Effects of Ancestry and Hypopigmenting Alleles on Skin Color in a Caribbean Native American Population

[1,2,*]Khai C Ang, [1,2]Victor A Canfield, [1,2]Tiffany C Foster, [1,2]Katherine P Reid, [3]Shou L Leong, [4,5,6]Yuka I Kawasawa, [4,7]Dajiang Liu, [8]John Hawley, [1,2,4,5*]Keith C Cheng

[1]Department of Pathology, [2]Jake Gittlen Laboratories for Cancer Research, [3]Department of Family & Community Medicine, [4]Department of Biochemistry and Molecular Biology, [5]Department of Pharmacology, [6]Institute of Personalized Medicine, [7]Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA, USA; [8]Salybia Mission Project, Dominica.

*Corresponding authors

Abstract

Shared ancestry suggests the sharing of one or more hypopigmenting skin color alleles between Native Americans and East Asians. Quantitating the effect sizes of candidate alleles requires an admixed study population with African but minimal European admixture. We genotyped 458 individuals (~15%) from a Caribbean Native American population, the Kalinago from the Commonwealth of Dominica, using Illumina Infinium Omni2.5-8 BeadChip. Admixture-based ancestry was confirmed to be shared with East Asians at K=3. Ancestry at K=6 was 55% Native American 32% African, and 11% European. Skin pigmentation ranged from 20 to 80 melanin units, averaging 46. Three albino individuals were homozygous for a multi-nucleotide polymorphism *OCA2*^{NW273KV} of African origin whose single allele effect size was -8 melanin units. European hypopigmenting allele frequencies for *SLC24A5*^{A111T} and *SLC45A2*^{L374F} were 0.14 and 0.05, with effect sizes per allele of -6 and -3, respectively. Native American ancestry contributed an estimated 21.8 to 28.5 melanin units.

Introduction

Skin pigmentation is a highly heritable polygenic trait influenced by health and environment[1]. European variants that lighten pigmentation include non-ancestral coding polymorphisms in *SLC24A5*[2–4] and *SLC45A2*[2,5,6] that are nearly fixed. However, the genetic basis for lighter skin pigmentation in Native American and East Asian populations, who carry the same ancestral alleles as Africans at these loci, has yet to be established. Dark skin was the ancestral phenotype for anatomically modern humans[3,7], whose evolution towards lighter skin at higher latitudes occurred independently in eastern and western Eurasia[4,8–10] potentially driven by a UV-dependent photoactivation step in the formation of vitamin D[11–14].

Our interest in the genetic basis of variation in skin pigmentation in Native American and East Asian populations led us to seek indigenous populations of the Western Hemisphere with sufficient African, but low enough European admixture to allow the mapping of Native American skin color alleles. Native Americans share common ancestry with East Asians[15,16], diverging before 15 kya[17–19], but the extent to which these populations share pigmentation variants remains to be determined. High European admixture is characteristic of most Native American

populations[17,18,20,21], complicating the characterization of pigmentation variants specific to Native Americans.

Prior to European contact, the Caribbean islands were inhabited by populations who migrated from the northern coast of South America[22–27]. During the Colonial period, large numbers of Africans were introduced into the Caribbean as slave labor[24,26]. As a consequence African and European admixture, the Native American ancestry now contributes only a minor portion of the ancestry of most Caribbean islanders[25,26,28,29]. The islands of Dominica and St. Vincent were the last colonized by Europeans, in the late 1700s[24,30,31]. In 1903, the British granted 15 km^2 (3,700 acres) on the eastern coast of Dominica as a reservation for the Kalinago, who were then called "Carib". When Dominica gained Independence in 1978, legal rights and a degree of protection from assimilation were gained by the inhabitants of the Carib Reserve (redesignated *Kalinago Territory* in 2015)[24]. The Kalinago, numbering about 3,000 living within the Territory[32], consider themselves to be of primarily Native American and African ancestry.

Early in our genetic, pedigree and phenotypic survey of the Kalinago, we noted one individual with albinism, and discovered that two others were known to reside in the Territory. Oculocutaneous albinism (OCA) is a recessive trait associated with visual system abnormalities and hypopigmentation of skin, hair, and eyes[33–37], that is caused by mutations in any of several autosomal pigmentation genes[37–48]. Here, we report on ancestry, distribution of measured skin color, identification of an albinism allele, and the hypopigmenting effects of this allele and the European *SLC24A5*^{A111T} and *SLC45A2*^{L374} alleles in a 15% sampling of Kalinago (Figure S1).

Results & Discussion

Our search for a population admixed for Native American/African ancestries led us to meetings in 2007 with faculty at the University of the West Indies in Kingston, Jamaica to discuss the history of indigenous populations in the Caribbean. There, we learned about the "Carib" population in the Commonwealth of Dominica. Observations from our initial trip to Dominica confirmed that the Kalinago skin color distribution and facial appearance were visually intermediate between those of East Asians and West Africans. Pursuit of the genetic studies described here required learning about oral and written histories, detailed discussion with community leadership, and IRB approval from Ross University (until Hurricane Maria in 2017, the largest medical school in Dominica), the Department of Health of the Commonwealth of Dominica, and relationship-building with three administrations the Kalinago Council over 15 years.

Population Sample

Our DNA and skin-color sampling program encompassed 458 individuals, representing 15% of the population of the territory and including the three known albino individuals. Ages ranged from 6 to 93 (Table S1 and Figure S2). We were able to obtain genealogical information for about half of the parents (243 mothers and 194 fathers); community-defined ancestry (described as 'black,' 'Kalinago,' or 'Mixed') for both parents was obtained for 426 individuals (92% of sample), including 221 from which DNA samples were obtained.

Kalinago Ancestry

The earliest western mention of the Kalinago (originally as "Caribs") was in Christopher Columbus's journal dated 26th November 1492[24]. Little is known about the detailed cultural and genetic similarities and differences between them and other pre-contact groups such as the Taino in the Caribbean. African admixture in the present Kalinago population derived from the African slave trade, but we were unable to identify written historical record that includes specific regional origin or well-defined contributions from other groups. The population's linguistics were uninformative, as they speak the same French-based Antillean Creole spoken on the neighboring islands of Guadeloupe and Martinique (in addition to English).

To study Kalinago population structure, we combined our Kalinago SNP genotype data with HGDP[49] data and analyzed using ADMIXTURE (Figures 1 and S3) as described in Methods. Shared ancestry of the Kalinago with East Asians at K=3 suggests the potential sharing of one or more lighter skin alleles between those populations. At K=4 and higher, a Native American component (that predominates in Kalinago) separates from the East Asian component. Consistent with prior work, an Oceanian component appears at K=5, and a Central & South Asian component appears at K=6[49]. These are minor components in our Kalinago sample (average <1%) (Table S3). On average, the Kalinago show 55% Native American, 32% African, and 11% European ancestry. The individual with least admixture has 94% Native American and 6% African. Principal component analysis (Figure S4) provides additional insight into the relationships between Kalinago and potential source populations.

Our analysis of Kalinago ancestry reflect considerably more Native American and less European ancestry than the Kalinago samples of Benn Torres et al[26] from outside of the Kalinago Territory (Figure 2A) and those admixed populations from the 1000 Genomes Project (1KGP)[29]. Some Western Hemisphere Native Americans reported in Reich et al[17] have varying proportions of European but very little African admixture (Figure 2B) . Overall, the Kalinago have more Native American and less European ancestry than any other Caribbean population.

