Genetics and "Race"

A Broad Question

- Obviously, humans are genetically diverse
  - Eyes, skin, hair, height, weight, etc., etc.
  - Long-recognized to be highly heritable
- How does this diversity align with "racial" stereotypes?

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

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

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," "Micro-nesians," New Britons, New Caledonians, New Hebrideans, Palauans, Papuans, "Poly-nesians," Saipanese, Samoans, Solomon Islanders, Tongans, Trukese, Yapese

Australian Aborigines

Shannon information measure

$$(2) \qquad H = -\sum_{i=1}^{n} p_i \ln_2 p_i.$$

This latter measure is widely used to characterize species diversity in community ecology, and since I am performing a kind of taxonomic analysis here, I will use H. The calculation of H is somewhat eased by published tables of $p\ln_2 p$ (Dolanský and Dolanský, 1952). In line with our requirements for a

Dolanský, L., and M. P. Dolanský. 1952. Table of $\log_2 1/P$, $p \cdot \log_2 1/p$, and $p \cdot \log_2 1/p + (1-p) \cdot \log_2 1/(1-p)$. Technical Report 227, Research Laboratory of Electronics. Cambridge, Massachusetts Institute of Technology

Shannon information measure

$$H = - \sum_{i}^{n} p_i \log_2 p_i$$

Within populations $\quad = \dfrac{H_{pop}}{H_{species}}$

Between populations in races $\quad = \dfrac{H_{race} - H_{pop}}{H_{species}}$

Between races $\quad = \dfrac{H_{species} - H_{race}}{H_{species}}$

A productive bus ride

"And so I thought, "Well, we've got enough of this data, let's see what it tells us about the differences between human groups." And so I just looked into the literature, and that literature was in books and so on.

And so one day I was going to give a lecture, I think it was in Carbondale, Illinois, or somewhere south. I was working in Chicago at the time. So I took a couple of these books with me and a pad of paper, and a table of logarithms which I needed for this purpose, and a little hand calculator, and I sat on this bus trip for three or four hours looking at the books, picking out the data, looking it up in the table of logarithms, doing a calculation, and writing it down in tables…"
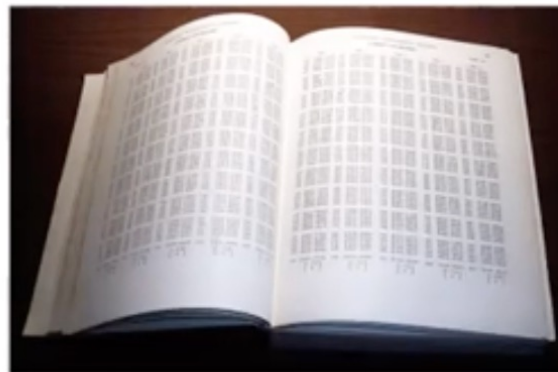
A productive bus ride

"And when I got back after the round trip I had all the data I needed to write the paper about how much human genetic variation there was, and so I did it. And that's been repeated in recent years using DNA and so on. You always get the same result.

Shows you it's worthwhile being afraid to fly, by the way, because you have lots of time on a bus to work."

- Credits to: Andrew Barry, Larry Edelman and the Race: The Power of an Illusion interview transcripts

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

Gene	Total $H_{species}$	Proportion		
		Within Populations	Within Races Between Populations	Between Races
Hp	.994	.893	.051	.056
Ag	.994	.834	—	—
Lp	.639	.939	—	—
Xm	.869	.997	—	—
Ap	.989	.927	.062	.011
6PGD	.327	.875	.058	.067
PGM	.758	.942	.033	.025
Ak	.184	.848	.021	.131
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

Arbitrary choices, enduring results

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.

