Anthropological Genetics in the 21st Century: Introduction

MICHAEL H. CRAWFORD

In this introduction to this special issue of Human Biology titled “Anthropological Genetics in the 21st Century,” I briefly review the historical roots of anthropological genetics and the origins and current results of the Human Genome Project and the Human Genome Diversity Program and discuss their influence on anthropological genetics; I then extrapolate the directions and trends of this field as we enter the third millennium.

All 8 of the contributions to this special issue of Human Biology summarize the current developments in the field of anthropological genetics and delineate future trends and directions. Six of the 8 contributions are based on presentations given at a symposium titled “Anthropological Genetics in the 21st Century” presented at the 14th International Congress of Anthropological and Ethnological Sciences, held in Williamsburg, Virginia, on July 27, 1998. Two new contributions (by Barbujani and Beall) were added after the symposium to fill in some of the major lacunae.

Historical Background

Although the roots of genetics, such as Mendel’s plant breeding experiments, the development of Darwin’s theory of natural selection, and Weismann’s demonstration of the isolation of germ plasm, grew in the 19th century, the field of human genetics began to blossom only in the second half of the 20th century. The dawn of the 20th century saw the rediscovery of Mendel’s laws of inheritance, rapidly followed by the concept of genetic equilibrium, enunciated independently by Castle, Hardy, and Weinberg. Bateson and Garrod independently gave impetus to studies of gene action and the field of biochemical genetics. Their research culminated in the early 1950s with the unraveling of the genetic code by Watson and Crick. By the 1930s much of the groundwork for population genetics was laid by the brilliant insights of Fisher, Haldane, and Wright. In the 1950s and 1960s Li, Crow, and Morton added further refinements and flourishes to the theory of population genetics.

1Department of Anthropology, University of Kansas, Lawrence, KS 66045.

Copyright © 2000 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: HUMAN GENOME DIVERSITY, FUNCTIONAL GENOMICS, HUMAN GENOME PROGRAM
It was at this time that the early anthropological geneticists, such as Roberts, Lasker, Livingstone, and Pollitzer, applied theory from population genetics to human aggregates of anthropological interest. In particular, Roberts’s original research on Tristan da Cunha provided valuable insight into the role of unique historical events in the molding of the gene pool of isolated human populations (Roberts 1968).

In the mid-1960s gene pools of human populations were being characterized primarily through the use of allele frequencies of blood group loci (secondary gene products) and/or proteins (primary gene products). Because blood typing dated back to the turn of the 20th century with the pioneering research on the ABO blood group system by Karl Landsteiner, much more data were available on blood group than protein geographic distributions and variation. With the discovery by Smithies (1955, 1959) of the molecular sieving effect through starch-gel electrophoresis, there was a flurry of activity as human geneticists and anthropologists applied this new method to a wide variety of populations (Crawford 1973). Similarly, in the 1980s, with the addition of pH gradients to gel electrophoresis for the separation of proteins, isoelectric focusing (IEF) further revealed the richness of human variation and dramatically altered the measurable genetic variation (Crawford 1987). Thus each technological and methodological development revealed the existence of greater amounts of genetic variation contained in the human genome.

**DNA and Methodological Breakthroughs.** In the 1970s and 1980s methodological breakthroughs facilitated the direct manipulation and study of DNA. One of these breakthroughs was improvement in the methods of extraction of DNA from blood and tissue. Originally, DNA extraction involved the homogenization of biological tissue, protein denaturing, and then ultracentrifugation. Toxic solvents (phenols and chloroform) and detergents were used on tissues such as placenta and liver, which are highly rich in DNA. These early DNA extraction methods were followed by the development of cell lysis procedures and protein denaturation through the use of proteolytic enzymes. A plethora of commercial kits are now in existence, and they expedite the extraction process from a maximum of 24 hr to less than 3 hr [see Smith (1998) for a review of the methods of DNA extraction and purification].