More specifically, the 55% Native American ancestry observed for the Kalinago is far greater than the reported 13% in Puerto Rico[18], 10-15% for Tainos across the Caribbean[50], and 8% for Cubans[51]. Samples from Jamaica and the Lesser Antilles by Benn-Torres et al.[25,26] yielded an average of 7.7% Native American ancestry (range 5.6% to 16.2%), with the highest value from a population of Kalinago ancestry outside the reservation in Dominica. Relevant to the potential mapping of Native American lighter skin color alleles, the Kalinago has the lowest European ancestry compared to other reported Caribbean Native Americans in St. Kitts (8.2%), Barbados (11.5%) and Puerto Rico (71%)[26]. Potential reasons for the high percentage of Native American ancestry in the Kalinago likely include their segregation within the 3,700-acre Kalinago Territory in Dominica granted by the British in 1903, and the Kalinago tradition of women marring non-Kalinago being required to leave the Territory; non-Kalinago spouses of Kalinago men are allowed to move to the Territory (KCC, KCA, Personal Communication with Kalinago Council, 2014). These factors help to explain why the Benn-Torres[26] Kalinago samples, collected outside the Kalinago territory, show lower fractional Native American ancestry.

During our fieldwork, members of the Kalinago community were noted to characterize themselves and others in terms of perceived ancestry as “black,” “Kalinago,” or “mixed,” based primarily on phenotype. These folk categorizations were broadly supported by differences in admixture (Figures S5, S6).

Kalinago Skin Color Variation

Melanin index unit (MI) calculated from skin reflectance (see Methods) was used as a quantitative measure of melanin pigmentation[52,53]. The MI in the Kalinago ranged from 20.7 to 79.7 (Figure 3), averaging 45.7. The three Kalinago albino individuals sampled had the lowest values (20.7, 22.4 and 23.8). Excluding these, the MI ranged between 28.7 to 79.7 and averaged 45.9. For comparison, the MI averaged 25 and 21 for people of East Asian and European ancestry, respectively, as measured with the same equipment in our laboratory[52,54]. This range is similar to that of another indigenous population related to East Asians, the Senoi of Peninsular Malaysia (MI 24 to 78; mean = 45.7)[52]. The Senoi are believed include admixture from the darker-skinned Malaysian Negritos whose pigmentation is darker (mean = 55) than that of the average Kalinago. In comparison, the average MI was 53.4[55] for Africans in Cape Verde and 59 for African-Americans[56].

An OCA2 albinism allele in the Kalinago

Oculocutaneous albinism (OCA) is a genetically determined hypopigmentation of the skin, hair and eye associated with nystagmus, reduced visual acuity, foveal hypoplasia and strabismus[57,58]. The three sampled albino individuals had pale skin (MI 20.7, 22.4 and 23.8 vs. 29-80 for non-albinos), showed nystagmus, and reported photophobia and high susceptibility to sunburn. In contrast to the brown irides and black hair of most Kalinago, including their parents, the albino individuals had golden blonde hair and grey irides with varying amounts of green and blue.

Whole exome sequencing of one albino individual and one parent (obligate carrier) identified 12 variant alleles in 7 oculocutaneous albinism genes (or regions) that were heterozygous in the parent and homozygous for a non-reference allele in the albino individual (summarized in Table S4A); none was a nonsense or splice site variant. Five of the twelve potential candidate mutations found by this approach were intronic, one was synonymous, one was located in the 5’UTR, and three were in 3’UTR (Table S4B). Two missense variants were found in *OCA2*: SNP rs1800401 (c.913C>T or p.Arg305Trp in exon 9), *R305W*, and multi-nucleotide polymorphism rs797044784 in exon 8 (c.819_822delCTGGinsGGTC; p.Asn273_Trp274delinsLysVal), *NW273KV*.

Of 458 Kalinago *OCA2* genotypes, 26 carried *NW273KV* and 60 carried *R305W* (Table 1). Only *NW273KV* homozygotes were albino, and neither of the two *R305W* homozygotes (who were either heterozygous or homozygous ancestral for *NW273KV*) were albino. Notably, one Kalinago individual who is homozygous derived for *R305W* mutation but homozygous ancestral for *NW273* has an MI of 72, among the darkest in the entire population. Therefore, *R305W* is not an albinism allele. The frequency of *R305W* is > 0.10 in some African, South Asian, and European populations[29]. Notably, the black hair and dark eyes of Kalinago *R305W* homozygotes without the *NW273KV* suggest the possibility that the *in silico* predictions by SIFT, Polyphen 2.0 and PANTHER that *R305W* is a likely pathogenic variant[59] are incorrect . The observed patterns of

zygosity suggest that *NW273KV* arose on the background of a haplotype carrying the widespread *R305W* variant.

To address the population origin of the *NW273KV* mutation in *OCA2*, chromosome 15 genotypes of three albino individuals and seven carriers were queried for regions exhibiting homozygosity and identity-by-descent; in parallel, local genomic ancestry was estimated. Overlapping regions of homozygosity in albino individuals 2 and 3 defined a shared albino haplotype of about 11 Mb (Figure 4). Comparison to local genomic ancestry for the three albino individuals shows that this albino haplotype is of African origin. The minimal region defined by homozygosity in albino 1 is 1.7 Mb in length and includes seven genes in addition to *OCA2*.

The sole prior reported instances of albinism including the *NW273KV* variant involves a compound heterozygote of African-American descent[60,61]. The conservation of the NW sequence among vertebrates and its inclusion in a potential N-linked glycosylation site[62] that is eliminated by the mutation support the pathogenicity of this mutation. The *NW273KV* frequency in our sample (0.03) translates into a Hardy-Weinberg albinism frequency (p^2 = 0.0009) of ~1 per 1000, as observed (3 in a population of about 3000). Examination of publicly available data reveals three $OCA2^{NW273KV}$ heterozygotes in the 1000 Genome Project, a pair of siblings from Barbados (ACB) and one individual from Sierra Leone (MSL)[29]. These three individuals share a haplotype of 1.5 Mb but no more than 1.04 Mb are shared with the Kalinago albinos.

One of the three albinos was also heterozygous for $SLC24A5^{A111T}$, but his skin and hair color was not lighter than that of the other two albinos, who were homozygous for the ancestral allele at $SLC24A5^{A111}$; this observation is consistent with epistasis of *OCA2* albinism over $SLC24A5^{A111T}$. Eight sampled non-albino individuals had combinations of lighter hair, fair skin, and lighter irides, among these, seven were heterozygous for $SLC24A5^{A111T}$, and four were heterozygous for the albino variant. A precise understanding of the phenotypic effects of the combinations of these hypopigmenting alleles will require further study.

Contribution of Native American Ancestry to Kalinago Skin Color Variation

The primary target of this work is quantification of the contribution of Native American ancestry to skin pigmentation. Answering this question required identifying the Kalinago as a Native American population admixed for African Ancestry. To minimize interference of pigmentation lightening alleles specific to Europeans, we confirmed the relatively small 15% European ancestry of this population, and now need to identify individuals lacking either of the two known hypopigmenting variants fixed in Europeans, $SLC24A5^{A111T}$ and $SLC45A2^{L374F}$. In addition, we must identify individuals that are also ancestral for the $OCA2^{NW273KV}$ albinism allele in this population.

To control for the effects of the known European pigmentation loci, all Kalinago samples were genotyped for two known pigmentation polymorphisms of European origin, $SLC24A5^{A111T}$ and $SLC45A2^{L374F}$. The phenotypic effects of these variants and $OCA2^{NW273KV}$ are shown in the histograms of Figure 5. Each variant decreases melanin pigmentation, and homozygotes are lighter than heterozygotes. The greatest effect is seen in the $OCA2^{NW273KV}$ homozygotes (the

albino individuals), as previously noted. The frequencies of the derived alleles of *SLC24A5*^{A111T} and *SLC45A2*^{L374F} in the Kalinago sample are 0.14 and 0.05, respectively.