- Similar results found in:

  - Classical/allozyme markers: Nei and Roychoudhury (1972, 1974); Latter (1973, 1980);

  - HLA loci: Bodmer (1975), Ryman, Chakraborty, Nei (1983)

  - RFLPs and microsatellites: Barbujani et al (1997); Jorde et al (2000), Watkins et al (2003), Rosenberg et al (2002,2003)

  - SNPs: Li et al (2008); Hunley et al (2016), …

  - Previous analyses by Cavalli-Sforza et al (1966) (see Bodmer 2018)

- The 8.3%/6.3% varies across studied but 85%-90% within populations is commonly found

  - Barbujani et al's 84.5 to Lewontin's 85.4: "he got the last two numbers wrong"

The impactful last paragraphs

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 randonly 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.

Fitroy's Apportionment of Diversity

- Edwards quotes Fitzroy:

  - "In the course of years spent in various quarters of the world, I have had opportunities of leisurely considering people from all the principal countries. I have read much of what has been written, during late years, on the subject of their resemblance, or difference; and the conclusion to which I have been obliged to come is – that there is far less difference between most nations, or tribes (selecting any two for the comparison), than exists between two individuals who might be chosen out of either one of those nations or tribes; colour and hair alone excepted."

  - "Fitzroy's acute observations led him to reject the prevalent view expressed in the *Dictionnaire Classique*, but he did not go on to conclude that 'racial classification is now seen to be of virtually no genetic or taxonomic significance …, no justification can be offered for its continuance'[2]. "

  - Point: Edwards does not critique the variance partitioning, it's Lewontin's attack on taxonomic significance that upsets him.

Teaching the apportionment result...

