

# A scan for genetic determinants of human hair morphology: *EDAR* is associated with Asian hair thickness

Akihiro Fujimoto<sup>1</sup>, Ryosuke Kimura<sup>5,\*</sup>, Jun Ohashi<sup>1</sup>, Kazuya Omi<sup>1</sup>, Rika Yuliwulandari<sup>1,6</sup>, Lilian Batubara<sup>6</sup>, Mohammad Syamsul Mustofa<sup>7</sup>, Urai Samakkarn<sup>8</sup>, Wannapa Settheetham-Ishida<sup>9</sup>, Takafumi Ishida<sup>2</sup>, Yasuyuki Morishita<sup>3</sup>, Takuro Furusawa<sup>4</sup>, Minato Nakazawa<sup>10</sup>, Ryutaro Ohtsuka<sup>11</sup> and Katsushi Tokunaga<sup>1</sup>

<sup>1</sup>Department of Human Genetics, Graduate School of Medicine, <sup>2</sup>Department of Biological Sciences, Graduate School of Science, <sup>3</sup>Department of Molecular Pathology, Graduate School of Medicine and <sup>4</sup>Division for International Relations, The University of Tokyo, Hongo, Tokyo, Japan, <sup>5</sup>Department of Forensic Medicine, Tokai University School of Medicine, Isehara, Kanagawa, Japan, <sup>6</sup>Pharmacology Department and, <sup>7</sup>Biology Department, Yarsi University, Central Jakarta, DKI Jakarta, Indonesia, <sup>8</sup>Rawai Health Centre, Rawai, Muang, Phuket, Thailand, <sup>9</sup>Department of Physiology, Faculty of Medicine, Khon Kaen University, Mittraphab Road, Khon Kaen, Thailand, <sup>10</sup>Socio-Environmental Health Sciences, Graduate School of Medicine, Gunma University, Maebashi-City, Gunma, Japan and <sup>11</sup>National Institute for Environmental Studies, Onogawa, Tsukuba-City, Ibaraki, Japan

Received September 11, 2007; Revised November 13, 2007; Accepted December 2, 2007

**Hair morphology is one of the most differentiated traits among human populations. However, genetic backgrounds of hair morphological differences among populations have not been clarified yet. In addition, little is known about the evolutionary forces that have acted on hair morphology. To identify hair morphology-determining genes, the levels of local genetic differentiation in 170 genes that are related to hair morphogenesis were evaluated by using data from the International HapMap project. Among highly differentiated genes, *ectodysplasin A receptor (EDAR)* harboring an Asian-specific non-synonymous single nucleotide polymorphism (1540T/C, 370Val/Ala) was identified as a strong candidate. Association studies between genotypes and hair morphology revealed that the Asian-specific 1540C allele is associated with increase in hair thickness. Reporter gene assays suggested that 1540T/C affects the activity of the downstream transcription factor NF- $\kappa$ B. It was inferred from geographic distribution of 1540T/C and the long-range haplotype test that 1540C arose after the divergence of Asians from Europeans and its frequency has rapidly increased in East Asian populations. These findings lead us to conclude that *EDAR* is a major genetic determinant of Asian hair thickness and the 1540C allele spread through Asian populations due to recent positive selection.**

## INTRODUCTION

There are numerous physiological and morphological variations in humans and some of them are diverged between populations. Besides skin color and facial features, hair morphology is one of the most distinctive traits among human populations, and classical classification of human populations

was based on such visible traits. In previous comparative studies of hair morphology among human populations, the differences between Asians, Europeans and Africans were observed in diameter, shape of cross-section and fiber, mechanical properties, and hair moisture (1,2). Most notably, African hair is more twisted than Asian and Caucasian hair, and Asian hair has a larger and more circular cross-section

\*To whom correspondence should be addressed at: Department of Forensic Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan. Tel: +81 463931121 ext. 2630; Fax: +81 463920284; Email: rkimura@is.icc.u-tokai.ac.jp

than African and Caucasian hair (1). Older genetic studies in some ethnic groups suggested that only a fairly small number of genes determine hair morphology (3–5). However, genes associated with common variation in hair morphology have not been clarified yet. Furthermore, little is known about the evolutionary forces that have acted on hair morphology.

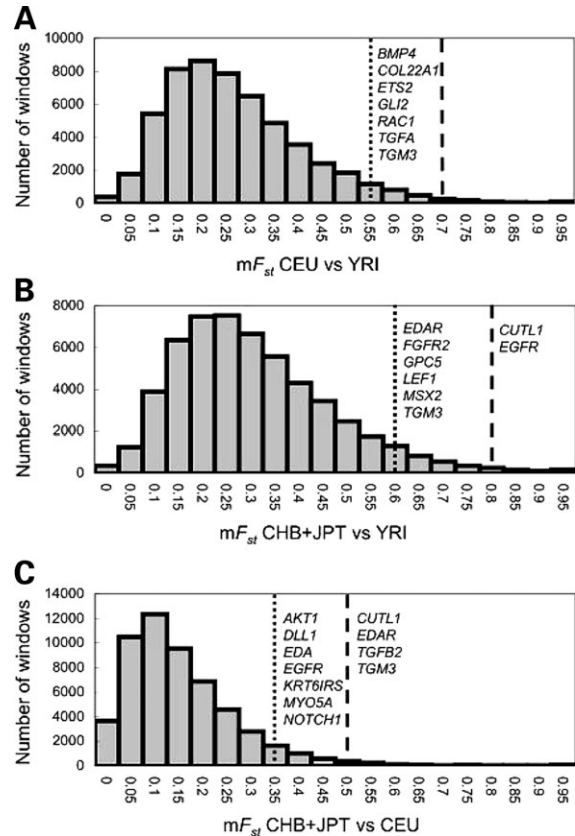
To identify candidates of hair morphology-related genes, it is useful to adopt a simple approach based on genetic differentiation between populations. Since hair morphology is highly differentiated between African, European, and Asian populations and determined by a small number of genes, a remarkable difference in the allele frequencies between these populations are expected to be found at loci that largely contribute to hair morphology. In fact, it has been reported that the genes involved in skin color (6,7), lactase tolerance (8), and malaria resistance (9) showed a high differentiation between populations.

The main aims of this study are to identify genes largely contributing to the differentiation of hair morphology and to elucidate the evolutionary history of hair morphology. Here, we performed (i) a population genetics-based analysis using a genome wide single nucleotide polymorphism (SNP) database to find candidate genes possibly related to hair morphology, (ii) an association test between hair morphology and polymorphisms in the most likely candidate gene, (iii) a functional analysis of the polymorphism that was associated with hair morphology and (iv) an evolutionary analysis of the gene. From these analyses, we revealed that *ectodysplasin A receptor (EDAR)* is associated with Asian hair thickness and that the Asian-specific 1540C allele has been subjected to positive selection.