The higher frequency of *SLC24A5*^{A111T} compared to *SLC45A2*^{L374F} frequencies is not explained solely by European admixture, given that most Europeans are fixed for both alleles. This deviation can, however, be accounted for by the involvement of source populations that have a lower frequency of *SLC45A2*^{L374F} than *SLC24A5*^{A111T}. The 0.03 excess of *SLC24A5*^{A111T} over *SLC45A2*^{L374F} frequency indicates a non-negligible frequency of *SLC24A5*^{A111T} in one or more African source population. At 32% AFR ancestry, this corresponds to an average *SLC24A5*^{A111T} frequency of about 0.09 for the AFR source populations. Although many sub-Saharan West African populations (the likeliest source of AFR ancestry in the Kalinago)[63] have lower *SLC24A5*^{A111T} frequencies, a similar frequency is observed in the Mende of Sierra Leone (MSL)[29,], while some West African populations such as Hausa and Mandinka have frequencies exceeding 0.10[64,65]. The 0.06 deficit of *SLC45A2*^{L374F} frequency compared to EUR ancestry (11%) corresponds to an average *SLC45A2*^{L374F} frequency in the European source population close to 0.5. This is far below the frequency of 0.82 observed the 2015 Genomes Project Spanish population sample (IBS)[29]. It should be noted that the major component in North African and Middle Eastern populations is not distinguished from Europeans in our analysis; these populations (and also inhabitants of Andalusia in Spain) have a wide range of *SLC45A2*^{L374F} frequencies dropping considerably below that of IBS[64,65]. We are unable with existing information to definitively account for the higher frequency of *SLC24A5*^{A111T} over that of *SLC45A2*^{L374F}.

To investigate the relative contributions of genetic variation to skin color, we performed genome wide association analyses using an additive model for Melanin Unit, conditioning on sex, ancestry, and genotypes of *SLC24A5*^{A111T}, *SLC45A2*^{L374F} and *OCA2*^{NW273KV}. These analysis omitted the three albino individuals. We found that sex, all three genotyped polymorphisms, and the first principal component were statistically significant (effect sizes shown in Table 2). Effect sizes were about -6 units (per allele) for *SLC24A5*^{A111T}, -3 units for *SLC45A2*^{L374F} and -8 units for the first *OCA2*^{NW273KV} allele. When controlling for *OCA2*^{NW273KV} status, *OCA2*^{R305W} had no detectable effect on skin color (not shown).

The effect size for *SLC24A5*^{A111T} is consistent with previously reported results of -5 melanin units for African-Americans[3,8] and -5.5 for the admixed inhabitants of the Cape Verde islands[66]. Reported effect sizes for continental Africans are both higher and lower (-7.7 in [67] and -3.6 in [68]). The estimated effect size in the CANDELA study (GWAS of combined admixed populations from Mexico, Brazil, Columbia, Chile and Peru)[10], only about -2 melanin units, is an outlier. For *SLC45A2*^{L374F}, significance was found in Beleza et al. (2013)[69] and Adhikari et al. (2019)[10] and for African Americans but not the African Caribbean subsample in Horton *et al.* (2007)[8].

Our estimate that a single *OCA2*^{NW273KV} allele causes about -8 melanin units of skin lightening is the first reported population-based effect size measurement for any albinism allele. To study the effect of homozygosity, we applied the estimated parameters to the three albinos, who were lighter by an average of 10 units than predicted by the additive model ($p < 0.0033$, 1-tailed t-test). An additive model for skin color in albinos is rejected, where the nature of the non-linearity or epistasis is unclear.

The strong dependence of pigmentation on ancestry for individuals lacking hypo-pigmenting alleles *SLC24A5*^{A111T}, *SLC45A2*^{L374F} and *OCA2*^{NW273KV} is depicted in Figure 6. Positive deviations from the best fit are apparent at both high and low NAM ancestry, but these do not change the conclusions that AFR ancestry contributes to darker skin compared to NAM ancestry, and more importantly, that there are skin-lightening variants of Native American origin.

To estimate the contribution of Native American ancestry to skin hypopigmenting alleles, we analyzed the 276 samples without the *SLC24A5*^{A111T}, *SLC45A2*^{L374F} and *OCA2*^{NW273KV} polymorphisms plotted in Figure 6. A conservative estimate is represented by the difference in average melanin unit between Kalinago with less than 20% Native American ancestry, and Kalinago of more than 80% Native American ancestry: 65.6 vs 43.8, or 21.8. An alternative estimate is 28.5, based on the difference in average melanin unit between Kalinago with >80% African ancestry (72.3) and Kalinago with >80% Native American ancestry (43.8). The shared ancestry between Native Americans and East Asians shown by admixture analysis at K=3 (Figure 1) suggests that one or more hypopigmenting alleles may be shared between these ancestries.

Methods

Ethics Statement

The study was reviewed and approved by the Kalinago council and institutional review boards of Penn State University (29269EP), Ross University, and the Dominica Ministry of Health (H125). Informed consent was obtained from each participant enrolled in the study, and in the case of minors, consent was also obtained from a parent or guardian.

Recruitment

Participants from among the Kalinago populations were recruited with the help of nurses from the Kalinago Territory in 2014. Recruitment took place throughout the territory's 8 hamlets. Place and date of birth, reported ancestry of parents and grandparents, number of siblings, and response to sun exposure (tanning ability, burning susceptibility) were obtained by interview. Hair color and texture and eye color (characterized as black, brown, gray, blue, green, hazel, no pigment) were noted visually but not measured quantitatively.

Skin Reflectometry

Skin reflectance was measured using a Datacolor CHECK^{PLUS} spectrophotometer and converted to melanin unit as we have previously described[52,53]. To minimize the confounding effects of sun exposure and body hair, skin color measurements were measured on each participant's inner arm, and the average of triplicate measurements was generated. Measurements at this location are generally used as an approximation for constitutive skin pigmentation[69,70]. Before skin color measurements were taken, alcohol wipes were used to minimize the effect of dirt and/or oil. In order to minimize blanching due to occlusion of blood from the region being measured, care was taken not to apply only sufficient pressure to the skin to prevent ambient light from entering the scanned area[71].

DNA Collection

Saliva samples were collected using the Oragene Saliva kit, and DNA was extracted using the prepIT.L2P kit, both from DNA Genotek (Ottawa, Canada). DNA integrity was checked by agarose gel electrophoresis and quantitated using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Further quantification was done using Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA) as needed, following manufacturer instructions.

Genotyping

Oculocutaneous albinism variants previously identified in African and Native Americans[38,39,45,46] were amplified by PCR in all albino individuals as well as control samples using published conditions. Selected alleles of *SLC24A5, SLC45A2*, and *OCA2* were amplified in all sampled individuals as described in Table S2. Amplicons generated by 30 cycles of PCR using an Eppendorf thermocycler were sequenced (GeneWiz, South Plainfield, NJ) and the chromatograms viewed using Geneious software.

Illumina SNP genotyping using the Infinium Omni2.5-8 BeadChip was performed for all the individuals sampled. This was performed in three cohorts, using slightly different versions of the array, and the results combined. Due to ascertainment differences between the cohorts, analysis is presented here only for the combined sample. After quality control to eliminate duplicates and monomorphic variants, and to remove variants and individuals with genotype failure rates > 0.05, 358 Kalinago individuals and 1 638 140 unique autosomal SNPs remained.

Whole exome sequencing of albino and obligate carriers

In order to identify the causative variant for albinism in the Kalinago, 2 samples (one albino and one parent) were selected for whole exome sequencing. Following shearing of input DNA (1 microgram) using a Covaris E220 Foucsed-ultrasonicator (Woburn, MA), exome enrichment and library preparation was done using the Agilent SureSelect V5+UTR kit (Santa Clara, CA). The samples were sequenced at 50x coverage using a HiSeq 2500 sequencer (Illumina, San Diego, CA).

