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

Xin Huang¹*, Sijia Wang¹,², Li Jin¹,²,* and Yungang He³,*

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 around 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 (0.0065/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.

KEY WORDS: Population genetics, Natural selection, Human evolution, Human pigmentation, Complex traits

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, MYOSA – conveying vesicles and organelles within the cytoplasm (Abdel-Malek et al., 2001; Rebeck 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; Donnelly et al., 2012; Beleza 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
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 (genetic) 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

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

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

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

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

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

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

where \( p_i(L_1, L_2) \) is the frequency of the combination carrying two alleles associated with one possible outcome of a trait, such as light pigmentation; \( q_i(L_1, L_2) \) 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 \( p_i(L_1, L_2) = p_i(L_1)p_i(L_2) \) and \( q_i(L_1, L_2) = q_i(L_1)q_i(L_2) \). 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:

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

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

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

Let

\[ s_1 = \left( \frac{\ln p_1 - \ln p_r}{q_1 - q_r} \right)/t_1, \]

\[ s_2 = \left( \frac{\ln p_2 - \ln p_w}{q_2 - q_w} \right)/t_2, \]

\[ s_{10} = \left( \frac{\ln p_w - \ln p_r}{q_w - q_r} \right)/t_{10}, \]

then we have

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

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

\[ \mathbf{d} = (d_{21} t_1 \quad d_{31} t_1 \quad \ldots \quad d_{55} t_5)^{\top}, \]

\[ \mathbf{s} = (s_1 \quad s_2 \quad \ldots \quad s_{10})^{\top}. \]

\[ \mathbf{T} = \begin{bmatrix} -t_1 & t_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & t_1 - t_2 \\ -t_1 & 0 & t_3 & 0 & 0 & 0 & 0 & 0 & 0 & t_2 - t_3 \\ -t_1 & 0 & 0 & t_4 & 0 & 0 & 0 & 0 & 0 & t_3 - t_4 \\ -t_1 & 0 & 0 & 0 & t_5 & 0 & 0 & 0 & 0 & t_4 - t_5 \\ -t_1 & 0 & 0 & 0 & 0 & t_6 & 0 & 0 & 0 & t_5 - t_6 \\ -1 & 0 & 0 & 0 & 0 & 0 & t_7 & 0 & 0 & t_6 - t_7 \\ 0 & -t_2 & t_3 & 0 & 0 & 0 & 0 & t_7 - t_8 & 0 & 0 \\ 0 & -t_2 & 0 & t_4 & 0 & 0 & 0 & t_7 - t_8 & 0 & 0 \\ 0 & 0 & -t_3 & 0 & t_5 & 0 & 0 & t_7 - t_8 & 0 & 0 \\ 0 & 0 & 0 & -t_4 & 0 & t_6 & 0 & t_7 - t_8 & 0 & 0 \\ 0 & 0 & 0 & 0 & -t_5 & 0 & t_7 - t_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -t_6 & t_7 - t_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & t_7 - t_8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & t_7 - t_8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & t_7 - t_8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & t_7 - t_8 \end{bmatrix} \]

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

\[ \mathbf{d} = \mathbf{T} \mathbf{s}. \]

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 \( \sigma^2 \). 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 \( \text{Pr}(\sum_{i=1}^{10} s_i^2 \geq \sum_{i=1}^{10} |s_i|^2 | \text{NE}) > \text{Pr}(\sum_{i=1}^{10} s_i^2 \geq \sum_{i=1}^{10} |s_i|^2 | \text{NE}) \), if \( |s_i|^2 \leq |s_{10}|^2 \) for solutions \( a \) and \( b \). Here, \( |s_i|^2 \) is the norm of a vector \( s \). 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:

\[ \mathbf{s} = \mathbf{T}^\top \mathbf{d}, \]

where \( \mathbf{T}^\top \) is the Moore-Penrose inverse of \( \mathbf{T} \).
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., 2013; Guenther et al., 2014; Murray et al., 2015; Yang et al., 2016; Hart et al., 2008; Eiberg et al., 2008; Branicki et al., 2009; Edwards et al., 2010; Han et al., 2008; Kayser et al., 2008; Sturm et al., 2008; Sulem et al., 2005; Stokowki et al., 2007; Miller 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; 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 results also reveal the selective pressure on light pigmentation 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 ($s_3=0.00650$/generation). This shared selection is also supported by other studies, revealing that ASIP, BNC2, and...
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 diversities 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 (DDBI 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 (0.0057744 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 (DDBI) 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 F3 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 10 SNPs and an r² 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 F3 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 (Nₑ=10,000), and the other parameters were set to the default values. Ten SNPs (rs110400, rs11547464, rs12203592, rs1800407, rs1805005, rs1805006, rs1805007, rs1805008, rs1805009, rs7465330) 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 (r²<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, rs11320064, rs12350739, rs6058017, rs12821256, rs1393350, rs1426654, rs642744, rs6510760, rs1129038, rs2228479, rs35264875, rs12869399, rs26722, rs16891982, rs885479, rs2378249, rs1042602, rs916977, rs1805005, rs1805006, rs1805007, rs1805008, rs1805009, rs7465330) 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.
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 coefficients into that of alleles as follows. Let the frequency of the ancestral allele A be 1, and the relative fitness of the derived allele a is e^r. When r is close to 0, we can approximate e^r as 1+r as using the Taylor series. The fitness of genotype aa becomes (1+r)^2=1+2r+r^2≈1+2s, and the fitness of genotype Aa is 1+r=1+0.5s. If s is the selection coefficient in SLiM 2, then s=2r; 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
\[ \tilde{d}_{ij} = \left( \frac{\ln p_i - \ln p_j}{q_i} \right) / t \]

