may have left the most widely recognized imprint in the geological record, but the biogenic accumulation of methane in the Archean atmosphere may have been more important, protecting life from freezing.

The idea of a methane greenhouse is not new, but Habicht *et al*. (*9*) present the first model that is conclusive with respect to the residence time of methane in the atmosphere. Today, the residence time of methane is about 10 years, but in an Archean anoxic environment it might have been 10,000 years (*19*). This time scale requires a continuous supply of methane to keep the methane greenhouse going. Methanogenic bacteria could have provided a continuous flow of methane, building up a high methane concentration in the atmosphere as long as the ocean sulfate concentration remained low.

The studies of Habicht *et al*. (*9*) and Farquhar *et al*. (*8*) impose important constraints on the Archean sulfur cycle. The discovery of mass-independent sulfur isotope effects in sulfide inclusions makes an oxygen-free atmosphere, flooded by UV light, a virtual certainty. Low sulfate concentration in the oceans seems to be essential for an early methane greenhouse.

However, to achieve a conclusive picture of the Archean environment, an independent proxy for atmospheric methane is needed. We must also improve our understanding of how the oxidation of the oceans relates to the rise of oxygen in the atmosphere. Iron and molybdenum isotopes may provide information about changing condition in the oceans. Another source of information may be the Cassini-Huygens mission (now on its way to Saturn), which will provide detailed insights into the composition of Titan's atmosphere. Titan could hold answers to Earth's evolution and may even provide tantalizing clues of life-bearing chemistry.

References

1. H. D. Holland, Ed., *The Chemical Evolution of the Atmosphere and Oceans* (Princeton Univ. Press, Princeton, NJ, 1984).
2. H. D. Holland, *Geochem. News* **100**, 20 (1999).
3. H. Ohmoto, *Geochem. News* **93**, 12 (1997).
4. J. Farquhar, H. Bao, M. Thiemens, *Science* **289**, 756 (2000).
5. J. Farquhar, S. Airieau, J. Savarino, M. Thiemens, *11th V. M. Goldschmidt Conference*, Hot Springs, VA, 20 to 24 May 2001 (abstract 3402).
6. S. J. Mojzsis *et al.*, *11th V. M. Goldschmidt Conference*, Hot Springs, VA, 20 to 24 May 2001 (abstract 3185).
7. A. Bekker *et al.*, *12th V. M. Goldschmidt Conference*, Davos, Switzerland, 18 to 23 August 2002 (abstract), A64.
8. J. Farquhar, B. A. Wing, K. D. McKeegan, J. W. Harris, P. Cartigny, M. H. Thiemens, *Science* **298**, 2369 (2002).
9. K. S. Habicht, M. Gade, B. Thamdrup, P. Berg, D. E. Canfield, *Science* **298**, 2372 (2002).
10. J. Farquhar, J. Savarino, S. Airieau, M. H. Thiemens, *J. Geophys. Res.* **106**, 32829 (2001).
11. C. Sagan, C. Chyba, *Science* **276**, 1217 (1997).
12. P. Cartigny, J. W. Harris, M. Javoy, *Science* **280**, 1421 (1998).
13. M. Chaussidon, F. Albarede, S. M. F. Sheppard, *Nature* **330**, 242 (1987).
14. C. S. Eldridge, W. Compston, I. S. Williams, J. W. Harris, J. W. Bristow, *Nature* **353**, 649 (1991).
15. D. E. Canfield, *Nature* **396**, 450 (1998).
16. ———, K. S. Habicht, B. Thamdrup, *Science* **288**, 658 (2000).
17. A. A. Pavlov *et al.*, *J. Geophys. Res.* **105**, 11981 (2000).
18. D. C. Catling, K. J. Zahnle, C. P. McKay, *Science* **293**, 839 (2001).
19. J. F. Kasting, J. L. Siefert, *Science* **296**, 1066 (2002).

PERSPECTIVES: HUMAN GENETICS

Mapping Human History

Mary-Claire King and Arno G. Motulsky

The DNA of modern humans contains a record of the travels and encounters of our ancestors. The genotypes of people living today are the result of ancient human migrations, the continuous appearance of new mutations, selection by climate and infection for genetic alleles that conferred a survival advantage, and mating patterns determined by cultural norms. By sampling genotypes from people across the globe, geneticists have reconstructed the major features of our history: our ancient African origin, migrations out of Africa, movements and settlements throughout Eurasia and Oceania, and peopling of the Americas (*1*–*5*). As genomic technology has improved, these analyses of genotype have successively incorporated new markers: blood groups (*2*), protein polymorphisms (*2*), mitochondrial DNA sequences (*1*), Y chromosome haplotypes (*3*), and highly variable nuclear microsatellite markers (*4, 5*).

