Human Races: Classifying People vs Understanding Diversity

Guido Barbujani*

Dipartimento di Biologia, Università di Ferrara, via L. Borsari 46, I-44100 Ferrara, Italy

Abstract: The idea that all humans naturally belong to one of a few biological types or races that evolved in isolation was unchallenged for centuries, but large-scale modern studies failed to associate racial labels with recognizable genetic clusters. Recently, the conclusions of those studies have been questioned by authors who argue that racial classification has objective scientific bases and is indispensable in epidemiology and genetics. However, no classification is useful if the classification units are vague or controversial, and no consensus was ever reached on the number and definition of the human races. The available studies show that there is geographic structure in human genome diversity, and that it is possible to infer with reasonable accuracy the continent of origin from an individual's multilocus genotype. However, clear-cut genetic boundaries between human groups, which would be necessary to recognise these groups as relatively isolated mating units which zoologists would call races, have not been identified so far. On the contrary, allele frequencies and synthetic descriptors of genetic variation appear distributed in gradients over much of the planet, which points to gene flow, rather than to isolation, as the main evolutionary force shaping human genome diversity. A better understanding of patterns of human diversity and of the underlying evolutionary processes is important for its own sake, but is also indispensable for the development of diagnostic and therapeutic tools designed for the individual genotype, rather than for ill-defined race-specific genotypes.

Key Words: Human diversity, population structure, geographic variation, gene flow, selection, isolation, risk factors.

INTRODUCTION

As of May 12th, 2005, the New York Times forum on human origins (http://forums.nytimes.com/top/opinion/readersopinions/forums/science/index.html) contains 25,341 messages, a large number of them dealing with the existence of human races. The quality of the messages ranges from scientifically sound to deeply prejudiced, but it is undeniable that they show a strong interest, not simply for human evolution in general, but for the race issue in particular. A similar debate has been going on on the columns of major scientific journals [1, 2]. For example, Burchard *et al*. [3] argued that although races have been used for social discrimination, the concept remains indispensable for prediction of an individual's genotype, so much so that progress in genetic research may be hampered by the refusal of racial categorization. By contrast, Cooper *et al*. [4] maintained that racial labels are useless for reliable prediction of health risk because genetic variation is continuous and discordant with racial classifications. Both in the popular and in the specialized press, two separate questions interfere with each other, namely whether humans are naturally subdivided in biological races, and whether it is socially acceptable to speak of races *provided races exist*.

These are different, if related, questions, and only the former can be addressed in quantitative, reproducible terms. Therefore, in this paper I shall try as much as possible to skip the latter, and I shall split the former, which is complicated, in two more easily tractable questions. As we know, people from different places look different and part of these differences are doubtless rooted in the people's genes. Is it accurate, then, to assign individuals to races, thus describing our species as essentially discontinuous at the genetic level, or do schemes of racial classification miss some crucial aspects of human biodiversity? And, if racial classification turns out to be inaccurate, does it contain anyway useful information for practical purposes, or is it better to resort to other concepts, e.g. population?

To address these questions, even unsophisticated approaches seem to work at a superficial level; everybody can tell, say, the average Swede from the average Senegalese. However, this approach, which in its more developed forms is called typologic and has a long history in anthropology [5, 6], does not lead us very far. When the task is not trivial, i.e. when individuals are to be assigned to several potential races, different people reach different conclusions, and social status is known to affect the way people are categorized in races [7]. Therefore, a less simplistic approach is necessary. An additional complication is, extensive migration in the last centuries has relocated many members of our species and so, just by looking at them, nobody can tell with certainty, say, a New Yorker from an inhabitant of Johannesburg. Thus, to look for races it is necessary to focus on the fraction of humankind that has been affected only mildly by recent migration and admixture [8]. But ultimately, if human races exist, and if one wants to use races to predict health risks [3], it must be possible: (a) to agree on a race list; (b) to place races on the world's map, and (c) to associate each race with diagnostic alleles or haplotypes.

In the first section of this paper I shall discuss a few definitions of race. The second section will review the main historical attempts to identify the human races. Modern meth-

*Address correspondence to this author at the Dipartimento di Biologia, Università di Ferrara, via L. Borsari 46, I-44100 Ferrara, Italy; Tel: +39 0532 291312; Fax: +39 0532 249761; E-mail: g.barbujani@unife.it

1389-2029/05 $50.00+.00

ods to understand whether proposed races are also distinct genetically, and the results of their application, are outlined in the three following sections, dealing, respectively, with descriptive approaches, estimations of genetic variances, and inferences of population structure. After a section contrasting the roles of subdivision and isolation by distance in shaping human biodiversity, I shall come back to the question whether racial classification may be useful for some practical purposes. The final sections highlight some priorities for future research in human biodiversity.

WHAT IS A RACE

The concept of race is somewhat elusive [see Refs. 9 and 10], and has long been. In one of the clearest attempts to define it, introducing geographic variation in his classical text *Systematics and the Origin of Species* [11], Ernst Mayr makes a distinction between species in which biological changes from population to population are continuous, and species in which groups of populations with different character combinations are separated by borders. In the latter, the entities separated by borders are subspecies or geographic races. Similarly, in classical human genetic or physical anthropology textbooks races are envisaged as large populations of individuals who evolved together, share a significant fraction of their genes, and hence can be distinguished from other races by their common gene pool [12] or by different alleles fixed in each [13]. Under both definitions, races are necessarily separated by borders of increased biological variation.

A schematic representation of alternative patterns of genetic variation in a species is in Fig. **1**. In principle, variation among individuals for *n* traits can be represented in an *n*-dimensional space, in which differences for each trait are represented along one of the *n* axes. That is complicated, and hence it is common to project phenotypic or genotypic variation in a two-dimensional space, so that each individual's genotype or phenotype is envisaged as a point in a plane. In species whose groups are well subdivided genetically and separated by borders, there will be little ambiguity as for the group each individual belongs to (left side of Fig. **1**). On the contrary, if group definition is based on criteria other than phenetic or genetic, cultural for example, the groups will overlap in the genotypic/phenotypic space and, as a consequence, many phenotypes or genotypes will fall in areas corresponding to two or more groups (right side of Fig. **1**). In the former case, the variances among groups will be large with respect to the variances between individuals of the same group, in the latter case they will be small. "Small" and "large" are rather vague concepts, but, if necessary, we shall resort to more specific definitions later.

Classical population-genetics theory and empirical data show that large genetic differences among groups develop if reproductive barriers separate these groups. In the presence of reproductive barriers, that is, under isolation, genetic drift, affecting independently each group, will reduce the group's internal variation (because the alleles that are lost are not reintroduced by gene flow) and will lead groups to diverge from each other (because different alleles are lost in each group and different mutations occur in each group) [14]; divergence will generate genetic discontinuities, i.e. boundaries, between groups. Conversely, when groups or populations exchange substantial numbers of migrants, the effects of divergence will be opposed by those of gene flow. As a consequence, genetic variation between populations and groups will tend to be continuous, and zones of abrupt genetic change will be unlikely [15, 16]. Therefore, the borders mentioned in Mayr's definition can be regarded as a result of independent genetic drift in groups that separated early in evolution, afterwards connected by little (or no) migration for much of their history.