## RESULTS AND DISCUSSION

### Interpopulation differentiation of genes involved in hair morphogenesis

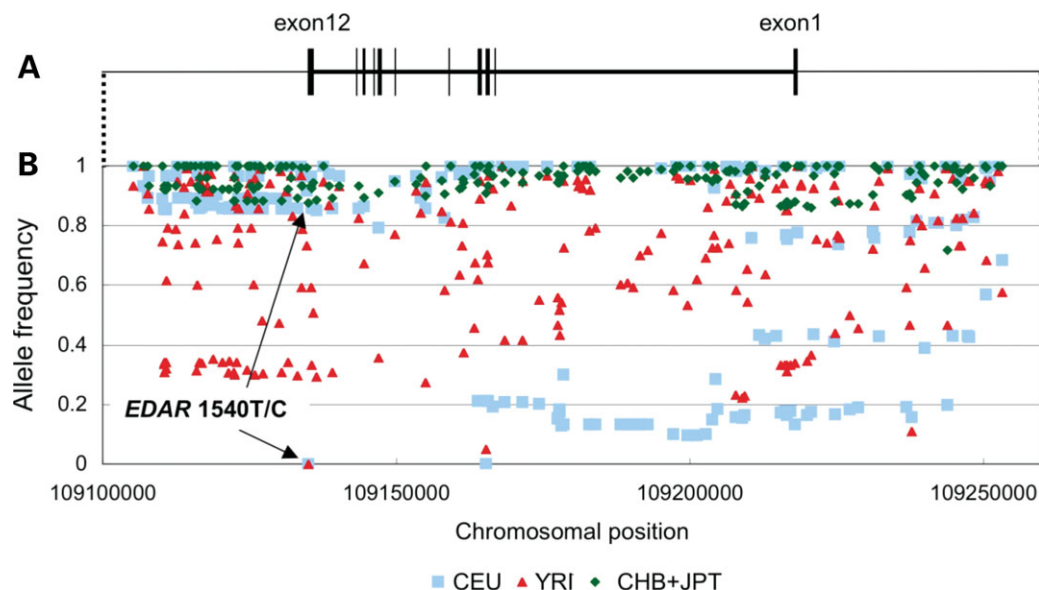
We first selected 170 genes that are related to hair morphogenesis, based on previous reports (10) and databases such as Online Mendelian Inheritance in Man (OMIM) and Gene Ontology (GO) (see Materials and Methods and Supplementary Material, Table S1). To examine the differentiation of these genes, we analyzed SNP data of the International HapMap Project (60 from Yorba in Ibadan, Nigeria, YRI; 60 from the CEPH population of northern and western European ancestry in Utah, CEU; 90 each from Han Chinese in Beijing and Japanese in Tokyo, CHB+JPT) (11). We then divided the sequence of each gene into 50 kb windows, calculated  $F_{ST}$  for each SNP between each combination of two populations, and obtained the maximum value of  $F_{ST}$  in each window ( $mF_{ST}$ ) (Supplementary Material, Table S1). We also obtained the empirical distribution of  $mF_{ST}$  across the entire genomic windows to determine the 95th and 99th percentiles as critical values (Supplementary Material, Table S2). Although such an outlier approach may not assure the detection power and/or may include the false positives, it is very concise and effective for the first scanning of candidate loci. As a result, we found that 21 and 5 genes have higher  $mF_{ST}$  values than the 95th and 99th percentiles, respectively (Fig. 1 and Supplementary



**Figure 1.** Empirical distributions of  $mF_{ST}$  and genes showing high  $mF_{ST}$ . (A) CEU versus YRI, (B) CHB+JPT versus YRI, (C) CHB+JPT versus CEU. Dashed line: 99th percentile; dotted line: 95th percentile.

Material, Table S3). Three of these genes (*CUTL1*, *EDAR* and *EGFR*) were differentiated both between CEU and CHB+JPT and between YRI and CHB+JPT, while *TGM3* was differentiated between all three populations. Therefore, these genes were strong candidates for being hair morphology-determining genes.

Among these candidate genes, we focused on *EDAR* for the subsequent studies since *EDAR* not only showed one of the highest  $mF_{ST}$  between CEU and CHB+JPT (Supplementary Material, Table S3) but also had a highly differentiated non-synonymous SNP (rs3827760), which has previously been suspected to be a target of positive selection (12–14). This non-synonymous SNP is located at the 1540th nucleotide from the transcription start site (370th amino acid from the translation start codon) and thus is called 1540T/C (370Val/Ala) here. A comparison with the sequence of chimpanzee (accession: XM\_525853) indicated that 1540C is the derived allele. The allele frequency of 1540C reaches 87.6% in CHB+JPT, whereas no 1540C is observed either in CEU or in YRI (Fig. 2B) (11). Such an Asian-specific allele with high frequency was not found elsewhere in the *EDAR* region (Fig. 2B). These observations imply that 1540C arose after the divergence of Asians from Europeans. We resequenced the complete coding exons and partial flanking introns of *EDAR* for 24 individuals from the HapMap samples, but did not detect any other differentiated SNP in the coding regions (Supplementary Material, Table S4). As shown in



**Figure 2.** Structure and allele frequencies of SNPs in *EDAR*. (A) Structure of *EDAR*; (B) the frequency of major allele in CHB+JPT in each population.

