The Company of **Biologists**

RESEARCH ARTICLE

Dissecting dynamics and differences of selective pressures in the evolution of human pigmentation

Xin Huang[1,*], Sijia Wang[1,2], Li Jin[1,2,*] and Yungang He[3,*]

ABSTRACT

Human pigmentation is a highly diverse and complex trait among populations and has drawn particular attention from both academic and non-academic investigators for thousands of years. Previous studies detected selection signals in several human pigmentation genes, but few studies have integrated contribution from multiple genes to the evolution of human pigmentation. Moreover, none has quantified selective pressures on human pigmentation over epochs and between populations. Here, we dissect dynamics and differences of selective pressures during different periods and between distinct populations with new approaches. We use genotype data of 19 genes associated with human pigmentation from 17 publicly available datasets and obtain data for 2346 individuals of six representative population groups from across the world. Our results quantify the strength of natural selection on light pigmentation not only in modern Europeans (0.0259/generation) but also in proto-Eurasians (0.00650/generation). Our results also suggest that several derived alleles associated with human dark pigmentation may be under positive directional selection in some African populations. Our study provides the first attempt to quantitatively investigate the dynamics of selective pressures during different time periods in the evolution of human pigmentation.

This article has an associated First Person interview with the first author of the article.

INTRODUCTION

Human pigmentation – the colour of human skin, hair, and eye – is one of the most diverse traits among populations. Its obvious diversity has attracted attention from both academic and non-academic investigators for thousands of years, as noted by Charles Darwin one century ago (Darwin, 1889) and as noticed by ancient Egyptians more than 4000 years ago (Norton, 2005). Why human pigmentation diverges, however, remains a central puzzle in human biology (Rees and Harding, 2012). Some researchers have proposed that the diversity of human pigmentation is adapted for the global difference in ultraviolet radiation (UVR) and driven by natural selection (Jablonski and Chaplin, 2000; Barsh, 2003; Parra, 2007; Jablonski and Chaplin, 2010). Dark skin may prevent sunburn amongst individuals in low latitude areas with high UVR, while light skin may protect infants against rickets in high latitude areas with low UVR (Jablonski and Chaplin, 2000; Parra, 2007; Chaplin and Jablonski, 2009; Jablonski and Chaplin, 2014, 2017; Cuthill, et al., 2017; Hochberg and Hochberg, 2019; Wolf and Kenney, 2019). Human pigmentation, especially skin pigmentation, is one of the traits that are under strong natural selection during the human dispersal out of Africa, because it is the first barrier between human body and living environment. A better understanding of how natural selection shapes the diversity of human pigmentation could provide relevant and beneficial information for public health (Jablonski and Chaplin, 2000; Barsh, 2003; Parra, 2007).

During the last decade, many studies have applied statistical tests to detect signals of natural selection in several human pigmentation genes (Izagirre et al., 2006; Voight et al., 2007; Lao et al., 2007; Myles et al., 2007; Norton et al., 2007; Pickrell et al., 2009; Beleza et al., 2013; Hider et al., 2013). These genes encode different proteins, including: signal regulators – ASIP, KITLG, MC1R – stimulating the melanogenic pathway; possible enhancers – BNC2, HERC2 – regulating pigmentation gene expression; important enzymes – TYR, TYRP1, DCT – converting tyrosine into melanin; putative exchangers – OCA2, SLC24A4, SLC24A5, SLC45A2, TPCN2 – controlling the environment within melanosomes; and an exocyst complex unit and molecular motor – EXOC2, MYO5A – conveying vesicles and organelles within the cytoplasm (Abdel-Malek et al., 2001; Rebbeck et al., 2002; Duffy et al., 2004, 2007; Graf et al., 2005; Sulem et al., 2007, 2008; Anno et al., 2008; Han et al., 2008; Ito and Wakamatsu, 2008; Kayser et al., 2008; Sturm and Duffy, 2012; Visser et al., 2012, 2014; Guenther et al., 2014). These proteins work at different stages of the melanogenic pathway, illustrating that human pigmentation is a complex trait affected by multiple genes with different roles.

Previous studies applied two groups of methods to detect natural selection. One group of methods detects unusually long extended haplotype homozygosity (Izagirre et al., 2006; McEvoy et al., 2006; Voight et al., 2007; Lao et al., 2007; Myles et al., 2007; Norton et al., 2007; Pickrell et al., 2009; Donnelly et al., 2012; Beleza et al., 2013). The other group of methods identifies extremely local population differentiation (Izagirre et al., 2006; Lao et al., 2007; Myles et al., 2007; Norton et al., 2007; Pickrell et al., 2009; Hider et al., 2013). By applying both groups of methods, previous studies have aimed to interpret the evolution of individual pigmentation genes; however, few studies have integrated contributions from multiple genes to the evolution of human pigmentation. Moreover, none of these studies have quantitatively investigated the historical selective pressures of pigmentation genes during different epochs and compared the

[1]Chinese Academy of Sciences Key Laboratory of Computational Biology, Chinese Academy of Sciences-Max Planck Society Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [2]State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200433, China. [3]Key Laboratory of Medical Epigenetics and Metabolism, Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China.

*Authors for correspondence (xinhuang.res@gmail.com; lijin@fudan.edu.cn; heyungang@fudan.edu.cn)

X.H., 0000-0002-9918-9602

differences of selective pressures between distinct populations. Therefore, it is necessary to perform an extensive quantification of selective pressures on human pigmentation using a creative approach.

RESULTS
The model
On the basis of a previous study (He et al., 2015), we measure selective pressures by (genic) selection coefficients. For any single nucleotide polymorphism (SNP) L, we can estimate the expectation of the selection (coefficient) difference per generation between populations i and j by

$$d_{ij}(L) = \left[\ln \frac{p_i(L)}{q_i(L)} - \ln \frac{p_j(L)}{q_j(L)} \right] / t_{ij}, \qquad (1)$$

where p and q are the frequencies of derived and ancestral alleles in a population, respectively; and t_{ij} is the divergence time of the populations i and j from their most recent common ancestor. Details of the calculations are described elsewhere (He et al., 2015).

We can extend Eqn 1 to summarise selection differences of multiple SNPs by assuming additive fitness and weak linkage disequilibrium. In general, for a trait with n SNPs, the expectation and variance of the total selection difference in this trait is

$$d_{ij}(L_1, L_2, \ldots, L_n) = \sum_{i=1}^{n} d_{ij}(L_k)$$

$$\text{var}[d_{ij}(L_1, L_2, \ldots, L_n)] = \sum_{i=1}^{n} \text{var}[d_{ij}(L_k)]. \qquad (2)$$

Here, we take two bi-allelic SNPs as an example. Using Eqn 1, we can estimate the selection difference of a trait affected by two SNPs between populations i and j using

$$d_{ij}(L_1, L_2) = \left[\ln \frac{p_i(L_1, L_2)}{q_i(L_1, L_2)} - \ln \frac{p_j(L_1, L_2)}{q_j(L_1, L_2)} \right] / t_{ij},$$

where $p(L_1, L_2)$ is the frequency of the combination carrying two alleles associated with one possible outcome of a trait, such as light pigmentation; $q(L_1, L_2)$ is the frequency of the combination carrying two alleles associated with another outcome of the same trait, such as dark pigmentation. With linkage equilibrium between L_1 and L_2, we have $p(L_1, L_2){=}p(L_1)p(L_2)$ and $q(L_1, L_2){=}q(L_1)q(L_2)$. Thus

$$d_{ij}(L_1, L_2) = d_{ij}(L_1) + d_{ij}(L_2).$$

We further assume the distribution of selection difference in each SNP is independent. Therefore, the variance of $d_{ij}(L_1, L_2)$ is

$$\text{var}[d_{ij}(L_1, L_2)] = \text{var}[d_{ij}(L_1)] + \text{var}[d_{ij}(L_2)].$$

The confidence interval (CI) of $d_{ij}(L_1, L_2)$ becomes $d_{ij}(L_1, L_2) \pm 1.96 \sqrt{\text{var}[d_{ij}(L_1, L_2)]}$.

