Briefing Book

for the Hearing on

The Human Genome Diversity Project

Before the

Committee on Governmental Affairs

United States Senate

April 26, 1993

National Center for Human Genome Research

National Institutes of Health
Public Health Service
Department of Health and Human Services

TABLE OF CONTENTS

If you have any questions concerning the hearing, please contact Cora Yamamoto or David Sandler of Senator Akaka's staff at [redacted]

Sincerely,

John Glenn

JOHN GLENN
U.S. Senator

Daniel K. Akaka

DANIEL K. AKAKA
U.S. Senator

SHANE MERZ
OFFICE OF SENATOR DANIEL K. AKAKA
720 HART SENATE OFFICE BUILDING
WASHINGTON, DC 20510-1103
Phone [redacted]

FAX COVER SHEET

TO: Ann Rufo

PHONE:

FAX: [redacted] ✓

FROM: SHANE MERZ [redacted]

DATE: April 16, 1993

REMARKS: Please call me if you have questions or comments about the issues to be discussed at the Human Genome hearing on April 26.

We are sending you _3_ pages, including this cover sheet. If you do not receive all of the pages, please contact our office at the phone number listed above.

DRAFT - April 21, 1993

Statement of Francis S. Collins, M.D., Ph.D.
Director
National Center for Human Genome Research

Senate Committee on Governmental Affairs
Hearing on the Human Genome Diversity Project
April 26, 1993

Mr. Chairman, Senator Akaka, and members of the Committee, it is my pleasure to appear before you today as the new Director of the National Center for Human Genome Research (NCHGR) of the National Institutes of Health (NIH) and to have the opportunity to talk to you about the Human Genome Project. I am enormously pleased to be at the helm of what I consider to be the single most important scientific endeavor we have ever embarked upon. By the end of the 15-year project, we hope to have produced detailed maps of all the human chromosomes and determined the sequence of the 3 billion pairs of nucleotide bases that make up human DNA. This information will be stored in databases that will allow researchers to have access to any region of the human genome right at their fingertips. The multitude of benefits we are witnessing already from the Human Genome Project is only the beginning of what I believe will be a revolution in molecular medicine and human biology.

Goals of the Human Genome Project

The Human Genome Project is an international research effort that has the goal of analyzing the structure of human DNA and determining the location of the estimated 100,000 genes located on 23 pairs of human chromosomes. In the United States, the Human Genome Project is managed principally by two government agencies, the NIH

and the Department of Energy (DOE), who together have set specific scientific goals in an initial five-year plan to manage this historic effort. You will also be hearing testimony today from Dr. David Galas of the DOE.

The primary mission of the Human Genome Project is to develop research tools--chromosome maps, DNA sequence information, laboratory technology, and computer databases--that will allow researchers to find and analyze genes faster, more easily, and more cheaply. These tools will have tremendous benefits for biomedical research and make important contributions to a variety of research projects. Of primary interest are the extraordinary medical benefits that will result from our ability to understand the genetic basis of health and disease. Gene discovery gives researchers the opportunity to study the function of the gene and its role in cell biology. This knowledge will revolutionize our strategies to diagnose, treat, and even prevent many diseases.

Medical Benefits

The pace of disease gene discovery has increased substantially because of the research tools developed by the Human Genome Project. We have seen the discovery of genes responsible for genetic diseases such as cystic fibrosis, neurofibromatosis type I and II, fragile X, and most recently, you may have heard about the discovery of the gene responsible for Huntington's disease. In collaboration with another of today's witnesses, Dr. Mary-Claire King, my colleagues and I are also zeroing in on a gene that causes breast cancer--I expect that the gene will be located within the next year. It is a gene that 1 out of 200 women inherit; these women have an 85% chance of getting breast

cancer and have an increased risk of ovarian cancer as well. We hope soon to be able to offer diagnostic testing for women at risk, and eventually develop ways to treat and prevent this type of early-onset breast cancer.

The isolation of the cystic fibrosis gene provides a good example of the progress that can be made in understanding a disease once the gene is isolated. I was part of a team of researchers who isolated the gene in 1989, and I am excited to report that we have already begun to design new, highly specific drugs to treat cystic fibrosis and are in the process of developing gene therapy techniques. The ability to begin gene therapy trials less than four years after the gene was discovered is a dramatic example of the power of the new tools of human molecular genetics. I am exceptionally pleased that the leadership of NIH has recently established a Division of Intramural Research within the NCHGR which will focus on technologies for finding disease genes and developing subsequent DNA diagnostics and gene therapies.

The Human Genome Project is cutting edge basic science, and is providing the research tools to accelerate our understanding of the biological and molecular processes that permit human life to develop and function. This will profoundly affect our ability to understand the molecular basis of disease and will greatly improve our ability to diagnose, treat, and prevent many common diseases resulting from malfunctions or mutations in our genes. Such diseases represent a major fraction of the chronic conditions that account for most of the health care costs today.

The Human Genome Diversity Project

The Human Genome Diversity Project, which we are here to discuss today, is one of the many research projects that will be greatly benefitted by the genome analysis tools being developed by the Human Genome Project. The objectives of the Human Genome Diversity Project are to collect, analyze, and preserve genetic samples from a host of vanishing human populations. This project has the potential of giving us new knowledge about human origins, evolutionary history, and genetic diversity; it may also eventually lead to a better understanding of the frequency and susceptibility to disease among diverse populations. It is timely to discuss the undertaking of the Human Genome Diversity Project; I am mindful that researchers wish to collect the samples now before we lose the opportunity due to the further breakdown of geographical barriers, war, famine, or disease.

The Human Genome Diversity Project is beyond the mission of the federally funded Human Genome Project. The NIH and DOE are facing challenges to accomplish the goals we set out to achieve within the next 15 years. We have a long road ahead to complete the genetic and physical maps of all the human chromosomes, and it is essential that there be further improvements in sequencing technology if we wish to sequence the entire human genome quickly and efficiently. We have been able to make significant progress because we have clearly defined our mission. At this time, the Human Genome Project's major contribution to the Human Genome Diversity Project would be for us to continue to develop the research tools that will allow genetic

diversity research, as well as other scientific research disciplines, to proceed in a cost-effective way. There may be aspects of technology development, for example high throughput genotyping of many DNA samples, which will benefit both the Human Genome Project and genetic diversity studies, and we look forward to exploring those complementary areas.

I believe the Human Genome Diversity Project is a valuable international endeavor, and that the NIH should cooperate in this project. We are pleased that the NCHGR was able to participate in and contribute funding to the National Science Foundation's grant to the Human Genome Organization (HUGO) Committee for Human Genetic Diversity for a series of workshops on the Human Genome Diversity Project. The National Institute of General Medical Sciences at NIH and the DOE were also contributors. The most recent workshop took place on February 16-18, 1993, on the NIH campus in Bethesda, Maryland, and it included discussions on the ethical issues surrounding genetic diversity research.

Ethical, Legal, and Social Issues

The information generated by the Human Genome Diversity Project will further our knowledge about our human origins and evolution, but it also will raise some challenging ethical, legal, and social issues that must be identified and addressed before the project begins. The February workshop raised several of these important issues including: (1) the ethical issues raised in doing biomedical research in developing countries and insuring the protection of human subjects (for example, it is difficult to

ascertain what "informed consent" means in other cultures to know if an individual is voluntarily participating in the sample collection process); (2) the legal issues raised by the possible commercial value of the project's samples or results; and (3) the social-political issues surrounding the possible misuse or misinterpretation of the information generated. Research concerning human genetic differences always merits careful attention to avoid notions of superiority or inferiority among diverse populations.

The Human Genome Project has also faced challenging issues related to the use of human genetic information. At the NCHGR, the Ethical, Legal, and Social Implications Branch was created to define these issues and to develop initial policy options to address them. Five percent of our budget is devoted to the activities of this branch. Our experience has shown that the need is great to examine the ethical, legal, and social issues--alongside with the scientific research--in order to minimize any adverse social consequences resulting from the generation of genetic information.

I would recommend that the Human Genome Diversity Project, as it is organized and funded, develop mechanisms to address the ethical and legal challenges the project will encounter. I would further suggest that an advisory group be established to monitor the research and the use of the information it generates. My staff and I would be happy to provide consultation and advice based on our own experience in this endeavor.

I would be pleased to answer any questions that you may have.