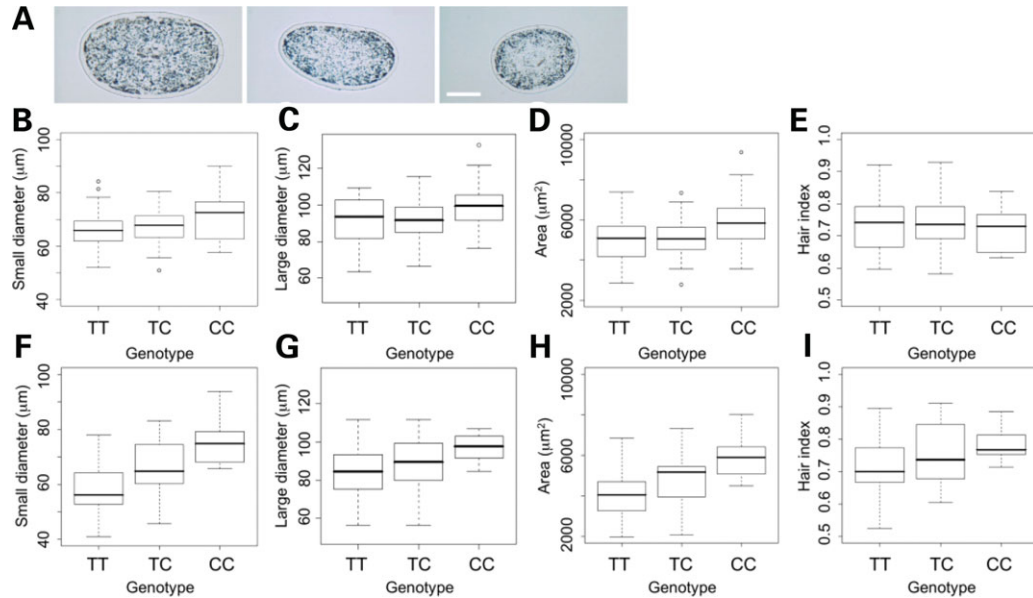
Figure 2, highly differentiated SNPs were densely scattered from the 5' region to the intron 4. Above all, an SNP, rs922452, showed the highest  $F_{ST}$  value between CEU and CHB+JPT in the *EDAR* region (Supplementary Material, Table S3). In addition, a previous study observed that three SNPs in the 5' region within 2 kb of the transcription start site, i.e. -1430G/A, 62C/T and 930A/G (previously described as 173, 1663 and 2531), which were in almost absolute LD, have frequency difference of over 85% between populations (14). However, these SNPs were not analyzed in the HapMap project.

#### Association between *EDAR* polymorphisms and hair morphology

To examine the association of polymorphisms in *EDAR* with hair morphology, two South-east Asian populations were subjected to following analyses since various hair phenotypes were observed within these populations. We recruited 121 unrelated individuals of the Indonesian (IDN), who are inhabitants in the west part of Java Island in Indonesia, and 65 unrelated individuals of the Thai-Mai (THM), who are called 'sea gypsies' in Thailand. Cross-sections made from five hairs per individual were prepared, and large diameter, small diameter, and area of hairs were measured under microscopy (Fig. 3A and Supplementary Material, Table S5). We then calculated the hair index, i.e. the ratio of small diameter to large diameter, as a measure of hair shape, which is usually smaller in curly or frizzy hair than in straight hair (5). Summary statistics for the measurements of hair sections are shown in Supplementary Material, Table S5. We genotyped 1540T/C in the exon 12, rs922452 in the intron 3, and -1430G/A as a representative for the highly differentiated SNPs in the 5' region. The number of individuals with each genotype was presented in Table 1. These genotype frequencies were consistent with the expectation from Hardy-Weinberg equilibrium except for -1430G/A in IDN ( $P = 1.6 \times 10^{-3}$ ).

Using analysis of variance (ANOVA), we found significant associations of 1540T/C with small diameter and area in both IDN and THM, and with large diameter in IDN (Fig. 3). In particular, area exhibited a strong association with 1540T/C ( $P = 5.5 \times 10^{-3}$  in IDN and  $P = 9.5 \times 10^{-4}$  in THM). The mean area for the TT, TC and CC genotypes was 4986, 5100 and 5927  $\mu\text{m}^2$ , respectively, in IDN, and 4060, 4844 and 5924  $\mu\text{m}^2$ , respectively, in THM. The rs922452 also showed significant association with area ( $P = 7.0 \times 10^{-3}$  in IDN and  $P = 0.017$  in THM), but  $P$ -values were higher than these of 1540T/C. This indicated that 1540T/C is the causative polymorphism. Because the rs922452 was in LD with 1540T/C ( $D' = 1.0$ ,  $R^2 = 0.42$  in IDN and  $D' = 0.94$ ,  $R^2 = 0.40$  in THM), the association of rs922452 was likely to be caused by LD. In contrast, -1430G/A did not show any significant association. Since the highly differentiated SNPs in the 5' region are within 2.5 kb and in strong LD in Chinese-descendant population ( $R^2 > 0.9$ ) (Seattle SNP database), these SNPs are expected to be in strong LD also in the Southeast Asian populations tested. Therefore, the promoter region is thought to be irrelevant to the phenotype.

We further analyzed the contribution of *EDAR* variants to hair morphology, considering effects of other factors. For this purpose, sex, age, ethnicity, and the number of an allele were entered into multiple regression analyses as independent variables. The analyses on IDN suggested that older age shows significantly smaller area and shorter large diameter, and that female shows significantly smaller area, shorter large diameter and higher hair index (Table 2). Although IDN individuals were ethnically classified into Java or Sunda, ethnic difference had an influence only on hair index (Table 2). In these analyses, 1540T/C showed significant association with area, small diameter and large diameter (Table 2). These results suggested that the Asian-specific 1540C allele is associated with thicker hair, but not with hair index. If we assumed recessive model for 1540C (TT and TC: 0; CC: 1),



**Figure 3.** *EDAR* 1540T/C and hair morphology. (A) Examples of hair cross-sections, bar = 40  $\mu\text{m}$ . (B)–(E) For IDN: (B) small diameter – ANOVA  $P = 0.032$ ; (C) large diameter – ANOVA  $P = 0.018$ ; (D) cross-section area – ANOVA  $P = 5.5 \times 10^{-3}$ ; (E) Hair index. (F)–(I) For THM: (F) small diameter – ANOVA  $P = 5.7 \times 10^{-5}$ ; (G) large diameter; (H) cross-section area – ANOVA  $P = 9.5 \times 10^{-4}$ ; (I) hair index.

