Annals of human genetics

doi: 10.1111/j.1469-1809.2011.00695.x

Genome-Wide Patterns of Genetic Distances Reveal Candidate Loci Contributing to Human Population-Specific Traits

João Pedro de Magalhães[1]* and Alex Matsuda[2]†

[1]*Integrative Genomics of Ageing Group, Institute of Integrative Biology, University of Liverpool, Liverpool, UK*
[2]*Escola Superior de Biotecnologia, Porto, Portugal*

Summary

Modern humans originated in Africa before migrating across the world with founder effects and adaptations to new environments contributing to their present phenotypic diversity. Determining the genetic basis of differences between populations may provide clues about our evolutionary history and may have clinical implications. Herein, we develop a method to detect genes and biological processes in which populations most differ by calculating the genetic distance between modern populations and a hypothetical ancestral population. We apply our method to large-scale single nucleotide polymorphism (SNP) data from human populations of African, European and Asian origin. As expected, ancestral alleles were more conserved in the African populations and we found evidence of high divergence in genes previously suggested as targets of selection related to skin pigmentation, immune response, senses and dietary adaptations. Our genome-wide scan also reveals novel candidates for contributing to population-specific traits. These include genes related to neuronal development and behavior that may have been influenced by cultural processes. Moreover, in the African populations, we found a high divergence in genes related to UV protection and to the male reproductive system. Taken together, these results confirm and expand previous findings, providing new clues about the evolution and genetics of human phenotypic diversity.

Keywords: Genetic variation, genomics, human evolution, Out-of-Africa hypothesis, population genetics, selection

Introduction

The evolution of human populations is a fascinating but controversial topic. Arguably the only point of consensus is the Out-of-Africa hypothesis, the idea that modern humans originated in Africa and then migrated into the rest of the world (Stringer & McKie, 1996; Tishkoff & Kidd, 2004; Handley et al., 2007). Several studies using mitochondrial DNA or polymorphic *Alu* insertions support the Out-of-Africa hypothesis and suggest bottlenecks in the exodus of modern humans from Africa (Cann et al., 1987; Batzer et al., 1994; Watkins et al., 2001; Watkins et al., 2003). Population bottlenecks during recent human evolution caused founder effects which contributed to genetic and phenotypic differences between populations (Tishkoff & Kidd, 2004; Ramachandran et al., 2005). Moreover, as human populations adapted to new environments around the world, including new climates, pathogens, and food sources, they became subject to unique evolutionary pressures (Tang et al., 2007; Barreiro et al., 2008). Cultural and even social differences may have also driven the evolution of population-specific traits (Laland et al., 2010). The evolutionary history of each population thus shapes its genetics and a combination of factors, including chance and natural selection, contributed to the present phenotypic diversity of human populations.

There is great interest in determining the genetic basis for differences between populations. They may provide clues about the evolutionary history of our species. Moreover, in biomedical research and from a clinical perspective, knowledge of genetic differences between populations are relevant as they could aid in the diagnosis and treatment of several conditions, particularly now in the era of personalized medicine (Burchard et al., 2003; Tishkoff & Kidd, 2004).

†Present address: Departments of Anesthesia and Medicine, and Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

*Corresponding author: João Pedro de Magalhães, Biosciences Building, Crown Street, Liverpool L69 7ZB, UK. Tel: +44 151 7954517; Fax: +44 151 7954408; E-mail: jp@senescence.info

Recent large-scale projects have provided data on millions of genetic variants–in particular single nucleotide polymorphisms (SNPs)–from human populations originating in different geographical locations. A number of methods have been developed to detect positive selection in the human genome and applied to large-scale SNP data (Akey et al., 2002; Weir et al., 2005; Voight et al., 2006; Myles et al., 2007; Sabeti et al., 2007; Tang et al., 2007; Williamson et al., 2007; Barreiro et al., 2008; Myles et al., 2008; Johansson & Gyllensten, 2008; Coop et al., 2009; Pickrell et al., 2009; Chen et al., 2010). These and other previous studies have identified a number of candidate loci under selection in human populations, some of which have been shown to be phenotypically relevant. Briefly, one of the best studied loci is the lactase gene (*LCT*). The ability to digest lactose in adulthood is higher in Northern European populations, which has been suggested as an adaptation accompanying dairy farming. This hypothesis is supported by evidence of recent positive selection in *LCT* in Europeans (Bersaglieri et al., 2004). Skin pigmentation is another well-studied adaptation with several genes involved in skin pigmentation–such as *SLC45A2* and *SLC24A5*–exhibiting evidence of recent positive selection in Europeans (Graf et al., 2005; Lamason et al., 2005; Myles et al., 2007). Finally, the Duffy blood group locus (*DARC*), a membrane protein in which a mutation confers resistant to vivax malaria, has long been considered a likely target of natural selection in African populations with high incidence of vivax malaria (Hamblin & Di Rienzo, 2000). Therefore, to name just three well-studied examples, lactose tolerance, skin pigmentation, and resistance to vivax malaria are three key phenotypic differences between populations likely driven by adaptations of human populations to new environments.

Studies focusing on population differences, such as those using the fixation index (F_{st}) as a measure of population differentiation and methods employing haplotype structure like extended haplotype homozygosity, by and large focus on detecting signatures of natural selection. While natural selection played a major role in determining differences between modern human populations, however, chance and demographic effects also played a crucial role. Indeed, random genetic drift and founder effects contributed to population differences that may have implications in medical genetics (Tishkoff & Kidd, 2004; Hartl & Clark, 2007). In this work, we wanted to take advantage of large-scale SNP datasets to determine candidate loci of population-specific traits without assumptions of whether genes were targets of selection or merely diverged by chance between populations. Therefore, we develop a simple method to infer population-specific genetic divergence by calculating the genetic distance between modern populations and a hypothetical ancestral population.

We apply our model-free method to large-scale SNP data from human populations of African, European, and Asian geographic origins. Initially we employ data obtained by Perlegen Sciences (Hinds et al., 2005) and later validate these results using data from the International HapMap Project (Frazer et al., 2007). Our genome-wide scan allowed us to identify outliers, genes and biological functions in which human populations most diverge from the ancestral state, which include several previously known candidates of recent positive selection as well as loci not previously identified as targets of selection but that may be related to population-specific phenotypes and disease susceptibility.

Materials and Methods

Data Sources and Selection

SNP data from the Perlegen and HapMap datasets were used in this study. The Perlegen dataset consists of genome-wide SNPs from three different populations of three different geographic regions: 23 African Americans (Afr-Am), 24 European Americans (Eur-Am), and 24 Han Chinese (Han-Ch) (Hinds et al., 2005). Only class A Perlegen data were used. This consists of SNPs randomly identified by Perlegen Sciences and thus minimizes any ascertainment bias. HapMap phase II data were also obtained from populations of African, European, and Asian origins: 90 individuals of European ancestry living in Utah, USA (CEU), 90 individuals from the Yoruba in Ibadan, Nigeria (YRI), 45 individuals from Beijing, China (CHB), and 45 individuals from Tokyo, Japan (JPT). Data from Chinese and Japanese populations were combined to derive a single Asian population (ASN) with the frequencies of each SNP calculated from the combined population.

SNPs, allele frequencies, and reference human and ancestral alleles were downloaded from SPSmart (Amigo et al., 2008) which in turn aggregates data from other sources such as dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/). A hypothetical homozygous ancestral population was constructed from the ancestral state of SNPs, which come from dbSNP (Amigo et al., 2008) and are primarily inferred from the chimpanzee genome. SNPs for which no ancestral allele was available were not considered, though this occurred for only ~5% of SNPs. Only autosomal chromosomes were analyzed.

Overall, 1,175,330 SNPs from Perlegen class A were used in this work. Of these, 873,180 were found in HapMap phase II and were employed for the complementary analyses using the HapMap dataset.

Chromosomal locations and gene symbols were obtained from Ensembl (http://www.ensembl.org/).

Estimating Genetic Distances

Genetic distances were calculated using the average genetic distance (D_A), as described (Nei, 1987; Hughes et al., 2008). There is a vast literature on methods to calculate genetic distances between populations, though previous results have suggested strong correlations between different measures of genetic distances in large-scale SNP analyses for human populations (Akey et al., 2002). D_A was chosen because it is a simple and intuitive measure that can be easily applied with our hypothetical ancestral population, as described below.

For a given biallelic SNP locus the genetic distance (d) is given by

$$d = 1 - (\sqrt{x_1 y_1} + \sqrt{x_2 y_2}),$$

where x_1 and y_1 are the frequencies of the first allele and x_2 and y_2 are the frequencies of the second allele in a modern population (x_1 and x_2) and in a hypothetical ancestral population (y_1 and y_2). Because allele frequencies in the hypothetical ancestral population are always assumed to be 1 for the ancestral allele and 0 for the derived allele, d is effectively given by

$$d = 1 - (\sqrt{x}),$$

where x is the frequency of the ancestral allele in the modern population and thus d will have a value between 0 and 1.

