
		MATP-Y5R	GTGCTATGGCTAGGTAAAGGG	
15	16400425	MATP-Y5F2	AGCCAAGTTGACCTGCTAGA	*Msp*I
		MATP-Y5R2	TGAAATCACCCAGCTTAATGCC	
17	16388182	SALPR-F	ACGTCAAAGCCGACTTTCTCC	*Pvu*II
		SALPR-R	GAGAGTTGAGCTGCAGGCAA	

The primers for the SNPs 10, 11, 13 and 16 were described previously (Yuasa *et al*. 2004).

analysis. In the African samples prior to PCR the whole genome was amplified using REPLI-g (Qiagen, Hilden, Germany) because of their small DNA volumes. Primers (Table 2) for the specific amplification of fragments encompassing mutation sites were designed on the basis of the genomic sequence (NT_006576). PCR was performed in a volume of 12 μl containing 20 ng genomic DNA, 2.5 pmol of each primer, 200 μM of each dNTP and 0.3 U HotStarTaq polymerase (Qiagen). Cycle conditions were 95°C for 15 min, then 30-35 cycles of 94°C for 30 sec, 55-56°C for 30 sec, 72°C for 40 sec, and a final extension step of 10 min at 72°C, in a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). Products were digested with restriction enzymes (Table 2). The digested fragments were separated by 6% or 7.5% polyacrylamide gel electrophoresis and then visualized by ethidium bromide staining.

Statistical Analysis

Allele frequencies were estimated by direct counting, and conformity to Hardy-Weinberg equilibrium was tested using the χ^2 test with 1 df. Wright's F_{ST} statistic was used to estimate the proportion of variation attributable to differences in the L374F mutation among 17 populations, and in each SNP among the African, Japanese and German populations (Hartl & Clark, 1997). Linkage disequilibrium (LD) between pairs of polymorphic sites was measured using the statistics D' (Lewontin, 1964) and r^2 (Hill & Robertson, 1968). Haplotype diversity and frequency for multiple loci were estimated by the expectation-maximization (EM) method using the Alrequin program (Schneider *et al*. 2000).

Results

Distribution of F374 Allele in *SLC45A2*

Table 1 summarizes the distribution of L374F genotypes and frequencies of the F374 allele in a total of 2050 unrelated individuals from 20 populations, including previously reported data (Nakayama *et al*. 2002; Yuasa *et al*. 2004). A significant deviation from Hardy-Weinberg equilibrium was observed in the Italian and Turkish populations due to an excess of homozygotes and a deficiency of heterozygotes. Although the reasons for this are not known, this deviation may be affected by assortative mating and/or Wahlund's principle (Hartl & Clark, 1997). The results showed a marked difference in the distribution of the F374 allele among various populations. Germans had the highest estimated frequency. French and Italians showed a significantly lower frequency than Germans. Turks were characterized by an intermediate value. Indians and Bangladeshis showed a much lower frequency than Europeans and Turks. Khalhas and Buryats from Mongolia revealed similar values to each other, and were somewhat higher than Bangladeshis. In contrast, the F374 allele was much less common in Han Chinese and Indonesian populations, and completely absent in Africans and Japanese (Nakayama *et al*. 2002). F_{ST} was estimated to be 0.74 among the 20 populations (see Table 1).

SNP Genotyping

A total of 17 SNPs, which are scattered throughout an 80 kb genomic region containing the complete *AMACR*, *SLC45A2* and *RLN3R1* genes (Fig. 2), were investigated in three populations. Table 3 shows the frequencies of mutant alleles, which were defined in this study as minor alleles found in the African and then Japanese populations. Striking differences were observed in allele frequency among the three populations. The German population was characterized by the lowest average heterozygosity for 12 SNPs in the *SLC45A2* gene. Interestingly, the heterozygosity of 4 out of 5 SNPs in the two adjacent genes, *AMACR* and *RLN3R1*, was rather higher in Germans than in Japanese. Of all the 17 SNPs the L374F mutation revealed the highest F_{ST} value (0.94). The SNPs near the L374F mutation site also revealed relatively high values in comparison with the SNPs in the 5′-flanking region of the *SLC45A2* gene. Low F_{ST} values were generally observed in the two adjacent genes.