Based on Eqn 2, we develop a new approach for dissecting historical selective pressures over epochs of the human evolutionary history. We simplify the evolutionary history of six representative human populations as a binary tree (Fig. 1). We can further divide the selection difference between paired populations into multiple terms if there are multiple branches between them. We further assume the selective pressure on each branch is constant over time. Let k denote the most recent common ancestor of i and j, we can divide d_{ij} in Eqn 1 into separate terms:

$$d_{ij}t_{ij} = \left(\ln \frac{p_i}{q_i} - \ln \frac{p_k}{q_k} \right) - \left(\ln \frac{p_j}{q_j} - \ln \frac{p_k}{q_k} \right).$$

For example, using the notations and demographic model in Fig. 1, we can estimate the total selection difference between East and West Africans as

$$d_{21}t_1 = -\left(\ln \frac{p_1}{q_1} - \ln \frac{p_r}{q_r} \right) + \left(\ln \frac{p_2}{q_2} - \ln \frac{p_w}{q_w} \right)$$
$$+ \left(\ln \frac{p_w}{q_w} - \ln \frac{p_r}{q_r} \right).$$

Let $s_1 = \left(\ln \frac{p_1}{q_1} - \ln \frac{p_r}{q_r} \right) / t_1, s_2 = \left(\ln \frac{p_2}{q_2} - \ln \frac{p_w}{q_w} \right) / t_2, s_{10} = \left(\ln \frac{p_w}{q_w} - \ln \frac{p_r}{q_r} \right) / (t_1 - t_2)$, then we have

$$d_{21}t_1 = -s_1 t_1 + s_2 t_2 + s_{10}(t_1 - t_2).$$

Therefore, we can represent the selection difference between paired populations as a combination of selection coefficients during different time periods. Using matrix algebra and the notations in Fig. 1, we can obtain Eqn 3, where

$$\mathbf{d} = \begin{pmatrix} d_{21}t_1 & d_{31}t_1 & \ldots & d_{65}t_5 \end{pmatrix}^{tr}$$

$$\mathbf{s} = \begin{pmatrix} s_1 & s_2 & \ldots & s_{10} \end{pmatrix}^{tr}$$

$$\mathbf{T} = \begin{bmatrix}
-t_1 & t_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & t_1-t_2 \\
-t_1 & 0 & t_3 & 0 & 0 & 0 & 0 & 0 & t_2-t_3 & t_1-t_2 \\
-t_1 & 0 & 0 & t_4 & 0 & 0 & 0 & t_3-t_4 & t_2-t_3 & t_1-t_2 \\
-t_1 & 0 & 0 & 0 & t_5 & 0 & t_4-t_5 & t_3-t_4 & t_2-t_3 & t_1-t_2 \\
-t_1 & 0 & 0 & 0 & 0 & t_5 & t_4-t_5 & t_3-t_4 & t_2-t_3 & t_1-t_2 \\
0 & -t_2 & t_3 & 0 & 0 & 0 & 0 & 0 & t_2-t_3 & 0 \\
0 & -t_2 & 0 & t_4 & 0 & 0 & 0 & t_3-t_4 & t_2-t_3 & 0 \\
0 & -t_2 & 0 & 0 & t_5 & 0 & t_4-t_5 & t_3-t_4 & t_2-t_3 & 0 \\
0 & -t_2 & 0 & 0 & 0 & t_5 & t_4-t_5 & t_3-t_4 & t_2-t_3 & 0 \\
0 & 0 & -t_3 & t_4 & 0 & 0 & 0 & t_3-t_4 & 0 & 0 \\
0 & 0 & -t_3 & 0 & t_5 & 0 & t_4-t_5 & t_3-t_4 & 0 & 0 \\
0 & 0 & -t_3 & 0 & 0 & t_5 & t_4-t_5 & t_3-t_4 & 0 & 0 \\
0 & 0 & 0 & -t_4 & t_5 & 0 & t_4-t_5 & 0 & 0 & 0 \\
0 & 0 & 0 & -t_4 & 0 & t_5 & t_4-t_5 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -t_5 & t_5 & 0 & 0 & 0 & 0
\end{bmatrix}.$$

Using matrix algebra, we can represent the selection differences of all the paired populations in Fig. 1 as

$$\mathbf{d} = \mathbf{Ts}. \qquad (3)$$

To obtain the optimal solution, we propose a probabilistic approach following the principle of minimum evolution (Cavalli-Sforza and Edwards, 1967; Edwards, 1996). Under neutral evolution (NE), we consider each estimated s as an independent random variable following a normal distribution with a mean of zero and a variance of σ^2. For each solution with ten variables, the summation below follows a chi-square distribution with ten degrees of freedom:

$$\sum_{i=1}^{10} \frac{s_i^2}{\sigma_i^2} \sim \chi^2(10).$$

Therefore, we have $\text{Pr}(|\mathbf{s}|^2 {>} |\mathbf{s}_a|^2 | \text{NE}) {\geq} \text{Pr}(|\mathbf{s}|^2 {>} |\mathbf{s}_b|^2 | \text{NE})$, if $|\mathbf{s}_a|^2 {\leq} |\mathbf{s}_b|^2$ for solutions a and b. Here, $|\mathbf{s}|^2$ is the norm of a vector $\mathbf{s}$. In other words, we can choose the most conservative solution with the least deviation from neutral evolution using a probabilistic approach. Thus, the optimal solution $\mathbf{s}^*$ is the minimum norm solution of Eqn 3:

$$\mathbf{s} = \mathbf{T}^+ \mathbf{d}, \qquad (4)$$

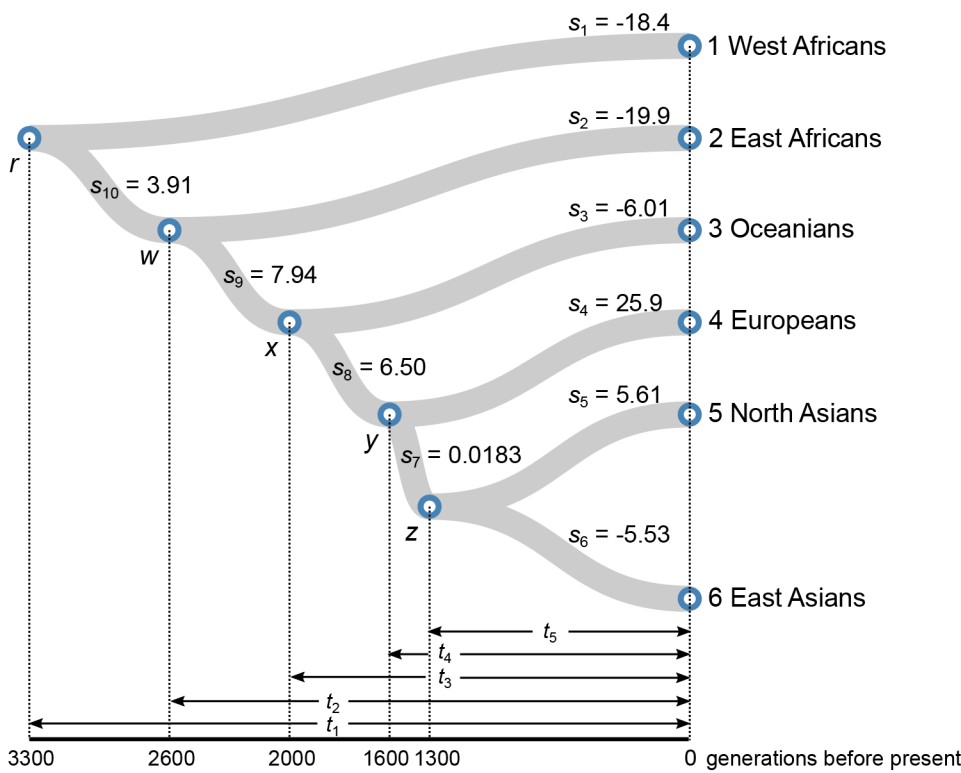where $\mathbf{T}^+$ is the Moore-Penrose inverse of $\mathbf{T}$.

Fig. 1. Time-varied selective pressures on an evolutionary tree. We model the evolutionary history of six representative human populations as a binary tree. Here, s_i (i=1, 2, …, 10) denotes the selection coefficient of the i-th epoch and can be obtained by estimating selection differences between paired populations. Divergence times between populations are based on previous studies (Schaffner et al., 2005; Oppenheimer, 2012; Schiffel and Durbin, 2014; Mondal et al., 2016). We assume one generation time is ∼30 years. We obtain the optimal solution deviated least from neutral evolution using a probabilistic approach. The numbers (×10^{-3}/generation) on the branches are the optimal solution. In the solution, numbers larger than zero suggest natural selection favoured light pigmentation, while those less than zero indicate natural selection preferred dark pigmentation.

Selective pressures over epochs

We applied our new approach with genotype data of worldwide populations from 17 publicly available datasets (Table S1). After data preparation (Materials and Methods), we obtained 2346 individuals and grouped them into six population groups based on their geographic locations (Table S2). We also selected 52 SNPs in 19 genes for analysis due to their association with human pigmentation in published genome-wide association studies (GWAS) or phenotype prediction models (Table S3; Rebbeck et al., 2002; Bonilla et al., 2005; Graf et al., 2005; Lamason et al., 2005; Stokowki et al., 2007; Miller et al., 2007; Anno et al., 2008; Han et al., 2008; Kayser et al., 2008; Sturm et al., 2008; Sulem et al., 2008; Eiberg et al., 2008; Branicki et al., 2009; Edwards et al., 2010; Branicki et al., 2011; Donnelly et al., 2012; Visser et al., 2012; Hart et al., 2013; Jacobs et al., 2013; Praetorius et al., 2013; Walsh et al., 2013; Guenther et al., 2014; Murray et al., 2015; Yang et al., 2016; Ainger et al., 2017; Crawfold et al., 2017). We then used Eqn 2 with 30 SNPs not in strong linkage disequilibrium (r^2<0.8) to estimate the total selection differences on human pigmentation (Materials and Methods). The maximum differences were observed between Europeans and the two African populations, while the minimum difference was observed between West and East Africans (Table 1). The estimated 95% confidence intervals (CI) indicate we cannot rule out the possibility that genetic drift caused the difference

between East and West Africans, as well as between Oceanians and East Asians (Table 1). We further assessed the significance levels of the observed selection differences by randomly sampling 10,000 sets of 30 SNPs in the genome and obtained the empirical distributions of population differences (Fig. S1). The differences from random sets of SNPs are close to zero, which is consistent with a recent study (Fortier et al., 2019, preprint) that suggests no genome-wide difference in the strength of natural selection between human populations. Whereas those from SNPs associated with human pigmentation are significantly departure from zero (Fig. S1), indicating that most population differences on SNPs associated with human pigmentation are possibly contributed by natural selection.

