GENETIC ANALYSIS OF AFRICAN POPULATIONS: HUMAN EVOLUTION AND COMPLEX DISEASE

Sarah A. Tishkoff and Scott M. Williams‡§*

Africa is one of the most ethnically and genetically diverse regions of the world. It is thought to be the ancestral homeland of all modern humans, and is the homeland of millions of people of the recent African diaspora. Because of the central role of African populations in human history, characterizing their patterns of genetic diversity and linkage disequilibrium is crucial for reconstructing human evolution and for understanding the genetic basis of complex diseases.

MODERN HUMANS
Homo sapiens sapiens. Anatomically modern humans who appear in the fossil record ~150,000–200,000 years ago and who are descended from *Homo erectus.*

HAPLOTYPE
A set of genetic markers that is present on one chromosome.

**Department of Biology, University of Maryland, College Park, Maryland 20742, USA. ‡Department of Microbiology, Meharry Medical School, Nashville, Tennessee 37208, USA. §Division of Cardiovascular Medicine, 383 PRB, and Program in Human Genetics, Vanderbilt University Medical Center, Nashville, Tennessee 37232, USA. Correspondence to S.A.T. or S.W.*
e-mails: tishkoff@umail.umd.edu; smwilliams@mail.mmc.edu
doi:10.1038/nrg865

Sub-Saharan African populations have a rich and important history from which we can learn about our origins as a species and the way in which genetic variation affects human phenotypes, including complex disease. Africa is thought to be the ancestral homeland of all MODERN HUMANS, and is the more recent homeland of millions of individuals whose ancestors were brought to Europe and to the Americas as slaves. The study of the levels and patterns of genetic diversity among the multitude of ethnically diverse African populations will shed light on many questions about human evolutionary history and the genetic basis of phenotypic variation. However, despite the important contributions that studies of African populations can make, these populations have been understudied compared with non-African populations.

One of the largest 'migration' events in recent human history has been that of millions of Africans who were brought as slaves to the Americas and Europe in the past several hundred years. The slaves, who originated from ethnically diverse African populations (predominantly from West Africa), admixed with each other, as well as with people of mainly European descent, making their gene pool genetically heterogeneous. People of the African diaspora often have high prevalences of certain complex diseases, such as hypertension, diabetes and prostate cancer, for which the genetic bases are poorly understood.

To better understand the genetic basis of such complex genetic diseases in both Africans and non-Africans, it is important to study African populations. Such studies will be especially important if common diseases turn out to be caused by common susceptibility alleles that are likely to be old and, therefore, present in African populations (as discussed in more detail below). In cases where this hypothesis — the common disease/common variant (CD/CV) hypothesis — does not hold to be true, but where many less-frequent, and possibly population-specific, alleles predispose to disease, it will be useful to know about possible susceptibility alleles in different populations. The high levels of genetic diversity in African populations and their demographic history also make these populations particularly informative for the fine mapping of complex genetic diseases (as discussed later)[1]. Furthermore, many of the environmental risk factors that trigger certain complex diseases are not as common in African populations[2,3], which makes it more feasible to differentiate genetic from environmental risk factors for these diseases. Because complex diseases are becoming more prevalent in Africa, as are the known environmental risk factors that are associated with urban and more sedentary lifestyles, it is imperative that we study African populations and their genetic diversity.

In this review, we discuss the distribution of genetic diversity across sub-Saharan Africa, HAPLOTYPE structure and LINKAGE DISEQUILIBRIUM (LD) in African

Table 1 | **Comparison of genetic diversity and population subdivision in Africa, Europe and Asia**

Locus	Marker system	Africa (*N*)	Europe (*N*)	Asia (*N*)	Highest genetic diversity*	F_{ST}§ Africa	F_{ST}§ Europe	F_{ST}§ Asia	References
MtDNA	HVSI	72	120	63	Africa	0.088	0.045	0.032	19
MtDNA	HVSII	72	120	63	Africa	0.092	0.013	0.017	19
NRY	6 STRPs	72	120	63	Africa	0.026	0.602	0.092	19
NRY	43 biallelic	360	507	1415	Africa‡	0.220	0.128	0.271	25
Autosome	60 STRPs	72	120	63	Africa	0.024	0.023	0.007	53
Autosome	45 STRPs	216	246	302	Africa	0.033	0.015	0.018	32
Autosome	30 RFLPs	72	120	63	Europe	0.027	0.013	0.017	19
Autosome	13 Alus	72	120	63	Africa	0.017	0.022	0.009	19
Autosome	8 Alus	176	334	359	Africa	0.088	0.011	0.058	33
PLAT	STRP1/Alu	924	352	386	Africa	3.18‖	2.31‖	2.09‖	30
PLAT	STRP2/Alu	1030	410	422	Africa	3.28‖	1.22‖	1.66‖	30

*For each study, regions with the highest levels of genetic diversity are indicated (Africa, Europe or Asia). The extent of differentiation among populations in each region is indicated by the measure F_{ST} or an equivalent measure. ‡Based on pairwise sequence differences. §Values shown were estimated using F_{ST} or the equivalent measures of G_{ST} or Φ_{ST}. ‖Subdivision estimated using a likelihood ratio statistic defined in REF. 30. HVSI/II, hypervariable sequence I/II; mtDNA, mitochondrial DNA; *N*, number of chromosomes sampled; *NRY*, non-recombining region of the Y chromosome; *PLAT*, tissue plasminogen activator; RFLP, restriction-fragment length polymorphism; STRP, short tandem-repeat polymorphism.

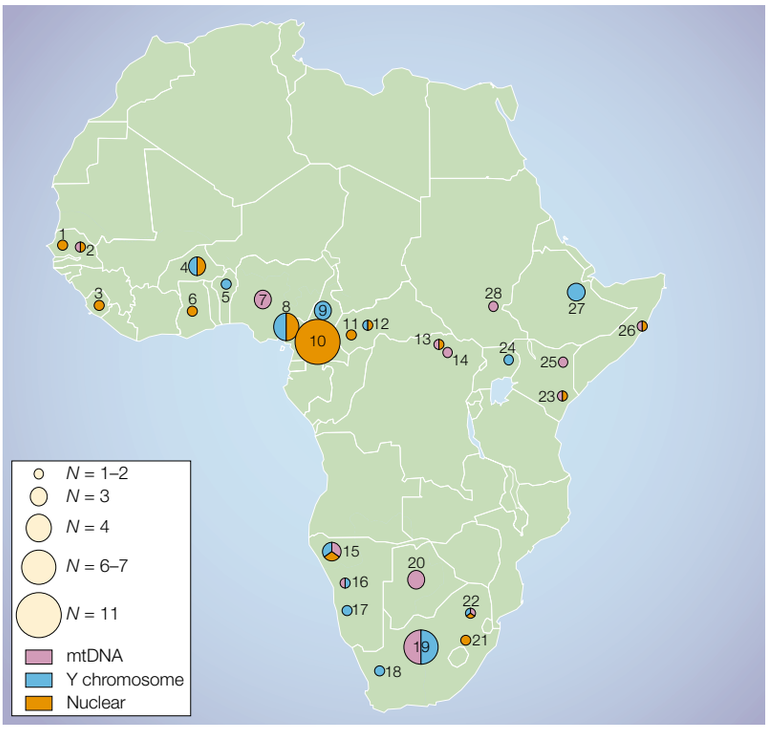


Figure 1 | **Distribution of non-biomedical studies of sub-Saharan African genetic diversity.** Only recent studies that include a sample size of 20 or more individuals from clearly defined ethnic groups are included. The size of each circle relates to the number (*N*) of ethnic groups that were sampled. Circles of one colour indicate ethnic groups that were typed for only one type of marker. Multicoloured circles indicate ethnic groups that were typed for several marker types. Numbers indicate references: 1 (REFS 17,116), 2 (REFS 27,29,30), 3 (REF. 42), 4 (REF. 117), 5 (REF. 45), 6 (REF. 42), 7 (REFS 20,21), 8 (REFS 117,118), 9 (REF. 61), 10 (REFS 42,117–119), 11 (REF. 119), 12 (REFS 25,27,29,30,36), 13 (REFS 17,27,29,30,36), 14 (REF. 63), 15 (REFS 25–27,29,30,33,63,118,120), 16 (REFS 25,63,120), 17 (REFS 120,121), 18 (REF. 120), 19 (REFS 25,63,64,120), 20 (REF. 63), 21 (REF. 33), 22 (REFS 33,117), 23 (REFS 20,21,27,29,30), 24 (REF. 26), 25 (REFS 20,21,63), 26 (REFS 20,21,27,29,30), 27 (REFS 46,61) and 28 (REF. 122). mtDNA, mitochondrial DNA.

genomes, and the use of this information in continuing studies of human evolution and complex genetic disease. The impact of infectious diseases in sub-Saharan Africa, also a topic of great importance, is not discussed here as it has recently been reviewed elsewhere[4].