DRAFT - April 16, 1993

Statement of Francis S. Collins, M.D., Ph.D.
Director
National Center for Human Genome Research

Senate Committee on Governmental Affairs
Hearing on the Human Genome Diversity Project
April 26, 1993

Mr. Chairman, Senator Akaka, and members of the Committee, it is my pleasure to appear before you today as the new Director of the National Center for Human Genome Research (NCHGR) of the National Institutes of Health (NIH) and to have the opportunity to talk to you about the Human Genome Project. I am enormously pleased to be at the helm of what I consider to be the single most important scientific endeavor we have ever embarked upon. By the end of the 15-year project, we will have produced detailed maps of all the human chromosomes and determined the sequence of the 3 billion pairs of nucleotide bases that make up human DNA. This information will be stored in databases that will allow researchers to have access to any region of the human genome right at their fingertips. The multitude of benefits we are witnessing already from the Human Genome Project is only the beginning of what I believe will be a revolution in molecular medicine and human biology.

Goals of the Human Genome Project

The Human Genome Project is an international research effort that has the goal of analyzing the structure of human DNA and determining the location of the estimated 100,000 genes located on 23 pairs of human chromosomes. In the United States, the Human Genome Project is managed principally by two government agencies, the NIH

and the Department of Energy (DOE), who together have set specific scientific goals in an initial five-year plan to manage this historic effort. You will also be hearing testimony today from Dr. David Galas of the DOE.

The primary mission of the Human Genome Project is to develop research tools--chromosome maps, DNA sequence information, laboratory technology, and computer databases--that will allow researchers to find and analyze genes faster, more easily, and more cheaply. These tools will have tremendous benefits for biomedical research and make important contributions to a variety of research projects. Of primary interest are the extraordinary medical benefits that will result from our ability to understand the genetic basis of health and disease. Gene discovery gives researchers the opportunity to study the function of the gene and its role in cell biology. This knowledge will revolutionize our strategies to diagnose, treat, and even prevent many diseases.

Medical Benefits

The pace of disease gene discovery has increased substantially because of the research tools developed by the Human Genome Project. We have seen the discovery of genes responsible for genetic diseases such as cystic fibrosis, neurofibromatosis type I and II, fragile X, and most recently, you may have heard about the discovery of the gene responsible for Huntington's disease. In collaboration with another of today's witnesses, Dr. Mary-Claire King, my colleagues and I are also zeroing in on a gene that causes breast cancer--I expect that the gene will be located within the next year. It is a gene that 1 out of 200 women inherit; these women have an 85% chance of getting breast

cancer and have an increased risk of ovarian cancer as well. We hope soon to be able to offer diagnostic testing for women at risk, and eventually develop ways to treat and prevent this type of early-onset breast cancer.

The isolation of the cystic fibrosis gene provides a good example of the progress that can be made in understanding a disease once the gene is isolated. I was part of a team of researchers who isolated the gene in 1989, and I am excited to report that we have already begun to design new, highly specific drugs to treat cystic fibrosis and are in the process of developing gene therapy techniques. The ability to begin gene therapy trials less than four years after the gene was discovered is a dramatic example of the power of the new tools of human molecular genetics. I am pleased that the NCHGR has recently established a Division of Intramural Research on the NIH campus which will focus on technologies for finding disease genes and developing subsequent DNA diagnostics and gene therapies.

The Human Genome Project is cutting edge basic science, and is providing the research tools to accelerate our understanding of the biological and molecular processes that permit human life to develop and function. This will profoundly affect our ability to understand the molecular basis of disease and will greatly improve our ability to diagnose, treat, and prevent many common diseases resulting from malfunctions or mutations in our genes. Such diseases represent a major fraction of the chronic conditions that account for most of the health care costs today.

The Human Genome Diversity Project

The Human Genome Diversity Project, which we are here to discuss today, is one of the many research projects that will be greatly benefitted by the genome analysis tools being developed by the Human Genome Project. The objectives of the Human Genome Diversity Project are to collect, analyze, and preserve genetic samples from a host of vanishing human populations. This project has the potential of giving us new knowledge about human origins, evolutionary history, and genetic diversity; it may also eventually lead to a better understanding of the frequency and susceptibility to disease among diverse populations. It is timely to discuss the undertaking of the Human Genome Diversity Project; I am mindful that researchers wish to collect the samples now before we lose the opportunity due to the further breakdown of geographical barriers, war, famine, or disease.

The Human Genome Diversity Project is, at this point, beyond the current goals and budget of the federally funded Human Genome Project. The NIH and DOE are facing challenges to accomplish the goals we set out to achieve within the next 15 years. We have a long road ahead to complete the genetic and physical maps of all the human chromosomes, and it is essential that there be further improvements in sequencing technology if we wish to sequence the entire human genome quickly and efficiently. Our task has been made more difficult by the fact that the NIH's component of the Human Genome Project has never received the funding that was originally projected. We have been able to make significant progress because we have clearly defined our mission. At

this time, the Human Genome Project's major contribution to the Human Genome Diversity Project would be for us to continue to develop the research tools that will allow genetic diversity research, as well as other scientific research disciplines, to proceed in a cost-effective way. There may be aspects of technology development, for example high throughput genotyping of many DNA samples, which will benefit both the Human Genome Project and genetic diversity studies, and we look forward to exploring those complementary areas.

I believe the Human Genome Diversity Project is a valuable international endeavor, and that the NIH should cooperate in this project. We are pleased that the NCHGR was able to participate in and contribute funding to the National Science Foundation's grant to the Human Genome Organization (HUGO) Committee for Human Genetic Diversity for a series of workshops on the Human Genome Diversity Project. The National Institute of General Medical Sciences at NIH and the DOE were also contributors. The most recent workshop took place on February 16-18, 1993, on the NIH campus in Bethesda, Maryland, and it included discussions on the ethical issues surrounding genetic diversity research.

Ethical, Legal, and Social Issues

The information generated by the Human Genome Diversity Project will further our knowledge about our human origins and evolution, but it also will raise some challenging ethical, legal, and social issues that must be identified and addressed before the project begins. The February workshop raised several of these important issues

including: (1) the ethical issues raised in doing biomedical research in developing countries and insuring the protection of human subjects (for example, it is difficult to ascertain what "informed consent" means in other cultures to know if an individual is voluntarily participating in the sample collection process); (2) the legal issues raised by the possible commercial value of the project's samples or results; and (3) the social-political issues surrounding the possible misuse or misinterpretation of the information generated. Research concerning human genetic differences always merits careful attention to avoid notions of superiority or inferiority among diverse populations.

The Human Genome Project has also faced challenging issues related to the use of human genetic information. At the NCHGR, the Ethical, Legal, and Social Implications Branch was created to define these issues and to develop initial policy options to address them. Five percent of our budget is devoted to the activities of this branch. Our experience has shown that the need is great to examine the ethical, legal, and social issues--alongside with the scientific research--in order to minimize any adverse social consequences resulting from the generation of genetic information.

I would recommend that the Human Genome Diversity Project, as it is organized and funded, develop mechanisms to address the ethical and legal challenges the project will encounter. I would further suggest that an advisory group be established to monitor the research and the use of the information it generates. My staff and I would be happy to provide consultation and advice based on our own experience in this endeavor.

I would be pleased to answer any questions that you may have.

WITNESS LIST

Hearing on the Human Genome Diversity Project
Committtee on Governmental Affairs
United States Senate
April 26, 1993

Francis S. Collins, M.D., Ph.D.
Director
National Center for Human Genome Research
National Institutes of Health

David J. Galas, Ph.D.
Associate Director for Health and Environmental Research
Department of Energy

Roger Herdman
Acting Director
Office of Technology Assessment

Cora Marrett
Assistant Director
Social, Behavioral, and Economic Sciences
National Science Foundation

Luigi Luca Cavalli-Sforza, M.D.
Professor, Department of Genetics
School of Medicine
Stanford University, Stanford, CA

Mary-Claire King, Ph.D.
Professor of Genetics and Epidemiology
School of Public Health
University of California, Berkeley, CA

Ruth Kirschstein, M.D. *
Director
National Institute of General Medical Sciences
National Institutes of Health

* Tentative

FACSIMILE COVER SHEET

U.S. Department of Energy
Washington, D.C. 20585

FAX No.: (301) 903-5051

Verification No. (301) 903-3764

Date: 4/15/93

TO: Elke Jordan
NIH

FAX No.: [redacted]

[redacted]

David A. Smith, Ph.D.
Director, Health Effects
and Life Sciences Research Division, OHER, ER-72

Office Tel No.: [redacted]

MESSAGE:

This is the latest version - right out of the computer at 11:30 today (4/15/93)

Dave

NUMBER OF PAGES (INCLUDING THIS PAGE) 12

PLEASE CALL (301) 903-3764 (OR SENDER) IF RE-TRANSMITTAL IS NECESSARY.

