Genomic Views of Human History

Kelly Owens and Mary-Claire King

New tools of genomic analysis shed light on historical puzzles. Migrations of ancient peoples, differences in migration patterns of males and females, historical demography of cultures with ancient roots, and patterns of human genetic diversity are increasingly the focus of integrated analysis by historians, anthropologists, and geneticists.

In 1959, C. P. Snow gave a lecture (*1*) in which he described two cultures, humanities ("literature") and science ("technology"), and bemoaned the inability of members of either to communicate with members of the other. His concern was primarily sheer ignorance of literature and history by technically trained people, and of science and technology by people taught humanities and (he later added) social science. Perhaps the greatest credit to Snow is that 40 years later, people in both cultures remain aware of his insights and are trying to address them. Efforts of teachers to do so continue worldwide. What Snow might not have envisioned was the extent to which modern technology would provide tools to help address questions posed by humanists. The two cultures may ultimately converge at questions of mutual interest, which arose from common curiosity, to be answered with tools from very different disciplines. As geneticists, we have been particularly drawn to the possibility of applying genomic analysis to questions of human history. Molecular genetics has begun to revolutionize the study of human evolution. Analysis of human genomes now offers the possibility of understanding movements and events of more recent human history. Geneticists' analyses of records written in our DNA can complement historians' analyses of records written by human observers. The questions we can pose are bounded only by our imaginations. How did languages begin and how do they survive? How are new cultures established and how are they isolated? Who moves across cultures? How do people adapt to new places? What is the role of women in pioneer societies? Genomic approaches may contribute to explorations of many of these issues.

Most efforts to apply genomic analysis to historical questions are still carried out by biologists rather than historians and hence are frequently oriented toward medicine and anthropology rather than cultural history. Our purpose here is to illustrate the types of historical questions that have recently been addressed by genomic analysis. We hope that in the future more collaboration between history and genetics will add to the scope and depth of the questions investigated. Also, genomic analysis of populations with vastly different cultural histories has already raised complex issues of the rights of individuals and groups, which are now being addressed by efforts to bring cultural understanding to bear on questions of law and genetics (*2*).

Departments of Genetics and Medicine, Box 357720, University of Washington, Seattle, WA 98195, USA.

Human Migrations

When people move, they take their genes along and pass them on to their descendants in their new homes. Thus, every present-day population retains clues to its ancient roots. Common ancestries can be confirmed and human migrations traced by comparing DNA sequences of present-day populations. Early migrations of modern humans out of Africa have been traced by analysis of DNA sequences (*3*, *4*). More recent human migrations have been followed through genetic trails as well. A lovely example is the application of statistical analysis of classical polymorphisms to the question of ancient migrations within Europe (*5*). Genotypes at 95 alleles were determined for several hundred individuals from 26 present-day European populations. Allele frequencies among populations were summarized by principal components, five of which explained 75% of European genetic variation. Values of each component vary in gradients across Europe. Most fascinating, these components, constructed statistically from genotype frequencies, reflected archaeological reconstructions of major human migrations (*5*). In particular, the values of the most important component reflected the paths of Neolithic farmers from the Middle East (i) westward across the Mediterranean to present-day Greece, Italy, and Spain, and (ii) northwest through Europe, to present-day Germanic-speaking areas of northern Germany, northwestern Europe, and England. Lapps in northern Scandinavia and other Uralic people in Finland and northern Russia were represented by a second component that may reflect both south-north migration and selection by climate. Migrations from the steppes of the lower Volga and Don rivers by the early pastoral nomads who first domesticated the horse could be represented by the third component. Celtic populations in Britain, Ireland, and Brittany were characterized by extreme values of a fourth component, and the Basque population by a fifth, reflecting historic, geographic, and linguistic isolation.

