

Opinion

Human genomic variation

Todd R Disotell

Address: Department of Anthropology, New York University, 25 Waverly Place, New York, NY 10003, USA. E-mail: todd.disotell@nyu.edu

Published: 10 November 2000

Genome **Biology** 2000, **1**(5):comment2004.1–2004.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2000/1/5/comment/2004>

© Genome**Biology**.com (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

The recent completion and assembly of the first draft of the human genome, which combines samples from several ethnically diverse males and females, provides preliminary data on the extent of human genetic variation.

On June 26, 2000 at the White House, Craig Venter, Celera Genomics' president and chief scientific officer, announced that the complete human genome had been assembled, using the whole-genome shotgun-sequencing method, in only nine months [1]. But what did he mean by 'the' human genome? In fact, the Celera research group sequenced a composite genome composed of three females and two males who identified themselves as African-American, Asian, Caucasian, and Hispanic. During his announcement, Venter explained that this sampling and the sequences generated from it, "help illustrate that the concept of race has no genetic or scientific basis." Numerous articles have appeared in the popular press since then with titles such as '*Do Races Differ? Not Really, Genes Show*' [2]. Do Celera's data indeed demonstrate this?

Celera's method of whole-genome shotgun sequencing allowed for the rapid discovery of hundreds of thousands of single nucleotide polymorphisms (SNPs). SNPs are less variable than microsatellite markers which have previously been widely used to characterize human molecular variation and evolution (see, for example, [3,4]), although they are also much more common and less mutationally complex [5,6]. In September 1998, the US National Center for Biotechnology Information (NCBI) created the Single Nucleotide Polymorphism database (dbSNP [7,8]) in order to gather widely disparate research groups' efforts into a common format that was readily accessible. Two years on, Celera launched its SNP database [9]. The first release of this database (September 2000) contains 2.4 million unique SNPs that are not found in the public databases [9]. So, taking these together with 400,000 non-redundant

SNPs from the public databases, there now are over 2.8 million SNPs characterized throughout the human genome. Will this new resource tell us anything new about human variation?

Prior to the SNP-gathering efforts, what was known about the patterns of human variation? Since the initial discoveries of blood-group typing during the first half of the twentieth century, blood-group analyses have led to various characterizations of the number of distinct human groups, such as Snyder's seven-way division [10]. Livingstone's work on the genetics and distribution of hemoglobin variants related to malaria and sickle-cell anemia exemplified much of the work of the 1950s and 1960s [11]. Lewontin [12], using a population geneticist's perspective, analyzed in 1972 the then relatively large body of data on blood groups and protein variants representing a total of 17 loci. He found that 85% of all human variation is found between individuals within a nation or tribe. An additional 8% is found between populations within races and only 6% between the races (defined in the broadest tripartite sense - Caucasoid, Negroid, and Mongoloid in Lewontin's parlance). This pattern seemingly contradicts our visual perceptions of the differences between groups from different areas around the world. Is it an artifact of using protein and blood-group data? The answer is no. Barbujani *et al.* [13] surveyed 16 populations from around the world for 109 DNA markers (30 microsatellites and 79 restriction-fragment length polymorphisms, RFLPs) and found that 84.4% of worldwide variation in these markers could be found between members of the same population while less than 10% of the variation was between major races.

What accounts for this pattern of diversity? It arises because, as it turns out, humans have a rather interesting evolutionary history. While some paleoanthropologists have been proposing a relatively recent origin for our species, it was not until the late 1980s that analyses of molecular data began to reveal just how recent was our shared common ancestor. The formerly generally accepted - though theoretically improbable - multiregional theory of modern human origins saw our ancestors migrating out of Africa over one million years ago into the different regions of the Old World. Anatomically modern humans then evolved in each of these regions from those original archaic migrants, although enough gene flow took place between the regions to maintain the continuity of our species.

Cann, Stoneking, and Wilson [14], after analyzing mitochondrial DNA restriction patterns from 145 individuals representing populations from around the world, inferred that all human mitochondrial DNA traces its common ancestry to an African population that lived around 200,000 years ago. Although there were analytical problems with this and subsequent studies, recent studies based on mitochondrial DNA continue to reach this same conclusion (see [15] and references therein). Hammer [16] and numerous researchers sequencing and typing the Y chromosome have reached similar conclusions of an approximately 200,000 year-old African origin for modern human Y chromosome variation. Many studies of autosomal variation, including those of Bowcock *et al.* [3] and Jorde *et al.* [17], offer support for a recent African origin. Molecular estimates suggest that populations did not expand out of Africa until around 100,000 years ago (see [18] for a recent discussion) which is consistent with the human fossil record [19]. Thus, there has been nearly twice as long for variation to accumulate in sub-Saharan Africa as in the rest of the world.

This then leads us back to Celera's sample. What strategy would best represent world-wide genetic diversity? A formal project has been proposed, the Human Genome Diversity Project [20], with the aim of collecting samples and data from a wide range of populations throughout the world thought to best represent human diversity. While not without controversy [21,22], this proposal is still being modified and developed to deal with various ethical concerns. Until such sampling is available, Celera's five-person sample should be viewed as only the first step in characterizing human diversity.