**Table 1.** Genotype and allele frequencies for *EDAR* 1540T/C and –1430G/A

Population	Genotype frequencies			Allele frequencies	
	1540TT	1540TC	1540CC	1540T	1540C
IDN ( $n = 121$ )	57 (47.1%)	46 (38.0%)	18 (14.9%)	66.1%	33.9%
THM ( $n = 65$ )	33 (50.8%)	25 (38.4%)	7 (10.8%)	70.0%	30.0%
	–1430GG	–1430GA	–1430AA	–1430G	–1430A
IDN ( $n = 121$ )	49 (40.5%)	40 (33.1%)	32 (26.4%)	57.0%	43.0%
THM ( $n = 65$ )	28 (43.1%)	27 (41.5%)	10 (15.4%)	63.8%	36.2%
	rs922452GG	rs922452GA	rs922452AA	rs922452G	rs922452A
IDN ( $n = 116$ )	26 (22.4%)	51 (44.0%)	39 (33.6%)	44.4%	55.6%
THM ( $n = 65$ )	19 (29.2%)	29 (44.6%)	17 (26.2%)	51.5%	48.5%

**Table 2.** Multiple regression analyses for hair morphology in IDN ( $n = 121$ )

Explanatory variables	Area			Small diameter			Large diameter			Hair index		
	RC	$F$	$P$ -value	RC	$F$	$P$ -value	RC	$F$	$P$ -value	RC	$F$	$P$ -value
Age	–21.7	6.03	0.016	–	–	–	–0.211	4.57	0.035	–	–	–
Sex (M: 0, F: 1)	–570	6.28	0.014	–	–	–	–6.85	7.30	0.0079	0.041	6.23	0.014
Ethnicity (Sunda: 0, Java: 1)	–	–	–	–	–	–	–	–	–	0.037	4.27	0.041
<i>EDAR</i> 1540T/C (TT: 0, TC: 1, CC: 2)	360	7.45	0.0073	2.18	6.32	0.013	3.34	5.17	0.025	–	–	–

A stepwise method ( $F_{\text{IN}}$  and  $F_{\text{OUT}}$ : 4) was used in the multiple regression analyses. RC: regression coefficient; –: a variable excluded during the stepwise procedure.

the 1540T/C showed more significant association with area ( $P = 2.9 \times 10^{-3}$ ).

Multiple regression analyses were also performed on THM individuals since they are originated from two different ethnics, Urak Lawoi and Moken. The analysis only on Urak Lawoi ( $n = 37$ ) showed that age and 1540T/C were significantly associated with area, but sex was not (Table 3). When

the individuals with Moken admixture were added ( $n = 65$ ), lower  $P$ -values were observed, while ethnicity showed no association (Table 3). Contrary to the result of IDN, the recessive model showed weaker association ( $P = 3.5 \times 10^{-3}$ ) than the allelic model. The regression coefficients, which denote the effect of the 1540C allele on increasing area, were also different between THM and IDN (Tables 2 and 3). Combining

**Table 3.** Multiple regression analyses for the area of hair cross-section in THM ( $n = 65$ )

Explanatory variables	Urak Lawoi ( $n = 37$ )			THM ( $n = 65$ )		
	RC	$F$	$P$ -value	RC	$F$	$P$ -value
Age	-27.5	4.92	0.033	-29.1	8.76	0.0043
Sex (M: 0, F: 1)	-	-	-	-	-	-
Ethnicity (Urak Lawoi: 0, Moken: 1)	NI	NI	NI	-	-	-
<i>EDAR</i> 1540T/C (TT: 0, TC: 1, CC: 2)	809	8.09	0.0075	740	11.77	0.0011

A stepwise method ( $F_{IN}$  and  $F_{OUT}$ : 4) was used in the multiple regression analyses. RC: regression coefficient; -: a variable excluded during the stepwise procedure; NI: a variable not included in the analyses. Ethnicity of the individuals mixed between Urak Lawoi and Moken were represented by 1/4, 1/2 or 3/4 depending on grandparents' ethnicities.

**Table 4.** Multiple regression analyses for the area of hair cross-section in IDN and THM ( $n = 186$ )

Explanatory variables	<i>EDAR</i> 1540T/C			<i>ABCC11</i> rs17822931		
	RC	$F$	$P$ -value	RC	$F$	$P$ -value
Age	-28.5 <sup>a</sup>	20.3	$1.1 \times 10^{-5}$	-31.9	23.8	$2.4 \times 10^{-6}$
Sex (M: 0, F: 1)	-463 <sup>a</sup>	8.17	0.0048	-482.7	8.11	0.0049
Ethnicity (IDN: 0, THM: 1)	-	-	-	-	-	-
<i>EDAR</i> 1540T/C (TT: 0, TC: 1, CC: 2)	491 <sup>a</sup>	18.4	$2.8 \times 10^{-5}$	NI	NI	NI
<i>ABCC11</i> rs17822931 (CC: 0, CT: 1, TT: 2)	NI	NI	NI	-	-	-

A stepwise method ( $F_{IN}$  and  $F_{OUT}$ : 4) was used in the multiple regression analyses. RC: regression coefficient; -: a variable excluded during the stepwise procedure; NI: a variable not included in the analyses.

<sup>a</sup> $R^2$  values of age, sex and 1540T/C were 0.088, 0.035 and 0.080, respectively.

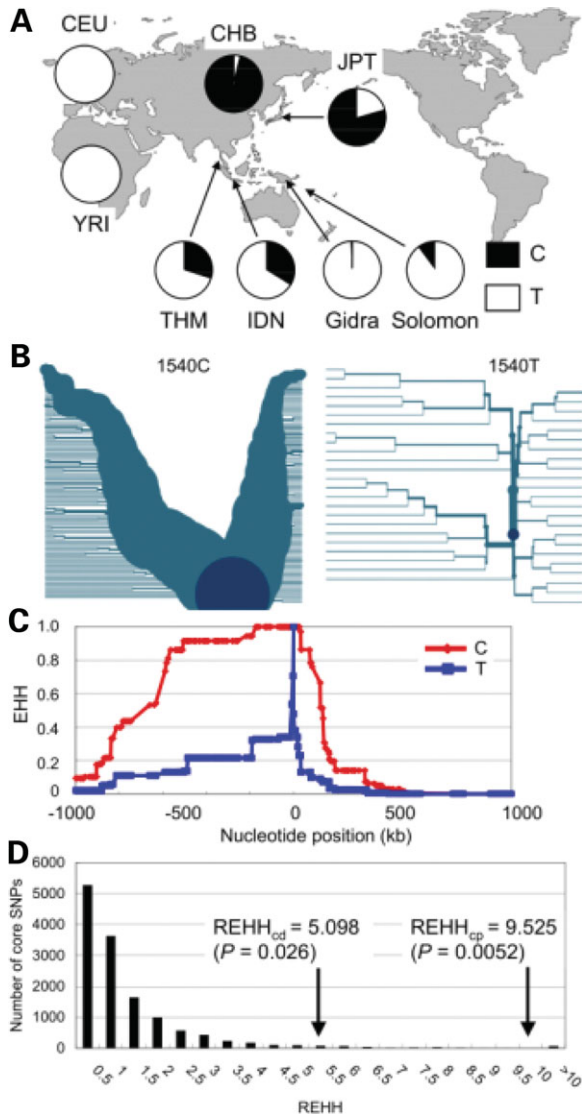
IDN and THM, we could obtain a further lower  $P$ -value in the correlation between 1540T/C and area of hair cross-section ( $P = 2.8 \times 10^{-5}$ ) (Table 4). The effect of the 1540C allele on increasing area was estimated to be  $491 \mu\text{m}^2$ . Also in this analysis, no association between ethnicity and area was observed. Furthermore, to consider the skew of allele frequency due to population stratification, we genotyped an SNP, rs17822931 in *ABCC11* gene, that is highly differentiated between Asian and other populations and is unlikely to be related to hair morphogenesis (15). Indeed, multiple regression analysis including this SNP in addition to age, sex and ethnicity as independent variables showed no association between the SNP and area (Table 4). These results suggest that population stratification, if any, is not responsible for the association observed in *EDAR* 1540T/C.