To see if such borders exist in the geographic distribution of human genome variation, in principle we can choose among an estimated 10 million single-nucleotide polymorphisms (SNPs) [17]. In practice, we can only study a fraction of them, and it is not clear which genome regions would be best suited for a description of global human biodiversity. It

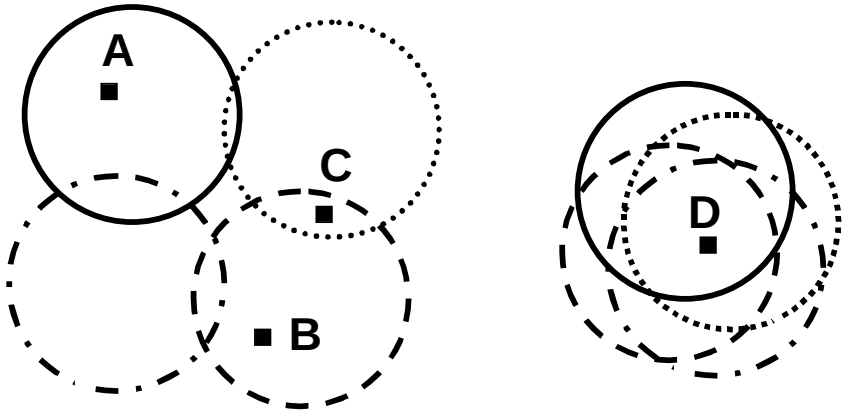

Fig. (1). Schematic representation of population groups as circles in a phenotypic or genotypic space. On the left, a species subdivided in distinct (possibly racial) groups. In this case, for most individuals there is no ambiguity as for the group they belong to; A and B are clearly part of the groups represented by a continuous (A) or a dashed (B) line, and only for C there is uncertainty. On the right, a species whose groups are not distinct biologically. The phenotype or genotype of most individuals, such as D, is compatible with assignment with more than one group (in fact, in this example, to all groups).

is clear, however, that genetic similarity at one or a few loci does not always imply shared ancestry, because even distantly related populations may be similar for certain traits if they evolved under the same selective pressures. Based on skin colour, possibly the most popular phenotype for folk racial classification, populations of sub-Saharan Africa, Southern India, Australia and Melanesia cluster together, yet these populations occur in distant branches of evolutionary trees based on their genes [18, 19]. Their common, dark skin color reflects evolutionary convergence, due to the common selection against light skin color in distantly-related populations occupying areas of intense UV radiation [20, 21]. The main molecular target of selection was probably the melanocortin 1 receptor gene (MC1R), for which variation in light-skinned populations has been interpreted as evidence that loss of pigment increases fitness in areas of low solar radiation [21], or simply that the selective pressure decreased when humans expanded from Africa into milder climates [22]. Thus, although skin color is genetically transmitted, skin color similarity is a poor indicator of shared ancestry [23], as are as other traits that evolved under convergent selection.

These observations imply that a detailed understanding of the evolutionary relationships and subdivision among populations can only be based on traits whose variation reflects the consequences of genome-wide (genetic drift and migration) rather than locus-specific (selection) evolutionary pressures [14, 24]. Therefore, the ideal data are a vast collection of allele frequencies and DNA sequences from worldwide-distributed individuals. However, human races have been the subject of intense scientific investigation long before genetics existed

LISTS OF RACES

Even early theorists of human classification noticed that morphological traits alone do not lead to a stable classification of humans in discrete groups. Thus, in most cases, races were defined by a combination of morphological as well as geographic, social and cultural criteria. Constant in those studies was the idea that human racial groups had long evolved independently from each other, as made explicit, for instance, by Coon [25]. As a consequence, a few major biological clusters were expected to exist, each corresponding to a morphological type dwelling in a specific area of the world. Members of these typological races would be similar to each other, and different from members of other groups.

A more detailed description of the rationale of racial classification systems, and of the related debate on monogenism versus polygenism (in today's terms, mono- or polyphyletic origin of humans) is in Cohen [26]. For a comparison with diversity in other species of large-bodied mammals, see Templeton [27]. According to some modern authors [see e.g. Refs. 28-30], the classical analyses of morphological traits, such as skeletal measures or skin color, suggest a clear racial subdivision. On the contrary, a review of the relevant literature shows that that is not the case [26, 31]. Indeed, although until 1962 nobody explicitly raised doubts on the existence of biological races in humans [32], studies of human morphology from Linnaeus to current times reached no consensus on which races exist, and which populations belong to which race. Lists compiled by serious scientists include anything between three and 200 different races [33], and it is impossible for me to identify in these lists anything that can be called 'common concepts' of race. The admittedly incomplete scheme of Table **1** is based on Refs. [8, 26, 29, 32]. In summary, human types can be, and have been, proposed and described, but to be useful, any classification ought to be based on non-arbitrary classification units into which objects (human individuals in our case) can be placed with minimum ambiguity. By contrast, different studies of human morphology failed to reach an agreement on the number and definition of the human races.

Table 1. Lists of Human Races

Author	No. of races	Races proposed
Linnaeus (1735)	6	Europaeus, Asiaticus, Afer, Americanus, Ferus, Monstruosus
Buffon (1749)	6	Laplander, Tartar, South Asian, European, Ethiopian, American
Blumenbach (1795)	5	Caucasian, Mongolian, Ethiopian, American, Malay
Cuvier (1828)	3	Caucasoid, Negroid, Mongoloid
Deniker (1900)	29	
Weinert (1935)	17	
Von Eickstedt (1937)	38	
Biasutti (1959)	53	
Coon (1962)	5	Congoid, Capoid, Caucasoid, Mongoloid, Australoid
US Office of Management and Budget (1997)	5	African-American, White, American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander
Risch et al. (2002) Fig. 1	5	African, Caucasian, Pacific islanders, East Asian, Native American
Risch et al. (2002) Table 3	5	African Americans, Caucasians, Hispanic Americans, East Asians, Native Americans

ARE RACES A BIOLOGICAL REALITY? DIAGNOSTIC ALLELES

There are many ways to test whether putative races correspond to genetically distinct sets of individuals; the simplest require no assumptions, and cluster genotypes on the basis of their similarity. By these approaches one can represent genetic relationships among individuals and populations through evolutionary trees [19], or by means of bidimensional projections generated by techniques of multivariate analysis.

Another simple approach is looking for diagnostic alleles by which individuals might be assigned to groups. Typically, these are rare, population-specific alleles associated with disease, and that is a problem, because pathologic alleles are diagnostic in a sense, but they are present in just a few individuals. Consider Tay-Sachs disease, whose incidence is notoriously high, with a carrier frequency around 1/26 [34] among Ashkenazi Jews, versus 1/145 to 1/166 in other populations of European descent [35]. It is possible to define the Ashkenazi Jews as those with the highest frequency of the Tay-Sachs allele [1], although as other populations, e.g. the Irish [35] show comparable carrier frequencies. But, more to the point, 25/26 Ashkenazi Jews do not carry the Tay-Sachs allele, and hence in this way one defines a set of subjects at risk, not a race.

Alleles related with skin colour do not seem to be diagnostic either. Aside from the issue of convergent evolution, discussed elsewhere in this paper, apparently skin colour is too complicated a trait to be a valid, general predictor of genomic ancestry. Although one major gene affecting this trait, MC1R, has been identified, the effects of other genes can obscure its effects, to the point that individuals with identical MC1R genotype show very different skin and hair colours [21]. It comes as no surprise that when measures of skin pigmentation were compared with AIMs (Ancestry Informative Markers, i.e. panels of alleles [36, 37] showing large frequency differences between populations of different continents) a correlation was found in the US whites and blacks [38], but not in rural Brazilian populations [39].

These and other descriptive approaches comparing alleles in individuals or populations lack statistical power, and only seldom do they lead to statistical tests able to discriminate between alternative hypotheses. If one wants to explicitly test genetic hypotheses, as we do, it seems necessary to resort to assumptions, either on the putative races to be compared, or on Hardy-Weinberg equilibrium in groups that the numerical analyses will then define. Such model-based approaches fall in two classes, namely those based on the estimation of genetic variances among predefined groups, and those aimed at inferring population structure from genetic data.

ARE RACES A BIOLOGICAL REALITY? GENETIC VARIANCES

Under this approach, one starts from a list of races, gathers information on several genes of their putative members, and estimates measures of genetic diversity. The global variance observed is eventually apportioned at three levels by comparing the variances between: (i) individuals of the same population, (ii) populations of the same race, and (iii) different races.

Lewontin [40] quantified genetic variances by estimating the Shannon information measure, *H*, which is similar to the gene diversity index [41]. Initially, three values of *H* were calculated, respectively around the average allele frequencies of each population (*H*pop), the average allele frequency across populations of a race (*H*race), and the average allele frequency for the entire species (*H*species). The global genetic diversity was then partitioned at the three levels as follows:

Variance within populations = *H*pop / *H*species

Variance between populations, within races = (*H*race – *H*pop) / *H*species

(the effect of belonging to different populations of the same race)

Variance between races = (*H*species – *H*race) / *H*species.

