Yamaguchi, K, Watanabe, C, Kawaguchi, A, Sato, T, Naka, I, Shindo, M, Moromizato, K, Aoki, K, Ishida, H and Kimura, R

Association of melanocortin 1 receptor gene (MC1R) polymorphisms with skin reflectance and freckles in Japanese.

Original article

Title: Association of melanocortin 1 receptor gene (*MC1R*) polymorphisms with skin reflectance and freckles in Japanese

Authors and Institutions:

Kyoko Yamaguchi[1,2*], Chiaki Watanabe[1], Akira Kawaguchi[2], Takehiro Sato[2], Izumi Naka[3], Misaki Shindo[2], Keiichi Moromizato[2], Kenichi Aoki[4], Hajime Ishida[2], Ryosuke Kimura[1*]

[1] Transdisciplinary Research Organization for Subtropics and Island Studies, University of the Ryukyus, Okinawa, Japan

[2] Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan

[3] Department of Molecular and Genetic Epidemiology, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan

[4] Department of Biological Sciences, The University of Tokyo, Tokyo, Japan

Total Number of Pages: 36

Number of Figures: 3

Number of Tables: 3

Number of Supplementary Figures: 6

Number of Supplementary Tables: 5

Abbreviated title: MC1R and skin reflectance in Japanese

**Corresponding Authors:*

Kyoko Yamaguchi

Faculty of Medicine, University of the Ryukyus

Uehara 207, Nishihara, Nakagami, Okinawa 903-0215, JAPAN

Phone: (098)895-1102

Email: kyoko-y@umin.ac.jp

Dr. Ryosuke Kimura

Transdisciplinary Research Organization for Subtropics and Island Studies, University of the Ryukyus

Uehara 207, Nishihara, Nakagami, Okinawa 903-0215, JAPAN

Phone:(098)895-1121

Email: rkimura@lab.u-ryukyu.ac.jp

ABSTRACT

Most studies on the genetic basis of human skin pigmentation have focused on people of European ancestry, and only a few studies have focused on Asian populations. We investigated the association of skin reflectance and freckling with genetic variants of *Melanocortin 1 Receptor* (*MC1R*) in Japanese. DNA samples were obtained from a total of 653 Japanese individuals (ages 19-40) residing in Okinawa; skin reflectance measured using a spectrophotometer and freckling status were determined for each individual. A lightness index (L*) and freckling status were not correlated with age, BMI, or ancestry (Ryukyuan or Main Islanders of Japan). Among the 10 nonsynonymous variants that were identified by direct sequencing of the coding region of *MC1R*, two variants—R163Q and V92M—with the derived allele frequencies of 78.6% and 5.5%, respectively, were most common. Multiple regression analysis showed that allele 163Q and the presence of nonsynonymous rare variants (allele frequencies <5%) were significantly associated with an increase in sex-standardized skin lightness (L* of CIELAB) of the inner upper arm. Relative to the 92V allele, the 92M allele was significantly associated with increased odds of freckling. This is the first study to show an association between the 163Q allele and skin reflectance values; this association indicated that light-toned skin may have been subject to positive selection in East Asian people.

Keywords: skin pigmentation/freckles/*MC1R*/Japanese.

INTRODUCTION

The human body is enveloped by skin, which serves as the interface for interaction between the body and its environment. Skin forms a barrier against microbial invasion and protects the body from damage caused by chemicals, heat, and ultra violet (UV) radiation. One important role of human skin is to control body temperature via sweating and subsequent heat loss.[11]Another role of skin is to mediate vitamin D (VD) synthesis when sun exposure is adequate. [2] As the interface between the body and the environment, all of these functions of the skin may arguably be regarded as adaptations to the different types of environments surrounding our ancestors, shaping the geographical variation in skin characteristics, including pigmentation in modern human.[3, 4] The apparent geographical variation in skin pigmentation has attracted both public and academic attention, and has been studied in many research areas such as evolutionary biology, clinical medicine, health science, physiology, and molecular genetics[5-8], as well as sociology and psychology.[9-11]

Various hypotheses have been proposed to explain the evolution of skin pigmentation; we summarize these hypotheses as follows. Dark skin is believed to be favored by natural selection near the equator where UV radiation is high because darker skin protects the body from over-heating and sunburns[12], as well as from UV-induced damage to DNA and folic acid.[3] At high latitudes including Europe, light skin may be selected for its ability to allow synthesis of sufficient VD even under

conditions of low UV radiation, in contrast, dark skin may lead to increased susceptibility to rickets at high latitudes.[2, 13] Jablonski and Chaplin (2000) reported a correlation between the amount of cutaneous melanin and the UV radiation levels among indigenous populations; this correlation was explained by the balance between the physiological requirements for VD synthesis and folate protection from UV.[4] Another explanation for the depigmentation of skin is neutral evolution; specifically, the reduction in exposure to UV radiation for populations that migrated away from the equator has removed the constraints maintaining dark skin.[14] However, recent findings from molecular and theoretical genetics indicate that random genetic drift alone is not sufficient to account for the rate of evolution of light skin. A long time before these proposals, Darwin (1871) hypothesized that the diversity of human skin color had arisen by sexual selection through a process of mate choice especially in high-latitude areas.[15] Therefore, sexual selection, as well as natural selection, has to be considered in investigating the evolution of human skin pigmentation.[16, 17]

Human skin reflectance is a quantitative genetic trait with a continuum from dark to light. Traditionally, color measuring studies on humans have used a set of color panels or color categories such as von Luschan's chromatic scale and Broca's color panel.[12, 18, 19] Later, the Fitzpatrick scale was introduced to categorize the sun-tanning response of skin into six categories.[20] As technology advanced, the color panels were replaced by spectrophotometers.[21, 22] Questionnaire-based evaluation such as self-reported categories (fair/pale, medium, or olive/dark) and Fitzpatrick scale

are still used even in genome-wide association studies (GWASs) focused on people of European descent.[23] However, methods that depend upon arbitrary categories for skin color or the Fitzpatrick scale may not be valid for East Asians; for example, all of the participants in the study by Motokawa *et al*. (2006) were sorted into the "medium" category and skin types II to IV.[24] Therefore, spectrophotometers, which yield objective and consistent color measurements, would be more suitable for measuring the skin reflectance of East Asians.

The *Melanocortin 1 receptor* (*MC1R*, MIM 15555) gene is very important in human skin pigmentation. *MC1R* is located on human chromosome 16q24.3, and it encodes a single exon of 954 bp and a protein of 317 amino acids.[25] The protein MC1R is a G-protein coupled seven transmembrane receptor for α-melanocyte simulation hormone (α-MSH). When α-MSH binds to the amino acid sequence His-Phe-Arg-Trp (TFRW) of MC1R, which is located on the extracellular surface a melanocyte, subsequent activation of adenylate cyclase facilitates accumulation of cAMP inside the melanocyte; this accumulation ultimately promotes eumelanin (dark pigment) synthesis.[26, 27] Therefore, MC1R activation plays a key role in skin pigmentation because it results in a switch from pheomelanin (light and immature pigment) synthesis to eumelanin synthesis.[28]

Molecular evidence indicates 1) that *MC1R* alleles underwent purifying or negative selection in Africa where UV radiation is high throughout the year, 2) that relaxation of constraint drove neutral evolution of *MC1R* in high-latitude regions such as Europe[29], and 3) that certain alleles

possibly underwent directional selection.[30] Therefore, the coding region of *MC1R* is highly conserved in Africans, while over 30 nonsynonymous and synonymous variants have been reported in populations of European ancestry.[31] In Asian populations, more than 10 variants have been reported, of which the V92M and R163Q variants are observed at high frequencies relative to the frequencies observed in non-Asian populations.[32-34] Recent genome scans for selective sweeps have revealed that strong positive selection has acted on pigmentation-related genes.[35-37] Coop *et al*. (2009) suggested a signature of an "East Asian sweep" on R163Q (rs885479) based on F_{ST} and XP-EHH in genome-wide SNP data from the HapMap, the CEPH-Human Genome Diversity Panel, and Perlegen dataset.[38]

The functions of the two variants, V92M and R163Q, have been studied mainly with regard to freckles and lentigines in Asians and to red hair, pale skin, and freckling phenotypes in Europeans. A study of a Japanese population reported that the 92M and 163R alleles are associated with solar lentigines and freckles.[39] In people of European descent, three alleles—V60L, V92M, and R163Q—are reportedly low-penetrance red hair color variants designated "*r*", and four alleles—D84E, R151C, R160W, D294H—designated "*R*" are high-penetrance red hair color variants.[40] The effect of *R* alleles accounted for the genome-wide significant association between SNPs in a region around *MC1R* and hair color or tanning ability of skin.[23, 41] In a study by Duffy *et al.* (2004), *r* alleles were associated with a 0.9% increase in skin reflectance on the inner upper arm,

but the increase was smaller than that associated with *R* alleles (1.9%).[42] In Asians, the associations of V92M and R163Q with skin reflectance have rarely been studied. An exception is a study on a Tibetan population; the main effects of these two variants were not significant, but an interaction between V92M and a SNP in *OCA2* was associated with an increase in lightness (L*) of skin on the inner upper arm.[43]

The association of *MC1R* variants with skin reflectance still remains unclear in most East Asian populations including the entire Japanese population; however, some researchers have suggested that positive selection has acted on *MC1R* alleles in East Asian populations.[38] Therefore, we investigated the effect of *MC1R* variants on skin pigmentation in a large sample of Japanese people by analyzing association between *MC1R* variants and skin reflectance objectively quantified with a spectrophotometer as well as freckling.

MATERIALS AND METHODS

Subjects

The subjects consisted of 653 healthy volunteers (364 males and 289 females). The age of subjects ranged from 19 to 40 years old (Mean=23.62, SD=4.62). Informed consent was obtained from each

individual. This study was approved by the research ethics committee of the University of the Ryukyus. The participants were asked to voluntarily disclose their sex, age, and the place of birth of their four grandparents. In this study, “Ryukyu Islands” refers to the southernmost islands in Japan, which include the Okinawa Islands, Sakishima Islands, and Amami Islands. The rest of Japan archipelago was defined as “Main Islands of Japan” (Figure 1). Ryukyu index was defined as the number of grandparents who are from the Ryukyu Islands, ranging from 0 (none of the grandparents were from the Ryukyu Island) to 4 (all four grandparents were from the Ryukyu Islands). The body mass index (BMI) of each subject was calculated from body height and weight, which were measured with an anthropometer and a portable digital scale, (mean=21.58, SD=3.10 for females; mean=23.20, SD=3.61 for males).