Pairwise LD in MATP

Pairwise LD values in the Japanese and German populations were evaluated between major SNPs in *SLC45A2* by D' and r^2. Fig. 3 illustrates substantial differences between D' and r^2. The D' values were generally much higher than the r^2 values, as D' values equal to 1.0 are caused by the presence of only three out of four possible haplotypes for a pair of loci. Although rare alleles with frequencies <5% do not have sufficient statistical power for LD detection (Lewontin, 1995; Goddard *et al*. 2000), the pairwise values between SNPs 10–15 were relatively high in both populations, suggesting that this 15-kb region forms a haplotype block.

Haplotype Analysis

Haplotypes were constructed on the basis of the genotype data from 12 SNPs in *SLC45A2*. Table 4 lists the 32 haplotypes and their frequencies, as estimated by the EM algorithm, with phase-unknown samples. This procedure has been shown to estimate common haplotype frequencies accurately when the Hardy-Weinberg assumption is fulfilled (Fallin & Schork, 2000; Tishkoff

Table 3 Frequency, heterozygosity and F_{ST} of SNPs in Africans, Japanese (Tottori) and Germans (Munich)

SNP No	SNP position on NT_006576	GenBank SNP number	Region	Mutation	Frequency of mutant*			Heterozygosity			F_{ST}
					African n = 17	Japanese n = 103	German n = 92	African n = 17	Japanese n = 103	German n = 92	
AMACR											
1	16459765	rs34688	intron 1	c > t	0.5000	0.0874	0.1196	0.515	0.160	0.212	0.195
2	16458616	rs34687	intron 1	g > a	0.4706	0.0874	0.1196	0.513	0.160	0.212	0.172
3	16450989	rs2287939	exon 4	C > T (S201L)	0.1176	0.1408	0.2989	0.214	0.243	0.421	0.043
4	16447049	rs840409	intron 4	c > g	0.0000	0.0874	0.0761	0.000	0.160	0.141	0.029
	Average heterozygosity							0.311	0.181	0.247	
SLC45A2											
5	16435295	rs732740	intron 1	t > c	0.0000	0.0000	0.0054	0.000	0.000	0.011	0.004
6	16434674	rs250413	intron 1	c > t	0.0000	0.1650	0.0109	0.000	0.277	0.022	0.103
7	16430294	rs181832	intron 2	t > c	0.3235	0.1650	0.0217	0.451	0.277	0.043	0.108
8	16426848	rs3776549	intron 2	g > a	0.2647	0.2087	0.0489	0.401	0.332	0.094	0.058
9	16422112	rs3756462	intron 2	t > c	0.0000	0.2087	0.0489	0.000	0.332	0.094	0.101
10	16415976	rs26722	exon 3	G > A (E272K)	0.0588	0.3835	0.0326	0.110	0.475	0.063	0.191
11	16406617	rs2287949	exon 4	G > A (T329T)	0.0882	0.2379	0.0054	0.166	0.364	0.011	0.094
12	16404484	rs250417	intron 4	g > c	0.2059	0.6408	0.0326	0.337	0.463	0.063	0.316
13	16403799	rs16891982	exon 5	G > C (L374F)	0.0000	0.0000	0.9620	0.000	0.000	0.074	0.944
14	16402809	rs40132	intron 5	t > c	0.0294	0.4078	0.0272	0.059	0.485	0.053	0.245
15	16400425	rs35394	intron 5	a > g	0.0294	0.4029	0.0272	0.059	0.483	0.053	0.240
16	16396933	rs3733808	exon 7	G > C (V507L)	0.0000	0.0049	0.0000	0.000	0.010	0.000	0.003
	Average heterozygosity							0.132	0.292	0.048	
RLN3R1											
17	16388182	rs42868	5′-flanking	c > g	0.3529	0.0437	0.1848	0.471	0.084	0.303	0.102
Average heterozygosity in three loci								0.194	0.253	0.110	

*Mutant alleles are defined as minor alleles found in Africans in this study.

