



ORIGINAL RESEARCH

Ambient Temperature is A Strong Selective Factor Influencing Human Development and Immunity



Lindan Ji^{1,2}, Dongdong Wu¹, Haibing Xie¹, Binbin Yao³, Yanming Chen³
David M. Irwin^{1,4,5}, Dan Huang^{6,7}, Jin Xu^{3,*}, Nelson L.S. Tang^{6,7,*},
Yaping Zhang^{1,8,*}

¹ State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

² Department of Biochemistry, Medical School of Ningbo University, Ningbo 315211, China

³ Department of Preventive Medicine, Medical School of Ningbo University, Ningbo 315211, China

⁴ Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5S 1A8, Canada

⁵ Banting and Best Diabetes Centre, University of Toronto, Toronto, ON M5S 1A8, Canada

⁶ Department of Chemical Pathology, and Laboratory for Genetics of Disease Susceptibility, Li Ka Shing Institute of Health Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China

⁷ KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming 650223, China

⁸ Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China

Received 5 January 2019; revised 22 August 2019; accepted 29 November 2019

Available online 19 August 2020

Handled by James Cai

KEYWORDS

Solar radiation;
Ambient temperature;
Natural selection;
Development;
Immunity

Abstract Solar energy, which is essential for the origin and evolution of all life forms on Earth, can be objectively recorded through attributes such as climatic **ambient temperature** (CAT), ultraviolet radiation (UVR), and sunlight duration (SD). These attributes have specific geographical variations and may cause different adaptation traits. However, the adaptation profile of each attribute and the selective role of solar energy as a whole during human evolution remain elusive. Here, we performed a genome-wide adaptation study with respect to CAT, UVR, and SD using the Human Genome Diversity Project-Centre Etude Polymorphisme Humain (HGDP-CEPH) panel data. We singled out CAT as the most important driving force with the highest number of adaptive loci (6 SNPs at the genome-wide 1×10^{-7} level; 401 at the suggestive 1×10^{-5} level). Five of the six genome-wide significant adaptation SNPs were successfully replicated in an independent Chinese population ($N = 1395$). The corresponding 316 CAT adaptation genes were mostly involved in **development** and **immunity**. In addition, 265 (84%) genes were related to at least one genome-wide association

* Corresponding authors.

E-mail: xujin1@nbu.edu.cn (Xu J), nelsontang@cuhk.edu.hk (Tang NLS), zhangyp@mail.kiz.ac.cn (Zhang Y).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2019.11.009>

1672-0229 © 2020 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

study (GWAS)-mapped human trait, being significantly enriched in anthropometric loci such as those associated with body mass index (χ^2 ; $P < 0.005$), immunity, metabolic syndrome, and cancer (χ^2 ; $P < 0.05$). For these adaptive SNPs, balancing selection was evident in Euro-Asians, whereas obvious positive and/or purifying selection was observed in Africans. Taken together, our study indicates that CAT is the most important attribute of solar energy that has driven genetic adaptation in development and immunity among global human populations. It also supports the non-neutral hypothesis for the origin of disease-predisposition alleles in common diseases.

Introduction

Solar radiation is an essential environmental factor in almost all ecosystems, and it is fundamental to the survival of the majority of life forms, including human beings [1]. Three attributes of solar radiation, including light intensity (ultraviolet radiation, UVR), climatic ambient temperature (CAT), and photoperiod (day/night cycle indicated by sunshine duration, SD), have been identified as important components of geographic variation in solar radiation exposure [2]. Together, these three attributes have shaped and characterized distinct habitats in which specific adaptive mechanisms could occur.

Physiological studies have been intensively carried out to investigate the morphological adaptation to solar radiation in humans. Sunlight intensity is the most commonly researched, and it is found to correlate well with the worldwide variation in human skin pigmentation [3]. Proposed driving mechanisms include photosynthesis of vitamin D [4] associated with ultraviolet B (UVB), photolysis of folate [5] with UVA, or a combination of the two. In addition, overexposure to sunlight, particularly UVB, is a recently well-recognized etiologic agent for skin cancer, which suggests a candidate protective mechanism of melanin [6]. CAT has a close association with the morphology of the human body, including body size, limb proportion, and pelvis form [7,8], consistent with Bergmann's and Allen's rules [9,10]. That is, in wide-ranging homeotherms, larger body size and shorter appendages will be found in the colder parts of the range, and vice versa. Furthermore, both elevated basal metabolic rates in arctic populations as a local adaptation to cold stress [11] and short stature in African and American tropical forest residents due to inefficient thermoregulation [12] suggest a thermal adaptation mechanism in response to CAT. By contrast, daily fluctuations and seasonal changes in these three attributes, especially in the photoperiod as indicated by SD, help entrain biological rhythms, which are of fundamental importance for human immunity and physiology [13–15]. Circadian misalignment involved in shift work has been implicated to increase the risk of metabolic syndrome, immune disorders, and many other diseases [16]. Therefore, a biological rhythm adaptation in response to SD is also strongly suggested as a human adaptation mechanism to solar radiation.

Also, genetic studies have been increasingly used to help elucidate the mechanisms of human physiological adaptation to solar radiation. To date, genes including *ASIP*, *ATRN*, *DCT*, *HERC2*, *IRF4*, *MC1R*, *OCA2*, *SLC24A5*, *SLC45A2*, *TYR*, and *TYRP1* have been identified to correlate with skin color [17]. However, these genes were revealed mainly through candidate genetic divergence studies between dark- and light-skinned populations. Recently, identification of sun-exposure-dependent gene expression quantitative trait loci (eQTLs) in *SLC45A2* confirmed human adaptation to local

variant levels of sunlight [18]. To cope with the harsh tropical forest environment, Africans and Amerindians may follow different routes through convergent genome-wide signals related to lipid metabolism, the immune system, body development, and RNA polymerase III transcription initiation [12]. By contrast, local adaptation to cold stress and a high marine-derived fat diet in a Greenlandic Inuit population led to the selection of the *FADS* region, which encodes fatty acid desaturases [19]. Interestingly, the cold-responsive *FADS* region is closely associated with human body height and weight [19]. However, only a few genes related to biological rhythms have been identified under selection. Among them, genetic polymorphisms of the *PER2* gene demonstrate a significant geographic distribution, which is suggestive of local directional selection [20]. Our previous work examined the melatonin pathway and found that the *MTNR1B* gene encoding melatonin receptor 2 has undergone the natural selection of population residential SD [15].

Although several aforementioned adaptive mechanisms have been suggested for single attribute—photosynthesis of vitamin D and/or photolysis of folate together with recently suggested protective mechanism of melanin for UVR; a thermal adaptation mechanism in response to CAT; and a biological rhythm adaptation in response to SD—none of the present studies has investigated the three attributes of solar radiation simultaneously. Thus far, the current physiological and genetic data are still insufficient to elucidate the overall selective role of solar radiation as an integrative environmental niche. In particular, it is unknown (1) which attribute of solar radiation played the dominant role, (2) which category of genes adapted most to these environmental attributes, and (3) what the underlying selection mechanism was.

To answer these questions, we conducted a series of analyses using an extensive genetic dataset generated from 40 sub-populations in the worldwide Human Genome Diversity Project-Centre Etude Polymorphisme Humain (HGDP-CEPH) panel ($N = 764$) [21]. First, through a series of genome-wide adaptation studies (GWAdS), we investigated which physical attribute of solar radiation, *i.e.*, CAT, SD, or UVR, has the strongest selective effect on human genome diversity. Genome-wide significant GWAdS SNPs were replicated by genotyping a cohort of Chinese individuals sampled from broad geographic ranges and by *in-silico* online analysis of dbCLINE database [22]. Next, we further characterized categories of genes showing adaptation to these attributes through Gene Ontology (GO) annotation [23] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [24]. Additionally, putative GWAdS SNPs were mapped against the GWAS Catalog [25] to explore potential association with known traits. Their effects on gene regulation and transcriptional regulation were retrieved from Genotype-Tissue Expression (GTEx) and HaploReg [26,27]. Finally, we

examined the underlying selection mechanisms through analyses of derived allele frequencies (DAFs) [28], Wright's fixation index (F_{ST}) [29], and integrated haplotype scores (iHSs) [30]. Collectively, the results of this study suggest that solar radiation may have a selective role in human development and immunity genes mainly through CAT.