is affected by population migration and substructure in theory. Here, \( \tilde{p}_i \) and \( \tilde{q}_i \) are the observed derived and ancestral allele frequencies in the population \( j \), respectively; \( \tilde{p}_j \) and \( \tilde{q}_j \) are the observed derived and ancestral allele frequencies in the population \( i \), respectively; and \( t \) is the divergence time from populations \( i \) and \( j \) to their most recent common ancestor. We demonstrate that \( \tilde{d}_{ij} \) 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 \( a>b, c>d \), then \( ac>bd \) and \( a+b/c+b < a/b \).

Proof: Let \( a>b, c>d \), then \( ac>bd \). Therefore, \( ac+b>bd+ab \), which is the same as \( a(b+c)>a(b+d) \).

Therefore, \( a+b/c+b < a/b \).

Inequality 2: If \( a_1>0, a_2>0, \ldots, a_n>0, \widetilde{a} = \frac{\sum_{i=1}^{n} a_i}{n} \), then \( \max(a_1, a_2, \ldots, a_n) \geq \frac{n}{\sum_{i=1}^{n} a_i/n} \geq a \).

Proof: Let \( a_{\max}=\max(a_1, a_2, \ldots, a_n) \), then
\[ a_{\max} - a = \frac{\sum_{i=1}^{n} a_{\max} - a_i}{n} \geq 0, \]
because \( a_{\max} \geq a \).

Let \( a_{\min}=\min(a_1, a_2, \ldots, a_n) \), then
\[ a - a_{\min} = \frac{\sum_{i=1}^{n} a_i - a_{\min}}{n} \leq 0, \]
because \( a_{\min} \leq a \).

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

The effect of migration

Suppose there is \( \alpha \) proportion of individuals in the population \( i \), which actually come from the population \( j \); also, there is \( \beta \) proportion of individuals in the population \( j \), which actually come from the population \( i \). Here, \( \alpha, \beta \in [0, 1/2] \), because we assume migrants should not become the majority of another population. Then we have
\[
\tilde{d}_{ij} = \frac{(1-\alpha)p_i + \alpha p_j}{q_j} - \frac{(1-\beta)p_j + \beta p_i}{q_i}.
\]

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, \( \tilde{d}_{ij} = \frac{(1-\alpha)p_i + \alpha p_j}{q_j} - \frac{(1-\beta)p_j + \beta p_i}{q_i} \).

The effect of substructure

Scenario 1: The population \( j \) has \( k \) subpopulations.

If the population \( j \) has \( k \) subpopulations, then \( \tilde{p}_j = \frac{\sum_{h=1}^{k} N_{jh} p_{jh}}{N_j} \). Here, \( p_{jh} \) is the derived allele frequency in the subpopulation \( h \) of the population \( j \). And \( N_j \) is the population size of the population \( j \). \( N_{jh} \) is the size of the population \( j \). We denote the minimum of \( p_{jh} \) as \( \min(p_{jh}) \). Because \( q_{jh}=1-p_{jh} \), \( \max(q_{jh})=1-\min(p_{jh}) \).

Based on Inequality 2, \( \tilde{d}_{ij} = \frac{\sum_{h=1}^{k} N_{jh} p_{jh}}{N_j} \).

Therefore, \( \tilde{d}_{ij} > \min(p_{jh}) \).

Scenario 2: The population \( j \) has \( l \) subpopulations.

If the population \( i \) has \( l \) subpopulations, then \( \tilde{p}_j = \frac{\sum_{h=1}^{l} N_{ih} p_{ih}}{N_j} \). We denote the maximum of \( p_{ih} \) as \( \max(p_{ih}) \). Then \( \tilde{d}_{ij} = \frac{\sum_{h=1}^{l} N_{ih} p_{ih}}{N_j} \).

Therefore, \( \tilde{d}_{ij} > \max(p_{ih}) \).

