

PERSPECTIVE

The Human Genome Project and Molecular Anthropology

Mark Stoneking¹

Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania 16802

It has been ~10 years since the idea of determining the complete nucleotide sequence of the human genome was first proposed as a “big-science” project to the scientific community at the 1986 meeting at Cold Spring Harbor Laboratory on the molecular biology of *Homo sapiens* (Bodmer 1986). The ensuing decade has seen the realization of this goal as a worthy endeavor, and an enormous amount of progress has been made in developing new technologies and in determining a physical map of the genome. It appears as if the Human Genome Project (HGP) is now in the home stretch, with the consensus emerging that everything is in place to determine the sequence itself (Marshall and Pennisi 1996). Although the HGP was initially promoted on the basis of what the resulting knowledge would contribute to efforts to alleviate human misery caused (or contributed to) by our genes, the purpose of this paper is to preview what a complete human DNA sequence might mean for molecular anthropology.

Molecular anthropology is defined here as the use of molecular genetic techniques to address questions that anthropologists are interested in concerning human evolution and diversity. Molecular anthropology encompasses the analysis and interpretation of the following: (1) patterns of molecular genetic variation in various contemporary human populations; (2) patterns of molecular genetic variation in nonhuman primate populations; (3) comparisons of human and nonhuman primate genes; and (4) the retrieval of genetic information directly from ancient specimens. The purpose of this article is not to review the present state of molecular anthropology, as that has been ably done in several recent edited volumes (Boyce and Mascie-Taylor 1996; Chadwick and Cardew 1996; Jackson et al. 1996). Rather, what follows is a speculative preview of what the HGP might contribute to molecular anthropology.

Mitochondrial DNA (mtDNA) has been used extensively by molecular anthropologists, and the

MtDNA Genome Project was completed in 1981 with the publication of the complete sequence of the human mtDNA genome (Anderson et al. 1981). In considering what the HGP might mean for molecular anthropologists, then, it is instructive to consider what the availability of a complete human mtDNA sequence has meant for studies of human mtDNA variation and evolution.

Human DNA Variation and Evolution: Lessons from mtDNA

The complete human mtDNA sequence has greatly facilitated the analysis and interpretation of human mtDNA variation and evolution. During the pre-PCR era of molecular anthropology, the complete human mtDNA sequence made possible detailed, high-resolution restriction site maps by the sequence-comparison mapping method (Cann et al. 1984). By comparing the DNA fragments generated by a particular restriction enzyme for a particular DNA sample to the DNA fragment that would be predicted from the complete sequence, any differences could be localized quickly to a particular location in the mtDNA genome, and frequently the exact nucleotide substitution responsible for the polymorphism could be inferred. A recent African origin of the human mtDNA ancestor was first proposed from such high-resolution restriction site studies (Cann et al. 1987), and restriction site analysis continues to be useful in tracing the evolutionary history of particular human populations (Stoneking et al. 1990; Wallace and Torroni 1992; Torroni et al. 1993; Chen et al. 1995).

With the advent of PCR, the availability of a complete mtDNA sequence has meant that it is a trivial matter to design primers for amplifying and either sequencing or otherwise analyzing variation in any segment of mtDNA that might be of interest (Schurr et al. 1990; Stoneking et al. 1991; Vigilant et al. 1991), up to and including the entire mtDNA genome (Cheng et al. 1994; Horai et al. 1995). This has not only greatly facilitated the analysis of mtDNA variation within and among human popu-

¹E-MAIL mas23@psu.edu; FAX (814) 863-1474.

STONEKING

lations, but it has also made it much easier to determine mtDNA sequences from other species for evolutionary comparisons (Ruvolo et al. 1991; Horai et al. 1992; Morin et al. 1994). Thus, the HGP will aid similarly in analyzing human variation and evolution by making it just as simple to design the primers to amplify any stretch of nuclear DNA that is of interest.

Selection of Appropriate Polymorphisms

The HGP will permit a more rational approach to selecting appropriate regions of nuclear DNA to analyze, depending on what question is being addressed. Historically, the selection of polymorphisms for studies of human nuclear genetic diversity and evolution has been based largely on what is available and what is popular, whereas the discovery of such markers has largely been driven by human disease concerns. For example, the first polymorphism was the ABO blood groups, which were discovered at the beginning of this century by Landsteiner in his efforts to understand and prevent blood transfusion reactions, and were subsequently used by the Hirszfields in their pioneering studies of human genetic diversity (Mourant 1983). As an aside, for those who would question the value of human diversity studies, it is worth pointing out that the inheritance of the ABO blood groups was not worked out by family studies but, rather, by an elegant application of population genetic theory by Bernstein that made use of the data collected by the Hirszfields (Crow 1993).