(the effect of belonging to different races)

With the advent of techniques for the direct study of DNA, it became possible to incorporate in the genetic variances a measure of molecular differentiation between alleles, such as the number of substitutions (for sequence polymorphisms), or allele-length differences (for microsatellites). Much like in Lewontin's approach, AMOVA, a nonparametric method for the analysis of variance suitable for molecular data [42], subdivides genetic diversity into three hierarchical components: between individuals within populations, between populations of the same group, and between groups. The significance of each component is then tested nonparametrically by reassigning each individual or population to a random geographic location, according to three resampling schemes. The molecular variances are recalculated, and the procedure is repeated many times, so as to obtain empirical null distributions for the three variances, against which the observed values are finally compared.

The actual values of these variances do not contain much useful information, because they also depend on the level of genetic polymorphism at the loci considered, and on the population samples available. That is why it is convenient to report, rather than the actual variances, the proportion of the total variance observed at each level.

In the earlier studies, seven-race [41-43] or three-race classification systems [44] were chosen, whereas more recent analyses compared continents. In all cases, individual differences between members of the same population accounted for about 85% of the global human variance; belonging to different populations added between 3 and 8% to that value, and to different races or continents between 6 and 11% (Table **2**). These proportions, inferred from protein variation, and hence from coding regions whose variation may differ from that observed at random genome locations, appeared at first counterintuitive. However, successive analyses of DNA polymorphisms, whether SNP, microsatellite, or insertion/deletion polymorphisms, both for coding and non-coding genome regions, confirmed them with remarkable precision [45-50], especially the finding that each population harbours a large share, on average close to 85%, of the global human diversity (Table **2**).

Table 2. Estimated Fractions of the Global Human Diversity at Three Hierarchical Levels of Population Subdivision.

Polymorphism	Reference	N of loci	Within population	Between populations, within race or continent	Between races or continents
Proteins[a]	40	17	85.4	8.3	6.3
Proteins[a]	43	18	85.5	5.5	9.0
Proteins[b]	44	25	86.0	2.8	11.2
mtDNA	42	(HV-I)[c]	75.4	3.5	21.1
Autosomal DNA	45	109	84.4	4.7	10.8
mtDNA	46	(HV-I and –II)[c]	81.4	6.1	12.5
Y chromosome	46	10	35.5	11.8	52.7
Autosomal DNA	47	90	84.8	1.6	13.6
mtDNA	47	(HV-I)[c]	71.5	6.1	23.4
Y chromosome	47	10	83.3	18.5	-1.8[f]
Alu insertions	48	21	82.9	8.2	8.9
Y chromosome	48	14	42.6	17.3	40.1
Beta-globin	48	1	79.4	2.8	17.8
Autosomal DNA	49	377	94.1	2.4	3.6
Median, all loci[d]			82.5	5.8	11.7
Median, autosomal[e]			86.2	3.8	10.0

[a] Seven races considered

[b] Three races considered

[c] HV-I and HV-II are the hypervariable regions I and II, respectively, of the mitochondrial genome control region.

[d] This value was obtained by considering all studies equally informative, calculating the median among their results, and normalizing it by dividing by 1.007 Because Ref. 19 contains roughly as many loci as all other combined studies, the median estimated giving equal weight to each locus would be very close to the upper bound of the range of within-population variance, and to the lower bounds of the between-population and between-continent variances in that study, respectively 93.8, 2.5, 3.9.

[e] Normalization obtained by dividing by 0.987

[f] Because two variances are obtained by subtraction, it may occasionally happen that one of them takes a negative value. This value (which, however, is not significantly different from 0) means that, on average, members of different continents do not differ for those markers more than members of the same continent.

Higher diversity between continents was observed for polymorphisms mapping on the Y chromosome (with one exception; Table **2**). There is no simple mathematic relationship linking the population sizes with the value of the genetic variance between populations, but the impact of genetic drift is inversely related to the effective population size. Because only one Y-chromosome (and one copy of the mitochondrial DNA, hereafter mtDNA) is potentially transmitted by a couple to their offspring, as opposed to four copies of each autosome, the effective population size for the Y-chromosome and for mtDNA is one-fourth of the autosomal population size. Therefore, one expects to find higher variances between populations from the analysis of uniparentally-transmitted traits. In this sense, what is surprising of these results, and still calls for an explanation, is not the fact that variances between continents are high for the Y chromosome, but that they are not for mtDNA. Possible additional factors causing differences between the patterns shown by the Y-chromosome and mtDNA include reduced male mobility [46], polygyny [51], the tendency of some mtDNA lineages to evolve faster than others [52], uncertainties on mitochondrial mutation patterns and rates [53], and systematic errors due to a biased ascertainment of polymorphic DNA sites [54].

On the other hand, the lowest variances between continents were observed in Rosenberg *et al*.'s study of 377 autosomal microsatellite polymorphisms [49], typed in the cell lines of the CEPH (Centre pour l'Etude de Polymorphisme Humaine at the Foundation Jean Dausset, Paris) [55]. However, using a different, and more appropriate, mutational model, diversity in the CEPH data appeared distributed much like in the previous studies [50]. Other Figs., estimated from the same data, show an important implication of these findings. Overall, 4199 different alleles are documented in the 377 loci of Ref. 43. Of those alleles, 66% are shared at least by Africa, Asia and Europe, whereas only 7.4% are continent-specific [49]. The small fraction of continent-specific alleles is in agreement with the results of two large-scale studies based high-throughput methods of genotyping. Gabriel *et al*. [56] sequenced 1.5 million bases of DNA in African, Asian and European individuals for a study of haplotype blocks, i.e. 20 to 50 kb regions of the genome that appear mildly affected, if at all, by historical recombination.

Less than 2% of haplotype blocks appeared restricted to Asia, 2% appeared restricted to Europe, 25% were Africa-specific, and the rest were shared among continents, with more than 50% occurring worldwide. Hinds *et al*. [57] genotyped 1,586,383 SNPs in US individuals of African, Asian and European descent. 'Private' SNPs, i.e. alleles restricted to one population, were 14% in African-Americans, and between 2% and 3% in Asian-Americans and European-Americans, so that more than 80% of the alleles appeared cosmopolitan or shared between continents. Therefore, fixed differences, even between geographically remote populations, appear extremely rare in the human genome. Note that by sampling a few individuals from distant locations one tends to overestimate the variances between continents; individuals from intermediate regions would probably show intermediate genetic features and increase the fraction of shared alleles (see also Ref. 30). Despite that, the study based on the largest collection of loci so far [49] and those entailing the broadest genotyping efforts so far [56, 57], concur in indicating that extensive allele and haplotype sharing across continents are the rule, not the exception.

Coming back to the scheme of Fig. **1**, if we want to represent the human species as a set of groups in the phenotypic or genotypic space, differences among these groups account, on average, for 10% of the total, and diversity within each group represents some 85% of the total. Thus, it seems that the terms "large" and "small" that we used previously are sufficient to describe these results; variation between groups is much smaller than variation within groups, and there is no doubt that a sensible description of human diversity is in the right part of Fig. **1**.

The small variances between populations do not mean that all populations are equal, but imply that any clustering of populations will be based on small genetic differences. Using eight different methods of assignment on five datasets representing population from all over the world, genotypes whose origin was temporarily disregarded could be allocated to the right continent with an error ≥30% using autosomal markers, and ≥27% using the Y-chromosome [48]. Predictably, the precision of the assignment decreased at the subcontinental level.

The effects of the low, but nonzero, variances between continents and populations are evident in many human evolutionary trees. In a study of 80 independent DNA loci, seven clusters of populations were found [58], which overlap, but only in part, with those identified in comparable studies [59], including those described in the following section [16, 48, 49, 60], but do not correspond to any of the racial classification systems of table **1**. In trees describing the relationships between individual genotypes, based on shared STR alleles [16] or on sequence comparisons between individuals typed for 14.4 kilobases at the angiotensinogen gene (AGT) [61, 62] there is a general, if imperfect, tendency of individuals of similar geographic origin to cluster together. Globally, these results mean that knowing where an individual comes from tells us something on that individual's genes, and people coming from the same place have a higher chance to share the same alleles than people from other places. This finding is not new; in classical protein studies, with sufficient data, significant differences could be found between even adjacent pairs of populations [63-65]. However, the Yanomama [63], the Sardinians [64] and the Basques [65] are regarded as genetic isolates, for which nobody so far has proposed the status of races. Similarly, local reproductive isolation and endogamy, such as those common among Hindu castes, have been shown to cause substantial DNA differences even among groups dwelling in the same town [66]. But studies of protein or DNA polymorphisms have not shown so far is that there are clear-cut geographic discontinuities in the distribution of human genome diversity, and that clusters found for one set of markers will stay the same when different markers are considered.