Skin reflectance

Skin reflectance of each subject was measured using a spectrophotometer (Konica Minolta CM-600d) between June 2009 and March 2011. This device measures the reflectance of an object with a light-receiving component and converts the data into various color systems such as CIE L*a*b* and spectral curve. We took the measurement on the inner upper arm of each subject under resting conditions. Each participant extended their left arm so that the inner side faced upward, to avoid error due to the arm positioning[44], and the midpoint between the armpit and the elbow was

measured.

Evaluation of Freckling

Freckles are small (1-3 mm) light brown macules on the face, arms, and back; freckles become pronounced by sun exposure and fade during winter.[45, 46] In Japanese people, freckles are dark brown and can be as large as 5 mm in diameter.[47] We evaluated freckles on a head shot of each participant, and considered freckles to be “present” in individuals with multiple macules on the nasal dorsum and upper cheek, and to be “absent” in individuals without macules on the nasal dorsum.

Genotyping and Haplotyping

Genomic DNA was extracted from a blood or saliva sample from each participant using the Gentra Puregene Blood Core Kit or the QIAamp DNA Blood Mini Kit (QIAGEN), following the manufacture’s protocols. Genotyping was performed using the direct-sequencing method. We designed primers (listed in Table S1) by using a free online program, Primer3, and referring to the GenBank sequence NG_012026. Nested PCR was performed to amplify the open reading frame of *MC1R* in a 20-μL total volume containing 0.8 units of FastStart Taq DNA polymerase (Roche), 1×PCR buffer + $MgCl_2$, and 200 μM of each dNTP. The cycling parameters were as follows: 95ºC×1 min, (95ºC×30 sec, 60ºC×30 sec, 72ºC×90 sec) for 40 cycles, 72ºC×7 min, and 20ºC for

storage. We used a different enzyme for the nested PCR of some samples; these reactions were performed in a 20-µL total volume containing 0.5 units TaKaRa ExTaq DNA polymerase (Takara Biotechnology), 1×PCR buffer, and 200 µM of each dNTP. The cycling parameters in this case were as follows: 94ºC×1 min, (98ºC×10 sec, 55ºC×30 sec, 72ºC×60 sec) for 40 cycles, 72ºC×1 min, and 4ºC for storage. The PCR products were sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on either an ABI PRISM® 310 or 3100 Genetic Analyzer (Applied Biosystems); the sequences were then analyzed with CodonCode Aligner (CodonCode Corporation). *MC1R* haplotypes were estimated with PHASE version 2.1[48, 49] then a haplotype network was drawn using the median-joining method with the software, NETWORK (fluxus-engineering.com).[50]

Selective neutrality of *MC1R* coding region was tested using Tajima's *D* with the software Arlequin ver. 3.5.1.3.[51, 52] The significance of the Tajima's *D* was tested by the *P*-value obtained as the proportion of simulated *D* less than or equal to the observation in 1000 simulations using a coalescent simulation algorithm (one-sided test).

Statistical Analyses

The statistical analyses were performed using IBM® SPSS® Statistics version 19. We adopted a significance level of 0.05 unless otherwise stated. For multiple testing, the false discovery rate

(FDR) was controlled at a level of 0.05 according to the method suggested by Benjamini and Hochberg (BH method), which has been found to be effective for quantitative trait loci (QTL) analysis.[53, 54]

In the sex-stratified analysis, the association between L* values and genotypes were tested separately for females and males. Then, the results for females and males were combined by meta-analyses using two methods: the Fisher's r-to-z transformation method that combines sex-specific standardized regression coefficients, and the inverse normal method that pools multiple *P*-values independently of effect sizes. [55]

In addition, L* values were adjusted for sex by standardizing them into z-scores separately for males and females. Multiple regression analyses were conducted to examine associations between *MC1R* variants and the z-scores of L* values on inner upper arms (ZL*), and covariates were controlled for in these analyses.

For the variants with allele frequencies no less than 0.05, genotypes were coded as the number of derived alleles (AA=0, AD=1, DD=2; A=ancestral allele, D=derived allele) assuming the additive (co-dominant) model where having DD rather than AA is twice as likely to affect the L* as is having AD rather than AA. The nonsynonymous variants with frequencies below 0.05 were combined as "rare variants", and a variable, designated RARE, was defined as the number of rare variants carried by the individual, since the sample size of each rare variant was too small to be

separately tested. We further examined the dominant model where having the genotype AD or DD rather than AA is equally likely to affect the L* (genotype coding: AA=0, AD=1, DD=1), and the recessive model where only DD affects the L* (AA=0, AD=0, DD=1).

The covariates included age (AGE), ancestry (Ryukyu Index), and BMI as well as the month when each measurement took place. The 212 females and 219 males were of Ryukyu index=4, and 43 females and 95 males were of Ryukyu index=0. The other categories for Ryukyu index had small sample sizes (2, 47, 19 for Ryukyu Index=1, 2, and 3, respectively; 16 participants had at least one grandparent with an unknown birthplace). For month, a dummy variable JAN was set equal to 1 if the measurement was made in January (n=19) and equal to 0 otherwise. Similarly, six additional dummy variables JUN, JUL, AUG, OCT, and NOV (n=60, 110, 58, 28, and 108, respectively) were defined in a manner identical to the JAN variable, setting measurements made in December as baseline because the sample size was the largest (n=269). No measurement was taken in February, April, May, and September. Only one individual was measured in March, so we included it in January.

The relationship between these variables and ZL* were examined with Spearman's rank correlation, t-test, Mann-Whitney U test, Kruskal-Wallis test, and multiple regression. In the multiple regression analysis, we fitted the data to a model

$$ZL^* = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + E$$

where X_1, X_2,…,X_k are independent variables. Multicollinearity was assessed by variance inflation factor (VIF). The best model was determined using a stepwise regression procedure with the probability-of-*F*-to-enter ≤0.05 and the probability-of-*F*-to-remove ≥0.10. We also tested the involvement of interaction that is described as a product term of two independent variables, such as X_1X_2, in the model. Analyses that included L* values outliers basically yielded the same results as analyses that excluded these outliers; therefore, we included the outliers to detect the effect of variants on the extreme ends of data.

In advance of the association analysis of *MC1R* with freckling, a Spearman's rank correlation coefficient (rho) was calculated between freckling and covariates of two continuous variables, age and BMI. For the categorical variables, including SEX, Ryukyu index, and month, a Pearson's chi-square test for categorical variables was conducted. The odds ratio (OR) for freckling of the minor allele at each SNP relative to the sample consensus allele was calculated. A 95 % confidence interval for each OR was obtained by OR±1.96SE, to test whether the OR is significantly different from 1. OR values that are not significantly different from 1 indicate that there was no association between the allele and freckling. For the logistic regression, we fitted the data to a model

$$\ln[\text{odds}(\text{freckle}=\text{present})] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + E$$

where X_1, X_2,…, X_k are independent variables. The same stepwise criteria used for the linear multiple regression analysis described above were used to select the best model. The additive

(co-dominant) model was assumed, and the genotypes were treated as continuous variables. The OR for X_k was given by the exponential function, $\exp(\beta_k)$.

We also conducted association analyses of the 92M and 163Q alleles with skin reflectance and freckles, following the procedure used in Duffy *et al*. (2004). [42] We defined a variable, designated *r*, as the total number of the 92M and 163Q alleles. The percentage reflectance at wavelength 650nm was used as the dependent variable instead of ZL*.

RESULTS

Mean L* values were 67.07 and 64.58, for females and males respectively. The equality of variances was rejected (SD=2.47 for female, 3.30 for male, F=15.27, *P*=1.03E-04), and a t-test for the equality of means showed a significant difference between sexes in mean L* values (equal variance not assumed, *P*=4.71E-26). Therefore, L* values were standardized for each sex and pooled for the following analyses.

We then examined the effect of covariates on ZL*, in advance of the multiple regression. AGE and BMI were not correlated with ZL* (AGE: Spearman's rho=0.09, *P*=0.818; BMI: rho=0.038, *P*=0.334), and hence excluded from the regression. We compared Ryukyuans (Ryukyu index=4) and Main Island Japanese (Ryukyu index=0), and found no significant difference in ZL*

between the two groups (t-test, equal variances assumed, P=0.855). Therefore, Ryukyu index was also excluded from the multiple regression model. When ZL* was regressed on the dummy variables relating to month, two such variables, JUN and JUL, were included in the best model as defined using the stepwise method; therefore, these two dummy variables were combined as JUN_JUL to be included in the subsequent regression analysis as a covariate.

As a result of sequencing the coding region of *MC1R*, we found 20 variants, 10 of which were nonsynonymous mutations; the rest were synonymous mutations. A single synonymous and two nonsynonymous variants were observed at derived allele frequencies higher than 5 % (5.5% for V92M, 78.6% for R163Q, and 8.2% for 942A>G). There was no significant deviation from Hardy-Weinberg equilibrium. The remaining nonsynonymous variants were observed only in the heterozygous state, and summed into the dummy variable RARE, which took either of two values 0 or 1 (Table 1). We did not find the variants that were strongly associated with pigmentary traits in the studies targeting people of European descendent.[42, 56, 57] . Figure 2 shows the haplotype network based on the estimated haplotypes of the sample. As shown in the largest circle with many branches, the majority of the estimated haplotypes had the 163Q allele. For the test of selective neutrality, Tajima's D was calculated as -1.66 with *P*-value of 0.015, which suggested a deviation from neutrality and the possibility of purifying or positive selection on *MC1R*.