We then solved the linear system (Eq. 4) with the observed selection differences on human pigmentation (Table 1). Our estimate shows that the modern European lineage had the largest selective pressure (s_4=0.0259/generation) on light pigmentation than the other branches (Fig. 1), suggesting that recent natural selection favoured light pigmentation in Europeans. Recent studies using ancient DNA could support our observation of recent directional selection in Europeans (Wilde et al., 2014; Mathieson et al., 2015). Our results also reveal the selective pressure on light pigmentation in the ancestral population of Europeans and East Asians (s_8=0.00650/generation). This shared selection is also supported by other studies, revealing that *ASIP*, *BNC2*, and

Table 1. Selection differences with 95% CI associated with human pigmentation between populations (×10^{-3}/generation)

	WAF	EAF	OCN	EUR	NAS	EAS
WAF	0	−3.626±3.728	−17.077±6.544	−34.059±5.353	−23.723±6.276	−19.331±6.003
EAF	3.626±3.728	0	−17.076±8.294	−38.628±6.814	−25.504±7.987	−19.936±7.654
OCN	17.077±6.544	17.076±8.294	0	−28.016±8.595	−10.959±8.990	−3.721±8.422
EUR	34.059±5.353	38.628±6.814	28.016±8.595	0	21.321±9.287	30.372±8.790
NAS	23.723±6.276	25.504±7.987	10.959±8.990	−21.321±9.287	0	11.137±8.681
EAS	19.331±6.003	19.936±7.654	3.721±8.422	−30.372±8.790	−11.137±8.681	0

Populations are abbreviated as follows: WAF, West Africans; EAF, East Africans; OCN, Oceanians; EUR, Europeans; NAS, North Asians; EAS, East Asians.

KITLG were under directional selection before the divergence of ancestral Europeans and East Asians (Donnelly et al., 2012; Beleza et al., 2013). We further applied SLiM 2 to examine whether the optimal solution could reproduce the observed selection differences (Haller and Messer, 2017) (Table 1). We set up a human demographic model according to previous studies and used the optimal solution as selection coefficients during different periods (Materials and Methods). The simulated selection differences are close to the data and little affected by the initial frequency of the beneficial allele (Fig. S2). This also illustrates that though we assume genic selection, our model could approximate genotypic selection in diploids (Materials and Methods).

Selection differences between populations

We also separately quantified selection differences of individual SNPs associated with human pigmentation (Table S4) using Eq. 1. Ten SNPs were removed because of their low derived allele frequencies among populations in our data (Materials and Methods). Statistical tests suggest that selective pressures in many loci differed significantly between populations ($P<0.05$). The remaining 42 SNPs were categorised into five groups (Fig. 2).

In the first group, derived alleles may be affected by Eurasian-shared selection (Fig. 2A). Among these SNPs, rs6119471 (*ASIP*) has large selection differences between Eurasians and Africans (Table S4). The derived allele of rs6119471 (*ASIP*) is almost fixed across Eurasians but maintains a low frequency in Africans (Fig. S4). Recent studies applied this SNP to predict dark/non-dark pigmentation phenotype in human (Spichenok et al., 2011). This may be explained by different selective pressures on this SNP among populations. Our results also indicate that two SNPs in *MC1R* (rs2228479 and rs885479) largely differ between Eurasians and Africans (Table S4). Previous studies used variants in *MC1R* to solve a long-standing puzzle, regarding whether light pigmentation in low UVR areas is caused by directional selection or the relaxation of selective pressures (Rana et al., 1999; Harding et al., 2000; Wilde et al., 2014). The relaxation of selective pressures would suggest that the diversity of *MC1R* variants increased in Eurasians due to the lack of selective constraints. In this scenario, the genetic diversity of *MC1R* variants could be largely attributed to genetic drift. In contrast, directional selection would suggest that *MC1R* variants were under positive selection in Eurasians. In this scenario, genetic drift cannot explain the genetic divergence of *MC1R* between Africans and Eurasians. Our statistical results show that the divergences of rs2228479 and rs885479 between Eurasians and Africans are highly significant departure from neutral evolution (Table S4), suggesting that directional selection is the more likely explanation. Experimental evidence suggests that the derived allele of rs2228479 could cause lower affinity for alpha-melanocyte stimulating hormone than the ancestral allele (Xu et al., 1996). Another study showed that the derived allele of rs885479 carries a lower risk of developing freckles and severe solar lentigines than the ancestral allele in East Asians (Motokawa et al., 2007). These studies revealed the potential roles of these *MC1R* variants in pigmentation phenotypes.

In the second group, derived alleles may be affected by African-specific selection (Fig. 2B). All these derived alleles are in/near two genes (*DDB1* and *MFSD12*) and were recently associated with human dark pigmentation (Crawfold et al., 2017). The previous study (Crawfold et al., 2017) did not find signals of positive selection at *MFSD12* using Tajima's D or iHS. Our method (He et al., 2015; Huang et al., 2019) shows that these SNPs in *MFSD12* differ significantly between Africans and Eurasians, possible signals of directional selection (Table S4). From the first and second groups, we can observe that directional selection not only affects derived alleles associated with light pigmentation in Eurasians, but also influences derived alleles associated with dark pigmentation in Africans. This observation suggests that human pigmentation is under directional selection with diversifying orientations among different populations. Thus, the previous view that the dark pigmentation in Africans is the result of purifying selection on ancestral alleles is incomplete.

The third and fourth groups display European- and Asian-specific selection, respectively (Fig. 2C and D). One notable SNP is rs1426654 (*SLC24A5*), which had the largest selection difference between Europeans and East Asians in our study (0.005774/ generation). Previous studies reported that this SNP is under strong directional selection in Europeans (Izagirre et al., 2006; Voight et al., 2007; Lao et al., 2007; Myles et al., 2007; Norton et al., 2007). Another notable SNP is rs1800414 (*OCA2*), which has large selection differences between East Asians and other populations. This reveals a potential role of rs1800414 (*OCA2*) on light pigmentation in East Asians. Several studies have suggested directional selection on this SNP in East Asians (Edwards et al., 2010; Yang et al., 2016). These large selection differences indicate the significant contributions of these SNPs to light pigmentation in Europeans and East Asians, respectively. Other SNPs in these groups also support the hypothesis that recent natural selection for light pigmentation independently occurred in Europeans and Asians since they diverged (Norton et al., 2007; Edwards et al., 2010; Yang et al., 2016). Interestingly, Oceanians comprise both African-specific (*DDB1*) and Asian-specific (*OCA2*) selection. However, due to limited sample size of Oceanians in our data from publicly available resources (Table S2), it should be cautious to interpret these results. It would be helpful to analyse larger datasets of Oceanians to confirm our observation.

The last group includes the five remaining SNPs (Fig. 2E), which exhibit specific selection differences between limited population pairs. Among them, the derived allele of rs1800401 (*OCA2*) and the ancestral allele of rs12896399 (*SLC24A4*) are both associated with dark pigmentation (Table S2). Only rs12896399 (*SLC24A4*) differs significantly between West Africans and Eurasians (Table S4). This may be a selection signal associated with dark pigmentation in West Africans, again indicating possible genetic diversity within African populations. We note that rs35264875 (*TPCN2*) and rs12821256 (*KITLG*) might be affected by selection in both East Africans and Europeans. A recent study showed that rs12821256 might have large effect on the skin pigmentation in South Africans (Martin et al., 2017). The other two SNPs, rs3829241 (*TPCN2*) and rs642742 (*KITLG*), also differ between Eurasians and Africans (Fig. 2A). These similar patterns of *TPCN2* and *KITLG* might suggest some connection between them.

