Novel genomic signals of recent selection in an Ethiopian population

Fasil Tekola-Ayele1, Adebowale Adeyemo1, Guanjie Chen1, Elena Hailu2, Abraham Aseffa2, Gail Davey3, Melanie J Newport3 and Charles N Rotimi1

The recent feasibility of genome-wide studies of adaptation in human populations has provided novel insights into biological pathways that have been affected by adaptive pressures. However, only a few African populations have been investigated using these genome-wide approaches. Here, we performed a genome-wide analysis for evidence of recent positive selection in a sample of 120 individuals of Wolaita ethnicity belonging to Omotic-speaking people who have inhabited the mid- and high-land areas of southern Ethiopia for millennia. Using the 11 HapMap populations as the comparison group, we found Wolaita-specific signals of recent positive selection in several human leukocyte antigen (HLA) loci. Notably, the selected loci overlapped with HLA regions that we previously reported to be associated with podoconiosis—a geoclimatic lymphedema of the lower legs common in the Wolaita area. We found selection signals in PPARA, a gene involved in energy metabolism during prolonged food deficiency. This finding is consistent with the dietary use of enset, a crop with high-carbohydrate and low-fat and -protein contents domesticated in Ethiopia subsequent to food deprivation 10,000 years ago, and with metabolic adaptation to high-altitude hypoxia. We observed novel selection signals in CDKAL1 and NEGR1, well-known diabetes and obesity susceptibility genes. Finally, the SLC24A5 gene locus known to be associated with skin pigmentation was in the top selection signals in the Wolaita, and the alleles of single-nucleotide polymorphisms rs1426654 and rs1834640 (SLC24A5) associated with light skin pigmentation in Eurasian populations were of high frequency (47.9%) in this Omotic-speaking indigenous Ethiopian population.

European Journal of Human Genetics (2015) 23, 1085–1092; doi:10.1038/ejhg.2014.233; published online 5 November 2014

INTRODUCTION

Adaptive response to geographically restricted selective pressures, such as diet, climate, and pathogen burden, is one of the drivers of population differences in frequencies of disease-associated genetic variants. Advantages variants that enhance survival against these selective pressures in human history rise to high frequency, and may now increase risk for chronic diseases as a consequence of lifestyle and ecological changes. This may have contributed to differences in the prevalence of diseases such as type 2 diabetes and malaria, and traits such as lactase persistence. Identifying genomic loci that have been the subject of selection and discerning their relationship with disease phenotypes is instructive in understanding the adaptive genetic basis of population-level differences in the prevalence of common diseases.

Several genome-wide scans for recent positive selection have identified genomic loci displaying signatures of natural selection. However, only a few African populations have been included in these studies. The rich genetic, ecological, and socio-cultural diversity of African populations is favorable for the detection of locally restricted and shared selection signals, and for elucidation of putative selective forces.

The Wolaita are one of the indigenous Ethiopian populations that have inhabited the mid- and high-land areas of southern Ethiopia for several thousand years, and predominantly speak Wolaita, an Omotic branch of the Afroasiatic language family. Compared with the HapMap samples, the Wolaita (WETH) had the smallest genetic differentiation from Kenyan Maasai (MKK) and the largest differentiation from Japanese (JPT), and lay at the farthest end of the African genetic cluster nearest to MKK and farthest from Nigerian Yoruba (YRI).

In the present study, we performed a genome-wide analysis to identify regions displaying evidence of recent positive selection in a sample of 120 WETH individuals. Our analyses revealed strong evidence of recent positive selection in the human leukocyte antigen (HLA) locus and in genes involved in energy metabolism. We also found that loci selected possibly for survival against pathogens and food deficiency in the past overlapped with genome-wide association study (GWAS) loci linked with protection to podoconiosis, and increased risk for metabolic disorders.

METHODS

Data sets
We used IlluminaHap 610 Bead Chip genotype data of 120 randomly selected individuals from the Wolaita ethnic group from Southern Ethiopia (WETH; Supplementary Figure 1) who were recruited to serve as controls in a GWAS of podoconiosis. Of the 551,840 autosomal single-nucleotide polymorphisms (SNPs) in the raw data set, 39,948 SNPs failed data quality filters (Supplementary Methods). The remaining 511,892 SNPs that passed quality control were merged with the HapMap data set that contained 1,440,616 SNP genotypes in 1,184 individuals from 11 populations. A total of 464,642 SNPs were common to both data sets in 1,075 unrelated individuals (120 from WETH 1184 individuals from 11 populations. A total of 464,642 SNPs were common to both data sets in 1,075 unrelated individuals (120 from WETH...
and 955 unrelated individuals in HapMap 3.2). The ethno-geographical breakdown of the HapMap sample has been described in Supplementary Methods (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap3r2_B36/).  