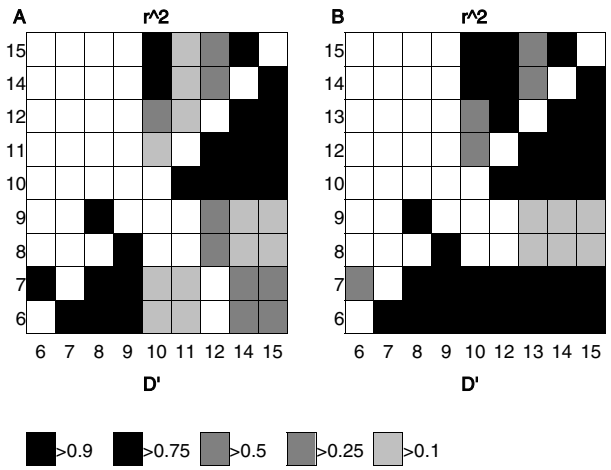

Figure 3 Pairwise LD analysis of the *SLC45A2* gene region in Japanese (A) and Germans (B), evaluated by r^2 and D' (above and below the diagonal, respectively). Number shows the SNPs.

et al. 2000). Although SNPs 8 and 9 investigated in the Japanese samples showed a slight deviation from Hardy-Weinberg equilibrium ($\chi^2 = 4.33; 0.05 > p > 0.02$), data from these SNPs were included in the estimation of haplotypes. A total of 32 haplotypes were observed and classified into 4 groups based on SNPs 10-15. In the Japanese population 17 haplotypes were observed belonging to the 3 groups, and all major haplotypes in each group (hp1, hp15 and hp25) shared a nucleotide sequence of TCTGT for SNPs 5-9. Group 2, consisting of the F374-containing haplotypes, accounted for most of the haplotypes in Germans. The L374-containing haplotypes found in Germans were a subset of those in the Japanese. These findings indicate less allelic complexity in Germans than in Japanese. Haplotype diversity was estimated as 0.815 ± 0.051 for Africans,

			Frequency		
Group	Haplotype	Sequence of the SNPs 5-16	African	Japanese	German
1	hp1	TCTGTGGGGTAG	0.449	0.245	
	hp2	TCCGTGGGGTAG	0.116		
	hp3	TCCATGGGGTAG	0.109		
	hp4	TCTATGGGGTAG	0.085		
	hp5	TCCGTAGGGTAG	0.034		
	hp6	TTCGTGGGGTAG		0.056	
	hp7	TCTACGGGGTAG		0.053	0.005
	hp8	TCTGTGAGGTAG		0.005	
2	hp9	TCTGTGGGCTAG			0.896
	hp10	TCTACGGGCTAG			0.039
	hp11	TTCGTGGGCTAG			0.011
	hp12	TCCGTGGGCTAG			0.005
	hp13	CCCGTGGGCTAG			0.005
	hp14	TCTGTAGGCTAG			0.005
3	hp15	TCTGTGACGTAG	0.048	0.102	0.005
	hp16	TCTATGACGTAG	0.031		
	hp17	TCCATGACGTAG	0.010		
	hp18	TCTACGACGTAG		0.059	
	hp19	TTCGTGACGTAG		0.055	
	hp20	TCTGTGACGTAC		0.005	
	hp21	TTCACGACGTAG		0.003	
	hp22	TTCGTAACGTAG		0.010	
	hp23	TCTGTGGCGTAG	0.064		
	hp24	TCCGTAGCGTAG	0.025		
4	hp25	TCTGTAGCGCGG		0.256	0.023
	hp26	TCTACAGCGCGG		0.078	0.004
	hp27	TTCGTAGCGCGG		0.039	
	hp28	TCCATGGCGCGG	0.029		
	hp29	TCTGTGGCGCGG		0.014	
	hp30	TCTACGGCGCGG		0.013	
	hp31	TCTGTGGCGCAG		0.005	
	hp32	TTCACGGCGCGG		0.003	
Total			1.000	1.000	1.000