Stepwise multiple regression analysis with genetic factors determined the best model

$$\widehat{ZL^*} = -0.29 + 0.17(\text{R163Q}) + 0.43(\text{RARE}),$$

where the effect of 163Q allele and rare variants were significant (P=0.016 and P=0.018, respectively) (Figure 3) but V92M and 942A>G were not significant (P=0.093 and 0.067, respectively). The interaction term between R163 and RARE (R163Q×RARE) was insignificant (P=0.67) in the model. R163Q and RARE remained significant even after controlling FDR by BH method with the second lowest P-value being less than 0.025 (=0.05×2/4) although they were not significant after Bonferroni correction (significance level=0.05/4=0.0125). The adjusted R^2 was 0.012, therefore, the model explained only 1.2% of the total variability in L* of inner upper arm for L* values that had been standardized by sex.[55]

A sex-stratified analysis showed that the regression coefficients for R163Q and RARE were 0.127 (P=0.036) and 0.085 (P=0.158) in females, while 0.095 (P=0.081) and 0.101 (P=0.064) in males (Table S2). Although the regression coefficients were insignificant in the sex-stratified analysis except for R163Q in females (P=0.036), the meta-analysis showed that the pooled standardized coefficients were significant for R163Q (P=0.0059) even after Bonferroni correction and for RARE (P=0.018) even after controlling FDR by BH method (Table S3).[53, 55] Pooled P-values using the inverse normal method also supported the significance (R163Q: P=0.0033; RARE: P=0.011) even after Bonferroni correction as well as controlling FDR (Table S3).[55]

When the month variables were added, stepwise multiple regression analysis determined

the best model to be

$$\widehat{ZL^*} = -0.42 + 0.57(\text{JUN\_JUL}) + 0.16(\text{R163Q}) + 0.39(\text{RARE}).$$

The dummy variable for month (JUL_JUN), the number of 163Q alleles (R163Q), and the presence of rare variants (RARE) remained to have significant coefficients (*P*=3.14E-10, 0.016 and 0.026, respectively) (Table 2). The adjusted R^2 was 0.072, indicating that the best model explained 7.2% of the total variability in ZL*. Interaction terms between each variable (JUN_JUL×R163Q, JUN_JUL×RARE, and R163Q×RARE) were not significant when added to the best model (*P*=0.804, 0.832, and 0.418, respectively). Although V92M was excluded from the best model, the regression coefficient was 0.056 (*P*=0.196) when added in the model.

The model above was developed based on the assumption of the additive (co-dominant) model for the R163Q variant. We next examined the dominant and recessive models. For the dominant model, the stepwise regression analysis determined the best model with only JUN_JUL as significant descriptive variables. For the recessive model, the best model determined by the stepwise multiple regression analysis was

$$\widehat{ZL^*} = -0.29 + 0.57(\text{JUN\_JUL}) + 0.20(\text{R163Q}) + 0.39(\text{RARE})$$

with the adjusted R^2 of 0.073, indicating a fit as good as the additive model (Table 2).

Based on the estimated haplotypes, we divided the rare variants into two groups, 163R+rare and 163Q+rare, depending on the allele at R163Q located on the same haplotype as the

nonsynonymous rare variants. Only one haplotype, the one bearing 120T, was contained in 163R+rare; in contrast, seven different haplotypes were contained in 163Q+rare. When we conducted a stepwise multiple regression analysis with 163R+rare and 163Q+rare instead of RARE, neither remained in the best model (P=0.141 for 163R+rare and P=0.067 for 163Q+rare). Pair-wise comparison by Mann-Whitney's *U*-test showed a significant difference only between "no rare variant" and 163Q+rare at the genotype 163R/Q (Figure 3).

In the multiple regression analysis with *r*, the total number of the 92M and 163Q alleles, we obtained the following best model

$$\widehat{650NM} = 50.08 + 1.82(\text{JUN\_JUL}) - 1.45(\text{SEX}) + 0.75(r) + 1.26(\text{RARE})$$

where 650NM was a percentage reflectance at wavelength 650nm, SEX was coded as 0 for females and 1 for males (*P*-values for *t*-test on each variable in the model: 7.28E-09, 1.29E-07, 3.76E-03, and 3.74E-02, respectively). The linear regression analysis with *r* as the only independent variable showed the regression coefficient of 0.65 (P=1.18E-02, 95% CI =0.15-1.15).

None of the covariates showed a significant correlation with freckling. The Chi-square test indicated that freckles occurred independent of SEX (Chi-square=1.45, df=1, P=0.228) and Ryukyu Index (only 0 and 4, Chi-square=3.41, df=1, P=0.065). The t-test showed the equal means of AGE and BMI in the two freckle groups, "present" and "absent". The ORs for freckling for the minor alleles relative to the sample consensus alleles are shown in Table 3. Only V92M showed a

significant change in the odds for freckling; individuals with the 92M allele had the OR of 3.11 relative to individuals who are 92V homozygotes (95% confidence interval (CI) =1.18-8.17). Similarly, only V92M exhibited significant coefficient (P=0.030) in the logistic regression model

$$\ln[\widehat{odds(freckle)}] = -3.48 + 1.00(\text{V92M})$$

with neither R163Q, 942A>G, nor RARE contributing significantly when added. The OR for freckling of the 92M allele was obtained as an exponential value of the regression coefficient (1.00), which was 2.72 (95% CI=1.10-6.71). In the logistic regression model with *r* as a descriptive variable, the regression coefficient for *r* was 0.95 (P=0.100) which is equivalent to OR of 2.58 (95% CI=0.83-7.96) and not significant.

DISCUSSION

It has been said that Ryukyuan had dark skin compared to Main Island Japanese. In an anthropological study on Okinawan people, Suda (1940) concluded that people from Okinawa Island had darker skin on the forehead than people from Kagoshima (Figure 1) when assessed with Broca's color palette.[58] On the other hand, Torii (1904) found no difference in skin color on the inner upper arm as defined by Broca's color palette between 77 female students from Okinawa and 30 from the Main Islands of Japan, all of whom resided in Okinawa Island.[59] The reason he stated for the

similarity in skin color was that their skin was protected from sunlight, covered with clothes with sleeves, which was privilege for the middle to upper class people at that time. Similarly, our study did not find statistically a significant difference in the lightness of skin at the inner upper arm between peoples of Ryukyuan and Main Island Japanese descent for those living in the same environment, Okinawa Island. Taken together, these studies indicate that the apparent difference in skin tone between Okinawan and Main Islanders might not be congenital in origin, but more likely be due to the difference in UV radiation levels between Okinawa and the Main Islands. This conclusion is reinforced by data of the Japan Meteorological Agency that Naha in Okinawa Island is exposed to higher UV radiation level throughout the year than Kagoshima in Kyushu Island, one of the Main Islands; average of daily cumulative standard erythema dose (DCSED) between 1994 and 2008 was 2.84 kJ/m^2 in Naha, and average DCSED from 1997 to 2004 was 2.41 kJ/m^2 in Kagoshima.

A sexual dimorphism in skin reflectance was clearly observed in our data; the mean for females was higher than that for males. This pattern was consistent with the pattern widely reported in previous studies.[4, 17, 60] That females show a lower variance in L* than males would also indicate either usage of sunscreen or avoidance of exposure to UV among females; however, the reflectance values from the inner upper arm are not supposed to be strongly affected by sun exposure.[12] This finding may reflect the trend among Japanese females living in Okinawa to desire a light skin tone,

as documented in other cities in East Asia.[61] We found no significant association between skin reflectance and age, which may be explained by the majority of the participants being in early adulthood.

As a result of the regression analysis with the month variables, we found that JUN and JUL were associated with a lighter skin tone. This could be explained by a time lag between the UV exposure and tanning response, and between tanning and depigmentation. According to the data of the Japan Meteorological Agency, a monthly average of daily cumulative standard erythema dose (DCSED) in Naha is the highest in July (Figure S1). It is possible that the skin of inner upper arms would accumulate melanin towards the end of high UV months and would remain pigmented until it slowly fades out.

Association analysis between *MC1R* variants and skin reflectance showed that the 163Q allele had the effect of increasing the lightness of skin at inner upper arm in an additive or recessive mode. Duffy *et al.* (2004) suggested an additive mode for the effect of *r* on skin reflectance of Europeans at wavelength 650nm, and the regression coefficient for *r* (0.9 with 95% CI=0.2-1.6) was similar to that found in the present study (0.7 with 95% CI=0.20-1.20).[42] In the multiple regression model with only *MC1R* variants, R163Q and RARE, the adjusted R^2 was 0.012, indicating that the model explained 1.2% of the total variability in L* of inner upper arm for L* values that were standardized by sex. This finding is consistent with our expectation that each variant contributes a

small amount to a quantitative trait such as skin pigmentation. Examination of the contribution of genes other than *MC1R* to skin pigmentation in Asian populations is awaited.

We also found a significant association between V92M and freckling in a logistic regression (OR=2.72, 95% CI=1.10-6.71); this finding was consistent with findings from a study by Motokawa *et al.* (2007) (OR=3.08, 95% CI=1.35-6.56).[39] However, the association between freckling and R163Q was not significant in our study (OR=1.12, 95% CI=0.54-2.4), a finding that was not consistent with their results (OR=2.08, 95% CI=1.01-4.12). In addition, when V92M and R163Q were combined as *r*, our logistic regression analysis did not show a significant effect of *r* on freckling although Motokawa *et al*. (2007) and Duffy *et al*. (2004) found it significant.[39, 42] Further examination is required to conclusively determine the effect of R163Q on freckling.

The haplotype network we constructed had a star-like composition of haplotypes surrounding the R163Q variant.[62] The negative value (-1.66) obtained for Tajima's *D*, showing an excess of observed rare variants relative to the expectation, indicated a possibility of recent demographic expansion or positive selection. Voight *et al.* (2005) obtained an average Tajima's *D* of 0.18 for Han Chinese (n=15) based on the resequencing data of 50 unlinked autosomal noncoding regions.[63] Their result suggests that demographic history in East Asia has shifted the *D* from zero toward the positive direction, which is the opposite direction of our *D*; therefore, the value we obtained for *MC1R* cannot be explained by demography. In addition, the star-like composition was

observed only at R163Q; therefore, this pattern may support that the frequency of 163Q alleles has increased recently. Thus, our result indicated the involvement of selection on the 163Q allele; this conclusion is consistent with a finding from a study by Coop *et al.* (2009) that showed a selective sweep on the same allele.[38] The present study showed an association between the 163Q allele and light skin tone in East Asia; this association indicated positive selection on light-toned skin among East Asians.

Although our findings indicated that positive selection on *MC1R* had occurred, it is difficult to identify the force of selection. When considering sexual selection on the light-toned skin, whether the change in skin reflectance is discernible or not is important. The amount of change in L* predicted by our multiple regression model is 0.82 (2×ZL*×2.47) for females and 1.08 (2×ZL*×3.30) for males, comparing 163Q homozygotes (163Q/Q) to 163R homozygotes (163R/R). A similar amount of change is predicted for the nonsynonymous rare variants (0.97 for females and 1.30 for males). These differences are distinguishable by eye if the two shades are placed close to each other, according to the definition adopted by CIELAB, the color space we used.[64] Psychological studies on perception of skin color would help in understanding the relationship between the statistically significant change in L* of skin and discernible change in lightness.