**Tests for signatures of recent positive selection**  
We performed the integrated haplotype score (iHS, which identifies partial sweeps) and cross-population extended haplotype homozygosity (XP-EHH, which identifies complete sweeps) tests. The SNP genotypes were phased using fastPHASE v 1.4. The haplotype inference and missing genotype estimation error rates in WETH were similar to or less than those in other African populations (detailed in Supplementary Methods, Supplementary Tables 1 and 2). We performed 11 XP-EHH tests comparing WETH with each HapMap sample and one iHS test for WETH (Supplementary Methods). For every SNP, we computed $F_{ST}$ using the method of Reynolds, Weir and Cockerham, and assessed statistical significance using a Bonferroni-corrected permutation test $P$-value as described elsewhere.  

**Pathway enrichment analysis**  
We identified all genes within 100 kb up- and down-stream of SNPs in the top 0.1% iHS score and performed pathway enrichment analysis using PANTHER (http://www.pantherdb.org/).  

**RESULTS**  
**Signatures of recent positive selection**  
A total of 453 SNPs were found in the 0.1% tail of the empirical distribution of iHS; SNPs in the HLA locus were overrepresented (82/453, $P<0.0001$). Moreover, 6 of the 10 SNPs with the highest iHS scores were in the HLA locus (Table 1; Supplementary Tables 3 and 4). The XP-EHH test found 21 SNPs that were selected for in WETH compared with each HapMap sample; the HLA locus was over-represented with 12 out of the 21 SNPs located near HLA-DQB1 and BTNL2 genes (Table 2). Other selected loci include NCAM1, which is involved in immunity, and ASTN2, which is expressed in the brain and previously reported to be in the top 0.1% selected loci in the Ethiopian Amhara and Ari populations.  

$F_{ST}$ test identified 27 SNPs with statistically significant differentiation between WETH and all HapMap samples (Table 3, Supplementary Table 5). The list included the obesity risk locus NEGR1 previously shown to be highly differentiated in sub-Saharan Africans. The $F_{ST}$ values and iHS scores of the HLA loci had overlapping distributions (Figure 1). The SNP with significantly differentiated $F_{ST}$ between WETH and all HapMap populations in the HLA locus (rs2233971, a missense mutation in C6orf15) also had the highest iHS in WETH. Nucleotide substitution (rs2233971 = chr6. hg19: g.31080323G > T) in the C6orf15 gene resulting in arginine to serine substitution is predicted to be 'possibly damaging' by PolyPhen. We also found significant $F_{ST}$ differentiations in ASTN2 and HLA-DRA loci, replicating the XP-EHH and iHS findings.  

Re-construction of founder haplotypes using haploPS (details in Supplementary Methods) found 101 signals of selection (Supplementary Table 6), and several overlaps with the iHS, XP-EHH, and $F_{ST}$ signals (Supplementary Tables 7–9). To assess whether**

**Table 1 Number of SNPs showing evidence of recent positive selection in WETH**  

| Population | Total (n=21,157) | Top 0.1% (n=453) |
|------------|-----------------|-----------------|
| MKK | 22,248 | 2984 (13.4) | 87 (0.4) |
| LWK | 23,146 | 3532 (15.3) | 130 (0.6) |
| YRI | 22,872 | 3450 (15.1) | 120 (0.5) |
| ASW | 23,629 | 3348 (14.2) | 130 (0.6) |
| CEU | 21,856 | 1863 (8.5) | 54 (0.2) |
| TSI | 21,703 | 1968 (9.1) | 69 (0.3) |
| CHB | 21,521 | 1612 (7.5) | 65 (0.3) |
| CHD | 21,422 | 1552 (7.2) | 60 (0.3) |
| JPT | 21,536 | 1563 (7.3) | 78 (0.4) |
| MEX | 22,210 | 1857 (8.4) | 45 (0.2) |
| GIH | 21,771 | 1850 (8.5) | 50 (0.2) |

**Common to HapMap East African populations**  
MKK and LWK 7655 1600 (20.9) 60 (0.8) 45 22 (48.9) 2 (4.4)  
**Common to all HapMap continental African populations**  
MKK, LWK, YRI 4747 1211 (25.5) 53 (1.1) 13 5 (38.5) 0  
**Common to all HapMap African ancestry populations**  
MKK, LWK, YRI, ASW 3156 910 (28.8) 52 (1.6) 7 4 (57.1) 0  
**Common to all HapMap populations**  
11 HapMap populations 21 5 (23.8) 1 (4.8) 0 0 0  