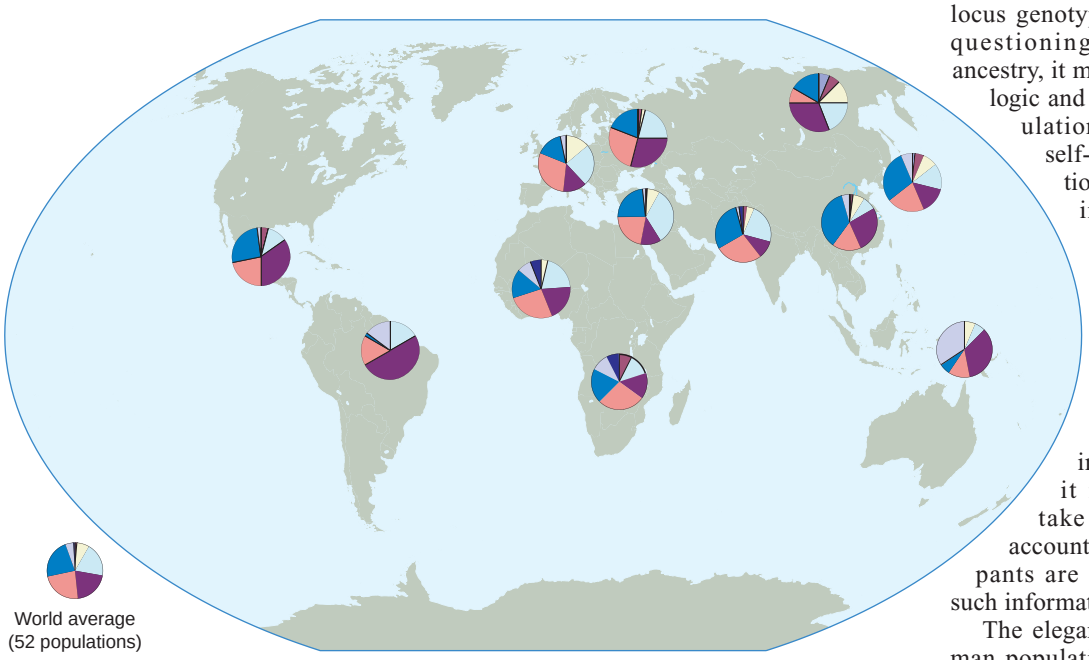
The most recent contribution to this literature is by Rosenberg *et al*. (*6*) on page 2381 of this issue. These investigators explored the genetic structure of human populations using highly variable markers on the human autosomes of individuals from different parts of the world. The genotyped markers were microsatellite short tandem repeat sequences that do not encode any expressed genes and are generally selectively neutral. The populations studied were defined by geography, language, and culture, and participating individuals were well rooted in their populations, with several generations of ancestors known to have lived in the same locale as the participant. Genotypes from more than a thousand individuals were evaluated by a statistical method that defines clusters of people on the basis of genetic similarity at multiple loci, without using prior information about ancestry. In this method, individuals are assigned to clusters probabilistically (*5, 7*). Individuals may have significant probabilities of membership in more than one cluster due either to genetic similarities of groups or to ancestral intergroup matings. The world map (see the figure) illustrates variation at one microsatellite marker in 12 populations. This marker has four common alleles, each of which appears in all populations. Rare alleles are shared by fewer populations. Few alleles are unique to only one population. No allele is population specific.

Previous genetic analyses of human history have consistently suggested that most human genetic variation is due to differences among individuals within populations rather than to differences among populations (*4, 8*). The Rosenberg *et al*. analysis of many more markers and many more people confirms this result: 93 to 95% of genetic variation is due to genetic differences among individuals who are members of the same population and only 3 to 5% of genetic variation is due to differences among the major population groups.

The power of the method lies in the construction of clusters on the basis of accumulated small differences in allele frequencies across many markers and many people. Statistical clustering of genotypes—composed of 4682 alleles from 377 markers in 1056 individuals from 52 populations—yields groups corresponding to major geographic regions of the world [see figure 1 in (*6*)]. Creation of two clusters reflects ancient human origins in Africa and rapid expansion throughout Eurasia, and migrations to the Americas from East Asia. Creation of five clusters yields groups corresponding to five major geographic regions of the world: Africa, Eurasia (Europe, the Middle East, Central and South Asia), East Asia, Oceania, and America. There is excellent agreement between membership of individuals in these clusters and their self-identified regions of origin. Similar results were obtained by the same statistical approach based on fewer populations and fewer markers [Table 2 of (*5*)].

The authors are in the Department of Medicine (Medical Genetics) and Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. E-mail: mcking@u.washington.edu

Population substructure could be consistently identified within some geographic regions but not others. Within Africa, for example, analysis consistently yielded the same four subclusters: Mbuti Pygmies, Biaka Pygmies, San peoples, and speakers of Niger-Kordofanian languages (Bantu, Yoruba, and Mandenka populations). In contrast, within Europe, multiple analyses were not consistent. Many more individuals will need to be included to sort out European demographic history.