Genetic diversity in Africa

Africa contains tremendous cultural, linguistic and genetic diversity, and has more than 2,000 distinct ethnic groups and languages (see online link to Ethnologue). Studies using mitochondrial (mt)DNA and nuclear DNA markers consistently indicate that Africa is the most genetically diverse region of the world (TABLE 1). However, most studies report only a few markers in divergent African populations, which makes it difficult to draw general conclusions about the levels and patterns of genetic diversity in these populations (FIG. 1). Because genetic studies have been biased towards more economically developed African countries that have key research or medical centres, populations from more underdeveloped or politically unstable regions of Africa remain undersampled (FIG. 1). Historically, human population genetic studies have relied on one or two African populations as being representative of African diversity, but recent studies show extensive genetic variation among even geographically close African populations, which indicates that there is not a single 'representative' African population.

From the efforts of an international Single Nucleotide Polymorphism (SNP) Consortium, more than 4 million SNPs have now been identified, ~1 per 1–2 kb (REF. 5). However, there is a bias towards the identification of SNPs in non-African populations (see online link to the dbSNP summary). The identification of SNPs across geographically and ethnically diverse human populations is important because the frequency of SNPs can vary substantially between

Box 1 | **Three main models of modern human evolution**

There are three main theories for the evolution of modern humans: the multiregional model, the recent African origin (RAO) model and the assimilation model (all reviewed in REF. 112**).**

The multiregional model proposes that there was no single geographical origin for modern humans but that, after the radiation of *HOMO ERECTUS* **from Africa into Europe and Asia ~800,000–1.8 million years before present (yr BP), there were independent transitions in regional populations from *H. erectus* to *Homo sapiens*. This model is supported primarily by the continuity of certain morphological traits in the fossil record (for example, the robust cheekbones observed in *H. erectus* fossils from Southeast Asia and in modern Australian aborigines), which indicate that modern populations evolved over very long periods of time in the regions where they are found today. Simultaneous evolution from *H. erectus* to *H. sapiens* in dispersed populations could have been achieved through extensive gene flow between populations, requiring a large** EFFECTIVE POPULATION SIZE **to sustain gene flow among geographically diverse populations.**

The RAO model proposes that all non-African populations descend from a *H. sapiens* ancestor that evolved in Africa 100,000–200,000 yr BP. This ancestor then spread throughout the world, replacing archaic *Homo* populations (for example, the Neanderthals). This model is supported by the fossil record, as the earliest modern human fossils were found in Africa and the Middle East, dating to 90,000–120,000 yr BP (REFS 60,112)**. The RAO model predicts that all genetic lineages derive from a recent common African ancestor and that non-African populations should carry a subset of the genetic variation present in modern African populations.**

The assimilation (or hybridization) model proposes that gene flow between the early human populations was not equal over time and space. This model allows for some gene flow between modern humans that migrated from Africa and archaic populations (for example, the Neanderthals) outside Africa. So, the evolution of modern humans could have been due to a blending of modern characters derived from African populations with local characteristics in archaic Eurasian populations. This model predicts that the modern gene pool derives from variable contributions of genes from archaic African and non-African populations.

populations, as will their usefulness as markers for gene-mapping studies[6]. In addition, the density of SNPs needed to map complex diseases is likely to vary across populations with distinct demographic histories[6–8].

The strategy for most efficiently using SNPs to map complex disease genes depends on various parameters that are, at present, not fully known[9]. For example, the CD/CV hypothesis states that common genetic diseases are affected by common disease susceptibility alleles at a few loci that are at high frequency across ethnically diverse populations[7,9–12]. These common alleles are likely to be old and to predate the divergence of human populations. If this is so, complex disease genes that are mapped in African populations by association between SNPs and disease phenotype could have important implications for understanding the genetic risk factors for many ethnic groups. If, however, some complex diseases are influenced by rare susceptibility alleles at many loci, as indicated by simulation studies[10] and, if these alleles are geographically restricted owing to mutation, GENETIC DRIFT or region-specific selection pressure, then characterization of SNP diversity, haplotype structure and LD in populations of African origin will be particularly important for studies of the complex diseases that are prevalent in populations of recent African descent[9].

LINKAGE DISEQUILIBRIUM (LD). The condition in which the frequency of a particular haplotype for two loci is significantly greater than that expected from the product of the observed allelic frequencies at each locus.

HOMO ERECTUS
Species of the genus *Homo* from which modern humans descend, which originated ~1.8 million years ago.

EFFECTIVE POPULATION SIZE (N_e). The theoretical number of breeding individuals, the genetic variation of which can be explained solely by mutation and genetic drift. N_e is related to, but never exceeds, actual population size (N), is most strongly influenced by bottleneck events and reflects the size of a population at its minimum.

A recent African origin of modern humans

Knowledge of human evolutionary history will help us to understand extant patterns of variation and to use those patterns in identifying alleles or genotypes that predispose to disease. Most archaeological and genetic data support a recent African origin (RAO) model of human evolution, although the topic remains hotly debated (BOX 1). However, variations of the RAO model that consider admixture between modern humans who migrated from Africa and archaic populations outside Africa, such as the Neanderthals, cannot be completely ruled out[13–15]. A thorough survey of African genetic diversity will be important for testing models of modern human origins and for determining how phenotypic variation arose.

Studies of protein polymorphisms[16], as well as studies of mtDNA[17–23], Y-chromosomal[19,23–26], autosomal[19,27–35] and X-chromosomal[36–39] DNA variation, indicate that African populations are the most variable and ancestral, as expected under an RAO model (TABLE 1). Phylogenetic analyses of mtDNA[17,18,21,40] and Y-chromosomal haplotypes[24–26] indicate that the most ancestral lineages are African specific and that all non-African lineages can be derived from a single ancestral African haplogroup, consistent with the RAO model. Africans have the largest number of population-specific autosomal[27–31,34,35,41–43], X-chromosomal[38,44] and mtDNA[17–21,23,40] haplotypes, with non-African populations harbouring only a subset of the genetic diversity present in Africa, as would be expected from a genetic bottleneck during migration out of Africa (FIG. 2).

The primary split between Africans and non-African populations is estimated to have occurred from 44,000 to 200,000 years before present (yr BP)[22,27,33,41,45]. An analysis of nuclear autosomal haplotype variability at three genes — *CD4*, myotonic dystrophy 1 (*DM1*) and tissue plasminogen activator (*PLAT*) — in the same set of 33 globally diverse populations, found that all non-African populations have a similar pattern of haplotype variability. They also have a subset of the haplotypic variability that is present in Ethiopian and Somalian populations, which is, in itself, a subset of the variability that is present in other sub-Saharan African populations[27,29,30]. These observations indicate that populations in northeast Africa might have diverged from the rest of sub-Saharan Africa early in the history of modern African populations and that a subset of this northeast African population migrated out of Africa and populated the rest of the globe (FIG. 2). Analysis of mtDNA[18,20,21,40] and Y-chromosome diversity[24–26,46] support a single East African source of migration out of Africa.

Demographic history of African populations

Extant levels and patterns of genetic diversity and LD are influenced by demographic factors, such as fluctuations in population size and structure, and admixture and migration, as well as by gene-specific factors, including selection, mutation rates, recombination and GENE CONVERSION. Mapping genes for complex

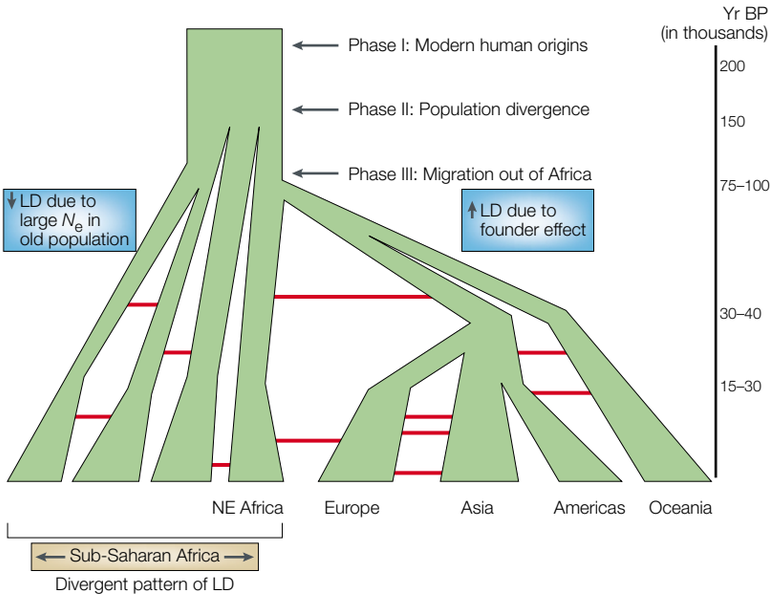


Figure 2 | **Model of human demographic history.** Northeast African populations differentiated from other sub-Saharan African populations early in African history. A small subset of this population migrated out of Africa in the past 100,000 years and rapidly expanded throughout a broad geographical region. Ancestral African populations have maintained a large and subdivided population structure throughout much of their evolutionary history, resulting in fewer sites being in linkage disequilibrium (LD) and in divergent patterns of LD, compared with non-African populations. The bottleneck event that is associated with the founding of non-African populations resulted in reduced genetic variation, greater LD in the genome and an increase in the size of haplotype blocks. Rapid population expansion resulted in the maintenance of this LD. Red lines that connect populations indicate gene flow. N_e, effective population size; Yr BP, years before present. Modified with permission from REF. 58 © (1998) Wiley.

disease relies on having some *a priori* knowledge of haplotype structure and LD to identify polymorphisms that are associated with disease genes. As described in BOX 2, haplotype structure and LD are a product of the age and demographic history of a population. Understanding the demographic history, genetic diversity and haplotype structure of African populations is, therefore, important for designing and interpreting gene-mapping studies in populations of recent African descent and for reconstructing modern human origins and the expansion of modern humans out of Africa, as we discuss below.