The *fastq* files were aligned back to Human Reference Genome GRCh37 (HG19) using BWA[72] and bowtie[73]. Candidate SNP polymorphisms were identified using GATK's UnifiedGenotyper[74], while the IGV browser was used to examine the exons of interest for indels[75,76]. Variants with low sequence depth (< 10) in either sample were excluded from further consideration.

Computational analysis

Association analysis, basic statistics, and merges with other datasets were performed using plink 1.9 [77,78]. Phasing and analysis of regions of homozygosity by descent and identity by descent were performed with Beagle 4.1[79,80], using 1000 Genomes Project (1KGP) phased data[29] as reference.

The genotyped individuals were randomly partitioned into nine subsets of 50 or 51 individuals (n=50 subsets) in which no pair exhibited greater than second-order relationship (PI_HAT > 0.25 using --genome command in plink). Using the same criteria, a maximal subset of 184 individuals was also generated (n=184 subset).

Principle components analysis (PCA) was performed using the smartpca program (version 13050) in the eigensoft package[81]. For comparison to HGDP populations, Kalinago samples were projected onto principal components calculated for the HGDP samples alone. For use as covariates in association analyses, the n=184 subset was used to generate the PCA, and the remaining individuals were projected onto the same axes.

Admixture analysis was performed using the ADMIXTURE program[82,83]. Each of the nine n=50 subsets was merged with the N=940 subset of HGDP data[84,85] for analysis and the outputs combined.

For association analysis, we removed the three albino individuals from the analysis. In addition to the entire remaining sample, we also analyzed the n=184 subset and each n=50 subset; the latter results were combined using METAL[86]. P-values were adjusted for statistic inflation by genomic control (median statistic method).

Statistical analysis of pigmentary effect of albinism involved fitting parameters to an additive model for the sample containing carriers but lacking albinos, applying the same model to the albino individuals, and comparing residuals for the albinos and the other individuals.

Local ancestry analysis of the region containing the albinism allele was performed using the PopPhased version of rfmix (v1.5.4) with the default window size of 0.2 cM[87]. A subset of 1KGP data served as reference haplotypes for European and African populations, and the Native American ancestry segments of the admixed samples as determined by Martin et al.[88] were combined to generate synthetic Native American reference haplotypes.

Acknowledgments

We would like to thank the Kalinago Council, Dominica Ministry of Health, nurses at the Kalinago Territory, Salybia Mission Project and the Kalinago community for their assistance and participation in this study. We would also like to acknowledge faculty of Ross University, Portsmouth, Dominica (now Bridgetown, Barbados), especially Drs. Gerhard Meisenberg (retired) and Liris Benjamin of Ross University in helping us to obtain the necessary IRB approval for fieldwork. We would also like to acknowledge the Hershey Rotary Club, Jake Gittlen Laboratories for Cancer Research and Department of Pathology for funding portions of this project. We would also like to acknowledge members of the Cheng Lab for their constructive comments and input.

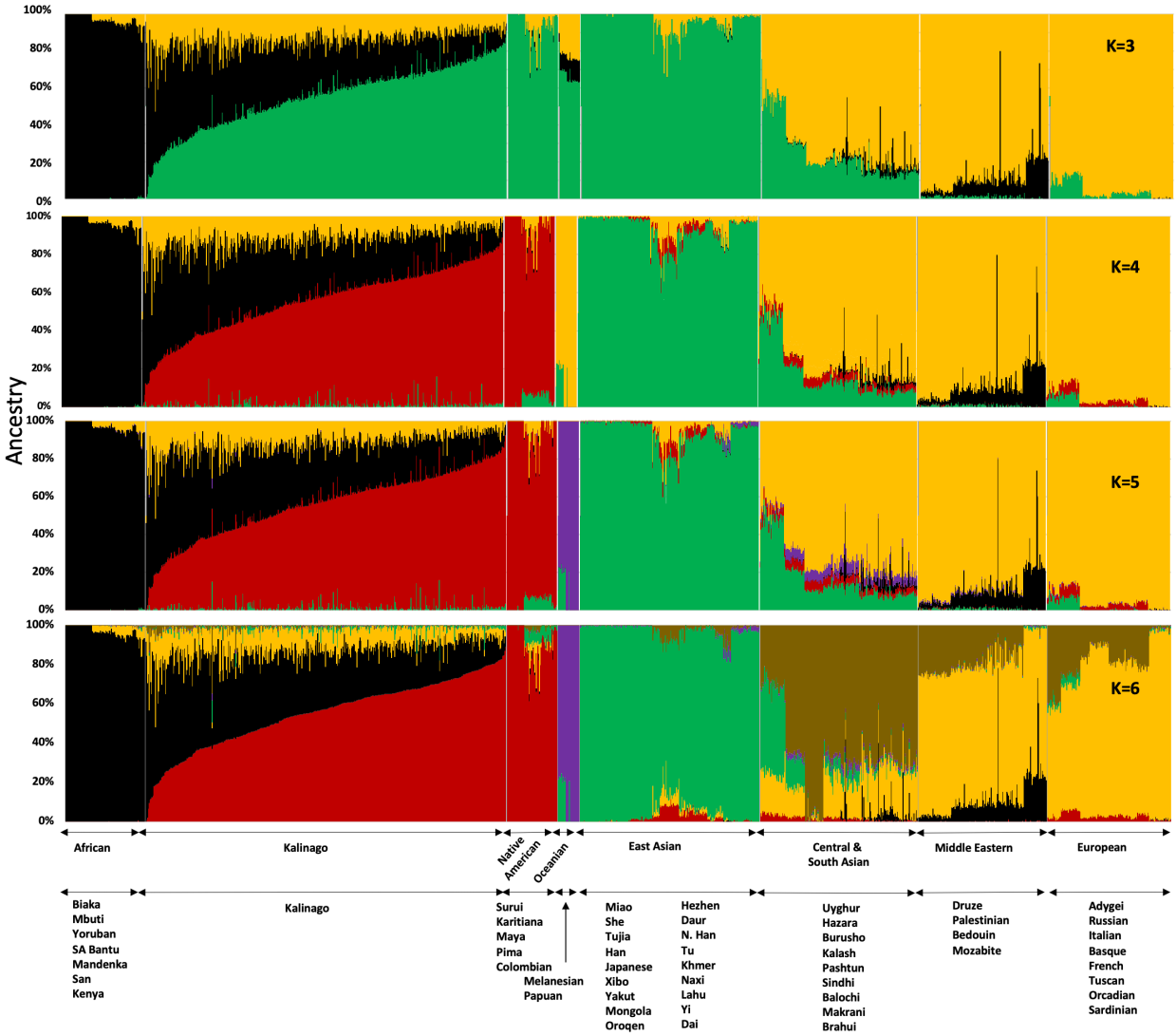

Figure 1: Admixture analysis of Kalinago compared with Human Genome Diversity Project populations. Results are depicted using stacked bar plots, with one column per individual. At K=3, the Kalinago, Native Americans, Oceanians, and East Asians fall into the same cluster. At K=4, the Kalinago and the Native Americans are separated from the East Asians.