**Abbreviations:** iHS, integrated haplotype score; SNP, single-nucleotide polymorphism; XP-EHH, cross-population extended haplotype homozygosity.  
*Population groups include Wolaita from Wolaita zone, Ethiopia (WETH), African ancestry in Southwest USA (ASW), Luyha in Webuye, Kenya (LWK), Maasai in Kenya (MKK), Yoruba in Ibadan, Nigeria (YRI), Tuscans in Italy (TSI), Utah residents with Northern and Western European ancestry from the CEPH collection (CEU), Mexican ancestry in Los Angeles, California (MEX), Gujarati Indians in Houston, Texas (GIH), Han Chinese in Beijing, China (CHB), Chinese in Metropolitan Denver, Colorado (CHD), Japanese in Tokyo, Japan (JPT).
the allele driving the selection signal was carried by the longest haplotype, as a proof-of-principle, we compared haplotype lengths in the HLA-DRA locus around rs3129882 using HapFinder.23 We observed that the length of the longest haplotype increased significantly when the core haplotype frequency decreased from 25 to 20%, suggesting the rs3129882 (G) variant (frequency = 22.9%) or nearby sequence variants may be driving the selection signal (Supplementary Figure 2).

Functional prediction and pathway enrichment analyses

Functional predictions of the SNPs in the top 0.1% tail of the empirical distribution of iHS are presented in Figure 2. Eleven SNPs were exonic, of which eight were missense variants in genes implicated in nervous system development (MAPT/SPPL2C), neuropsychiatric responses (ANKK1), fertilization, muscle development and neurogenesis (AADM28), and genes in the HLA locus (HLA-DOB and ZSCAN12; Table 4).

The ‘T-cell activation’ PANTHER pathway and the ‘mammary gland development’, ‘cellular defense response’, ‘response to stimulus’, and ‘antigen processing and presentation’ PANTHER biological processes showed Bonferroni-corrected statistically significant enrichment (Supplementary Tables 10 and 11).

We found several selection signals in the PPARA gene locus in the XP-EHH test; SNP rs5767743 with the highest iHS score in this locus also had XP-EHH > 2 when comparing WETH with CEU, GIH, JPT, MEX, MKK, and TSI. PPARA was a component of the PANTHER-enriched ‘carbohydrate metabolic process’ term.

Targets of recent positive selection and common traits

To identify positively selected loci that presumably enhanced survival against pathogens and food shortage in human history, but presently increase risk for chronic diseases, we explored overlaps between the identified selection signals and loci associated with podoconiosis and type 2 diabetes. We also explored selection signals known to have a role in skin pigmentation.

Podoconiosis

We compared podoconiosis GWAS loci12 with this study’s selection signals and found that the HLA region containing the 12 XP-EHH SNPs selected for among WETH overlapped with that of the top 10 GWAS SNPs.12 Pair-wise LD calculations showed modest correlation (r² = 0.56) between two SNPs with strong XP-EHH (rs9275141 and rs2856695; Table 2) and SNP rs17612858 (chr6, hg19:g.32620622A>T) that has the best GWAS signal for podoconiosis. This correlation was stronger than the average LD in the HLA locus in a 30 kb window (r² = 0.20) and between adjacent SNPs (r² = 0.33). The frequency of the TA haplotype that carries the positively selected T alleles of rs9275141 and rs2856695 and the podoconiosis-protective rs17612858 A allele was higher than expectation under linkage equilibrium among non-podoconiosis controls.
We searched for overlaps between genes in the top 0.1% iHS and Type 2 diabetes "p = 0.001) and podoconiosis cases (0.42 vs 0.25, P < 0.001; Supplementary Tables 12 and 13).

type 2 diabetes

We found a novel selection signal at rs9348453, an intronic SNP in SLC24A5 (chr15.hg19:g.48426484A>G) and rs1834640 (chr15.hg19:g.48392165A>G) implicated in skin pigmentation in Eurasian populations were within the top 1 and 3% iHS in WETH, and the A alleles of both SNPs implicated in light skin pigmentation had high frequency (47.9%) in WETH.