African population size. Levels of genetic variation in modern populations reflect the history of their population size. Studies of nuclear sequence diversity in humans consistently estimate an effective population size (N_e) of ~10,000, and studies of mtDNA diversity estimate a haploid N_e of ~5,000 (REF. 47). But estimates of human N_e are usually based on pooled data from African and non-African populations, which are biased towards larger samples from non-Africans and so could result in a biased estimate of N_e, particularly if non-African populations went through severe bottlenecks during migration out of Africa. Therefore, it is not surprising that studies of African populations, in which a range of markers were used[30,33,48,49], have reported a larger N_e in African populations.

Despite uncertainties in the estimates of long-term N_e in humans, it is clear that, after an initial speciation event from *Homo erectus* to modern *Homo sapiens*, humans spread across a broad geographical region and rapidly increased in population size in the past 50,000–100,000 years. An even more rapid population expansion probably occurred after the development and spread of agriculture in the past 10,000 years[16]. Analyses of pairwise mtDNA sequence divergence in globally diverse populations indicate that African populations expanded in size earlier than either Asian or European populations: 99,000 yr BP versus 52,000 yr BP and 23,000 yr BP in Asia and Europe, respectively[47,50]. An earlier African expansion is also supported by studies of highly variable microsatellites[51–53].

The patterns of sequence variation of autosomal and X-chromosomal genes do not give such a clear indication of rapid population growth[54,55]. Approximately 50% of the available nuclear DNA data indicate an excess of intermediate-frequency haplotypes rather than an excess of low-frequency haplotypes, as predicted by a model of rapid population growth[54,55]. However, when African samples pooled by geographical region were analysed separately from non-African samples, more loci showed evidence for rapid population growth[48]. It has been suggested that the pattern of variation that is observed in some nuclear genes could be the result of historical BALANCING SELECTION, which predicts an excess of haplotypes at intermediate frequencies[48,55]. Many genes, such as β-globin (*HBB*), Duffy (*FY*), pyruvate dehydrogenase E1 α-subunit (*PDHA1*) and melanocortin 1 receptor (*MC1R*), that do not show evidence of population growth are potential targets for selection. By analysing nuclear sequence variability from several loci sampled from ethnically diverse African populations, these conflicting patterns of human demographic history should eventually be resolved.

African populations are highly subdivided. The classic measure of POPULATION SUBDIVISION is F_{ST}, which ranges in value from zero to one; an F_{ST} value of zero indicates no differentiation between populations. Estimates of F_{ST} in different geographical regions (Africa, Europe and Asia) for protein polymorphisms, blood groups, restriction-fragment length polymorphisms (RFLPs) and autosomal short tandem-repeat polymorphisms (STRPs), indicate that only 11–16% of observed variation is due to differences among populations (F_{ST} = 0.11–0.16), consistent with a recent common ancestry and/or sufficient gene flow to maintain population similarities[19]. Global F_{ST} estimates based on variation in mtDNA (F_{ST} = 0.24–0.27)[19] and the Y chromosome (F_{ST} = 0.30–0.64)[19,25,26,56,57] are higher, possibly owing to the smaller effective size for these loci, which would cause more genetic drift[19].

There have been relatively few studies of population subdivision in sub-Saharan Africa, and those that have been done do not yield consistent results (TABLE 1). This is probably due to the types of marker and the types of analysis used, as well as the number and geographical range of the African populations studied. Phylogenetic

GENETIC DRIFT
The random fluctuation that occurs in allele frequencies as genes are transmitted from one generation to the next. This is because allele frequencies in any sample of gametes that perpetuate the population might not represent those of the adults in the previous generation.

GENE CONVERSION
A specific type of recombination, which results in non-reciprocal genetic exchange, in which the sequence of one DNA strand is used to alter the sequence of the other.

BALANCING SELECTION
The maintenance of genetic polymorphism owing to selection.

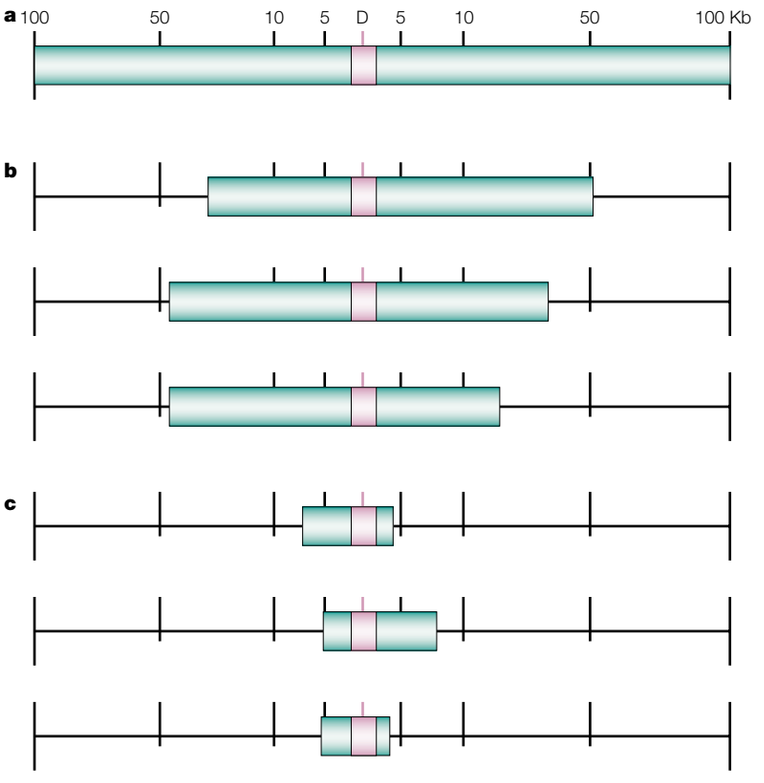
POPULATION SUBDIVISION
A population sample that is not genetically homogeneous but, rather, is composed of more than one subpopulation with low levels of gene flow.

Box 2 | **Linkage disequilibrium and population demography**

Mapping disease genes by association requires the identification of linkage disequilibrium (LD) between a marker and a disease phenotype, making it vitally important to know the underlying levels and patterns of LD in a given study population to avoid erroneous conclusions. Several studies[8,27,29,30,65] of African populations have indicated that levels and patterns of LD in these populations differ from those in non-African populations owing to the age of African populations, admixture with other African and non-African populations, and historical differences in population size and substructure.

A disease mutation (shown in violet) that occurs on a single haplotype background will initially be in complete LD with flanking markers on that chromosome (see panel **a**). In each generation, LD between a marker and a disease allele decays owing to recombination between the sites, and also because of the effects of mutation and gene conversion at marker loci. Young populations, and those that have undergone recent bottlenecks (as probably occurred during the migration of ancestral humans out of Africa), will have haplotype blocks of large to moderate size (panel **b**, shown in green)[8,65]. (In panel **b**, LD extends up to 50 kb in at least one direction.) In older and larger African populations, in which there has been more recombination, the size of haplotype blocks will probably be smaller (in panel **c**, the haplotype blocks extend for less than 10 kb)[8,65].

LD can also be established by a founder event, with the strength and extent of the LD depending on the severity and length of the bottleneck event[7,66,67]. Population substructure increases LD owing to a smaller effective population size and to higher levels of genetic drift in subdivided populations[67]. So, if a pooled sample derived from several African populations was analysed, spurious LD would be detected, even if the haplotypes in each subpopulation were in linkage equilibrium. This could lead to erroneous conclusions about the association between genetic markers and disease phenotype[67]. Small populations of stable size are expected to show LD between closely linked loci as a result of increased genetic drift, and larger populations will have fewer sites in LD[66,67,113]. New mutations are less likely to be in LD in growing populations owing to the smaller effect of genetic drift, but allelic associations that exist before population expansion might persist for a longer period of time in an expanding population than in a population of constant size[66,113].



analyses of haplotypes from Y chromosomes[24–26,46], mtDNA[17,18,21], autosomes[28] and X chromosomes[38] are indicative of a long history of population subdivision in Africa. Values of F_{ST} estimated from mtDNA are usually much higher in Africa than elsewhere (TABLE 1). Genotyping studies of autosomal SNPs, STRPs or Alu elements in diverse African populations have also reported higher F_{ST} values for African populations[19,32,33], but the difference between regions is not as striking (TABLE 1). However, studies of autosomal haplotype variability, which might be more informative than studying unlinked loci to assess population subdivision, show much stronger evidence of high levels of genetic diversity among African populations[27–30,38]. This is particularly true for haplotypes that contain highly variable STRP markers[27,29,30], whereas haplotypes composed of high-frequency SNP markers are more likely to be old and shared among African populations[37]. Inconsistencies in estimates of F_{ST} for autosomal, Y-chromosomal and mtDNA markers might be due to the smaller N_e of the Y chromosome and mtDNA, to selection acting on mtDNA and/or Y-chromosomal DNA or different male and female migration histories[19,26,57,58].