Average d values of SNPs in a gene give the average genetic distance (D_A). Because the distribution of SNPs can vary between regions of the genome and nearby SNPs will have stronger linkage disequilibrium (LD), considering each SNP as an independent observation would be incorrect. Therefore, D_A was first determined for blocks of 5 kb and then the D_A of the gene was calculated from the average D_A of all its blocks. The 5 kb window size was chosen because it was deemed to have an appropriate balance between number of SNPs and strong LD. Although the criteria is admittedly arbitrary, SNPs within 5 kb of each other are in strong LD (Shifman et al., 2003; Hinds et al., 2005) and blocks of 5 kb have on average ~3 (HapMap), ~3.5 (Perlegen) SNPs per block. That said, the results using a window size of 2.5 kb or 10 kb were very consistent with those obtained with 5 kb (not shown).

Because the goal of this project was to detect functions and processes associated with population differences, a gene-centric approach was employed in which D_A was calculated for each gene using SNPs within the gene's coding region plus SNPs 50 kb from the coding region on both sides. SNPs within 50 kb of a gene will tend to be under moderate LD (Shifman et al., 2003; Hinds et al., 2005).

As in other similar studies (Voight et al., 2006; Pickrell et al., 2009), only common SNPs were considered as these are more likely to differentiate populations. This meant only SNPs with a minor allele frequency >5% in at least one of the three human populations were used (Frazer et al., 2007). SNPs in which the minor alleles were different between two of the populations were also kept in the analysis, however.

Statistical Tests

Genes that have significantly higher D_A values than the average would be candidates for specifying population differences (Fig. 1). To identify such genes in each of the modern human populations, the cumulative Poisson distribution was employed. First, the average D_A for all blocks in all genes between the ancestral population and Afr-Am, Eur-Am, and Han-Ch was calculated to be, respectively, 0.190, 0.230, and 0.236. For the HapMap dataset, the average D_A between the ancestral population and YRI, CEU, and ASN was, respectively, 0.184, 0.226, and 0.230. Based on genome-wide averages, the cumulative Poisson distribution was then used to calculate for each gene the probability that it has a higher D_A in a given modern–ancestral population pair than would be expected by chance, after correcting for the number of blocks (n_{blocks}) in the gene:

$$P(x \geq k) = \sum_{j=k}^{k_{\max}} \frac{e^{-\lambda}\lambda^j}{j!},$$

where $k = \frac{D_A n_{\text{blocks}}}{\overline{D}_A}$, which represents the D_A of the gene in the population pair over the average D_A for all genes ($\overline{D}_A$) multiplied by n_{blocks}; λ is similarly estimated from the average D_A of the two other modern populations for the gene over the average for all genes multiplied by n_{blocks} while $k_{\max} = \frac{n_{\text{blocks}}}{\overline{D}_A}$ (i.e., when $D_A = 1$).

In effect, the k metric is equivalent to the number of derived alleles that are present in a given modern human population in a given gene corrected for the average across the genome. The P-value derived from the Poisson function represents the probability that a given gene in a population has a higher representation of derived alleles than of ancestral alleles when compared to other populations. It does not allow, however, a direct rejection of a null hypothesis. Instead, it is a metric designed to rank genes and detect the most extreme outliers that are potentially involved in population-specific phenotypes. When calculating P-values for Eurasians, D_A is simply the average between Eur-Am and Han-Ch, while λ is estimated based only on the Afr-Am population. Overall, we calculated D_A values for 17,344 genes, which had at least three blocks, using the Perlegen data and 17,038 genes using HapMap.

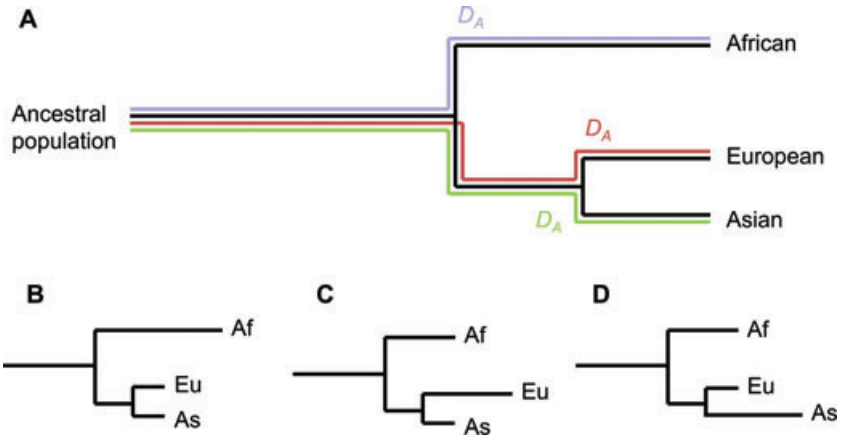

Figure 1 Schema of our method to detect population-specific genetic divergence. The genetic distance (D_A) between a hypothetical ancestral population and modern human populations–in the case of this work of African, European, and Asian origins–is calculated from the genome-wide SNP data (A). D_A is then calculated for individual genes. Genes for which D_A is larger in African (B), European (C), or Asian (D) populations than expected by chance are detected as population-specific genetic divergence and are thus candidates for specifying phenotypic differences in that population.

To identify candidate pathways, processes, and functions in each population, Gene Ontology (GO) annotation, which describes how gene products behave in a cellular context (Ashburner et al., 2000), was employed. The cumulative Poisson distribution described earlier was employed to derive *P*-values for GO categories. For a given GO category, D_A and n_{blocks} were calculated from the sum of D_A and n_{blocks} from all genes associated with that category. In other words, each GO category can be viewed as the concatenated sequence of all genes associated with it, as in other studies (de Magalhaes & Church, 2007). A complementary functional enrichment analysis of genes at the $P < 0.05$ cutoff was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID, Huang da et al., 2009).

A false discovery rate (FDR) analysis was performed to estimate the number of both genes and GO categories expected to be significant by chance. Simulations–using the Fisher-Yates shuffle algorithm to randomly permute SNP allelic frequencies of each block–were performed to determine, by chance, how many genes and GO categories below a given *P*-value threshold one is expected to find. Based on these simulations, an FDR-adjusted *P*-value (*q*-value), defined as the number of expected false positives over the number of significant results (Storey & Tibshirani, 2003), was estimated for each gene and each GO category.

The algorithm was implemented in the Perl language. Statistical tests were performed in Microsoft Excel 2007 and in SPSS 17 (SPSS Inc., Chicago, IL, USA).

Results

Algorithm to Detect Population-Specific Genetic Differentiation

The aim of this work is to identify genes and processes that are candidates for explaining phenotypic variation between populations. To detect such genes, the underlying principle behind our method is that loci that contribute to population-specific traits will have genetic patterns that are more distinct from states found in ancestral populations. Therefore, we first calculated for all genes the average genetic distance (D_A) between each modern human population studied in this work and a hypothetical ancestral population (Fig. 1). We first employed SNP data from Perlegen Sciences genotyped from Afr-Am, Eur-Am, and Han-Ch individuals. Since our goal was to identify functions and processes that differ between populations, only SNPs within 50 kb of coding regions were included in our calculations of D_A. Ancestral alleles were used to derive a hypothetical homozygous ancestral population (see Materials and Methods).

Our conceptually simple method is based on detecting genes that have a larger D_A in a given population than expected by chance (Fig. 1). Therefore, to calibrate our measurements, we determined the genome-wide genetic distances between the modern Perlegen populations and our hypothetical ancestral population. As expected, our results show that D_A between the ancestral population and Afr-Am is shorter

Table 1 Selection of the most highly divergent genes in each population.