We compared several populations in 1540C frequency in order to estimate the contribution of the 1540T/C polymorphism to the difference in hair thickness between populations. In a previous paper, it has been reported that the mean area of hair cross-sections was  $3857 \mu\text{m}^2$  in Caucasians and  $4274 \mu\text{m}^2$  in Africans (1), which is similar to the value of 1540TT individuals in THM but lower than that in IDN. In addition, we measured the area of 12 Japanese individuals with unknown genotype and the mean area was  $5639 \mu\text{m}^2$ . When the allele frequency of 1540C in JPT (79.5%) was considered, the mean area of 12 Japanese is in agreement with the expectation calculated from the mean area of each genotype under the assumption of Hardy-Weinberg equilibrium ( $5494 \mu\text{m}^2$  from THM,  $5618 \mu\text{m}^2$  from IDN). Since it has been reported that the diameter of Melanesian hair is similar to that of African and European hair (16), we were also interested

in the allele frequency of 1540C in Melanesia. We therefore genotyped 1540T/C in two Melanesian populations, the Gidra in the New Guinea island and the Solomon islanders. The allele frequencies of 1540C were 1.0% in the Gidra and 10.4% in the Solomon islanders (Fig. 4A). The low 1540C frequency in the Melanesian populations is consistent with the thin phenotype of their hairs. These results support that *EDAR* is a genetic determinant of Asian hair thickness and can explain a large part of the difference of hair thickness between Asians and other populations.

#### Functional analysis of *EDAR* 1540T/C

*EDAR*, ectodysplasin A receptor, is a member of the tumor necrosis factor receptor family. The disruption of *EDAR* in humans causes ectodermal dysplasia, which is characterized by abnormal morphogenesis of teeth, hair and eccrine sweat glands (10,17,18). The 1540T/C (370Val/Ala) polymorphism is located in the death domain, that is, the intracellular domain necessary to interact with *EDAR*-binding death domain adapter protein, *EDARADD* (19,20). It is known that *EDAR/EDARADD* interaction results in the activation of the downstream transcription factor NF- $\kappa$ B, and this molecular pathway plays a key role in formation of hair placode (10). Therefore, a possibility is that 1540T/C affects the NF- $\kappa$ B activity through an altered efficiency of interaction between *EDAR* and *EDARADD*. To compare the 1540T and C alleles in the ability of NF- $\kappa$ B activation, we performed luciferase assays using HeLa and 293A cells. These cells were transfected with *EDAR* expression vectors carrying each allele (pEF1-*EDAR*-1540T and pEF1-*EDAR*-1540C),



**Figure 4.** Evolutionary history of *EDAR* 1540T/C. (A) Geographical distribution of 1540C; (B) the extended haplotype frequencies at various distance from 1540T/C; (C) EHH values at various distances from 1540T/C in CHB+JPT; (D) empirical distribution of REHH values for alleles with frequencies of  $87.6 \pm 2.5\%$  at 0.25 cM distance from the SNPs on chromosome 2 in CHB+JPT. REHH values of 1540C on both centromere-proximal ( $REHH_{cp}$ ) and centromere-distal sides ( $REHH_{cd}$ ) are shown.

reporter plasmids with the luciferase gene under control of the five NF- $\kappa$ B promoter elements (pNF- $\kappa$ B-LUC plasmid) and internal control vectors (pRh-TK vector). The NF- $\kappa$ B/luciferase activities were compared among 1540T, 1540C and 1540T+C (artificial heterozygote). We observed significant differences in relative luciferase activities between 1540T and 1540C (*t*-test: HeLa cell  $P = 5.7 \times 10^{-3}$ , 293A cell  $P = 1.9 \times 10^{-5}$ ) and between 1540T and 1540T+C (293A cell  $P = 5.8 \times 10^{-5}$ ) (Fig. 5). Interestingly, the C allele that was associated with thicker hair showed lower relative luciferase activities than the T allele in both the cell lines. These results indicated that the amino acid replacement in the death domain causes a functional change and results in the lower activity of NF- $\kappa$ B. Although the relation between

NF- $\kappa$ B activation level and hair thickness has not been clear, it has been reported that steroid induces NF- $\kappa$ B suppression as well as hair regrowth (21,22), which supports that the lower NF- $\kappa$ B level may be associated with higher activity of hair formation. A previous study has suggested that the amino acid change of 1540T/C (Val/Ala) is a quite conservative and is predicted to have 'benign' effect (14). Taking it into account that the death domain in *EDAR* is completely conserved between mouse and human (data not shown), we may be able to interpret that such a conservative amino acid change on the domain with a important function can cause a mild effect on the phenotype.

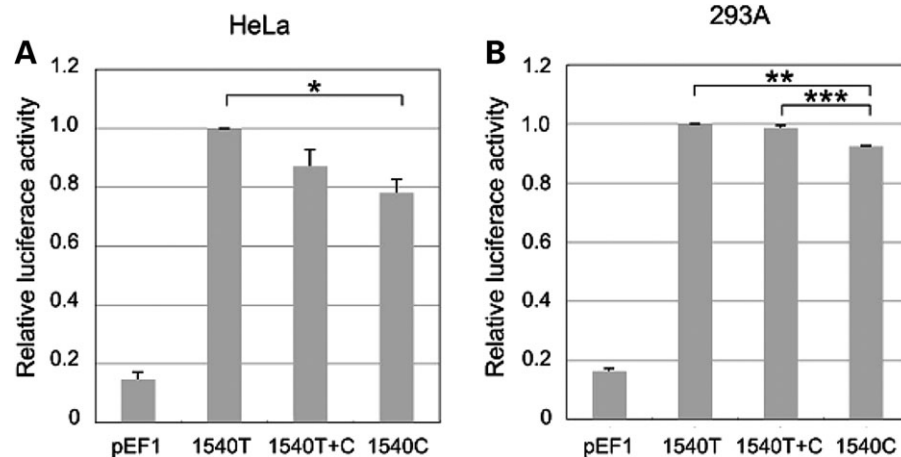
### Evolutionary history in *EDAR* 1540T/C