Population structures in the Japanese archipelago have been observed, especially between Ryukyu and Main Island Japanese.[65] As we found no significant difference in neither skin

reflectance nor freckling between Ryukyuans (Ryukyu index=4) and Main Island Japanese (Ryukyu index=0), genetic factors that strongly affect pigmentation are unlikely to be highly differentiated between these subpopulations. To further investigate the effects of population stratification on our association study, we examined other SNPs (rs3827760, rs853975, rs350886, rs1048610) that are differentiated among populations across the world in a similar manner to *MC1R* R163Q (Figures S2-S6), and found no association of these SNPs with pigmentary traits (Table S4). Therefore, population stratification, if any, had only a small effect on our association study.

Our results suggested that two major nonsynonymous variants, V92M (rs2228479) and R163Q (rs885479), had different influence on skin phenotype; V92M was associated with freckling while R163Q was associated with the lightness of skin. These two variants are located in different domains of MC1R; V92M is in the second transmembrane domain of 7 transmembrane domains, whereas R163Q is in a cytoplasmic loop. SIFT, a software to predict the effect of amino acid substitution on protein function based on the conservativeness of amino acid in the closely related sequences, predicted that V92M and R163Q would not damage the function of MC1R with scores of 0.68 and 0.21, respectively.[66] According to PolyPhen-2, another tool that predicts impact of amino acid substitution on the structure of proteins and the subsequent functional change, both of the 163Q and 92M mutations were predicted to be benign with scores of 0.015 and 0.004, respectively, whereas the rare variants were predicted to be “possibly damaging” or “probably damaging” (scores

ranged from 0.656 to 0.999) except for V140M (score=0.390) (Table S5).[67] This is consistent with the finding that the effect of RARE on skin reflectance was twice as much as R163Q in our multiple regression analyses. However, an *in vitro* study showed that the 92M allele caused impaired MC1R function; in contrast, the 163Q allele had almost the same response to α-MSH as the human consensus allele.[32] These results may explain the difference in the effect of the V92M and R163Q alleles on skin pigmentation.

As discussed above, the 163Q allele does not cause any change in the MC1R ligand binding amino acid sequence; hence, the allele may retain a normal ability to promote eumelanin (dark pigment) synthesis through activation of adenylate cyclase and the subsequent accumulation of cAMP in a melanocyte cell. An *in vitro* study on the function of *MC1R* variants also indicates that 163Q functions normally when compared to the consensus.[32] Motokawa *et al*. (2008) investigated polymorphisms in the promoter region of *MC1R* linked to R163Q among 247 Japanese.[68] They found that the promoter haplotypes with a combination of -490T, -445A, and -226T, 98.5% of which bore the 163Q allele in the coding region, showed lower OR for initiating freckles and solar lentigines than did the human consensus haplotype (-490C, -445G, -226A). They suggested that the -490C>T variant might be responsible for the decreased promoter activity, based on the finding by Moro *et al*. (1999) that a deletion from -517 to -447 damages the promoter activity.[69] Therefore, impaired function in the promoter region, caused by the variant(s) in linkage disequilibrium with

R163Q, may explain the association between the 163Q allele and the lightness of skin, which we found in this study. Further functional studies are needed to reveal the mechanism.

In summary, we found that the 163Q allele contributed to the lightness of skin in Japanese people. The effect of 92M in increasing the odds of freckling was also confirmed. These results explain a part of the difference in constitutive skin pigmentation between Asian and non-Asian populations as well as the north to south gradient in pigmentation across Asia. Association studies on admixed populations or comparative studies in allele frequencies among different populations are required to further elucidate the evolution of skin pigmentation among the peoples of Asia.

ACKNOWLEDGEMENTS

We thank the participants of this study and those who helped to collect samples and data. This study was partly supported by the Rising Star Program at University of the Ryukyus and a Grant-in-Aid for Young Scientists (A) and for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to RK).

References

1. Gray, H., Standring, S., Ellis, H. & Berkovits, B. K. B. *Gray's anatomy: the anatomical basis of clinical practice*. 39th edn. (Elsevier/ Churchill Livingstone: Edinburgh, 2005).

2. Murray, F. G. Pigmentation, sunlight, and nutritional disease. *Amer. Anthrop.* **36**, 438-445 (1934).

3. Jablonski, N. G. The evolution of human skin and skin color. *Annu. Rev. Anthrop.* **33**, 585-623 (2004).

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

5. King, R. & Summers, C. Albinism. *Dermatol. Clin.* **6**, 217 - 228 (1988).

6. Parra, E. J. Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am. J. Phys. Anthropol.* **Suppl 45**, 85-105 (2007).

7. Relethford, J. H. Hemispheric difference in human skin color. *Am. J. Phys. Anthropol.* **104**, 449-457 (1997).

8. Sturm, R. A. Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* **18**, R9-R17 (2009).

9. Frost, P. European hair and eye color - A case of frequency-dependent sexual selection? *Evol.*

Hum. Behav. **27**, 85-103 (2006).

10. van den Berghe, P. L. & Frost, P. Skin color preference, sexual dimorphism and sexual selection: A case of gene culture co-evolution? *Ethn. Racial Stud.* **9**, 87-113 (1986).

11. Wagatsuma, H. The social perception of skin color in Japan. *Daedalus.* **96**, 407-443 (1967).

12. Robins, A. H. *Biological perspectives on human pigmentation*. (Cambridge University Press: Cambridge, 1991).

13. Loomis, W. F. Skin-pigment regulation of Vitamin-D biosynthesis in man. *Science.* **157**, 501-506 (1967).

14. Brace, L. Structural reduction in evolution. *Am. Nat.* **97**, 39-49 (1963).

15. Darwin, C. *The decent of man, and selection in relation to sex*. (John Murray: London, 1871).

16. Jones, D. C., Brace, L., Jankowiak, W., Laland, K. N., Musselman, L. E., Langloris, J. H., et al. Sexual selection, physical attractiveness, and facial neoteny: Cross-cultural evidence and implications [and comments and reply]. *Curr. Anthrop.* **36**, 723-748 (1995).

17. Aoki, K. Sexual selection as a cause of human skin colour variation: Darwin's hypothesis revisited. *Ann. Hum. Biol.* **29**, 589-608 (2002).

18. Broca, P. *Instructions générales pour les récherches anthropologiques à faire sur le vivant*. (Masson: Paris, 1879).

19. von Luschan, F. *Beiträge zur Völkerkunde der Deutschen Schutzgebieten*. (Deutsche

Buchgemeinschaft: Berlin, 1897).

20. Fitzpatrick, T. B. Soleil et peau. *Journal de Médecine Esthétique* **2**, 33-34 (1975).

21. Lasker, G. W. Photoelectric measurement of skin color in a Mexican Mestizo population. *Am. J. Phys. Anthropol.* **12**, 115-122 (1954).

22. Garrard, G., Harrison, G. A. & Owen, J. J. T. Comparative spectrophotometry of skin colour with EEL and photovolt instruments. *Am. J. Phys. Anthropol.* **27**, 389-395 (1967).

23. Han, J. L., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4**, e1000074 (2008).

24. Motokawa, T., Kato, T., Hongo, M., Ito, M., Takimoto, H., Katagiri, T., et al. Characteristic MC1R polymorphism in the Japanese population. *J. Dermatol. Sci.* **41**, 143-145 (2006).

25. Gantz, I., Yamada, T., Tashiro, T., Konda, Y., Shimoto, Y., Miwa, H., et al. Mapping of the gene encoding the melanocortin-1 (alpha-melanocyte stimulating hormone) receptor (MC1R) to human chromosome 16q24.3 by Fluorescence in situ hybridization. *Genomics.* **19**, 394-395 (1994).

26. Holder, J. R., Bauzo, R. M., Xiang, Z., Scott, J. & Haskell-Luevano, C. Design and pharmacology of peptoids and peptide-peptoid hybrids based on the melanocortin agonists core tetrapeptide sequence. *Bioorg. Med. Chem. Lett.* **13**, 4505-4509 (2003).

27. Garcia-Borron, J. C., Sanchez-Laorden, B. L. & Jimenez-Cervantes, C. Melanocortin-1 receptor structure and functional regulation. *Pigment Cell Res.* **18**, 393-410 (2005).

28. Barsh, G. S. The genetics of pigmentation: from fancy genes to complex traits. *Trends Genet.* **12**, 299-305 (1996).

29. Harding, R. M., Healy, E., Ray, A. J., Ellis, N. S., Flanagan, N., Todd, C., et al. Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66**, 1351-1361 (2000).

30. Rana, B. K., Hewett-Emmett, D., Jin, L., Chang, B. H. J., Sambuughin, N., Lin, M., et al. High polymorphism at the human melanocortin 1 receptor locus. *Genetics.* **151**, 1547-1557 (1999).

31. Makova, K. & Norton, H. Worldwide polymorphism at the MC1R locus and normal pigmentation variation in humans. *Peptides.* **26**, 1901-1908 (2005).

32. Nakayama, K., Soemantri, A., Jin, F., Dashnyam, B., Ohtsuka, R., Duanchang, P., et al. Identification of novel functional variants of the melanocortin 1 receptor gene originated from Asians. *Hum. Genet.* **119**, 322-330 (2006).

33. Peng, S., Lu, X. M., Luo, H. R., Xiang-Yu, J. G. & Zhang, Y. P. Melanocortin-1 receptor gene variants in four Chinese ethnic populations. *Cell Res.* **11**, 81-84 (2001).

34. Yao, Y. G., Lu, X. M., Luo, H. R., Li, W. H. & Zhang, Y. P. Gene admixture in the silk road region of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes Genet. Syst.* **75**, 173-178 (2000).

35. Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., et al. Positive natural selection in the human lineage. *Science*. **312**, 1614-1620 (2006).

36. Voight, B. F., Kudaravalli, S., Wen, X. Q. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, 446-458 (2006).

37. Kimura, R., Fujimoto, A., Tokunaga, K. & Ohashi, J. A practical genome scan for population-specific strong selective sweeps that have reached fixation. *PLoS One*. **2**, e286 (2007).

38. Coop, G., Pickrell, J. K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., et al. The role of geography in human adaptation. *PLoS Genet.* advance online publication, doi:10.1371/journal.pgen.1000500

39. Motokawa, T., Kato, T., Hashimoto, Y. & Katagiri, T. Effect of Val92Met and Arg163Gln MC1R gene on freckles and solar variants of the lentigines in Japanese. *Pigment Cell Res.* **20**, 140-143 (2007).

40. Sturm, R. A., Duffy, D. L., Box, N. F., Chen, W., Smit, D. J., Brown, D. L., et al. The role of melanocortin-1 receptor polymorphism in skin cancer risk phenotypes. *Pigment Cell Res.* **16**, 266-272 (2003).