Results

Results of the GWAdS

For the 516,663 common SNPs, after Pearson bivariate correlation and multiple linear regression analysis, CAT had a greater number of significantly associated SNPs (six SNPs at the genome-wide 1×10^{-7} level; a total of 401 SNPs at the suggestive 1×10^{-5} level) than did UVR (two genome-wide SNPs; 233 suggestive SNPs) and SD (one suggestive SNP) (Figure 1). The six genome-wide CAT-associated SNPs were in or near five genes, UL16 binding protein 3 (*ULBP3*), keratin 31 (*KRT31*), long intergenic non-protein coding RNA 112 (*LINC00112*), dopamine receptor D3 (*DRD3*), and netrin G1 (*NTNG1*), with two hits at *ULBP3*. The two significant UVR SNPs were linked to NAD kinase (*NADK*) and insulin-like growth factor binding protein 3 (*IGFBP3*). All these SNPs were still significant after false discovery rate (FDR) corrections (modified positive false discovery rate, pFDR; $P < 0.01$, Tables S1–S3).

Replication of the genome-wide adaptation signals in Chinese populations

After using a haplotype tagging technique and a minor allele frequency (MAF) filter, five genome-wide significant CAT-associated SNPs and one UVR SNP were experimentally validated in 1395 Chinese individuals. Neither significant deviation from Hardy-Weinberg equilibrium (HWE) nor gender or age difference in allele frequencies was observed ($P > 0.05$).

Among the five replicated CAT-associated SNPs, the rs13729C allele frequency in *ULBP3* was negatively associated with the annual average CAT ($r = -0.555$, $P = 0.032$), whereas rs1444041T in *NTNG1* (representing rs11185115A, $r^2 = 0.954$) and rs9825563A in *DRD3* had marginally negative associations (Figure S1A; Table S4). They did show consistent trends as observed in the HGDP-CEPH populations (Figure S1B; Table S5); that is, all three DAFs increase with declining annual average CAT. When additional temperature indexes were examined, all five SNPs, except rs1107877 in *KRT31*, were all significantly associated with annual average CAT range (Pearson bivariate correlation; $P < 0.05$). For these four SNPs, rs12626864 in *LINC00112* and rs13729 in *ULBP3* were associated with an annual average maximum CAT and annual average minimum CAT, respectively. In addition, except rs12626864, the remaining three SNPs were associated with annual average extremely minimum CAT (Table 1). However, no SNP was significantly associated with

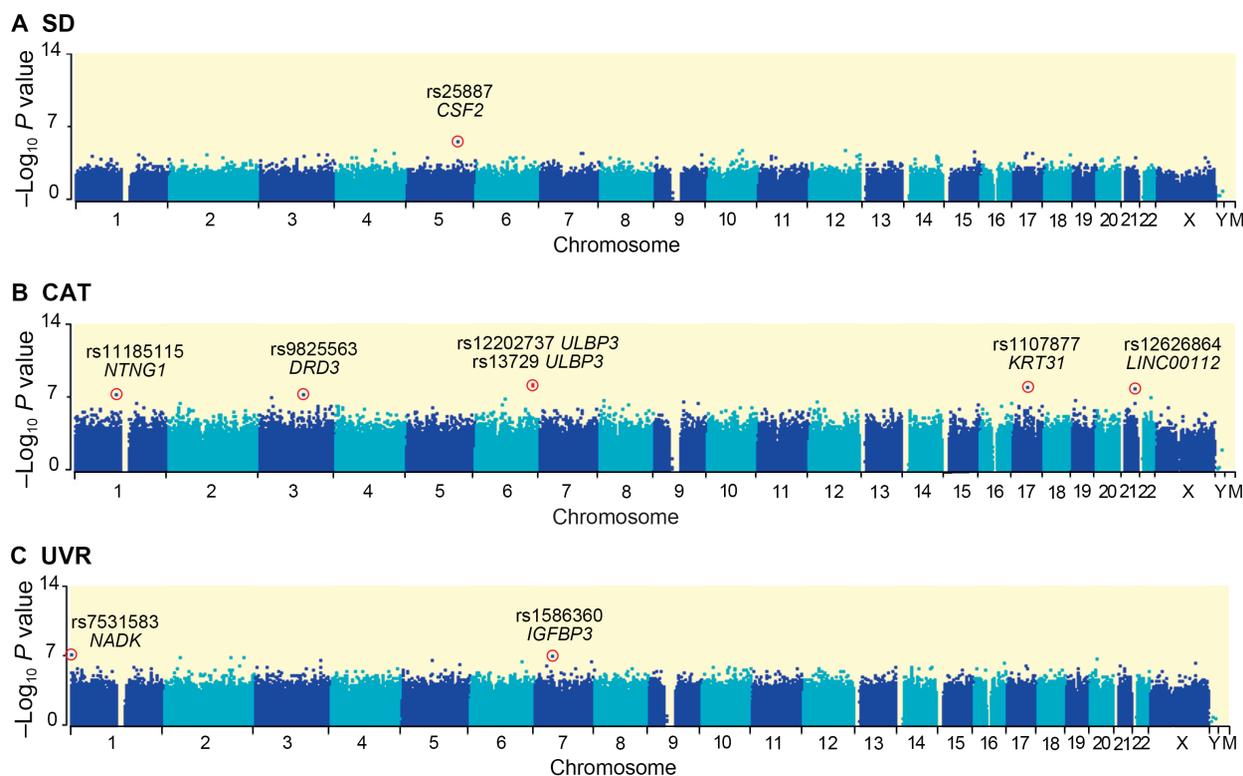


Figure 1 Manhattan plots showing the significance of association of all 516,663 SNPs with CAT, SD, and UVR

SNPs are plotted on the x-axis according to their position on each chromosome; associations with the three environmental factors are indicated on the y-axis (as $-\log_{10} P$ value). Six CAT-associated SNPs ($P \leq 1 \times 10^{-7}$), one SD-associated SNP ($P \leq 1 \times 10^{-5}$) and two UVR-associated SNPs ($P \leq 1 \times 10^{-7}$) are indicated with circles. CAT, climatic ambient temperature; SD, sunlight duration; UVR, ultraviolet radiation.

Table 1 Replication study for five CAT-associated SNPs in Chinese populations

Original SNP	Tag SNP	Gene	\bar{T}	Max \bar{T}	Min \bar{T}	\bar{T} range	Extreme min \bar{T}	Extreme max \bar{T}
rs13729	rs13729	<i>ULBP3</i>	√		√	√	√	
rs12202737								
rs1107877	rs1107877	<i>KRT31</i>						
rs12626864	rs12626864	<i>LINC00112</i>		√		√		
rs9825563	rs9825563	<i>DRD3</i>				√	√	
rs11185115 [†]	rs1444041	<i>NTNG1</i>				√	√	

Note: Significant associations between SNPs and CAT variables are indicated using √ (Pearson bivariate correlation; $P < 0.05$). rs12202737 is in complete linkage disequilibrium with rs13729 ($r^2 = 1$) in HapMap CHB population. rs11185115 within *NTNG1* is tagged by SNP rs1444041 ($r^2 = 0.954$) due to lack of appropriate SNPstream primers. \bar{T} means annual average CAT. CAT, climatic ambient temperature (°C).

annual average extremely maximum CAT. On the other hand, the six genome-wide CAT SNPs, except rs9825563 in *DRD3* being associated with intensive agriculture, were confirmed to be all significantly associated with minimum temperature (winter) in dbCLINE database (Table 2), which examined 61 worldwide populations including the 52 HGDP-CEPH populations using a different transformed rank statistic method [22].