To assess the evolutionary history of 1540C, we examined the extended haplotype homozygosity (EHH) (23) for 1540T/C based on the haplotype data from Phase I of the International HapMap Project (11). The extended haplotype frequencies and EHH at various distances from 1540T/C in CHB+JPT are shown in Figure 4B and C. Although 1540C has a higher allele frequency than 1540T, EHH of 1540C decays more gently than that of 1540T. To evaluate the significance of the EHH value, we performed the long-range haplotype (LRH) test (23). 1540C obviously deviated from the empirical distribution ( $P = 5.2 \times 10^{-3}$  for the centromere-distal side and  $P = 0.026$  for the centromere-proximal side) (Fig. 4D). The slower rate of LD decay, as well as high population differentiation in 1540C, indicates that the frequency of 1540C rapidly increased in the East Asian populations by recent positive selection. Carlson *et al.* (14) also reported that Tajima's *D*, and Fay and Wu's *H* values in this region demonstrated deviation from the neutral expectation.

Although it is difficult to specify the target site of positive selection in the region around *EDAR* because of the strong LD, the results of our association and functional analyses imply that *EDAR* 1540T/C is the target of positive selection. A possible explanation about evolutionary force against *EDAR* is cold tolerance. Since hair can play an important role in the protection of the head against coldness by preventing heat exhalation, the thicker hair of 1540C carriers may have been advantageous in cold climates in the north part of Asia. An alternative possibility is that functional changes on *EDAR* may affect another trait. For example, since disruption of *EDAR* results in abnormal teeth morphogenesis (10), it is possible that the functional change between 1540T and C also have an influence on teeth morphology, which is known to have diverged phenotypes among populations (24) that might be caused by adaptation to the local diets.

### CONCLUSION

This is the first report about the genetic basis of the common hair morphological difference and its molecular evolution. We provided a clear evidence of the association between the 1540T/C polymorphism of *EDAR* and hair thickness. The evolutionary analysis suggested that the derived allele, 1540C, increased in Asian populations by recent positive selection. However, mode of inheritance in the phenotype is



**Figure 5.** Luciferase assay for the NF- $\kappa$ B activity. The effects of the 1540T/C were examined in the two cell lines HeLa and 293A. Relative luciferase activities were standardized as fold activities upon each cells transfected by pEF1-EDAR-1540T (ancestral type). Values represent the means  $\pm$  SE of three independent transfections, each with triplicate determinations. (A) HeLa cell; *t*-test: 1540T versus 1540C  $P = 0.0057$  (\*) (B) 293A cell; *t*-test: 1540T versus 1540C  $P = 1.9 \times 10^{-5}$  (\*\*), 1540T+C versus 1540C  $P = 5.8 \times 10^{-5}$  (\*\*\*). *P*-values for *t*-tests were adjusted by Holm's method.

still controversial. In our study, different populations showed different values of regression coefficients in the association studies. As further studies, we need to consider the effect of other genes and gene–gene interactions as well as to examine other populations. In particular, *EDA* that encode the ligand of EDAR was listed as another candidate gene in this study and should be a target in a future study.

As shown here, a simple population genetics analysis based on local genetic differentiation enabled us to identify a genetic determinant of hair morphology that can explain a large part of difference in hair thickness between Asians and other populations. Although the genetic basis of the difference in hair frizziness between populations still remains to be elucidated, it will be revealed in the same manner and this trait may be involved in some of these genes that showed high differentiation in our study. In addition, it is possible to find new associations between genes and other highly differentiated human traits such as pigmentation and body composition (13,25). Furthermore, population genetics-based studies can contribute to genome-wide case-control association studies on diseases with different prevalence among populations, such as obesity, diabetes and hypertension: the combination of the two strategies would allow us to identify the susceptibility genes more efficiently. Such scans for candidate genes, and the follow-on association and functional studies, will become more important tools for identifying the loci related to phenotypic variations in human populations, and will provide us with advanced knowledge about the history of human adaptations to local environments.

## MATERIALS AND METHODS

### A scan based on interpopulation genetic differentiation

We used SNP data of 210 unrelated individuals from Phase II of the International HapMap Project (60 from Yorba in Ibadan, Nigeria, YRI; 60 from the CEPH population of northern and western European ancestry in Utah, CEU; 90 each from Han Chinese in Beijing and Japanese in Tokyo,

CHB+JPT) (11). To estimate the levels of population differentiation in genes related to hair morphogenesis, the following procedures were adopted. We first selected 170 candidate genes related to hair morphogenesis, based on a review paper of Schmidt-Ullrich and Paus (10) and databases such as GO: <http://www.geneontology.org/> and OMIM: <http://www.ncbi.nlm.nih.gov/Omim/>. These candidate genes included hair keratins, keratin-related proteins, genes related to hair abnormalities and hair formation in human and/or mouse. We divided the nucleotide sequence of each gene into 50 kb windows, putting the center of the gene at the center of a window. When the length of the gene was longer than 50 kb, further windows were added on the both centromere-proximal and -distal sides until the entire gene is covered by the windows. When a 50 kb window included a part of the gene region, the window was regarded as connected with the candidate gene. Second, we calculated  $F_{ST}$  between CEU, YRI and CHB+JPT for each SNP as a measure of population differentiation. We selected only polymorphic SNPs that had been genotyped in all the populations. Next, the maximum value of  $F_{ST}$  in each 50 kb window ( $mF_{ST}$ ) was obtained. We also obtained the empirical distribution of  $mF_{ST}$  for the entire genomic region to determine the 95th and 99th percentiles. The genes showing  $mF_{ST}$  higher than the 95th or 99th percentiles were considered as candidate genes. We also determined the derived alleles of the SNPs with highest  $F_{ST}$  in these genes by comparisons with chimpanzee genome sequence.