41. Nan, H., Kraft, P., Qureshi, A. A., Guo, Q., Chen, C., Hankinson, S. E., et al. Genome-Wide Association Study of Tanning Phenotype in a Population of European Ancestry. *J. Invest.*

Dermatol. **129**, 2250-2257 (2009).

42. Duffy, D. L., Box, N. F., Chen, W., Palmer, J. S., Montgomery, G. W., James, M. R., et al. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum. Mol. Genet.* **13**, 447-461 (2004).

43. Akey, J. M., Wang, H., Xiong, M., Wu, H., Liu, W., Shriver, M. D., et al. Interaction between the melanocortin-1 receptor and P genes contributes to inter-individual variation in skin pigmentation phenotypes in a Tibetan population. *Hum. Genet.* **108**, 516-520 (2001).

44. Buckley, W. R. & Grum, F. Reflection spectrophotometry—use in evaluation of skin pigmentary disturbances. *Arch. Dermatol.* **83**, 249-261 (1961).

45. Bastiaens, M., ter Huurne, J., Gruis, N., Bergman, W., Westendorp, R., Vermeer, B. J., et al. The melanocortin-1-receptor gene is the major freckle gene. *Hum. Mol. Genet.* **10**, 1701-1708 (2001).

46. Habif, T. P. *Clinical dermatology : a color guide to diagnosis and therapy* 4th edn. (Mosby: Philadelphia, 2004).

47. Sato, K. in *Color atlas of dermatology* Vol. 3 (eds. Suzuki, H.&Kanzaki, T.), 58, (Kodansha, Tokyo 2009).

48. Stephens, M. & Donnelly, P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **73**, 1162-1169 (2003).

49. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978-989 (2001).

50. Bandelt, H. J., Forster, P. & Rohl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37-48 (1999).

51. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564-567 (2010).

52. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* **123**, 585-595 (1989).

53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery eate - a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.).* **57**, 289-300 (1995).

54. Benjamini, Y. & Yekutieli, D. Quantitative trait loci analysis using the false discovery rate. *Genetics.* **171**, 783-790 (2005).

55. Tango, T. *Introduction to meta-analysis*. (Asakura shoten: Tokyo, 2002).

56. Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443-1452 (2007).

57. Eriksson, N., Macpherson, J. M., Tung, J. Y., Hon, L. S., Naughton, B., Saxonov, S., et al. Web-Based, Participant-Driven Studies Yield Novel Genetic Associations for Common Traits.

PLoS Genet. **6**, (2010).

58. Suda, A. Anthropometry of the Ryukyu Islanders. *J. Anthrop. Soc. Nippon.* **55**, 39-63 (1940).

59. Torii, R. Skin color of Okinawan people. *J. Anthrop. Soc. Nippon.* **20**, 44-56 (1904).

60. Frost, P. Human skin color: a possible relationship between its sexual dimorphism and its social perception. *Perspect. Biol. Med.* **32**, 38-58 (1988).

61. Li, E. P. H., Min, H. J., Belk, R. W., Kimura, J. & Bahl, S. Skin lightening and beauty in four asian cultures. *Adv. Consum. Res.* **35**, 444-449 (2008).

62. Forster, P., Torroni, A., Renfrew, C. & Rohl, A. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol. Biol. Evol.* **18**, 1864-1881 (2001).

63. Voight, B. F., Adams, A. M., Frisse, L. A., Qian, Y., Hudson, R. R. & Di Rienzo, A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 18508-18513 (2005).

64. Luo, M. R., Cui, G. & Rigg, B. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Res. Appl.* **26**, 340-350 (2001).

65. Yamaguchi-Kabata, Y., Tsunoda, T., Kumasaka, N., Takahashi, A., Hosono, N., Kubo, M., et al. Genetic differences in the two main groups of the Japanese population based on autosomal SNPs and haplotypes. *J. Hum. Genet.* **57**, 326-334 (2012).

66. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on

protein function using the SIFT algorithm. *Nat Protoc.* **4**, 1073-1081 (2009).

67. Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. A method and server for predicting damaging missense mutations. *Nat Methods.* **7**, 248-249 (2010).

68. Motokawa, T., Kato, T., Hashimoto, Y., Takimoto, H., Yamamoto, H. & Katagiri, T. Polymorphism patterns in the promoter region of the MC1R gene are associated with development of freckles and solar lentigines. *J. Invest. Dermatol.* **128**, 1588-1591 (2008).

69. Moro, O., Ideta, R. & Ifuku, O. Characterization of the promoter region of the human melanocortin-l receptor(MC1R) gene. *Biochem. Biophys. Res. Commun.* **262**, 452-460 (1999).

Table 1. Genotype and allele frequency of *MC1R* variants observed in this study.

Position	Amino Acid Position	rs #	Observed Genotype Frequencies			Call Sum	Derived Allele Frequencies (%)
			AA	AD	DD		
3G>A	M1M		643	1	0	644	0.08
189C>T	I63I		650	1	0	651	0.08
200G>A	R67Q	rs34090186	648	4	0	652	0.31
215C>T	P72L		651	1	0	652	0.08
249G>A	S83S		649	3	0	652	0.23
274G>A	V92M	rs2228479	581	68	2	651	5.53
330T>C	A110A		641	2	0	643	0.16
359T>C	I120T	rs33932559	619	24	0	643	1.87
366G>A	V122V		643	1	0	644	0.08
418G>A	V140M		643	1	0	644	0.08
488G>A	R163Q	rs885479	31	214	401	646	78.64
540C>T	I180I		643	2	0	645	0.16
660C>A	G220G		647	1	0	648	0.08
665C>A	A222D		647	1	0	648	0.08
682C>T	Q228STOP		645	1	0	646	0.08
687C>T	R229R		645	1	0	646	0.08
861C>G	I287M		645	1	0	646	0.08
923C>T	T308M		646	1	0	647	0.08
942A>G	T314T	rs2228478	531	98	3	632	8.23

AA= homozygote of ancestral allele, AD= heterozygote of ancestral and derived alleles, DD= homozygote of derived alelle.

Table 2. The coefficients of best models determined by stepwise multiple regression analyses.

Genotype Model	Independent Variables	Unstandardized Coefficients		Standardized Coefficients			Adjusted	ANOVA	
		B	SE	Beta	t	*P*	R^2	F	*P*
Additive Model	(Constant)	-0.42	0.12		-3.60	3.5E-04	0.072	17.03	1.2E-10
	JUL_JUL	0.57	0.09	0.25	6.40	3.1E-10			
	R163Q	0.16	0.07	0.10	2.40	1.6E-02			
	RARE	0.39	0.18	0.09	2.23	2.6E-02			
Recessive Model	(Constant)	-0.29	0.07		-4.23	2.6E-05	0.073	17.18	1.0E-10
	JUN_JUL	0.57	0.09	0.25	6.39	3.2E-10			
	R163Q	0.20	0.08	0.10	2.49	1.3E-02			
	RARE	0.39	0.18	0.09	2.25	2.5E-02			

Stepwise criteria was probability-of-F-to-enter<=0.050 and probability-of-F-to-remove>=0.100. Dependent variable is sex-adjusted Z-score of upper arm L* (ZL*). Independent variables are months at measurement (JUN_JUL), the number of 163Q allele (R163Q), the number of 92M allele (V92M), the number of 942>G allele (942A>G) and rare variants (RARE). V92M and 942A>G were excluded from the best models.

Table 3. Odds ratios for freckling of the minor alleles relative to the major alleles.

	Freckles				
	Absent	Present	*P* *	OR	(95%CI)
RARE					
0	580	20		1	
1	32	2	0.62	1.81	(0.41 - 8.09)
V92M					
92V/92V	564	17		1	
92V/92M, 92M/92M	64	6	**2.86E-02**	**3.11**	**(1.18 - 8.17)**
R163Q					
163Q/163Q	386	15		1	
163R/163Q	207	7	0.82	0.87	(0.35 - 2.17)
163R/163R	30	1	1.00	0.86	(0.11 - 6.72)
T314T(942A>G)					
942A/942A	514	17		1	
942A/942G, 942G/942G	96	5	0.56	1.57	(0.57 - 4.37)

* *P* -value of two-tailed fisher's exact test

Figure 1 Map of the islands of Japan divided in two groups, Ryukyu Islands and Main Islands of Japan.

In this study, Ryukyu Islands comprised Okinawa Islands (including Okinawa Island where the subjects resided), Sakishima Islands, and Amami Isalnds. Despite administration by Okinawa Prefecture, Daito Islands were excluded from Ryukyu Islands because of the genetic background of the residents who largely descended from relatively recent settlers from Hachijo Islands.

Figure 2 *MC1R* haplotype network based on the estimated haplotypes of 1306 chromosomes evaluated in this study.