Historical migration events in Africa. Historical migration and subsequent admixture have shaped patterns of LD and haplotype structure in African populations[59]. Ancient African migration events are complex, owing to historical fluctuations in geology, ecology and climate, including periods of glaciation and warming, that have affected population expansion and contraction[60]. However, historical migration events can be discerned through analysis of mtDNA and Y-chromosomal haplotype lineages (FIG. 3). African mtDNA has three main lineages — L1, L2 and L3 — which have an estimated COALESCENCE date of 126,000–165,000 yr BP (REFS 17,18,20,21). The L1 lineage is the most ancient and is present in the San population from South Africa and the Biaka Pygmies from the Central African Republic, which are two of the most genetically divergent populations in Africa[18]. L2 and L3 diverged from L1 ~60,000–103,000 yr BP (REFS 17,18,21). The L2 lineage is present in Mbuti Pygmies from the Democratic Republic of the Congo and in west African Bantu-speaking populations. The L3 lineage is widely dispersed throughout east Africa but is rare elsewhere in sub-Saharan Africa[17,18,21]. Phylogenetic analysis indicates that the L3 haplogroup is the precursor of non-African mtDNA haplotypes[18] and that a subset of this lineage (L3a) travelled out of Africa in the ancestors of modern Eurasians ~60,000–80,000 yr BP (REF. 21). The L3a lineage also occurs at a high frequency in Ethiopian mtDNA[40], which supports the proposal that modern humans migrated out of Africa through Ethiopia[27,29,30].

Y-chromosomal haplotype analyses have shown that the most ancestral Y-chromosomal haplogroup is present in east African Sudanese and Ethiopians, as well as in southern African Khoisan-speaking populations, such as the San[24,25,46,61,62]. These findings support the mtDNA data, which indicate that the San population is the most genetically divergent population in Africa, and

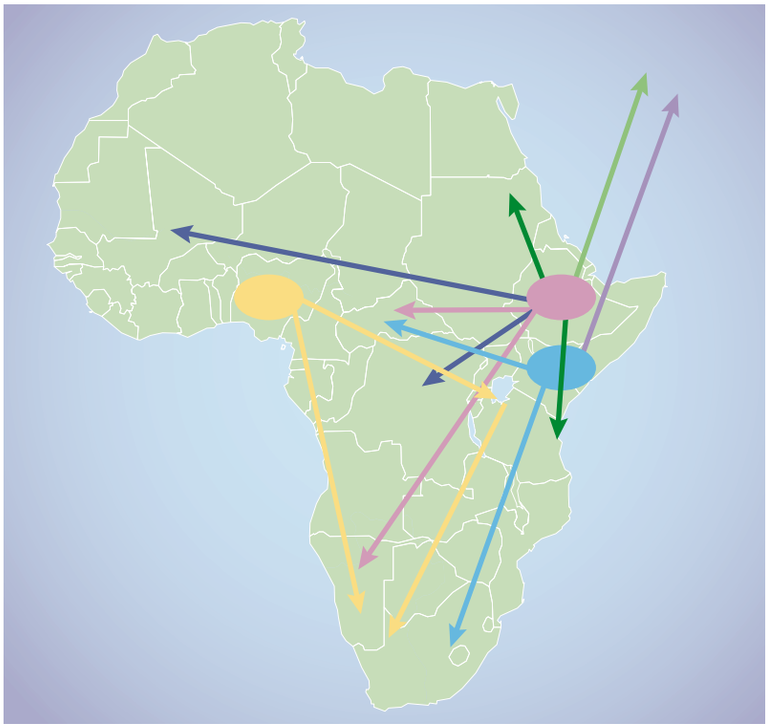


Figure 3 | **Major migration events in Africa.** Light blue arrows indicate expansion of an ancestral Y-chromosomal lineage (I and II) from east Africa into southern and central Africa in the past 100,000 years, and the light purple arrow indicates the expansion of a subset of the Y-chromosomal lineage II out of Africa in the past 44,000–100,000 years. Pink arrows indicate expansion of the L1 mitochondrial DNA (mtDNA) lineage into southern and central Africa in the past 120,000 years, the dark blue arrows indicate expansion of the L2 mtDNA lineage into central and western Africa in the past 30,000–70,000 years, and the dark green arrows indicate expansion of the L3 mtDNA lineage throughout east Africa. The light green arrow indicates the expansion of the L3a subhaplogroup out of Africa in the past 70,000–100,000 years. The yellow arrows indicate the recent expansion of Bantu speakers from a postulated homeland in Cameroon into southern Africa in the past 3,000 years, as evidenced by both mtDNA and Y-chromosomal haplotype variation.

also protein and archaeological data, which together indicate that Khoisan-speaking populations originated in east Africa and migrated into southern Africa >10,000 yr BP (REFS 16,61). All Y-chromosomal haplotypes in non-African populations are derived from a single Y-chromosomal lineage from Africa, which reflects an expansion event out of Africa as recently as 44,000 yr BP (REF. 24).

Archaeological, linguistic and genetic data indicate that several important long-range migration events have occurred in Africa in the past several thousand years, shaping the pattern of genetic variation in modern African populations. The most significant of these is the expansion of Bantu-speaking agricultural populations from a postulated homeland in Cameroon into southern Africa (either directly south and/or through the Great Lakes region of east Africa) in the past 3,000 years (REF. 63) (FIG. 3), which is supported by both mtDNA[63] and Y-chromosome data[25,56,62].

There have also been recent migration events of non-Africans, particularly those of semitic origin, along the east coast of Africa, and subsequent admixture with indigenous African populations. For example, in the Bantu-speaking Lemba population of Zimbabwe, who claim to have Jewish ancestry[64], Y-chromosome lineages have been found that are of recent Jewish descent[64].

So, population history in Africa is likely to be a complex web of population diversification that involves population expansions, contractions, fragmentation, and differential levels and patterns of gene flow (FIGS 2,3). An analysis of genome-wide genetic variation in diverse African populations is required to understand better the genetic structure of these populations.

Linkage disequilibrium in African populations

African demographic history has shaped patterns of genetic variation and LD, and these patterns have important implications for mapping complex disease genes in populations of African descent (BOX 2). As mentioned above, the success of finding an association between a disease gene and variable markers or haplotypes depends largely on knowing the underlying patterns of LD in the population that is being studied[1,11].

Haplotype analysis of the *CD4* locus in 42 globally diverse human populations (including 16 from sub-Saharan Africa) has shown that sub-Saharan African populations have patterns of LD that are distinct from non-African populations[27]. The African populations have more haplotypes, lower levels of LD between alleles and more divergent patterns of LD (alleles that are in positive association in one population might be in negative association in another). This initial finding has been supported by haplotype analyses at the *DM1* and *PLAT* loci in the same populations[29,30], and by two recent studies of long-range LD between SNP markers at several nuclear loci in African and non-African populations[8,65], both of which used high-frequency SNPs that had been identified in non-African populations. Reich *et al.*[65] found that LD in a Nigerian (Yoruba) population extends an average distance of only 5 kb, as opposed to the 60 kb that is observed in a European-descended Mormon population. Gabriel *et al.*[8] examined LD between closely spaced SNPs in the same Yoruban and Mormon populations, as well as in Japanese, Chinese and African-American populations. The average size of haplotype blocks (regions with high levels of LD) was estimated to be 11 kb in the Yoruban and African-American samples, and 22 kb in the European and Asian samples. Both studies found that haplotype block size differed markedly across genomic regions, as well as between populations, which shows the need to construct a detailed empirical haplotype map of the genome across ethnically diverse populations[8]. The degree of bias caused by the use of SNPs identified in non-African populations in these studies requires further analysis[8,9].

The simplified model of demographic history, shown in FIG. 2, can explain the divergent patterns of LD in African, compared with non-African, populations. Ancestral African populations should have fewer sites in LD and shorter blocks of LD compared with non-Africans — first, because they have maintained a larger N_e, and second, because there has been more time for

F_{ST}
A measure of population subdivision that indicates the proportion of genetic diversity found between populations relative to the amount within populations.

