14

The Apportionment of Human Diversity

R. C. LEWONTIN

*Committee on Evolutionary Biology,
University of Chicago,
Chicago, Illinois*

INTRODUCTION

It has always been obvious that organisms vary, even to those pre-Darwinian idealists who saw most individual variation as distorted shadows of an ideal. It has been equally apparent, even to those post-Darwinians for whom variation between individuals is the central fact of evolutionary dynamics, that variation is nodal, that individuals fall in clusters in the space of phenotypic description, and that those clusters, which we call demes, or races, or species, are the outcome of an evolutionary process acting on the individual variation. What has changed during the evolution of scientific thought, and is still changing, is our perception of the relative importance and extent of intragroup as opposed to intergroup variation. These changes have been in part a reflection of the uncovering of new biological facts, but only in part. They have also reflected general sociopolitical biases derived from human social experience and carried over into "scientific" realms. I have discussed elsewhere (Lewontin, 1968) long-term trends in evolutionary doctrine as a reflection of long-term changes in socioeconomic relations, but even in the present era of Darwinism there is considerable diversity of opinion about the amount or importance of intragroup variation as opposed to the variation between races and species. Muller, for example (1950), maintained that for sexually reproducing species, man in particular, there was very little genetic variation within populations and that most men were homozygous for wild-type genes at virtually all their loci. On such a view, the obvious genetical differences in morphological and physiological characters between races are a major component of the total variation within the species.

Dobzhansky, on the other hand (1954) has held the opposite view, that heterozygosity is the rule in sexually reproducing species, and this view carries with it the concomitant that population and racial variations are likely to be less significant in the total species variation.

As long as no objective quantification of genetic variation could be given, the problem of the relative degree of variation within and between groups remained subjective and necessarily was biased in the direction of attaching a great significance to variations between groups. This bias necessarily flows from the process of classification itself, since it is an expression of the perception of group differences. The erection of racial classification in man based upon certain manifest morphological traits gives tremendous emphasis to those characters to which human perceptions are most finely tuned (nose, lip and eye shapes, skin color, hair form and quantity), precisely because they are the characters that men ordinarily use to distinguish individuals. Men will then be keenly aware of men group differences in such characters and will place strong emphasis on their importance in classification. The problem is even more pronounced in the classification of other organisms. All wild mice look alike because we are deprived of our usual visual cues, so small intergroup differences in pelage color are seized upon for subspecific identification. Again this tends to emphasize between-group variation in contrast to individual variation.

In the last five years there has been a revolution in our assessment of inherited variation, as a result of the application of molecular biological techniques to population problems. Chiefly by use of protein electrophoresis, but also by immunological techniques, it has become possible to assess directly and objectively the genetic variation among individuals on a locus by locus basis. The techniques do not depend upon any *a priori* judgments about the significance of the variation, nor upon whether the variation is between individuals or between groups, nor do they depend upon how much or how little variation is actually present (Hubby and Lewontin, 1965). As a result, the original question of how much variation there is within populations has now been resolved. In a variety of species including Drosophila, mice, birds, plants, and man, it is the rule, rather than the exception, that there is genetic variation between individuals within populations. For example, Prakash et al. (1969) found 42% of a random sample of loci to be segregating in populations of *D. pseudoobscura*, producing an average heterozygosity per locus per individual of 12%. A study of a number of populations of *Mus musculus* by Selander and Yang (1969) gave almost identical results. Two analyses for man, one on enzymes by Harris (1970) and one on blood groups by Lewontin (1967), give respective estimates of 30% and 36% for polymorphic loci within populations, and 6% and 16% for heterozygosity per gene per individual.

The existence of these objective techniques for the assessment of genetic variation, and their widespread application in recent years to large numbers of populations, in conjunction with older information on the distribution of human

blood group genes, makes it possible to estimate, from a random sample of genetic loci, the degree of variation within and between human populations and races, and so to put the comparative differentiation within and between groups on a firm quantitative basis.