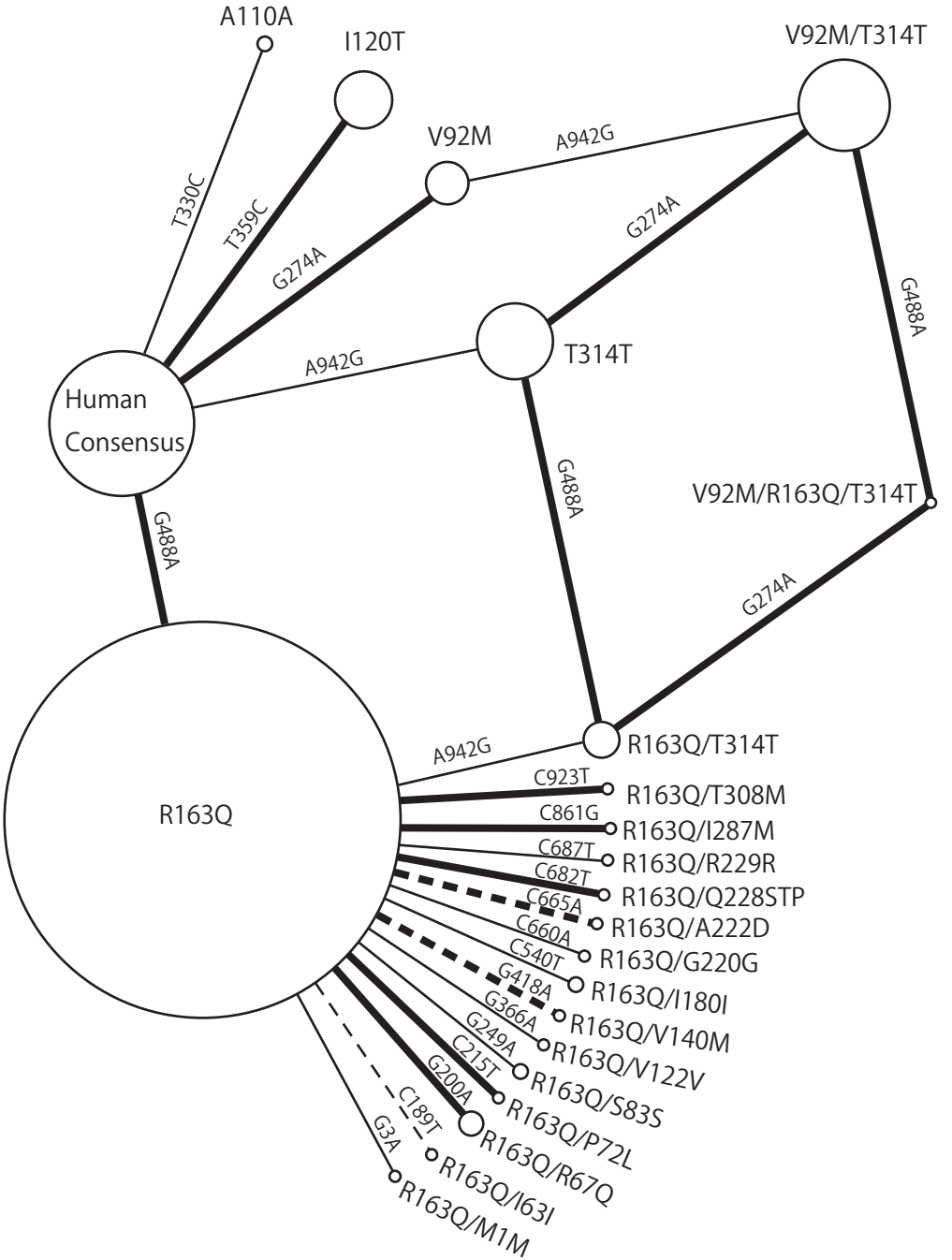
Each circle represents a haplotype with its amino acid change(s) annotated by or in the circle. The size of each circle reflects frequency of the haplotype. The smallest circle is equivalent to 1 out of 1306 chromosomes whereas the largest circle contains 931 chromosomes. Base changes are annotated along the lines connecting circles, which are broad for nonsynonymous. The three haplotypes estimated with the probability of 0.5 with R163Q are indicated with the dotted lines.

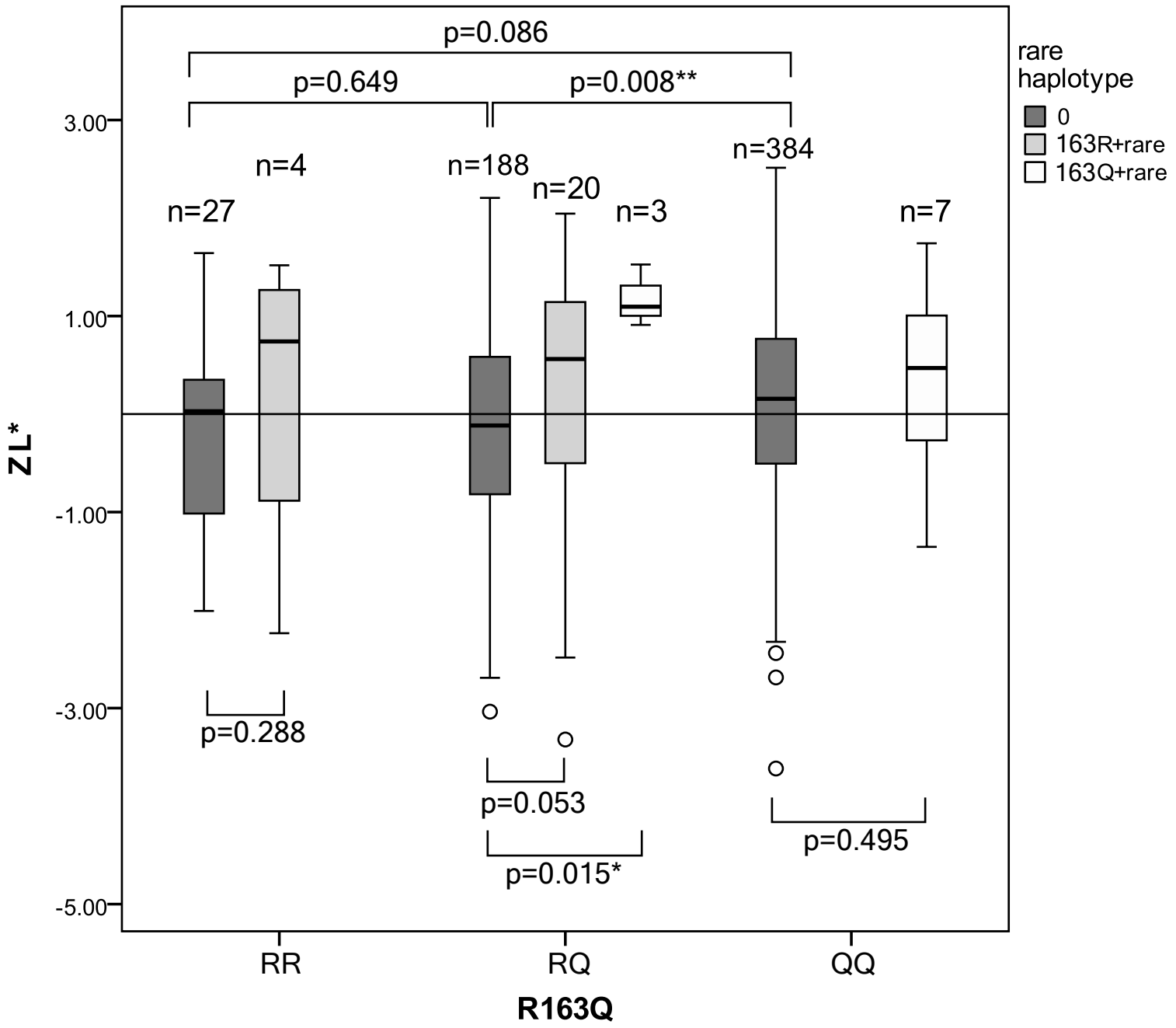
Figure 3 Boxplot of L* (lightness) categorized by R163Q genotype and clustered by rare haplotypes.

Z-score of L* at the inner upper arm was categorized by R163Q genotypes and clustered by the

haplotypes of rare variants at *MC1R.* The haplotype "163R+rare" means the rare variant is located on the same haplotype as the 163R allele, "163Q+rare" means that the haplotype contains the 163Q allele, and "0" means rare variant is absent. The haplotypes estimated with the probability of 0.5 were assumed to be "163Q+rare". The bottom of the box is the 25th percentile, the center line is the median, and the top of the box is the 75^{th} percentile. The bars from the box represents the range of data, except for ones with black dots where bars extend as far as 1.5 times longer than the box and data beyond the end of bar are outliers marked as open dots. The Kruskal-Wallis test rejected the null hypothesis of equal population median for R163Q genotypes when rare haplotype=0 (P=0.012), as well as for rare haplotypes when R163Q=RQ (P=0.010). Pair-wise tests for equal medians were performed with the Mann-Whitney U test and P-values are shown in the figure. .


RUSSIA
Hokkaido
CHINA
KOREA
JAPAN
Honshu
Shikoku
Kyushu
MAIN ISLANDS of JAPAN
Japanese Archipelago
Kagoshima
RYUKYU ISLANDS
Naha
Okinawa Island
Daito Islands


A110A
I120T
V92M/T314T
V92M
A942G
T330C
T359C
G274A
G274A
G488A
T314T
A942G
Human
Consensus
G488A
V92M/R163Q/T314T
G274A
G488A
A942G
R163Q/T314T
C923T
R163Q/T308M
R163Q
C861G
R163Q/I287M
C687T
R163Q/R229R
C682T
R163Q/Q228STP
C665A
R163Q/A222D
C660A
R163Q/G220G
C540T
R163Q/I180I
G418A
R163Q/V140M
G366A
R163Q/V122V
G249A
R163Q/S83S
C215T
R163Q/P72L
G200A
R163Q/R67Q
C189T
R163Q/I63I
G3A
R163Q/M1M


p=0.086
p=0.649
p=0.008**
rare haplotype
0
163R+rare
163Q+rare
3.00
1.00
-1.00
-3.00
-5.00
ZL*
n=27
n=4
n=188
n=20
n=3
n=384
n=7
p=0.288
p=0.053
p=0.015*
p=0.495
RR
RQ
QQ
R163Q

Supplementary Figures

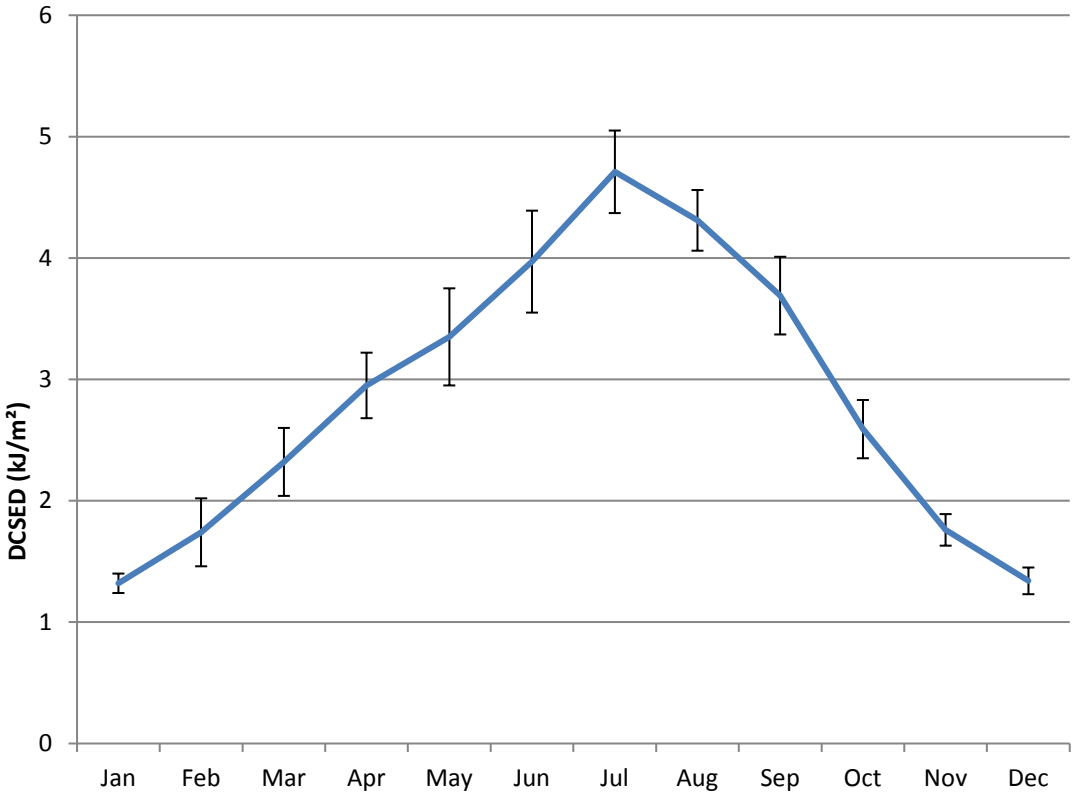


Figure S1. Monthly average of daily cumulative standard erythema dose (DCSED) in Naha. The graph was created based on the data of the Japan Meteorological Agency (http://www.data.kishou.go.jp/obs-env/uvhp/cie_monthave_nah.html#sansyo). An average value for each month is based on the data from 1994 to 2008.