ARE RACES A BIOLOGICAL REALITY? INFERRING POPULATION STRUCTURE

This approach differs from the previous one, in that groups or races are not assumed from the start, but inferred from the pattern of genetic variation in the data. A likelihood-based approach, implemented in the software package Structure, was used in several recent studies [67]. No particular mutational model is assumed. Each individual's genotype is regarded as a mixture of contributions originating in k population clusters (I shall use the term 'clusters' for the sake of clarity, although in the original paper these clusters are referred to as populations), and $q(i)$ is the fraction of the genes of that individual that come from the i-th cluster. The analysis is run for different values of k, whose likelihood is eventually compared. Once k has been estimated or somehow defined, each individual's genotype is supposed to have been generated by drawing alleles, in different proportions, from the k clusters. Under the assumption that each of the populations is in Hardy-Weinberg equilibrium, allele frequencies are estimated for each cluster by a Monte Carlo-Markov Chain algorithm, and the vector $q(1), q(2)…q(k)$ is then evaluated for each individual, regardless of her/his geographical provenance. In this way, ultimately one can estimate the probability of every individual genotype to belong to each of the inferred clusters.

Analysing by Structure two datasets of Alu insertions, Romualdi *et al*. found evidence for three (two worldwide distributed, one Eurasian) and four (African-Oceanian, Asian-American, European, Eurasian) non-overlapping clusters, respectively [48]. The distribution of X-chromosome microsatellites showed four clusters (Africa, Western Eurasia, China, New Guinea, with most Ethiopians falling in the second cluster) which are uncorrelated with variation at six autosomal loci of pharmacogenetic relevance typed in the same individuals [60]. In two other studies [49, 68] apparently the most likely k was not estimated from the data, but rather given arbitrary values between 2 and 6. With k=6 Rosenberg *et al*. [49] found genetic clusters corresponding to (1) Africa, (2) Europe, Western Asia and part of Central Asia; (3) the Kalash of Pakistan; (4) East Asia and part of Central Asia; (5) Oceania; and (6) South America. Bamshad *et al*. [68] observed a separation between Africa and Eurasia with k=2, a split between Asia and Europe with k=3, and two African clusters with k=4, confirming that variation within Africa exceeds that among other continents [57, 69, 70].

To summarize, almost each of the studies mentioned in this section was based on a different collection of populations and individuals, and hence it is difficult to jointly inter-

pret their results. All of these studies identified a geographical structure, but each time a different structure, with different genetic datasets suggesting different clusterings of the individuals [48, 68]. The results of Ref. 49 have been interpreted as showing that there are six major genetic groups in humankind corresponding to common notions of races [71] and hence that self-reported ancestry can facilitate the assessment of epidemiological risk, but that does not seem the case, for three reasons. Firstly, the studied subjects were not asked to self-report their ancestry, and there is no guarantee that they regard themselves as external researchers do. Secondly, the six clusters do not correspond to any previously proposed racial classification (Table **1**). Thirdly, the clusters found in other studies are different [30, 48, 60]. Neither the studies apportioning diversity nor those describing population structure have made it possible: (a) to agree on a race list; (b) to place races on the world's map, and (c) to associate each race with diagnostic alleles or haplotypes. Despite extensive use of the race concept in various fields of medicine (see e.g. Refs. 72-74), the claim that clusters inferred from genetic data coincide closely with groups defined by self-identified racial or continental ancestry [29] is in open contradiction with the available evidence.

SUBDIVISION AND ISOLATION BY DISTANCE

The above conclusions may seem at odds with the fact that large differences are consistently observed among ethnic groups in the US, or comparing samples coming both from the US and from other locations in Africa and Eurasia [3, 8, 37, 68, 71]. However, the US ethnic groups cannot possibly be regarded as representative samples of the populations from the five continents. Indeed, their relationships with their European, African, Asian and native American ancestors are often unclear, and the way people are classified in the US gives little consideration to admixture, so that, just to mention an example, children of a Norwegian and of a Nigerian are classified as African-Americans [10]. In addition, many US studies include Hispanics [8], a problematic category that does not fit any traditional or common definition of race [26], comprising people whose mother tongue is Spanish or Portuguese. Hispanics are individuals with extremely variable physical aspect and skin colour, whose genomes are often mosaics of contributions originating in two or three continents, who are identified as Hispanics only in the country where they immigrated in the last decades, and who would never define themselves as such in their place of origin. In short, no serious conclusion about human biodiversity can be drawn from studies including Hispanics [8, 37], and other samples of recent immigrants can be considered only if there is evidence that they represent one specific community, and not a heterogeneous assemblage of many. Of course, in some contexts it makes sense to compare the genetic features of groups labelled on the basis of their language or of other non-biological features of theirs, for instance for forensic identification purposes. However, these data provide no information on the global structure of the human population.

The second reason why differences among US groups do not suggest the existence of deep subdivision among the main human groups is that, as we saw in the section on genetic variances, allele-frequency differences exist between all populations, including communities separated by short geographic distances, or by cultural barriers at geographical distance zero [66]. With large sample sizes, these differences reach statistical significance. Clearly, recent immigrating groups coming together from different continents are going to have different allele frequencies, and there is no doubt about that. The question is not whether human group A can somehow be shown to differ from group B, because the answer is invariably positive. The question is whether these differences are distributed continuously or discontinuously or, in other words, whether humans show the biological subdivision that would justify clustering of its members into discrete races.

To find a convincing answer, one must study population samples from all over the world. In the broadest meta-analysis so far, multilocus allele frequencies, summarised by means of synthetic components and plotted on the map, showed continuous, gradual change in all continents [75]. Recently, Serre and Pääbo [76] addressed the same question at the DNA level by reanalysing the 377 loci typed in Ref. 49. They resampled random individuals from that dataset, so as to eliminate as far as possible the effect of the clustering of genotypes in populations that is likely to result from the sampling of geographically discrete populations. They confirmed that the direct analysis of discrete samples tends to suggest the existence of clusters, roughly corresponding to continents. However, these clusters disappeared when the unit of analysis became the randomly-sampled individual, indicating that discontinuities are probably an artefact due to the discontinuous design of that and other studies. Therefore, assigning individual genotypes to groups apparently conceals the continuous nature of human diversity and entails a high degree of arbitrariness, although local differences exist and may be of evolutionary or medical significance.

Continuous genetic variation in the geographic space, such as observed in Refs. 73, 74, and in several papers cited therein, is the expected product of isolation by distance [15], loosely defined as the process whereby gene flow between communities is inversely related with their geographic distance. Because drift affects populations independently, but spatially close populations tend to exchange more genes than distant populations, the overall effect of isolation by distance is a decline of genetic similarity at increasing spatial distances. On the contrary, if there are reproductive barriers, genetic differences will not be simply proportional to geographic distances, and patterns of genetic variation will not fit models of isolation by distance. In other words, population subdivision and isolation by distance are alternative models, only the former causing the onset of genetic boundaries between populations [16]. John Relethford applied a formal model of isolation by distance for the analysis of classical protein polymorphisms, microsatellite DNA markers and craniometric measures taken from worldwide-distributed samples. An excellent fit was observed for the three classes of data, suggesting that patterns of human diversity can largely be accounted for by the simple interaction of drift and geographically-structured gene flow [76]. Jointly taken, these results confirm the existence of geographic structuring in the human genome, and the absence of clear boundaries suggests that, as a rule, migrational networks were more important than reproductive barriers in determining the observed patterns of human biodiversity.

IS RACIAL CLASSIFICATION USEFUL

Our species is comparatively young, and its current diversity can be traced back to one or a few founder populations that expanded from Africa not long ago [78, 79]. Under that scenario, and given the evidence for extensive gene flow [80, 81] it is unsurprising that, as we saw in the previous sections, genetic differences between continents be small, geographic variation be continuous, and populations that appear to form a cluster when studied for certain markers do not cluster together when analyzed for other markers. Human population had too little time, and too intense migrational exchanges, to develop racial differences. Therefore, we can define groups based on the physical aspect of people, but the alleles found in these groups are not reliable predictors of variation at other independently-transmitted loci [see especially Ref. 60]. By studying genotypes we can identify with good approximation the geographic origin of most individuals, but humans do not come in neat racial packages. If races are defined as entities separated by boundaries [11], characterized by common gene pools [12] and fixed allelic differences [13], there is no such thing in humans.