FACSIMILE COVER SHEET

11:30am
4/15/93

DRAFT
April 13, 1993

Statement of David J. Galas

Associate Director for Health and Environmental Research

Office of Energy Research

U.S. Department of Energy

before the

Committee on Governmental Affairs

United States Senate

April 26, 1993

Approved by: _____
William Happer

Date: _____

Mr. Chairman and Members of the Committee:

I am very pleased to have the opportunity to discuss the historic Human Genome Project with you today. In many ways, the Human Genome Project (HGP) epitomizes the promise of the biological sciences for the future. The impact on our lives of the vast amount of new knowledge of the living world is just beginning to be realized--it will inevitably transform the biological sciences, medicine, as well as agriculture, food manufacture, chemical technology, and many other areas in ways that are difficult to predict. The project is even now producing a strong shift in the biomedical sciences, from descriptive phenomenology to fundamental understanding, and changes in the practice of medicine will not be far behind. One of the most significant of these latter changes will be a shift in emphasis from curative to preventative medicine--from therapy to prevention. This shift will soon begin to occur because the research that focuses on the fundamental genetic determinants of the functioning of the human body will enable more and more detailed understanding of the nature and causes of human diseases. This will, in turn, allow effective prevention, early intervention and more benign and less costly treatment. This knowledge will include an understanding of the propensities of different individuals for different diseases and susceptibilities to health-endangering practices and environmental influences.

The Department of Energy (DOE) has sponsored the Human Genome Project (HGP) from its inception. At that time, it was recognized that the essential technology which would enable us to undertake this historic project was within our reach, and that immense potential benefits to biology and medicine could be realized. The DOE and the National Institutes of Health (NIH) jointly

support this national research effort, carefully coordinating our respective programs.

The HGP is a project that also holds within its promise some significant issues and questions concerning the nature and uses of human genetic information. While the promise of human genome research is enormous, some of the ethical, legal, and social issues associated with this new knowledge are serious and difficult. The HGP has supported an unprecedented effort--more than 3 percent of its budget is devoted to the effort--to study these issues and examine the ethical, legal and social consequences of the coming deluge of genetic information and the new, associated technologies. These issues include the proper legal restraints on discrimination based on genetic information and the rights to privacy of genetic information and their consequences. It has been well recognized that ethical, legal, and social impacts will accompany greater knowledge about the human genome and that some attempts need to be made to anticipate, and to try to ameliorate, these impacts. Another issue has to do with the proper distribution for the public good of intellectual property rights for genetic material and information--patents on genes.

We consider this historic project to be of immense potential value to the American people and to the world. Thus, we strongly support the expeditious transfer of the benefits of this research and of the associated technologies to the public, which of necessity involves the patent system. This issue has

been discussed elsewhere, and I will not comment further here on the gene patenting issue.

Goals and Progress of the Human Genome Project

The general approach to elucidating the genetic contents of the genome is to physically "map" the chromosomes and then sequence large parts of it. Mapping one of the 24 distinct human chromosomes means producing a linear series of DNA fragments, containing genes, that extend collectively from one end of that chromosome to the other. This would then enable the location, isolation and characterization of all of the individual genes and functional sites in that chromosome. The over-arching goal of the HGP is to create a knowledge base of unprecedented detail and complexity as a resource for scientific investigations that will enable subsequent research to be immensely more effective and efficient. The process of mapping chromosomes is well underway and, it is fair to say, is making spectacular progress. Never before has so much genetic information been gained, and so many genes located, identified or characterized so rapidly. Never before has the technical means to gain information been more promising. Never before has so much biological knowledge been generated as during the past few years of genome research. No matter how strongly I emphasize it, you will undoubtedly be surprised in the next few years by the sheer rate at which new information is produced. Let me cite a few recent examples to illustrate this point. Hardly a week goes by now that a discovery in human genetics, relevant to some heritable disease, does not appear in the popular press. Recently, genes involved in muscular

dystrophy, fragile-X syndrome, Huntington's disease, "Lou Gehrig's disease" and many others have been discovered because of advances in human genetic knowledge. These discoveries are the harbingers of much more new knowledge that will be gained because of the HGP. The resource of knowledge of genomic maps, for example, now allows disease-related genes to be found much more efficiently. This prospect of new knowledge offers great promise and opportunity to many different branches and disciplines of the biological and medical sciences.

The original and explicit intent for the Human Genome Project (as arrived at by the DOE Office of Energy Research and the NIH National Center for Human Genome Research, and defined in the 1988 DOE-NIH Memorandum of Understanding and the interagency five-year plan which commenced on October 1, 1990), is clear. Broadly speaking, the goals are twofold and include the mapping and sequencing of the entire human genome, along with the development of advanced technologies and instrumentation to achieve these ends. They also include the development of the informatics capabilities to manage, access, and analyze the resulting data. It is the aim of the U.S. Human Genome Project to accelerate future biological science by building the tools, material resources, and infrastructure so that other branches of science, to be determined by the aggressiveness and imagination of their practitioners, can make the greatest use of these new methods and tools. In other words, the HGP is directed at building a scientific resource of unprecedented complexity and power. Many scientific research enterprises will make use of this resource: medical genetics, for the diagnosis and treatment of disease; and fundamental biology,

for the detailed understanding of biological function, to give two examples. Surely, population genetics and anthropology will also be among the early exploiters of the new genetics made possible by the HGP.

Human Genetic Diversity

Let me turn now to the specific emphasis of the present hearing--the question of the nature of human genetic diversity and the research intended to elucidate it. The fundamental scientific questions concerning the nature and full extent of human genetic diversity are very important. The origins and histories of human populations, and the extent to which the genome can and does vary among the present human population, are momentous and fascinating scientific issues. The area of human molecular population genetics has recently attracted much attention, and the opportunities facilitated by new methods, technologies and approaches made possible by advances in molecular biology and genetics are indeed exciting prospects. It is now possible to distinguish the relatively distant genetic origins of individuals, in addition to being able to distinguish individuals from one other. The former is based on differences in the human genome that persist throughout large groups of people with common heritage. The latter is sometimes referred to as DNA forensics, and is based on relatively recent genetic changes that result in differences among individuals. It is important to recognize, however, the natural boundary between the building of a powerful scientific resource, which is the HGP, and a broad field of scientific investigation that encompasses physical anthropology, human evolution and population genetics.

For several reasons, it would be a serious mistake to ignore this distinction and attempt to bring this sort of research initiative under the umbrella of the HGP. The importance and breadth of the questions of human genetic diversity are such that this research should be constituted as an effort that is clearly distinct from the HGP, both in organization and funding. The most direct reason is that the goals of the research are clearly distinct and, in some ways, derivative of the genome project. The HGP can provide an invaluable resource for planning, designing and launching a research program focused on human diversity, but such a program would clearly not fit the central goals of the HGP. To divert any part of the HGP to address these issues now would be a great disservice to other areas of biology and medicine that are likewise dependent on the acquisition of the knowledge resources being derived from the genome project, and which could make similar claims to scarce sources of support. In addition, it must be noted that the more focused and timely the acquisition of the HGP data is, the more effective a human genetic diversity project will be. Thus, any diversion from the goals of the HGP would, in the long view, detract from diversity studies as well as others.

There are, then, substantive reasons for keeping diversity studies distinct from the HGP. Let me now turn for a moment to some considerations on the planning for such a project. First, I will discuss an issue that is distinct from the scientific questions. Before starting an initiative like this one, serious thought must be given to the impact on the population to be studied, especially in the area of ethical, legal, and social implications of the

research itself and the information it could produce. These are serious issues. Privacy, discrimination and legal and social issues in the research locales are significant considerations in planning such a project, and must not be slighted. Of course, since all federal agencies have subscribed to a "Federal Policy for the Protection of Human Subjects" (10 CFR Part 745)--and part of this rule is that all human subjects, anywhere in the world, must be treated as they would be in this country--all subjects and all information deriving from this research would have to be treated with the great care specified in those rules. In addition, Third World countries may not feel that providing unlimited access to their "genetic resources" without assurances of various kinds, perhaps including compensation, is in their best interests. Also, how other research workers, unconnected with the immediate research but having access to the data, might make use of biological materials and information created by this project is a serious issue and must be considered. These difficult issues have not received much attention to date. It is clear to me that it would be irresponsible not to have these issues clearly addressed before such research ensues.