COALESCENCE
The convergence of alleles, in a modern population, back in time to a common ancestor.

recombination and mutation to decrease levels of LD (BOX 2). However, when we examine patterns of LD in single, distinct African populations, we anticipate that patterns of association among pairs of alleles might vary due to the stochastic effects of drift in subdivided populations. The larger number of sites in LD, and the extended size of haplotype blocks, in non-African populations is probably due to a founding event that occurred during the expansion of modern humans out of Africa in the past 100,000 years (REFS 27,29,30). A population bottleneck will produce a reduction in the number and diversity of haplotypes in a founder population. During this founding event, the particular pattern of pairwise allelic association might have differed from the parental African population, depending on the genetic constitution of the founders relative to that of the parental gene pool. Although LD is expected to increase after a founding event, with the strength and extent of the LD depending on the severity and length of the bottleneck event[7,66,67], it might occasionally remain the same or decrease owing to the stochastic effect of genetic drift during population founding[30]. So, it is not surprising that several studies have not observed a marked difference in LD between African and non-African populations[35,68,69].

LD that has been established after a founding event is likely to be maintained in rapidly expanding non-African populations. This model is supported by recent studies in European populations, which show that LD can extend over regions up to 173 kb (REFS 8,70). However, because of European demographic history, haplotype blocks are likely to be smaller and LD patterns are likely to vary across ethnically diverse African populations[8,65]. If this is true, a denser SNP map will be required to study African populations. Although these LD patterns make African populations less useful for the initial stages of association mapping compared with recent founder populations, in which high levels of LD are a feature, they are likely to make African populations more useful for fine-scale mapping of complex disease genes[1,12,65,71]. However, more realistic models of human demographic history need to be considered to account for complex patterns of population expansions and contractions, migration, admixture, population-specific selection and subdivision.

Genetic influences on disease in Africa

The prevalence of disease types differs between the developed and the developing worlds; non-communicable diseases are most prevalent in the developed world, whereas infectious or communicable disease is most prevalent in the developing world, including sub-Saharan Africa. Both communicable and non-communicable diseases have genetic components. Genetic variation has been shown to affect susceptibility to various infectious diseases[4]. In sub-Saharan Africa, the most common genetic diseases have evolved as a by-product of selection to reduce susceptibility to infection — for example, the haemoglobinopathies, which protect individuals from malaria and affect *ca.* 250 million people worldwide (see online link to the WHO report entitled 'Services for the prevention and management of genetic disorders and birth defects in developing countries'). About 75% of the global incidence of haemoglobinopathy is in sub-Saharan Africa (>200,000 births per year)[72].

In addition to the prevalence of infectious diseases, several complex diseases are becoming more common in Africa, presumably due to an increasingly urbanized Western lifestyle. Populations of sub-Saharan African ancestry living outside Africa have a high prevalence of several complex diseases, including hypertension[73,74], diabetes[75–77], obesity[78,79] and prostate cancer. By comparing populations in Africa to populations of African ancestry outside Africa, we might be able to identify both the environmental and the genetic risk factors that contribute to certain complex diseases. Differences in the genetic architecture of populations can cause different responses to the same environmental risks, owing to variations in allele frequencies at interacting loci[80,81]. Such differences might result in the failure to replicate associations at a single locus[81,82], which emphasizes the need to do several cross-population comparisons.

Hypertension. People of sub-Saharan African descent in North America are disproportionately affected by hypertension[73]. For example, African Americans have a 1.5–2-fold increase in hypertension prevalence compared with European-descended Americans, with the largest difference being between African-American and European-American women[83]. African-American females have an age-adjusted prevalence of 35.9% (based on data from 1988 to 1994) compared with 19.7% in female Americans of European descent. The reason for this disparity in disease prevalence among ethnic groups is unknown, but several theories have been proposed. These theories include: differences in genetic predisposition (caused by varying frequencies of common disease susceptibility alleles — the CD/CV hypothesis); the presence of population-specific genes or alleles that predispose to disease; and differences in exposure to environmental risks (including those related to socio-economic status, such as diet)[84,85]. Until more data are collected, the cause of the disparity remains controversial. Nonetheless, several studies indicate a genetic component in the aetiology of hypertension, with HERITABILITY ranging from 45 to 68% in people of sub-Saharan African descent[86,87].

Despite the known high prevalence of hypertension in people of African descent that live in western/urban environments, especially in the United States, it was not considered to be a major disease in sub-Saharan Africa until recently[88]. Studies of hypertension in both rural and urban settings in Nigeria and Cameroon place the incidence of hypertension at 15–20% (REF. 73), and as high as 20–33% in Tanzania[89]. In both studies, urban subjects and women had the highest prevalence. These and other studies document the increasing prevalence of hypertension in sub-Saharan Africa, especially in urban environments.

Several studies have addressed the genetics of this complex disease both in African populations and in populations of sub-Saharan African descent outside

HERITABILITY
The fraction of the phenotypic variance due to genetic variance.

QUANTITATIVE TRAIT LOCUS (QTL). A genetic locus that is identified through the statistical analysis of complex traits (such as body weight). These traits are typically affected by more than one gene and by the environment.

ANGIOTENSINOGEN
The precursor molecule that is cleaved by angiotensin-I-converting enzyme into angiotensin II.

Na^+/H^+ EXCHANGER
The system through which extracellular sodium ions are traded for intracellular hydrogen ions to adjust or maintain cellular pH, as well as cell volume. This exchanger has been implicated in the development of essential hypertension.

Africa. For example, the heritability of plasma levels of angiotensin-I-converting enzyme (ACE) and angiotensin II (AGT11) — which are both molecules in the renin–angiotensin system that lead to vasoconstriction and increased blood pressure — is 67 and 77%, respectively, in Nigeria, and 18% for both in African Americans[90]. These studies show that the heritability of certain complex traits varies between populations of African ancestry. Therefore, comparisons among populations of African ancestry will be necessary to understand the role and interaction of genetic predisposition and environmental risk in this, and presumably other, complex diseases.

One example of the use of an African-descended population to localize a QUANTITATIVE TRAIT LOCUS (QTL) that putatively contributes to hypertension has been reported[71]. Variation in ACE plasma levels had previously been associated with genetic variation in or near the *ACE* locus, but the exact site of the QTL had been difficult to pinpoint. Studies in a German cohort showed a significant association between ACE levels and polymorphisms in a 13-kb region in the *ACE* gene, but results from a Jamaican sample narrowed the region to less than 3 kb. This supports the idea that populations with more genetic diversity and lower levels of LD can be used to finely map disease loci[71].

Association studies of single genes and their role in hypertension in populations of African descent have not been so successful and some have produced conflicting results; for example, positive associations of ANGIOTENSINOGEN (*AGT*) variants with hypertension have not been replicated in all studies[82,91–93]. There are several possible explanations for this. One is that the positive association is spurious. Another is that either gene–gene and/or gene–environmental interactions have a role in determining the hypertension phenotype[80,94]. The potential role of gene–gene interactions has been supported by a finding in a Ghanaian population, in which single genes that show no evidence of association with hypertension when analysed independently do show evidence of association when multilocus interactions are considered[82]. This finding supports the idea that the genetics of complex disease can be the product of alleles at several loci that interact to predispose to disease and highlights the importance of studying the underlying genetic structure of each population.

Diabetes and obesity. Diabetes and obesity are also common conditions in populations of the African diaspora[75,76,78,79], but are rarer in Africa. For example, the prevalence of type 2 diabetes is 1–2% in Nigeria, but 10–13% in African Americans and Afro-Caribbeans[75,77]. This difference will probably decrease over the next few decades as Africans adopt a more westernized lifestyle[88]. Obesity is beginning to pose an important health risk in urban South-African Blacks[95]. In Tanzanian populations, the prevalence of obesity is 17.4 and 21.5% in women and men, respectively, in an urban environment, and 12.7 and 6.2% for women and men, respectively, in a rural environment[96]. In urban South Africa, the prevalence of obesity in women has risen to almost Western levels[95]. Obesity poses a health risk because of its causative role in several associated diseases, such as cardiovascular disease, type 2 diabetes and hypertension[97,98].

African Americans are more overweight than are Americans of European descent or native Africans (for example, Nigerians and Tanzanians)[78,79,96,99,100]. Despite the high heritability measures of obesity-related traits in many African-descended populations[100], few genes have been identified that have a substantial role in human obesity. One gene that is known to have a physiological role in fat metabolism and an association with obesity is the human uncoupling protein 3 (*UCP3*) gene. An exon splice-donor mutation in *UCP3* is associated with both reduced fat oxidation and severe obesity in the Gullah community of South Carolina, USA (a group of African descent with very low levels of European admixture[101]) and the Mende tribe of Sierra Leone[102]. This mutation, which has a frequency of 10% in the Mende tribe and in the Gullah mentioned above, and another *UCP3* polymorphism, have African origins and might promote fat storage, which leads to obesity in a Western/industrialized environment[102]. Other studies of *UCP3*, as well as of *UCP2*, which is tightly linked to *UCP3*, support the conclusion that there is an association between certain genetic variants in these genes with resting energy expenditure and/or body composition in African Americans[103,104]. In one of these studies[103], the findings in African-American women did not replicate in women of European descent, which raises the question of what other genetic and environmental factors might account for this difference between ethnic groups.