Gene[1]	Chr	Start	End	D_A Anc-Afr[2]	D_A Anc-Eur[3]	D_A Anc-Han[4]	*P*-value
Genes divergent in Eur-Am							
KCNH7	2	162886163	163453274	0.195	0.362	0.257	0.000
MYEF2	**15**	**46168921**	**46307850**	**0.207**	**0.490**	**0.208**	**0.000**
ZMYM4	**1**	**35457155**	**35710131**	**0.187**	**0.340**	**0.165**	**0.002**
SLC24A5	**15**	**46150461**	**46271881**	**0.223**	**0.410**	**0.201**	**0.003**
MYST4	10	76205346	76512645	0.144	0.304	0.213	0.004
ATP6V1H	8	54740669	54968403	0.208	0.435	0.283	0.007
PAWR	12	78459878	78658921	0.172	0.355	0.227	0.007
BMP2K	4	79866556	80102363	0.241	0.379	0.247	0.008
LCT	**2**	**136211885**	**136361220**	**0.118**	**0.267**	**0.151**	**0.008**
Genes divergent in Han-Ch							
EXOC6B	2	72206621	72956685	0.181	0.272	0.412	0.000
POLR3B	12	105225619	105478104	0.188	0.212	0.430	0.000
C6orf173	6	126652946	126761447	0.143	0.181	0.642	0.000
HERC1	**15**	**61637871**	**61963200**	**0.193**	**0.220**	**0.399**	**0.000**
THADA	2	43261479	43726689	0.212	0.279	0.383	0.001
RTTN	**18**	**65772025**	**66073942**	**0.184**	**0.319**	**0.410**	**0.002**
XKR6	**8**	**10741075**	**11146258**	**0.166**	**0.200**	**0.313**	**0.003**
Genes divergent in Afr-Am							
TRIM10	6	30177705	30286690	0.408	0.244	0.266	0.003
SYT14	1	208128161	208454259	0.154	0.108	0.109	0.008
CADM3	1	157358040	157489556	0.199	0.136	0.119	0.010
SLC35D2	9	98072834	98235797	0.396	0.282	0.221	0.015
DARC	**1**	**157389721**	**157492914**	**0.209**	**0.126**	**0.141**	**0.017**
Genes divergent in Eurasians (but not highly divergent in Eur-Am or Han-Ch)							
CSPP1	8	68089157	68321048	0.274	0.664	0.779	0.001
RB1	13	47725884	48004027	0.132	0.319	0.264	0.001
KCNT2	1	194411536	194894122	0.168	0.301	0.324	0.001
EPHA6	3	97966311	99000235	0.178	0.319	0.299	0.001
ZEB1	10	31598148	31906740	0.171	0.355	0.342	0.001

Anc-Afr, Ancestral-African; Anc-Eur, Ancestral-European; Anc-Han, Ancestral-Chinese.

[1]Genes highlighted in bold have been previously shown to be under selection using other methods (see text for references).

[2]D_A between ancestral and African American population.

[3]D_A between ancestral and European American population.

[4]D_A between ancestral and Han Chinese population.

than for Eur-Am and Han-Ch, respectively 0.190, 0.230, and 0.236. The differences in D_A between all the populations were highly significant ($P < 0.001$ using a Wilcoxon signed-rank test). These results are in line with many previous findings showing that the genetic distance between Eurasian populations and Africans is larger than the differences between Eurasian populations (Hughes et al., 2008), in accordance with the African origin of modern humans. D_A was consistent between chromosomes (the relative standard deviation for the three populations was ~2%), and thus genome-wide D_A values were used as reference for detecting outliers.

Because the distribution of SNPs is not consistent across the genome, D_A was calculated from the average of the genetic distance of individual SNPs in blocks of 5 kb. As further detailed in the Methods, the D_A of a given gene was then calculated from the average D_A of its 5 kb blocks. To detect candidate genes in each population, we employed the cumulative Poisson distribution to calculate *P*-values and identify genes in which D_A is larger than expected by chance–after adjusting for the number of blocks in each gene. Expected D_A values are estimated from the other population pairs to account for variation across the genome (see Materials and Methods). The *P*-value derived from the Poisson distribution, however, does not allow a direct rejection of a null hypothesis but rather provides a metric to rank candidate genes.

Contrary to many previous similar works (in particular methods based on haplotype structure), our method is not

a test for positive selection. Phenotypically relevant differences between populations could be due to chance events like neutral drift and population bottlenecks that are not under selection. Our method thus aims to detect candidate loci for population-specific traits without any intrinsic assumption of their origin. As further discussed ahead, however, it is plausible that the most striking differences between populations found using our method were driven by selection and can thus be seen as candidate regions for recent positive selection.

Genes that Differ Most between Populations

The method described allowed us to detect genes in our three studied populations and in Eurasians (Eur-Am + Han-Ch) for which D_A is disproportionally higher in a given population when compared to the others, after correcting for the averages across the whole genome. At $P < 0.05$, we detected 74 genes in Afr-Am, 98 in Eur-Am, 147 in Han-Ch, and 218 genes in Eurasians (see Tables S1–S4). Simulations were performed to estimate that, by chance, 34 genes were expected in Afr-Am, Eur-Am, and Han-Ch, and 52 in Eurasians. Although it is encouraging to detect more genes than expected by chance, it is important to note that population demographic effects are not considered in our statistical test. Besides, since our method is not a test for selection it does not attempt to exclude random events. In fact, genes relevant to population-specific differences could occur purely by chance. This is why, as indicated, our *P*-value is a metric designed to detect and rank outliers, which can be considered candidates for population-specific differences.

As expected, a large number of the top genes detected using our method had been reported as candidates of selection by previous studies. *LCT*, for example, had a larger genetic distance in Europeans than in other populations (Fig. 2). Further examples of candidates of selection in Europeans that were rediscovered using our method include genes associated with skin pigmentation like *SLC45A2* and *SLC24A5* (Graf et al., 2005; Lamason et al., 2005; Barreiro et al., 2008), *RAB3GAP1* and *BCAS3* (Sabeti et al., 2007), *MYEF2* and *ARHGAP26* (Johansson & Gyllensten, 2008), and *ZMYM4* (Pickrell et al., 2009) (Tables 1 and S2). *DARC* was the only candidate gene under selection in Africans rediscovered using our method (Table 1). Many previous candidates of selection in Asian populations were rediscovered, however, including *XKR6* (Deng et al., 2008), *ABCC11*, *SYTL3*, and *RTTN* (Barreiro et al., 2008), *EDAR*, *HERC1*, and *SLC30A9* (Sabeti et al., 2007) (Tables 1 and S3). Interestingly, some genes previously identified as candidates in Europeans or Asians, such as *PPARD* (Voight et al., 2006), *DOCK4* (Johansson & Gyllensten, 2008), *TOP2B*, and *DNAH6* (Tang et al., 2007), were rediscovered in our study as significant in Eurasians (see Table S4). Among the top genes in Eurasians was also *C21orf34* ($P = 0.023$; Table S4) which has been previously associated with selection in populations outside of Africa (Pickrell et al., 2009). As far as we could tell, there were no contradictory findings of genes being detected by our method as candidates in one population having been previously reported to be under selection in another population included in our study. Overall, these results demonstrate the power of our method to detect biologically relevant results.

Although it is encouraging to verify that many of the genes detected by our method had been suggested as candidates of selection previously, these results are merely confirmatory. One major interest of this analysis is the identification of new genes and processes that may help explain population differences. Interestingly, several genes were identified by our method as candidates for explaining population differences for the first time (to our knowledge). These new candidate genes for specifying population-specific traits are described below.

Top genes in Europeans included *MYST4*, a histone acetyltransferase involved in gametogenesis (McGraw et al., 2007), a potassium voltage-gated channel (*KCNH7*) as well as poorly studied genes like *ATP6V1H* and *BMP2K* (Table 1). In Asians, the top genes were *EXOC6B*, an exocyst complex component, *POLR3B*, *C6orf173*, and *THADA*, a gene associated with thyroid adenoma (Rippe et al., 2003). In Afr-Am, we detected fewer genes and even top genes had higher *P*-values (Tables 1 and S1). A few genes in the TRIM cluster on chromosome 6 were detected, close to the major histocompatibility complex. Other genes detected in Afr-Am included *CADM3*, *SLC35D2*, and *SYT14*. Finally, a number of candidate genes not significant among Eur-Am or Han-Chn were detected among Eurasians (Table 1). Among the top genes were retinoblastoma 1 (*RB1*), *CSPP1*, and *KCNT2* which encodes a potassium channel.

A selection of the most significant genes for each population is shown in Table 1. Tables with significant genes, at the $P < 0.05$ threshold, for each population are available in Tables S1–S4 with the full dataset available online (http://genomics.senescence.info/evolution/human_populations.html).