I have argued that the human diversity research should not be considered to be a part of the HGP. I have also pointed out that there are several societal obstacles that must be overcome to carry out such research in an effective and responsible manner. I would like to mention now that there are significant scientific issues that need to be considered further before this research project should be considered for significant support.

While the differences between two human genomes can be significant in what they could tell us about their respective genetic origins, the differences between two human genomes appear to be, at any one position, of the order of one part in one thousand. There are regions that are highly variable, changing rapidly so that only close relatives share the same sequence--the basis for DNA forensics. There are also much more slowly varying regions that are shared by most individuals with relatively distant genetic heritage. What this means is that very careful considerations must go into devising a plan for sampling a human population before it is likely to be useful. The genetic differences being looked for must be anticipated to some extent in order to plan the sampling properly. If one is searching for characteristic genetic differences that can mark historical and recent pre-historical populations, one must be very careful to understand the genetic markers one is studying, and tailor the sampling scheme to the genetics. It has also been suggested that very important information about medical genetics is to be gained by sampling diverse human populations. While this is certainly true in principle, it is also clear that this kind of information is unlikely to derive from a program that takes a few samples from a widely diverse set of groups and individuals. The point of this discussion is simply that a great deal remains to be done to arrive at a robust plan for such a program if it is likely to be useful to science and to humanity.

Because of our interest in fostering the productive use of genome research in related areas, the DOE, as well as the NIH and the National Science Foundation (NSF), have provided some support to interested scientists to hold several

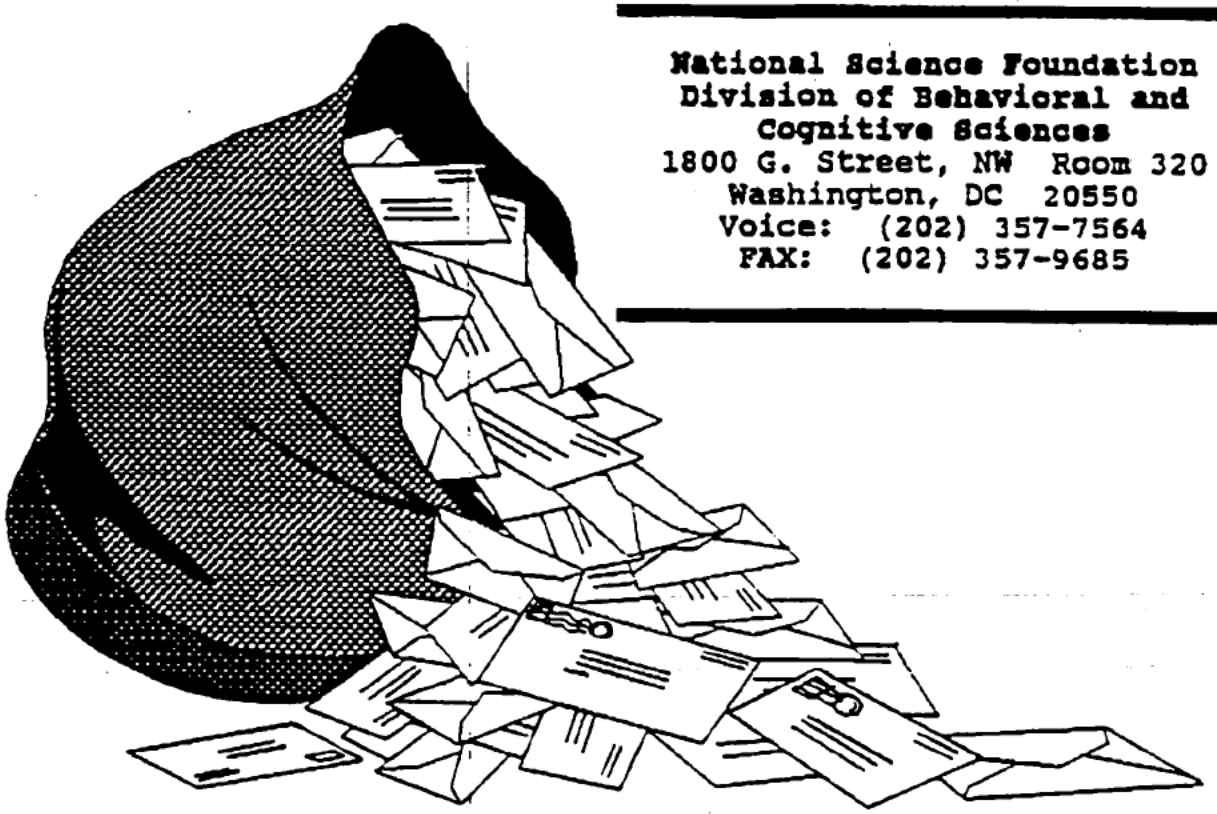
planning meetings over the course of the past year concerning the study of human genetic diversity. It is reasonably clear to me that this planning process has not yet succeeded in devising an incisive set of goals and objectives, a convincing rationale, and a clear set of immediately achievable goals. This can clearly be achieved, but at the moment the plans remain immature. The important point, however, is that when the scientific questions and the ethical and social issues are properly dealt with in planning this research, the project should be considered only on its own merits, and certainly as distinct from the HGP. If this planning is done well, the project may well contribute substantially to understanding our biological heritage and our history--invaluable contributions to human knowledge. This represents another area that the HGP can contribute to by providing the fundamental resource for human diversity research.

Conclusion

The Human Genome Project, the historic, coordinated, international effort to map and sequence the genetic material of our species, is well underway and is making great progress. There are many reasons that this enterprise was undertaken; its ultimate utility should extend to every branch of biological science, medicine, and biotechnology. Profound questions about human biology, human origins, human development, and human disease, will become answerable. To accomplish this in a reasonable time period and with the anticipated efficiency, which results from adherence to a careful plan, we must focus our attention and our resources on the goals at hand. The beneficiaries of the

HGP will be every branch of biology and medicine. I support the broad aims of those interested in human genetic diversity studies and encourage them to make use of the powerful tools, resources, and vast knowledge deriving from the Human Genome Project. However, these studies should compete for funding on their own merits and should not be a part of the Human Genome Project.

This concludes my prepared testimony. I would be happy to answer your questions.



National Science Foundation
Division of Behavioral and Cognitive Sciences
1800 G. Street, NW Room 320
Washington, DC 20550
Voice: (202) 357-7564
FAX: (202) 357-9685

FAX COVER SHEET

Transmitted to:

NAME: E. JORDAN

LOCATION: BETHESDA MD

[redacted]

Transmitted by:

NAME: Joseph B. Friedlander

PHONE: [redacted]

DATE: 4/7/93 TIME: 5:30

We are transmitting 4 pages of material, excluding this cover sheet. If you do not receive all 5 pages, please call as soon as possible.

Statement on the Human Genome Diversity Project

Cora Marrett, Assistant Director
Social, Behavioral, and Economic Sciences
National Science Foundation

* Relation to Human Genome Project.

Recent remarkable advances in molecular genetics have enabled the launching of the Human Genome Project, whose mission was the decoding of the total package of genetic material, or DNA, in a human. This massive effort is well underway and is almost five years old. It is funded by the National Institutes of Health and the Department of Energy, and currently has an annual budget of approximately $180 million. It is still many years away from completion.

Once the Genome Project produces an entire DNA sequence, we will still not have extensive information on the hereditary diversity that exists among people - the very large differences that exist among people within ethnic groups or populations, and the smaller differences that characterize differences among ethnic groups or populations.

The Human Genome Diversity Proposal is a developing effort of a group of physical and cultural anthropologists, archaeologists, geneticists, epidemiologists, and linguists to collect, organize, and analyze sufficient DNA from different people and populations around the world to answer two major issues.

*Objectives of the Human Genome Diversity Project

The first issue is understanding recent human evolution. By laying out the specific DNA differences among individuals world-wide, it will be possible to reconstruct the relationships and origins of the different human populations - the ancient origins of different Native American groups, for example, of Polynesians, of different European and African groups, and so forth.

The second issue is the hereditary basis for differences in human susceptibility to diseases. We are learning weekly the genetic determination of a many comparatively rare diseases - Huntington's Chorea, cystic fibrosis, and so on. The list is now thousands long. For a very few of these, we know the exact location and DNA sequence of the gene involved.

In addition, we know that the more common diseases, afflicting millions of people, often involve both genetic predispositions and environmental effects - well established cases include breast cancer, diabetes, bowel cancer, and at least certain cardiovascular conditions. These problems, and others as well, vary in their risk from one family to the next and one ethnic group to the next. Heredity clearly plays a role. Learning more about the gene or genes that place people at higher risk for these will inevitably help in their treatment and prevention.