Another study examined the frequency of the *825T* allele in the G protein β3-subunit (*GNB3*) gene in rural and urban Africa, China and Germany[105]. The *825T* allele is associated with hypertension in Caucasians and affects Na^+/H^+ EXCHANGER activity, which

Box 3 | Ethical considerations for research in African populations

The importance of genetic research in Africa should not override ethical considerations. Researchers in Europe and/or North America need to address two sets of considerations — those that our own funding agencies and societies mandate and those of the countries and communities in which we work. The former require, among other things, informed consent, and the latter require addressing unique cultural and political concerns, which can be achieved by involving local researchers and potential participants to sensitize foreign researchers to local values. Input from local researchers and participants can also help to shape the research to local interests and needs[114] and prevent the 'hit-and-run' researcher image by providing continuity. The inclusion of local researchers also helps to develop important research infrastructure and expertise at local and national levels. Finally, we need to avoid 'coercive' practices, and coercion needs to be defined by both local and Western standards.

It is also important to give back as much as possible to the communities in which we work, including reporting back on the results of our studies. The needs of many of the local communities are significantly different to what we might normally encounter in the West. Ultimately, taking an ethical approach to studies of African populations will increase the co-operation of the people most affected by this work, the local participants. These tactics will increase the chances of research studies being successful, which will provide value both to the local communities and to the larger scientific community[115].

HUMAN LEUKOCYTE ANTIGEN (HLA). Also known as major histocompatibility complex (MHC). A glycoprotein found on the surface of antigen-presenting cells that presents antigen for recognition by helper T cells.

is elevated in obese but not lean individuals. *GNB3* is also associated with adipogenesis[105]. The *825T* allele is significantly associated with increased weight in non-Africans[105], and was also significantly more common in the Zimbabwean and South-African Black samples than in non-Africans. However, the association with weight in the African samples was not statistically significant because of the low frequency of the alternative (*825C*) allele. Overall, the results indicate that *825T* might be associated with obesity in all populations investigated in this study, but confirmation of this will require the analysis of larger sample sizes from African populations. Urban–rural prevalence differences indicate that environmental factors are also important in the incidence of obesity in these populations.

Diabetes is also increasing in prevalence in sub-Saharan Africa[88]. Several studies have looked at candidate genes for type 1 diabetes, but fewer have examined type 2 diabetes, which prompts a search for appropriate African populations in which to study this disease.

Type 1, or insulin-dependent, diabetes is characterized by the destruction of the insulin-producing cells of the pancreas and is associated with several HUMAN LEUKOCYTE ANTIGEN (HLA) class II alleles. Variation in the HLA complex has been studied in several parts of Africa. Studies in a Bantu population from Cameroon support an association between several HLA class II alleles (for example, *DRB1*03*, *DQA1*0301* and *DQB1*0602*) and type 1 diabetes[106]. These results replicated some of the earlier findings of disease association with *DQA1* and *DQB1* alleles in Senegal[107,108]. The Cameroon study, however, suffered from small numbers of type 1 diabetes patients ($N = 10$) and requires further confirmation.

Associations were also observed between HLA class II alleles and type 1 diabetes in Zimbabwean and South-African Blacks[109,110]. Several, but not all, of the haplotypes that are associated with type 1 diabetes in these two populations are identical. However, it is important to note that the alleles that associate in these studies are different to those in the Cameroonian and Senegalese populations[106–108]. These findings confirm the association of HLA class II with type 1 diabetes, but the underlying basis of this association is obviously complex and variable across populations.

Type 2, or non-insulin-dependent, diabetes is characterized by insulin resistance, adult age of onset and association with obesity. The health impact of type 2 diabetes is substantial, with its incidence predicted to increase to 300 million cases by 2025, with most of the increase occurring in the developing world[111]. The current prevalence of type 2 diabetes in Africa (~1–2%) is lower than in populations of African descent in industrialized nations (~11–13%). The genetic basis of type 2 diabetes is not as well defined as that for type 1 diabetes and is a subject of continuing collaborative research between US, Nigerian and Ghanaian investigators[76].

Conclusions

There is much to be learned from the genetics of sub-Saharan African populations about human origins and evolution, and about the origin and nature of human complex disease. At present, we have little understanding of the genetic structure of sub-Saharan populations and the genetic basis of complex disease in African populations because very few studies have been conducted in African ethnic groups. Our lack of knowledge about the contribution of genetics to disease in African populations extends to single-locus diseases, which are rarely reported in African populations (see link to the Online Mendelian Inheritance in Man database). Research activity has traditionally been biased towards the study of non-African populations, and our knowledge of even the most fundamental information about the genetic basis of disease in Africa is quite limited. Increased funding and resources for studying genetic diversity in Africa is needed to reconstruct human evolutionary history, to dissect the genetic basis of resistance and susceptibility to disease, and to design better drugs for all people. A shared set of DNA resources and the establishment of an African genetic database would help to provide researchers with common information and would facilitate studies of several loci in the same set of African populations. Studies also need to meet strict ethical standards and to involve both local researchers and study participants (BOX 3). We are at a pivotal time in the field of human genetics as we have various new tools to dissect long-standing and important questions. To answer them we will need to include studies of a wider variety of human populations, especially those from Africa.

1. Jorde, L. B., Watkins, W. S. & Bamshad, M. J. Population genomics: a bridge from evolutionary history to genetic medicine. *Hum. Mol. Genet.* **10**, 2199–2207 (2001).
2. Kaufman, J. S., Owoaje, E. E., Rotimi, C. N. & Cooper, R. S. Blood pressure change in Africa: case study from Nigeria. *Hum. Biol.* **71**, 641–657 (1999).
3. Forrester, T., Cooper, R. S. & Weatherall, D. Emergence of Western diseases in the tropical world: the experience with chronic cardiovascular diseases. *Br. Med. Bull.* **54**, 463–473 (1998).
4. Cooke, G. S. & Hill, A. V. Genetics of susceptibility to human infectious disease. *Nature Rev. Genet.* **2**, 967–977 (2001).
5. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
6. Chakravarti, A. To a future of genetic medicine. *Nature* **409**, 822–823 (2001).
7. Reich, D. E. & Lander, E. S. On the allelic spectrum of human disease. *Trends Genet.* **17**, 502–510 (2001).
8. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
Recent and extensive data on haplotype structure and LD in African, European and Asian populations, showing shorter haplotype blocks in African populations.
9. Weiss, K. M. & Clark, A. G. Linkage disequilibrium and the mapping of complex human traits. *Trends Genet.* **18**, 19–24 (2002).
10. Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69**, 124–137 (2001).
11. Chakravarti, A. Population genetics — making sense out of sequence. *Nature Genet.* **21**, 56–60 (1999).
12. Wright, A. F., Carothers, A. D. & Pirastu, M. Population choice in mapping genes for complex diseases. *Nature Genet.* **23**, 397–404 (1999).
13. Nordborg, M. On the probability of Neanderthal ancestry. *Am. J. Hum. Genet.* **63**, 1237–1240 (1998).
14. Wall, J. D. Detecting ancient admixture in humans using sequence polymorphism data. *Genetics* **154**, 1271–1279 (2000).
15. Templeton, A. Out of Africa again and again. *Nature* **416**, 45–51 (2002).
16. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (eds) in *The History and Geography of Human Genes* 413 (Princeton Univ. Press, New Jersey, 1994).
17. Chen, Y. S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S. & Wallace, D. C. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57**, 133–149 (1995).
18. Chen, Y. S. *et al.* mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations. *Am. J. Hum. Genet.* **66**, 1362–1383 (2000).

An extensive analysis of RFLP variation in mtDNA from ethnically diverse African populations. The data from this study support an east African source for migration out of Africa.