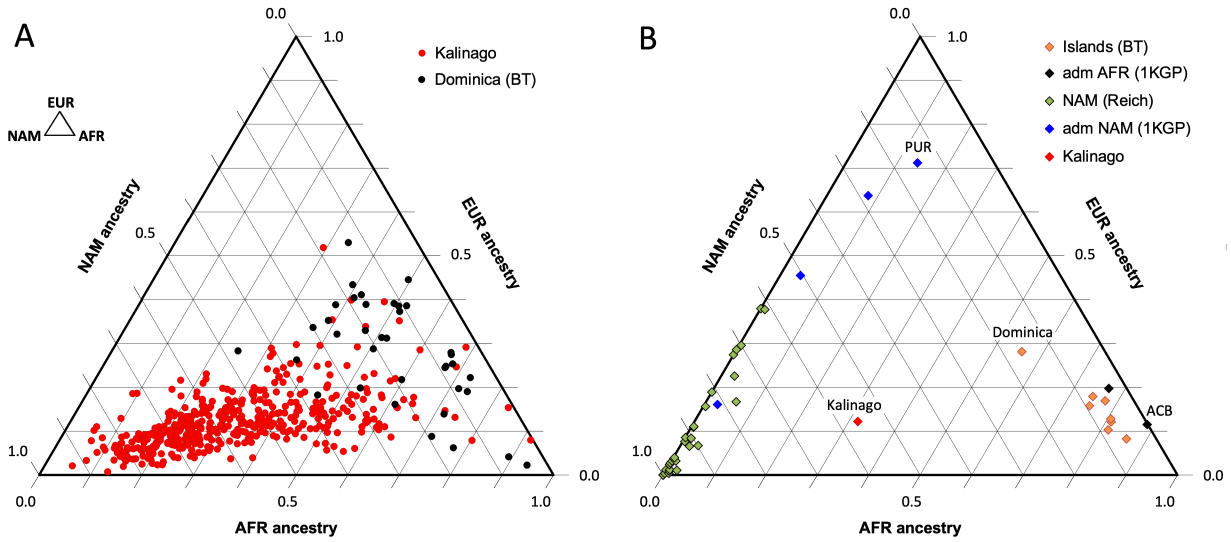

Figure 2. Comparison of Kalinago ancestry with that of other populations in the Western Hemisphere. Ternary plots show estimated proportions of African (AFR), European (EUR) and Native American (NAM) ancestry. **A,** Comparison of individuals (n=452, omitting 6 individuals with EAS > 0.1) genotyped in this study to individuals (n=38) from southern Dominica sampled by Benn Torres et al[26] **B,** Comparison of population averages. Kalinago, this study (n=458); Islands (BT) indicates Caribbean islanders reported in Benn Torres et al[26], with Dominica labeled; admixed (adm) AFR (1KGP) and admixed NAM (1KGP) represent admixed populations from[29], with Caribbean samples PUR (Puerto Rico) and ACB (Barbados) labeled; and AMR (Reich) indicates mainland Native American samples reported in Reich et al[17]. Inset shows ancestries at vertices.

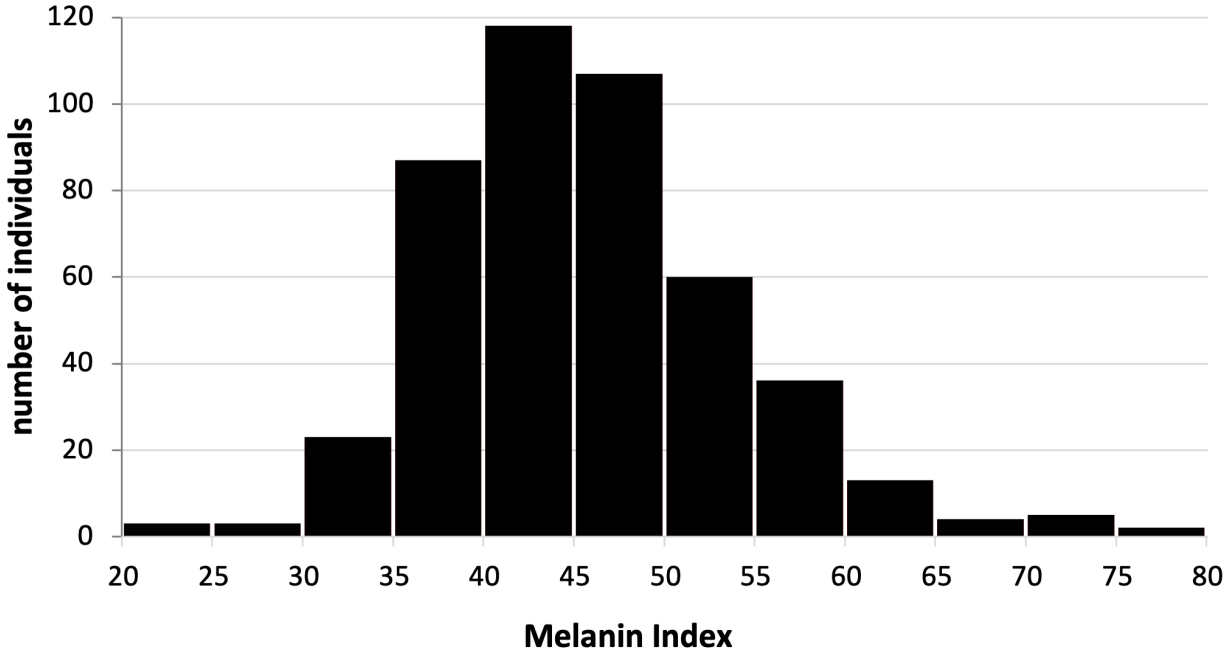

Figure 3: Skin color distribution of the Kalinago from Commonwealth of Dominica. We collected 462 Kalinago who live in the Kalinago Reservation. Each participant was asked a set of questions about their ancestry, gave their saliva sample, and have their skin color measured under their arm.

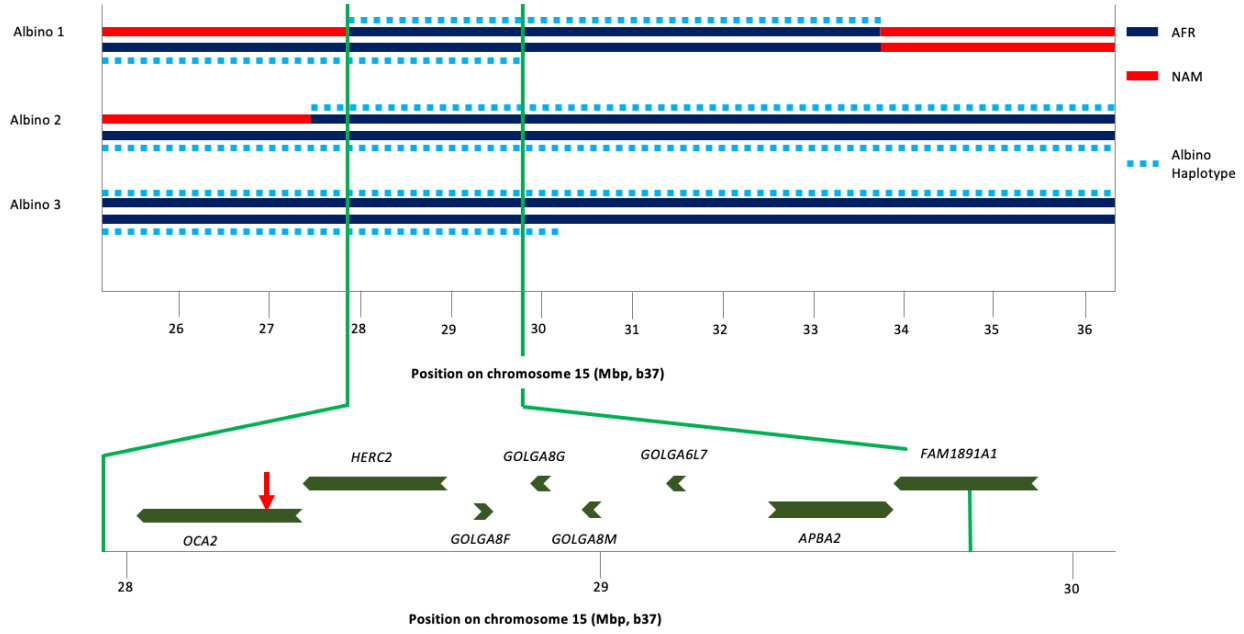

Figure 4. Haplotype analysis for three albino individuals. Inner two lines indicate NAM (red) or AFR (black) ancestry; no EUR ancestry was found in this genomic region. For this local ancestry analysis, the region shown here consisted of 110 non-overlapping segments with 7 to 346 SNPs each (mean 65). The deduced extent of shared albino haplotype (dotted light blue lines) is indicated on each chromosome. Minimum homozygous region (determined by albino individual 1) shared by all three albino individuals is shown at expanded scale below. Genes in this region are labeled, and the position of the *NW273KV* polymorphism in *OCA2* is indicated by the red arrowhead.