Another methodological breakthrough was the development of hybridization methods, which are based on the application of heat to double-stranded DNA and which permit the comparison of DNA strands. This use of heat and cooling set the stage for the development of polymerase chain reaction (PCR) techniques. A third breakthrough was the discovery that bacteria produced restriction enzymes that cleaved specific sequences. This permitted the recognition of heritable restriction sites and the subdivision of long chains of DNA into more manageable fragments. And last, PCR methodology was developed to amplify and copy regions of the genome demarcated by primers.
Table 1. Differences in Methodological Approaches between Human Genetics and Anthropological Genetics

<table>
<thead>
<tr>
<th>Anthropological Genetics</th>
<th>Human Genetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broader, biocultural perspective on genetic-environmental interactions</td>
<td>Mechanisms and processes, particularly in disease</td>
</tr>
<tr>
<td>Population focus, pedigrees used to measure familial resemblance</td>
<td>Families of probands, twins, and twin families</td>
</tr>
<tr>
<td>Small, reproductively isolated populations, often non-Western</td>
<td>Large, urban, and clinical samples</td>
</tr>
<tr>
<td>Culturally homogeneous populations</td>
<td>Samples may be heterogeneous by race, socioeconomic factors, occupation, lifestyle</td>
</tr>
<tr>
<td>Sampling representative of normal variation in a population</td>
<td>Sampling often based on clinical ascertainment</td>
</tr>
<tr>
<td>Attempts made to characterize and measure the environment</td>
<td>Environmental variation is rarely assessed; it is assumed that $c^2 = 1 - h^2$</td>
</tr>
<tr>
<td>Study of normal variation in complex traits</td>
<td>Dichotomy of disease vs. normality</td>
</tr>
</tbody>
</table>

The development of this elegant method permitted scientists to experiment with minute quantities of DNA.

The Rise of Anthropological Genetics

Anthropological genetics was formalized as a field of inquiry during the 1970s and 1980s, with the initial twin foci of population structure and genetic-environmental interactions in complex traits (Crawford and Workman 1973; Mielke and Crawford 1980; Crawford and Mielke 1982). Table 1 summarizes the methodologies of anthropological genetics and compares them to the approaches used in human genetics. Although both fields study the genetics of humans, anthropological geneticists tend to focus on normal variation in non-Western reproductively isolated populations. Most practitioners of human genetics are more disease oriented, examining processes in families of probands.

In the late 1980s DNA markers became powerful tools for the reconstruction of human phylogeny. In particular, because of its maternal transmission, absence of recombination, and abundance in skeletal tissue, mtDNA proved to be a useful chronometer. Changes in the mtDNA molecules result entirely from the accumulation of mutations over time. The first contribution to this special issue by O’Rourke et al. provides a state-of-the-art overview of ancient mtDNA variation in native American skeletal populations. O’Rourke et al. observe considerable variation in the haplogroup frequencies of both contemporary and ancient populations and suggest that the patterns observed in contemporary populations have their roots in pre-Columbian
times. In the next millennium the history and phylogeny of our species will be explored with much greater precision and will be elucidated through the application of molecular genetics to osseous and mummified materials.

**Human Genome Project**

During the late 1980s, the National Academy of Sciences and the Office of Technology Assessment jointly recommended a concerted human genome program, with the goal of mapping and sequencing the 3 billion base pairs (bp) that constitute the human genome. This recommendation was followed in 1990 first by a 5-year and then by a 15-year plan, presented to Congress by the National Institutes of Health and the Department of Energy. This monumental undertaking, originally scheduled for completion in 2005, has already resulted in major repercussions on the biological sciences and anthropological genetics.

**Where Are We in the Sequencing of the Genome?** To date, the major accomplishments of the Human Genome Project include the publication in 1992 of a low-resolution genetic linkage map for the entire human genome. This was followed in 1995 by the publication of high-resolution physical maps for chromosomes 16 and 19 and moderate-resolution maps for chromosomes 3, 11, 12, and 22. A detailed human physical map containing 52,000 STSs (short 200–300-bp sequence tagged sites of DNA, occurring once in the human genome with known location and base sequences) has been constructed (Collins et al. 1998). These STSs are useful for localizing and orienting the mapping and sequence data and serve as landmarks on the physical map of the human genome (Wang et al. 1998). A detailed human physical map (locations of identifiable landmarks on DNA, e.g., restriction enzyme cutting sites and genes) was prepared. In 1996–1997 the sequences of the human T-cell receptor region were completed, and high-resolution physical maps for chromosomes X and 7 were completed. Stimulated by the potential threat of the human genome being sequenced by a commercial enterprise and bolstered by additional funding, Collins et al. (1998) proposed a new plan that would complete the full sequencing by the year 2003. Thus, as anthropological genetics enters the 21st century, the sequencing of the human genome should be complete.