Do males and females migrate in the same ways? Genetic analyses of sequences of mitochondrial DNA (mtDNA), Y chromosome markers, and nuclear loci carried out in the past few years suggest that the migration rates of males and females have been dramatically different for much of human history. Mitochondrial DNA is purely maternally inherited, with virtually no recombination; Y chromosomes are purely paternally inherited, with no recombination on the Y-specific region of the chromosome. Hence each offers a gender-specific record of the past. Sequences on autosomes reveal the combined impact of migration of males and females. Y chromosome variants are far more localized geographically than are alleles of mtDNA or autosomal markers (*6*). That is, most genetic variation of both autosomal loci and mtDNA sequences is within each population, not between populations. An average population from anywhere in the world includes 85% of all human variation at autosomal loci (*7*) and 81% of all human variation in mtDNA sequences (*8*). Differences among populations from the same continent contribute another 6% of variation; only 9 to 13% of genetic variation differentiates populations from different continents (Fig. 1). In contrast, an average population includes only 36% of the world's Y chromosome variation; most (53%) variation on the Y lies in populations from different continents (*6*). Most of this discrepancy is explained by higher migration rates among females than males. When females relocate to the birthplaces of their spouses, children are born close to the birthplaces of their fathers but farther from the birthplaces of their mothers. Most individual females do not move far, but over hundreds of generations, the genetic effects of their movements accumulate, leading to the observed migration patterns.

Genetic Perspectives on Cultural History

Genomic analysis can reveal the historical demography of cultures with ancient roots. Genomic analysis can also indicate how current populations are related to each other, and often the extent and timing of their contacts. These questions have interested geneticists and historians from cultures in all parts of the world.

How have people and genes moved along ancient trade routes? The Kazakh, Uighur, and Kirghiz populations of central Asia live

along the ancient Silk Road, the trade route between Europe and Asia that flourished between approximately 200 B.C. and 400 A.D. Analysis of mtDNA sequences of these populations suggests that they are descended from people moving from Europe to Asia and vice versa more than 2000 years ago, albeit long after the early human migrations out of Africa (*9*). Comparison of mtDNA and Y chromosome sequences again indicates more female migration, which is consistent with recognized paternal-clan exogamy in this region (*10*). Furthermore, Y chromosome variation in part parallels language differences among these populations, whereas mtDNA variation does not. Y chromosome data from central Asia and from other regions of the world (*11*) suggest that genetic differences at linguistic boundaries are due primarily to male, rather than female, isolation.

The history of the Nile River Valley poses a somewhat different genetic puzzle (*12*). The Nile River Valley is more than 2000 km long from southern Sudan through Egypt and less than 10 km wide for most of its length. It has existed in its present geographic and climatic form for more than 10,000 years, with civilizations described in written records going back more than 5000 years. The question is whether the Nile has been a "genetic corridor" or whether the cultural and linguistic differences of populations along the river posed a barrier to migration. Analysis of mtDNA sequences from Egyptian, Nubian, and Sudanese populations living from the northern to the southern extent of the river valley revealed a smooth gradient of genetic distance paralleling geographic distance, with no evident barriers. Furthermore, genetic distances were consistent with earlier migrations from north to south rather than from south to north. These data are consistent with descriptions by ancient Egyptian historians, who chronicled Pharaonic colonization of Nubia during the Middle Kingdom (1991–1785 B.C.) and the New Kingdom (1490–1437 B.C.) as well as later conquests of Egypt by Nubian kings during the 25th Dynasty (730–655 B.C.). It remains to be seen whether genetic analysis of Y chromosomal and autosomal alleles will confirm the ancient accounts. Stay tuned.

Historical, linguistic, and genetic controversies surround the question of the origin of the Ashkenazi Jews (*13*). For hundreds of years, theories of Ashkenazi roots have been drawn from analysis of language, history, and literature. Genetic evidence contributes as well, including genomic characterization of ancestral lineages of specific Ashkenazi alleles. For example, genetic analysis of a founder allele for idiopathic torsion dystonia, an autosomal dominant and incompletely penetrant disease, indicated that the mutation occurred approximately 300 to 500 years ago in what was then the northern part of the Jewish Pale of settlement (*14*). The recent origin and current high frequency of this deleterious allele permitted historical demographic analysis of the Ashkenazi population. The data suggest that present-day Ashkenazim descend from a relatively small number of ancestors, perhaps 10,000 to 20,000 in Poland and neighboring states at the beginning of the 16th century and even fewer 1000 years ago. The data are consistent with historical investigations that suggest that the wealthier classes of this population expanded rapidly despite many losses in subsequent pogroms. The combination of a small founding population and rapid expansion over 1000 years are sufficient to explain the presence, purely by chance, of multiple alleles unique to this population: Any allele from a small founding population that survives to the present will be, by necessity, at reasonably high frequency. In Finland as well, rapid expansion of a small founder population in less than 500 years, has led to persistence of at least 30 alleles that are quite common among Finns and unique to the Finnish population. Genetic drift, rather than selective advantage to carriers, is the most plausible explanation for existence of unique alleles in both populations (*15*).