For UVR-associated rs7531583 in *NADK*, the G allele frequency was associated with summer UVR dose (June, July, and August, $P < 0.05$) instead of annual average total UVR dose. However, the two UVR SNPs failed to be replicated in the dbCLINE. Only rs7531583 was found to be associated with absolute latitude (transformed rank statistic 0.002).

Potential functional implications of the genome-wide significant signals

In-silico analysis was used to annotate these SNPs in the GTEx project [26] and HaploReg database [27] for known tissue eQTLs and functional noncoding genomic regions, respectively (Tables S6 and S7). Five out of six CAT (except rs11185115 of *NTNG1*) and both of the UVR-associated signals might have functional regulation on transcription.

Five out of the six CAT signals (except rs11185115 from *NTNG1* gene) are predicted to affect known transcription factor regulatory motifs: rs12202737 and rs13729 are putative *cis*-eQTLs for *ULBP3* and *RAET1L*, while rs1107877 of *KRT31* acts as a *cis*-eQTL for *KRT32* and *KRTAP3-2*. For the two UVR signals, rs7531583 of *NADK* could *cis*-regulate itself and five nearby genes including *CALML6*, *CDK11A*, *MMP23A*, *RPI-140A9.1*, and *TMEM52*; rs1586360 from *IGFBP3* could regulate *IGFBP1*. In addition, they both were predicted to affect transcription factor motifs for a given proximal gene.

Categories of genes with significant solar radiation signals

Although only 8 SNPs reached genome-wide significance, 629 SNPs are putatively adaptive ($P \leq 1 \times 10^{-5}$). At this level, 316 CAT-associated genes, 191 UVR-associated genes, and a single SD-associated gene were identified. Among them, the SD-associated *CSF2* was also targeted by UVR, while 23 other UVR-associated genes, including the genome-wide significant *IGFBP3*, were also targeted by CAT (Figure S2). Overall, a total of 484 non-redundant genes were mapped to GO annotations and KEGG pathways. In comparison with the null distribution, no significant KEGG pathway was observed, but GO

Table 2 Replication study for six CAT-associated SNPs in dbCLINE

Original SNP	Gene	Environmental variable	Bayes factor	Transformed rank statistic
rs13729	<i>ULBP3</i>	Min \bar{T} (winter)	6.19	0.045
rs12202737	<i>ULBP3</i>	Min \bar{T} (winter)	6.25	0.040
rs1107877	<i>KRT31</i>	Min \bar{T} (winter)	11.40	0.016
rs12626864	<i>LINC00112</i>	Min \bar{T} (winter)	4.46	0.033
rs9825563	<i>DRD3</i>	Intensive agriculture	2.16	0.025
		Polar	4.58	0.032
rs11185115	<i>NTNG1</i>	Min \bar{T} (winter)	5.63	0.043

Table 3 GO annotations for CAT-associated genes and significant categories

Category	Bonferroni <i>P</i> value	Percentage of related genes (%)	Adjusted fold enrichment	χ^2 <i>P</i> value
Biological process				
GO:0048869 cellular developmental process	4.49×10 ⁻⁵	17.30	1.82	5.21×10 ⁻⁶
GO:0048856 anatomical structure development	5.19×10 ⁻⁵	16.63	1.56	2.55×10 ⁻⁹
GO:0030154 cell differentiation	7.09×10 ⁻⁵	22.25	1.83	2.72×10 ⁻⁵
GO:0048731 system development	2.87×10 ⁻⁴	20.67	1.54	3.36×10 ⁻¹¹
GO:0009653 anatomical structure morphogenesis	1.35×10 ⁻³	11.91	1.87	8.83×10 ⁻⁷
GO:0007399 nervous system development	2.14×10 ⁻³	25.39	1.94	9.06×10 ⁻¹²
GO:0032502 developmental process	2.55×10 ⁻³	23.37	1.40	7.12×10 ⁻⁴
GO:0045595 regulation of cell differentiation	6.54×10 ⁻³	6.97	2.58	8.99×10 ⁻³¹
GO:0007275 multicellular organismal development	8.31×10 ⁻³	8.09	1.38	1.57×10 ⁻⁷
GO:0048513 organ development	0.03	11.91	1.52	1.27×10 ⁻⁸
GO:0050793 regulation of developmental process	0.05	8.09	2.14	6.02×10 ⁻²³
Cell Component				
GO:0016020 membrane	0.03	46.52	1.24	1.75×10 ⁻³
GO:0044425 membrane part	0.04	42.25	1.26	3.34×10 ⁻⁵
INTERPRO				
IPR013098:Immunoglobulin I-set	6.15×10 ⁻³	4.00	5.03	1.71×10 ⁻¹⁰
SP_PIR_KEYWORDS				
membrane	3.96×10 ⁻³	43.33	1.22	1.02×10 ⁻⁶
UP_SEQ_FEATURE				
domain:Ig-like C2-type 1	0.04	4.33	4.19	2.26×10 ⁻⁹
domain:Ig-like C2-type 2	0.04	4.33	4.17	2.26×10 ⁻⁹

Note: Bonferroni *P* values are calculated from the modified one-tail Fisher Exact Test, supplied by the DAVID online analysis. Percentage of related genes is the percentage of each category genes found among all CAT-associated genes. Adjusted fold enrichment derives from the ratio of the percentages for each category genes found among CAT-associated genes and all 17,169 genes from the Illumina 650Y platform. Similarly, χ^2 *P* values are calculated according to the numbers of related genes for CAT-associated genes and for the 17,169 genes. Here, number and percentage of each GO category for all 17,169 Illumina 650Y platform genes work as null distributions.

terms involving developmental processes and immunity were significantly enriched (χ^2 ; $P < 5 \times 10^{-4}$, Table S8).

When examined separately, only the 316 CAT-associated genes showed a significant clustering with developmental processes and immunity after genomic control (Table 3). More importantly, the adjusted fold-enrichment increased significantly compared with the data obtained using the full 484 genes, with a newly emerging GO category: 048513 organ development. Specially, the fold enrichment is significantly higher (χ^2 ; $P < 5 \times 10^{-9}$) in categories for immunoglobulin I-set domain (5.03), immunoglobulin C2-set domain (4.33), regulation of cell differentiation (2.58), and regulation of developmental process (2.14). To this end, we focused on the CAT-associated signals with respect to potential GWAS traits and evolutionary analyses to examine the underlying selective mechanisms.

GWAS mapped traits or human diseases

Owing to the various linkage disequilibrium patterns revealed by different sets of tagging SNPs among human populations,

we only examined potential phenotype effect for the 316 CAT-associated genes in the GWAS Catalog. Among them, 265 genes were associated with a GWAS trait or human disease. To be more specific, 102 genes were associated with human anthropometric traits including height (21), weight (5), obesity (46), waist-to-hip ratio (8), fat (7), adipose (11), and body mass index (BMI, 32), showing significant enrichment compared with the two null distributions of all GWAS genes and HGDP-CEPH genes (χ^2 ; $P < 0.001$). When examined separately, waist-hip ratio and obesity (early onset extreme) showed the most significant enrichment (χ^2 ; $P < 0.005$, Table S9).