Because of the conservative nature of DNA and the wealth of information about the function of specific regions contained in the genomes of simpler organisms, researchers first sequenced the simplest genomes. In 1995 Fleischmann et al. were the first to sequence the entire genome of a free-living organism, the bacterium *Haemophilus influenza* (1,830,137 bp). This was followed by Fraser et al. (1995) publishing the complete DNA sequence of the smallest bacterium, *Mycoplasma genitalium* (580,070 bp), displaying
the minimum number of genes needed for independent existence. In 1996 the genome was sequenced for *Methanococcus jannaschii* (confirming the existence of a third major branch of life, the Archaen) (Bult et al. 1996). In 1997 the *Escherichia coli* genome was sequenced (Blattner et al. 1997), followed by the deciphering of the complete genome sequence of *Mycobacterium tuberculosis* (Cole et al. 1998). The *M. tuberculosis* genome is made up of 4,411,529 bp and approximately 4,000 coding genes, with an atypically large number involved in the production of enzymes. Similarly, sequencing of the genomes of the fruitfly (*Drosophila*) and the mouse should also be completed by 2003. This wealth of sequence data should aid in the reconstruction of the evolution of the genome and provide functional genomics with a running start.

**Who Will Own the Genome?** In May 1998 a new partnership was formed between Perkin-Elmer Corporation and J. Craig Venter, a DNA sequencing specialist. The new company promises to complete the sequencing of the human genome in 3 years for $300 million, a fraction of the cost envisaged by the government-sponsored Human Genome Project. This attempt by the private sector to sequence the human genome is being made possible because Perkin-Elmer contributed 230 advanced sequencers with the potential to identify 100 million bases of sequence data per day. Venter promised that within 1 year 99% of the genome will be completed. The following 2 years would be spent fitting together pieces and filling in the gaps. However, there are numerous skeptics who predict that the large gaps created by this method will not be successfully filled through the use of Venter’s whole-genome shotgun approach (i.e., chopping the genome into overlapping pieces and reassembling them into a complete genome using computer programs).

One possible result of this privately funded sequencing of the human genome will be an attempt by pharmaceutical companies to patent any genes of pharmacological interest. The president of Perkin-Elmer was quoted in *Science* as saying that there are about 100 to 300 such patentable genes. In my opinion the patenting of genes may further undermine the worldwide Human Genome Diversity Program. Suspicion and fear already exist that rare DNA sequences of individuals will be patented and that researchers will exploit the genes of native peoples around the world. I believe that some legislation or at least a legally binding agreement is in order to prevent the patenting of any raw DNA data. All sequencing data should be added to public databases and made available to both the public and the private sectors.

**Human Genome Diversity**

Despite the accomplishments of the Human Genome Project, the program is focused on a generic genome that until recently ignored the variation in human populations generated by the forces of evolution. Yet the rate of
polymorphism in the human species has been estimated as about 1 variant per 500 bp with about 15% of the variation consisting of insertions or deletions (Hieter and Boguski 1997). These figures were derived from an analysis of 290 kb of the human beta-T-cell receptor locus for which more than 1 haplotype was sequenced (Hieter and Boguski 1997). In 1991 Cavalli-Sforza announced an initiative to survey genomic variation across the entire human species. Although this bold plan was supported by most population and anthropological geneticists, there was some opposition from a few anthropologists and indigenous organizations, who accused the organizers of the Human Genome Diversity Program of racism and the exploitation of third world genomes. Until the ethical and legal questions could be resolved, the basic research envisaged by the Human Genome Diversity Program simmered on a back burner.

Because of the controversy surrounding the Human Genome Diversity Program, a National Research Council committee (chaired by William J. Schull) was convened in early 1996 to assess the scientific value, technical aspects, and organizational requirements for a systematic worldwide survey of human genetic variability and the legal and social implications of such a survey. In its final report, titled *Evaluating Human Genetic Diversity*, the committee recommended several strategies for the project, including the collection of samples from unrelated individuals within populations (National Research Council 1998). The identity of these donors cannot be linked to the samples, thereby protecting individual privacy but eliminating the possibility of uncovering any disease genes in these samples. A sampling of populations on a worldwide basis would measure human genetic diversity and permit the reconstruction of human phylogeny. However, the biomedical research cannot be appended to the goals of the Human Genome Diversity Program. Recently, the National Institutes of Health added a diversity component to the Human Genome Project by characterizing SNPs ("snips"—single nucleotide polymorphisms) in 4 major US population categories.