Genetic data also demonstrate the Middle Eastern origin of the Ashkenazim, who share specific alleles with other Middle Eastern Jewish and non-Jewish groups. Southern and Central European populations subsequently contributed other alleles to the Ashkenazim. For example, disequilibrium analysis of human leukocyte antigens DR and DQ indicates that Ashkenazi Jews are most closely related to other Jews, next most closely to other Mediterranean populations, and less closely to Central Europeans, though carrying some Central European alleles (*16*). Genetic evidence also supports the oral tradition that the Lemba, who are now Bantu-speaking people of southern Africa, derive from Jews who migrated from the Middle East to Yemen 2700 years ago and from Yemen to southern Africa 2400 to 2000 years ago. More than 50% of Lemba Y chromosomes carry haplotypes that are common among Jewish populations but absent in their African neighbors (*17*). Genetic analysis has also confirmed the distinctiveness of the Cohanim, or traditional Jewish priesthood. Y chromosome haplotypes differ between Cohanim, whether Ashkenazi or Sephardic, and other Jewish men. Indeed, most Cohen males carry a unique Y haplotype (*18*).

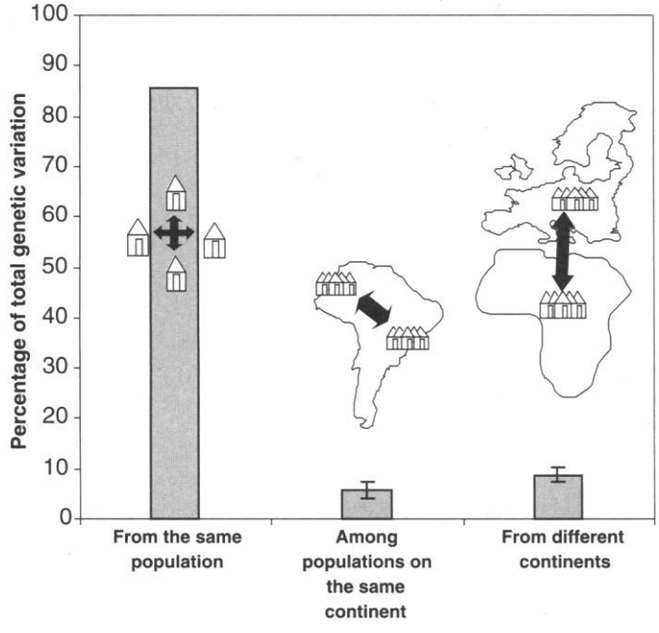

Fig. 1. Percentage of human genetic variation within and between populations.

Genetics, History, and Race

The above examples demonstrate the power of genetics to reveal historical migrations and demography. However, genetic differences of populations from different continents represent only about 10% of human genetic diversity; no major genetic discontinuities across populations have been observed (*7*). (The Y chromosome, which differs considerably among populations, encodes far less than 1% of male, human genes.) More than 80% of genetic variation is between individuals of the same population, even in small or isolated populations (Fig. 1). Most human genetic variation antedates the migration of modern humans

out of Africa (*5*, *19*, *20*). The possibility that human history has been characterized by genetically relatively homogeneous groups ("races"), distinguished by major biological differences, is not consistent with genetic evidence.