19. Jorde, L. B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).
An extensive analysis of nucleotide diversity in the D-loop of mtDNA from ethnically diverse African populations, showing more diversity in Africa.
20. Watson, E. *et al.* mtDNA sequence diversity in Africa. *Am. J. Hum. Genet.* **59**, 437–444 (1996).
21. Watson, E., Forster, P., Richards, M. & Bandelt, H. J. Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.* **61**, 691–704 (1997).
22. Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
23. Jorde, L. B. *et al.* Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.* **57**, 523–538 (1995).
24. Underhill, P. A. *et al.* Y chromosome sequence variation and the history of human populations. *Nature Genet.* **26**, 358–361 (2000).
25. Hammer, M. F. *et al.* Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203 (2001).
26. Hammer, M. F. *et al.* Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**, 427–441 (1998).
27. Tishkoff, S. A. *et al.* Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. *Science* **271**, 1380–1387 (1996).
28. Alonso, S. & Armour, J. A. A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. *Proc. Natl Acad. Sci. USA* **98**, 864–869 (2001).
29. Tishkoff, S. A. *et al.* A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. *Am. J. Hum. Genet.* **62**, 1389–1402 (1998).
30. Tishkoff, S. A. *et al.* Short tandem-repeat polymorphism/Alu haplotype variation at the *PLAT* locus: implications for modern human origins. *Am. J. Hum. Genet.* **67**, 901–925 (2000).
31. Kidd, K. K. *et al.* A global survey of haplotype frequencies and linkage disequilibrium at the *DRD2* locus. *Hum. Genet.* **103**, 211–227 (1998).
32. Calafell, F., Shuster, A., Speed, W. C., Kidd, J. R. & Kidd, K. K. Short tandem repeat polymorphism evolution in humans. *Eur. J. Hum. Genet.* **6**, 38–49 (1998).
33. Stoneking, M. *et al.* Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res.* **7**, 1061–1071 (1997).
34. Armour, J. A. *et al.* Minisatellite diversity supports a recent African origin for modern humans. *Nature Genet.* **13**, 154–160 (1996).
35. Kidd, J. R. *et al.* Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, *PAH*, in a global representation of populations. *Am. J. Hum. Genet.* **66**, 1882–1899 (2000).
36. Zietkiewicz, E. *et al.* Nuclear DNA diversity in worldwide distributed human populations. *Gene* **205**, 161–171 (1997).
37. Zietkiewicz, E. *et al.* Genetic structure of the ancestral population of modern humans. *J. Mol. Evol.* **47**, 146–155 (1998).
38. Labuda, D., Zietkiewicz, E. & Yotova, V. Archaic lineages in the history of modern humans. *Genetics* **156**, 799–808 (2000).
39. Harris, E. E. & Hey, J. X chromosome evidence for ancient human histories. *Proc. Natl Acad. Sci. USA* **96**, 3320–3324 (1999).
40. Quintana-Murci, L. *et al.* Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nature Genet.* **23**, 437–441 (1999).
41. Tishkoff, S. A., Kidd, K. K. & Clark, A. G. in *Trinational Workshop on Molecular Evolution* (eds von Haeseler, A. & Uyenoyama, M.) 181–198 (Duke Univ. Publ. Group, Durham, North Carolina, 1998).
42. Tishkoff, S. A. *et al.* Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
43. Jin, L. *et al.* Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *Proc. Natl Acad. Sci. USA* **96**, 3796–3800 (1999).
44. Kaessmann, H., Heissig, F., von Haeseler, A. & Paabo, S. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nature Genet.* **22**, 78–81 (1999).
45. Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L. & Feldman, M. W. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl Acad. Sci. USA* **92**, 6723–6727 (1995).
46. Semino, O., Santachiara-Benerecetti, A. S., Falaschi, F., Cavalli-Sforza, L. L. & Underhill, P. A. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* **70**, 265–268 (2002).
47. Jorde, L. B., Bamshad, M. & Rogers, A. R. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* **20**, 126–136 (1998).
48. Wall, J. D. & Przeworski, M. When did the human population size start increasing? *Genetics* **155**, 1865–1874 (2000).
49. Sherry, S. T., Harpending, H. C., Batzer, M. A. & Stoneking, M. Alu evolution in human populations: using the coalescent to estimate effective population size. *Genetics* **147**, 1977–1982 (1997).
50. Sherry, S. T. *et al.* Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* **66**, 761–775 (1994).
51. Kimmel, M. *et al.* Signatures of population expansion in microsatellite repeat data. *Genetics* **148**, 1921–1930 (1998).
52. Reich, D. E. & Goldstein, D. B. Genetic evidence for a Paleolithic human population expansion in Africa. *Proc. Natl Acad. Sci. USA* **95**, 8119–8123 (1998).
53. Relethford, J. H. & Jorde, L. B. Genetic evidence for larger African population size during recent human evolution. *Am. J. Phys. Anthropol.* **108**, 251–260 (1999).
54. Przeworski, M., Hudson, R. R. & Di Rienzo, A. Adjusting the focus on human variation. *Trends Genet.* **16**, 296–302 (2000).
55. Harpending, H. & Rogers, A. Genetic perspectives on human origins and differentiation. *Annu. Rev. Genomics Hum. Genet.* **1**, 361–385 (2000).
56. Poloni, E. S. *et al.* Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am. J. Hum. Genet.* **61**, 1015–1035 (1997).
57. Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in humans. *Nature Genet.* **20**, 278–280 (1998).
58. Mountain, J. L. Molecular evolution and modern human origins. *Evol. Anthropol.* **7**, 21–37 (1998).
An informative review of molecular studies of modern human origins.
59. Briscoe, D., Stephens, J. C. & O'Brien, S. J. Linkage disequilibrium in admixed populations: applications in gene mapping. *J. Hered.* **85**, 59–63 (1994).
60. Lahr, M. M. & Foley, R. A. Towards a theory of modern human origins: geography, demography and diversity on recent human evolution. *Yb. Phys. Anthropol.* **41**, 137–176 (1998).
An important discussion of how historical climatic changes in Africa have influenced the emergence of modern humans and variation in modern African populations.
61. Scozzari, R. *et al.* Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am. J. Hum. Genet.* **65**, 829–846 (1999).
62. Passarino, G. *et al.* Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am. J. Hum. Genet.* **62**, 420–434 (1998).
63. Soodyall, H., Vigilant, L., Hill, A. V., Stoneking, M. & Jenkins, T. mtDNA control-region sequence variation suggests multiple independent origins of an 'Asian-specific' 9-bp deletion in sub-Saharan Africans. *Am. J. Hum. Genet.* **58**, 595–608 (1996).
64. Thomas, M. G. *et al.* Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba — the 'Black Jews of Southern Africa'. *Am. J. Hum. Genet.* **66**, 674–686 (2000).
65. Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
66. Kruglyak, L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* **22**, 139–144 (1999).
67. Pritchard, J. K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**, 1–14 (2001).
A comprehensive review of models of LD and the impact of demographic history and selection on levels and patterns of LD.
68. Peterson, R. J., Goldman, D. & Long, J. C. Effects of worldwide population subdivision on *ALDH2* linkage disequilibrium. *Genome Res.* **9**, 844–852 (1999).
69. Mateu, E. *et al.* Worldwide genetic analysis of the *CFTR* region. *Am. J. Hum. Genet.* **68**, 103–117 (2001).
70. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. High-resolution haplotype structure in the human genome. *Nature Genet.* **29**, 229–232 (2001).
71. Zhu, X. *et al.* Localization of a small genomic region associated with elevated ACE. *Am. J. Hum. Genet.* **67**, 1144–1153 (2000).
A demonstration that populations of African descent can be used to fine-map QTLs, where the use of non-African populations had failed to do so as precisely.
72. Weatherall, D. J. & Clegg, J. B. Inherited haemoglobin disorders: an increasing global health problem. *Bull. World Health Organ.* **79**, 704–712 (2001).
73. Cooper, R. *et al.* The prevalence of hypertension in seven populations of west African origin. *Am. J. Public Health* **87**, 160–168 (1997).
74. Kaufman, J. S., Durazo-Arvizu, R. A., Rotimi, C. N., McGee, D. L. & Cooper, R. S. Obesity and hypertension prevalence in populations of African origin. The Investigators of the International Collaborative Study on Hypertension in Blacks. *Epidemiology* **7**, 398–405 (1996).
75. Cooper, R. S. *et al.* Prevalence of NIDDM among populations of the African diaspora. *Diabetes Care* **20**, 343–348 (1997).
76. Rotimi, C. N. *et al.* In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. *Ann. Epidemiol.* **11**, 51–58 (2001).
77. Rotimi, C. N. *et al.* Prevalence of diabetes and impaired glucose tolerance in Nigerians, Jamaicans and US blacks. *Ethnic Dis.* **9**, 190–200 (1999).
78. Colilla, S., Rotimi, C., Cooper, R., Goldberg, J. & Cox, N. Genetic inheritance of body mass index in African-American and African families. *Genet. Epidemiol.* **18**, 360–376 (2000).
79. Rotimi, C. N. *et al.* Distribution of anthropometric variables and the prevalence of obesity in populations of west African origin: the International Collaborative Study on Hypertension in Blacks (ICSHIB). *Obesity Res.* **3** (Suppl. 2), 95–105 (1995).
80. Moore, J. H. & Williams, S. M. New strategies for identifying gene–gene interactions in hypertension. *Ann. Med.* **34**, 88–95 (2002).
81. Templeton, A. R. in *Epistasis and the Evolutionary Process* (eds Wade, M. J., Brodie, E. D. & Wolf, J. B.) 41–57 (Oxford Univ. Press, New York, 2000).
82. Williams, S. M. *et al.* Combinations of variations in multiple genes are associated with hypertension. *Hypertension* **36**, 2–6 (2000).
This study shows the importance of examining several genes simultaneously for association with complex diseases, such as hypertension, even when independent analysis of genes does not indicate an association.
83. Eberhardt, M. S. *et al. Health United States, 2001 with Urban and Rural Health Chartbook* (National Center for Health Statistics, Hyattsville, Maryland, 2001).
84. Williams, D. R. Black–White differences in blood pressure: the role of social factors. *Ethnic Dis.* **2**, 126–141 (1992).
85. James, S. A. in *Hypertension in Blacks: Epidemiology, Pathophysiology and Treatment* (eds Hall, W. D., Saunders, E. & Shulman, N. B.) (Yearbook Medical Publ., Chicago, Illinois, 1985).
86. Rotimi, C. N. *et al.* Maximum-likelihood generalized heritability estimate for blood pressure in Nigerian families. *Hypertension* **33**, 874–878 (1999).
87. Gu, C. *et al.* Familial resemblance for resting blood pressure with particular reference to racial differences: preliminary analyses from the HERITAGE Family Study. *Hum. Biol.* **70**, 77–90 (1998).
88. Unwin, N. *et al.* Noncommunicable diseases in sub-Saharan Africa: where do they feature in the health research agenda? *Bull. World Health Organ.* **79**, 947–953 (2001).
89. Edwards, R. *et al.* Hypertension prevalence and care in an urban and rural area of Tanzania. *J. Hypertens.* **18**, 145–152 (2000).
90. Cooper, R. S. *et al.* Heritability of angiotensin-converting enzyme and angiotensinogen: a comparison of US blacks and Nigerians. *Hypertension* **35**, 1141–1147 (2000).
An important observation that documents the role of environmental variability across populations and among individuals, and its effect on the ability to detect genetic factors.
91. Caulfield, M. *et al.* Linkage of the angiotensinogen gene locus to human essential hypertension in African Caribbeans. *J. Clin. Invest.* **96**, 687–692 (1995).
92. Rotimi, C. *et al.* Angiotensinogen gene in human hypertension. Lack of an association of the 235T allele among African Americans. *Hypertension* **24**, 591–594 (1994).
93. Rotimi, C. *et al.* Hypertension, serum angiotensinogen, and molecular variants of the angiotensinogen gene among Nigerians. *Circulation* **95**, 2348–2350 (1997).