With the sponsorship of the Human Genome Diversity Program by both the National Science Foundation and the National Institutes of Health in the United States, the evaluation of human diversity on a worldwide basis should preoccupy anthropological geneticists well into the 21st century. A focused research program of this sort should also underwrite the development of anthropological genetics and anthropology departments, thus attracting graduate students and faculty into this fledgling field.

**Functional Genomics**

Once the human genome has been successfully sequenced, the function of more than 100,000 genes will have to be identified. This is the most challenging future in the post–Human Genome Project era. Elucidating the genetic
basis of human phenotypes (both simple and complex) is a daunting task. Knowing the sequence of individual genes does not reveal anything about the complexities of life. A gene directs the synthesis of protein, but a given sequence does not reveal how proteins interact, how cells and tissues communicate, how organs come into being, or how evolution works. These tasks should preoccupy the molecular and anthropological geneticists of the next millennium.

Where do we start in identifying the functions of the 3 billion bp that will be known by the year 2003? Dryja (1997) pointed out that currently 5,000 or more phenotypes are known and an additional 60,000 to 70,000 transcriptional units have been estimated. He suggested 2 approaches for identifying the connections between specific genes and phenotypes: the phenotype-based approach and the gene-oriented approach.

**Phenotype-Based Approach.** The phenotype-based approach, or reverse genetics method, begins with a phenotype and concludes with the identification of the responsible gene or genes. Individuals having a particular phenotype are identified and analyzed. Clues pointing to the responsible genes are obtained from biochemical and genetic analyses. The genes are evaluated until a gene is found with sequence abnormalities specific for that phenotype. This method has been used for most human gene-disease correlations. For the phenotype-oriented approach the essential intermediate step is the determination of the approximate chromosomal location for the gene in question. Once the appropriate chromosomal location is identified, the functional evaluation should be limited to only the genes in that region (i.e., candidate genes). In phenotypes with a specific biochemical abnormality, the scope of the gene scan can be narrowed to those genes expressing proteins in the defective pathway.

Two papers in this special issue review phenotype-based approaches. In the second contribution to this special issue Blangero and his colleagues from the Southwest Foundation for Biomedical Research review the methodologies for mapping quantitative trait loci (QTLs). This group of researchers has been on the cutting edge of QTL mapping through their application of variance components-based methodologies. They have already revealed some of the QTLs responsible for complex phenotypes, such as skin pigmentation, stature, and numerous polygenic diseases in Mexican American populations. During the next millennium, we should be able to discover the function of many genes and their locations that influence our normal and disease complex phenotypes.

Terwilliger and Göring, in the third contribution to this special issue, review the strategies used in linkage analysis and the mapping of genes. They stress the importance of experimental design in the study of complex phenotypes because this strategy increases power for gene mapping. They eval-
uate the use of population isolates and admixture mapping, both approaches long considered within the scope of anthropological genetics.

**Gene-Oriented Approach.** The gene-oriented approach, or candidate gene approach, is a less powerful method and begins with the isolation of a gene followed by the partial characterization of its gene product. Research is directed toward uncovering what complex human phenotype would result from mutations of that gene. Kamboh et al. (1996) used this approach to examine the relationship between apolipoprotein genes (*APOA4, APOH, and APOE*) and quantitative plasma lipid levels in Evenki reindeer herders. The results indicated that the well-established association between the *APOE*4 allele and elevated plasma cholesterol levels in European populations was not observed among the Evenki, despite their having a comparable frequency of the *APOE*4 allele. In fact, in Evenki women the *APOE*4 allele was associated with significantly lower low-density lipoprotein cholesterol levels compared with the *APOE*3 allele. The apparent absence of the positive association of the *APOE*4 allele in this study may be due to significantly lower plasma cholesterol levels in the Evenki compared with American and European whites. The most likely explanation for this finding is a gene-diet interaction that modulates the effects of the *APOE* polymorphism. Despite consuming a diet that contains a large amount of reindeer meat, the Evenki do not have a high saturated fat intake, which contrasts with industrialized Western societies. The Evenki also exhibit high levels of energy expenditure, which adds to the explanation of why they have low cholesterol levels despite having a similar frequency of the atherogenic *APOE*4 allele.