THE GENES

Of the 35 or so blood group systems in man, 15 are known to be segregating with an alternative form in frequency greater than 1% in some human populations. (For a summary, see Lewontin, 1967.) Of these, 9 systems have been characterized in enough populations to make them useful for our purposes. They are listed in Table 1 together with the extremes of gene frequency known over the whole range of human populations. I use the concept of "system" rather than "gene" here since it is uncertain whether the MNS system is a single locus with four alleles (as I treat it here) or two closely linked loci with two alleles each. The same ambiguity exists for the Rhesus group, which, again, I treat as a single locus with multiple alleles. For the Rh system, there are many more alleles known than the six listed, but most studies have not had available the full range of antisera, especially anti-D^u, anti-e and anti-d, so that the six classes used here include some confounding of subclasses. All the blood group data upon which the present calculations have been made are taken from Mourant (1954), Mourant et al. (1958), and Boyd (1950).

A second group of loci that have more recently been surveyed are serum proteins and red blood cell enzymes (Table 1). In contrast to the blood groups, which are detected by immune differences, the serum proteins and RBC enzymes are studied by electrophoretic techniques, different alleles producing proteins with altered electrophoretic mobility. A full discussion of these methods is given by Harris (1970), who was the first to use it for population genetic purposes in man; and by Giblett (1969), who also gives extensive information on the distribution of alleles in different human populations. It is from this latter source that the data for this paper are taken.

THE SAMPLES

The amount of world survey work carried out for the different genes obviously varies considerably. For Xm only four populations are reported: a Norwegian, a U.S. white, a U.S. black, and an Easter Island sample; while for the ABO system literally hundreds of populations in all regions of the world had been sampled by the time Mourant's 1954 compilation was made. In the case of the better known blood groups such as ABO, Rh, and MNS, there is an *embarras de richesse*, and some small sample of population is included in the present calculation. Since our object is to look at the distribution of genic diversity

Table 1. Human Genes or "Systems" Included in this Study and Extremes of Allele Frequency in Known Populations

Locus		Allele	Frequency Range	Extreme Populations
Haptoglobin	(Hp)	Hp^1	.09 - .92	Tamils-Lacondon
Lipoprotein	(Ag)	Ag^x	.23 - .74	Italy-India
Lipoprotein	(Lp)	Lp^a	.009- .267	Labrador-Germany
	(Xm)	Xm^a	.260- .335	Easter Is.-U.S. Blacks
Red Cell Acid Phosphatase	(APh)	p^a	.09 - .67	Tristan da Cunha-Athabascan
		p^b	.33 - .91	Athabascan-Tristan da Cunha
		p^c	0- .08	Many
6-phosphogluconate dehydrogenase	(6PGD)	PGD^A	.753-1.000	Bhutan-Yucatan
Phosphoglucomutase	(PGM_1)	PGM_1	.430- .938	Habbana Jews-Yanomama
Adenylate kinase	(AK)	AK^2	0- .130	Africans, Amerinds-Pakistanis
Kidd	(Jk)	JK^a	.310-1.000	Chinese-Dyaks, Eskimo
Duffy	(Fy)	Fy^a	.061-1.000	Bantu-Chenchu, Eskimo
Lewis	(Le)	Le^b	.298- .667	Lapps-Kapinga
Kell	(K)	K	0- .063	Many-Chenchu
Lutheran	(Lu)	Lu^a	0- .086	Many-Brazilian Amerinds
P		P	.179- .838	Chinese-West Africans
MNS		MS	0- .317	Oceanians-Bloods
		Ms	.192- .747	Papuans-Malays
		NS	0- .213	Borneo, Eskimo-Chenchu
		Ns	.051- .645	Navaho-Palauans
Rh		CDe	0- .960	Luo-Papuans
		Cde	0- .166	Many-Chenchu
		cDE	0- .308	Luo, Dyak-Japanese
		cdE	0- .174	Many-Ainu
		cDe	0- .865	Many-Luo
		cde	0- .456	Many-Basques
ABO		I^A	.007- .583	Toba-Bloods
		I^B	0- .297	Amerinds, Austr. Abo.-Toda
		i	.509- .993	Oraon-Toba