94. Kardia, S. L. Context-dependent genetic effects in hypertension. *Curr. Hypertens. Rep.* **2**, 32–38 (2000).
95. Walker, A. R., Adam, F. & Walker, B. F. World pandemic of obesity: the situation in Southern African populations. *Public Health* **115**, 368–372 (2001).
96. Aspray, T. J. *et al.* Rural and urban differences in diabetes prevalence in Tanzania: the role of obesity, physical inactivity and urban living. *Trans. R. Soc. Trop. Med. Hyg.* **94**, 637–644 (2000).
97. Bray, G. A. Complications of obesity. *Ann. Intern. Med.* **103**, 1052–1062 (1985).
98. Folsom, A. R. *et al.* Implications of obesity for cardiovascular disease in blacks: the CARDIA and ARIC studies. *Am. J. Clin. Nutr.* **53** (Suppl.), 1604–1611 (1991).
99. Burke, G. L. *et al.* Obesity and overweight in young adults: the CARDIA study. *Prev. Med.* **19**, 476–488 (1990).
100. Luke, A. *et al.* Heritability of obesity-related traits among Nigerians, Jamaicans and US black people. *Int. J. Obes. Relat. Metab. Disord.* **25**, 1034–1041 (2001).
101. Parra, E. J. *et al.* Ancestral proportions and admixture dynamics in geographically defined African Americans living in South Carolina. *Am. J. Phys. Anthropol.* **114**, 18–29 (2001).
102. Argyropoulos, G. *et al.* Effects of mutations in the human uncoupling protein 3 gene on the respiratory quotient and fat oxidation in severe obesity and type 2 diabetes. *J. Clin. Invest.* **102**, 1345–1351 (1998).
103. Kimm, S. Y. *et al.* Racial differences in the relation between uncoupling protein genes and resting energy expenditure. *Am. J. Clin. Nutr.* **75**, 714–719 (2002).
104. Yanovski, J. A. *et al.* Associations between uncoupling protein 2, body composition, and resting energy expenditure in lean and obese African American, white, and Asian children. *Am. J. Clin. Nutr.* **71**, 1405–1420 (2000).
105. Siffert, W. *et al.* Worldwide ethnic distribution of the G protein β3 subunit *825T* allele and its association with obesity in Caucasian, Chinese, and Black African individuals. *J. Am. Soc. Nephrol.* **10**, 1921–1930 (1999).
106. Mbanya, J. C., Sobngwi, E. & Mbanya, D. N. *HLA-DRB1*, *-DQA1*, *-DQB1* and *DPB1* susceptibility alleles in Cameroonian type 1 diabetes patients and controls. *Eur. J. Immunogenet.* **28**, 459–462 (2001).
107. Cisse, A. *et al.* Distribution of *HLA-DQA1* and *-DQB1* alleles and *DQA1–DQB1* genotypes among Senegalese patients with insulin-dependent diabetes mellitus. *Tissue Antigens* **47**, 333–337 (1996).
108. Chauffert, M. *et al.* Susceptibility to type 1 diabetes in the Senegalese population is linked to *HLA-DQ* and not *TAP* and *LMP* genes. *Diabetes Care* **20**, 1299–1303 (1997).
109. Garcia-Pacheco, J. M. *et al.* Distribution of *HLA-DQA1*, *-DQB1* and *DRB1* alleles in black IDDM patients and controls from Zimbabwe. *Tissue Antigens* **40**, 145–149 (1992).
110. Pirie, F. J., Hammond, M. G., Motala, A. A. & Omar, M. A. HLA class II antigens in South African Blacks with type I diabetes. *Tissue Antigens* **57**, 348–352 (2001).
111. King, H., Aubert, R. E. & Herman, W. H. Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. *Diabetes Care* **21**, 1414–1431 (1998).
112. Relethford, J. H. *Genetics and the Search for Modern Human Origins* (John Wiley & Sons, New York, 2001).
113. Slatkin, M. Linkage disequilibrium in growing and stable populations. *Genetics* **137**, 331–336 (1994).
114. Benatar, S. R. & Singer, P. A. A new look at international research ethics. *BMJ* **321**, 824–826 (2000).
115. Emanuel, E. J., Wendler, D. & Grady, C. What makes clinical research ethical? *JAMA* **283**, 2701–2711 (2000).
116. Rando, J. C. *et al.* Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Annls Hum. Genet.* **62**, 531–550 (1998).
117. Scozzari, R. *et al.* Differential structuring of human populations for homologous X and Y microsatellite loci. *Am. J. Hum. Genet.* **61**, 719–733 (1997).
118. Spedini, G. *et al.* The peopling of sub-Saharan Africa: the case study of Cameroon. *Am. J. Phys. Anthropol.* **110**, 143–162 (1999).
119. Destro-Bisol, G. *et al.* Microsatellite variation in Central Africa: an analysis of intrapopulational and interpopulational genetic diversity. *Am. J. Phys. Anthropol.* **112**, 319–337 (2000).
120. Spurdle, A. & Jenkins, T. The Y-specific p21A1/TaqI polymorphism occurs in four major population groups. *Hum. Hered.* **43**, 31–34 (1993).
121. Forster, P. *et al.* A short tandem repeat-based phylogeny for the human Y chromosome. *Am. J. Hum. Genet.* **67**, 182–196 (2000).
122. Krings, M. *et al.* mtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? *Am. J. Hum. Genet.* **64**, 1166–1176 (1999).

Acknowledgements

The authors thank the reviewers C. Freund and E. Clayton for useful comments and/or discussion on earlier versions of this paper, and K. Powell and H. Mortensen for assistance with figure 1. S.A.T. was supported by a Burroughs Welcome Fund Career Award, a David and Lucile Packard Career Award and a US National Science Foundation grant. S.M.W. was supported by the US National Institutes of Health.

Online links

DATABASES

The following terms in this article are linked online to:

Cancer.gov: http://www.nci.nih.gov/cancer_information
prostate cancer

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink
ACE | *AGT* | ATRAP | *CD4* | *DM1* | *FY* | *GNB3* | *HBB* | *MC1R* | *PDHA1* | *PLAT* | renin | *UCP2* | *UCP3*

OMIM: http://www.ncbi.nlm.nih.gov/Omim
type 1 diabetes | type 2 diabetes

FURTHER INFORMATION

Ethnologue: http://www.ethnologue.com

dbSNP summary:
http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi

Encyclopedia of Life Sciences
Insulin action: molecular basis of diabetes | Population genetics of modern human evolution

Online Mendelian Inheritance in Man database:
http://www.ncbi.nlm.nih.gov/Omim

Sarah Tishkoff's lab: http://www.life.umd.edu/biology/tishkofflab

Single Nucleotide Polymorphism (SNP) Consortium:
http://snp.cshl.org

WHO report: Services for the prevention and management of genetic disorders and birth defects in developing countries:
http://www.who.int/ncd/hgn/reppub_malta.htm

Access to this interactive links box is free online.