Downloaded from http://bio.biologists.org/ by guest on March 21, 2021
Scenario 3: The population $i$ has $l$ subpopulations, and the population $j$ has $k$ subpopulations.

Based on scenarios 1 and 2, we have
\[
\ln \frac{\max(p_i)}{\min(p_j)} > \ln \frac{\min(p_i)}{\max(p_j)} > 0.
\]

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 is the largest difference between subpopulations.

Acknowledgements

X.H. thanks Dr Minxian Wang, Dr Yuchen Wang, Dr Haiyi Lou and Dr Lin Tang for comments on the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: X.H., S.W., Y.H.; Methodology: X.H., Y.H.; Software: X.H.; Validation: X.H.; Formal analysis: X.H.; Investigation: X.H., Y.H.; Visualization: L.J., Y.H.; Project administration: X.H.; Funding acquisition: L.J., Y.H.

Funding

This work was supported by grants from National Natural Science Foundation of China [31871255 and 91331108 to S.W.]; 31322030 and 91331109 to Y.H.; National Thousand Young Talents Award, the Max Planck-CAS Paul Gerson Unna Independent Project Grant [2017SHZDZX01]. S.W. was also awarded by the National Thousand Talents Program Leadership Award, and open projects from the State Key Research Group Leadership Award, and open projects from the State Key Laboratory of Genetic Engineering at Fudan University.

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), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/EIG), bwh.harvard.edu/plink/), EIGENSOFT 6.0.1 (https://github.com/DReichLab/
Human skin pigmentation as revealed from analyses of single nucleotide polymorphisms. Signs of positive selection in genes associated with human skin pigmentation candidate genes in populations of East Asian ancestry. Exploring signatures of positive selection in candidate loci for skin pigmentation in Andamanese provides insights into ancient human migration into Asia and adaptation. Improved eye- and skin-color prediction based on 8 SNPs. The chemistry of mixed melanogenesis—pivotal roles of dopaquinone. The evolution of the human iris color gene. A scan for selection in candidate loci for skin pigmentation in humans. Mol. Biol. Evol. 23, 1697-1706. doi:10.1093/molbev/ms030 Comprehensive candidate gene study highlights UGTA1 and BNC2 as new genes determining continuous skin color variation in Europeans. Human Skin Pigmentation Variation: A Phenotypic, Genotypic and Evolutionary Perspective. Pigment Cell Res. 15, 15058. doi:10.1038/hgv.2015.58 Data on skin color: A comparison of data from the United States and Europe. Human Skin Pigmentation Variation: A Phenotypic, Genotypic and Evolutionary Perspective. Pigment Cell Res. 11, 238-242. doi:10.1016/j.ptcr.2005.11.017 A method for testing and estimating selection differences between populations. J. Hum. Evol. 43, 107-127. doi:10.1016/S1095-8312(02)00062-3 Current skin pigment variation in a large number of populations. Mol. Biol. Evol. 19, 1643-1650. doi:10.1093/mbie/19.6.1643 Multivariate analysis of global human genetic data. Science 314, 1113-1117. doi:10.1126/science.1134923 Human pigmentation: a multivariate approach to the study of a complex trait. Proc. Natl. Acad. Sci. USA 102, 15084-15089. doi:10.1073/pnas.0507558102 A genetic model for human skin coloration. Proc. Natl. Acad. Sci. USA 100, 15543-15548. doi:10.1073/pnas.2327208100 Evolutionary perspective in ricketts and implications for public health. Am. J. Phys. Anthropol. 45, 805-813. doi:10.1002/ajpa.20277 Improved eye- and skin-color prediction based on 8 SNPs. J. Invest. Dermatol. 132, 1547-1557. doi:10.1038/jid.2011.358 Evolutionary perspective in ricketts and adaptation. Mol. Biol. Evol. 29, 2386-2393. doi:10.1093/molbev/msr020 Evolutionary perspective in ricketts and implications for public health. Am. J. Phys. Anthropol. 45, 805-813. doi:10.1002/ajpa.20277 Improved eye- and skin-color prediction based on 8 SNPs. J. Invest. Dermatol. 132, 1547-1557. doi:10.1038/jid.2011.358
The 1000 Genomes Project Consortium

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, 1015-1025. 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 Wurbach, 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., Jarmans, 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.180

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, 2151-2162. doi:10.1101/gr.095341.109

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., Kaysers, 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 Kaysers, 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., Wolstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branci, W. and Kaysers, M. (2013). The HtriAplex 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., Kaysers, M., Unterländer, M., Hafffelder, N., Potekhina, I. D., Schier, W., Thomas, M. G. et al. (2014). Direct evidence for positive selection of skin, hair, and eye pigment 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., Varendri, K., Sahakyan, H., Khusainova, R. et al. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 299, R491-R501. doi:10.1152/ajpregu.00136.2019

Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litiñov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovska, 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