Interestingly, other GWAS traits including immunity, cancer, mental function, and mental disorders also showed significant enrichment (χ^2 ; $P < 0.05$, Table S9). Here, immunity associated traits mainly include acne (severe), parasitemia in *Trypanosoma cruzi* seropositivity, inflammatory biomarkers, allergic rhinitis, immune response to smallpox (secreted IFN-alpha), diisocyanate-induced asthma, inflammatory skin, and bowel diseases. For cancers, the main types involved are prostate cancer and colorectal cancer. For mental disorders, the

main related types are referred to eating disorders (purging via substances), alcohol dependence, bipolar disorder and schizophrenia, major depressive disorder, schizophrenia, *etc.*

Evolutionary analyses of the CAT-associated SNPs

Among the 401 suggestive SNPs, 49 (46 ancestral alleles) were completely fixed in all four examined African populations, with additional SNPs partially fixed in at least one African population. All these SNPs were polymorphic in the Asian, European, and Middle Eastern populations, with the exception of Siberians (Table S10).

The DAF spectrum of the CAT-associated SNPs in four continental populations, Asian, European, Middle Eastern, and Sub-Saharan African populations merged from HGDP-CEPH populations (Table S11), was first analyzed. In Asians, the DAF distribution skewed toward intermediate-frequency derived alleles ($2pq > 0.40$, $P < 0.01$), with a significant scarcity of rare derived alleles ($DAF \leq 0.10$, $P < 5 \times 10^{-8}$) and

high-frequency derived alleles ($DAF > 0.90$, $P < 0.01$) relative to control SNPs (Figure 2). By contrast, the DAF distribution in the African populations, especially for the three tropical populations (Kenya, Nigeria, and Senegal), skewed toward rare derived alleles ($P < 1 \times 10^{-16}$, Figure 2). For Europeans, a non-significant excess of intermediate-frequency derived alleles was also observed while no specific DAF distribution pattern was observed in Middle Eastern populations (Figure S3). In the HapMap populations, we observed similar trends as the HGDP-CEPH populations for Asians (ASN, comprised of CHB and JPT) and Africans (YRI), with a weaker statistical significance ($P < 0.01$, Figure S4). On the contrary, Caucasians (CEU) as Europeans showed a statistical significance ($P < 0.01$). Taken together, DAF distribution patterns propose balancing selection as the major evolutionary driver for both Asians and Europeans, and in Africans if considered the large proportion of fixed SNPs positive (complete sweeps) or purifying selection was suggested.

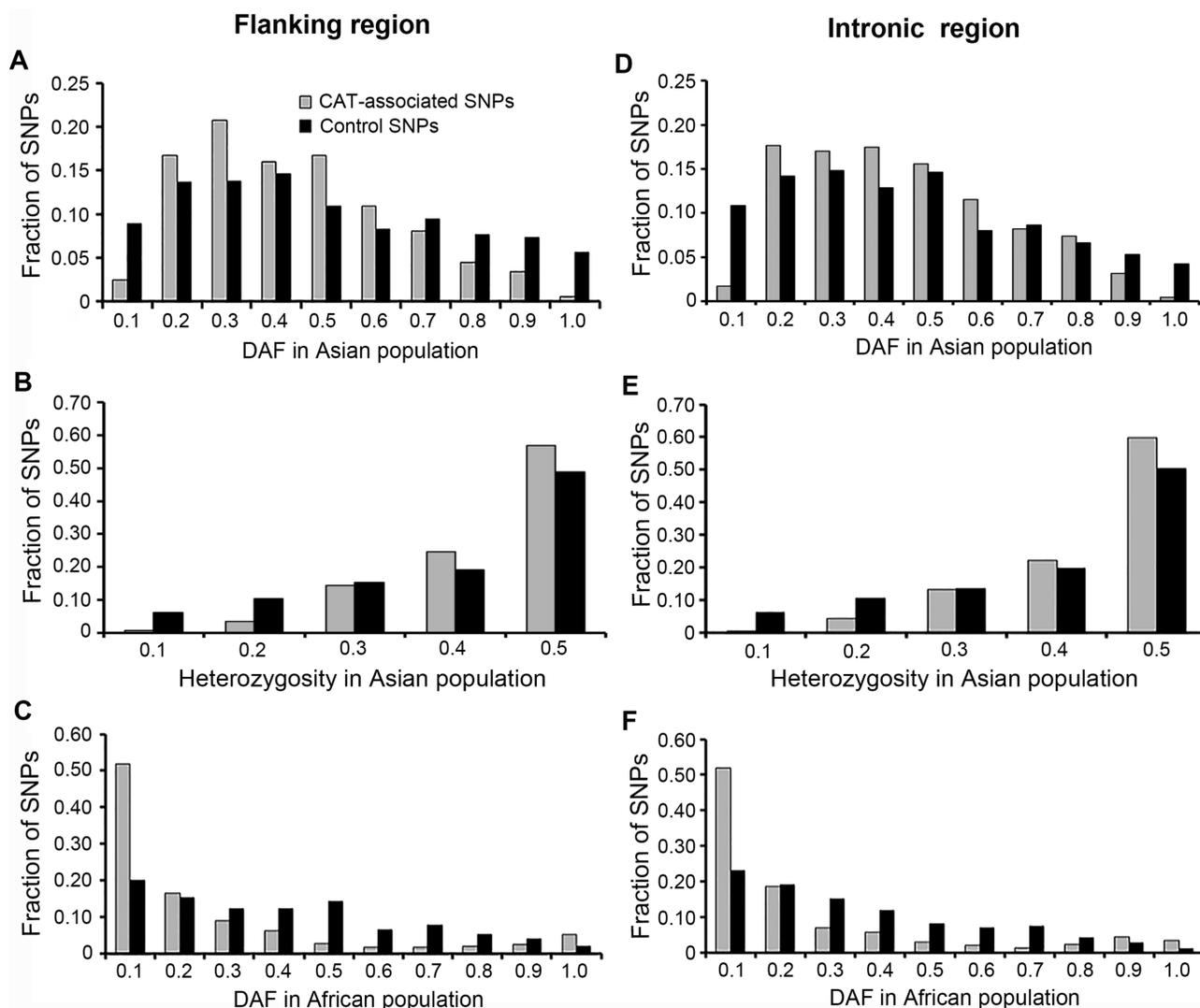


Figure 2 DAF spectrums of CAT-associated flanking region and intronic region SNPs in Asian and African populations DAF distributions of flanking region SNPs in Asian (A) and African populations (C). DAF distributions of intronic region SNPs in Asian (D) and African populations (F). Heterozygosity plots of Asian population for flanking region SNPs (B) and intronic region SNPs (E). DAF, derived allele frequency.

For the 12 HGDP-CEPH constituent populations, those residing in China, Japan, Russia, and Siberia were also suggested to have experienced balancing selection, whereas all four African populations could have experienced positive or purifying selection events (Figures S5 and S6). For the remaining populations, the French and Italian populations each showed a non-significant trend of balancing selection, whereas the populations residing in Israel and Pakistan remained ambiguous (Figure S7). Therefore, a clear world-wide selection pattern was revealed, indicating geographically specific genetic adaptation to CAT.

Next, F_{ST} was examined to measure the degree of population differentiation and to identify local adaptation signals. Among the four HGDP-CEPH continental populations, according to F_{ST} results derived from control SNPs, the population differentiation pattern was generally consistent with their speculated split time [31], with the European and Middle Eastern populations having the lowest F_{ST} value. However, for CAT-associated SNPs, the pattern was distorted with Asian and European populations having the lowest differentiation (Table S12). This trend was further confirmed in the HapMap data (Table S13). Several CAT SNPs were found to display high population differentiation values, namely between Asians and Africans (0.346–0.231) or between Europeans and Africans (0.277–0.218) as listed in Table S14.