throughout the species, I have tried to include what would appear to be *a priori* representatives of the range of human diversity. But how does one do that? Do the French, the Danes, and the Spaniards, say, cover the same range of density as the Ewe, Batutsi, and Luo? How many different European nationalities should be included as compared with how many African peoples or Indian tribes? There is, moreover, the problem of weighting. The population of Japan is vastly larger than the Yanomama tribes of the Orinoco. Should each population be given equal weight, or should some attempt be made to weight each by the proportion of the total species population that it represents? Such weighting would clearly decrease any total measure of human diversity since it would reduce effectively

to zero the contribution of all of the small, isolated and usually genetically divergent groups. It would also decrease the proportion of all human diversity calculated to be between populations, for the same reason. In this paper I have chosen to count each population included as being of equal value and to include, as much as possible, equal numbers of African peoples, European nationalities, Oceanian populations, Asian peoples, and American Indian tribes. Both of these choices will maximize both the total human diversity and the proportion of it that is calculated betweeen populations as opposed to within populations. This bias should be born in mind when interpreting the results.

A second methodological problem arises over the question of racial classification. In addition to estimating the within-and between-population diversity components, I attempt to break down the between-population components into a fraction within and between "races." Despite the objective problems of classification of human population into races, anthropological, genetical, and social practice continues to do so. Racial classification is an attempt to codify what appear to be obvious nodalities in the distribution of human morphological and cultural traits. The difficulty, however, is that despite the undoubted existence of such nodes in the taxonomic space, populations are sprinkled between the nodes so that boundary lines must be arbitrary. No one would confuse a Papuan aboriginal with any South American Indian, yet no one can give an objective criterion for where a dividing line should be drawn in the continuum from South American Indians through Polynesians, Micronesians, Melanesians, to Papuans. The attempts of Boyd (1950) and Mourant (1954) to use blood group data and other genetic information for racial classification illustrate that, no matter what the form of the data, the method of classification remains the same. Obvious and well differentiated stereotypes are set up representing well-differentiated population groups. Thus, the inhabitants of Europe speaking Indo-European languages, the indigenes of sub-Saharan Africa, the aborigines of North and South America, and the peoples of mainland East and Southeast Asia, become the modal groups for Caucasian, Negroid, Amerind, and Mongoloid races. Then by the use of linguistic, morphological, historical, and cultural information, all those not yet included are assorted by affinity into these original classes or, in the case of particularly divergent groups like the Australian aborigines, set up as separate races or subraces. In such a scheme, some populations always create difficulties. Are the Lapps Caucasians or do they belong with the Turkic peoples of Central Asia to the Mongoloid race? Linguistically they are Asians; morphologically they are ambiguous; they have the ABO and Lutheran blood group frequencies typical of Europeans but their Duffy, Lewis, Haptoglobin, and Adenylate-kinase gene frequencies are Asian. Their MNS blood group is clearly non-Asian but also is a very poor fit to European frequencies. Similar great difficulties exist for Hindi-speaking Indians and Urdu-speaking Pakistanis. They are, genetically, the mixture of Aryans, Persians, Arabs, and Dravidians that history tells us they should be.

For the purpose of this paper there are two alternatives. Racial classification could be done entirely from evidence external to the data used here (i.e., linguistic, historical, cultural, and morphological). This convention would then decrease the calculated diversity between races and increase the within-race, between-population component, since it would lump together, in one race, groups that are genetically divergent. The alternative would be to use internal evidence only and establish the racial lines that maximize the similarity of the populations with races. The difficulty of such a procedure is that it has no end. The between-race component would be maximized if every population were made a separate race! Even a reasonable application of this method would require that Indians and Arabs each be made separate races and that Oceania be divided into a number of such groups. I have chosen a conservative path and have used mostly the classical racial groupings with a few switches based on obvious total genetic divergence. Thus, the question I am asking is, "How much of human diversity between populations is accounted for by more or less conventional racial classification?" Table 2 shows the racial classification used in this paper. I have made seven such "races" adding South Asian aborigines and Oceanians to the usual four races, also segregating off the Australian aborigines with the Papuan aborigines. Not all the populations listed under each race are sampled for every gene, but the racial classification was, of course, consistent over all genes.