DISCUSSION

Compared with previous studies (Izagirre et al., 2006; Voight et al., 2007; Lao et al., 2007; Myles et al., 2007; Norton et al., 2007; Pickrell et al., 2009; Beleza et al., 2013; Hider et al., 2013; Wilde et al., 2014), our study has three advantages. First, our approach considers the fluctuation of selective pressures over epochs, an important factor in evolution (Crow and Kimura, 2009) that was ignored by previous studies. Our results provide more information about the dynamics of selective pressures during human evolution. Second, we summarise selective pressures based on multiple human pigmentation genes (Eq. 2), while previous studies usually tested selection signals in individual human pigmentation genes. Moreover, we simultaneously interpret selective pressures in multiple populations, whereas previous studies separately
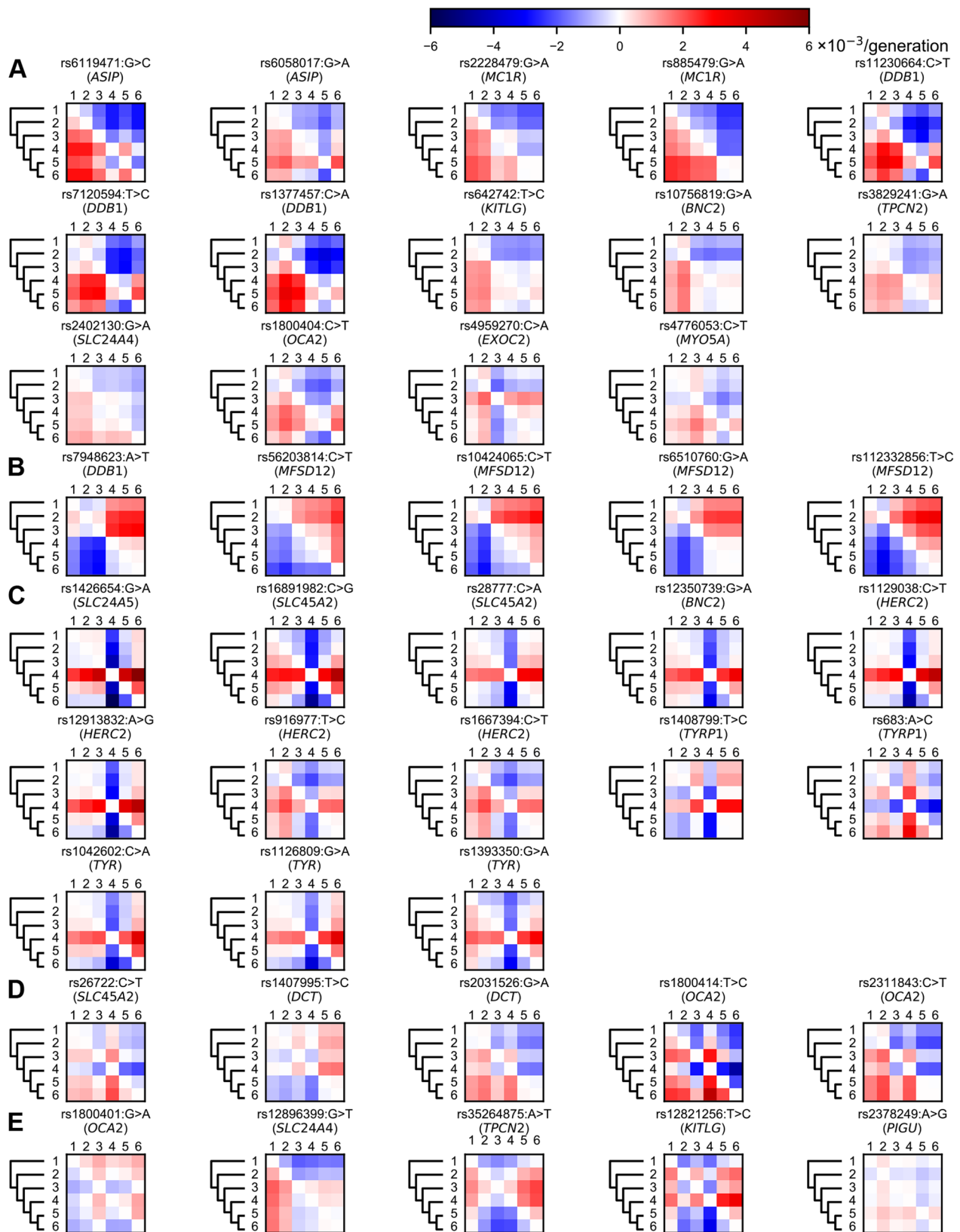
Fig. 2. Selection differences in individual loci between populations. We used Eqn 1 to quantify the selection differences of 42 SNPs associated with human pigmentation, and categorised them into five kinds of patterns: (A) derived alleles affected by Eurasian-shared selection; (B) derived alleles affected by African-specific selection; (C) alleles affected by European-specific selection; (D) alleles affected by Asian-specific selection; and (E) others. Red colour indicates selective pressures of populations in rows are larger than those in columns; blue colour indicates selective pressures of populations in rows are smaller than those in columns. Populations are abbreviated as follows: 1, West Africans (n=493); 2, East Africans (n=59); 3, Oceanians (n=32); 4, Europeans (n=701); 5, North Asians (n=114); 6, East Asians (n=947).

investigated selection signals in single population. Third, we do not need to assume population continuity as in those ancient DNA studies (Wilde et al., 2014; Mathieson et al., 2015), because our study is based on genetic data from only present-day populations.

We note that our investigation has several limitations. First, our model is based on the infinite population size model. The limited sample size would affect our results, therefore, we grouped populations into large population groups based on their geographic locations to mitigate the effect of sample size. Analysis of data with larger sample size could improve our estimate, as more and more genomic datasets become available. Second, although we chose the solution that deviates least from neutral evolution as the optimal solution, we cannot exclude the possibility of other solutions. This reflects the difficulty of analysing historical selective pressures, which is a well-recognised challenge in population genetics (Crow and Kimura, 2009). Our solution provides a first step toward resolving the dynamics of selection in the evolution of human pigmentation. This solution may be improved by combining both ancient and modern human genetic data, as well as by using a Bayesian framework for inference. Adding more population groups would also possibly improve the solution, because this would provide more constraints in the linear system (Eq. 4). Third, our results may be affected by a severe bottleneck. A recent study (Terhorst et al., 2017) suggests a more severe Out-of-Africa bottleneck in human evolutionary history than in the model used in our simulation. This would probably reduce the selection differences between Eurasians and Africans, leading to an underestimation of selective pressures. Fourth, our results may also be affected by population migration and sub-structure. We used knowledge from previous studies, principle component analysis and F_3 test to rigorously prune potential admixed populations, including South Asians, Central Asians, the Middle East People and Americans. Removing these populations would lose information of selective pressures on human pigmentation in these lineages; however, as a first step to explore the historical selective pressures in the evolution of human pigmentation, we focused more on reducing the bias induced by population admixture. New methods explicitly accounting for population admixture would be helpful to provide more comprehensive view on the dynamics of selective pressures during the evolution of human pigmentation. Besides, we demonstrate that our estimate provides lower bounds of selection differences on human pigmentation when migration or sub-structure exists (Materials and Methods). Fourth, the human pigmentation SNPs used in our study may be biased. For example, our results indicate small genetic differences on human pigmentation between Oceanians and East Asians (Table 1), while recent studies (Martin et al., 2017) suggest Oceanians are darker than East Asians in skin pigmentation using melanin index. One possible reason is that some Oceanian-specific or East-Asian-specific SNPs are missing. This is because we selected candidates based on results from published GWAS or phenotype prediction models, and most of these studies used samples with European ancestry (Sirugo et al., 2019). More studies on non-European populations could resolve this missing diversity and enhance our knowledge on the evolution of human pigmentation. Finally, we noticed that our model is a simple model. Other biological factors, such as linkage disequilibrium between SNPs, sexual selection, and different levels of vitamin D among human populations, may be possible to be integrated into a more comprehensive model based on this simple model.

To summarise, we extended an established method (He et al., 2015) to dissect dynamics of selective pressures over epochs. Our study provides the first attempt to resolve time-varied selective pressures in the evolution of human pigmentation. Our study also provides information on differences of selective pressures between distinct population groups. Further studies are in progress to verify our present views on the evolution of human pigmentation.

MATERIALS AND METHODS
Data preparation
Seventeen datasets (Li et al., 2008; Teo et al., 2009; Behar et al., 2010; Rasmussen et al., 2010; The, 1000 Genomes Project Consortium, 2010; The International HapMap 3 Consortium, 2010; Metspalu et al., 2011; Pagani et al., 2012; Yunusbayev et al., 2012; Di Cristofaro et al., 2013; Fedorova et al., 2013; Xing et al., 2013; Kovacevic et al., 2014; Raghavan et al., 2014; Yunusbayev et al., 2015; Mondal et al., 2016; Pagani et al., 2016) containing genotype data from worldwide human populations were obtained from the listed resources (Table S1). After downloading, all the genotype data were liftovered to genomic coordinates using the Human Reference Genome Hg19. A merged dataset containing 6531 individuals was obtained after removing duplicated and related individuals. After merging, SNPs with call rate less than 0.99 or individuals with call rate less than 0.95 were removed. SNPs in strong linkage disequilibrium were further removed by applying a window of 200 SNPs advanced by 25 SNPs and an r^2 threshold of 0.8 (–indep-pairwise 200 25 0.8) in PLINK 1.7 (Purcell et al., 2007). This LD-pruning was applied to each population separately. The remaining 61,597 SNPs were used for further analysis. In order to mitigate the bias induced by population migration, potential admixed populations, such as the Middle East People and South Asians, were excluded according to previous studies (Li et al., 2008; Teo et al., 2009; Behar et al., 2010; Rasmussen et al., 2010; The 1000 Genomes Project Consortium, 2010; The International HapMap 3 Consortium, 2010; Metspalu et al., 2011; Pagani et al., 2012; Yunusbayev et al., 2012; Di Cristofaro et al., 2013; Fedorova et al., 2013; Xing et al., 2013; Kovacevic et al., 2014; Raghavan et al., 2014; Yunusbayev et al., 2015; Mondal et al., 2016; Pagani et al., 2016), principal component analysis (PCA) using SMARTPCA (version: 13050) from EIGENSOFT (version: 6.0.1) (Patterson et al., 2006; Price et al., 2006), and F_3 test using ADMIXTOOLS (version: 3.0) (Patterson et al., 2012). Finally, 2346 individuals were obtained and divided into six groups according to their geographic regions for further analysis. These groups are West Africans, East Africans, Oceanians, Europeans, North Asians and East Asians. The PCA plot (Fig. S3) shows that these 2346 individuals were properly separated into six population groups.