Finally, iHSs were examined to identify recent and generally segregating selection sweeps. More SNPs were fixed in the African populations compared with controls, with no available iHSs (χ^2 ; $P < 5 \times 10^{-5}$). Additionally, distinct and shared positive selection signals were detected among different continental populations. For the HapMap populations, one SNP (rs174556 in *FADS1* gene) was shared between the ASN and YRI populations, and two SNPs (rs657672 in *CACNB1* and rs16823913 in *CCL20*) were shared between the ASN and CEU populations (Table S15). When only genes were considered, three genes (*CACNB1*, *CCL20*, and *CD180*) were shared in ASN and CEU populations, while only *FADS1* was shared between ASN and YRI populations (Figure S8A). In HGDP-CEPH populations, the signals were more focused on fewer genes, with one SNP (rs10942089 in *CDH10*) shared among the East Asian, Middle Eastern, and South Asian populations; another SNP (rs16960758 in *SLC12A1*) among the European, Middle Eastern, and South Asian populations; and a SNP (rs11208527 in *RAVER2*) between the European and Middle Eastern populations (Table S16). For selected genes, *CDH10* was shared among East/South Asian and Middle Eastern populations, while *SLC12A1* was shared between South Asian, Middle Eastern, and European populations. This left the African population with three distinct genes (*CLASP2*, *MAGI3*, and *SLITRK5*) (Figure S8B). Lastly, when the HapMap and HGDP-CEPH results were combined, only rs16960758 in *SLC12A1* for Europeans and rs439022 in *SLITRK5* for Africans, which are both eQTLs, were mutually verified in both datasets.

Discussion

As the major source of the Earth's energy, the Sun has played a key role in the development as well as maintenance of a rich and diverse biosphere. Therefore, it has long been proposed as a strong driving force for the origin and subsequent evolution

of human beings [32]. Morphological and physiological studies have revealed accumulating ethnic and clinal variations associated with each of the three main attributes of solar radiation: CAT, UVR, and SD [22,33,34]. However, no study thus far has examined these three factors simultaneously and the existing genetic studies can only explain a small fraction of the underlying heritability. This study sought to provide a full elucidation of the evolutionary role of solar radiation as a whole through a series of analyses with respect to CAT, UVR, and SD.

CAT is the main driving force of solar radiation

Traditional selection analyses tended to identify 'hard sweeps' that *de novo* advantageous mutations with large effective size spread rapidly to fixation, usually agnostic to the underlying selective pressures. However, such 'hard sweeps' are rare. More adaptations may be expressed by the so-called 'soft sweeps' in which subtle allele frequency shifts occur in standing variation [35]. For this type, the environmental correlation approach provides a good alternative, further pinpointing the underlying driving force [36]. In this study, we combined the two types of methodologies together. First, the environmental correlation approach was used to select candidate loci for each attribute. Next, standard selection methods including F_{ST} , iHS, and DAF were used to determine the selective scenario more specifically. By doing this, the signals and selection patterns revealed would be more accurate.

Furthermore, population structure is an important and unavoidable confounding factor. Generally, principal component (PC)-based analyses are used for correction, although the number of PCs included can be arbitrary and subjective. In this study, we carefully controlled for potential population structure in three ways. First, we included geographic coordinates, which have been confirmed as a good representative for human demographic history [37] as covariates in the multiple linear regression analysis for each SNP. Second, we corrected for null models derived from either genomic controls or deliberately defined signals. Finally, and most importantly, we verified our genome-wide signals experimentally using genetically more homogenous Chinese populations residing in multiple solar radiation environments and through confirmation in the dbCLINE database, which includes worldwide populations. By using these strategies, the identified selective signals are more likely to be truly related to UVR, CAT, and SD themselves.

Notably, CAT, especially an extremely minimum annual average temperature, is finally identified as the main force driving solar radiation selection of human genes involving development and immunity.

Development and immunity genes are mostly selected by solar radiation

Previous studies have described a close association between human morphology and CAT, in accordance with Bergmann's and Allen's rules [9,10]. In addition, annual average temperature has been found to be significantly correlated with body mass and relative sitting height [38]. These morphological variations are all related to efficient thermoregulation, from heat conservation to dissipation [33]. In this study, biological pro-

cesses that occur at multiple levels, including cell differentiation, anatomical structure morphogenesis, organ development, and regulation of developmental processes, all showed significant enrichment in CAT-associated signals. This observation is supported by Fricot et al.'s work. They tested the same HGDP dataset with multiple temperature indexes through a latent factor mixed model and found a significant enrichment of GO terms associated with human developmental processes [39]. Next, genome-wide signals provide further information about the effect of CAT upon human development. Among the five genes derived from the six genome-wide CAT-associated signals, *ULBP3* and *KRT31* are involved in hair development [40,41], while *DRD3* and *NTNG1* are involved in neuronal development [42,43]. All six signals, except rs1107877 in *KRT31*, were successfully validated; five of them were significantly associated with the annual average CAT range and four were significantly associated with an annual average extremely minimum CAT. Previous studies on human skeletal variation have shown that the minimum temperature consistently had a stronger effect on shaping human morphology than maximum temperature [44,45], and this was genetically supported by our results since most of our genome-wide CAT SNPs were more significantly associated with the annual average extremely minimum temperature or minimum temperature (winter). More interestingly, when the 316 CAT-associated genes were checked for GWAS traits, anthropometric traits, including height, weight (weight, obesity, adipose, and fat), waist-to-hip ratio, and BMI, demonstrated significant enrichment, especially in waist-hip ratio and obesity. Therefore, our overall results provide strong genomic evidence for Bergmann's and Allen's rules, further highlighting the role of low CAT upon human development.

Global variations in CAT have affected the distribution and abundance of disease vectors; accordingly, the resultant difference in infectious disease patterns could exert selective pressure on the human immune system [46]. This expectation has been partially verified as the pathogen selection hypothesis; that is, the local burden of infective agents would shape the variability of human genome, especially for immune-related genes [47]. The identified immunity clustering GO terms and enrichment of immunity-relevant GWAS-mapped traits within the CAT-associated SNPs provide genomic evidence linking CAT and immune system variations directly. This conclusion is also supported at the individual gene level. Because of pleiotropy, the protein encoded by *ULBP3* is also a ligand for the NKG2D receptor, and it can activate multiple signaling pathways in primary NK cells to produce cytokines and chemokines [48]. In addition, *DRD3*-encoded dopamine receptor D3 is important for modulating T cell function by inhibiting the activated T cell receptor-induced cell proliferation and secretion of IL-2, IFN- γ , and IL-4 [49]. Together, these results not only reveal a significant association between CAT and human immunity, but they also help elucidate CAT as the underlying environmental basis for the pathogen selection hypothesis.

In addition, UVR-associated signal *NADK* is involved in most reductive biosynthetic reactions and antioxidant defense systems by controlling the NADPH concentration [50], whereas *IGFBP3* is associated with osteoblast differentiation and bone development [51]. The UVR-associated signals can thus be genetically attributed to the developmental process and immunity as well.