If D_A values were to follow a neutral model then a similar distribution would be expected for genic (defined as SNPs within 50 kb of a coding region) and nongenic regions (defined as all SNPs not in genic regions), as argued by others (Voight et al., 2006; Barreiro et al., 2008). To test this assumption, we determined extreme D_A values between modern populations by calculating the difference between D_A values in two populations (e.g., D_A in Afr-Am – D_A in Eur-Am)

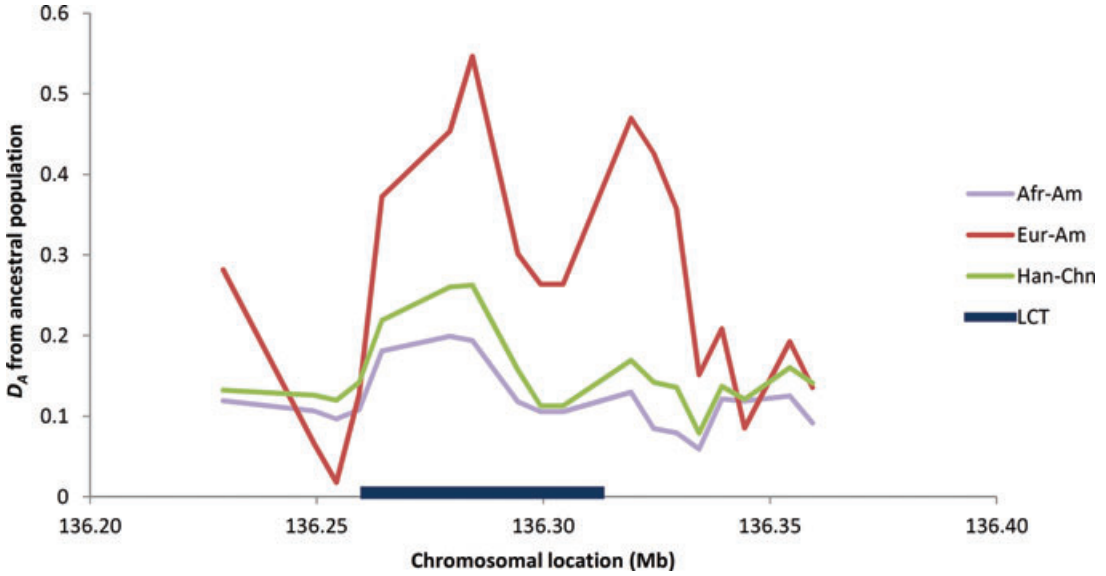

Figure 2 Genetic distance (D_A) in the lactase locus between a hypothetical ancestral population and African American (Afr-Am, in light blue), European American (Eur-Am, in red), and Han Chinese populations (Han-Ch, in green). A rolling average is used across blocks in the chromosomal region. The location of the lactase gene (*LCT*) is represented in dark blue.

Table 2 Selected top GO categories in Afr-Am ($q < 0.1$ and $n \geq 3$).

GO ID	Name	GO type[1]	*n* Genes	D_A Anc-Afr[2]	D_A Anc-Eur[3]	D_A Anc-Han[4]	*q*-Value
GO:0005882	Intermediate filament	C	98	0.209	0.241	0.239	<0.01
GO:0006636	Unsaturated fatty acid biosynthetic process	P	4	0.261	0.261	0.217	<0.01
GO:0005922	Connexon complex	C	17	0.185	0.195	0.197	0.025
GO:0008527	Taste receptor activity	F	10	0.249	0.269	0.244	0.047
GO:0016458	Gene silencing	P	3	0.228	0.204	0.215	0.068
GO:0042613	MHC class II protein complex	C	13	0.196	0.201	0.209	0.060
GO:0019902	Phosphatase binding	F	5	0.197	0.216	0.220	0.062
GO:0006816	Calcium ion transport	P	87	0.193	0.227	0.229	0.061
GO:0019717	Synaptosome	C	34	0.193	0.224	0.224	0.058
GO:0030539	Male genitalia development	P	8	0.229	0.240	0.225	0.083
GO:0007635	Chemosensory behavior	P	6	0.222	0.232	0.215	0.088
GO:0050909	Sensory perception of taste	P	21	0.222	0.246	0.239	0.082

Anc-Afr, Ancestral-African; Anc-Eur, Ancestral-European; Anc-Han, Ancestral-Chinese.
[1]C: Cellular component; F: Molecular function; P: Biological process.
[2]D_A between ancestral and African American population.
[3]D_A between ancestral and European American population.
[4]D_A between ancestral and Han Chinese population.

for all blocks in genic and nongenic regions. Most blocks had a D_A difference between population pairs in the range of 0.01–0.15 with <1% of extreme blocks with a difference above 0.5. Interestingly, however, the proportion of extreme blocks was considerably higher in genic than in nongenic regions (Fig. 3). These results show that far more extreme D_A values are observed in genic regions, suggesting that the patterns observed are not random and that at least some of the results observed in terms of differentiated genes between populations are due to selection.

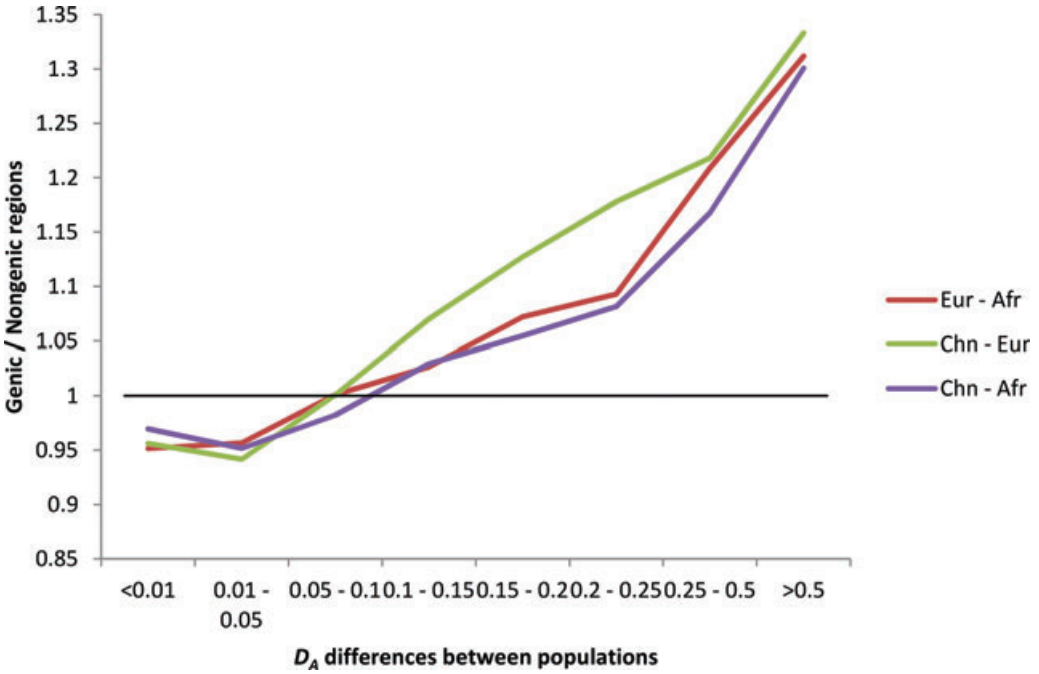

Figure 3 Comparison between the distribution of differences in genetic distance (D_A) between pairs of the studied modern populations in genic versus nongenic regions. The percentage of blocks with a D_A difference between two populations in genic regions is divided by the percentage in nongenic regions and plotted according to the D_A difference. Pairwise comparisons are shown between African American and European American (Eur-Afr, in red), Han Chinese and European American (Chn-Eur, in green), and Han Chinese populations and African American (Chn-Afr, in purple).

Candidate Processes and Functions Related to Population-Specific Phenotypes

In addition to identifying new candidate genes, the goal of this work is to identify processes, functions, and mechanisms that may be related to population-specific phenotypes. Therefore, we performed an analysis using GO annotation (Ashburner et al., 2000). As detailed in the Methods, D_A values were calculated from the concatenated sequences of genes associated with each GO category. Because this analysis looks for deviations from general patterns by combining values from different genes often located in different regions, significant GO categories are more likely explained by selection acting on specific processes than patterns in individual genes. Simulations were performed to estimate FDRs and P-values adjusted to what would be expected by chance (q-values, see Materials and Methods).

At $P < 0.05$, 21 categories were significant in Afr-Am, 14 in Eur-Am, and 32 in Han-Ch. A selection of significant categories that distanced themselves in each population are indicated in Tables 2–5 with all significant results available in Tables S5–S8. Similarly to the results observed for individual genes, several categories are related to those previously reported before as candidates for selection in human populations (Akey et al., 2002; Voight et al., 2006; Williamson et al., 2007; Barreiro et al., 2008; Pickrell et al., 2009). These include categories related to senses, like taste (GO:0008527), which is among the top categories in Afr-Am, or audition (GO:0042491) and olfaction (GO:0007608 and GO:0004984) in Han-Ch, as well as categories related to immune response (e.g., GO:0042613 in Afr-Am and GO:0009405 in Han-Ch). Categories possibly related to dietary adaptations, like unsaturated fatty acid biosynthetic process (GO:0006636 in Afr-Am), were also among top categories.

Among Afr-Am (Table 2), it is interesting to note male genitalia development (GO:0030539) as a candidate GO category. Another novel (to our knowledge) category was UV protection (GO:0009650), even though it was borderline significant with $q = 0.11$ (and $P = 0.049$; Table S5).