Perhaps things will change when we have more data on genome diversity. However, increases in our knowledge of variation will result in a greater discriminating power only if there is something to discriminate. At present, the labels used to classify people do not appear to reflect a recognizable biological subdivision of our species. Should these results be confirmed, no technical progress will be able to provide us with a biologically stable classification of human groups.

Is there any practical reason to maintain some kind of racial classification, then, even though its biological meaning is, at best, unclear? There seem to be three arguments in favour of this view. One is that races are important for gene hunting because alleles that cause monogenic disorders are enriched in, for example, Mexicans in Texas, Ashkenazi Jews and the inhabitants of Tristan da Cunha [1]. However, none of these populations can possibly enjoy the status of a race, according to any of the definitions listed at the beginning of this manuscript. They are populations, a concept of population is all we need to study their pathologic alleles, and the presence of these alleles cannot dictate medical treatment for an entire continent, whose inhabitants will mostly carry the common alleles [82]. Few multigenic diseases are well understood at the genetic level, but some populations can conceivably carry certain combinations of predisposing alleles in linkage disequilibrium. However, once again, we have no evidence that these combinations characterise broad groups which could possibly be called races, whereas there is reason to believe they tend to occur in specific genetic isolates [83]. The same seems the case for allele frequencies at loci of pharmacogenomic interest, whose extreme values are observed in groups such as "Japanese" or "Mediterranean people" [84] that were never proposed as races, to the best of my knowledge. However, the clearest demonstration of the ambiguous relationships between racial classification and disease comes from Ioannidis *et al*.'s [85] meta-analysis of studies of association between disease and genes in different groups that the authors of the original papers defined as 'racial'. More than half, 43 of the 83 studies considered, showed overall significant differences for at least one of the continental groups considered, but the relative risks, expressed as odds ratios, differed significantly among continents in a much smaller fraction of cases, 14%. Ioannidis and collaborators interpret this finding as a consequence of the fact that not only genetic variation is greater within than between groups, but so are additional factors of risk, such as those related with lifestyle and environment [85]. In summary, it seems that the word race is and has been used loosely by authors who really meant populations. In this case the problem is only semantic and can easily be fixed, perhaps resorting to the expression "genetically differentiated populations" when necessary.

The second argument is that, by and large, commonly used racial labels reflect the underlying genetic structure of humankind [75]. Therefore, skin color and body measures would allow inferences on the likely allelic state at untyped genes [2, 8, 71]. Clearly,10% difference between continents (Table **1**), less than 5% continent-specific polymorphisms in Asia or Europe [56, 57], lack of correlation between skin color and African ancestry [39] and inconsistent clusterings of populations inferred from different genes [48, 49, 60, 68] leave little hope of accuracy. Predicting allelic state at a locus based on variation at other genotypic or phenotypic traits seems more complicated and error-prone than actually typing the locus of interest. But an additional, and so far irresolvable, problem is that no "commonly used racial labels" exist, because scientists never agreed on those labels (Table **1**), and because people are classified differently in different cultures [86]. Whereas in the USA Japanese and Chinese people would be considered to be "Asian", in apartheid South Africa the former were considered to be white and the latter Asian [87]. In Japan, the majority of the population considers the ethnic group burakumin as biologically distinct from them, but in the USA both the Japanese majority and minority groups would be considered part of the same race. [88]. The Bribri of Costa Rica refer to themselves as *bribri*, which means "the people", whereas all others are *ña*, which has a very negative connotation (L. Madrigal, pers. comm.). That classification is subjective, of course, but not necessarily more so are than the many and conflicting Western classifications.

If races don't exist, why are forensic scientists so good at finding them?, asked Sauer [89], and this is the third argument in favour of a racial classification. By considering separate databases for different ethnic groups, forensic scientists minimize the probability of unfair decisions against members of minorities [90]. However, this seems yet another case in which people say race but mean something else. If races are biological realities, i.e., if they are subspecies, they must be the same everywhere, whereas forensic race catalogs differ across countries. In the UK we find White-skinned European, Afro-Caribbean, Indian Subcontinent, South East Asian and Middle Eastern: (www.forensic.gov.uk/forensic/foi/ foi_docs/43L_Commonplace_characteristics.pdf), only two of which (the 1st and the 4th) correspond to races in the USA (Table **1**). Which list is right? My answer is that neither gives a sensible, all-purpose description of human diversity, but both can work if one is to categorize people in specific urban areas, where boundaries between groups are sharper than at the world scale and certain groups are underrepresented or absent altogether. Once again, a concept of popu-

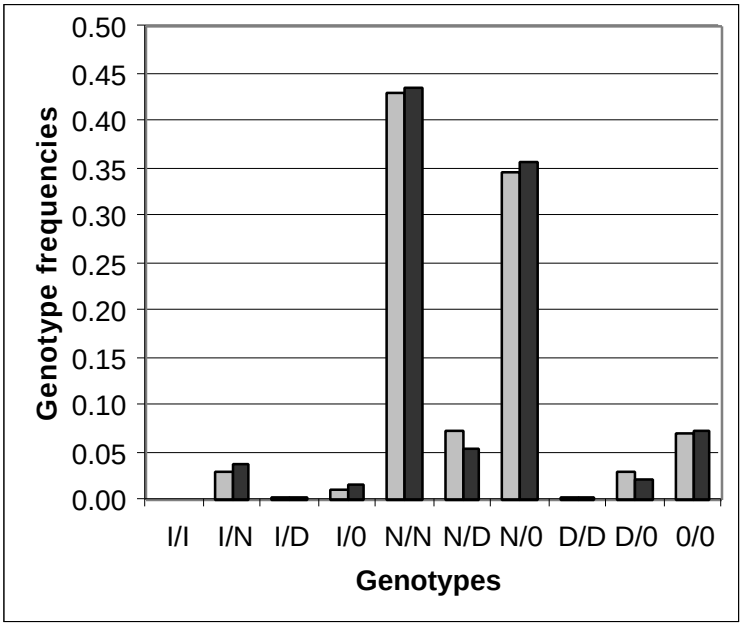

Fig. (2). Frequencies of CYP2D6 genotypes in samples of Asians (grey bars, N=1352) and Europeans (black bars, N= 7602) [91]. Genotypes are defined in terms of alleles increasing the metabolic rate (I), decreasing it (D), associated with normal metabolic rates (N), or with no enzyme production (0).

lation diversity seems all that is needed for forensic identification.

Contrary to the claim that racial stereotypes capture some meaningful aspects of biological variation, the available data suggest that a good way to predict whether certain individuals will have certain health risks or will benefit from pharmaceutical treatment is to study their genes. For example, Fig. **2** shows the distribution of the genotypes of CYP2D6, a locus coding for one of the main drug-metabolizing enzymes of the cytochrome, in two large samples of Asians and Europeans [91]. Depending on the genotype, people benefit from standard treatment (normal metabolizers), have side-effects (slow metabolizers), or eliminate the drug before it has exerted its action (fast metabolizers). Genotype frequencies are almost identical in the two groups, but this not the main point here. Even if allele frequencies differed, it is clear that both samples contain the whole spectrum of genotypes, and hence in both populations there will be fast, normal and slow metabolizers. That leaves little hope to develop different drugs, or drug dosages, specific for the Asian or the European market. Much more productive, and now technically conceivable, is concentrating the efforts on genetic typing of individuals, which in turn may lead to tailoring specific pharmacological treatment for different classes of metabolizers [92], no matter where in the world they live.

WHAT NEXT

As usual, we need more data. High-throughput methods for the rapid characterization of thousands of polymorphisms have opened unprecedented opportunities to better understand human genome diversity. However, to be informative, these studies must be conducted according to criteria that permit comparison and cross-validation of data. This means that a relatively small number of markers, typed in a globally-distributed sample, will be more informative than more markers, typed in a few regions of the world or, even worse, in samples of immigrants of different origins collected in Europe or in the US. The CEPH diversity cell lines have proved to be of crucial importance for that purpose. As time passes, and more loci are being typed, the possibilities increase to achieve a comprehensive picture of our genome diversity. However, the CEPH samples do not include any populations from India, nor from other areas of the world of crucial evolutionary importance. Extending the sampling is thus an important research priority [10].