THE MEASURE OF DIVERSITY

The basic data are the frequencies of alternative alleles at various loci (or supergenes) in different populations. The problem is to use these data to characterize diversity. One ordinarily thinks of some sort of analysis of variance for this purpose, an analysis that would break down genetic variance into a component within population, between populations, and between races. A moment's reflection, however, will reveal that this is an inappropriate technique for dealing with allelic frequencies since, when there are more than two alleles at one locus, there is no single well-ordered variable whose variance can be calculated. If there are two alleles at a locus, say A_1 and A_2, they can be assigned random variable values, say 0 and 1, respectively, and the variance of the numerical random variable could be analyzed within and between populations. If there are three alleles, however, this trick will not work, for if we assigned random variable values, say 0, 1, and 2 to three alleles A_1, A_2, and A_3, we would get the absurd result that a population with equal proportions of A_1 and A_3 would have a greater variance than are those with equal proportions of A_1 and A_2, and A_2 or A_3.

Table 2

Inclusive List of All Populations Used For Any Gene
in this Study by the Racial Classification Used in this Study

Caucasians

Arabs, Armenians, Austrians, Basques, Belgians, Bulgarians, Czechs, Danes, Dutch, Egyptians, English, Estonians, Finns, French, Georgians, Germans, Greeks, Gypsies, Hungarians, Icelanders, Indians (Hindi speaking), Italians, Irani, Norwegians, Oriental Jews, Pakistani (Urdu-speakers), Poles, Portuguese, Russians, Spaniards, Swedes, Swiss, Syrians, Tristan da Cunhans, Welsh

Black Africans

Abyssianians (Amharas), Bantu, Barundi, Batutsi, Bushmen, Congolese, Ewe, Fulani, Gambians, Ghanaians, Hobe, Hottentot, Hututu, Ibo, Iraqi, Kenyans, Kikuyu, Liberians, Luo, Madagascans, Mozambiquans, Msutu, Nigerians, Pygmies, Sengalese, Shona, Somalis, Sudanese, Tanganyikans, Tutsi, Ugandans, U.S. Blacks, "West Africans," Xosa, Zulu

Mongoloids

Ainu, Bhutanese, Bogobos, Bruneians, Buriats, Chinese, Dyaks, Filipinos, Ghashgai, Indonesians, Japanese, Javanese, Kirghiz, Koreans, Lapps, Malayans, Senoy, Siamese, Taiwanese, Tatars, Thais, Turks

South Asian Aborigines

Andamanese, Badagas, Chenchu, Irula, Marathas, Naiars, Oraons, Onge, Tamils, Todas

Amerinds

Alacaluf, Aleuts, Apache, Atacameños, "Athabascans", Ayamara, Bororo, Blackfeet, Bloods, "Brazilian Indians," Chippewa, Caingang, Choco, Coushatta, Cuna, Diegueños, Eskimo, Flathead, Huasteco, Huichol, Ica, Kwakiutl, Labradors, Lacandon, Mapuche, Maya, "Mexican Indians," Navaho, Nez Percé, Paez, Pehuenches, Pueblo, Quechua, Seminole, Shoshone, Toba, Utes, "Venezuelan Indians," Xavante, Yanomama

Oceanians

Admiralty Islanders, Caroline Islanders, Easter Islanders, Ellice Islanders, Fijians, Gilbertese, Guamians, Hawaiians, Kapingas, Maori, Marshallese, Melanauans, "Melanesians," "Micronesians," New Britons, New Caledonians, New Hebrideans, Palauans, Papuans, "Polynesians," Saipanese, Samoans, Solomon Islanders, Tongans, Trukese, Yapese