In Eur-Am, the top categories were related to chromatin and histone acetylation (GO:0016568 and GO:0016573), in line with previous results (Voight et al., 2006). Another top category was gamete generation (GO:0007276), though again selection in genes related to reproduction has been found in similar such studies (Akey et al., 2002; Voight et al., 2006). In

Table 3 Selected top GO categories in Eur-Am ($q < 0.1$ and $n \geq 3$).

GO ID	Name	GO type[1]	*n* Genes	D_A Anc-Afr[2]	D_A Anc-Eur[3]	D_A Anc-Han[4]	*q*-Value
GO:0016568	Chromatin modification	P	98	0.178	0.236	0.231	<0.01
GO:0007276	Gamete generation	P	6	0.186	0.315	0.261	<0.01
GO:0016573	Histone acetylation	P	7	0.181	0.266	0.223	<0.01
GO:0016407	Acetyltransferase activity	F	9	0.180	0.259	0.216	<0.01
GO:0008135	Translation factor activity, nucleic acid binding	F	4	0.219	0.365	0.297	<0.01
GO:0004707	MAP kinase activity	F	12	0.178	0.243	0.215	<0.01
GO:0043353	Enucleate erythrocyte differentiation	P	4	0.165	0.279	0.249	0.011
GO:0045651	Positive regulation of macrophage differentiation	P	4	0.150	0.275	0.246	0.012
GO:0007518	Myoblast cell fate determination	P	3	0.232	0.344	0.276	0.090

Anc-Afr, Ancestral-African; Anc-Eur, Ancestral-European; Anc-Han, Ancestral-Chinese.
[1]C: Cellular component; F: Molecular function; P: Biological process.
[2]D_A between ancestral and African American population.
[3]D_A between ancestral and European American population.
[4]D_A between ancestral and Han Chinese population.

this context, mating behavior (GO:0007617) also caught our attention ($P = 0.040$ and $q = 0.11$; Tables S6).

Categories related to exocytosis were the top categories in Han-Ch (GO:0000145 and GO:0006904), which is consistent with the genic analysis given that an exocyst component was the top gene in this population. Another top category was meiotic recombination (GO:0007131). Of interest was also the mitochondrial alpha-ketoglutarate dehydrogenase complex (GO:0005947) involved in oxidative stress and neurodegeneration (Gibson et al., 2005). As detailed in Table 4, other top categories included ubiquitin-protein ligase activity (GO:0004842) and behavior (GO:0007610).

In Eurasians, 56 categories passed $P < 0.05$. Top categories included fatty acid transport (GO:0015908) and caspase activity (GO:0030693), thus mostly recapitulating the categories for other populations and previous studies (Table 5).

Strikingly, categories related to nervous system development were top candidates in all three populations: synaptosome (GO:0019717) in Afr-Am, nerve growth factor receptor signaling pathway (GO:0048011), and generation of neurons (GO:0048699) in Han-Ch, and though slightly below our statistical threshold ($q = 0.12$) dendritic spine (GO:0043197) in Eur-Am.

We also performed a functional enrichment analysis among significant genes ($P < 0.05$) for each population using DAVID (Huang da et al., 2009). Although some functions and processes were found enriched, such as exocytosis in Han-Ch, the only significant category after FDR correction was zinc finger region among Afr-Am (Tables S17–S20). These results show that our GO enrichment methodology, that in essence gives a different weight to each gene contributing to the GO categories, outperforms more traditional approaches.

Comparison with Patterns from HapMap Data

Although encouraged by our findings, we were concerned about the population admixture in the African American group and whether this could have been a cause for the lower number of significant genes in this group. Besides, the particular history of African Americans (e.g., slavery) could have shaped the patterns observed in our results. Therefore, we repeated our analysis using HapMap phase II data, which like Perlegen includes populations from African (YRI), European (CEU), and Asian (ASN) origins but with the African population originating in Nigeria (Frazer et al., 2007). For this analysis we only considered HapMap SNPs in class A Perlegen data to minimize any ascertainment bias, thus the total number of SNPs was lower than when using the Perlegen data. Data for Chinese and Japanese populations was merged into a single Asian population (see Materials and Methods).

The results from HapMap were largely in accordance with the results obtained using the Perlegen dataset. Briefly, the D_A between the ancestral population and YRI was again shorter than for CEU and ASN, respectively 0.184, 0.226, and 0.230. We found slightly fewer statistically significant genes, however, for European and Asian populations than using the Perlegen dataset, respectively, 84 and 132 at $P < 0.05$ (Tables S10 and S11), and more were expected by chance

Table 4 Selected top GO categories in Han-Ch ($q < 0.1$ and $n \geq 3$).

GO ID	Name	GO type[1]	*n* Genes	D_A Anc-Afr[2]	D_A Anc-Eur[3]	D_A Anc-Han[4]	*q*-value
GO:0000145	Exocyst	C	5	0.190	0.247	0.314	< 0.01
GO:0006904	Vesicle docking during exocytosis	P	17	0.188	0.231	0.269	< 0.01
GO:0048011	Nerve growth factor receptor signaling pathway	P	5	0.202	0.256	0.328	< 0.01
GO:0042491	Auditory receptor cell differentiation	P	6	0.197	0.246	0.303	< 0.01
GO:0003682	Chromatin binding	F	69	0.189	0.235	0.254	< 0.01
GO:0007608	Sensory perception of smell	P	303	0.194	0.223	0.242	0.023
GO:0004842	Ubiquitin-protein ligase activity	F	112	0.186	0.234	0.247	0.023
GO:0008017	Microtubule binding	F	36	0.206	0.250	0.276	0.023
GO:0004984	Olfactory receptor activity	F	296	0.195	0.222	0.242	0.024
GO:0006512	Ubiquitin cycle	P	337	0.182	0.225	0.235	0.025
GO:0007131	Meiotic recombination	P	18	0.182	0.223	0.251	0.025
GO:0009405	Pathogenesis	P	6	0.220	0.297	0.348	0.027
GO:0048699	Generation of neurons	P	4	0.227	0.254	0.316	0.048
GO:0030246	Carbohydrate binding	F	10	0.170	0.212	0.249	0.051
GO:0007610	Behavior	P	15	0.189	0.251	0.273	0.063
GO:0000724	Double-strand break repair via homologous recombination	P	11	0.178	0.233	0.262	0.065
GO:0001942	hair follicle development	P	3	0.191	0.226	0.292	0.080
GO:0005947	Mitochondrial alpha-ketoglutarate dehydrogenase complex	C	4	0.174	0.240	0.280	0.083

Anc-Afr, Ancestral-African; Anc-Eur, Ancestral-European; Anc-Han, Ancestral-Chinese.
[1]C: Cellular component; F: Molecular function; P: Biological process.
[2]D_A between ancestral and African American population.
[3]D_A between ancestral and European American population.
[4]D_A between ancestral and Han Chinese population.

(65.6) than in the Perlegen analysis. Besides, the top genes for CEU and ASN were among the significant genes already identified in the Perlegen data for Eur-Am and Han-Ch. As for GO categories (Tables S14 and S15), the top categories largely recapitulated the results for Eur-Am and Han-Ch. Therefore, while it is encouraging that the results obtained from the Perlegen populations were corroborated in CEU and ASN, the analysis of HapMap did not substantially add to our previous findings. Our results for all genes, however, are available in the Supporting Information.

For the African population the results obtained using HapMap data were better than those using Perlegen. At $P < 0.05$, 170 genes were detected in YRI when 65.6 would be expected by chance. Even at $P < 0.01$, 29 genes were identified (2.7 would be expected by chance) which is a considerable improvement over the results obtained using Perlegen data. Of these 29 genes, 11 had been identified in the Perlegen analysis. Perhaps the admixture in the Perlegen Afr-Am population lowers significance levels and the number of significant genes. The top gene in YRI was *KIAA1267*, a poorly studied gene that is in the same region of the microtubule-associated protein tau (*MAPT*), also among our candidates. One gene re-identified was *RSBN1* since it was previously shown to have a signal for selection in YRI (Voight et al., 2006). Another gene that caught our attention was *STRBP*, the spermatid perinuclear RNA binding protein. This gene was also detected in the Perlegen analysis but with only $P = 0.042$ when in YRI the signal is much stronger ($P = 0.001$). Other genes detected with strong evidence for higher divergence in YRI include *CADM3*, *ANKRD17*, *COX18*, *GJB2*, and *POU1F1* (Tables 6 and S9).

A larger number of GO categories (64 when 3.3 would be expected by chance) were also detected in YRI (Table S13). The results were largely consistent, however, with the Perlegen analysis with by and large the same type of categories being detected. Categories that caught our attention in the Perlegen analysis were re-identified such as UV protection (GO:0009650) and male genitalia development

Table 5 Selected top GO categories in Eurasians ($q < 0.05$ and $n \geq 3$).