Table 4 Haplotypes and their frequencies in African, Japanese and German

0.881 ± 0.017 for Japanese and 0.261 ± 0.043 for Germans, and the mean number of pairwise differences was 1.586 ± 0.090 for Africans, 3.50 ± 1.79 for Japanese and 0.58 ± 0.47 for Germans. One major haplotype (hp9) differed by one nucleotide in SNP 13 from hp1; hp9 must have arisen following a G-to-C transversion in hp1. Minor alleles in each group occurred due to recombination events, as revealed by the presence of the same sequences in SNPs 5-9. In the German populations recombination between both F374-containing and L374-containing haplotypes must have occurred prior to the disappearance of the L374-carrying haplotypes.

Discussion

In this study the distribution of the F374 allele of the human *SLC45A2* gene, encoding MATP, was confirmed to differ greatly among three major human populations. No other alleles like the F374 allele have been identified from many Caucasian-specific alleles observed to date. The F374 allele shows near-fixation in Germans. White South Africans showed a frequency of 0.89 due to admixture of several European ethnic groups, and contained 7% non-European genes (Nakayama *et al*. 2002). It is possible to estimate the allele frequency in their ancestors (p) using the model of one-way migration (Hartl & Clark, 1997):

$$p' = (1 - m)p + mp^*, \quad (1)$$

where p' is the frequency in white South Africans (0.89), p^* is the frequency in non-Europeans (0), and m is the present proportion of copies of *SLC45A2* that were derived from non-Europeans (0.07). The allele frequency in ancestors of white South Africans (p) was calculated to be 0.957; this value is comparable to that found in Germans, suggesting that northern Europeans, or Germanic-speaking people, share high frequencies of the F374 allele. In comparison with Europeans Turks, Indians and Bangladeshis showed significantly low frequencies of the F374 allele, which gradually declined from Germany to Bangladesh. It seems that these frequencies are associated with latitude, ultraviolet radiation levels and skin colour. The F374 allele was also found in East and Southeast Asia, excluding Japanese populations and a Han-Chinese population from Wuxi. It is likely that there was gene flow from Europe and the Middle East to East Asia along the Silk Road and/or its northern and southern regions. Western Eurasian-specific haplogroups of mitochondrial DNA were observed at a frequency of 14.3% in Mongolians from Xinjiang Province, China (Yao *et al*. 2004). Khalhas from Mongolia and Buryats from Siberia have a fairly high frequency of mitochondrial DNA haplotypes of European origin (Pakendorf *et al*. 2003). Thus, extensive gene admixture has occurred between Asians and Europeans, and we suggest that the F374 allele is substantially specific to so-called Caucasoid populations, including Europeans, west and south Asians and north Africans.