Once a broader collection of samples is available, covering in greater detail India, Northern Asia and the Americas, it will be important to look for genetic boundaries at the appropriate geographical scale, testing whether the locations suggested by most genes coincide. Should that prove the case, then the observation that certain genes are similar (or dissimilar) in two groups of populations could be taken as suggesting that other genes, including those of medical interest, are likely to be similar (or dissimilar) between those groups. Conversely, should that not prove to be the case, as most available studies indicate, better medical care should be sought by concentrating efforts on rapid and efficient genotyping of individuals, rather than on predictions based on skin colour, self-assessed ethnicity and other inaccurate proxies of genetic relatedness.

The study of world-distributed samples is also necessary to begin to understand the genetic bases of human phenotypic variation, both morphological and in disease susceptibility. We are still far from developing theoretical frame-

works enabling us to deal with the complexity of the multifactorial, polygenic transmission mechanisms [93] of these traits. However, a few major loci involved in the inheritance of complex traits have been identified (MC1R among them, Ref. 21) and more will be in the future. Analyses of the global patterns of variation for these genes will represent an indispensable step towards a deeper understanding of phenotypic diversity, a theme on which important insight is also being provided by studies of animals and plants [94].

CONCLUDING REMARKS

Species that are subdivided in essentially isolated reproductive units, such as some bats [95] tend to form races, species where gene flow prevails such as the coyote [96] tend to show continuous variation. A number of intermediate possibilities exist and are documented in the animal kingdom, but apparently humans are closer to coyotes than to bats, whereas other large-bodied mammals (including chimpanzee and bonobo [97, 98], Grant's gazelle [99], elephant [27] and roe deer [100]) have higher inter-population variances than humans, although their habitat is much more restricted.

Still, race is a social reality, and as such it will continue affecting our life. Racial categorization has a long history, and may be related to a deep-rooted psychological need to quickly identify potential enemies and allies [101]. However, the biological reality is different and, for humans, it is one of continuous variation [75], clines, and genetic boundaries that cross the geographic space without surrounding and thus defining specific isolated groups of populations [102]. If we are to understand human diversity, and if we are to exploit the potential represented by the ever-increasing genomic data for mapping and cloning of disease-causing genes [103], race is neither an accurate nor a useful concept, unless it is used in such a loose sense as to mean population, in which case a rigorous usage of words is advisable. Both if we want to understand human evolution, and if our aim is to treat multifactorial pathologies, there is no alternative to coming to terms with the continuous nature of human genetic diversity.

ACKNOWLEDGEMENTS

This paper benefited from extensive discussion with David Balding and, indirectly, Anthony Edwards (who do not share most of my views), Lorena Madrigal, who suggested a number of changes and provided precious guidance with the anthropological literature, an anonymous referee, Elise Belle, Laurent Excoffier, David Goldstein, and Giorgio Bertorelle. I thank them all.

REFERENCES

[1] Phimister, E.G. Medicine and the racial divide. *New Engl. J. Med.* **2003,** *348***,** 1081-1082.

[2] Collins, F.S. What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era. *Nature Genet.* **2004,** *36***,** S13-S15.

[3] Burchard, E.G.; Ziv, E;, Coyle, N.; Gomez, S.L.; Tang, H.; Karter, A.J.; Mountain, J.L.; Perez-Stable, E.J.; Sheppard, D and Risch, N. The Importance of race and ethnic background in biomedical research and clinical practice. *New Engl. J. Med.* **2003,** *348***,** 1170-1175.

[4] Cooper, R.S.; Kaufman, J.S. and Ward R. Race and genomics. *New Engl. J. Med.* **2003,** *348***,** 1166-1170.

[5] Relethford, J.H. *The Human Species. An Introduction to Biological Anthropology*., 4th ed. McGraw-Hill, New York, 2001,

[6] Washburn, S.L. The study of race. *Am. Anthropol.* **1962,** *65***,** 521-531.

[7] Weeks, M. and Lupfer, M.B. Complicating race: the relationship between prejudice, race, and social class categorizations. *Pers. Soc. Psychol. Bull*. **2004,** *30***,** 972-984.

[8] Risch, N.; Burchard, E.; Ziv, E.; Tang, H. Categorization of humans in biomedical research: Genes, race and disease. *Genome Biol.* **2002,** *3***,** 2007.1-2007.11.

[9] Keita, S.O.Y.; Kittles, R.A.; Royal, C.D.M.; Bonney, G.E.; Furbert-Harris, P.; Dunston, G.M. and Rotimi, C.N. Conceptualizing human variation.. *Nature Genet.* **2004,** *36***,** 517-520.

[10] Kittles, R.A. and Weiss, K.M. Race, ancestry and genes: Implications for defining disease risk. *Annu. Rev. Genomics Hum. Genet*. **2003,** *4***,** 33–67.

[11] Mayr, E. *Systematics and the Origin of Species*, 3rd edn, Columbia University Press, New York, 1947, 98-106.

[12] Vogel, F. and Motulsky, A.G. *Human Genetics: Problems and Approaches*, 2nd edn, Springer-Verlag, Berlin, 1986, 534.

[13] Boyd, WC. *Genetics and the Races of Man: An Introduction to Modern Physical Anthropology*. Little Brown, Boston, 1950,

[14] Goldstein, D.B. and Chikhi, L. Human migrations and population structure: what we know and why it matters. *Annu. Rev. Genomics Hum. Genet.* **2002,** *3***,** 129-152.

[15] Wright, S. Isolation by distance. *Genetics* **1943,** *28***,** 139-156.

[16] Barbujani, G.; Oden, N.L. and Sokal, R.R. Detecting areas of abrupt change in maps of biological variables. *Syst. Zool*. **1989,** *38***,** 376-389.

[17] Kruglyak L; Nickerson, D. Variation is the spice of life. *Nature Genet*. **2001,** *27***,** 234-236.

[18] Cavalli-Sforza, L.L.; Menozzi, P; Piazza, A; Mountain, J. Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* **1988,** *85***,** 6002-6006.

[19] Bowcock, A.M; Ruiz-Linares, A.; Tomfohrde, J; Minch, E; Kidd, J.R. and Cavalli-Sforza L.L. .High-resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **1994,** *368***,** 455-457

[20] Jablonski, N.G. and Chaplin, G. The evolution of human skin coloration. *J. Hum. Evol.* **2000,** *39***,** 57-106.

[21] Rees, J.R. Genetics of hair and skin color. *Annu. Rev. Genet*. **2003,** *37***,** 67–90.

[22] Harding, R.M.; Healy, E.; Ray, A.J.; Ellis, N.S.; Flanagan, N.; Todd, C.; Dixon, C.; Sajantila, A.; Jackson, I.J.; Birch-Machin, M.A.; Rees, J.L. Evidence for variable selective pressures at MC1R. *Am. J. Hum. Genet.* **2000,** *66***,** 1351-1361.

[23] Relethford, J.H. Apportionment of global human genetic diversity based on craniometrics and skin color *Am. J. Phys. Anthropol.* **2002,** *118***,** 393-398.

[24] Luikart, G.; England, P.R.; Tallmon, D.; Jordan, S. and Taberlet, P. The power and promise of population genomics: From genotyping to genome typing. *Nature Rev. Genet.* **2003,** *4***,** 981-994.

[25] Coon, C.S. *The origins of races.* Knopf, New York, 1962,

[26] Cohen, C. Les races humaines en histoire des sciences. In: *Aux Origines d'Homo sapiens*, **1991,** ed.s. Hublin J.J. and Tillier, A.M., Presses Universitaires de France, Paris, 9-56.

[27] Templeton, A. Human races: A genetic and evolutionary perspective. *Am. Anthropol.* **1999,** *100***,** 632-650.

[28] Miller, E.M. Review of History and Geography of Human Genes, by Cavalli-Sforza L.L. *et al*. *Mankind Quart.* **1994,** *35***,** 71-108.

[29] Mountain, J.L. and Risch, N. Assessing genetic contributions to phenotypic differences among 'racial' and 'ethnic' groups. *Nature Genet.* **2004,** *36***,** S48-S53.

[30] Tishkoff, S. and Kidd, K.K. Implications of biogeography of human populations for 'race' and medicine. *Naure Genet*. **2004,** *36***,** S21-S27.