Australian Aborigines

$$\text{Between populations in races} = \frac{H_{race} - H_{pop}}{H_{species}}$$

$$\text{Between races} = \frac{H_{species} - H_{race}}{H_{species}}$$

Table 4. Proportion of Genetic Diversity Accounted for Within and Between Populations and Races

Gene	Total $H_{species}$	Within Populations	Within Races Between Populations	Between Races
			Proportion	
Hp	.994	.893	.051	.056
Ag	.994	.834	–	–
Lp	.639	.939	–	–
Xm	.869	.997	–	.011
Ap	.989	.927	.062	.067
6PGD	.327	.875	.058	.025
PGM	.758	.942	.033	.131
Ak	.184	.848	.021	.048
Kidd	.977	.741	.211	.048
Duffy	.938	.636	.105	.259
Lewis	.994	.966	.032	.002
Kell	.189	.901	.073	.026
Lutheran	.153	.694	.214	.092
P	1.000	.949	.029	.022
MNS	1.746	.911	.041	.048
Rh	1.900	.674	.073	.253
ABO	1.241	.907	.063	.030
Mean		.854	.083	.063

The results are quite remarkable. The mean proportion of the total species diversity that is contained within populations is 85.4%, with a maximum of 99.7% for the Xm gene, and a minimum of 63.6% for Duffy. Less than 15% of all human genetic diversity is accounted for by differences between human groups! Moreover, the difference between populations within a race accounts for an additional 8.3%, so that only 6.3% is accounted for by racial classification.

This allocation of 85% of human genetic diversity to individual variation within populations is sensitive to the sample of populations considered. As we have several times pointed out, our sample is heavily weighted with "primitive" peoples with small populations, so that their H_O values count much too heavily compared with their proportion in the total human population. Scanning

Table 3 we see that, more often than not, the H_{pop} values are lower for South Asian aborigines, Australian aborigines, Oceanians, and Amerinds than for the three large racial groups. Moreover, the total human diversity, $H_{species}$, is inflated because of the overweighting of these small groups, which tend to have gene frequencies that deviate from the large races. Thus the fraction of diversity within populations is doubly underestimated since the numerator of that fraction is underestimated and the denominator overestimated.

When we consider the remaining diversity, not explained by within-population effects, the allocation to within-race and between-race effects is sensitive to our racial representations. On the one hand the over-representation of aborigines and Oceanians tends to give too much weight to diversity between races. On the other hand, the racial component is underestimated by certain arbitrary lumpings of divergent populations in one race. For example, if the Hindi and Urdù speaking peoples were separated out as a race, and if the Melanesian peoples of the South Asian seas were not lumped with the Oceanians, then the racial component of diversity would be increased. Of course, by assigning each population to separate races we would carry this procedure to the *reductio ad absurdum*. A *post facto* assignment, based on gene frequencies, would also increase the racial component, but if this were carried out objectively it would lump certain Africans with Lapps! Clearly, if we are to assess the meaning of racial classifications in genetic terms, we must concern ourselves with the usual racial divisions. All things considered, then, the 6.3% of human diversity assignable to race is about right, or a slight overestimate considering that H_{pop} is overestimated.

It is clear that our perception of relatively large differences between human races and subgroups, as compared to the variation within these groups, is indeed a biased perception and that, based on randomly chosen genetic differences, human races and populations are remarkably similar to each other, with the largest part by far of human variation being accounted for by the differences between individuals.

Human racial classifcation is of no social value and is positively destructive of social and human relations. Since such racial classification is now seen to be of virtually no genetic or taxonomic significance either, no justification can be offered for its continuance.

REFERENCES

Boyd, W. C. 1950. Genetics and the Races of Man. Boston, D. C. Heath and Co.

Dolanský, L., and M. P. Dolanský. 1952. Table of log$_2$ 1/P, p·log$_2$ 1/p, and p·log$_2$ 1/p + (1-p)·log$_2$ 1/(1-p). Technical Report 227, Research Laboratory of Electronics. Cambridge, Massachusetts Institute of Technology

Dobzhansky, Th. 1954. A review of some fundamental concepts and problems of population genetics. Sympos. Quant. Biol., 20:1-15.