Subsequent polymorphisms involving gene products in the blood that were initially discovered by the medical genetic community [such as human leukocyte antigen (HLA), immunoglobulin types, and isozymes] we later used in human diversity studies without much consideration given to the suitability of these polymorphisms for the questions that were addressed. This trend continued with the discovery of DNA polymorphisms, including restriction fragment-length polymorphisms (RFLPs), variable number of tandem repeat (VNTR) polymorphisms, and short tandem repeat (STR) polymorphisms, although in some cases these polymorphisms have provided useful information (Bowcock et al. 1994; Mountain and Cavalli-Sforza 1994; Goldstein et al. 1995).

However, as our knowledge of properties of the human genome and the various types of polymorphisms grows, a more rational selection of markers, appropriate to the questions being addressed, is be-

coming possible. And there is plenty of precedence for this in the mtDNA field; it is well known that different mtDNA regions evolve at different rates, and hence are not equally suitable for all studies. For analyzing variation within human populations, the hypervariable segments of the noncoding mtDNA control region provide the most information, whereas for reconstructing primate evolution, more slowly evolving regions such as the mitochondrial rRNA genes are more appropriate—the control region evolves too rapidly to be informative over millions of years of evolutionary history, whereas the rRNA genes evolve too slowly to reveal how variation is apportioned within and between human populations.

It is thus expected that HGP will permit molecular anthropologists to select regions of the genome with evolutionary properties that can be utilized to the greatest advantage. This has already begun to some extent. For example, one type of polymorphism that is proving to be extremely useful for reconstructing human evolution are *Alu* insertion polymorphisms, where the polymorphism consists of the presence or absence of an *Alu* element at a particular chromosomal location (Batzer et al. 1991). Because the insertion of an *Alu* element at a particular locus is a unique event, and because *Alu* elements are rarely deleted (and even then, the deletion is never exact, but leaves behind a footprint), the presence of an *Alu* element marks a chromosomal region as being identical by descent for which the ancestral state is known; these properties have been used to demonstrate that *Alu* insertion polymorphisms indicate a recent African origin of modern humans (Batzer et al. 1994). Similarly, when one considers why mtDNA has proven to be so useful in reconstructing human evolution and population history, it is because not only is mtDNA highly variable, but the lack of recombination provides a framework for interpreting the high levels of variability. This suggests that the most informative nuclear DNA regions for reconstructing human evolution would have these same properties, and two recent studies have focused on analyzing haplotypes that consist of highly variable nuclear DNA regions that are closely linked to (Tishkoff et al. 1996) or encompass (Armour et al. 1996) more slowly evolving polymorphisms. The more slowly evolving polymorphisms enable much more detailed insight into the highly variable polymorphisms, and both studies provide strong support for a recent African origin of modern humans.

Another fruitful source of markers for molecular anthropologists will arise out of the disease gene

studies that will be facilitated by the HGP. As mentioned previously, the HGP is usually promoted by the scientific community in terms of how it will advance our knowledge of genes that contribute to abnormal (disease) variation in humans. Although molecular anthropologists are more concerned with "normal" variation in humans, there is one area in which the two overlap, and that is in tracing the origin, history, and spread of known disease-causing mutations, as, for example, the $\Delta F508$ mutation in the cystic fibrosis gene (Morrall et al. 1994). The expected proliferation of our knowledge of similar such mutations and how they are distributed among human populations is thus another contribution of the HGP to molecular anthropology.