The large differences in distribution of the F374 allele are comparable to that of the FY^*O allele (the null allele at the Duffy blood group locus), with an F_{ST} value of 0.78 found in Sub-Saharan Africans (Cavalli-Sforza *et al*. 1994). Exceptionally high F_{ST} values are a potential indicator of the effects of directional selection (Lewontin & Krakauer, 1973; Bowcock *et al*. 1991). Because MATP plays an important role in melanin synthesis the high frequency of the F374 allele among Europeans may be a response to selection for the lower amount of solar ultraviolet radiation, and associated with depigmentation. The highest frequencies of the F374 allele in northern parts of Europe do not necessarily indicate its origin and spread. European ancestors, after expanding toward the higher latitudes, may have acquired higher frequencies of this allele to adapt to lower levels of solar ultraviolet radiation. Most of the L374-carrying haplotypes were lost during expansion. Haplotype analysis has also revealed a large difference in haplotype diversity between the Japanese and Germans. The F374-bearing haplotypes found in Germans shared the same basic sequence in SNPs 10-15 in a 15-kb region, which formed a haplotype block characterized by low haplotype diversity, strong associations between alleles, and rare recombination. These findings suggest that the L374F mutation occurred only once in the ancestry of the Caucasian group. The hp9 is probably a founder haplotype, constituting 93% of the F374-carrying haplotypes.

Relatively little is understood about the genes responsible for differences in skin, hair and eye colour among individuals within or between populations. Of several human pigmentation-related genes, the *MC1R* (melanocortin 1 receptor) gene also plays very important roles. Some mutations in *MC1R* are associated with red

hair, fair skin and freckles. Low diversity in *MC1R* was observed in Africans, whereas the diversity increases in Europeans and, to a lesser degree, in Asians (Rees, 2004; Sturm & Frudakis, 2004). In contrast, genetic diversity in *SLC45A2* was very low in Europeans in comparison to Africans and Japanese. Because the *SLC45A2* gene is known to be responsible for OCA4, some mutations would be expected to generate hypopigmentation. Although no direct experimental data has been obtained to date on the effect of the F374 allele on the activity of MATP in pigmentation, the large difference in distribution of the F374 allele and its haplotypes suggests that this allele may be an important factor in hypopigmentation in Caucasian populations.

Acknowledgements

We would like to thank the late Prof. Satoshi Horai for providing us with some African DNA samples. This study was supported in part by Grants-in-Aid for Scientific Research (to I.Y. and K.U.) from the Japan Society for the Promotion of Science.

References

Bennett, D. C. & Lamoreux, M. L. (2003) The color loci of mice - a genetic century. *Pigment Cell Res* **16**, 333–344.

Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenutos, L., Kidd, K.K. & Cavalli-Sforza, L.L. (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci USA* **88**, 839–843.

Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1994) *The history and geography of human genes*. Princeton University Press, Princeton, NJ.

Diamond, J. (2005) Geography and skin colour. *Nature* **435**, 283–284.

Dobashi, Y., Kido, A., Fujitani N., Susukida, R. & Oya, M. (2003) Population data of nine STR loci, D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818, D13S317 and D7S820, in Bangladeshis and Indonesians. *Forens Sci Int* **135**, 72–74.

Fallin, D. & Schork, N.J. (2000) Accuracy of haplotype frequency estimation for bialleic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet* **67**, 947–959.

Fukamachi, S., Shimada, A. & Shima, A. (2001) Mutations in the gene encoding B, a novel transporter protein, reduce melanin content in medaka. *Nature Genet* **28**, 381–385.

Goddard, K. A. B., Hopkins, P. J., Hall, J. M. & Witte, J.S. (2000) Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. *Am J Hum Genet* **66**, 216–234.

Harada, M., Li, Y. F., El-Gamil, M., Rosenburg, S. A. & Robbins, P.F. (2001) Use of an *in vitro* immunoselected tumor line to identify shared melanoma antigens recognized by HLA-A* 0201-restricted T cells. *Cancer Res* **61**, 1089–1094.

Hartl, D. L. & Clark, A.G. (1997) *Principles of population genetics*, 3rd editions. Sinauer Associates, Sunderland.

Hill, W. G. & Robertson, A. (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**, 226–231.

Inagaki, K., Suzuki, T., Shimizu, H., Ishii, N., Umezawa, Y., Tada, J., Kikuchi, N., Takata, M., Takamori, K., Kishibe, M., Tanaka, M., Miyamura, Y., Ito, S. & Tomita, Y. (2004) Oculocutaneous albinism type 4 is one of the most common types of albinism in Japan. *Am J Hum Genet* **74**, 466–471.