Data imputation
Genotypes of 19 human pigmentation genes with 500-kb flanking sequences on both sides were obtained from the genotype datasets. Haplotype inference and genotype imputation were performed on the selected genotypes using BEAGLE 4.1 (Browning and Browning, 2007, 2016) with 1000 Genomes phase 3 haplotypes as the reference panel. During phasing and imputation, the effective population size was assumed to be 10,000 (N_e=10,000), and the other parameters were set to the default values. Ten SNPs (rs1110400, rs11547464, rs12203592, rs1800407, rs1805005, rs1805006, rs1805007, rs1805008, rs1805009, rs74653330) were removed because of their low derived allele frequencies in our datasets after imputation (Fig. S4). Because rs12203592 (*IRF4*) was removed, 18 genes with the remaining 42 SNPs were used for further analysis.

Estimating selection differences between populations and selective pressures over epochs
We used Eqn 1 to estimate the selection differences of the remaining 42 SNPs. We then used Eqn 2 and selected 30 SNPs not in strong linkage disequilibrium (r^2<0.8) as well as known phenotypes to estimate the total selection differences on human pigmentation between populations. These SNPs were rs3829241, rs56203814, rs916977, rs1800414, rs10424065, rs6119471, rs1408799, rs11230664, rs4959270, rs1800401, rs2378249, rs1042602, rs12350739, rs6058017, rs12821256, rs1393350, rs1426654, rs642742, rs6510760, rs1129038, rs2228479, rs35264875, rs12896399, rs26722, rs16891982, rs885479, rs28777, rs1800404, rs10756819, rs2402130. To dissect selective pressures over epochs, we applied Eqn 4 with the total selection differences from the selected 30 SNPs and the divergence times shown in Fig. 1.

Reproducing the observed selection differences from the optimal solution

We used SLiM 2 (version: 2.6) (Haller and Messer, 2017) to simulate a demographic model of human evolution (Fig. S5) to examine whether the optimal solution could reproduce the observed selection differences. We varied the initial frequency of the beneficial allele with 0.001 and 0.01. We divided the optimal solution by 30 to obtain the average selection coefficient for each SNP, because we used 30 SNPs to estimate the total selection differences on human pigmentation. We used the effective population size of each population estimated by previous studies (McEvoy et al., 2011; Mezzavilla and Ghirotto, 2015). We set both the mutation rate and the recombination rate to 1×10^{-8} per generation per site. In each run, we simulated a fragment with 10^6 base pairs, and set the 50,000th site under selection. We repeated each set of parameters more than 10,000 times and analysed those results in which beneficial alleles were not fixed or lost in all the populations. We compared the average selection differences from simulation with the observed selection differences. We noticed that the selection coefficient in SLiM 2 measures differences in fitness between genotypes instead of alleles. We can transform the selection coefficient of genotypes into that of alleles as follows. Let the fitness of the ancestral allele A be 1, and the relative fitness of the derived allele a be e^s. When s is close to 0, we can approximate e^s as $1+s$ using the Taylor series. The fitness of genotype aa becomes $(1+s)^2=1+2s+s^2{\approx}1+2s$, and the fitness of genotype Aa is $1+s=1+0.5s'$. If s' is the selection coefficient in SLiM 2, then $s'=2$ s; and the dominance coefficient becomes 0.5. Simulations were performed in Digital Ocean (https://cloud.digitalocean.com/) Optimized Droplets. The information of these droplets is as follows: CPU, Intel® Xeon® Platinum 8168 Processor; Random-access memory, 64 GB; Operating system, Ubuntu 16.04.4×64.

The effects of population migration and substructure

In this section, we examine how the estimated selection difference $\hat{d}_{ij}=\left(\ln\dfrac{\hat{p}_i}{\hat{q}_i}-\ln\dfrac{\hat{p}_j}{\hat{q}_j}\right)/t$ is affected by population migration and substructure in theory. Here, $\hat{p}_i$ and $\hat{q}_i$ are the observed derived and ancestral allele frequencies in the population i, respectively; $\hat{p}_j$ and $\hat{q}_j$ are the observed derived and ancestral allele frequencies in the population j, respectively; and t is the divergence time from populations i and j to their most recent common ancestor. We demonstrate that $\hat{d}_{ij}$ provides a lower bound of selection difference between populations i and j when migration or substructure exists. We first provide two inequalities that will be used later.

Inequality 1: If $a>b>0$, $c>d>0$, then $ac>bd$ and $\dfrac{a+d}{b+c}<\dfrac{a}{b}$.

Proof: $a>b>0$, $c>d>0$, then $ac>bc$, $bc>bd$. Therefore, $ac>bd$. Furthermore, $ac+ab>bd+ab$, which is the same as $a(b+c)>b(a+d)$. Therefore, $\dfrac{a+d}{b+c}<\dfrac{a}{b}$.

Inequality 2: If $a_1>0$, $a_2>0$, L, $a_n>0$, $\bar{a}=\sum\limits_{i=1}^{n}a_i/n$, then $\max a_1, a_2, L, a_n\} \ge \bar{a}$ $\min a_1, a_2, L, a_n\} \le \bar{a}$.

Proof: Let $a_{\max}=\max\{a_1, a_2, L, a_n\}$, then

$$a_{\max}-\bar{a}=\left(\sum_{i=1}^{n}a_{\max}-\sum_{i=1}^{n}a_i\right)/n$$
$$=\sum_{i=1}^{n}(a_{\max}-a_i)/n\ge 0$$

because $a_{\max}\ge a_i$.

Let $a_{\min}=\min\{a_1, a_2, ..., a_n\}$, then

$$\bar{a}-a_{\min}=\left(\sum_{i=1}^{n}a_i-\sum_{i=1}^{n}a_{\min}\right)/n$$
$$=\sum_{i=1}^{n}(a_i-a_{\min})/n\le 0$$

because $a_{\min}\le a_i$.

For the proofs in below, we assume $\hat{d}_{ij}>0$ without loss of generality. If $\hat{d}_{ij}<0$, then we can exchange i and j, and still obtain $\hat{d}_{ji}>0$.

The effect of migration

Suppose there is α proportion of individuals in the population i, which actually come from the population j; also, there is β proportion of individuals in the population j, which actually come from the population i. Here, $\alpha, \beta\in\left[0,\dfrac{1}{2}\right)$, because we assume migrants should not become the majority of another population. Then we have

$$\hat{p}_i=(1-\alpha)p_i+\alpha p_j$$
$$\hat{q}_i=(1-\alpha)q_i+\alpha q_j$$
$$\hat{p}_j=(1-\beta)p_j+\beta p_i$$
$$\hat{q}_j=(1-\beta)q_j+\beta q_i$$

Here, p_i and q_i are the true derived and ancestral allele frequencies in the population i, respectively; p_j and q_j are the true derived and ancestral allele frequencies in the population j, respectively. Therefore,

$$\hat{p}_i\hat{q}_j=[(1-\alpha)p_i+\alpha p_j][(1-\beta)q_j+\beta q_i]$$
$$=(1-\alpha)(1-\beta)p_iq_j+\alpha(1-\beta)p_jq_j+\beta(1-\alpha)p_iq_i+\alpha\beta q_ip_j$$
$$\hat{q}_i\hat{p}_j=[(1-\alpha)q_i+\alpha q_j][(1-\beta)p_j+\beta p_i]$$
$$=(1-\alpha)(1-\beta)q_ip_j+\alpha(1-\beta)p_jq_j+\beta(1-\alpha)p_iq_i+\alpha\beta p_iq_j$$

Further, we have

$$\hat{p}_i\hat{q}_j-\hat{q}_i\hat{p}_j=(1-\alpha)(1-\beta)(p_iq_j-q_ip_j)+\alpha\beta(q_ip_j-p_iq_j)$$
$$=[(1-\alpha)(1-\beta)-\alpha\beta](p_iq_j-q_ip_j)$$
$$=(1-\alpha-\beta)(p_iq_j-q_ip_j)$$

Because $\alpha, \beta\in\left[0,\dfrac{1}{2}\right)$, then $1-\alpha-\beta>0$; and $\hat{p}_i\hat{q}_j>\hat{q}_i\hat{p}_j$, therefore, $p_iq_j-q_ip_j>0$.