Development of New Technology

Another way in which the HGP will enhance molecular anthropology is in the development of new technology. When I was briefly associated with the HGP at Lawrence Berkeley National Laboratory back in 1989, it was not clear what technology would be used to accomplish the final goal of a complete human DNA sequence. There was some speculation that some novel strategy would be ultimately employed, which would be useful only for generating a single sequence, for example, direct microscopic visualization of DNA. Fortunately for the molecular anthropology community, this has not turned out to be the case. The techniques that are being used to accomplish the final goal, that is, automated, fluorescence-based sequencing with robotic preparation and manipulation of sequencing templates, are readily adaptable to large-scale population analyses. These techniques are already making it possible to obtain data on a scale that was unheard of only a few years ago. For example, in 1983, the first study of human mtDNA variation detected by DNA sequencing was published (Aquadro and Greenberg 1983), involving just seven sequences of the 1-kb control region, and was (rightly) considered a technological tour de force; now, studies of human mtDNA sequence variation routinely include hundreds of individuals (Melton et al. 1995; Sajantila et al. 1995; Sykes et al. 1995). What these technological advances mean for molecular anthropologists is that the questions we ask need not be limited by the amount of data that can be gathered. In the near future it will be possible to analyze thousands or even tens of thousands of individuals, not just for mtDNA but also for long stretches of nuclear DNA.

The "New" Molecular Anthropology

These technological advances will, in turn, permit molecular anthropologists to consider novel kinds of questions. For example, polymorphism at a specific chromosomal location is influenced by the sequence context of the polymorphism in ways that we are just beginning to understand. Studies in other organisms have revealed a surprisingly good correlation between the extent of polymorphism and the amount of recombination in a chromosomal region (Begun and Aquadro 1992). This correlation could reflect either hitchhiking of polymorphisms closely linked to advantageous mutations, or background selection against weakly deleterious mutations, both of which are expected to have a larger effect in regions of low recombination. In any event, it is clear that to adequately explain patterns of variation at a particular locus, a lot more must be known about the surrounding sequence context of the locus, and both the HGP and the resulting technology will prove indispensable in generating and analyzing sequence context.

Another area of great interest to molecular anthropology that the HGP will facilitate is in understanding the genetic basis of "normal" human variation. Currently there is a sharp distinction between the physical anthropology of yesterday, which accumulated vast amounts of information on variation in such traits as dermatoglyphics (fingerprints), skin color, stature, body shape, skeletal traits, etc., and the molecular anthropology of today, which studies variation in genes that are not at all involved with such traits. The HGP will not, of course, reveal which genes influence such traits, but it will serve as a tool that will greatly facilitate the identification of such genes. It is anticipated that molecular anthropologists in the next decade will turn their attention to variation in these sorts of genes, thereby linking normal morphological variation to the underlying genetic variation. As discussed above for mtDNA, the HGP will also facilitate comparative sequencing of the genomes of nonhuman primates, and the identification of genes involved in the morphological differences apparent between humans and nonhuman primates. One of the enduring mysteries of human evolution is how we can differ so greatly at the morphological level from our nearest relative, the chimpanzee, when at the DNA sequence level humans and chimpanzees are ~99% identical. We should soon know if what separates us from chimpanzees are not differences in which genes are present in the DNA but, rather, differences in when and where the same genes are

STONEKING

expressed during development, as was suggested >20 years ago (King and Wilson 1975).

The HGP and the Human Genome Diversity Project

No discussion of the HGP and molecular anthropology would be complete without mentioning the Human Genome Diversity Project (HGDP). The HGDP is a proposal to sample, in a systematic fashion, the genetic diversity of human populations around the world (Cavalli-Sforza et al. 1991; Weiss et al. 1992). The HGDP has come in for its share of controversy, and a full discussion of the merits and drawbacks of the HGDP is beyond the scope of this review. Nevertheless, in considering what the HGDP might mean for molecular anthropology, it is instructive to consider what an analogous resource, the Centre d'Etude du Polymorphisme Humain (CEPH) families, has meant for human mapping and linkage studies. Undeniably, human mapping and linkage studies would have proceeded in the absence of the CEPH family resource, just as studies of human genetic diversity will continue even if the HGDP does not come to fruition. But just as undeniably, the CEPH family resource was extremely valuable in allowing mapping and linkage studies to proceed in a much more organized, coordinated, efficient, and rational manner than would have occurred otherwise. Although it remains to be seen what (if any) shape the HGDP will take, I and others have advocated a structure for the HGDP much like CEPH: That is, a facility that would consist of tens, or even hundreds, of thousands of DNA samples from populations from around the world that investigators could access to study variation in whatever gene or DNA segment they were interested. Although there are ethical and practical considerations that must be addressed, the obstacles to HGDP do not seem insurmountable, and (if done properly) the HGDP could become a resource as valuable to the molecular anthropology community and others interested in human diversity as CEPH has been to the human gene-mapping and linkage community.