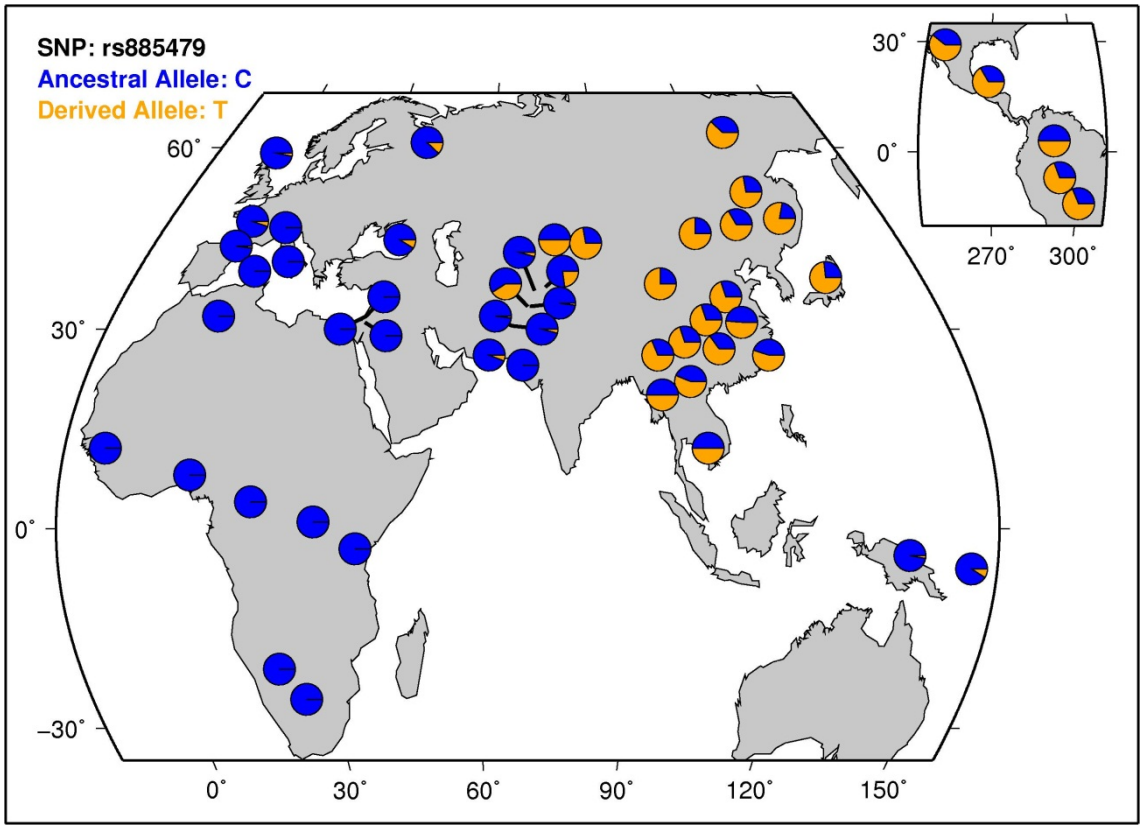


Figure S2. Geographic distribution of rs885479 across the Human Genome Diversity Panel CEPH (HGDP) populations. The map was created using HGDP selection browser (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/).

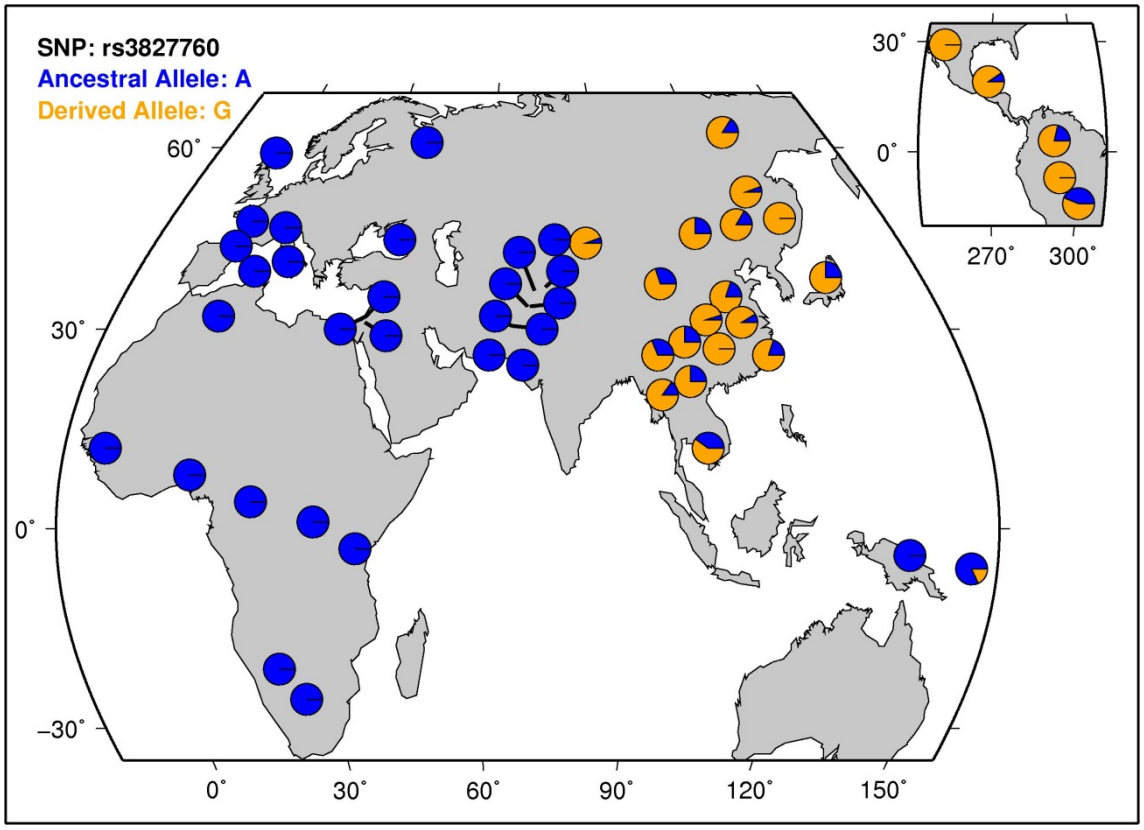


Figure S3. Geographic distribution of rs3827760 across the Human Genome Diversity Panel CEPH (HGDP) populations. This SNP is a nonsynonymous substitution in *EDAR* (V370A). The map was created using HGDP selection browser (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/) based on the allele frequencies imputed by fastPHASE.

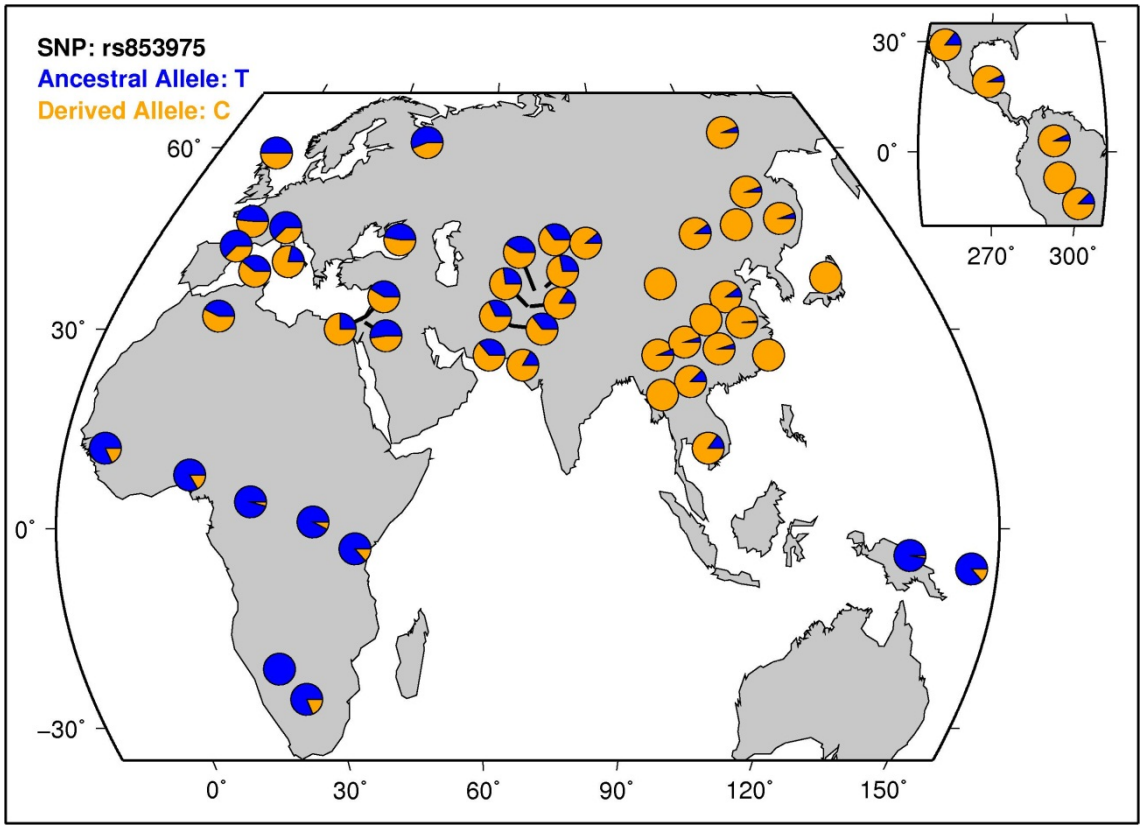


Figure S4. Geographic distribution of rs853975 across the Human Genome Diversity Panel CEPH (HGDP) populations. The map was created using HGDP selection browser (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/) based on the allele frequencies imputed by fastPHASE .