Given the patterns of human molecular variation and our evolutionary history described here, a scientifically - though perhaps not politically - more viable strategy would be to examine many more sub-Saharan Africans than non-Africans, because sub-Saharan African populations can be expected to represent the majority of all human variation. Furthermore, if samples are to represent different regions of the world, they should be gathered from those regions themselves, not from

within a population as mixed as that of the United States. About 20% of Americans have close relatives from racial groups different from their own, on the basis of household survey data [23]. Molecular estimates of the European contribution to the African-American gene pool range from lows of around 7% in Jamaica, up to 26% in some North American cities [24]. Ancestry among the population who identify as Hispanic can be very mixed, although generally with a very significant European component (see, for example, [25]). Asian Americans similarly have high intermarriage rates [23]. Thus, the self-identified ethnicity of Celera's donor pool in all likelihood dramatically over represents the European gene pool while underestimating the most variable region of the world, sub-Saharan Africa.

While Venter surely spoke from the heart when he stated that "We did this initial sampling, not in an exclusionary way, but out of respect for the diversity that is America"... (White House Press Conference, 26 June 2000), a more accurate sampling of human diversity needs to take into account our evolutionary history and known patterns of variation among current human populations. A human diversity project that takes into account the various ethical and legal issues raised has been made far more tractable by the groundwork laid by Celera and the publicly funded Human Genome Organization's impressive accomplishments, not only in sequencing 'the' human genome but also in beginning to use it as a map to discover the full extent of human genetic diversity.

References

1. **Celera Genomics Completes the First Assembly of the Human Genome (press release).** [http://www.celera.com/corporate/about/press_releases/celera062600_1.html]
2. Angier N. **Do races differ? Not really, genes show.** *New York Times*, 2000, August 22, pp. F1-F5.
3. Bowcock AM, Hebert JM, Mountain JL, Kidd JR, Rogers J, Kidd KK, Cavalli-Sforza LL: **Study of an additional 48 DNA markers in five human populations from four continents.** *Gene Geogr* 1991, **5**:151-173.
4. Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW: **Genetic absolute dating based on microsatellites and the origin of modern humans.** *Proc Natl Acad Sci USA* 1995, **92**:6723-6727.
5. Weiss KM: **In search of human variation.** *Genome Res* 1998, **8**:691-697.
6. Kruglyak L: **The use of a genetic map of biallelic markers in linkage studies.** *Nature Genet* 1997, **17**:21-24.
7. **dbSNP: A database of human single nucleotide polymorphisms** [<http://www.ncbi.nlm.nih.gov/SNP/>]
8. Smigielski EM, Sirotkin K, Minghong W, Sherry ST: **dbSNP: a database of single nucleotide polymorphisms.** *Nucl Acids Res* 2000, **28**:352-355.
9. **Celera SNP Reference Database** [<http://www.celera.com/corporate/snpdata.html>].
10. Synder LH: **Human blood groups: their inheritance and racial significance.** *Am J Phys Anthropol* 1926, **9**:233-263.
11. Livingstone FB: **Anthropological implications of sickle-cell gene distribution in West Africa.** *Amer Anthropol* 1958, **60**:533-562.
12. Lewontin RC: **The apportionment of human diversity.** In *Evolutionary Biology*, vol 6. Edited by Dobzhansky TH, Hecht MK, Steere WC. New York: Appleton-Century-Crofts; 1972
13. Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL: **An apportionment of human DNA diversity.** *Proc Natl Acad Sci USA* 1997, **94**:4516-4519.

14. Cann RL, Stoneking M, Wilson AC: **Mitochondrial DNA and human evolution.** *Nature* 1987, **325**:31-36.
15. Disotell TR: **Sex-specific contributions to genome variation.** *Curr Biol* 1999, **9**:R29-R31.
16. Hammer MF: **A recent common ancestry for human Y chromosomes.** *Nature* 1995, **378**:376-378.
17. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA: **The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data.** *Am J Hum Genet* 2000, **66**:979-988.
18. Disotell TR: **The southern route to Asia.** *Curr Biol* 1999, **9**:R925-R928.
19. Stringer CB, Andrews P: **Genetic and fossil evidence for the origin of modern humans.** *Science* 1988, **239**:233-263.
20. Cavalli-Sforza LL, Wilson AC, Cantor CR, Cook-Deegan RM, King MC: **Call for a worldwide survey of human genetic diversity: a vanishing opportunity for the Human Genome Project.** *Genomics* 1991, **11**:490-491.
21. Marks J: **The trouble with the Human Genome Diversity Project.** *Mol Med Today* 1998, **4**:243.
22. Wallace RW: **The Human Genome Diversity Project: medical benefits versus ethical concerns.** *Mol Med Today* 1998, **4**:59-62.
23. Goldstein JR: **Kinship networks that cross racial lines: the exception or the rule?** *Demography* 1999, **36**:399-407.
24. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD: **Estimating African American admixture proportions by use of population-specific alleles.** *Am J Hum Genet* 1998, **63**:1839-1851.
25. Long JC, Williams RC, McAuley JE, Medis R, Partel R, Tregellas WM, South SF, Rea AE, McCormick SB, Iwaniec U: **Genetic variation in Arizona Mexican Americans: estimation and interpretation of admixture proportions.** *Am J Phys Anthropol* 1991, **84**:141-157.