**DISCUSSION**

We found that the HLA locus and genes involved in immune response and metabolism are enriched for genomic signatures of selection among the Wolaita ethnic group from southern Ethiopia (WETH). The majority of the HLA selection signals found in this study were not detected in previous studies conducted in global populations including
change over time, recent positive selection can lead one allele to increase in frequency with the collective selection events ultimately maintaining multiple alleles at a locus.\textsuperscript{33} Moreover, the haplotype pattern around a locus under recent balancing selection can resemble an incomplete sweep of positive selection.\textsuperscript{33,43,44} Therefore, the LD-based tests we used detect recent balancing as well as positive selection, and currently available analysis methods have little power to distinguish these.\textsuperscript{35} The possible effect of long-term balancing selection on HLA diversity in Ethiopia has been shown by analysis of 61 global populations by Prugnolle \textit{et al.} and analysis of 535 populations by Sanchez-Mazas \textit{et al.}\textsuperscript{36,37} These studies found that HLA genetic diversity is positively correlated with pathogen diversity of a geographic region and inversely correlated with geographic distance from Ethiopia.\textsuperscript{36,37} This corroborates previous evidence suggesting the presence of more pathogens in Africa where humans have lived the longest, and in Ethiopia, where pathogen richness (the number of kinds of pathogens) is one of the greatest on a global scale.\textsuperscript{38} Historical accounts and archeological evidence indicate expansion of agriculture in the fertile Ethiopian highlands, and formation of urban centers at least as early as the fifth millennium BC.\textsuperscript{39} These markers of ancient civilization were linked with more settled life, high population density and poor hygienic conditions that facilitated spread of pathogens, which are still a significant cause of mortality and morbidity in the region. Taken together, these data suggest that pathogens are the strongest driving force behind the distinctive and highly enriched selection signals in the HLA locus that we found among this Ethiopian population.

We found signatures of positive selection in genes involved in metabolic processes including a novel selection signal in \textit{CDKAL1}, a gene that has been implicated in type 2 diabetes, pancreatic $\beta$-cell function, and insulin secretion.\textsuperscript{24,25,40–43} \textit{CDKAL1} inhibits the CDK5 protein leading to enhanced insulin secretion under conditions of high glucose levels.\textsuperscript{44,45} Therefore, reduced expression of \textit{CDKAL1} inhibits insulin secretion leading to an impaired response to glucotoxicity and increased risk for type 2 diabetes.\textsuperscript{46} The \textit{CDKAL1} rs9348453 ancestral A allele in the haplotype favored by selection had high frequency (470\%) in all population groups analyzed in this study. Characterization of sequence variation produced by the 1000G Project shows that rs9348453 has potential regulatory role in human skeletal muscle myoblast cell lines that are used to study diabetes and insulin resistance. Moreover, rs9348453 is correlated ($r^2 = 0.73$) with rs79915874 (chr6:hg19:g:21005146T\textsuperscript{4}C) that has been predicted to disrupt binding motifs of the hepatocyte nuclear factor 4 transcription factor, which is mainly expressed in the liver and pancreatic $\beta$ cells.
Table 4 Eight non-synonymous variants in the top 0.1% iHS$^a$

| SNP          | Genomic position | Derived allele freq | Gene     | Consequence to transcript | Protein residue change | PolyPhen score | SIFT score |
|--------------|------------------|--------------------|----------|---------------------------|------------------------|---------------|------------|
| rs11630901   | 15 41819367      | 4.94 A,G           | RPPA1    | Missense/splice region    | g. 41819367C          | 0.01$^a$   | 0.45$^d$ |
|             |                  |                    |          |                           | p.(Arg582Gly)         |              |            |
| rs12185268   | 17 43923683      | 3.55 G,A           | MAPT/SSPL2C | Missense           | g. 3923683A>G       | 1           | 0.001     |
|             |                  |                    |          |                           | p.(Ile471Val)         |              |            |
| rs12373139   | 17 43924130      | 3.55 G,A           | MAPT/SSPL2C | Missense           | g. 3924130G>A       | 0.63        | 0.001     |
|             |                  |                    |          |                           | p.(Gly620Arg)         |              |            |
| rs1800497    | 11 113270828     | 3.53 A,G           | ANKK1    | Missense              | g. 113270828G>Ap     | 1           | 0         |
|             |                  |                    |          |                           | (Glu713Lys)           |              |            |
| rs2232430    | 6 28359186       | 3.46 C,T           | ZSCAN12  | Missense              | g. 28359186C>T       | 0.42        | 0.005     |
|             |                  |                    |          |                           | p.(Arg294Lys)         |              |            |
| rs2233971    | 6 31080323       | 5.56 G,T           | C6orf15  | Missense              | g. 31080323G>T       | 0.52        | 0.52      |
|             |                  |                    |          |                           | p.(Arg45Ser)          |              |            |
| rs2621330    | 6 32781524       | 3.69 C,T           | HLA-DOB  | Missense              | g. 32781524C>T       | 0.95        | 0.02      |
|             |                  |                    |          |                           | p.(Val244Ile)         |              |            |
| rs7829965    | 8 24207438       | 3.75 A,G           | ADAM28   | Missense              | g. 24207438G>A       | 1           | 0         |
|             |                  |                    |          |                           | p.(Met684Ile)         |              |            |