CONCLUSION

The HGP will not solve the problems of world hunger, poverty, or increasing crime (despite the claims of some well-meaning but misguided proponents); it may not even cure cancer, heart disease, or other ailments that have some basis in our genes (but it will surely be a tool that will aid in alleviating these

problems). However, there is much for molecular anthropologists to cheer about, and I for one look forward to the completion of the HGP with great anticipation.

REFERENCES

- Anderson, S., A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, and I.G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**: 457–465.
- Aquadro, C.F. and B.D. Greenberg. 1983. Human mitochondrial DNA variation and evolution: Analysis of nucleotide sequences from seven individuals. *Genetics* **103**: 287–312.
- Armour, J.A.L., T. Anttinen, C.A. May, E.E. Vega, A. Sajantila, J.R. Kidd, K.K. Kidd, J. Bertranpetit, S. Pääbo, and A.J. Jeffreys. 1996. Minisatellite diversity supports a recent African origin for modern humans. *Nature Genet.* **13**: 154–160.
- Batzer, M.A., V.A. Gudi, J.C. Mena, D.W. Foltz, R.J. Herrera, and P.L. Deininger. 1991. Amplification dynamics of human-specific (HS) Alu family members. *Nucleic Acids Res.* **19**: 3619–3623.
- Batzer, M.A., M. Stoneking, M. Alegria-Hartman, H. Bazan, D.H. Kass, T.H. Shaikh, G.E. Novick, P.A. Ioannou, W.D. Scheer, R.J. Herrera, and P.L. Deininger. 1994. African origin of human-specific polymorphic Alu insertions. *Proc. Natl. Acad. Sci.* **91**: 12288–12292.
- Begun, D.J. and C.F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- Bodmer W.F. 1986. Human genetics: The molecular challenge. *Cold Spring Harbor Symp. Quant. Biol.* **51**: 1–13.
- Bowcock, A.M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J.R. Kidd, and L.L. Cavali-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**: 455–457.
- Boyce, A.J. and C.G.N. Mascie-Taylor, eds. 1996. *Molecular biology and human diversity*. Cambridge University Press, Cambridge, UK.
- Cann, R.L., W.M. Brown, and A.C. Wilson. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* **106**: 479–499.
- Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* **325**: 31–36.
- Cavalli-Sforza, L.L., A.C. Wilson, C.R. Cantor, R.M. Cook-Deegan, and M.-C. King. 1991. Call for a worldwide survey of human genetic diversity: A vanishing opportunity for the Human Genome Project. *Genomics* **11**: 490–491.