Humans on the move. Worldwide genetic variation at a neutral marker. Allele frequencies of one randomly chosen microsatellite marker reveal common alleles shared in all populations and the gradual and arbitrary differences in allele frequencies across geographic regions. Populations shown in this example are Yoruba and Bantu (Africa); French, Russians, Palestinians, and Pakistani Brahui (Eurasia); Han Chinese, Japanese, and Yakut (East Asia); New Guineans (Oceania); and Maya and Karitianans (America). Each color on the diagrams represents one of nine alleles of GGAA29H03 (D13S1493), which range in length from 219 to 255 base pairs. By accumulating small differences in allele frequencies from hundreds of such highly variable markers and hundreds of people, statistical methods reveal genetic clusters of Africans, Eurasians, East Asians, Pacific Islanders, and Americans, corresponding to major ancient human migrations (*6*).

The identification of clusters corresponding to the major geographic regions may depend on the sampling of individuals from well-defined, relatively homogeneous populations. If individuals were sampled from a worldwide "grid" (or a worldwide grid weighted by population density), the clusters might be much less precisely defined. Does the correspondence of worldwide genetic clusters and major geographic regions suggest borders around genetic clusters analogous to the physical borders—oceans, mountain ranges, and deserts—separating geographic regions? No. Both the results of Rosenberg and colleagues and those of previous studies (*1–5, 8*) indicate that unlike separations between geographic regions, differences in allele frequencies are gradual, without discontinuities between clusters. After thousands of years—if enough markers and people are studied—allele frequency differences are collectively adequate to create clusters that correspond to the major migrations of human history.

What are the implications of the Rosenberg *et al*. findings for medicine? The current medical literature increasingly includes studies exploring population differences in disease incidence or in efficacy or adverse responses to drug treatment (*5*, *9*). The rationale of these studies is that alleles influencing disease susceptibility or treatment response may differ in frequency across populations. Consequently, individuals would be better served if critical genotypes were taken into account when assessing disease risk or designing treatment regimens. In the absence of knowing the identities of the critical alleles, personal ancestry as indicated by study participants is often used as an initial but potentially misleading substitute.

The Rosenberg *et al*. data suggest that with the exception of ancient highly selected loci (for example, the Duffy null blood group, which confers complete protection against vivax malaria), very few alleles will be both confined to one population and common enough in that population to dictate medical management for the entire group. Instead, critical alleles influencing disease risk or response to treatment are likely to be either ancient, worldwide, and relatively common in many populations, or geographically localized and individually rare (*10*).

Differences among populations in disease frequency and treatment outcome certainly occur but may not be genetic in origin. Given that the major population origin of groups can be defined by multilocus genotype clustering (*5–7*) without questioning individuals about their ancestry, it may be tempting in epidemiologic and clinical studies to omit population characterization through self-reporting. However, correlations identified by the clustering method may be falsely ascribed to genes when in fact they had nothing to do with genetics but were caused by social, economic, or discriminatory factors limited to a genetically defined population cluster (*11*). To evaluate medically important group differences, it is therefore necessary to take all such risk factors into account. Patients and study participants are usually the best source of such information.

The elegant statistical analysis of human population structure by Rosenberg and colleagues reflects the major human migrations out of Africa, into Europe, across Asia, into Oceania, and to the Americas. By genotyping a large sample of an individual's alleles, it is possible to identify the migrations in which his or her ancestors participated. But the link between historical genetic demography and medically important risk is complex. Disease susceptibility may be genetic but not geographically clustered, or geographically clustered but not genetic, or neither, or both.

References

1. A. C. Wilson, R. L. Cann, *Sci. Am.* **266**, 68 (April 1992).
2. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
3. P. A. Underhill *et al.*, *Nature Genet.* **26**, 358 (2000).
4. G. Barbujani, A. Magagni, E. Minch, L. L. Cavalli-Sforza, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4516 (1997).
5. J. F. Wilson *et al. Nature Genet.* **29**, 265 (2001).
6. N. A. Rosenberg *et al. Science* **298**, 2381 (2002).
7. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945 (2000).
8. R. C. Lewontin, *Evol. Biol.* **6**, 381 (1972).
9. H. G. Xie, R. B. Kim, A. J. Wood, C. M. Stein, *Annu. Rev. Pharmacol. Toxicol.* **41**, 815 (2001).
10. J. K. Pritchard, N. J. Cox, *Hum. Mol. Genet.* **11**, 2417 (2002).
11. N. Risch, E. Burchard, E. Ziv, H. Tang, *Genome Biol.* **3**, comment2007 (2002).