$^a$Population groups include Wolaita from Wolaita zone, Ethiopia (WETH), African ancestry in Southwest USA (ASW), Luhya in Webuye, Kenya (LWK), Maasai in Kinyawa, Kenya (MKK), Yoruba in Ibadan, Nigeria (YRI), Tuscans in Italy (TSI), Utah residents with Northern and Western European ancestry from the CEPH collection (CEU), Mexican ancestry in Los Angeles, California (MXL), Gujarati Indians in Houston, Texas (GIH), Han Chinese in Beijing, China (CHB), Chinese in Metropolitan Denver, Colorado (CHD), Japanese in Tokyo, Japan (JPT).

$^b$Chromosome.

$^c$Alleles refer to ancestral and derived alleles, respectively. Ancestral state was taken from the ancestral state data released by the 1000G Project at ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/technical/reference/ancestral_alignments, constructed from a four-way primate alignment.

$^d$Possibly damaging by PolyPhen prediction.

$^e$Deletious by SIFT prediction.

Taken together, these findings suggest that CDKAL1 may be one of the key energy metabolism genes targeted by recent selection.$^5,47,48$ Investigation of this novel locus of selection using targeted-sequencing may provide new insights in the pathogenesis of diabetes and metabolic disorders.

The carbohydrate metabolic process term that was significantly enriched in PANTHER included the PPARA (peroxisome proliferator activated receptor alpha) gene. PPARA plays a key role in lipid and carbohydrate metabolism by direct regulation of numerous genes encoding enzymes and transport proteins that are important for glucose homeostasis and lipid metabolism.$^{49}$ PPARA is activated during energy deprivation and plays a key role in the management of energy stores during fasting and prolonged food deprivation.$^{50-53}$ Within the context of the historic dietary experiences of the Wolaita people, selection pressure acting on the PPARA gene in the Wolaita people who have inhabited the Ethiopian highlands for millennia. Similar to the effect of a high-carbohydrate and low-fat content diet, genetic and non-genetic adaptations to hypoxia at high altitudes lead to a shift from lipid to carbohydrate oxidation, promotion of lipid storage, and preference for anaerobic glycolysis for energy expenditure.$^{56}$ Consistent with this thinking, previous studies have shown that PPARA may be implicated in high-altitude adaptation in the Amhara population group from northern Ethiopia and in Tibetans.$^{21,57}$ Also, correlation between the selected PPARA haplotype and serum-free fatty acid levels has been observed among Tibetans.$^{56}$ In all, PPARA may have been targeted by reinforced selective forces from the nutritional content of the enset diet and high-altitude hypoxia; these observations may explain our findings of several PPARA selection signals in both the iHS and XP-EHH tests. Our finding of PPARA gene selection has relevance for cardio-metabolic diseases because genetic variants in PPARA have been found to be associated with blood lipid levels, lipoproteins, and type 2 diabetes.$^{58}$ Reduction in fatty acid oxidation through PPARA inhibition results in increased levels of stored and circulating lipids, a known risk factor for cardiovascular diseases and the metabolic syndrome.$^{59}$ This advantageous genetic adaptation of the past may fuel the rise in cardiovascular diseases in Ethiopians.$^{60,61}$ and perhaps other populations undergoing urbanization and transition toward lipid-rich diets and sedentary lifestyles. The cardio-metabolic effect of this genetic adaptation, which may have important clinical and public health implications, needs to be investigated further in other African and global populations.