How, then, does genetics explain the stereotypic features of "races": skin color, hair color and texture, and facial traits? These traits are quite literally superficial, in that they affect exposed surfaces of the body. It is reasonable to suggest that variation in these traits may reflect differential selection by climate in various parts of the world. Recent analysis of the melanocortin-stimulating hormone receptor gene (MC1R) suggests that various alleles of this single locus may underlie much of observed human variation in skin and hair color (*21*, *22*). This variation is largely due to varied amounts of eumelanin (brown and black melanins) and phaeomelanin (red and yellow melanins) produced by melanocytes. Eumelanin protects against ultraviolet (UV) radiation, whereas phaeomelanin may contribute to skin damage, including melanoma, induced by UV. The balance of melanins is regulated by melanocyte-stimulating hormone, which acts through its receptor. Amino acid sequence variants occur at multiple sites in the second transmembrane domain, the first extracellular domain, and the seventh transmembrane domain of the MC1R protein. Variation at these sites was found in more than 80% of individuals with red hair and fair skin that burns rather than tans, but in less than 4% of British or Irish individuals with skin that tans without burning, and in no African individuals. Among Asians, still other amino acid substitutions in MC1R are common. Nucleotide diversity at MC1R is several times higher than the average nucleotide diversity in human populations. High nucleotide diversity, coupled with common variation at nonsynonymous sites, suggest that MC1R variation is an adaptive response to selection for different alleles in different environments, possibly to differences in day length and hence available sunlight at different latitudes. If true, variation at this locus, which encodes evolutionarily important but superficial traits, has been the cause of enormous suffering. Variation in other traits popularly used to identify "races" is likely to be due to similarly straightforward mechanisms, involving limited numbers of genes with very specific physiological effects. Of course, prejudice does not require a rational basis, let alone an evolutionary one, but the myth of major genetic differences across "races" is nonetheless worth dismissing with genetic evidence.

References and Notes

1. C. P. Snow, *The Two Cultures and the Scientific Revolution, The Rede Lecture* (Cambridge Univ. Press, New York, 1959).
2. H. Greely, *Annu. Rev. Anthropol.* **27**, 473 (1998). ———, *Houston Law Rev.* **33**, 1431 (1997).
3. A. C. Wilson and R. L. Cann, *Sci. Am.* **266**, 68 (April 1992).
4. L. Jin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3796 (1999).
5. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994); A. Piazza *et al.*, *ibid.* **92**, 5836 (1995).
6. M. T. Seielstad, E. Minch, L. L. Cavalli-Sforza, *Nature Genet.* **20**, 278 (1998).
7. G. Barbujani *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4516 (1997).
8. L. Excoffier, P. E. Smouse, J. M. Quattro, *Genetics* **131**, 479 (1992).
9. D. Comas *et al.*, *Am J. Hum. Genet.* **63**, 1824 (1998).
10. A. Perez-Lezaun *et al.*, *ibid.* **65**, 208 (1999).
11. E. S. Poloni *et al.*, *ibid.* **61**, 1015 (1997).
12. M. Krings *et al.*, *ibid.* **64**, 1166 (1999).
13. Second International Symposium in Memory of Richard M. Goodman, National Foundation for Jewish Diseases and Tel Aviv University, Israel, June 1999.
14. N. Risch *et al.*, *Nature Genet.* **9**, 152 (1995).
15. A. G. Motulsky, *ibid.*, p. 99; A. de la Chapelle and F. A. Wright, *Proc. Natl. Acad. Sci. U.S.A.*, **21**, 12416 (1998).
16. J. Martinez-Laso *et al.*, *Tissue Antigens* **47**, 63 (1996).
17. A. B. Spurdle and T. Jenkins, *Am. J. Hum. Genet.* **59**, 1126 (1996).
18. K. Skorecki *et al.*, *Nature* **385**, 32 (1997).
19. L. B. Jorde *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3100 (1997).
20. H. Kaessmann *et al.*, *Nature Genet.* **22**, 78 (1999).
21. P. Valverde *et al.*, *ibid.* **11**, 328 (1995).
22. B. K. Rana *et al.*, *Genetics* **151**, 1547 (1999); H. B. Schioth *et al.*, *Biochem. Biophys. Res. Commun.* **260**, 488 (1999).

VIEWPOINT

Biosequence Exegesis

Mark S. Boguski

Annotation of large-scale gene sequence data will benefit from comprehensive and consistent application of well-documented, standard analysis methods and from progressive and vigilant efforts to ensure quality and utility and to keep the annotation up to date. However, it is imperative to learn how to apply information derived from functional genomics and proteomics technologies to conceptualize and explain the behaviors of biological systems. Quantitative and dynamical models of systems behaviors will supersede the limited and static forms of single-gene annotation that are now the norm. Molecular biological epistemology will increasingly encompass both teleological and causal explanations.