*Current Status of the Human Genome Diversity Proposal

I want to stress that The Human Genome Diversity Project is, at this point, still only a proposal. The Human Genome Organization (HUGO, and international scientific organization), established a committee in 1991 to develop a Human Genome Diversity Project. As part of that effort, the committee members in the United States, including Dr. Cavalli-Sforza and Dr. King, submitted a proposal to the Physical Anthropology Program at NSF to support a series of workshops that would develop a plan for amassing genetic, linguistic, medical, an environmental information on a diverse group of human populations from around the world. As the lead agency and program, NSF Physical Anthropology coordinated modest levels of funding for these workshops with selected other NSF programs, with the Center for Human Genome Research, the National Institute of General Medical Sciences, and the Department of Energy.

Three of four planned workshops have now been held. The first focussed on what an adequate sample of people would be to characterize a population. The second focussed on what particular populations were the best candidates for inclusion in the project. Over 400 populations were identified. The third focussed on ethical issues, biotechnical developments, and funding possibilities. The fourth will involve international possibilities for collaboration and funding.

Since the project clearly must be an international one, there has been an attempt from the outset to enlist cooperation and parallel funding from other countries. European members of the HUGO committee have worked to establish funding to support the HGD project, and are working to allow funding for European components of the HGD project. They have already secured some funding to begin work. Technical and scientific representatives of the E.C. and Japan have also been invited to the workshops, and individual scientists from a number of countries around the world have participated in the workshops. Certain private agencies likely to be interested in the project have also been approached by the organizers.

The structure of the project that is evolving will have, at its center, an International HGD Scientific Committee. A second committee of importance will be the HGD Ethics Committee, to monitor, assess and anticipate the complex ethical issues that will inevitably develop.

The project itself can be thought of as being divided into three parts.

The first part will be the collection phase. This involves the identification of appropriate populations or groups to be covered in the survey; the collection of required family, medical, demographic, cultural and ecological information on each group if it has not already been made; and the taking of blood samples from cooperating individuals. At this point, at least 400 groups world-wide are likely to be included.

The second part involves the transformation and storage of the blood samples themselves in central laboratories. Along with the other data collected on the populations, stored in computer files, these will make a permanent and inexhaustible source for future scientific and health research.

The third part is the analysis phase, which will extend for some time into the future.

As currently envisioned, the Human Genome Diversity Project can have the first two phases complete and have made a good start on the analysis phase within 5 years. The estimated cost will be a tiny fraction of the Human Genome Project, for reasons to be

discussed, and the cost to the U.S. Federal Agencies involved will be only a portion of those costs.

*Summary

This is an idea whose time has come. The great hereditary diversity of humanity is rapidly being eroded or obscured by the same forces that are transforming so much of our world today - increased ease of movement and migration, the penetration of heretofore isolated ecologies and environments by the world economy and its representatives, and changing diets and life styles accompanying these alterations. We need a systematic collection of information, genetic and otherwise, on the diversity of human populations before they become even more diluted by the growing global mobility of peoples. This will be an invaluable resource for understanding our origins and our hereditary accomodations and susceptibilities to the diseases that afflict us.

STANFORD UNIVERSITY SCHOOL OF MEDICINE
STANFORD, CALIFORNIA 94305-5120

FACSIMILE TRANSMISSION

DEPARTMENT OF GENETICS
(415) 723-5052 (Direct: 723-5804)

TO: ELKE JORDAN
GENOME OFC

DATE: [redacted]

FROM: LUIGI LUCA CAVALLI-SFORZA, M.D.
PROFESSOR OF GENETICS
FAX #: 1 415 725-1534

OF PAGES: 10, INCLUDING THIS PAGE.

MESSAGE:

Draft 4/14/93 Bibliography missing. Comments welcome LCS

Testimony before U.S.Senate regarding the Human Genome Diversity project, as requested by senator Daniel Akaka of Hawaii.

To be given on April 26, 1993,
By L. Luca Cavalli-Sforza, professor of Genetics (emer[redacted]tive)- Stanford University (Stanford, CA).

1.Generalities on aim of HGD.

The Human Genome Diversity (HGD) Project aims at surveying the genetic diversity of presently living humans. It proposes to do so by sampling primarily indigenous populations, whose origin is usually better known than that of individuals who emigrated in the past centuries to new places. These samples should come from an adequate number of individuals representing the human species, and used to form a repository ensuring unlimited survival of their genetic material, DNA, for future study. Special, but not exclusive attention will be paid to people of historical interest likely to disappear due to physical extinction or, more frequently, dispersal and loss of identity. These events take place at an alarming rate, paralleling the rate of economic development which causes them. The program is therefore urgent in order to avoid the irreversible loss of precious genetic information.

2.Preparation of cell lines

The HGD project makes use of the existence of modern techniques of conservation, which allow us to keep certain cells of an organism alive and capable of reproducing, thus generating potentially unlimited amounts of their DNA. No or very little DNA deterioration is experienced in these conditions. For this purpose, fresh samples of blood must be collected and transported shortly after collection to appropriately equipped laboratories, where certain blood cells (B lymphocytes) are transformed by Epstein-Barr virus, grown and stored under liquid nitrogen where they conserve indefinitely their viability and capacity to reproduce.

3. Choice of populations.

Accurate choice of populations and individuals is an essential part of this project, and for this aim we want to ensure the cooperation of anthropologists. A preliminary list of about 4-600 populations to be sampled has been established with their help. On the basis of discussions with population geneticists and statisticians, we plan to generate transformed cell lines from about 25 unrelated individuals per population. When anthropologists obtain the blood samples for preparation of transformed cell lines, we expect them to collect also at least 10 times more blood , saliva and/or hair samples from the same and related, neighboring populations. Such samples will not be subjected to transformation, but will be stored for studies requiring larger numbers of individuals for special purposes, for which smaller, limited amounts of DNA are sufficient, or less sophisticated, cheaper methods of DNA amplification can be employed.

4.Extent of diversity at DNA level.

Why study individual diversity, one may ask? It is necessary to study it because it is there and is far from trivial. It would be erroneous to think that the Human Genome (HG) exists in a single copy, repeated again and again in every individual. The opposite is true; one can say there are as many genomes as there are people , and the potential variety of human genomes is of the order of 1 followed by 1,000,000 zeros (There are 3 billion nucleotides in a single, haploid genome, like the DNA passed from a parent to a child, and assuming one every 1000 nucleotides is polymorphic, i.e. exists in two different forms, the number of potentially different genotypes is ~~abstract~~ 2 to power of 3 about millions). There are very unlikely to exist living human beings having exactly the same DNA. Even identical twins must differ from each other, because after duplication of the fertilized egg giving origin to both of them a dozen or more fresh mutations are expected to occur.

Prompted by the enormous magnitude of the task of analysing the whole human genome, the HG project was started in its simplest form, as the plan of studying a single copy of the genome (not necessarily all coming from the same individual). Some of the desirable problems which it was suggested would benefit from the effort can be thus answered . We are ourselves involved, in our lab, in research on two genetic diseases, and we are well aware of how useful it could be, and how much time we would spare if the HG project were already completed. Knowing the DNA sequence of the regions where we know or suppose the genes responsible for our diseases are located would greatly simplify our task.

It is true, however, that a number of important benefits could be gained from reaching a better understanding of individual variation. An important step forward could be made in the next five years with an expenditure that could be only 3% of the total cost of the Human Genome. The time is now ripe for starting as soon as possible an adequate analysis of diversity, and the optimal procedure is that of carrying it out together with the sampling of disappearing indigenous world populations.

5. Importance of studying diversity for genome analysis.

The general importance of studying individual DNA variation will be better appreciated considering that the real purpose of studying the sequence, i.e. the structure of the human genome, is that of inferring the function of its parts. This constitutes a considerable challenge. We cannot today predict the exact function of a DNA segment by simply determining its sequence; we can only make some informed guesses. One obviously hopes, however, to generate rules that will ease this task, and some have already been proposed. Not surprisingly, the study of variation among individuals of a DNA segment can contribute in an important way to understanding its function. In general, genes or their parts which are basic to the life of the cell or of the individual vary very little from one individual to the other and only in specific

ways, e.g. for nucleotides which affect only trivially the structure of the protein they synthesize. Lack of variation, or minimal presence of it indicates an "important" gene; changes in it are likely to determine serious abnormalities. The identification of a gene responsible for a disease usually requires detection of a clear association between changes in its structure and specific pathological phenomena. More generally, the function of a gene is determined by studying the variation in different persons of the specific function assumed to be influenced by the gene. Although some of this work can be carried out in vitro, the final test is the association of the external form or behavior (the "phenotype") of the individual and the DNA structure of the gene in question.