We found selection signals in WETH (an Omotic language speaking ethnic group) around the SLC24A5 gene implicated in skin pigmentation in European and West Asian populations.$^{1,15}$ A recent study
found these loci in the top 5% of selection signals for Semitic-Cushitic-speaking Ethiopian populations, but not in a combined analysis of three Omotic language speaking Ethiopian ethnic groups (Wolaita, Ari Agricultural, and Ari Blacksmith). Moreover, we found a higher frequency of the alleles associated with light skin pigmentation in Eurasian populations in WETH (A allele frequency of both SNPs = 47.9% in WETH vs 23% in the combined Omotic sample reported25). Our study had a larger sample of the Wolaita than the previous study (n = 120 vs n = 8). Moreover, the genetic, geo-climatic, and demographic differences between the Wolaita and the Ari may have masked the selection signal during combined analysis of the ethnic samples in the previous study.62 The SLC24A5 gene’s derived Alu111Thr allele (rs1426564 A allele) is present at low frequencies in other sub-Saharan African populations.63 Our finding of high frequency of this allele in an Omotic-speaking indigenous southern Ethiopian ethnic population that has little shared genetic ancestry with Eurasians62 strengthens previous suggestions of the need for studies to understand whether the derived alleles underlying the adaptive response originated in East Africa or Eurasia.63

We replicated several genes reported to be under selection by at least two genome-wide scans (Supplementary Table 14). Consistent with a study which demonstrated that inflammatory-disease susceptibility loci are enriched for signatures of recent positive selection,64 we identified selection signals in loci implicated in podocnosis. The findings suggest that positive selection has favored the haplotypes carrying the podocnosis-protective alleles. The presence of more risk alleles in podocnosis cases may be due to the effect of mate selection because individuals from podocnosis affected families are subject to social stigmatization and are excluded from marriage by non-affected community members who recognize that podocnosis is heritable.65-67

Overall, this study provides strong evidence of selection in the HLA locus and metabolism genes in an Ethiopian population. It is likely that the burden of a diverse array of pathogens and adaptations to dietary fluctuations may represent the strongest selective forces in the history of this African population. Furthermore, our findings of overlaps between several previously reported disease susceptibility GWAS loci and targets of recent positive selection demonstrate the usefulness of African population samples to elucidate the adaptive genetic basis for many complex diseases.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
The research project was supported by the Wellcome Trust (grant #079791), and the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health, in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is also supported by the National Institute of Diabetes and Digestive and Kidney Diseases. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank staff of the Mossey Foot Treatment and Prevention Association (now Mossey Foot International) for coordinating the fieldwork.