### Samples

Examinations of hair morphology were performed on two Southeast Asian populations since various hair phenotypes were observed within these populations. The subjects were 121 unrelated individuals in the west part of Java Island, Indonesia and 205 individuals including relatives in the Rawai village of Phuket, Thailand. Those Indonesian individuals were gathered from several villages and ethnically classified

into Sunda or Java. People in the Rawai village are minority ethnic people known as the Thai-Mai, which are composed of two groups, Urak Lawoi and Moken. They are also called 'sea gypsies' because of their past nomadic mode of life on the ocean (26). The Urak Lawoi, who are major in the Rawai village, are thought to have migrated to this area about 200 years ago. On the other hand, the settlement of the Moken occurred for past several decades. Therefore, ethnicity and family history were asked to the subjects in the Rawai village. From the 205 individuals of Thai-Mai, we selected 65 unrelated individuals for the association study. Blood and hair samples were obtained from them with informed consent. DNA was extracted from the blood samples with a standard method. For measurement of hair morphology, we used five hairs per individual to make cross-sections. The hair samples were embedded in paraffin or epoxy resin and cut into 1–9  $\mu\text{m}$  thick sections. When embedded in paraffin blocks, hair samples were tensioned to be vertical to the surface of the block to ensure perpendicular cutting of hair cross-sections. Then, large diameter, small diameter, and area of cross-sections were measured with an Aqua Cosmos/Basic system (Hamamatsu Photonics) attached to a Hamamatsu C4742-95 CCD camera. Large diameter was defined as the length of the largest axis, and small diameter was the length of the axis vertically passing the center of the largest axis. We also calculated the hair index, i.e. the ratio of small diameter to large diameter, which have been used as a measure of the hair shape: The hair index of curly or frizzy hair is usually smaller than that of straight hair (5). Hair samples from 12 Japanese volunteers with unknown genotype were also collected and analyzed in the same manner.

To examine the allele frequency of 1540T/C in Melanesian populations, we genotyped 48 Gidra and 48 Solomon Islanders. The Gidra are non-Austronesian-speaking Melanesian people of the southwestern lowlands of Papua New Guinea (27–30). The Solomon Islanders are Austronesian-speaking Melanesians (31).

### Variation screening and genotyping

We designed specific primers for the amplification of each exon and partial flanking intron of *EDAR* and analyzed 24 individuals from the HapMap samples (8 YRI, 8 CEU and 8 CHB individuals). After the polymerase chain reaction (PCR) amplification, we sequenced the PCR products on an ABI Prism 3100 automatic sequencer using BigDye terminator cycle sequencing ready reaction kit ver. 3.1 (PE Biosystems, Foster City, CA, USA). The *EDAR* 1540T/C and –1430A/G polymorphism were genotyped by PCR-direct sequencing. Additionally, we also genotyped an SNP, rs17822931 in *ABCC11* gene, as an ethnic marker that shows high differentiation between CHB+JPT and the others. The sequences of primers for PCR and sequencing are available on request.

### Luciferase assay

Total RNA was isolated from hair bulges of a Japanese individual with informed consent using Isogen RNA extraction reagent (Nippon Gene) and subjected to cDNA synthesis using oligo(dT) primers and ImProm-II reverse transcriptase

(Promega). The amino acid sequence of cDNA was identical with reference sequence (NM\_022336.2). DNA fragment including the coding region of *EDAR* 1540T was amplified by PCR from the resultant cDNAs using primer pair that contains either *EcoRI* or *XbaI* site: 5'-CCGGAATTCGGAGAGGATGGCCCATGTGG-3' and 5'-CTAGTCTAGAGGATGCAGCATGTGGCTGG-3'. The PCR product was digested with *EcoRI* and *XbaI*, and then subcloned into *EcoRI* and *XbaI* sites of pEF1/*Myc*-HisA vector (Invitrogen), and the resulted plasmid was designated pEF1-EDAR-1540T. pEF1-EDAR-1540C vector was generated from the pEF1-EDAR-1540T vector by QuikChange mutagenesis kit (Stratagene) using primers, 5'-AACTCTGAGAAGGCTGCTGTGAAAACGTGGCGC-3' and 5'-GCGCCACGTTTTCACAGCAGCCTTCTCAGAGTT-3' (mutated nucleotides are underlined). The NF- $\kappa$ B reporter assay was performed as follows. HeLa and 293A cells were plated into 24-well culture plates and transfected with 250 ng of pNF- $\kappa$ B-LUC plasmid (Stratagene), 50 ng of pRh-TK vector (Promega), and one of the followings: 300 ng of pEF1 empty vector (negative control), 300 ng of pEF1-EDAR-1540T, 300 ng of pEF1-EDAR-1540C, or 150 ng pEF1-EDAR-1540T and 150 ng pEF1-EDAR-1540C (artificial heterozygote), using lipofectamine 2000 reagent (Invitrogen) according to the manufacture's instructions. After 48-h incubation, the expression of luciferase was examined using a Dual-Luciferase reporter assay system (Promega).

### Statistical analyses

Allele frequency was estimated by gene counting. The agreement of genotype frequencies with Hardy–Weinberg expectations was tested by  $\chi^2$ -test. Comparisons of large diameter, small diameter, area, and hair index between genotypes were carried out using one-way ANOVA. The effects of genotypes, age, sex and ethnicity on hair morphology were examined by using multiple regression analysis with the stepwise procedure, where the criteria for variable selection ( $F_{\text{IN}}$ ) and rejection ( $F_{\text{OUT}}$ ) were set at 4.0. IDN ( $n = 121$ ), Urak Lawoi ( $n = 37$ ), THM (Urak Lawoi and Moken:  $n = 65$ ) and combination of IDN and THM ( $n = 186$ ) were used as the subjects for the regression analyses. In the analysis of THM, the ethnicity of an individual was evaluated based on his/her grandparents' ethnicities. In short, the ethnicity score was calculated as  $i/4$ , where  $i$  represents the number of Moken grandparent. Differences in relative luciferase activities were analyzed by using pair-wise  $t$ -test.

### Long-range haplotype test

We performed the LRH test on 1540T/C by using the haplotype data from Phase I (release 16c.1) of the International HapMap Project (11). EHH at varying distance was calculated with the SWEEP software (<http://www.broad.mit.edu/mpg/sweep/>) (23). To evaluate the statistical significance, we obtained the empirical distribution of relative EHH (REHH) at 0.25 centiMorgans (cM) distance on both centromere-proximal and -distal sides in CHB+JPT. Because REHH depends on the frequency of the tested allele (i.e. 87.6%), we calculated REHH values for all the alleles with frequency



ranging between 85.1 and 90.1% on chromosome 2 to draw the empirical distribution in CHB+JPT. The calculation of REHH was programmed in Visual Basic (Microsoft Excel).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

## ACKNOWLEDGEMENTS

We are deeply grateful to the people participated in this study.