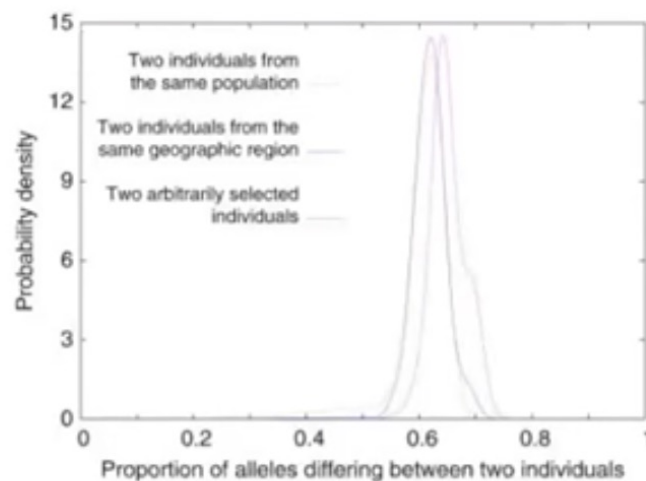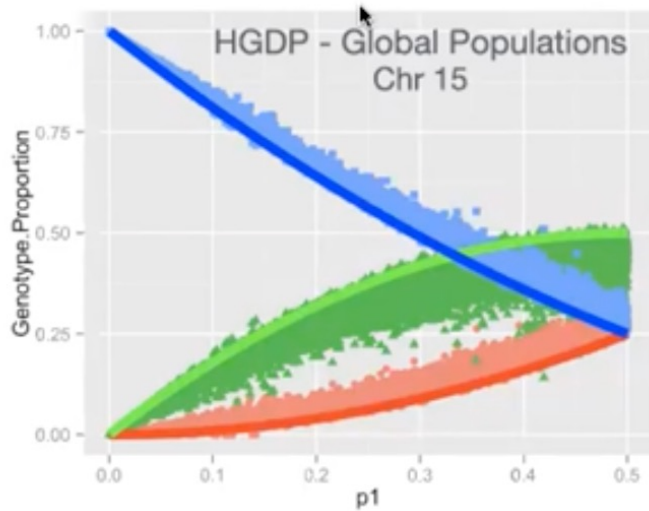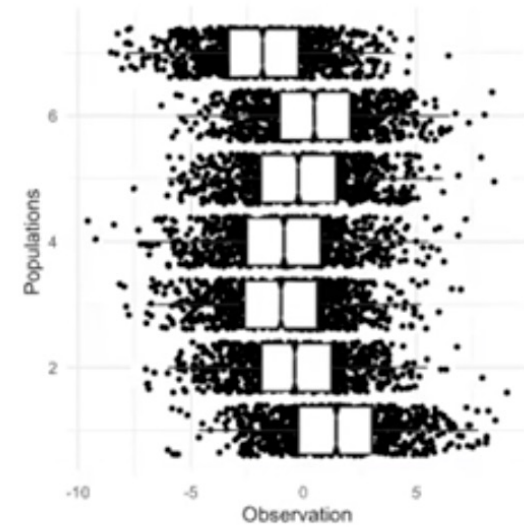MISCONCEPTION:
Human ancestral groups do not share many alleles.
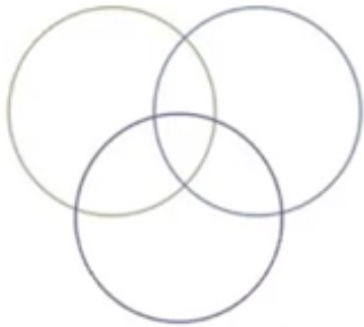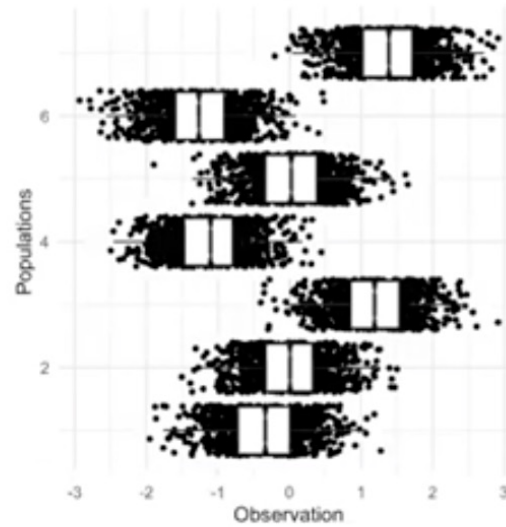
CORRECT MODEL:
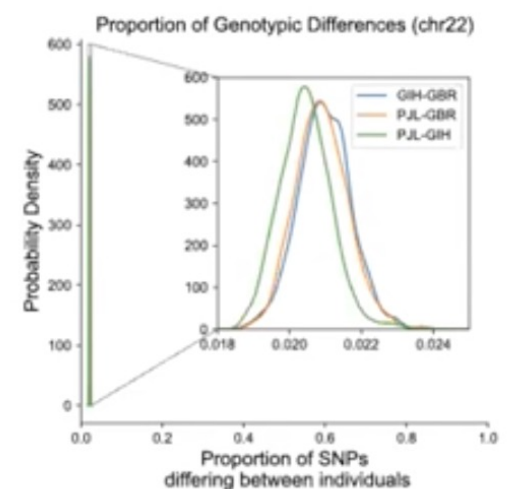Each region contains a majority of the alleles found worldwide.

Each circle represents the total number of alleles (gene variants) in the human genome found within a particular group.

African Ancestry   East-Asian Ancestry   European Ancestry

© Brian Donovan / BSCS Science Learning

Via ANOVA

HGDP - Global Populations Chr 15

Two individuals from the same population

Two individuals from the same geographic region

Two arbitrarily selected individuals

Proportion of alleles differing between two individuals

Rosenberg (2011; 2019)

Proportion of Genotypic Differences (chr22)

GIH-GBR
PJL-GBR
PJL-GIH

Proportion of SNPs differing between individuals

1000 Genomes data

Acknowledgement

Some slides and material taken from an excellent talk by John Novembre:

https://www.dropbox.com/s/20duu56k1y3pj88/novembre_recording_2021-09-29.mp4?dl=0

Highly recommended, and I'm not sure how permanent the above link is, so watch soon...

It was part of https://cehg.stanford.edu/evolgenome-seminars; Also recommended!