The sequences of proteins (and a few nucleic acids) had slowly been accumulating in the literature since Sanger's seminal work on the structure of insulin in the 1950s. However, the real catalyst for the expanding depth and scope of our knowledge of macromolecular sequences was the development of rapid DNA-sequencing technologies in 1975 (*1*). At that point, the direct determination of protein sequences began to be supplanted by conceptual translations (using the genetic code) of DNA sequences into their cognate gene products. For the next 15 years, sequence data continued to be so novel and revealing that each new example was reported in a peer-reviewed publication accompanied by a richly detailed interpretation derived from the biological hypothesis or context that led to the cloning of a particular gene in the first place. [In one case, which seems remarkable in hindsight, the cloning of the first mammalian messenger RNA (mRNA) in 1977 merited three separate publications in the journal *Cell*, one describing the 5′ untranslated region, another the 3′ untranslated region, and another the coding sequence of rabbit β globin mRNA.] Often there were many follow-up publications about particularly important genes or proteins that resulted in an even richer and more complete "annotation" (in the literature) of the biological processes in which a sequence was involved. This was the era of "functional cloning," the hallmark of which was "function first, sequence later." In that era, database similarity (or "homology") searching was a hit-or-miss activity, characterized by frequent misses and rare hits, the latter usually being extremely rewarding and informative and engendering the palpable excitement of ground-breaking discoveries (*2*). On the order of 10,000 mammalian genes have been functionally cloned over the past two decades.

In the late 1980s and early 1990s, the method of functional cloning was enriched by a powerful new approach, positional cloning (*3*). In this approach, one begins with a phenotype, proceeds through genetic linkage and physical

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD 20894, USA.

References and Notes

[2] **Legal, Ethical, and Social Issues in Human Genome Research**
Henry T. Greely
Annual Review of Anthropology, Vol. 27. (1998), pp. 473-502.
Stable URL:
http://links.jstor.org/sici?sici=0084-6570%281998%292%3A27%3C473%3ALEASII%3E2.0.CO%3B2-X

[4] **Distribution of Haplotypes from a Chromosome 21 Region Distinguishes Multiple Prehistoric Human Migrations**
Li Jin; Peter A. Underhill; Vishal Doctor; Ronald W. Davis; Peidong Shen; L. Luca Cavalli-Sforza; Peter J. Oefner
Proceedings of the National Academy of Sciences of the United States of America, Vol. 96, No. 7. (Mar. 30, 1999), pp. 3796-3800.
Stable URL:
http://links.jstor.org/sici?sici=0027-8424%2819990330%2996%3A7%3C3796%3ADOHFAC%3E2.0.CO%3B2-0

[7] **An Apportionment of Human DNA Diversity**
Guido Barbujani; Arianna Magagni; Eric Minch; L. Luca Cavalli-Sforza
Proceedings of the National Academy of Sciences of the United States of America, Vol. 94, No. 9. (Apr. 29, 1997), pp. 4516-4519.
Stable URL:
http://links.jstor.org/sici?sici=0027-8424%2819970429%2994%3A9%3C4516%3AAAOHDD%3E2.0.CO%3B2-7

LINKED CITATIONS

[15] **Linkage Disequilibrium Mapping in Isolated Populations: The Example of Finland Revisited**

Albert De La Chapelle; Fred A. Wright

Proceedings of the National Academy of Sciences of the United States of America, Vol. 95, No. 21. (Oct. 13, 1998), pp. 12416-12423.

Stable URL:

http://links.jstor.org/sici?sici=0027-8424%2819981013%2995%3A21%3C12416%3ALDMIIP%3E2.0.CO%3B2-G

[19] **Microsatellite Diversity and the Demographic History of Modern Humans**

Lynn B. Jorde; Alan R. Rogers; Michael Bamshad; W. Scott Watkins; Patrycja Krakowiak; Sandy Sung; Juha Kere; Henry C. Harpending

Proceedings of the National Academy of Sciences of the United States of America, Vol. 94, No. 7. (Apr. 1, 1997), pp. 3100-3103.

Stable URL:

http://links.jstor.org/sici?sici=0027-8424%2819970401%2994%3A7%3C3100%3AMDATDH%3E2.0.CO%3B2-P