GO ID	Name	GO type[1]	*n* Genes	D_A Anc-Afr[2]	D_A Anc-Eur[3]	D_A Anc-Han[4]	*q*-Value
GO:0000145	Exocyst	C	5	0.1901	0.2465	0.3138	<0.01
GO:0007243	Protein kinase cascade	P	64	0.1754	0.2274	0.2354	<0.01
GO:0045651	Positive regulation of macrophage differentiation	P	4	0.1501	0.2747	0.2459	<0.01
GO:0015908	Fatty acid transport	P	4	0.1931	0.325	0.3358	<0.01
GO:0016568	Chromatin modification	P	98	0.1782	0.2357	0.2308	<0.01
GO:0043550	Regulation of lipid kinase activity	P	3	0.1557	0.2741	0.2531	<0.01
GO:0004842	Ubiquitin-protein ligase activity	F	112	0.1863	0.2343	0.2468	<0.01
GO:0006091	Generation of precursor metabolites and energy	P	69	0.1992	0.2574	0.2682	<0.01
GO:0043353	Enucleate erythrocyte differentiation	P	4	0.165	0.2787	0.2487	<0.01
GO:0006512	Ubiquitin cycle	P	337	0.182	0.2252	0.2349	<0.01
GO:0030693	Caspase activity	F	15	0.2022	0.285	0.2868	<0.01
GO:0030168	Platelet activation	P	16	0.1867	0.2531	0.2624	<0.01
GO:0048185	Activin binding	F	3	0.1565	0.2532	0.2577	<0.01
GO:0004128	Cytochrome-b5 reductase activity	F	5	0.1665	0.2549	0.2647	<0.01
GO:0001515	Opioid peptide activity	F	5	0.1774	0.279	0.2708	<0.01
GO:0015871	Choline transport	P	3	0.21	0.3027	0.3442	<0.01

Anc-Afr, Ancestral-African; Anc-Eur, Ancestral-European; Anc-Han, Ancestral-Chinese.
[1]C: Cellular component; F: Molecular function; P: Biological process.
[2]D_A between ancestral and African American population.
[3]D_A between ancestral and European American population.
[4]D_A between ancestral and Han Chinese population.

(GO:0030539) as well as categories related to senses, immune responses, dietary adaptations, and nervous system development (Table S13). In the latter category type, axon (GO:0030424) and neuron differentiation (GO:0030182) were identified, which had not been found in the Perlegen analysis. Other categories in YRI not identified before that might be of interest were hair follicle morphogenesis (GO:0031069), artery morphogenesis (GO:0048844), and blood vessel remodeling (GO:0001974).

Full lists with significant GO categories at the $P < 0.05$ threshold for all HapMap populations are available in Tables S13–S16.

Discussion

There is a vast literature on methods to detect signatures of selection in human populations (Akey et al., 2002; Weir et al., 2005; Voight et al., 2006; Sabeti et al., 2007; Tang et al., 2007; Williamson et al., 2007; Myles et al., 2008; Coop et al., 2009 Pickrell et al., 2009). Our method is not necessarily more powerful than other methods, such as those employing F_{st} or haplotype structure, to detect selection. The main novelty of our study is that we scan for genetic differences based not on patterns of selection but rather on the genetic distance between each population and an inferred ancestral population. Because of this broader focus on population differences our study provides a different approach. Although many of our results are confirmatory, our work also reveals new candidate regions. In fact, the large number of genes and GO categories that overlap with previous findings–including those few genes for which considerable evidence exists of a phenotypic role in population-specific traits like *LCT*, *DARC*, *SLC45A2*, and *SLC24A5*–demonstrate that our method can detect genes that are phenotypically relevant.

The shorter genetic distance between the African populations (Afr-Am and YRI) and the hypothetical ancestral population is consistent with previous analyses of *Alu* insertions suggesting that the ancestral state is more common in African populations (Watkins et al., 2001; Watkins et al., 2003). Our analysis showing the preservation of human ancestral alleles in an African population, however, is based on a much larger number of polymorphisms. It is consistent with the Out-of-Africa hypothesis since demographic effects like bottlenecks

Table 6 Selection of the most highly divergent genes in the African (Yoruba) population.

Gene[1]	Chr	Start	End	D_A Anc-YRI[2]	D_A Anc-CEU[3]	D_A Anc-ASN[4]	*P*-value
KIAA1267	17	41413181	41675943	0.496	0.345	0.542	0.001
CADM3	1	157358040	157489556	0.227	0.129	0.104	0.001
STRBP	9	124877139	125120718	0.292	0.166	0.219	0.001
SCMH1	1	41215461	41530375	0.332	0.228	0.245	0.001
C5orf34	5	43472567	43600944	0.330	0.144	0.029	0.001
ANKRD17	4	74109369	74393366	0.162	0.069	0.059	0.002
BTRC	10	103053815	103357060	0.343	0.210	0.335	0.002
DARC	**1**	**157389721**	**157492914**	**0.241**	**0.119**	**0.122**	**0.002**
SYT14	1	208128161	208454259	0.193	0.126	0.108	0.002
CEP57	11	95113290	95255502	0.215	0.139	0.138	0.003
COX18	4	74089280	74204336	0.183	0.047	0.032	0.003
MAPT	17	41277624	41511547	0.471	0.341	0.522	0.004
C9orf56	9	125017303	125119075	0.297	0.114	0.152	0.004
GJB2	13	19609605	19715114	0.289	0.139	0.152	0.005
RSBN1	**1**	**114055977**	**114206593**	**0.289**	**0.210**	**0.171**	**0.006**
POU1F1	3	87341473	87458427	0.304	0.220	0.235	0.010

Anc-YRI, Ancestral-African; Anc-CEU, Ancestral-European; Anc-ASN, Ancestral-Asian.
[1]Genes highlighted in bold have been previously shown to be under selection using other methods (see text for references).
[2]D_A between ancestral and African (Yoruba) population.
[3]D_A between ancestral and European population.
[4]D_A between ancestral and Asian population.

and adaptations to new environments likely drove genetic differentiation in Eurasians.

It is important to note that our method of calculating *P*-values is model-free and does not account for population demographic effects such as bottlenecks during the migration of populations out of Africa and into Eurasia. Considering the several known targets of selection detected using our method, it is plausible that many genes detected using our method could be driven by selection–and hence should be considered as candidates in future studies–but it is possible that genes are highly divergent in a given population due to demographic effects. Nonetheless, even for alleles that became more prevalent in a given population due to chance, it is possible that these are phenotypically relevant. Ample evidence suggests that founder effects contributed to population differences (Ramachandran et al., 2005). For example, population bottlenecks have been suggested to have an effect on human phenotypic variation based on analyses of cranial traits (Manica et al., 2007). Its focus on genetic differences that could have been driven by selection and by demographic effects is one of the distinct aspects of our method when compared to genome-wide scans for selection and is the reason why our algorithm is not intrinsically designed to detect selection.

One problem that may affect some of our candidate genes, as it has for similar studies, is genetic hitchhiking (Smith & Haigh, 1974). For example, among our top ranked genes in Eur-Am, *RAB3GAP1* is located within less than 150 kb of *LCT*. *LCT* has a lower *P*-value and it is possible that–even though *RAB3GAP1* has been identified as a candidate gene under selection by another study (Sabeti et al., 2007)–*RAB3GAP1* has a strong signal due to its proximity to *LCT*. Another caveat of our study is that our ancestral population is artificial, since an ancestral homozygous population never actually existed, and this might contribute to increase the power of our method but also to generate false positives. Besides, we only employed data from three populations, which may not be fully representative of their geographical origins. In particular for genes and GO categories with the lowest signals, our results may not apply to all human populations in Africa, Europe, or Asia. Nonetheless, the consistent results obtained by applying our method to both Perlegen and HapMap datasets, which were derived from different sets of populations, increase our confidence that our algorithm can provide new insights on the migration of human populations from Africa, one of the most important evolutionary events of our species. With the recent explosion in large-scale SNP data for human populations, in particular driven by advances in DNA sequencing (de Magalhaes et al., 2010), our method could be useful to a broad range of studies and ours can be seen as only a proof-of-concept study. Given our method's balance of simplicity and precision it may be useful for many other studies, including those in nonhuman species.

Despite the overlap with known results, it is noteworthy that our algorithm allowed us to detect several new genes and processes as candidates for explaining population differences. These could be targets for further study to better understand

human evolutionary history and adaptation and the many new candidate genes identified using our method could lead to new insights. In particular, there were several genes in each of the populations with strong signals (Tables 1 and 6). These include *KIAA1267*, *CADM3*, *STRBP*, *SCMH1*, *MAPT*, and *SYT14* in Africans and *MYST4*, *KCNH7*, and *BMP2K* in Europeans. *KIAA1267* and *MAPT* are in the same region and have both been associated with neurodegenerative diseases (Tobin et al., 2008), though whether this is related to our results is unclear. *STRBP* could be another case of a gene related to reproduction under selection since evidence from mice suggests this gene is involved in spermatid RNA metabolism (Schumacher et al., 1995). Although not previously suggested to be under selection in human populations, the gene encoding the poorly studied bone morphogenic protein *BMP2K* has been previously suggested to be under selection in hominids (Arbiza et al., 2006). Finally, it is curious to note that *MYST4* and *NCOA2*, also in our top results (Table S2), have both been known to form fusion proteins in acute myeloid leukemia (Murati et al., 2004), though whether this is related to their strong signals in our study is unknown.