We also have $\dfrac{\hat{p}_i}{\hat{q}_i}=\dfrac{(1-\alpha)p_i+\alpha p_j}{(1-\alpha)q_i+\alpha q_j}=\dfrac{p_i+(\alpha/(1-\alpha))p_j}{q_i+(\alpha/(1-\alpha))q_j}$. Because $p_iq_j-q_ip_j>0$, we have $p_i\dfrac{\alpha}{1-\alpha}q_j>q_i\dfrac{\alpha}{1-\alpha}p_j$. From Inequality 1, we know $\dfrac{\hat{p}_i}{\hat{q}_i}<\dfrac{p_i}{q_i}$. Similarly, we also have $\dfrac{\hat{q}_j}{\hat{p}_j}=\dfrac{q_j+(\beta/(1-\beta))q_i}{p_j+(\beta/(1-\beta))p_i}<\dfrac{q_j}{p_j}$, therefore, $\dfrac{\hat{p}_i\hat{q}_j}{\hat{q}_i\hat{p}_j}<\dfrac{p_iq_j}{q_ip_j}$. According to the monotony of the logarithmic function, we have $0<\ln\dfrac{\hat{p}_i\hat{q}_j}{\hat{q}_i\hat{p}_j}<\ln\dfrac{p_iq_j}{q_ip_j}$; thus, $0<\hat{d}_{ij}<d_{ij}$. In other words, if migration exists between populations i and j, the estimated selection difference is lower than the true value.

The effect of substructure

Scenario 1: The population j has k subpopulations.

If the population j has k subpopulations, then $\hat{p}_j=\dfrac{\sum_k N_{jk}p_{jk}}{N_j}$. Here, p_{jk} is the derived allele frequency in the subpopulation k of the population j. And N_j is the population size of the population j, $N_j=\sum_k N_{jk}$. We denote the minimum of p_{jk} as $\min(p_{jk})$. Because $q_{jk}=1-p_{jk}$, then $\max(q_{jk})=1-\min(p_{jk})$. Based on Inequality 2, $\hat{p}_j>\min(p_{jk})$, $\hat{q}_j<\max(q_{jk})$. Therefore, $\ln\dfrac{\hat{p}_j}{\hat{q}_j}>\ln\dfrac{\min(p_{jk})}{\max(q_{jk})}$. We have $\ln\dfrac{\hat{p}_i}{\hat{q}_i}-\ln\dfrac{\min(p_{jk})}{\max(q_{jk})}>\ln\dfrac{\hat{p}_i}{\hat{q}_i}-\ln\dfrac{\hat{p}_j}{\hat{q}_j}>0$.

Scenario 2: The population i has l subpopulations.

If the population i has l subpopulations, then $\hat{p}_i=\dfrac{\sum_l N_{il}p_{il}}{N_i}$. We denote the maximum of p_{il} as $\max(p_{il})$. Then $\hat{p}_i<\max(p_{il})$, $\hat{q}_i>\min(q_{il})$. Therefore, $\ln\dfrac{\max(p_{il})}{\min(q_{il})}>\ln\dfrac{\hat{p}_i}{\hat{q}_i}$, and we have $\ln\dfrac{\max(p_{il})}{\min(q_{il})}-\ln\dfrac{\hat{p}_j}{\hat{q}_j}>\ln\dfrac{\hat{p}_i}{\hat{q}_i}-\ln\dfrac{\hat{p}_j}{\hat{q}_j}>0$.

Scenario 3: The population i has l subpopulations, and the population j has k subpopulations.

Based on scenarios 1 and 2, we have

$$\ln\frac{\max(p_{il})}{\min(q_{il})} - \ln\frac{\min(p_{jk})}{\max(q_{jk})} > \ln\frac{\hat{p}_i}{\hat{q}_i} - \ln\frac{\hat{p}_j}{\hat{q}_j} > 0.$$

In summary, if populations i and j have subpopulations, and their estimated selection difference is larger than zero, then at least one pair of their subpopulations has selection difference larger than zero. Moreover, the estimated difference is smaller than the largest difference between subpopulations.

Data availability
The publicly available genomic datasets used in this study can be found in Table S1 from the supplementary. The software used in this study are PLINK 1.7 (https://zzz.bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), ADMIXTOOLS 3.0 (https://github.com/DReichLab/AdmixTools), BEAGLE 4.1 (https://faculty.washington.edu/browning/beagle/b4_1.html), SLiM 2 (https://github.com/MesserLab/SLiM), and SeleDiff (https://github.com/xin-huang/SeleDiff).

Supplementary information
Supplementary information available online at
https://bio.biologists.org/lookup/doi/10.1242/bio.056523.supplemental

References
Abdel-Malek, Z. A., Scott, M. C., Furumura, M., Lamoreux, M. L., Ollmann, M., Barsh, G. S. and Hearing, V. J. (2001). The melanocortin 1 receptor is the principal mediator of the effects of agouti signaling protein on mammalian melanocytes. *J. Cell Biol.* **114**, 1019-1024.

Ainger, S. A., Jagirdar, K., Lee, K. J., Soyer, H. P. and Sturm, R. A. (2017). Skin pigmentation genetics for the clinic. *Dermatology* **233**, 1-15. doi:10.1159/000468538

Anno, S., Abe, T. and Yamamoto, T. (2008). Interactions between SNP alleles at multiple loci contribute to skin color differences between Caucasoid and Mongoloid subjects. *Int. J. Biol. Sci.* **4**, 81-86. doi:10.7150/ijbs.4.81

Barsh, G. S. (2003). What controls variation in human skin color? *PLoS Biol.* **1**, e27. doi:10.1371/journal.pbio.0000027

Behar, D. M., Yunusbayev, B., Mesplau, M., Metsplau, E., Rosset, S., Parik, J., Rootsi, S., Chaubey, G., Kutuev, I., Yudkovsky, G. et al. (2010). The genome-wide structure of the Jewish people. *Nature* **466**, 238-242. doi:10.1038/nature09103

Beleza, S., Santos, A. M., McEvoy, B., Alves, I., Martinho, C., Cameron, E., Shriver, M. D., Parra, E. J. and Rocha, J. (2013). The timing of pigmentation lightening in Europeans. *Mol. Biol. Evol.* **30**, 24-35. doi:10.1093/molbev/mss207

Bonilla, C., Boxill, L. A., Donald, S. A., Williams, T., Sylvester, N., Parra, E. J., Dios, S., Norton, H. L., Shriver, M. D. and Kittles, R. A. (2005). The 8818G allele of the agouti signaling protein (ASIP) gene is ancestral and is associated with darker skin color in African Americans. *Hum. Genet.* **116**, 402-406. doi:10.1007/s00439-004-1251-2

Branicki, W., Brudnik, U. and Wojas-Pelc, A. (2009). Interactions between *HERC2*, *OCA2* and *MC1R* may influence human pigmentation phenotype. *Ann. Hum. Genet.* **73**, 160-170. doi:10.1111/j.1469-1809.2009.00504.x

Branicki, W., Liu, F., van Duijn, K., Draus-Barini, J., Pośpiech, E., Walsh, S., Kupiec, T., Wojas-Pelc, A. and Kayser, M. (2011). Model-based prediction of human hair color using DNA variants. *Hum. Genet.* **129**, 443-454. doi:10.1007/s00439-010-0939-8

Browning, S. R. and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084-1097. doi:10.1086/521987

Browning, B. L. and Browning, S. R. (2016). Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116-126. doi:10.1016/j.ajhg.2015.11.020

Cavalli-Sforza, L. L. and Edwards, A. W. F. (1967). Phylogenetic analysis: models and estimation procedures. *Evolution* **21**, 550-570. doi:10.1111/j.1558-5646.1967.tb03411.x

Chaplin, G. and Jablonski, N. G. (2009). Vitamin D and the evolution of human depigmentation. *Am. J. Phys. Anthropol.* **139**, 451-461. doi:10.1002/ajpa.21079

Crawfold, N. G., Kelly, D. E., Hansen, M. E. B., Beltrame, M. H., Fan, S., Bowman, S. L., Jewett, E., Ranciaro, A., Thompson, S., Lo, Y. et al. (2017). Loci associated with skin pigmentation identified in African populations. *Science* **358**: eaan8433. doi:10.1126/science.aan8433

Crow, J. F. and Kimura, M. (2009). *An Introduction to Population Genetics Theory.* Caldwell, US: The Blackburn Press.

Cuthill, I. C., Allen, W. L., Arbuckle, K., Caspers, B., Chaplin, G., Hauber, M. E., Hill, G. E., Jablonski, N. G., Jiggins, C. D., Kelber, A. et al. (2017). The biology of color. *Science* **357**, eaan0221. doi:10.1126/science.aan0221

Darwin, C. (1889). *The Descent of Man, and Selection in Relation to Sex.* New York, US: D. Appleton and Company.