The underlying selection mechanism of solar radiation

As described in the recently modified Out-of-Africa dispersal model with minor introgression from archaic humans [52], human populations have lived in tropical Africa for the longest time; therefore, as they migrated and resided in temperate Eurasia and arctic Siberia, the declining CAT outside Africa would exert a new selective pressure [33]. Accordingly, our results suggest that the four African populations who lived within relatively invariant hot environments for the longest time (average CAT range from 27.0 °C–29.3 °C for the three tropical populations and 19.0 °C for South Africa) might have experienced ancient events of positive selection (complete sweeps), or they may have evolved under strict purifying selection. In addition, among the 401 CAT-associated SNPs, 49 were completely fixed, and more than half were nearly fixed in Africans. Taking into account these results along with the positively selected signals indicated by iHS and F_{ST} statistics for CAT SNPs, a selective scenario of positive selection upon several genes (by both hard sweep and soft sweep) together with events of purifying selection in other loci probably provide an optimal adaptation for African populations. As humans migrated out of Africa, the ancient heat-adapted genetic background encountered diverse CAT (average ranging from –10.0 °C in Siberia to 18.3 °C in Pakistan), particularly cold stress during the last glacial period [53]. Furthermore, the temperature fluctuates more fiercely throughout the year outside tropical Africa, usually with both hot summers and cold winters. Consistent with this change, five of the six CAT genome-wide signals were significantly associated with the annual average CAT range: four with annual average extremely minimum CAT and one with annual average maximum CAT. Therefore, under these conditions, balancing selection could probably result in more adaptation, even for the arctic Siberia population. Still, signals of positive selection, as indicated by iHS and F_{ST} , may also have contributed to this local adaptation to CAT sweeping variants from low to intermediate frequencies. Here, positive selection would tend to act on pre-existing variations (selection on standing variation – soft-sweeps), except for the Siberian population. The situation, for Israeli and Pakistani populations, however, is ambiguous and may reflect complex interactions between natural selection and demographic history. Further studies with larger sample sizes and from more subpopulations are needed to better elucidate their selection patterns.

As for the associations between the morphological variations and CAT, all are related to efficient thermoregulation from heat conservation to dissipation, as described in Bergmann's and Allen's rules. Among the 316 CAT-associated genes, additional signals relevant to thermoregulation, including those involved in hair formation (*KRT31*, *KRT32*, *KRT35*, *KRTAP4-1*, *COL22A1*, *CSMD1*, and *OVOL1*), temperature homeostasis (*ACOT11* and *PPARGC1A*), and body fluid balance (*SLC12A1*), were highlighted. Here, solute carrier family 12 member 1 encoded by *SLC12A1* mediates Na⁺, K⁺, and Cl[–] reabsorption. It also plays a vital role in the regulation of ionic balance and cell volume associated with sweating [54]. Recently, *SLC12A1* has been strongly suggested as a candidate gene for hypertension in a large-scale meta-analysis [55]. The derived rs16960758C allele in this gene may have undergone a strong positive selection among European, Middle

Eastern, and South Asian populations, as indicated by iHS. Indeed, the loss of hair and a concomitant increase in the number of sweat glands are suggested to have been advantageous for sensitive whole-body cooling during early human evolution [56]. Collectively, these results provide evidence of genetic adaptation owing to thermoregulation [57].

Unlike previous evolutionary studies, which found a larger fraction of nonsynonymous SNPs [22], the identified CAT-associated signals in this study tended to be located more in flanking regions and introns. The GTEx project and HaploReg database were searched for the functional impact of the genome-wide SNPs. In total, five out of six CAT-associated signals and two of the UVR-associated signals all have important functional impact, either as *cis*- or *trans*-regulate genes or as modifiers of associated regulatory motifs. Therefore, the selection of CAT could be through gene regulation.

Implication for future clinical studies

For the CAT-associated signals, the following gene categories showed significant enrichment: immunity (92), hypertension and blood pressure (14), mental function and mental disorder (97), and cancer (49). This provides a new perspective for elucidating the genetic architecture and pathogenesis of these complex traits, as demonstrated by the examples of infectious diseases [58] and essential hypertension [59]. Generally, two main models were suggested to explain these observations. First, as proposed in the ancestral-susceptibility model, the prevalent alleles, most of which are ancestral and adaptive in Africa, may increase disease susceptibility, whereas the other alleles, which are mostly derived ones, are gradually becoming advantageous owing to the changing environment [60]. This model has been well explained in essential hypertension through the thrifty gene hypothesis [59,61]. Second, genetic pleiotropy is responsible for both selected developmental processes and developmental disorders. This explanation is partially supported by the massive overlap between anthropometric genes and either mental or cancer related genes within the CAT signals (Figure S9). Meanwhile, selective disadvantages late in life accompanying a productive advantage at an early age will show themselves more clearly as human longevity increases [62]. Although the present results are insufficient to fully elucidate the evolutionary mechanisms for these complex traits, they offer an opportunity to connect associated genes and diseases with their underlying environmental factors, thereby providing new clinical perspectives.

Conclusion

In summary, this is the first systematic study of the evolutionary effects of solar radiation on human genome diversity. CAT, especially cold stress, was identified as the main driving force of selection by solar radiation. Together with two other attributes, CAT mostly selected for genes involved in human development and immunity, with thermal regulation being the major mechanism. Specifically, the enrichment in developmental genes provides direct genomic evidence for Bergmann's and Allen's longstanding rules, and the enrichment in immunity categories reveals the underlying environmental driving force for the pathogen selection hypothesis. Furthermore,

human populations residing in different environments and with different demographic histories showed region-specific selective patterns. This study thus provides an initial genetic insight into the evolutionary role of solar radiation in *Homo sapiens*; it also provides clues for future studies on humans and other species.

Materials and methods

Studied HGDP-CEPH human populations ($N = 764$)

Population genotype data were retrieved from the HGDP-CEPH dataset, which included 952 unrelated individuals from 52 populations [21]. Because more genetic data have accumulated for the HapMap YRI, CEU, CHB, and JPT populations that could be used for confirmation, all 40 HGDP-CEPH African, European, and Asian populations with specific climatic data were included (Biaka Pygmy, Cambodian, Mbuti Pygmy, Mozabite, and San were excluded owing to a lack of specific climatic data). Genetic data for these 40 populations, which were further merged into 12 integrated countries/regions ($N = 764$, Table S11) to enhance representativeness [15], were used for data analysis.

Replicating Chinese human populations ($N = 1395$)

To better control for the underlying population structure, the results from the HGDP-CEPH populations were validated in a replication sample set of Chinese populations (Table S17). A total of 971 Han individuals were sampled randomly from nine provinces of China, and the four Chinese ethnic minorities were from six resident areas (Figure S10, $N = 424$). The geographic origin of each individual was ascertained by self-reported precoded geographic tenancy for three generations, and only individuals without a history of family migration were used.

Climatological and geographic information

For HGDP-CEPH populations (Table S11), environmental data, including the daily average SD, CAT, and altitude, were extracted from the World Meteorological Organization [15] according to the sampling latitude and longitude data provided [21]. UVR (UV minimum erythema dose, UVMED) data were derived from readings taken from the NASA Total Ozone Mapping Spectrometer (TOMS) [3]. For replicating Chinese populations, climatological data (daily average SD; annual average CAT, maximum/minimum CAT, CAT range, extreme maximum/minimum CAT; annual and monthly average UVR dose) were collected from the China Meteorological Data Sharing Service System according to geographic information (altitude, latitude, and longitude) [15]. For all populations, if specific climatic information was not available, information for the closest neighboring site was used as a substitute. In addition, an average would be taken if samples were collected from more than one nearby location. All the climatic data were demonstrated in Tables S18 and S19, and all geographic data were transformed into natural logarithm form to ensure normality.

SNP genotyping for replication

Six genome-wide CAT-associated SNPs (rs1107877, rs11185115, rs12202737, rs12626864, rs13729, and rs9825563) were experimentally confirmed in Chinese populations. Genomic DNA was extracted from whole blood using the standard phenol/chloroform method. Among them, rs11185115 within *NTNG1* was replaced by a nearby SNP, rs1444041 ($r^2 = 0.954$), which was not in the original dataset, owing to a lack of the appropriate SNPstream primers; rs12202737 was represented by rs13729 from the same *ULBP3* gene because of their complete LD in the CHB population ($r^2 = 1$). In total, four SNPs (rs12626864, rs13729, rs1444041, and rs9825563) were genotyped using the Beckman SNPstream genotyping system [63], while one SNP (rs1107877) was genotyped by the PCR-RFLP method (Sau3AI, Takara, China). For 2 UVR-associated SNPs, rs1586360 was not genotyped owing to its low MAF in the CHB populations (MAF < 0.1). Only rs7531583 was genotyped using the same Beckman SNPstream genotyping system.