Taken together, our results demonstrate that our method is statistically sound. Even though we did not simulate demographic effects because our goal is not to detect selection but rather phenotypic changes, we used an empirical distribution to demonstrate that the patterns observed are not random. Under neutrality, the percentage of blocks at varying levels of differentiation between population pairs would be expected to be the same for genic and nongenic regions since these are under the same demographic effects. Since selection preferentially targets genic regions (Voight et al., 2006; Barreiro et al., 2008), our results showing a higher percentage of extreme D_A values in genic regions suggest that selection played a role in our observed results (Fig. 3).

Even if the results from our study are statistically sound, it is intrinsically difficult to demonstrate that even the genes with the strongest signals are phenotypically relevant. Nonetheless, our work provides important leads for further research avenues. Multiple candidate genes emerged in Han-Ch (Tables 1 and S3). Examples include *EXOC6B*, *POLR3B*, *C6orf173*, and *THADA*. The strong signal in GO categories related to exocytosis together with the fact that a component of the exocyst (*EXOC6B*) was the gene with the lowest *P*-value makes a strong case for selection in this protein complex. We speculate that the role of exocytosis in the nervous system–and neuronal development had strong signals in Han-Ch–may be responsible for these observations.

While the phenotypic role of individual genes in population differences may be difficult to assert, by looking at molecular pathways and processes we were able to detect many that may be related to population-specific traits and were perhaps under selection in specific populations. In fact, divergent GO categories are more likely to be caused by selection since they encompass concatenated sequences from multiple genes, usually in different genomic regions. Our results from GO categories are thus statistically robust and in line with results using other methods. They confirm previous findings suggesting selection in processes related to senses like taste, smell, and audition as well as immune response. Many of the categories we found related to lipid and fatty acid metabolism are likely the result of adaptations to new food sources, as reported by others (Voight et al., 2006). Therefore, several of our results are in accordance with previous findings indicating patterns of selection consistent with adaptation to new environments such as climates, pathogens, and sources of food.

Although our finding of GO categories related to reproduction as potential targets of selection has been reported before (Voight et al., 2006), we were intrigued to find evidence of higher divergence in genes related to male genitalia in Afr-Am. Genes related to male genitalia were also among our enriched GO categories in the HapMap African population, showing that these results are not due to the evolutionary history of African Americans. Because African Americans have been shown to have higher levels of testosterone than Caucasians (Ross et al., 1986; Winters et al., 2001), we speculate that these results could be phenotypically relevant.

We found evidence of divergence in multiple categories related to neuronal development in all three populations, in both Perlegen and HapMap analyses, yet more strongly in Han-Ch. Williamson et al. (2007) previously identified genes involved in nervous system development as candidates of recent selective sweeps, yet our results expand on these previous findings and provide additional genes and categories with evidence of being involved in population differentiation. These results are also in accordance with human evolutionary history. As populations migrated to new environments, adaptations in terms of behavior driven by climate, landscape, fauna, or flora might have resulted in adaptations at the level of the nervous system. These results may also be explained by cultural processes that could have played a role in shaping the human genome (Laland et al., 2010). Indeed, behavior was a significant GO category in Han-Ch which as far as we know is novel. Given the nature of our analysis, however, demographic effects alone could also explain these results.

Several other GO categories identified by us as candidates may provide clues about human evolution, adaptation, and genetics. To our knowledge, UV protection has not been associated with selection in African populations and, even though it may only be relevant to specific geographic regions, we found such evidence for both Afr-Am and YRI populations. We speculate that slightly deleterious derived alleles affecting UV protection drifted to higher frequencies in African populations given the protection from the sun offered by dark skin pigmentation.

In conclusion, we present a conceptually simple algorithm to detect genes and processes that most diverged in populations and apply it to data from three human populations from three continents. This work confirms and extends previous findings on the African origins of the human species, providing new candidate genes for specifying population-specific traits. Some of these genes may have been targets of selection while others may have been differentiated by chance. Our work also reveals putative processes under selection in each of the populations. Of particular note, our results showing evidence of differentiation in neuronal genes support the hypothesis that adaptations to new environments and/or cultural processes could have driven adaptations at the level of the nervous system. Besides, our results suggesting divergence in genes related to the male reproductive system in Africans are to our knowledge novel. These findings may help understand the genetics and evolution of human populations as well as the phenotypic and biological differences between them. Our method and results are therefore complementary and synergistic to those previously published and are one more step in understanding genetic differences between populations that could be important for human evolution and adaptation as well as medically relevant. A combination of methods will be necessary to understand the genetic basis of population phenotypic differences and our method is one more approach available to researchers.

Acknowledgements

The work of AM was supported by a fellowship from the Fundação Luso-Americana. JPM thanks the BBSRC (BB/H008497/1), the Ellison Medical Foundation, and a Marie Curie International Reintegration Grant within EC-FP7 for supporting work in his lab. Further thanks to Chris Tyler-Smith for useful suggestions; Steve Paterson, Harry Noyes, and Austin Hughes for comments on previous drafts; Jorge Amigo Lechuga for valuable assistance with SPSmart; and Domingos Magalhães for advice on the statistical analyses.

References

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* **12**, 1805–1814.

Amigo, J., Salas, A., Phillips, C., & Carracedo, A. (2008) SPSmart: Adapting population based SNP genotype databases for fast and comprehensive web access. *BMC Bioinform* **9**, 428.

Arbiza, L., Dopazo, J., & Dopazo, H. (2006) Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *PLoS Comput Biol* **2**, e38.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29.

Barreiro, L. B., Laval, G., Quach, H., Patin, E., & Quintana-Murci, L. (2008) Natural selection has driven population differentiation in modern humans. *Nat Genet* **40**, 340–345.

Batzer, M. A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D. H., Shaikh, T. H., Novick, G. E., Ioannou, P. A., Scheer, W. D., Herrera, R. J., & Deininger, P. L. (1994) African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci USA* **91**, 12288–12292.

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., Rhodes, M., Reich, D. E., & Hirschhorn, J. N. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* **74**, 1111–1120.

Burchard, E. G., Ziv, E., Coyle, N., Gomez, S. L., Tang, H., Karter, A. J., Mountain, J. L., Perez-Stable, E. J., Sheppard, D., & Risch, N. (2003) The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* **348**, 1170–1175.

Cann, R. L., Stoneking, M., & Wilson, A. C. (1987) Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.

Chen, H., Patterson, N., & Reich, D. (2010) Population differentiation as a test for selective sweeps. *Genome Res* **20**, 393–402.

Coop, G., Pickrell, J. K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., Myers, R. M., Cavalli-Sforza, L. L., Feldman, M. W., & Pritchard, J. K. (2009) The role of geography in human adaptation. *PLoS Genet* **5**, e1000500.

Deng, L., Zhang, Y., Kang, J., Liu, T., Zhao, H., Gao, Y., Li, C., Pan, H., Tang, X., Wang, D., Niu, T., Yang, H., & Zeng, C. (2008) An unusual haplotype structure on human chromosome 8p23 derived from the inversion polymorphism. *Hum Mutat* **29**, 1209–1216.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., De Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., Mccarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J.,

Mcvean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Yakub, I., Birren, B. W., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., Mclay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., Mcewen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., & Stewart, J. (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861.

Gibson, G. E., Blass, J. P., Beal, M. F., & Bunik, V. (2005) The alpha-ketoglutarate-dehydrogenase complex: A mediator between mitochondria and oxidative stress in neurodegeneration. *Mol Neurobiol* **31**, 43–63.

Graf, J., Hodgson, R., & Van Daal, A. (2005) Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. *Hum Mutat* **25**, 278–284.

Hamblin, M. T., & Di Rienzo, A. (2000) Detection of the signature of natural selection in humans: Evidence from the Duffy blood group locus. *Am J Hum Genet* **66**, 1669–1679.

Handley, L. J., Manica, A., Goudet, J., & Balloux, F. (2007) Going the distance: Human population genetics in a clinal world. *Trends Genet* **23**, 432–439.

Hartl, D. L., & Clark, A. G. (2007) *Principles of population genetics*. Sunderland, MA: Sinauer and Associates.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A., & Cox, D. R. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072–1079.