6. Some general guidelines for the study of diversity.

Five years after the beginning of this endeavour, the task of sequencing a single genome is still a formidable one, and it is sofar premature to think that one could, or would want to, sequence in full even only two genomes, which might seem the minimum necessary for learning about individual diversity. But techniques may make substantial progress in a few years, and what appears today an impossibile task may well be far less difficult a few years from now. It is also certain that the technologies being developed for study of the Human Genome are ideally suited for the analysis of individual diversity, and that there will be considerable reciprocal fertilization at the technological level between the developments of the mainstream Human Genome research, that of Human Genome Diversity, and the applications to medical genetics, which are a special aspect of the studies of indvidual diversity. At least initially, however, it is neither feasible nor necessary to study individual variation for the whole genome. Rather, we need to set up methods and learn rules that allow us to predict the individual variation to be expected in a given segment with minimal effort. This can be done by studying individual variation in a sample of DNA segments which need not be large. There is a parallel between the inferential effort necessary for predicting the function of a gene, or more generally, a DNA segment on the basis of its structure alone, i.e. its sequence, and that of predicting on the same basis its variation. They are both challenging tasks, and there are also close direct relationships between the two types of inference. I mentioned the rough guiding principle that the more important the function of a DNA segment, the less variation is seen in that segment among individuals. But this principle is not general enough. There exist DNA regions like for instance HLA, the genes we study because of their importance for the success of organ transplantation, and for the predisposition to certain autoimmune diseases, for which individual diversity is a biological necessity. Here the extreme individual variation of HLA is a prerequisite for its function.

The analysis of DNA samples from the world population for the purpose of studying individual diversity can be extended by present methods only to a small fraction of the whole human genome, given its cost. But the study should be planned so as to

make it as informative as possible. Let us assume that we want to analyse 10,000 DNA segments of length around 1000 nucleotides each in different parts of the genome. They should be chosen so as to represent various categories of genes and DNA segments, e.g. various types of repetitive DNA, coding and non coding DNA, and various other identifiable DNA categories. Although ideally, they should be studied on all the 10,000 or so individuals from which cell lines should be prepared, at the beginning analysis should be restricted to a smaller fraction of individuals (e.g. a minisample of 100 to 1000 individuals). The minisample could be chosen in such a way as to represent the world through a hierarchical stratification of the 10,000 individuals. It would thus be easy to evaluate the usefulness of extending the survey of individual variation to all the 10,000 cell lines, or may be the 100,000 or more of the enlarged collection of non tranformed samples. Only those DNA segments that show sufficient variation would be eligible for being studied on the larger samples. The majority of them would not have to be sequenced, but just tested with cheaper methods allowing the recognition of individual polymorphic (i.e. variable) sites. The collection of indigenous populations could be organized so as to have already after the first year a "minisample" of individuals and populations adequate for coverage of the whole world, which could be employed for the initial screening of variation. A preliminary collection of cell lines of a few hundred individuals from all over the world is already in existence and being used in preliminary observations, thanks to a pilot project set up in the last few years by the writer in collaboration with Ken and Judy Kidd of the Genetics Dept. of Yale University. Samples of this collection are available to all research workers through the Camden Repository of New Jersey.

Even limiting the analysis of individual variation to a small fraction of the genome a general picture can be obtained which will be useful to direct further investigations. We currently know very little of individual variation at the DNA level, but we are aware of the existence of segments which are much more variable than others. Their variation seems to have a precise biological meaning. Our knowledge of this aspect is still scarce, however. For instance we know only few examples of localized high variation, like that for genes that are important in transplantation (HLA), and mitochondrial DNA, but their number is bound to increase with a systematic search. It will be important from a biological point of view to establish if the high variation is due to high mutation, natural selection or other mechanisms. The search for highly variable DNA segments has, in my view, the highest priority because they are the most informative ones, but it will take time to detect them. A first phase of extensive preliminary investigations will therefore be necessary to choose appropriate, representative DNA segments to be studied. Neither in this initial phase nor in later ones will it be necessary to use full sequencing methods on all individuals for each DNA segment; many times cheaper and faster methods would be preferable at least for initial screening, especially in regions in which variation is usually low.

It is my belief that research workers engaged in genome research will in due time find an inexhaustible source of important new biological problems in the study of individual variation. HGD will take great advantage from, and also contribute in an important way to, the current trend towards automation of DNA sequencing and in general, testing genetic variation of medical importance. Our rapidly increasing arsenal of research robots can speed up considerably the testing of large numbers of individuals for specific DNA segments, but the first need is that of learning more about the extent, location and causes of individual variation, and optimal strategies of studying diversity.

7. Need of studying variation at a worldwide level.

Individual variation must be studied on a world sample, and not on a limited sample of individuals collected locally as has been sofar the case, because there exists intergroup as well as intragroup variation. We have recently demonstrated that limiting the study of variation to Caucasoids, as done in almost all the investigations of DNA variation available, has introduced a serious bias in the evolutionary interpretation of the data. Moreover, the full power of study of individual variation would not be reaped, if it were limited to Caucasoids. There is considerable ethnic variation in genetic disease and predisposition to disease, which is important in planning health surveys, dietary advice, search for donors for transplants, etc. It is necessary to explore the whole human species with a well designed sampling scheme. At the same time, it is necessary to avoid the dispersion of effort which has been characteristic of most work carried out in the pre DNA era, in which different research workers studied haphazard collections of markers on haphazard collections of populations. The result is that from published articles one can obtain only an extremely sparse matrix of genes x populations data, and the information that can be obtained is a very small fraction of what it could be if the immense effort of the thousands of research workers who volunteered their effort had proceeded in a more organized manner. A collection of cell lines made available to research workers in a central repository can help avoiding this waste. It will be necessary, however, to ensure that also some of the testing effort be made in a systematic way. Making a representative set of cell lines chosen so as to be a representative world sample available at especially low cost could be a simple way to ensure a more balanced and informative effort.

I will concentrate here on giving a short account of scientific benefits that this endeavour can generate for the study of history of human differentiation and evolution. We have already accumulated substantial knowledge on genetic variation of human populations by methods prior to the DNA era. But this information suffers from many uncertainties and from having been almost entirely unplanned and being therefore full of gaps which make its analysis difficult and inefficient. Even so, they have shown that the study of living populations allows to reconstruct important aspects of ancient human history which have helped to

explain archeological and linguistic information. Striking conclusions have thus been reached which have received wide interest, but need confirmation and extension. Modern methods of study of individual diversity using DNA have much more power, can examine a much greater range of genetic variation (potentially all), are much more efficient. They can be more easily automated than the older techniques, and considerable efforts in this direction are already en route. Enormous increases in precision and in the variety of historical problems to which one can have access are made possible by increasing the number of populations and the number of genes (DNA segments) studied sofar. Moreover, the analysis of DNA allows to obtain "fossil" information from mommies, bones etc. which has sofar been lacking. For the full use of fossil information, however, one needs to have adequate data on living populations from the same and neighboring areas, with which it can be compared.

Obtaining a representative sample of the world, even if one concentrates on indigenous peoples may be rewarding also in view of another application of DNA: the "fingerprinting" of individuals for forensic purposes. It has become clear that courts need better information on frequencies of genes of different populations in order to evaluate more correctly the probabilities of paternity or individual identity, etc. It may be objected that these frequencies are more frequently needed for urban populations, which could hardly be considered as indigenous and would not be directly part of our survey. It is possible, however, to extend knowledge of frequencies of indigenous peoples to those of special strata of urban populations, when something is known of the ethnic origins of the latter. It has been shown for instance that one can predict rather accurately the genetic background of white Americans given information on their surname and that of their parents or grandparents. This may be misleading only in a few cases of recent change of surname, or for Blacks who have standard English names, but these exceptions can usually be identified and be given special analytical treatment if necessary.

8. Genetic diversity and racism.

An important question is: can the study of intergroup genetic difference foster racism? My persuasion is that it will have an opposite effect. It is difficult to make full justice to this important subject in a short time, but more extensive analyses can be found elsewhere [xx]. It might be useful to begin by a definition of racism, which in my view is the persuasion that difference in success of different peoples or races is due to their inherent biological superiority. The idea is that this superiority is genetically determined and as such inherited and impossible to change (Notoriously, the superior race is usually one's own). The importance of socio-cultural inheritance in determining success (economic, political or military) is usually forgotten. Yet simple historical facts like the short duration of empires, and of the associated superiorities, suggest immediately that the background to success is socioeconomic and cultural; it cannot be genetic, because genetic processes are much slower, and

very little genetic change can take place in the very few centuries during which empires rise and fall. This is a very simple but powerful consideration that is not, however, heard frequently, if ever, simply because knowledge of genetic processes is confined to a few specialists.