[31] Brown, R.A. and Armelagos, G.J. Apportionment of racial diversity: A review. *Evol. Anthropol.* **2001,** *10***,** 24-40.

[32] Livingstone, F.B. On the nonexistence of human races. *Curr. Anthropol*. **1962,** *3***,** 279-281

[33] Armelagos, G.J. Racism and physical anthropology : Brues' review of Barkan's The Retreat of Scientific racism. *Am. J. Phys. Anthropol.* **1994,** *93***,** 381-383.

[34] Charrow, J.L. Ashkenazi Jewish genetic disorders. *Familial Cancer* **2004,** *3***,** 201-206.

[35] Branda, K.J.; Tomczak, J. and Natowicz, M.R. Heterozygosity for Tay-Sachs and Sandhoff diseases in non-Jewish Americans with ancestry from Ireland, Great Britain, or Italy. *Genet. Test.* **2004,** *8***,** 174-180.

[36] Parra E.J. ; Marcini, A. ; Akey, J. ; Martinson, J. ; Batzer, M.A. ; Cooper, R. ; Forrester, T. ; Allison, D.B. ; Deka, R. ; Ferrell, R.E. and Shriver, M.D. Estimating African American admixture proportions by use of population-specific alleles. Am. J. Hum. Genet. **1998,** 63, 1839-1851.

[37] Shriver, M.D.; Smith, M.W.; Jin, L.; Marcini, A.; Akey, J.M.; Deka, R and Ferrell, R.E. Ethnic-affiliation estimation by use of population-specific DNA markers. Am. J. Hum. Genet. **1997,** 60, 957-964

[38] Shriver, M.D.; Parra, E.J.; Dios, S.; Bonilla, C.; Norton, H.; Jovel, C.; Pfaff, C.; Jones, C.; Massac, A.; Cameron, N.; Baron, A.; Jackson, T.; Argyropoulos, G.; Jin, L.; Hoggart, C.J.; McKeigue, P.M. and Kittles, R.A. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum. Genet. **2003,** 112, 387-399

[39] Parra, F.C ; Amado, R.C.; Lambertucci, J.R.; Rocha, J.; Antunes, C.M. and Pena, S.D.J. Color and genomic ancestry in Brazilians. Proc. Natl. Acad. Sci. USA **2003,** 100, 177-182.

[40] Lewontin, R.C. The apportionment of human diversity. *Evol. Biol.* **1972,** *6***,** 381-398.

[41] Nei, M. and Kumar, S. Molecular evolution and phylogenetics Oxford University Press, Oxford, 2003,

[42] Excoffier, L.; Smouse, P. and Quattro, J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **1992,** *131***,** 479–491.

[43] Latter, B.D.H. Genetic differences within and between populations of the major human subgroups. *Am. Nat.* **1980,** *116***,** 220-237.

[44] Ryman, N.; Chakraborty, R. and Nei, M. Differences in the relative distribution of human gene diversity between electrophoretic and red and white cell antigen loci. *Hum. Hered.* **1983,** *33***,** 93-102.

[45] Barbujani, G; Magagni, A.; Minch, E; Cavalli-Sforza, L.L. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **1997,** *94***,** 4516–4519.

[46] Seielstad, M.T.; Minch, E. and Cavalli-Sforza, L.L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **1998,** *20***,** 278-280.

[47] Jorde, L.B. ; Watkins, W.S. ; Bamshad, M.J. ; Dixon, M.E. ; Ricker, C.E. ; Seielstad, M.T. and Batzer, M.A. The distribution of human genetic diversity: A comparison of mitochondrial, autosomal and Y-chromosome data. *Am. J. Hum. Genet.* **2000,** *66***,** 979-988.

[48] Romualdi, C.; Balding, D.; Nasidze, I.S.; Risch, G.; Robichaux, M.; Sherry, S.T.; Stoneking, M.; Batzer, M.A. and Barbujani, G. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.* **2002,** *12***,** 602-612.

[49] Rosenberg, N.A.; Pritchard, J.K. ; Weber, J.L. ; Cann, H.M. ; Kidd, K.K. ; Zhivotovsky, L.A. and Feldman, M.W. Genetic structure of human populations. *Science* **2002,** *298***,** 2381-2385.

[50] Excoffier, L. and Hamilton, G. Comment on "Genetic structure of human populations". *Science* **2003,** *300***,** 1877.

[51] Dupanloup, I.; Pereira, L.; Bertorelle G.; Calafell F.; Prata, M.J.; Amorim, A. and Barbujani, G. A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *J. Mol. Evol.* **2003,** *57***,** 85-97.

[52] Torroni, A.; Rengo, C.; Guida, V.; Cruciani, F.; Sellitto, D.; Coppa, A.; Calderon, F.L.; Simionati, B.; Valle, G.; Richards, M.; Macaulay, V. and Scozzari, R. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am. J. Hum. Genet.*. **2001,** *69***,** 1348-1356.

[53] Pakendorf, B. and Stoneking, M. Mitochondrial DNA and human evolution. *Annu.. Rev. Genomics Hum. Genet.* **2005,** *6***,** 165-183.

[54] Wilder, J.A.; Kingan, S.B.; Mobasher, Z.; Pilkington, M.M. and Hammer, M.F. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nature Genet.* **2004,** *36***,** 1122-1125.

[55] Cavalli-Sforza, L.L. The human genome diversity project: past, present and future. *Nature Rev. Genet.* **2005,** *6***,** 333-339.

[56] Gabriel, S.B.; Schaffner, S.F.; Nguyen, H.; Moore, J.M.; Roy, J.; Blumenstiel, B.; Higgins, J.; DeFelice, M.; Lochner, A.; Faggart, M.; Liu-Cordero, S.N.; Rotimi, C.; Adeyemo, A.; Cooper, R.; Ward, R; Lander, E.S.; Daly, M.J. and Altshuler, D. The structure of haplotype blocks in the human genome. *Science* **2002,** *296***,** 2225-2229.

[57] Hinds, D.A.; Stuve, L.L.; Nilsen, G.B.; Halperine, E.; Eskin, E.; Ballinger, D.G.; Frazer, K.A. and Cox D.R. Whole-genome patterns of common DNA variation in three human populations. *Science* **2005,** *307***,** 1072-1079.

[58] Kidd, K.K.; Pakstis, A.J.; Speed, W.C. and Kidd, J.R. Understanding human DNA sequence variation. *J. Hered.* **2004,** *95***,** 406-420.

[59] 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. and Jorde, L.B. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Res*. **2003,** *13***,**1607-1618.

[60] Wilson, J.F.; Weale, M.E.; Smith, A.C.; Gratrix, F.; Fletcher, B.; Thomas, M.G.; Bradman, N. and Goldstein, D.B. Population genetic structure of variable drug response. *Nat. Genet.* **2001,** *29***,** 265–269

[61] Jorde, L.B. and Wooding, S.P. Genetic variation, classification and 'race'. *Nature Genet.* **2004,** *36***,** S28-S33.

[62] Nakajima, T., Wooding, S., Sakagami, T., Emi, M., Tokunaga, K., Tamiya, G., Ishigami, T., Umemura, S., Munkhbat, B., Jin, F., Guan-Jun, J., Hayasaka, I., Ishida, T., Saitou, N., Pavelka, K., Lalouel, J.M., Jorde, L.B. and Inoue, I. Natural selection and population history in the human angiotensinogen gene AGT,: 736 complete AGT sequences in chromosomes from around the world. *Am. J. Hum. Genet.* **2004,** *74***,** 898-916.

[63] Ward R.H.; Gershowitz, H.; Layrisse, M. and Neel, J.V. The genetic structure of a tribal population, the Yanomama Indians XI. Gene frequencies for 10 blood groups and the ABH-Le secretor traits in the Yanomama and their neighbors; the uniqueness of the tribe. *Am. J. Hum. Genet.* **1975,** *27***,** 1-30.

[64] Barbujani, G. and Sokal, R.R. Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. *Am. J. Hum. Genet.* **1991,** *48***,** 398-411

[65] Bertranpetit, J. and Cavalli-Sforza, L.L. A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann. Hum. Genet.* **1991,** *55***,** 51-67.

[66] Wooding, S.; Ostler, C.; Prasad, B.V.; Watkins, W.S.; Sung, S.; Bamshad, M. and Jorde, L.B.. Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data. *Hum. Genet.* **2004,** *115***,** 221-229.