Di Cristofaro, J., Pennarun, E., Mazières, S., Myres, N. M., Lin, A. A., Temori, S. A., Metspalu, M., Metspalu, E., Witzel, M., King, R. J. et al. (2013). Afghan hindu kush: where Eurasian sub-continent gene flows converge. *PLoS ONE* **8**, e76748. doi:10.1371/journal.pone.0076748

Donnelly, M. P., Paschou, P., Grigorenko, E., Gurwitz, D., Barta, C., Lu, R.-B., Zhukova, O. V., Kim, J.-J., Siniscalco, M., New, M. et al. (2012). A global view of the *OCA2-HERC2* region and pigmentation. *Hum. Genet.* **131**, 683-696. doi:10.1007/s00439-011-1110-x

Duffy, D. L., Box, N. F., Chen, W., Palmer, J. S., Montgomery, G. W., James, M. R., Hayward, N. K., Martin, N. G. and Sturm, R. A. (2004). Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. *Hum. Mol. Genet.* **13**, 447-461. doi:10.1093/hmg/ddh043

Duffy, D. L., Montgomery, G. W., Chen, W., Zhao, Z. Z., Le, L., James, M. R., Hayward, N. K., Martin, N. G. and Sturm, R. A. (2007). A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am. J. Hum. Genet.* **80**, 241-252. doi:10.1086/510885

Edwards, A. W. F. (1996). The origin and early development of the method of minimum evolution for the reconstruction of phylogenetic trees. *Syst. Biol.* **45**, 79-91. doi:10.1093/sysbio/45.1.79

Edwards, M., Bigham, A., Tan, J., Li, S., Gozdzik, A., Ross, K., Jin, L. and Parra, E. J. (2010). Association of the *OCA2* polymorphism His615arg with melanin content in east asian populations: further evidence of convergent evolution of skin pigmentation. *PLoS Genet.* **6**, e1000867. doi:10.1371/journal.pgen.1000867

Eiberg, H., Troelsen, J., Nielsen, M., Mikkelsen, A., Mengel-From, J., Kjaer, K. W. and Hansen, L. (2008). Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum. Genet.* **123**, 177-187. doi:10.1007/s00439-007-0460-x

Fedorova, S. A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., Trofimova, N., Zhadanov, S. I., Kashani, B. H., Olivieri, A. et al. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* **13**, 127. doi:10.1186/1471-2148-13-127

Fortier, A. L., Coffman, A. J., Struck, T. J., Irby, M. N., Burguete, J. E. L., Ragsdale, A. P. and Gutenkunst, R. N. (2019). DFEnitely different: genome-wide characterization of differences in mutation fitness effects between populations. *bioRixv* doi:10.1101/703918

Graf, J., Hodgson, R. and van Daal, A. (2005). Single nucleotide polymorphisms in the *MATP* gene are associated with normal human pigmentation variation. *Hum. Mutat.* **25**, 278-284. doi:10.1002/humu.20143

Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A. and Kingsley, D. M. (2014). A molecular basis for classic blond hair color in Europeans. *Nat. Genet.* **46**, 748-752. doi:10.1038/ng.2991

Haller, B. C. and Messer, P. W. (2017). SLiM 2: flexible, interactive forward genetic simulations. *Mol. Biol. Evol.* **34**, 230-240. doi:10.1093/molbev/msw211

Han, J., Kraft, P., Nan, H., Guo, Q., Chen, C., Qureshi, A., Hankinson, S. E., Hu, F. B., Duffy, D. L., Zhao, Z. Z. et al. (2008). A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4**, e1000074. doi:10.1371/journal.pgen.1000074

Harding, R. M., Healy, E., Ray, A. J., Ellis, N. S., Flanagan, N., Todd, C., Dixon, C., Sajantila, A., Jackson, I. J., Birch-Machin, M. A. et al. (2000). Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **66**, 1351-1361. doi:10.1086/302863

Hart, K. L., Kimura, S. L., Mushailov, V., Budimlijia, Z. M., Prinz, M. and Wurmbach, E. (2013). Improved eye- and skin-color prediction based on 8 SNPs. *Croat. Med. J.* **54**, 248-256. doi:10.3325/cmj.2013.54.248

He, Y., Wang, M., Huang, X., Li, R., Xu, H., Xu, S. and Jin, L. (2015). A probabilistic method for testing and estimating selection differences between populations. *Genome Res.* **25**, 1903-1909. doi:10.1101/gr.192336.115

Hider, J. L., Gittelman, R. M., Shah, T., Edwards, M., Rosenbloom, A., Akey, J. M. and Parra, E. J. (2013). Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol. Biol.* **13**, 150. doi:10.1186/1471-2148-13-150

Hochberg, Z. and Hochberg, I. (2019). Evolutionary perspective in rickets and vitamin D. *Front. Endocrinol.* **10**, 306. doi:10.3389/fendo.2019.00306

Huang, X., Jin, L. and He, Y. (2019). SeleDiff: a fast and scalable tool for estimating and testing selection differences between populations. *J. Open Source Software* **4**, 1545. doi:10.21105/joss.01545

Ito, S. and Wakamatsu, K. (2008). Chemistry of mixed melanogenesis—pivotal roles of dopaquinone. *Photochem. Photobiol.* **84**, 582-592. doi:10.1111/j.1751-1097.2007.00238.x

Izagirre, N., García, I., Junquera, C., de la Rúa, C. and Alonso, S. (2006). A scan for signatures of positive selection in candidate loci for skin pigmentation in humans. *Mol. Biol. Evol.* **23**, 1697-1706. doi:10.1093/molbev/msl030

Jablonski, N. G. and Chaplin, G. (2000). The evolution of human skin coloration. *J. Hum. Evol.* **39**, 57-106. doi:10.1006/jhev.2000.0403

Jablonski, N. G. and Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proc. Natl. Acad. Sci. USA* **107** Suppl. 2, 8962-8968. doi:10.1073/pnas.0914628107

Jablonski, N. G. and Chaplin, G. (2014). Skin cancer was not a potent selective force in the evolution of protective pigmentation in early hominins. *Proc. R. Soc. B Biol. Sci.* **281**, 20140517. doi:10.1098/rspb.2014.0517

Jablonski, N. G. and Chaplin, G. (2017). The colours of humanity: the evolution of pigmentation in the human lineage. *Philos. Trans. R. Soc. B: Biol. Sci.* **372**, 20160349. doi:10.1098/rstb.2016.0349

Jacobs, L. C., Wollstein, A., Lao, O., Hofman, A., Klaver, C. C., Uitterlinden, A. G., Nijsten, T., Kayser, M. and Liu, F. (2013). Comprehensive candidate gene study highlights UGT1A and BNC2 as new genes determining continuous skin color variation in Europeans. *Hum. Genet.* **132**, 147-158. doi:10.1007/s00439-012-1232-9

Kayser, M., Liu, F., Janssens, A. C. J. W., Rivadeneira, F., Lao, O. van Duijn, K., Vermeulen, M., Arp, P., Jhamai, M. M., van Ijcken, W. F. J. et al. (2008). Three genome-wide association studies and a linkage analysis identify *HERC2* as a human iris color gene. *Am. J. Hum. Genet.* **82**, 411-423. doi:10.1016/j.ajhg.2007.10.003

Kovacevic, L., Tambets, K., Ilumäe, A.-M., Kushniarevich, A., Yunusbayev, B., Solnik, A., Bego, T., Primorac, D., Skaro, V., Leskovac, A. et al. (2014). Standing at the gateway to Europe - the genetic structure of Western Balkan populations based on autosomal and haploid markers. *PLoS ONE* **9**, e105090. doi:10.1371/journal.pone.0105090

Lamason, R. L., Mohideen, M.-A. P. K., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E. et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782-1786. doi:10.1126/science.1116238

Lao, O., de Gruijter, J. M., van Duijn, K., Navarro, A. and Kayser, M. (2007). Signatures of positive selection in genes associated with human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. *Ann. Hum. Genet.* **71**, 354-369. doi:10.1111/j.1469-1809.2006.00341.x

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L. et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100-1104. doi:10.1126/science.1153717

Martin, A. R., Lin, M., Granka, J. M., Myyrick, J. W., Liu, X., Sockell, A., Atkinson, E. G., Werely, C. J., Möller, M., Sandhu, M. S. et al. (2017). An unexpectedly complex architecture for skin pigmentation in Africans. *Cell* **171**, 1340-1353.e14. doi:10.1016/j.cell.2017.11.015

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M. et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503. doi:10.1038/nature16152

McEvoy, B., Beleza, S. and Shriver, M. D. (2006). The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Hum. Mol. Genet.* **15**, R176-R181. doi:10.1093/hmg/ddl217

McEvoy, B. P., Powell, J. E., Goddard, M. E. and Visscher, P. M. (2011). Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res.* **21**, 821-829. doi:10.1101/gr.119636.110

Metspalu, M., Romero, I. G., Yunusbayev, B., Chaubey, G., Mallick, C. B., Hudjashov, G., Nelis, M., Mägi, R., Metspalu, E., Remm, M. et al. (2011). Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* **89**, 731-744. doi:10.1016/j.ajhg.2011.11.010

Mezzavilla, M. and Ghirotto, S. (2015). Neon: an R package to estimate human effective population size and divergence time from patterns of linkage disequilibrium between SNPs. *J. Comput. Sci. Syst. Biol.* **8**, 37-44. doi:10.4172/jcsb.1000168

Miller, C. T., Beleza, S., Pollen, A. A., Schluter, D., Kittles, R. A., Shriver, M. D. and Kingsley, D. M. (2007). *cis*-Regulatory changes in *Kit ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**, 1179-1189. doi:10.1016/j.cell.2007.10.055

Mondal, M., Casals, F., Xu, T., Dall'Olio, G. M., Pybus, M., Netea, M. G., Comas, D., Laayouni, H., Li, Q., Majumder, P. P. et al. (2016). Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat. Genet.* **48**, 1066-1070. doi:10.1038/ng.3621

Motokawa, T., Kato, T., Hashimoto, Y. and Katagiri, T. (2007). Effect of Val92Met and Arg163Gln variants of the *MC1R* gene on freckles and solar lentigines in Japanese. *Pigment Cell Res.* **20**, 140-143. doi:10.1111/j.1600-0749.2007.00364.x

Murray, N., Norton, H. L. and Parra, E. J. (2015). Distribution of two OCA2 polymorphisms associated with pigmentation in East-Asian populations. *Hum. Genome Var.* **2**, 15058. doi:10.1038/hgv.2015.58

Myles, S., Somel, M., Tang, K., Kelso, J. and Stoneking, M. (2007). Identifying genes underlying skin pigmentation differences among human populations. *Hum. Genet.* **120**, 613-621. doi:10.1007/s00439-006-0256-4

Norton, H. L. (2005). Human Skin Pigmentation Variation: A Phenotypic, Genotypic and Evolutionary Perspective. *Ph.D. Dissertation*. Pennsylvania State University. Available from https://etda.libraries.psu.edu/catalog/6658.