Another consideration about racism is that genetic differences between human groups are small. We have already considerable information from a study of genetic markers with methods of the pre-DNA era. We have still limited information on intergroup diversity using data obtained directly on DNA, given that the relevant techniques have become available only in the last decade, but we already know enough to be confident that the picture thus obtained is very similar, and much more clear. The basic conclusion from the study of differences among groups is that they are small compared with those within the groups themselves. The aspiration to "race purity" of classical racism is absurd. A village, a small tribe show almost the same genetic variation among individuals as the whole world (Nei, Lewontin). Only human populations of very small islands which have been subjected for a long time to very close inbreeding (marriage among close relatives) show a moderate increase of genetic homogeneity. This is often accompanied by infertility, in agreement with the conclusions of countless animal (and to some extent plant) breeding experiments.

The statement that genetic differences among races are very small may seem to fly in the face of evidence known to everybody. If we look at the skin, eye, hair color, at the facial and body morphology of people who originated in different continents we can usually predict with accuracy their ethnic origin, simply relying on these superficial traits. Most such characters are homogeneous within the groups, and show sharp differences between the groups. This is just the opposite of what we conclude from a random, large sample of genes studied with pre-DNA or DNA methods, where almost without exception differences between people from the various continents are only quantitative, and of small magnitude with respect to that within groups, even within small towns or villages. For genes existing in various forms one finds usually the same forms may exist in almost all parts of the world, but only in different percentages, and the differences are rarely striking. Clearly, pigmentation and morphology of the human face and body have different behavior from random genes, and show a much greater racial differentiation. It seems important to discuss the reason of this difference, which is clear even if it is not widely known [Bodmer and LCS, 1976] and can explain this apparent contradiction. Everything we see and use in our diagnosis of ethnic origins of individuals in every day is a property of the surface of the body. In the last 50,000 years or so humans, after developing in the tropical, warmer parts of the world, have spread to the rest of it, and learnt to survive in extremely cold environments as, e.g., Siberia. This has involved a cultural as well as a biological adaptation. It is clear that all the characters which show a strong difference between races, as well as greater homogeneity within races, are connected with adaptation to climate, and have been so explained by physical

anthropologists (Coon). It is inevitable that they involve the surface and general shape of the body, because they form the interface between the external world (from the physical rigors of which we must protect ourselves), and the internal one, the temperature of which we try quite successfully to keep constant.

It is not surprising, therefore, that the surfaces of our bodies differ greatly depending on the climates in which we developed in the last tens of thousands of years. As the surface is the part of us which we see, its conspicuousness affects our judgment on questions of race, and prompts us to believe that all other differences are equally striking. But the truth is that they are not. We may add that they could hardly be, given that the evolutionary separation among humans living in the various continents has been relatively short and there has not been time to develop much divergence. Knowledge about the real genetic differences which exist can only help to defuse the race bomb.

9. The plight of indigenous populations.

Many indigenous populations are in conditions of extreme poverty, and are often abused and victimized in the process of economic development. Their plight is completely obscure, and it is very difficult to help them. Bringing it to public knowledge is the first need, and these investigations can do much towards this aim, given the interest shown, for instance, by many journalists and mass communication specialists. It would of course be impossibile to reach all people in need in the course of this program; there exist of the order of 5000 different populations in the world, based on the count of different languages in existence, and the program can reach about 10% of them. A large fraction of these is found in developing countries and will be in need. The program can bring light on the conditions of these people by making them known to a large public, and this is very likely to have positive effects. It can also generate greater public knowledge and interest in cultural diversity, and the desirability of maintaining it in suitable conditions. The United Nations are aware of all these needs and have called 1993 the year of the indigenous people. Given the terrible urgent problems which face the world today, the limited resources of the United Nations are fully committed in many parts of the globe, and it would be unrealistic to hope that they could do much to help indigenous populations. One way in which our program could give some support is in bringing relevant facts to the attention of the public of developed countries, and in special cases also to the attention of the public and the authorities of their own country.

10.HGD and identity.

We have found our project has generated enormous spontaneous interest among journalists and laymen. The program has the potential of discovering remote origins of people, if not or not always for individuals at least for groups, and this seems to strike important, to some extent unexpected positive emotional reactions. Most people know too little about their remote past, because its memory was lost after their ancestors migrated to

America, due to the ignorance and anger that often accompanies poverty or to the conditions under which they were forced to abandon their original abode. But knowledge of personal identity may be a basic interest of humans and an important component of self-esteem, and information of the kind which the program can provide may be cherished by many. It is well known that several cultures dedicate special effort to conserving their genealogies.

11.Spreading scientific knowledge.

A further consideration is that the HGD project can contribute to spread interest in the public for genetics and create a desire to learn more about our science. Greater scientific literacy is one of the important goals that should be promoted by a renewed effort in education. Recent international comparisons [Sci.Am.] show this is a major and urgent need especially in the United States. Helping to generate further knowledge of the kind promoted by HGD, the program can generate curiosity of many individuals, and become a powerful stimulus creating additional interest for a specific science, which is often taught in High school and should hopefully become even more popular.

Program Project

DIVERSITY FOR HUMAN DNA SEQUENCES

Director: L.L.Cavalli-Sforza,
Genetics Dept.
Stanford University.

DIVERSITY IN HUMAN DNA SEQUENCES

Principal Investigator: Cavalli-Sforza, L.L.

Subproject Core:	Cavalli-Sforza, L.L.
Subproject 1	New DNA Regions in Humans showing High Variation. Cavalli-Sforza, L.L.
Subproject 2	Analysis of Human Genetic Diversity. Slatkin, M.
Subproject 3	DNA Sequences for Studies of Human Genome Diversity. King, M-C
Subproject 4	Theory and Analysis of Molecular Variation. Feldman, M.
Subproject 5	HLA Class 1 Sequence Diversity in Indigenous People. Parham, P.

INTRODUCTION

Program Project on Human Genome Diversity for DNA sequences

Director: L.L.Cavalli-Sforza

ABSTRACT

Recognition of the need to study Human Genome Diversity has generated considerable interest in a program to obtain a sample of the world's indigenous populations. While this sample will be collected , it is urgent to decide which genes should be studied, a non trivial problem which will take some years. A major method of studying genetic diversity is that of analysing nucleotide sequences of DNA segments. Sequencing can be more efficient than an alternative approach, the study of point polymorphisms, provided the DNA segments studied show enough variation. Some of the most exciting results in human evolution (and also disease association in the case of HLA) have been obtained by sequencing studies of the control region of mitochondrial DNA and of HLA genes. There are no other regions known today which could be equally useful, and we are interested in finding more and studying them on a pilot sample of the world population. A minicollection of indigenous populations already in existence can be used for the purpose. Results thus obtained can be extremely useful for planning the analysis of larger samples to be collected in the future.

Regions with high variation must have a relatively high mutation rate and/or selection for diversity. There is a need for theoretical studies of these evolutionary processes which can be greatly stimulated if new, richer data become available. Recent theoretical developments already in existence which answer some of the analytical needs , as e.g. coalescent theory have hardly been used in humans. The feedback between theories and observations can be very profitable not only in physics but also in biology, especially if there is continuous contact between the interested scientists. The present program project aims at establishing a close collaboration among theoreticians and experimenters interested in the problems of evolution which can gain from the analysis of variation of DNA sequences at the individual level.

BRIEF HISTORY OF THIS PROGRAM PROJECT

The present program project has been operational for 12 years, beginning in 1980 with methods of searching for new polymorphisms using techniques of DNA analysis (to become later known as RFLPs) or monoclonal antibodies. In the last five years it was dedicated to a continuation of the study of DNA polymorphisms and associations with certain diseases (Wilson's disease, LCS; YACS and hybridization methods for studying DNA, Ron Davis; autoimmune diseases: multiple sclerosis, L.Steinman; lupus erithematosus, MC King), and for a minor part to certain evolutionary problems (1988-1991, immunoglobulin K, and since 1991 mitochondrial DNA sequences , LCS). In the spring 1992 we were able to obtain an ABI 373A automatic sequencer which has considerably speeded up our sequencing work.