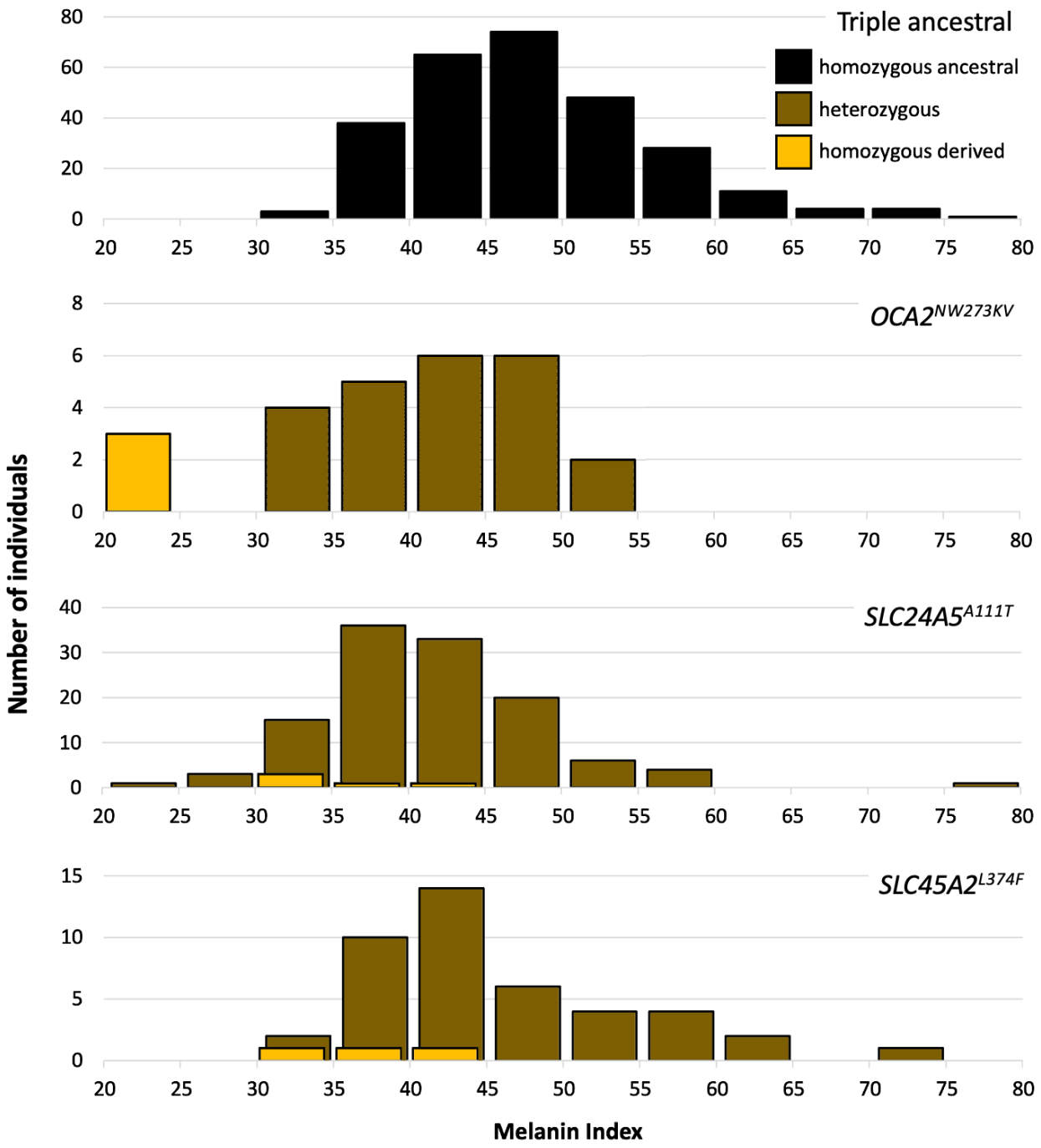

Figure 5. Skin color distribution of Kalinago samples according to genotype. Triple ancestral for three pigmentation alleles (*SLC24A5*^{111A}, *SLC45A2*^{374L} *and OCA2*^{273NW}) and derived (heterozygous or homozygous) for the indicated variant: *OCA2*^{NW273KV}; *SLC24A5*^{A111T}; and *SLC45A2*^{L374F}.