4 Hamblin MT, Di Rienzo A: Detection of natural selection in human populations: evidence from the Duffy blood group locus. Am J Hum Genet 2000; 66: 1669–1679.
5 Tishkoff SA, Reed FA, Ranciaro A et al: Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 2007; 39: 31–40.
6 Ayodo G, Price AL, Keinan A et al: Combining evidence of natural selection with association analysis increases power to detect malaria-resistance variants. Am J Hum Genet 2007; 81: 234–242.
7 Grossman SR, Andersen KG, Shlyakhter I et al: Identifying recent adaptations in large-scale genomic data. Cell 2013; 150: 703–713.
8 Karlsson EK, Kwiatkowski DP, Sabeti PC: Natural selection and infectious disease in human populations. Nat Rev Genet 2014; 15: 379–393.
9 Akey JM: Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res 2009; 19: 711–722.
10 Campbell MC, Tishkoff SA: African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet 2008; 9: 403–433.
11 Tekola-Ayele F, Adeyemo A, Aseeda A et al: Clinical and pharmacogenomic implications of genetic variation in a Southern Ethiopian population. Pharmacogenomics J 2015; 15: 101–108.
12 Tekola Ayele F, Adeyemo A, Finan C et al: HLA class II locus and susceptibility to podocnosis. N Engl J Med 2012; 366: 1200–1206.
13 Altshuler DM, Gibbs RA, Peltonen L et al: Integrating common and rare genetic variation in diverse human populations. Nature 2010; 467: 52–58.
14 Voight BF, Kudaravalli S, Wen X, Pritchard JK: A map of recent positive selection in the human genome. PLoS Biol 2006; 4: e72.
15 Sabeti PC, Varilly P, Fry B et al: Genomic-scale wide detection and characterization of positive selection in human populations. Nature 2007; 449: 913–918.
16 Schetter P, Stephens M: A fast and flexible statistical model for large-scale population-genome data: applications to inferring missing genotypes and haplotype phase. Am Hum Genet 2007; 81: 629–644.
17 Andres AM, Clark AG, Shimmin L, Boerwinkle E, Sing CF, Hixson JF: Understanding the accuracy of statistical haplotype inference with sequence data of known phase. Genet Epidemiol 2007; 31: 696–707.
18 Reynolds J, Weir BS, Cockerham CC: Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 1983; 105: 767–779.
19 Wagh K, Bhalia A, Alese G et al: Lactase persistence and lipid pathway selection in the Maasai. PLoS One 2012; 7: e44751.
20 Mi H, Muruganujan A, Thomas PD: PANTHER in 2013: modeling the evolution of gene function. Nucleic Acids Res 2013; 41: D377–438.
21 Scheinfeldt LB, Soi S, Thompson S et al: Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol 2012; 13: R1.
22 Liu X, Ong RT, Pillai EN et al: Detecting and characterizing genetic signatures of positive selection in global populations. Am J Hum Genet 2013; 92: 866–881.
23 Ong RT, Liu X, Poh WT, Sim X, Chia KS, Teo YY: A method for identifying haplotypes carrying the causative allele in positive natural selection and genome-wide association studies. Bioinformatics 2011; 27: 822–828.
24 Morris AP, Voight BF, Tesloich TM et al: Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012; 44: 981–990.
25 Saxena R, Elbers CC, Guo Y et al: Large-scale gene-centric meta-analysis across 39 type 2 diabetes loci identifies type 2 diabetes. Nat Genet 2013; 45: 410–425.
26 Barreiro LB, Lawal G, Quach H, Patin E, Quintana-Murci L: Natural selection has driven population differentiation in modern humans. Nat Genet 2008; 40: 340–345.
27 Jarvis JP, Scheinfeldt LB, Soi S et al: Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. PLoS Genet 2012; 8: e1002641.
28 Williamson SH, Hubisz MJ, Clark AG, Payser BA, Bustamante CD, Nielsen R: Localizing recent adaptive evolution in the human genome. PLoS Genet 2007; 3: e90.
29 Cao K, Moormann AM, Lyke KE et al: Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. Tissue Antigens 2004; 63: 293–325.
30 Bhalia G, Patterson N, Pasanovic B et al: Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. Am J Hum Genet 2011; 89: 368–381.
31 Barreiro LB, Quintana-Murci L: From evolutionary genetics to human immunology: how selection shapes host defense genes. Am J Hum Genet 2010; 87: 17–30.
32 Bubb KL, Bowe D, Buckley D et al: Scan of human genome reveals no new Loci under ancient balancing selection. Genetics 2006; 173: 2165–2177.
33 DiGiorgio M, Lohmueller KE, Nielsen R: A model-based approach for identifying signals of ancient balancing selection in genetic data. PLoS Genet 2014; 10: e1004561.
34 Solberg OD, Mack SJ, Lancaster AK et al: Balancing selection and heterogeneity across the classical human leucocyte antigen loci: a meta-analytic review of 497 population scans. Hum Immunol 2008; 69: 464–466.
35 Andres AM, Hubisz MJ, Indap A et al: Targets of balancing selection in the human genome. Mol Biol Evol 2009; 26: 2755–2764.
36 Prugnolle F, Manica A, Charpentier M, Guegan JF, Guerrier V, Bailly F: Pathogen contamination and worldwide HLA class I diversity. Curr Biol 2005; 15: 1022–1027.
37 Sanchez-Mazas A, Lemaître JF, Curnut M: Distinct evolutionary strategies of human leucocyte antigen loci in pathogen-rich environments. Philos Trans R Soc Lond B Biol Sci 2012; 367: 830–839.
38 Dunn RR, Davies TJ, Harris NC, Gavin MC: Global drivers of human pathogen richness and prevalence. Proc Biol Sci 2010; 277: 2587–2595.

39 Henze PB: Layers of Time: A History of Ethiopia. London: Palgrave Macmillan, 2004.

40 Okada Y, Kubo M, Ohmiya H et al: Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. Nat Genet 2012; 44: 302–306.

41 Kwak SH, Choi SH, Jung HS et al: Clinical and genetic risk factors for type 2 diabetes at early or late post partum after gestational diabetes mellitus. J Clin Endocrinol Metab 2013; 98: E744–E752.

42 Ng MC, Saxena R, Li J et al: Transferability and fine mapping of type 2 diabetes loci in African Americans: the Candidate Gene Association Resource Plus Study. Diabetes 2013; 62: 965–976.

43 Pascoe L, Tura A, Patel SK et al: Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. Diabetes 2007; 56: 3101–3104.

44 Ubeda M, Rakosilis JM, Habener JF: Inhibition of cyclin-dependent kinase 5 activity protects pancreatic beta cells from glucotoxicity. J Biol Chem 2006; 281: 28858–28864.

45 Wei FY, Nagashima K, Onshima H et al: Cdk5-dependent regulation of glucose-stimulated insulin secretion. Nat Med 2005; 11: 1104–1108.