With complex phenotypes, such as levels of plasma cholesterol, triglycerides, and high- and low-density lipoproteins, there is evidence of genetic-environmental interaction, such as lifestyle and nutritional patterns. This quantitative characterization of the environment is an area that is vastly overlooked by human geneticists who are examining genetic-environmental interactions. In the 21st century anthropological geneticists could contribute much to this area of research by developing methodologies for more fully defining environmental covariates and indexes. To date, covariates such as sex, age, education, and occupation are used with complex phenotypes. Yet adequate assessment of nutritional intake, accurate measures of activity patterns, and psychological variables are rarely used in studies of interaction between complex lipid phenotypes and the environment.

**Gene Geography.** Barbujani, in the fourth contribution to this special issue, provides a critical overview of the methodologies applied to the geographic distributions of genes. He stresses that the observed geographic patterns provide insight into population history, evolution of inherited disease, and the reconstruction of demographic processes. Yet he cautions readers about the application of old models to molecular data. When the Human
Genome Diversity Program becomes a reality in the next millennium and comparable molecular genetic data are collected worldwide, the formal analyses of the distribution of these genes should yield great insight into the evolutionary history of our species.

**Admixture.** In the next millennium, with ever greater human mobility and migrations generated by economic opportunities, ethnic conflicts, and ideological persecution, our understanding of admixture and its biological repercussions on human populations will be of great importance. There will be fewer and fewer isolated, so-called anthropological populations in existence as the world population increases at its current precipitous rate. Because of its colonial heritage and vast cultural and genetic diversity, Latin America displays considerable gene flow and hybridization. Sans, in the fifth article of this special issue, examines admixture studies of Latin America and traces possible trends as we enter the 21st century.

**Human Adaptability.** One of the most significant influences on the direction of human biology and anthropological genetics of the 1960s and 1970s was the International Biological Program (IBP). This program sponsored research on human populations of the Andes, the Amazon, and the Arctic. Recent applications of anthropological genetics to the data gathered in physiological adaptability studies greatly strengthen this field as it enters the next millennium. Little and Garutto, in the sixth contribution to this special issue, trace the early studies of human adaptability from the IBP to the present and suggest future directions for this field.

Beall, in the seventh contribution to the special issue, provides evidence of an effective synthesis of anthropological genetics and the physiological adaptation to high-altitude hypoxia. Her unique comparison of Tibetan and Andean population responses to the same physiological stresses indicates differences in the adaptive response. Her research shows the direction for future human adaptation studies that combine physiological measures with quantitative genetic and molecular approaches. Such synthetic methodologies will revolutionize the field of human adaptability.

**Behavioral Genetics.** In the last contribution to this special issue Gilger reviews some of the fundamental questions and approaches of human behavioral genetics as we enter the next millennium. Human behavioral genetics has always been a highly controversial and politically sensitive field. At times, research in behavioral genetics (particularly intelligence studies) has been used for political or social agenda. Yet behavioral and molecular genetic methods reveal the underlying genetic architecture of human behavior. Most critics are willing to accept the overwhelming evidence in support of genetic control or influence of behavior in animal models yet argue against human behavioral genetics. Gilger clearly demonstrates how statistical approaches
are being used in conjunction with developing molecular genetic methodologies to elucidate the genetics of complex behavioral phenotypes.

Conclusion

The third millennium offers anthropological geneticists unique research opportunities. With the completion of the sequencing of the human genome functional genomics will focus on gene expression, defining the role of each gene, and elucidating how the genome functions as a whole in the complex natural history of the human organism. Worldwide diversity studies should yield a rich harvest of thousands of polymorphic coding genes and may provide some insight into how these unique sequences affect gene products. Anthropological geneticists, because of their field orientation, biocultural focus, and skill in defining environmental covariates, should become important players in the next millennium.

Where do we go from here? A recent book by Horgan (1996), a science writer for *Scientific American*, based on interviews with eminent scientists, concluded that “science . . . has ended” and that “there will be no great revelations in the future comparable to those bestowed upon us by Darwin or Einstein or Watson or Crick” (p. 16). Have the basic or fundamental questions of biological science been answered, and will we in the future be merely filling in gaps? I believe that this pessimistic assessment of biological sciences overlooks the role of technological breakthroughs, such as those associated with the manipulation of DNA. We are currently experiencing the equivalent of an adaptive radiation following the technological discoveries linked with the human genome. New technology followed by new scientific approaches and new methodologies should generate a plethora of new fundamental genetic questions as we move into the third millennium.

*Received 4 October 1999.*

**Literature Cited**