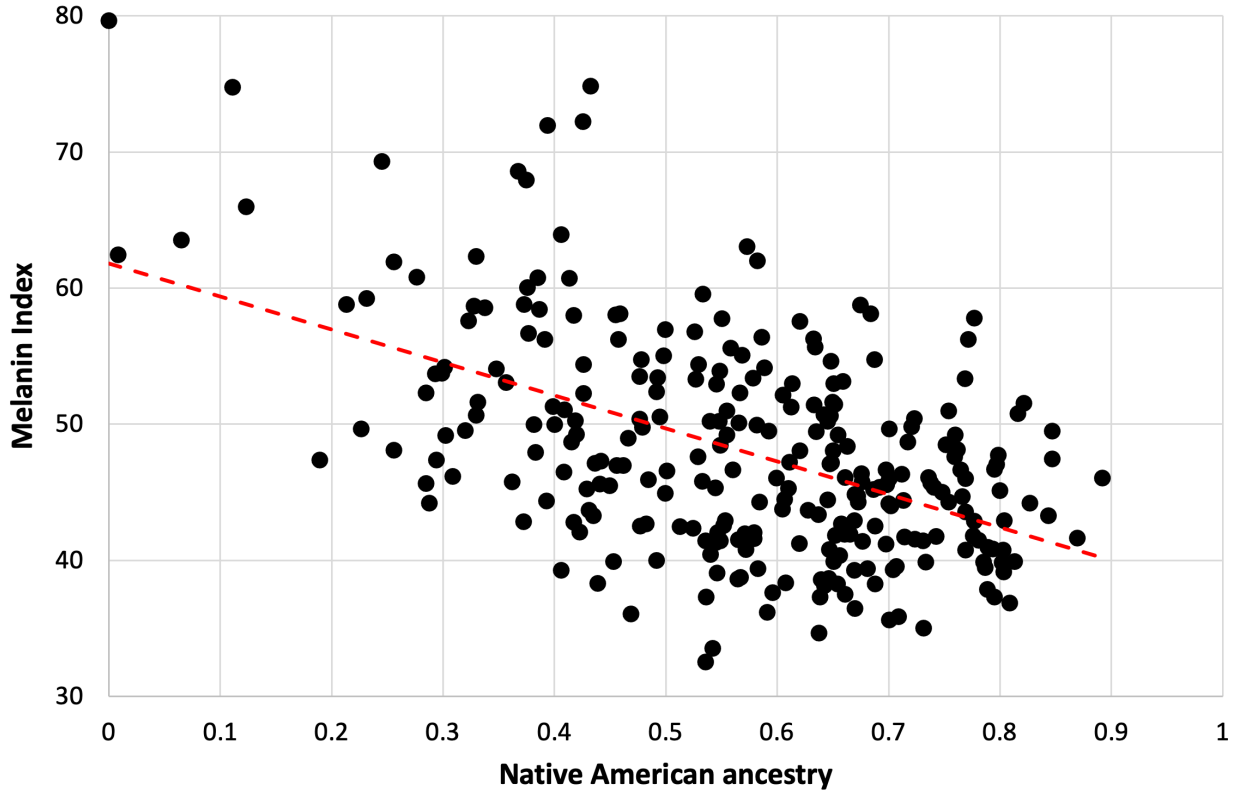

Figure 6. Dependence of Melanin Unit on ancestry for Kalinago. Only individuals who are ancestral for *SLC24A5*^{111A}, *SLC45A2*^{374L}, and *OCA2*^{273NW} alleles are shown (n=276). The dotted red line represents the best fit (linear regression).

Table 1. Albinism among *NW273KV* and *R305W* genotypes.

Allele/Genotype		***NW273KV* genotype**			
		Homozygous Ancestral[a]	Heterozygous	Homozygous Derived	Total
***R305W* genotype**	Homozygous Ancestral	398	0	0	398
	Heterozygous	33	22	0	55
	Homozygous Derived	1	1	3*	5
	Total	432	23	3*	458

[a] Ancestral=reference allele and derived=alternate allele for both variants.
* Albino phenotype.

Table 2. Effect sizes for covariates in full model with 10 Principal Components

covariate	Effect size[a]	Adjusted *p*-value
rs1426654 ($SLC24A5^{A111T}$)	-5.8	1.5E-12
rs16891982 ($SLC45A2^{L374F}$)	-2.8	0.015
albino allele ($OCA2^{NW273KV}$)	-7.8	2.5E-05
sex (male vs female)	2.4	0.0013

[a] per allele effect size, in melanin units, for *A111T* and *L374F*; effect of first allele for albino variant

References

1. Barsh, G. S. What controls variation in human skin color? *PLoS Biol* **1**, E27 (2003).
2. Soejima, M. & Koda, Y. Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *Int J Legal Med* **121**, 36–9 (2007).
3. Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–6 (2005).
4. Basu Mallick, C. *et al.* The Light Skin Allele of SLC24A5 in South Asians and Europeans Shares Identity by Descent. *PLoS Genet* **9**, e1003912 (2013).
5. Lucotte, G. & Yuasa, I. Near Fixation of 374l Allele Frequencies of the Skin Pigmentation Gene SLC45A2 in Africa. *Biochem Genet* **51**, 655–665 (2013).
6. Yuasa, I. *et al.* Distribution of the F374 allele of the SLC45A2 (MATP) gene and founder-haplotype analysis. *Ann Hum Genet* **70**, 802–11 (2006).
7. Jablonski, N. G. & Chaplin, G. The evolution of human skin coloration. *J Hum Evol* **39**, 57–106 (2000).
8. Norton, H. L. *et al.* Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol Biol Evol* **24**, 710–22 (2007).
9. Yang, Z. *et al.* A genetic mechanism for convergent skin lightening during recent human evolution. *Mol Biol Evol* msw003 (2016) doi:10.1093/molbev/msw003.
10. Adhikari, K. *et al.* A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nature Communications* **10**, 1–16 (2019).
11. Holick, M. F. The cutaneous photosynthesis of previtamin D3: a unique photoendocrine system. *J Invest Dermatol* **77**, 51–8 (1981).

12. Engelsen, O. The Relationship between Ultraviolet Radiation Exposure and Vitamin D Status. *Nutrients* **2**, 482–495 (2010).
13. Hanel, A. & Carlberg, C. Skin colour and vitamin D: An update. *Experimental Dermatology* **n/a**,.
14. Loomis, W. F. Skin-pigment regulation of vitamin-D biosynthesis in man. *Science* **157**, 501–506 (1967).
15. Derenko, M. *et al.* Origin and Post-Glacial Dispersal of Mitochondrial DNA Haplogroups C and D in Northern Asia. *PLoS One* **5**, (2010).
16. Tamm, E. *et al.* Beringian Standstill and Spread of Native American Founders. *PLoS ONE* **2**, (2007).
17. Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
18. Gravel, S. *et al.* Reconstructing Native American Migrations from Whole-Genome and Whole-Exome Data. *PLoS Genet* **9**, e1004023 (2013).
19. Moreno-Mayar, J. V. *et al.* Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* (2018) doi:10.1038/nature25173.
20. Klimentidis, Y. C., Miller, G. F. & Shriver, M. D. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am. J. Phys. Anthropol.* **138**, 375–383 (2009).
21. Brown, L. A. *et al.* Admixture Mapping Identifies an Amerindian Ancestry Locus Associated with Albuminuria in Hispanics in the United States. *J. Am. Soc. Nephrol.* **28**, 2211–2220 (2017).

22. Island Caribs. *Wikipedia, the free encyclopedia* (2016).

23. HARVEY, R. G. *et al.* FREQUENCY OF GENETIC TRAITS IN THE CARIBS OF DOMINICA. *Human Biology* **41**, 342–364 (1969).

24. Honychurch, L. *The Dominica Story: A History of the Island*. (Macmillan, 2012).

25. Torres, J. B. *et al.* Genetic Diversity in the Lesser Antilles and Its Implications for the Settlement of the Caribbean Basin. *PLOS ONE* **10**, e0139192 (2015).

26. Torres, J. B., Stone, A. C. & Kittles, R. An anthropological genetic perspective on Creolization in the Anglophone Caribbean. *Am. J. Phys. Anthropol.* **151**, 135–143 (2013).

27. Benn-Torres, J. *et al.* Admixture and population stratification in African Caribbean populations. *Ann. Hum. Genet.* **72**, 90–98 (2008).

28. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–73 (2010).

29. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

30. Honychurch, L. Review of The Lesser Antilles in the Age of European Expansion. *NWIG: New West Indian Guide / Nieuwe West-Indische Gids* **72**, 305–307 (1998).

31. A Brief History of the Caribbean: From the Arawak and Carib to the Present: Jan Rogozinski: 9780452281936: Amazon.com: Books. https://www.amazon.com/Brief-History-Caribbean-Arawak-Present/dp/0452281938.

32. Kalinago Territory. *Wikipedia* (2021).

33. Gargiulo, A. *et al.* Molecular and clinical characterization of albinism in a large cohort of Italian patients. *Invest Ophthalmol Vis Sci* **52**, 1281–9 (2011).

34. Grønskov, K., Brøndum-Nielsen, K., Lorenz, B. & Preising, M. N. Clinical utility gene card for: Oculocutaneous albinism. *European Journal of Human Genetics* **22**, (2014).

35. Gronskov, K., Ek, J. & Brondum-Nielsen, K. Oculocutaneous albinism. *Orphanet J Rare Dis* **2**, 43 (2007).

36. Hong, E. S., Zeeb, H. & Repacholi, M. H. Albinism in Africa as a public health issue. *BMC Public Health* **6**, 212 (2006).

37. Vogel, P. *et al.* Ocular albinism and hypopigmentation defects in Slc24a5-/- mice. *Vet Pathol* **45**, 264–79 (2008).

38. Stevens, G., Ramsay, M. & Jenkins, T. Oculocutaneous albinism (OCA2) in sub-Saharan Africa: distribution of the common 2.7-kb P gene deletion mutation. *Hum. Genet.* **99**, 523–527 (1997).