Jablonski, N. G. & Chaplin, G. (2000) The evolution of skin coloration. *J Hum Evol* **39**, 57–106.

Jobling, M. A., Hurles, M. E. & Tyler-Smith C. (2004) *Human evolutionary genetics-Origins, peoples & disease*. Garland, New York.

Kushimoto, T., Valencia, J. C., Costin, G. E., Toyofuku, K., Watanabe, H., Yasumoto, K., Rouzaud, F., Vieira, W. D. & Hearing, V. J. (2003) The melanosome: an ideal model to study cellular differentiation. *Pigment Cell Res* **16**, 237–244.

Lehman, A. L., Silvers, W. K., Puri, N., Wakamatsu, K., Ito, S. & Brilliant, M. H. (2000) The underwhite (uw) locus acts autonoumously and reduces the production of melanin. *J Invest Dermatol* **115**, 601–606.

Lewontin, R. C. (1964) Interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67.

Lewontin, R. C. (1995) The detection of linkage disequilibrium in molecular sequence data. *Genetics* **140**, 377–388.

Lewontin, R. C. & Krakauer, J. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195.

Nakayama, K., Fukamachi, S., Kimura, H., Koda, Y., Soemantri, A. & Ishida, T. (2002) Distinctive distribution of *AIM1* polymorphism among major human populations with different skin color. *J Hum Genet* **47**, 92–94.

Newton, J. M., Cohen-Barak, O., Hagiwara, N., Gardner, J. M., Davisson, M. T., King, R. A. & Brilliant, M.H. (2001) Mutations in the human orthologue of the mouse *underwhite* gene (*uw*) underlie a new form of oculocutaneous albinism, OCA4. *Am J Hum Genet* **69**, 981–988.

Pakendorf, B., Wiebe, V., Tarskaia, L. A., Spitsyn, V. A., Soodyall, H., Rodewald, A. & Stoneking, M. (2003) Mitochondrial DNA evidence for admixed origins of Central Siberian populations. *Am J Phys Anthropol* **120**, 211–224.

Rees, J. L. (2004) The genetics of sun sensitivity in humans. *Am J Hum Genet* **75**, 739–751.

Relethford, J. H. (2002) Apportionment of global human genetic diversity based on craniometrics and skin color. *Am J Phys Anthropol* **118**, 393–398.

Rundshagen, U., Zühlke, C., Opitz, S., Schwinger, E. & Käsmann-Kellner, B. (2004) Mutations in the MATP gene in five German patients affected by oculocutaneous albinism type 4. *Hum Mutat* **23**, 106–110.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin: A software for population genetics data analysis*. Ver 2.000. Genetics and Biometry Lab, Dept of Anthropology, University of Geneva.

Sturm, R. A. & Frudakis, T.N. (2004) Eye colour: portals into pigmentation genes and ancestry. *Trends Genet* **20**, 327–332.

Tishkoff, S. A., Pakstis, A. J., Ruano, G. & Kidd, K. K. (2000) The accuracy of statistical methods for estimation of haplotype frequencies: an example from the CD4 locus. *Am J Hum Genet* **67**, 518–522.

Yao, Y. G., Kong, Q. P., Wang, C. Y., Zhu, C. L. & Zhang, Y. P. (2004) Different matrilineal contributions to genetic structures of ethnic groups in the Silk Road region in China. *Mol Biol Evol* **21**, 2265–2280.

Yuasa, I., Umetsu, K., Watanabe, G., Nakamura, H., Endoh, M. & Irizawa, Y. (2004) MATP polymorphisms in Germans and Japanese: L374F mutation as a population marker for Caucasoids. *Int J Legal Med* **118**, 364–366.

Received: 26 April 2005
Accepted: 21 October 2005