46 Steinhordt D, Thorleifsson G, Brennström R et al: A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 2007; 39: 770–775.

47 Prentice AM, Hennig BJ, Fulford AJ: Evolutionary origins of the obesity epidemic: natural selection of thrifty genes or genetic drift following predation release? Int J Obes (Lond) 2008; 32: 1607–1610.

48 Nakayama K, Ogawa A, Miyashita H et al: Positive natural selection of TRIB2, a novel gene that influences visceral fat accumulation, in East Asia. Hum Genet 2013; 132: 201–217.

49 Lefebvre P, Chinet G, Fruchart JC, Staels B: Sorting out the roles of PPAR alpha in energy metabolism and vascular homeostasis. J Clin Invest 2006; 116: 571–580.

50 Kersten S, Seydoux J, Petres JM, Gonzalez FJ, Desvergne B, Wahli W: Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting. J Clin Invest 1999; 103: 1489–1498.

51 Hashimoto T, Cook WS, Qi C, Yeldandi AV, Reddy JK, Rao MS: Defect in peroxisome proliferator-activated receptor alpha inducible fatty acid oxidation determines the severity of hepatic steatosis in response to fasting. J Biol Chem 2000; 275: 28918–28928.

52 Leone TC, Weinheimer CJ, Kelly DP: A critical role for the peroxisome proliferator-activated receptor alpha (PPARalpha) in the cellular fasting response: the PPARalpha-null mouse as a model of fatty acid oxidation disorders. Proc Natl Acad Sci USA 1999; 96: 7473–7478.

53 Reddy JK, Hashimoto T: Peroxisomal beta-oxidation and peroxisome proliferator-activated receptor alpha: an adaptive metabolic system. Annu Rev Nutr 2001; 21: 193–230.

54 Brandt SA, Spring A, Hiebsch C et al: The "Tree Against Hunger". Enset-based Agricultural Systems in Ethiopia. Washington, DC: American Association for the Advancement of Science, 1997.

55 Ehtet C: On the antiquity of agriculture in Ethiopia. J Afr Hist 1979; 20: 161–177.

56 Ge RL, Simonson TS, Cooksey RC et al: Metabolic insight into mechanisms of high-altitude adaptation in Tibetans. Mol Genet Metab 2012; 106: 244–247.

57 Simonson TS, Yang Y, Huff CD et al: Genetic evidence for high-altitude adaptation in Tibet. Science 2010; 329: 72–75.

58 Shin MJ, Kanaya AM, Krauss RM: Polymorphisms in the peroxisome proliferator activated receptor alpha gene are associated with levels of apolipoprotein CIII and triglyceride in African-Americans but not Caucasians. Atherosclerosis 2008; 198: 313–319.

59 Kimura R, Takahashi N, Murota K et al: Activation of peroxisome proliferator-activated receptor-alpha (PPARalpha) suppresses postprandial lipemia through fatty acid oxidation in enterocytes. Biochem Biophys Res Commun 2011; 410: 1–6.

60 Tesfaye F: Epidemiology of Cardiovascular Disease Risk Factors in Ethiopia: The rural–urban gradient; PhD thesis Umeå: Umeå University, 2008.

61 Misingana A, Mariam DH, Ayata Y, Ayete K: Patterns of mortality in public and private hospitals of Addis Ababa, Ethiopia. BMC Public Health 2012; 12: 1007.

62 Pagani L, Kivisild T, Tarekweg A et al: Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet 2012; 91: 83–96.

63 Bezea S, Santos AM, McEvoy B et al: The timing of pigmentation lightening in Europeans. Mol Biol Evol 2013; 30: 24–35.

64 Raj T, Kuchroo M, Replige JM, Raychaudhuri S, Stranger BE, De Jager PL: Common risk alleles for inflammatory diseases are targets of recent positive selection. Am J Hum Genet 2013; 92: 517–525.

65 Yakob B, Deribe K, Davey G: High levels of misconceptions and stigma in a community highly endemic for poxvirus infection in southern Ethiopia. Trans R Soc Trop Med Hyg 2008; 102: 429–444.

66 Tora A, Davey G, Tadele G: A qualitative study on stigma and coping strategies of patients with poxvirus infection in Wolaita zone, Southern Ethiopia. Int Health 2011; 3: 176–181.

67 Tekola F, Bull S, Farsides B et al: Impact of social stigma on the process of obtaining informed consent for genetic research on poxvirus infection: a qualitative study. BMC Med Ethics 2009; 10: 13.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)