Norton, H. L., Kittles, R. A., Parra, E., McKeigue, P., Mao, X., Cheng, K., Canfield, V. A., Bradley, D. G., McEvoy, B. and Shriver, M. D. (2007). Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710-722. doi:10.1093/molbev/msl203

Oppenheimer, S. (2012). Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos. Trans. R. Soc. B: Biol. Sci.* **367**, 770-784. doi:10.1098/rstb.2011.0306

Pagani, L., Kivisild, T., Tarekegn, A., Ekong, R., Plaster, C., Romero, I. G., Ayub, Q., Mehdi, S. Q., Thomas, M. G., Luiselli, D. et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91**: 83-96. doi:10.1016/j.ajhg.2012.05.015

Pagani, L., Lawson, D. J., Jagoda, E., Mörseburg, A., Eriksson, A., Mitt, M., Clemente, F., Hudjashov, G., DeGiorgio, M., Saag, L. et al. (2016). Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* **538**, 238-242. doi:10.1038/nature19792

Parra, E. J. (2007). Human pigmentation variation: evolution, genetic basis, and implications for public health. *Am. J. Phys. Anthropol.* **45** Suppl, 85-105. doi:10.1002/ajpa.20727

Patterson, N., Price, A. L. and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* **2**, e190. doi:10.1371/journal.pgen.0020190

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D. (2012). Ancient admixture in human history. *Genetics* **192**, 1065-1093. doi:10.1534/genetics.112.145037

Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., Srinivasan, B. S., Barsh, G. S., Myers, R. M., Feldman, M. W. et al. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826-837. doi:10.1101/gr.087577.108

Praetorius, C., Grill, C., Stacey, S. N., Metcalf, A. M., Gorkin, D. U., Robinson, K. C., Otterloo, E. V., Kim, R. S. Q., Bergsteinsdottir, K., Ogmundsdottir, M. H. et al. (2013). A polymorphism in IRF4 affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155**, 1022-1033. doi:10.1016/j.cell.2013.10.022

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904-909. doi:10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575. doi:10.1086/519795

Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T. W., Jr., Orlando, L., Metspalu, E. et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of native Americans. *Nature* **505**, 87-91. doi:10.1038/nature12736

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

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J. S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R. et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762. doi:10.1038/nature08835

Rebbeck, T. R., Kanetsky, R. A., Walker, A. H., Holmes, R., Halpern, A. C., Schuchter, L. M., Elder, D. E. and Guerry, D. (2002). *P* gene as an inherited biomarker of human eye color. *Cancer Epidem. Biomar.* **11**, 782-784.

Rees, J. L. and Harding, R. M. (2012). Understanding the evolution of human pigmentation: Recent contributions from population genetics. *J. Invest. Dermatol.* **132**, 846-853. doi:10.1038/jid.2011.358

Schaffner, S. F., Foo, C., Gabriel, S., Reich, D., Daly, M. J. and Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* **15**, 1576-1583. doi:10.1101/gr.3709305

Schiffel, S. and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919-925. doi:10.1038/ng.3015

Sirugo, G., Williams, S. M. and Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell* **177**, 26-31. doi:10.1016/j.cell.2019.02.048

Spichenok, O., Budimlija, Z. M., Mitchell, A. A., Jenny, A., Kovacevic, L., Marjanovic, D., Caragine, T., Prinz, M. and Wurmbach, E. (2011). Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci. Int. Genet.* **5**, 472-478. doi:10.1016/j.fsigen.2010.10.005

Stokowski, R. P., Pant, P. V. K., Dadd, T., Fereday, A., Hinds, D. A., Jarman, C., Filsell, W., Ginger, R. S., Green, M. R., van der Ouderaa, F. J. et al. (2007). A genomewide association study of skin pigmentation in a South Asian population. *Am. J. Hum. Genet.* **81**, 1119-1132. doi:10.1086/522235

Sturm, R. A. and Duffy, D. L. (2012). Human pigmentation genes under environmental selection. *Genome Biol.* **13**, 248. doi:10.1186/gb-2012-13-9-248

Sturm, R. A., Duffy, D. L., Zhao, Z. Z., Leite, F. P. N., Stark, M. S., Hayward, N. K., Martin, N. G. and Montgomery, G. W. (2008). A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82**, 424-431. doi:10.1016/j.ajhg.2007.11.005

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., Manolescu, A., Karason, A., Palsson, A., Thorleifsson, G. et al. (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443-1452. doi:10.1038/ng.2007.13

Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Jakobsdottir, M., Steinberg, S., Gudjonsson, S. A., Palsson, A., Thorleifsson, G. et al. (2008). Two newly identified genetic determinants of pigmentation in Europeans. *Nat. Genet.* **40**, 835-837. doi:10.1038/ng.160

Teo, Y.-Y., Sim, X., Ong, R. T. H., Tan, A. K. S., Chen, J., Tantoso, E., Small, K. S., Ku, C.-S., Lee, E. J. D., Seielstad, M. et al. (2009). Singapore genome variation project: a haplotype map of three Southeast Asian populations. *Genome Res.* **19**, 2154-2162. doi:10.1101/gr.095000.109

Terhorst, J., Kamm, J. A. and Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303-309. doi:10.1038/ng.3748

The 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073. doi:10.1038/nature09534

The International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58. doi:10.1038/nature09298

Visser, M., Kayser, M. and Palstra, R.-J. (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome Res.* **22**, 446-455. doi:10.1101/gr.128652.111

Visser, M., Palstra, R. and Kayser, M. (2014). Human skin color is influenced by an intergenic DNA polymorphism regulating transcription of the nearby *BNC2* pigmentation gene. *Hum. Mol. Genet.* **23**, 5750-5762. doi:10.1093/hmg/ddu289

Voight, B. F., Kudaravalli, S., Wen, X. and Pritchard, J. K. (2007). Correction: A map of recent positive selection in the human genome. *PLoS Biol.* **5**, e147. doi:10.1371/journal.pbio.0050147

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Karnysz, A., Branicki, W. and Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7**, 98-115. doi:10.1016/j.fsigen.2012.07.005

Wilde, S., Timpson, A., Kirsanow, K., Kaiser, E., Kayser, M., Unterländer, M., Hollfelder, N., Potekhina, I. D., Schier, W., Thomas, M. G. et al. (2014). Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl. Acad. Sci. USA* **111**, 4832-4837. doi:10.1073/pnas.1316513111

Wolf, S. T. and Kenney, W. L. (2019). The vitamin D-folate hypothesis in human vascular health. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **317**, R491-R501. doi:10.1152/ajpregu.00136.2019

Xing, J., Wuren, T., Simonson, T. S., Watkins, W. S., Witherspoon, D. J., Wu, W., Qin, G., Huff, C. D., Jorde, L. B. and Ge, R.-L. (2013). Genomic analysis of natural selection and phenotypic variation in high-altitude Mongolians. *PLoS Genet.* **9**, e1003634. doi:10.1371/journal.pgen.1003634

Xu, X., Thörnwall, M., Lundin, L.-G. and Chhajlani, V. (1996). Val92Met variant of the melanocyte stimulating hormone receptor gene. *Nat. Genet.* **14**, 384. doi:10.1038/ng1296-384

Yang, Z., Zhong, H., Chen, J., Zhang, X., Zhang, H., Luo, X., Xu, S., Chen, H., Lu, D., Han, Y. et al. (2016). A genetic mechanism for convergent skin lightening during recent human evolution. *Mol. Biol. Evol.* **33**, 1177-1187. doi:10.1093/molbev/msw003

Yunusbayev, B., Metspalu, M., Järve, M., Kutuev, I., Rootsi, S., Metspalu, E., Behar, D. M., Varendi, K., Sahakyan, H., Khusainova, R. et al. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Mol. Biol. Evol.* **29**, 359-365. doi:10.1093/molbev/msr221

Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S. et al. (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet.* **11**, e1005068. doi:10.1371/journal.pgen.1005068