GWAS data analyses

For each of the 516,663 common SNPs (MAF ≥ 0.1) of the whole 660,918 SNPs in the original HGDP-CEPH dataset [21], allelic frequencies were calculated for the 12 HGDP-CEPH countries/regions. Pearson bivariate correlation was then carried out to evaluate the association between allele frequencies and each of the three solar radiation attributes. For each solar radiation attribute, any significant Pearson bivariate correlation was further confirmed by multiple linear regression analysis using the three geographic parameters and the other two attributes as covariates (SPSS for Windows; 16.0). A modified positive false discovery rate, pFDR, was explored to correct for whole genome testing [64], and the WGAViewer software [65] was used to draw Manhattan plots to show the locations of the significant signals (see the flowchart, Figure S11).

GWAS mapping and functional clustering

Associated genes were assigned according to the dbSNP database, and SNPs within an intergenic region were annotated with the closest gene symbol. Then, they were mapped to GO annotation terms [23] and KEGG pathways [24] using the publicly available program Database for Annotation, Visualization, and Integrated Discovery (DAVID) [66]. Here, the GO and KEGG results of all 17,169 HGDP-CEPH dataset genes were used as null distributions. In addition, these associated genes were checked in the NHGRI-EBI GWAS Catalog (<http://www.ebi.ac.uk/gwas/search>) through 20 May 2016 for potential phenotype effect. Similarly, the GWAS category distribution patterns for both the 12,456 total GWAS mapped genes and the 17,169 HGDP-CEPH genes were used as two separate null distributions.

Population genetics analysis

In addition to the HGDP-CEPH data, allele frequencies for the HapMap CHB, JPT, CEU, and YRI populations retrieved

from HapMart (HapMap_rel27 dat) were analyzed. First, in DAF analysis, SNPs for CAT-associated intronic and flanking region SNPs were performed separately, with specific controls defined from the same Illumina 650Y dataset. For flanking region SNPs, controls should be from unassociated genes; the physical distance should be similar to that of the associated SNP, with matches made in a one-to-one manner. For intronic SNPs, we first carried out a similar one-to-one matching. Then we performed 100 simulations by randomly re-sampling the intronic SNPs from the remaining unassociated genes, regardless of physical distance. For all these SNPs, the ancestral allele information was retrieved from dbSNP. Second, pairwise F_{ST} values [29] for each SNP were calculated to evaluate the degree of population differentiation and identify local adaptation signals. Third, for each SNP, the iHS was retrieved from Haplotter [30] for HapMap data and from the HGDP Selection Browser [67] for HGDP-CEPH data, with an $|iHS| \geq 2.0$ used to infer positive selection signals. Notably, the control SNPs defined in the DAF analysis were also used as controls in the F_{ST} and iHS analyses.

Statistical analysis for Chinese replication populations

HWE was tested through the software PEDSTATS [68]. Statistical analysis of the gender and age differences of allele frequencies was performed by χ^2 tests (SPSS for Windows; 16.0). Environmental associations were assessed using Pearson bivariate correlation and multiple linear regression analysis as described for the worldwide populations (SPSS for Windows; 16.0).

Ethical statement

Blood samples from all 1395 Chinese individuals for the confirmation experiment were obtained with written informed consent. The protocol for this study was reviewed and approved by the ethics committee of the Kunming Institute of Zoology, Chinese Academy of Sciences.

CRedit author statement

Lindan Ji: Investigation, Formal analysis, Writing - original draft, Writing - review & editing. **Dongdong Wu:** Formal analysis, Writing - original draft. **Haibing Xie:** Formal analysis, Writing - original draft. **Binbin Yao:** Formal analysis. **Yanming Chen:** Formal analysis. **David M. Irwin:** Writing - original draft, Writing - review & editing. **Dan Huang:** Formal analysis. **Jin Xu:** Conceptualization, Supervision, Formal analysis, Writing - original draft, Writing - review & editing. **Nelson L.S. Tang:** Conceptualization, Supervision, Formal analysis, Writing - original draft, Writing - review & editing. **Yaping Zhang:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgments

This work was supported by grants from the National Key R&D Program of China (Grant No. 2018YFE0203700), the National Natural Science Foundation of China (Grant No. 81402747), the Applied Research Project on Nonprofit Technology of Zhejiang Province (Grant No. LGF20H260009), and the KC Wong Magna Fund in Ningbo University, China. NLST was also supported by CUHK VC discretionary fund, China. We also thank Dr. [Yuqing Qiu](#) and Dr. Ling Wang for their help in the data analyses.

Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2019.11.009>.

ORCID

0000-0003-2043-1150 (Lindan Ji)
 0000-0001-7101-7297 (Dongdong Wu)
 0000-0003-4977-8270 (Haibing Xie)
 0000-0003-0035-8355 (Binbin Yao)
 0000-0002-2328-0944 (Yanming Chen)
 0000-0001-6131-4933 (David M. Irwin)
 0000-0003-3022-5772 (Dan Huang)
 0000-0002-4745-9850 (Jin Xu)
 0000-0002-3607-5819 (Nelson L.S. Tang)
 0000-0002-5401-1114 (Yaping Zhang)