HGP AND MOLECULAR ANTHROPOLOGY

- Chadwick, D. and G. Cardew, eds. 1996. *Variation in the human genome*. John Wiley and Sons, Chichester, UK.
- Chjen, Y.-S., A. Torroni, L.E. Excoffier, A. Santachiara-Benerecetti, and D.C. Wallace. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57**: 133–149.
- Cheng, S., R. Higuchi, and M. Stoneking. 1994. Complete mitochondrial genome amplification. *Nature Genet.* **7**: 350–351.
- Crow, J.F. 1993. Felix Bernstein and the first human marker locus. *Genetics* **133**: 4–7.
- Goldstein, D.B., A.R. Linares, L.L. Cavalli-Sforza, and M.W. Feldman. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci.* **92**: 6723–6727.
- Horai, S., Y. Satta, K. Hayasaka, R. Kondo, T. Inoue, T. Ishida, S. Hayashi, and N. Takahata. 1992. Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* **35**: 32–43.
- Horai, S., K. Hayasaka, R. Kondo, K. Tsugane, and N. Takahata. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci.* **92**: 532–536.
- Jackson, M., T. Strachan, and G. Dover, eds. 1996. *Human genome evolution*. BIOS Scientific Publishers, Oxford, UK.
- King, M.C. and A.C. Wilson. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.
- Marshall, E. and E. Pennisi. 1996. NIH launches the final push to sequence the genome. *Science* **272**: 188–189.
- Melton, T., R. Peterson, A.J. Redd, N. Saha, A.S.M. Sofro, J. Martinson, and M. Stoneking. 1995. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am. J. Hum. Genet.* **57**: 403–414.
- Morin, P.A., J.J. Moore, R. Chakraborty, L. Jin, J. Goodall, and D.S. Woodruff. 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**: 1193–1201.
- Morral, N., J. Bertranpetit, X. Estivill, V. Nunes, T. Casals, J. Gimenez, A. Reis, R. Varon-Mateeva, M. Macek, L. Kalaydjieva, D. Angelicheva, R. Dancheva, G. Romeo, M.P. Russo, S. Garnerone, G. Restagno, M. Ferrari, C. Magnani, M. Claustres, M. Desgorges, M. Schwartz, M. Schwarz, B. Dallapiccola, G. Novelli, C. Ferec, M. de Arce, M. Nemeti, J. Kere, M. Anvret, N. Dahl, and L. Kadasi. 1994. The origin of the major cystic fibrosis mutation ($\Delta F508$) in European populations. *Nature Genet.* **7**: 169–175.
- Mountain, J.L. and L.L. Cavalli-Sforza. 1994. Inference of human evolution through cladistic analysis of nuclear DNA restriction polymorphisms. *Proc. Natl. Acad. Sci.* **91**: 6515–6519.
- Mourant, A.E. 1983. *Blood relations*. Oxford University Press, Oxford, UK.
- Ruvolo, M., T.R. Disotell, M.W. Allard, W.M. Brown, and R.L. Honeycutt. 1991. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. *Proc. Natl. Acad. Sci.* **88**: 1570–1574.
- Sajantila, A., P. Lahermo, T. Anttinen, M. Lukka, P. Sistonen, M.-L. Savontaus, P. Aula, L. Beckman, L. Tranebjaerg, T. Gedde-Dahl, L. Issel-Tarver, A. Di Rienzo, and S. Pääbo. 1995. Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.* **5**: 42–52.
- Schurr, T.G., S.W. Ballinger, Y.-Y. Gan, J.A. Hodge, D.A. Merriweather, D.N. Lawrence, W.C. Knowler, K.M. Weiss, and D.C. Wallace. 1990. Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am. J. Hum. Genet.* **46**: 613–623.
- Stoneking, M., L.B. Jorde, K. Bhatia, and A.C. Wilson. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124**: 717–733.
- Stoneking, M., D. Hedgecock, R.G. Higuchi, L. Vigilant, and H.A. Erlich. 1991. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.* **48**: 370–382.
- Sykes, B., A. Leiboff, J. Low-Beer, S. Tetzner, and M. Richards. 1995. The origins of the Polynesians: An interpretation from mitochondrial lineage analysis. *Am. J. Hum. Genet.* **57**: 1463–1475.
- Tishkoff, S.A., E. Dietzsch, W. Speed, A.J. Pakstis, J.R. Kidd, K. Cheung, B. Bonne-Tamir, A.S. Santachiara-Benerecetti, P. Moral, M. Krings, S. Pääbo, E. Watson, N. Risch, T. Jenkins, and K.K. Kidd. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* **271**: 1380–1387.
- Torroni, A.S., T.G. Schurr, M.F. Cabell, M.D. Brown, J.V. Neel, M. Larsen, D.G. Smith, C.M. Vullo, and D.C. Wallace. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.* **53**: 563–590.
- Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A.C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253**: 1503–1507.
- Wallace, D.C. and A.S. Torroni. 1992. American Indian prehistory as written in the mitochondrial DNA: A review. *Hum. Biol.* **64**: 403–416.
- Weiss, K.M., K.K. Kidd, and J.R. Kidd. 1992. Human genome diversity project. *Evol. Anthropol.* **1**: 78–80.



The human genome project and molecular anthropology.

M Stoneking

Genome Res. 1997 7: 87-91

Access the most recent version at doi:[10.1101/gr.7.2.87](https://doi.org/10.1101/gr.7.2.87)

References This article cites 34 articles, 14 of which can be accessed free at:
<http://genome.cshlp.org/content/7/2/87.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement for ThruPLEX HV DNA-seq. The text "ThruPLEX® HV" is in white on a dark blue background, with "failproof DNA-seq of FFPE & cfDNA" below it. To the right is the Takara logo, which includes a stylized "T" and the word "Takara" in blue, with "Cantech Wako cellartis" in smaller text below.

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>