[67] Pritchard, J.K.; Stephens, M and Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000,** *155***,** 945-959.

[68] Bamshad, M.J.; Wooding, S.; Watkins, W.S.; Ostler, C.T.; Batzer, M.A. and Jorde, L.B. Human population genetic structure and inference of group membership. *Am. J. Hum. Genet.* **2003,** *72***,** 578-589.

[69] Yu, N.; Chen, F.C.; Ota, S.; Jorde, L.B.; Pamilo, P.; Patthy, L.; Ramsay, M.; Jenkins, T.; Shyue, S.K. and Li, W.H. Larger genetic differences within africans than between Africans and Eurasians. *Genetics* **2002,** *161***,** 269-274.

[70] Zietkiewicz, E.; Yotova, V.; Gehl, D.; Wambach, T.; Arrieta, I.; Batzer, M.; Cole, D.E.; Hechtman, P.; Kaplan, F.; Modiano, D.; Moisan, J.P.; Michalski, R. and Labuda, D. Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. *Am. J. Hum. Genet.* **2003,** *73***,** 994-1015.

[71] Bamshad, M.; Wooding, S.; Salisbury, B.A. and Stephens, J.C. Deconstructing the relationship between genetics and race. *Nature Rev. Genet.* **2004,** *5***,** 598-609

[72] Kimm, S.Y.; Glynn, N.W.; Aston, C.E.; Damcott, C.M. ; Poehlman, E.T.; Daniels, S.R. and Ferrell, R.E.. Racial differences in the relation between uncoupling protein genes and resting energy expenditure. *Am. J. Clin. Nutr.* **2002,** *75***,** 714–719.

[73] Acheson, L.S. Bone density and the risk of fractures: should treatment thresholds vary by race? *JAMA* **2005,** *293***,** 2151-2154.

[74] Patel, D.A.; Barnholtz-Sloan, J.S.; Patel, M.K.; Malone, J.M., Jr; Chuba, P.J. and Schwartz, K. A population-based study of racial and ethnic differences in survival among women with invasive cervical cancer: Analysis of Surveillance, Epidemiology, and End Results data. *Gynecol. Oncol.* **2005,** *97***,** 550-558.

[75] Cavalli-Sforza, L.L.; Menozzi, P.; Piazza, A. *The history and geography of human genes.* Princeton University Press, Princeton, 1994,

[76] Serre, D. and Pääbo, S. Evidence for gradients of human genetic diversity within and among continents. *Genome Res.* **2004,** *14***,** 1679-1685.

[77] Relethford, J.H. Global patterns of isolation by distance based on genetic and morphological data. *Hum. Biol.* **2004,** *76***,** 499-513.

[78] Tishkoff, S.A. and Williams S.M. **2002,** Genetic analysis of African populations: Human evolution and complex disease. *Nature Rev. Genet. 3***,** 611-621

[79] Harding, R.M. and McVean, G. A structured ancestral population for the evolution of modern humans. *Curr. Opin. Genet. Dev.* **2004,** *14***,** 667-674.

[80] Barbujani, G. and Goldstein, D.B. Africans and Asians abroad: genetic diversity in Europe. *Annu. Rev. Genomics Hum. Genet.* **2004,** *5***,** 119-150.

[81] Lahr, M.M. and Foley, R.A. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am. J. Phys. Anthropol.* **1998,** Suppl. *27***,** 137-176.

[82] King, M.C. and Motulsky, A.G. Human genetics. Mapping human history. *Science* **2002,** *298***,** 2342-2343.

[83] Arcos-Burgos, M. and Muenke, M. Genetics of population isolates. *Clin. Genet* **2002,** *61***,** 233–247

[84] Weber, W.W. Populations and genetic polymorphisms. *Mol. Diagn.* **1999,** *4***,** 299-307.

[85] Ioannidis, J.P.A.; Ntzani, E.E.; Trikalinos, T.A. 'Racial' differences in genetic effects for complex diseases. *Nature Genet.* **2004,** *36***,** 1312-1318.

[86] Foster, M.W. and Sharp, R.R. Race, ethnicity, and genomics: Social classifications as proxies of biological heterogeneities. *Genome Res.* **2002,** *12***,** 844-850.

[87] Ember, C. R. and Ember, M. *Cultural Anthropology*. 10th edition. Prentice Hall, New York, 2002,

[88] Kottak,C.P. *Anthropology: The Exploration of Human Diversity* 10th ed McGraw-Hill. New York, 2004,

[89] Sauer, N.J. Forensic anthropology and the concept of race: If races don't exist, why are forensic anthropologists so good at identifying them? *Soc. Sci. Med.* **1992,** *34***,** 107-111

[90] Jobling, M.A; Hurles, M.E. and Tyler-Smith, C. *Human Evolutionary Genetics*, **2004,** pp. 473-497, Garland Science

[91] Shimizu, T.; Ochiai, H.; Asell, F.; Yokono, Y.; Kikuchi, Y.; Nitta, M.; Hama Y.; Yamaguchi, S.; Hashimoto, M.; Taki, K.; Nakata, K.; Aida, Y.; Ohashi, A. and Ozawa, N. Bioinformatics research on inter-racial differences in drug metabolism. I. Analysis on frequencies of mutant alleles and poor metabolizers on CYP2D6 and CYP2C19. *Drug Metab. Pharmacokin.* **2003,** *18***,** 48-70

[92] Weber, W.W. Effect of pharamacogenetics on medicine. *Environ. Mol. Mutagen.* **2001,** *37***,** 179-184.

[93] Hirschhorn, J.N. and Daly, M.J.. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **2005,** *6***,** 95-108.

[94] Juenger, T.; Perez-Perez, J.M., Bernal, S and Micol, J.L.. Quantitative trait loci mapping of floral and leaf morphology traits in Arabidopsis thaliana: evidence for modular genetic architecture. *Evol. Dev.* **2005,** *7***,** 259-271.

[95] Miller-Butterworth, C.M.; Jacobs, D.S. and Harley, E.H. Strong population substructure is correlated with morphology and ecology in a migratory bat. *Nature* **2003,** *424***,** 187-191

[96] Vila, C.; Amorim, I.R.; Leonard, J.A.; Posada, D.; Castroviejo, J.; Petrucci-Fonseca, F.; Crandall, K.A.; Ellegren, H. and Wayne, R.K. Mitochondrial DNA phylogeography and population history of the grey wolf *canis lupus*. *Mol. Ecol.* **1999,** *8***,** 2089-103.

[97] Gagneux, P.; Wills, C.; Gerloff, U.; Tautz, D.; Morin, P.A.; Boesch, C.; Fruth, B.; Hohmann, G.; Ryder, O.A. and Woodruff, D.S. Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc. Natl. Acad. Sci. USA.* **1999,** *96***,** 5077-5082.

[98] Enard, W. and Pääbo, S. Comparative primate genomics. *Annu. Rev. Genomics Hum. Genet.* **2004,** *5***,** 351-378.

[99] Arctander, P.; Kat, P.W.; Aman, R.A. and Siegismund, H.R.. Extreme genetic differences among populations of Gazella granti, Grant's gazelle in Kenya. *Heredity* **1996,** *76***,** 465-475.

[100] Vernesi, C.; Pecchioli, E.; Caramelli, D.; Tiedemann, R.; Randi, E. and Bertorelle, G. The genetic structure of natural and reintroduced roe deer Capreolus capreolus, populations in the Alps and central Italy, with reference to the mitochondrial DNA phylogeography of Europe. *Mol. Ecol.* **2002,** *11***,** 1285-1297.

[101] Kurzban, R.; Tooby, J. and Cosmides, L.R. Can race be erased? Coalitional computation and social categorization. *Proc. Natl. Acad. Sci. USA* **2001,** *98***,** 15387-15392.

[102] Barbujani, G. and Sokal, R.R. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc. Natl. Acad. Sci. USA* **1990,** *87***,** 1816-1819.

[103] Jorde LB, Watkins WS, Bamshad MJ. Population genomics: a bridge from evolutionary history to genetic medicine. *Hum. Mol. Genet.* **2001,** *10***,** 2199-2207.

[104] Weatherall, D.J. Phenotype-genotype relationships in monogenic disease: Lessons from the thalassaemias. *Nat. Rev. Genet.* **2001,** *2***,** 245-255.