References

- [1] Mead MN. Benefits of sunlight: a bright spot for human health. *Environ Health Perspect* 2008;116:A160–7.
- [2] Ciarleglio CM, Ryckman KK, Servick SV, Hida A, Robbins S, Wells N, et al. Genetic differences in human circadian clock genes among worldwide populations. *J Biol Rhythms* 2008;23:330–40.
- [3] Jablonski NG, Chaplin G. Human skin pigmentation as an adaptation to UV radiation. *Proc Natl Acad Sci U S A* 2010;107:8962–8.
- [4] Baggerly CA, Cuomo RE, French CB, Garland CF, Gorham ED, Grant WB, et al. Sunlight and vitamin D: necessary for public health. *J Am Coll Nutr* 2015;34:359–65.
- [5] Juzeniene A, Grigalavicius M, Ma LW, Juraleviciute M. Folic acid and its photoproducts, 6-formylpterin and pterin-6-carboxylic acid, as generators of reactive oxygen species in skin cells during UVA exposure. *J Photochem Photobiol B* 2016;155:116–21.
- [6] Brenner M, Hearing VJ. The protective role of melanin against UV damage in human skin. *Photochem Photobiol* 2008;84:539–49.
- [7] Gruss LT, Schmitt D. The evolution of the human pelvis: changing adaptations to bipedalism, obstetrics and thermoregulation. *Philos Trans R Soc Lond B Biol Sci* 2015;370:20140063.
- [8] Ruff C. Variation in human body size and shape. *Annu Rev Anthropol* 2002;31:211–32.
- [9] Allen JA. The influence of physical conditions in the genesis of species. *Radic Rev* 1877;1:108–40.
- [10] Bergmann C. Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. *Göttinger Studien* 1847;3:595–708.
- [11] Leonard WR, Sorensen MV, Galloway VA, Spencer GJ, Mosher MJ, Osipova L, et al. Climatic influences on basal metabolic rates among circumpolar populations. *Am J Hum Biol* 2002;14:609–20.
- [12] Amorim CE, Daub JT, Salzano FM, Foll M, Excoffier L. Detection of convergent genome-wide signals of adaptation to tropical forests in humans. *PLoS One* 2015;10:e0121557.
- [13] Doherty CJ, Kay SA. Circadian control of global gene expression patterns. *Annu Rev Genet* 2010;44:419–44.
- [14] Dopico XC, Evangelou M, Ferreira RC, Guo H, Pekalski ML, Smyth DJ, et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nat Commun* 2015;6:7000.
- [15] Ji LD, Xu J, Wu DD, Xie SD, Tang NL, Zhang YP. Association of disease-predisposition polymorphisms of the melatonin receptors and sunshine duration in the global human populations. *J Pineal Res* 2010;48:133–41.
- [16] Gu F, Han J, Laden F, Pan A, Caporaso NE, Stampfer MJ, et al. Total and cause-specific mortality of U.S. nurses working rotating night shifts. *Am J Prev Med* 2015;48:241–52.
- [17] Liu F, Wen B, Kayser M. Colorful DNA polymorphisms in humans. *Semin Cell Dev Biol* 2013;24:562–75.
- [18] Kita R, Fraser HB. Local adaptation of sun-exposure-dependent gene expression regulation in human skin. *PLoS Genet* 2016;12:e1006382.
- [19] Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jorgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 2015;349:1343–7.
- [20] Cruciani F, Trombetta B, Labuda D, Modiano D, Torroni A, Costa R, et al. Genetic diversity patterns at the human clock gene *period 2* are suggestive of population-specific positive selection. *Eur J Hum Genet* 2008;16:1526–34.
- [21] Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008;319:1100–4.
- [22] Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, et al. Adaptations to climate-mediated selective pressures in humans. *PLoS Genet* 2011;7:e1001375.
- [23] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet* 2000;25:25–9.
- [24] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 2006;34:D354–7.
- [25] Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42:D1001–6.
- [26] Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015;348:648–60.
- [27] Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–4.
- [28] Chen K, Rajewsky N. Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 2006;38:1452–6.
- [29] Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002;12:1805–14.
- [30] Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol* 2006;4:e72.
- [31] Oppenheimer S. Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map. *Philos Trans R Soc Lond B Biol Sci* 2012;367:770–84.
- [32] Roberts DF. Climate and human variability. Menlo Park: Benjamin-Cummings Publishing Company; 1978.
- [33] Beall CM, Jablonski NG, Steegmann A. Human adaptation to climate: temperature, ultraviolet radiation, and altitude. In: Sara Stinson BB, O'Rourke D, editors. *Human biology: an evolutionary and biocultural perspective*. New York: John Wiley and Sons; 2012. p. 177–250.

- [34] Hut RA, Paolucci S, Dor R, Kyriacou CP, Daan S. Latitudinal clines: an evolutionary view on biological rhythms. *Proc Biol Sci* 2013;280:20130433.
- [35] Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 2010;20:R208–15.
- [36] Haasl RJ, Payseur BA. Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Mol Ecol* 2016;25:5–23.
- [37] Handley LJ, Manica A, Goudet J, Balloux F. Going the distance: human population genetics in a clinical world. *Trends Genet* 2007;23:432–9.
- [38] Roberts DF. Body weight, race and climate. *Am J Phys Anthropol* 1953;11:533–58.
- [39] Frichot E, Schoville SD, Bouchard G, Francois O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol* 2013;30:1687–99.
- [40] Betz RC, Petukhova L, Ripke S, Huang H, Menelaou A, Redler S, et al. Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nat Commun* 2015;6:5966.
- [41] Langbein L, Rogers MA, Winter H, Praetzel S, Beckhaus U, Rackwitz HR, et al. The catalog of human hair keratins. I. Expression of the nine type I members in the hair follicle. *J Biol Chem* 1999;274:19874–84.
- [42] Sun X, Zhao Y, Wolf ME. Dopamine receptor stimulation modulates AMPA receptor synaptic insertion in prefrontal cortex neurons. *J Neurosci* 2005;25:7342–51.
- [43] Song YS, Lee HJ, Prossellkov P, Itohara S, Kim E. Trans-induced cis interaction in the tripartite NGL-1, netrin-G1 and LAR adhesion complex promotes development of excitatory synapses. *J Cell Sci* 2013;126:4926–38.
- [44] Betti L, Balloux F, Hanihara T, Manica A. The relative role of drift and selection in shaping the human skull. *Am J Phys Anthropol* 2010;141:76–82.
- [45] Betti L, von Cramon-Taubadel N, Manica A, Lycett SJ. The interaction of neutral evolutionary processes with climatically-driven adaptive changes in the 3D shape of the human os coxae. *J Hum Evol* 2014;73:64–74.
- [46] Khasnis AA, Nettleman MD. Global warming and infectious disease. *Arch Med Res* 2005;36:689–96.
- [47] Fumagalli M, Cagliani R, Pozzoli U, Riva S, Comi GP, Menozzi G, et al. Widespread balancing selection and pathogen-driven selection at blood group antigen genes. *Genome Res* 2009;19:199–212.
- [48] Cosman D, Mullberg J, Sutherland CL, Chin W, Armitage R, Fanslow W, et al. ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. *Immunity* 2001;14:123–33.
- [49] Sarkar C, Basu B, Chakroborty D, Dasgupta PS, Basu S. The immunoregulatory role of dopamine: an update. *Brain Behav Immun* 2010;24:525–8.
- [50] Pollak N, Niere M, Ziegler M. NAD kinase levels control the NADPH concentration in human cells. *J Biol Chem* 2007;282:33562–71.
- [51] Li J, Jin D, Fu S, Mei G, Zhou J, Lei L, et al. Insulin-like growth factor binding protein-3 modulates osteoblast differentiation via interaction with vitamin D receptor. *Biochem Biophys Res Commun* 2013;436:632–7.
- [52] Arnason U. The Out of Africa hypothesis and the ancestry of recent humans: Cherchez la femme (et l'homme). *Gene* 2016;585:9–12.
- [53] Shakun JD, Carlson AE. A global perspective on Last Glacial Maximum to Holocene climate change. *Quat Sci Rev* 2010;29:1801–16.
- [54] Markadieu N, Delpire E. Physiology and pathophysiology of SLC12A1/2 transporters. *Pflugers Arch* 2014;466:91–105.
- [55] Liu C, Kraja AT, Smith JA, Brody JA, Franceschini N, Bis JC, et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* 2016;48:1162–70.
- [56] Jobling MA, Hurler M, Tyler-Smith C. Human evolutionary genetics: origins, peoples & disease. New York: Garland Pub; 2004.
- [57] Jablonski NG. The evolution of human skin and skin color. *Annu Rev Anthropol* 2004;33:585–623.
- [58] Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Riva S, Clerici M, et al. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *J Exp Med* 2009;206:1395–408.
- [59] Young JH, Chang YP, Kim JD, Chretien JP, Klag MJ, Levine MA, et al. Differential susceptibility to hypertension is due to selection during the out-of-Africa expansion. *PLoS Genet* 2005;1:e82.
- [60] Di Rienzo A, Hudson RR. An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends Genet* 2005;21:596–601.
- [61] Ji LD, Tang NL, Xu J. AGTR1 has undergone natural selection in Euro-Asian populations in relation to ambient temperature that predisposes Chinese populations to essential hypertension. *Int J Cardiol* 2016;209:278–80.
- [62] Williams GC. Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 1957:398–411.
- [63] Denomme GA, Van Oene M. High-throughput multiplex single-nucleotide polymorphism analysis for red cell and platelet antigen genotypes. *Transfusion* 2005;45:660–6.
- [64] Storey JD. A direct approach to false discovery rates. *J Roy Stat Soc Ser B (Stat Method)* 2002;64:479–98.
- [65] Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, et al. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* 2008;18:640–3.
- [66] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
- [67] Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 2009;19:826–37.
- [68] Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* 2005;21:3445–7.