Huang Da, W., Sherman, B.T., & Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57.

Hughes, A. L., Welch, R., Puri, V., Matthews, C., Haque, K., Chanock, S. J., & Yeager, M. (2008) Genome-wide SNP typing reveals signatures of population history. *Genomics* **92**, 1–8.

Johansson, A., & Gyllensten, U. (2008) Identification of local selective sweeps in human populations since the exodus from Africa. *Hereditas* **145**, 126–137.

Laland, K. N., Odling-Smee, J., & Myles, S. (2010) How culture shaped the human genome: Bringing genetics and the human sciences together. *Nat Rev Genet* **11**, 137–48.

Lamason, R. L., Mohideen, M. A., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., Jurynec, M. J., Mao, X., Humphreville, V. R., Humbert, J. E., Sinha, S., Moore, J. L., Jagadeeswaran, P., Zhao, W., Ning, G., Makalowska, I., Mckeigue, P. M., O'donnell, D., Kittles, R., Parra, E. J., Mangini, N. J., Grunwald, D. J., Shriver, M. D., Canfield, V. A., & Cheng, K. C. (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782–1786.

de Magalhaes, J. P., & Church, G. M. (2007) Analyses of human-chimpanzee orthologous gene pairs to explore evolutionary hypotheses of aging. *Mech Ageing Dev* **128**, 355–364.

de Magalhaes, J. P., Finch, C.E., & Janssens, G. (2010) Next-generation sequencing in aging research: Emerging applications, problems, pitfalls and possible solutions. *Ageing Res Rev* **9**, 315–323.

Manica, A., Amos, W., Balloux, F., & Hanihara, T. (2007) The effect of ancient population bottlenecks on human phenotypic variation. *Nature* **448**, 346–348.

McGraw, S., Morin, G., Vigneault, C., Leclerc, P., & Sirard, M. A. (2007) Investigation of MYST4 histone acetyltransferase and its involvement in mammalian gametogenesis. *BMC Dev Biol* **7**, 123.

Murati, A., Adelaide, J., Mozziconacci, M. J., Popovici, C., Carbuccia, N., Letessier, A., Birg, F., Birnbaum, D., & Chaffanet, M. (2004) Variant MYST4-CBP gene fusion in a t(10;16) acute myeloid leukaemia. *Br J Haematol* **125**, 601–604.

Myles, S., Somel, M., Tang, K., Kelso, J., & Stoneking, M. (2007) Identifying genes underlying skin pigmentation differences among human populations. *Hum Genet* **120**, 613–621.

Myles, S., Tang, K., Somel, M., Green, R. E., Kelso, J., & Stoneking, M. (2008) Identification and analysis of genomic regions with large between-population differentiation in humans. *Ann Hum Genet* **72**, 99–110.

Nei, M. (1987) *Molecular evolutionary genetics*. New York: Columbia University Press.

Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., Srinivasan, B. S., Barsh, G. S., Myers, R. M., Feldman, M. W. & Pritchard, J. K. (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19**, 826–837.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., & Cavalli-Sforza, L. L. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* **102**, 15942–15947.

Rippe, V., Drieschner, N., Meiboom, M., Murua Escobar, H., Bonk, U., Belge, G., & Bullerdiek, J. (2003) Identification of a gene rearranged by 2p21 aberrations in thyroid adenomas. *Oncogene* **22**, 6111–6114.

Ross, R., Bernstein, L., Judd, H., Hanisch, R., Pike, M., & Henderson, B. (1986) Serum testosterone levels in healthy young black and white men. *J Natl Cancer Inst* **76**, 45–48.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., Mccarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A.,

Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., De Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., Mccarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Sham, P. C., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., Mcvean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Johnson, T. A., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Yakub, I., Birren, B. W., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., Mclay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., Mcewen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., & Stewart, J. (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918.

Schumacher, J. M., Lee, K., Edelhoff, S., & Braun, R. E. (1995) Spnr, a murine RNA-binding protein that is localized to cytoplasmic microtubules. *J Cell Biol* **129**, 1023–1032.

Shifman, S., Kuypers, J., Kokoris, M., Yakir, B., & Darvasi, A. (2003) Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* **12**, 771–776.

Smith, J. M., & Haigh, J. (1974) The hitch-hiking effect of a favourable gene. *Genet Res* **23**, 23–35.

Storey, J. D., & Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440–9445.

Stringer, C., & Mckie, R. (1996) *African exodus: The origins of modern humanity*. London: Jonathan Cape.

Tang, K., Thornton, K. R., & Stoneking, M. (2007) A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol* **5**, e171.

Tishkoff, S. A. & Kidd, K. K. (2004) Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* **36**, S21–S27.

Tobin, J. E., Latourelle, J. C., Lew, M. F., Klein, C., Suchowersky, O., Shill, H. A., Golbe, L. I., Mark, M. H., Growdon, J. H., Wooten, G. F., Racette, B. A., Perlmutter, J. S., Watts, R., Guttman, M., Baker, K. B., Goldwurm, S., Pezzoli, G., Singer, C., Saint-Hilaire, M. H., Hendricks, A. E., Williamson, S., Nagle, M. W., Wilk, J. B., Massood, T., Laramie, J. M., Destefano, A. L., Litvan, I., Nicholson, G., Corbett, A., Isaacson, S., Burn, D. J., Chinnery, P. F., Pramstaller, P. P., Sherman, S., Al-Hinti, J., Drasby, E., Nance, M., Moller, A. T., Ostergaard, K., Roxburgh, R., Snow, B., Slevin, J. T., Cambi, F., Gusella, J. F. & Myers, R. H. (2008) Haplotypes and gene expression implicate the MAPT region for Parkinson disease: The GenePD Study. *Neurology* **71**, 28–34.

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006) A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72.

Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V., Batzer, M. A., Harpending, H. C., Rogers, A. R., & Jorde, L. B. (2001) Patterns of ancestral human diversity: An analysis of Alu-insertion and restriction-site polymorphisms. *Am J Hum Genet* **68**, 738–752.

Watkins, W. S., Rogers, A. R., Ostler, C. T., Wooding, S., Bamshad, M. J., Brassington, A. M., Carroll, M. L., Nguyen, S. V., Walker, J. A., Prasad, B. V., Reddy, P. G., Das, P. K., Batzer, M. A., & Jorde, L. B. (2003) Genetic variation among world populations: Inferences from 100 Alu insertion polymorphisms. *Genome Res* **13**, 1607–1618.

Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M., & Hill, W. G. (2005) Measures of human population structure show heterogeneity among genomic regions. *Genome Res* **15**, 1468–1476.

Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D., & Nielsen, R. (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genet* **3**, e90.

Winters, S. J., Brufsky, A., Weissfeld, J., Trump, D. L., Dyky, M. A. & Hadeed, V. (2001) Testosterone, sex hormone-binding globulin, and body composition in young adult African American and Caucasian men. *Metabolism* **50**, 1242–1247.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1 Top ranked genes in Afr-Am ($P < 0.05$) obtained using Perlegen data.
Table S2 Top ranked genes in Eur-Am ($P < 0.05$) obtained using Perlegen data.
Table S3 Top ranked genes in Han-Ch ($P < 0.05$) obtained using Perlegen data.
Table S4 Top ranked genes in Eurasians ($P < 0.05$) obtained using Perlegen data.
Table S5 Top ranked GO categories in Afr-Am ($P < 0.05$) obtained using Perlegen data.
Table S6 Top ranked GO categories in Eur-Am ($P < 0.05$) obtained using Perlegen data.
Table S7 Top ranked GO categories in Han-Ch ($P < 0.05$) obtained using Perlegen data.
Table S8 Top ranked GO categories in Eurasians ($P < 0.05$) obtained using Perlegen data.
Table S9 Top ranked genes in Afr ($P < 0.05$) obtained using HapMap data.
Table S10 Top ranked genes in Eur ($P < 0.05$) obtained using HapMap data.
Table S11 Top ranked genes in Asn ($P < 0.05$) obtained using HapMap data.
Table S12 Top ranked genes in Eurasians ($P < 0.05$) obtained using HapMap data.

Table S13 Top ranked GO categories in Afr ($P < 0.05$) obtained using HapMap data.
Table S14 Top ranked GO categories in Eur ($P < 0.05$) obtained using HapMap data.
Table S15 Top ranked GO categories in Asn ($P < 0.05$) obtained using HapMap data.
Table S16 Top ranked GO categories in Eurasians ($P < 0.05$) obtained using HapMap data.
Table S17 Functional enrichment in Afr-Am obtained from DAVID.
Table S18 Functional enrichment in Eur-Am obtained from DAVID.
Table S19 Functional enrichment in Han-Ch obtained from DAVID.
Table S20 Functional enrichment in Eurasians obtained from DAVID.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received: 27 August 2011
Accepted: 17 November 2011