*Conflict of Interest statement.* The authors declare no conflict of interest.

## FUNDING

This study is partly supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## REFERENCES

1. Franbourg, A., Hallegot, P., Baltenneck, F., Toutain, C. and Leroy, F. (2003) Current research on ethnic hair. *J. Am. Acad. Dermatol.*, **48**, S115–S119.
2. Khumalo, N.P., Doe, P.T., Dawber, R.P. and Ferguson, D.J. (2000) What is normal black African hair? A light and scanning electron-microscopic study. *J. Am. Acad. Dermatol.*, **43**, 814–820.
3. Losty, J.P. (1928) Hybridization among human races in South Africa. *Genetica*, **10**, 131.
4. Davenport, G.C. and Davenport, C.B. (1908) Heredity of hair form in man. *Am. Nat.*, **42**, 341.
5. Bean, R.B. (1911) Heredity of hair form among the Filipinos. *Am. Nat.*, **45**, 524.
6. Lamason, R.L., Mohideen, M.A., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., Jurynec, M.J., Mao, X., Humphreville, V.R., Humbert, J.E. *et al.* (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, **310**, 1782–1786.
7. Soejima, M., Tachida, H., Ishida, T., Sano, A. and Koda, Y. (2006) Evidence for recent positive selection at the human AIM1 locus in a European population. *Mol. Biol. Evol.*, **23**, 179–188.
8. Swallow, D.M. (2003) Genetics of lactase persistence and lactose intolerance. *Annu. Rev. Genet.*, **37**, 197–219.
9. Hamblin, M.T., Thompson, E.E. and Di Rienzo, A. (2002) Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.*, **70**, 369–383.
10. Schmidt-Ullrich, R. and Paus, R. (2005) Molecular principles of hair follicle induction and morphogenesis. *Bioessays*, **27**, 247–261.
11. The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
12. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R. *et al.* (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**, 913–918.
13. Kimura, R., Fujimoto, A., Tokunaga, K. and Ohashi, J. (2007) A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS ONE*, **2**, e286.
14. Carlson, C.S., Thomas, D.J., Eberle, M.A., Swanson, J.E., Livingston, R.J., Rieder, M.J. and Nickerson, D.A. (2005) Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.*, **15**, 1553–1565.
15. Yoshiura, K., Kinoshita, A., Ishida, T., Ninokata, A., Ishikawa, T., Kaname, T., Bannai, M., Tokunaga, K., Sonoda, S., Komaki, R. *et al.* (2006) A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat. Genet.*, **38**, 324–330.
16. Hrdy, D. (1973) Quantitative hair form variation in seven populations. *Am. J. Phys. Anthropol.*, **39**, 7–17.
17. Mikkola, M.L. and Thesleff, I. (2003) Ectodysplasin signaling in development. *Cytokine Growth Factor Rev.*, **14**, 211–224.
18. Monreal, A.W., Ferguson, B.M., Headon, D.J., Street, S.L., Overbeek, P.A. and Zonana, J. (1999) Mutations in the human homologue of mouse dl cause autosomal recessive and dominant hypohidrotic ectodermal dysplasia. *Nat. Genet.*, **22**, 366–369.
19. Headon, D.J., Emmal, S.A., Ferguson, B.M., Tucker, A.S., Justice, M.J., Sharpe, P.T., Zonana, J. and Overbeek, P.A. (2001) Gene defect in ectodermal dysplasia implicates a death domain adapter in development. *Nature*, **414**, 913–916.
20. Mustonen, T., Pispä, J., Mikkola, M.L., Pummila, M., Kangas, A.T., Pakkasjarvi, L., Jaatinen, R. and Thesleff, I. (2003) Stimulation of ectodermal organ development by Ectodysplasin-A1. *Dev. Biol.*, **259**, 123–136.
21. Ardite, E., Panes, J., Miranda, M., Salas, A., Elizalde, J.I., Sans, M., Arce, Y., Bordas, J.M., Fernandez-Checa, J.C. and Pique, J.M. (1998) Effects of steroid treatment on activation of nuclear factor kappaB in patients with inflammatory bowel disease. *Br. J. Pharmacol.*, **124**, 431–433.
22. Seiter, S., Ugurel, S., Tilgen, W. and Reinhold, U. (2001) High-dose pulse corticosteroid therapy in the treatment of severe alopecia areata. *Dermatology*, **202**, 230–234.
23. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**, 832–837.
24. Hanihara, T. and Ishida, H. (2005) Metric dental variation of major human populations. *Am. J. Phys. Anthropol.*, **128**, 287–298.
25. Myles, S., Somel, M., Tang, K., Kelso, J. and Stoneking, M. (2007) Identifying genes underlying skin pigmentation differences among human populations. *Hum. Genet.*, **120**, 613–621.
26. Ninokata, A., Kimura, R., Samakkarn, U., Settheetham-Ishida, W. and Ishida, T. (2006) Coexistence of five G6PD variants indicates ethnic complexity of Phuket islanders, Southern Thailand. *J. Hum. Genet.*, **51**, 424–428.
27. Bellwood, P. (1989) *The Colonization of The Pacific: Some Current Hypotheses*. Oxford University Press, Oxford.
28. Ohashi, J., Naka, I., Kimura, R., Tokunaga, K., Yamauchi, T., Natsuhara, K., Furusawa, T., Yamamoto, R., Nakazawa, M., Ishida, T. *et al.* (2006) Polymorphisms in the ABO blood group gene in three populations in the New Georgia group of the Solomon Islands. *J. Hum. Genet.*, **51**, 407–411.
29. Ohashi, J., Naka, I., Ohtsuka, R., Inaoka, T., Ataka, Y., Nakazawa, M., Tokunaga, K. and Matsumura, Y. (2004) Molecular polymorphism of ABO blood group gene in Austronesian and non-Austronesian populations in Oceania. *Tissue Antigens*, **63**, 355–361.
30. Ohashi, J., Naka, I., Tokunaga, K., Inaoka, T., Ataka, Y., Nakazawa, M., Matsumura, Y. and Ohtsuka, R. (2006) Brief communication: Mitochondrial DNA variation suggests extensive gene flow from Polynesian ancestors to indigenous Melanesians in the northwestern Bismarck Archipelago. *Am. J. Phys. Anthropol.*, **130**, 551–556.
31. Ohtsuka, R., Kawabe, T., Inaoka, T., Akimichi, T. and Suzuki, T. (1985) Inter- and intra-population migration of the Gidra in lowland Papua: a population-ecological analysis. *Hum. Biol.*, **57**, 33–45.