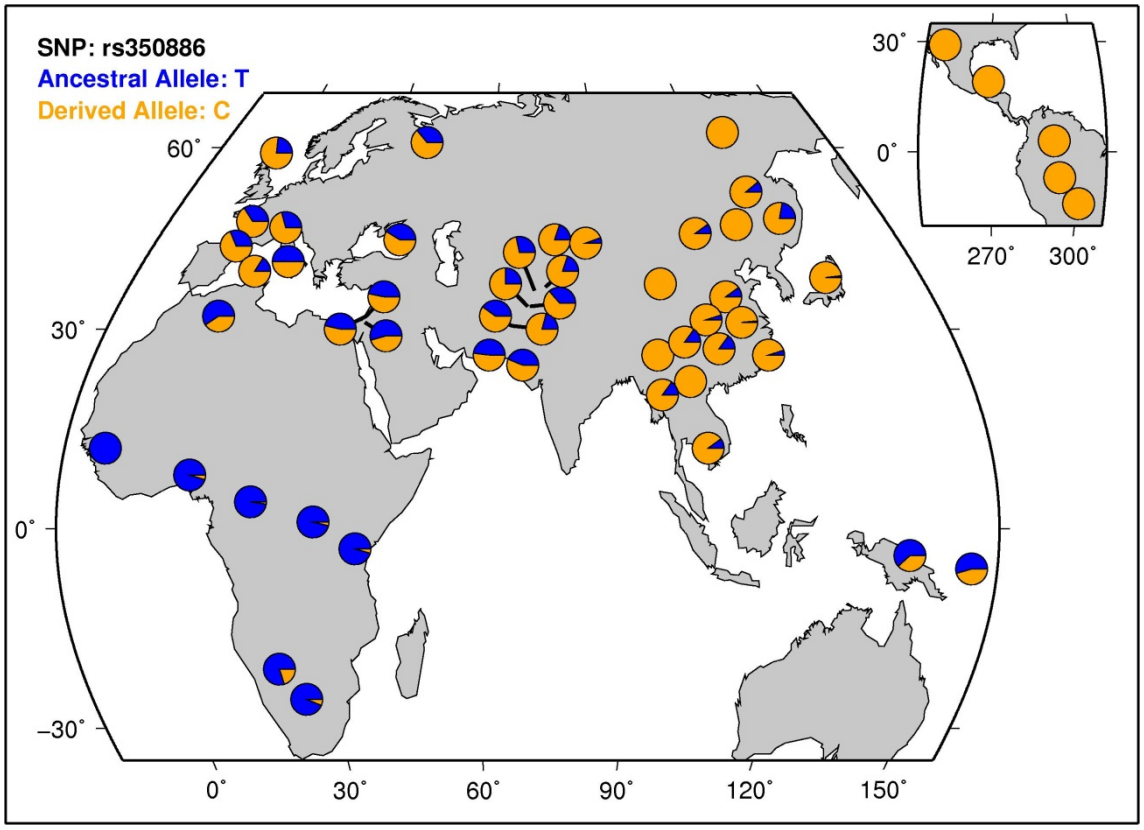


Figure S5. Geographic distribution of rs350886 across the Human Genome Diversity Panel CEPH (HGDP) populations. This SNP is located about 200bp downstream of the 3' end of *MAP2K2*. The map was created using HGDP selection browser (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/).

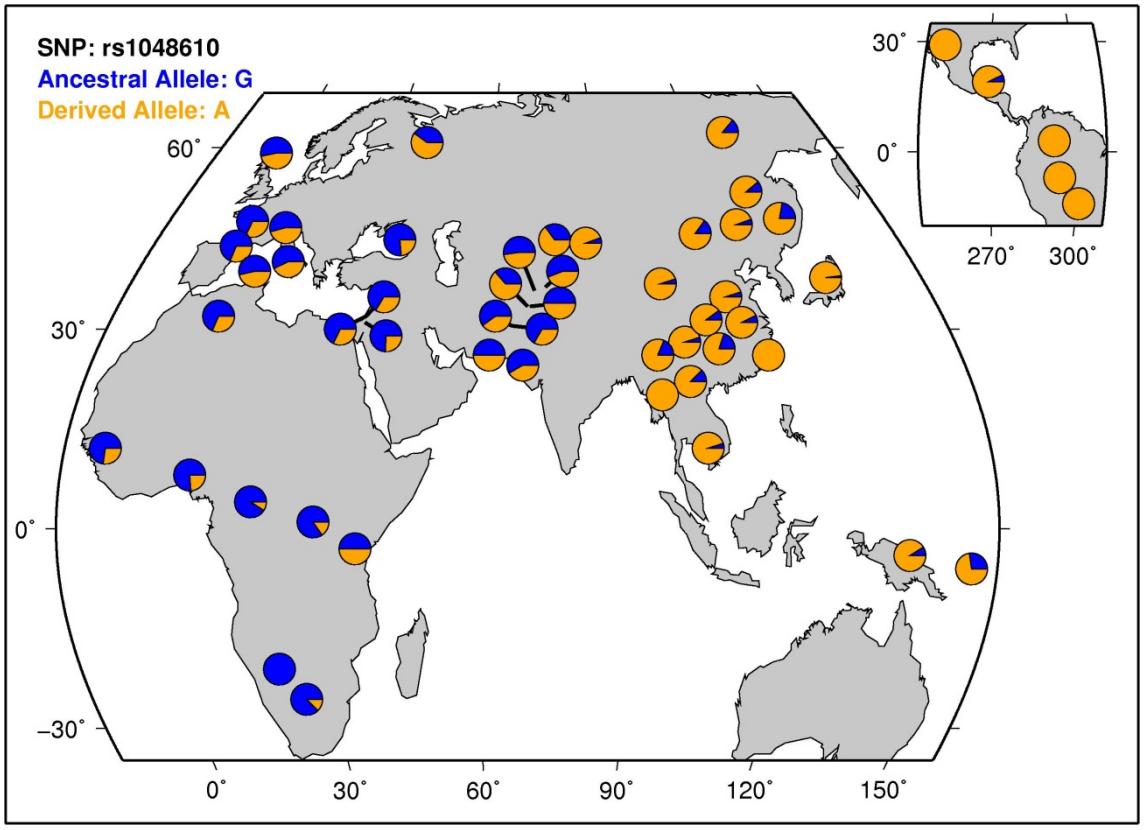


Figure S6. Geographic distribution of rs1048610 across the Human Genome Diversity Panel CEPH (HGDP) populations. This SNP is a synonymous substitution in *ADAM17*. The map was created using HGDP selection browser (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/) based on the allele frequencies imputed by fastPHASE.

Supplementary Tables

Table S1. Primers used in this study.

	Primers
For nested PCR	
1st PCR	1057_f (5’-CGAAATGTCCTGGGGACCTG-3’)
	2441_r (5’-AACGGGGACCAGGGAGGTAA-3’)
2nd PCR	1306_f (5’-CAGGACACCTGGAGGGGAAG-3’)
	2414_r (5’-CCCAGGGTCACACAGGAACC-3’)
For sequencing	
	1778_r (5’-CAGAGGCTGGACAGCATGGA-3’)
	2111_r (5’-GTGAGGGTGACAGCGCCTTT-3’)
	1890_f (5’-GGCCAGTGTCGTCTTCAGCA-3’)

Table S2. The results of sex-stratified multiple regression analyses.

Sex	Independent Variables	Unstandardized Coefficients		Standardized Coefficients			Adjusted	ANOVA	
		B	SE	Beta	t	*P*	R^2	F	*P*
Female	(Constant)	66.16	0.43		155.42	2.8E-275	0.019	2.72	6.7E-02
	R163Q	0.53	0.25	0.13	2.11	3.6E-02			
	RARE	1.01	0.71	0.09	1.42	1.6E-01			
Male	(Constant)	63.65	0.53		120.75	2.6E-284	0.016	2.74	6.6E-02
	R163Q	0.54	0.31	0.10	1.75	8.1E-02			
	RARE	1.40	0.75	0.10	1.86	6.4E-02			

Dependent variable: upper arm L*

Independent variables: the number of 163Q allele (R163Q), rare variants (RARE)

Additive models were assumed to explain genotype difference.

Table S3. The results of a meta-analysis combining the standardized coefficients and *P*-values from the sex-stratified multiple regression analysis.

Independent Variables	Pooled Standardized Coefficients※	95% Upper Confidence Interval	95% Lower Confidence Interval	*P*	Pooled *P* #
R163Q	0.109	0.032	0.186	5.9E-03	3.3E-03
RARE	0.094	0.016	0.171	1.8E-02	1.1E-02

Dependent variable: upper arm L*

Independent variables: the number of 163Q allele (R163Q) and rare variants (RARE)

※Fisher's r-to-z transformation method was used to pool the standardized regression coefficients of the separate models for females and males.

#Pooled *P*-values were calculated using the inverse normal method.

Table S4. The coefficients of linear regression models.

		Unstandardized Coefficients		Standardized Coefficients			Adjusted
SNP rs#	Gene	B	SE	Beta	t	*P*	R^2
rs3827760	EDAR	-.074	.058	-.050	-1.286	.199	.001
rs853975	-	-.054	.711	-.003	-.077	.939	-.002
rs350886	MAP2K2	.039	.123	.013	.322	.748	-.001
rs1048610	ADAM17	.104	.143	.029	.728	.467	-.001

Dependent variable: ZL*; n=644

Each SNP was entered as an independent variable in a single linear regression model to examine an association of the derived allele with skin reflectance. The derived alleles of these SNPs can be observed at high frequencies in East Asia as shown in Figures S3-S6.

Table S5. Prediciton of functional effects of *MC1R* variants observed in this study.

Position	Amino Acid substitution	rs #	PolyPhen score	Prediction※
200G>A	R67Q	rs34090186	0.744	Possibly Damaging
215C>T	P72L		0.997	Probably Damaging
274G>A	V92M	rs2228479	0.015	Benign
359T>C	I120T	rs33932559	0.656	Possibly Damaging
418G>A	V140M		0.390	Benign
488G>A	R163Q	rs885479	0.004	Benign
665C>A	A222D		0.870	Possibly Damaging
682C>T	Q228STOP			
861C>G	I287M		0.999	Probably Damaging
923C>T	T308M		0.979	Probably Damaging

※Functional effects were predicted using HumDiv model of PolyPhen-2.