39. King, R. A. *et al.* Tyrosinase gene mutations in oculocutaneous albinism 1 (OCA1): definition of the phenotype. *Hum Genet* **113**, 502–513 (2003).

40. Spritz, R. A., Fukai, K., Holmes, S. A. & Luande, J. Frequent intragenic deletion of the P gene in Tanzanian patients with type II oculocutaneous albinism (OCA2). *Am J Hum Genet* **56**, 1320–1323 (1995).

41. Stevens, G., van Beukering, J., Jenkins, T. & Ramsay, M. An intragenic deletion of the P gene is the common mutation causing tyrosinase-positive oculocutaneous albinism in southern African Negroids. *Am J Hum Genet* **56**, 586–591 (1995).

42. Edwards, M. *et al.* Association of the OCA2 Polymorphism His615Arg with Melanin Content in East Asian Populations: Further Evidence of Convergent Evolution of Skin Pigmentation. *PLOS Genetics* **6**, 1–8 (2010).

43. Gao, J. *et al.* Retrospective analysis in oculocutaneous albinism patients for the 2.7 kb deletion in the OCA2 gene revealed a co-segregation of the controversial variant, p.R305W. *Cell Biosci* **7**, (2017).

44. Woolf, C. M. Albinism (OCA2) in Amerindians. *Am. J. Phys. Anthropol.* **Suppl 41**, 118–140 (2005).

45. Yi, Z. *et al.* A 122.5-kilobase deletion of the P gene underlies the high prevalence of oculocutaneous albinism type 2 in the Navajo population. *Am. J. Hum. Genet.* **72**, 62–72 (2003).

46. Carrasco, A., Forbes, E. M., Jeambrun, P. & Brilliant, M. H. A splice site mutation is the cause of the high prevalence of oculocutaneous albinism type 2 in the Kuna population. *Pigment Cell & Melanoma Research* **22**, 645–647 (2009).

47. Grønskov, K. *et al.* Mutations in C10orf11, a Melanocyte-Differentiation Gene, Cause Autosomal-Recessive Albinism. *The American Journal of Human Genetics* **92**, 415–421 (2013).

48. Kausar, T., Bhatti, M., Ali, M., Shaikh, R. & Ahmed, Z. OCA5, a novel locus for non-syndromic oculocutaneous albinism, maps to chromosome 4q24. *Clin Genet* **84**, 91–93 (2013).

49. Cavalli-Sforza, L. L. The Human Genome Diversity Project: past, present and future. *Nature Reviews Genetics* **6**, 333–340 (2005).

50. Schroeder, H. *et al.* Origins and genetic legacies of the Caribbean Taino. *PNAS* 201716839 (2018) doi:10.1073/pnas.1716839115.

51. Marcheco-Teruel, B. *et al.* Cuba: Exploring the History of Admixture and the Genetic Basis of Pigmentation Using Autosomal and Uniparental Markers. *PLOS Genetics* **10**, e1004488 (2014).

52. Ang, K. C. *et al.* Skin color variation in Orang Asli tribes of Peninsular Malaysia. *PLoS One* **7**, e42752 (2012).

53. Diffey, B. L., Oliver, R. J. & Farr, P. M. A portable instrument for quantifying erythema induced by ultraviolet radiation. *Br J Dermatol* **111**, 663–72 (1984).

54. Tsetskhladze, Z. R. *et al.* Functional assessment of human coding mutations affecting skin pigmentation using zebrafish. *PLoS One* **7**, e47398 (2012).

55. Beleza, S. *et al.* The Admixture Structure and Genetic Variation of the Archipelago of Cape Verde and Its Implications for Admixture Mapping Studies. *PLOS ONE* **7**, e51103 (2012).

56. Shriver, M. D. *et al.* Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* **112**, 387–99 (2003).

57. van Geel, N. *et al.* Hypomelanoses in Children. *J Cutan Aesthet Surg* **6**, 65–72 (2013).

58. Dessinioti, C., Stratigos, A. J., Rigopoulos, D. & Katsambas, A. D. A review of genetic disorders of hypopigmentation: lessons learned from the biology of melanocytes. *Experimental Dermatology* **18**, 741–749 (2009).

59. Kamaraj, B. & Purohit, R. Computational Screening of Disease-Associated Mutations in OCA2 Gene. *Cell Biochem Biophys* **68**, 97–109 (2013).

60. Garrison, N. A. *et al.* P gene mutations in patients with oculocutaneous albinism and findings suggestive of Hermansky-Pudlak syndrome. *J Med Genet* **41**, e86–e86 (2004).

61. Lee, S.-T. *et al.* Mutations of the P Gene in Oculocutaneous Albinism, Ocular Albinism, and Prader-Willi Syndrome Plus Albinism. *New England Journal of Medicine* **330**, 529–534 (1994).

62. Rinchik, E. M. *et al.* A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. *Nature* **361**, 72–76 (1993).

63. Micheletti, S. J. *et al.* Genetic Consequences of the Transatlantic Slave Trade in the Americas. *The American Journal of Human Genetics* **0**, (2020).

64. Cheung, K. H. *et al.* ALFRED: a Web-accessible allele frequency database. *Pac Symp Biocomput* 639–50 (2000).

65. Rajeevan, H., Soundararajan, U., Kidd, J. R., Pakstis, A. J. & Kidd, K. K. ALFRED: an allele frequency resource for research and teaching. *Nucleic Acids Research* **40**, D1010–D1015 (2012).

66. Beleza, S. *et al.* The Timing of Pigmentation Lightening in Europeans. *Mol Biol Evol* **30**, 24–35 (2013).

67. Crawford, N. G. *et al.* Loci associated with skin pigmentation identified in African populations. *Science* eaan8433 (2017) doi:10.1126/science.aan8433.

68. Martin, A. R. *et al.* An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell* **171**, 1340-1353.e14 (2017).

69. Park, J.-H. & Lee, M.-H. A Study of Skin Color by Melanin Index According to Site, Gestational Age, Birth Weight and Season of Birth in Korean Neonates. *J Korean Med Sci* **20**, 105–108 (2005).

70. Choe, Y. B., Jang, S. J., Jo, S. J., Ahn, K. J. & Youn, J. I. The difference between the constitutive and facultative skin color does not reflect skin phototype in Asian skin. *Skin Res Technol* **12**, 68–72 (2006).

71. Fullerton, A. *et al.* Guidelines for measurement of skin colour and erythema. A report from the Standardization Group of the European Society of Contact Dermatitis. *Contact Dermatitis* **35**, 1–10 (1996).

72. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

73. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).

74. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

75. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178–192 (2013).

76. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration | Briefings in Bioinformatics | Oxford Academic. https://academic.oup.com/bib/article/14/2/178/208453.

77. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaSci* **4**, 7 (2015).

78. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

79. Browning, S. R. & Browning, B. L. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics* **81**, 1084–1097 (2007).

80. Browning, B. L. & Browning, S. R. Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data. *Genetics* **194**, 459–471 (2013).

81. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–909 (2006).

82. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–64 (2009).

83. Zhou, H., Alexander, D. & Lange, K. A quasi-Newton acceleration for high-dimensional optimization algorithms. *Stat Comput* **21**, 261–273 (2011).

84. Rosenberg, N. A. Standardized Subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, Accounting for Atypical and Duplicated Samples and Pairs of Close Relatives. *Annals of Human Genetics* **70**, 841–847 (2006).

85. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–4 (2008).

86. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

87. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics* **93**, 278–288 (2013).

88. Martin, A. R. *et al.* Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *The American Journal of Human Genetics* **100**, 635–649 (2017).