Since 1991. three of the applicants of this project (MF, MCK, LCS) have been associated with two other research workers from the U.S. and four from Europe as members of the committee for the study of Human Genome Diversity (chaired by LCS) , which is laying the foundations for an international effort to start a systematic collection and storage of DNA from indigenous human populations for evolutionary analysis. The first scientific meeting of the Human Genome Diversity project, dedicated to statistical and population-genetic problems of the sampling strategy was organised by M.Feldman c/o the Morrison Institute at Stanford.

The new version of the program project takes its inspiration from this new venture. Two of the activities of the latest version of the project are continued: the search for the gene responsible for Wilson's disease, and of associations for lupus erithematosus.

THE PRESENT PROJECT

The motivation of this new version of the program project has its roots in the recognition that the analysis of DNA sequences is the most thorough method for the study of the genome. Almost all of the evolutionary work done to date in humans did not use this approach, except for a few investigations of short DNA segments variable enough that sequencing is the best available method of analysis: mitochondrial DNA (mt-DNA) and Human leucocyte antigens (HLA). These are among the most exciting studies in human evolution; they should be deepened and extended to other variable regions. An additional consideration in favor of this approach is that DNA sequencing is already automated, unlike other conventional methods of studying genetic variation by analysis of individual sites, like RFLPs (restriction fragment length polymorphisms), and the other methods developed later

(PCRPs, SSCP, DGGE etc.). It is likely that given the impetus provided by the Human Genome Project, additional, substantial progress in automated sequencing will be made in a relatively short term. Only automation can generate the number of observations that are necessary for a satifactory study of human evolution.

High variation can be due to high mutation rates, as is likely to be the case for mt-DNA, or to natural selection for diversity, as is likely to be true of much, but perhaps not all, HLA variation. At the moment, no other segments with high variation have been investigated, but it is likely that other regions with high mutation rates exist. The empirical part of our study is aimed at locating appropriate DNA segments and start their evolutionary analysis. In the wake of the Human Genome Diversity Project, which should collect between the years 1994 and 1998 samples from a large number of human aboriginal populations, it seems very timely to search for DNA segments of special interest for evolutionary studies. The analysis of a reasonably large number of genes is an essential prerequisite for reaching reliable conclusions in human evolution.

Mutation rates observed in some regions of human DNA are even too high for these regions to be used in evolutionary studies (5%, Jeffreys et al. 1988). In unpublished collaborative work with Anne Bowcock, we have found that CA repeats may have mutation rates as high as 1/1000, and have proved to be very useful for certain evolutionary problems but perhaps not for all, because of their high mutation rate.

Mutation rates which are best for particular evolutionary investigations are likely to depend on the problem being considered. This is one of several questions requiring theoretical work. Another important theoretical aspect is that the analysis of evolutionary trees obtained with sequence data (or equivalent information with multiple polymorphisms centered on a short DNA segment, haplotype etc.) differs from the traditional one using gene frequencies. In the first case the unit studied is an individual (for a haploid constituent of the genome like mt-DNA or the Y chromosome), or a haplotype for the diploid part of the genome; in general, a single DNA segment. In the second case the unit of study is a population made of several individuals and the statistic employed is the relative frequency of an allele. Techniques of statistical analysis and the assumptions directing it differ substantially in the two cases. When comparing DNA segments, as often happens in comparisons between two species of each of which only one individual is analysed, or in the analysis of trees built on mt-DNA (mitochondrial DNA), differences are due entirely to mutation and selection. But when two or more populations are compared, also migration and drift come directly into the picture, and in intraspecific comparisons are likely to play the major role.

Unlike standard gene frequency data, where the population approach is the usual one, sequence information lends itself to being analysed in the two ways, i.e. using as a unit the sequence of an individual or a population of sequenced individuals. Nor-

mally in mt-DNA research the first approach only has been considered; but a comparison of the two approaches is useful or even necessary in certain circumstances. It may give special insight on the difference between the two, and help to pinpoint the roles of migration and drift.

The availability of information on individual variation of sequences, especially in Drosophila, has given rise to new theoretical developments. One of these, coalescent theory (Kingman , offers new approaches to a variety of old problems of population genetics, and deserves to be considered seriously and used in applications to human data. Balanced polymorphisms can hitchhike other unbalanced ones, as shown in Drosophila (see the "shifting windcw" method, McDonald and Kreitman 1991) . Similar studies in highly variable human regions could be useful. The distribution of timings of branchings in the tree of human mt-DNA can give information on the demographic history of a population (Slatkin and Hudson 1991, Rogers and Harpending 1992). This is usually studied indirectly, from the shape of the distribution of pairwise nucleotide differences between individuals. In our lab we have observed a striking confirmation of this fact (unpubl, see subproject 1). There are many other developments which come to mind when considering the application of old methods to new problems: e.g., the distribution of linkage disequilibria for studying the variation of mutation rates at the level of single nucleotides, which can be done particularly well in haploid sequences like mt-DNA or Y chromosomes, as well as that of recombination and gene conversion in diploid loci.

One can multiply the examples of theoretical problems, methods and results that arise from the availability of data on individual sequence variation in humans. For this field to develop vigorously it is essential to have close communication between theoreticians and experimenters. We believe it important, therefore, to concentrate on this problem both at the empirical and at the mathematical level, with continuous feedback between theory and observation. The structure of the new project is based on five subprojects: two strictly theoretical ones (Marc Feldman, Monty Slatkin) will concentrate on problems of mutation/selection/drift/migration in sequence analysis, by mathematical modelling, simulation and analysis of data, including observations made by the other subprojects; three experimental ones (Mary-Claire King, Peter Parham, Luca Cavalli-Sforza) will concentrate in the sequencing analysis of highly variable regions already known, or to be searched and detected. The core will assure the exchange of material and data, as well as the coordination of research among the experimental and the theoretical laboratories.

The five subprojects are assembled in this application in the following order:

CORE: L. Cavalli-Sforza (Stanford)

SUBPROJECT 1: L.Cavalli-Sforza (Stanford) :
New DNA regions in humans showing high variation.

SUBPROJECT 2: M. Slatkin (Berkeley):
Analysis of Human Genetic Diversity.

SUBPROJECT 3: M.C. King (Berkeley):
DNA sequences for studies of Human Genome Diversity

SUBPROJECT 4: M. Feldman (Stanford):
Theory and analysis of molecular variation

SUBPROJECT 5: P.Parham (Stanford):
HLA class I sequence diversity in indigenous people

SPECIFIC AIMS

1) Indigenous populations to be studied will include:
a) 15-20 world populations, such as those which exist as cultures of B lymphocytes transformed by EB virus (to which we will refer to as "immortalized lines") and available to LCS' lab. They are listed in table 1 of the Core.
b) The Human Genome Diversity Project may soon add many more, and a fraction of the new populations can perhaps be used to supplement and replace some of those already in culture, so as to have a small but excellent sample of the aboriginal world population. We do not intend, however, to study systematically all the populations to be collected and immortalized by the Human Genome Diversity Project, which may be of the order of 500, with a total of 10,000 individuals. We prefer to consider our program as a pilot study, preparing preliminary information for planning later systematic work on a larger number of individuals and populations. The later study may enjoy the possibility of using techniques faster and cheaper than those available now, but the sequences to be studied have to be discovered and tested in a preliminary way beforehand.
c) for the study of some well identified sequences as for mt-DNA and possibly also HLA, collections of small blood samples, rather than immortalised cell cultures, or even simply hair roots etc. are adequate and many are already available.
The number of individuals per population, and in some cases also the populations that will be analysed will vary depending on sequences analysed. In most cases, the same populations and individuals will be analysed for different sequences. Appropriate DNA samples will be made available by the core to the experimental laboratories, which will also effect routine sequencing of specific regions on request of experimental laboratories.

2) With current techniques, analysis of individual variation by sequencing can be efficient only in regions of high variation. Elsewhere analysis of single polymorphic sites is preferable. The only two known highly variable groups of sequences (mt-DNA and HLA) will be studied in depth on opportune population samples.

3) We plan to search for new regions of high variation, from an empirical point of view, using several different methods. Both sex (X and Y) chromosomes and autosomes will be included. There is still no precise idea of which sequences to use beyond the few already known; they have to be discovered, and easy methods to find them will be important. This work seems especially timely in view of the likely availability of a much greater collection of world's aboriginal population DNAs, hopefully beginning a year from now. There is at the moment a lack of genetic markers which could be usefully studied on these populations. Presently available ones, including point polymorphisms like standard RFLPs, PCRPs etc. have a variety of drawbacks and it would be very difficult if one had to indicate now a battery of markers to be studied on the new populations to be collected soon.

4) Considerable biological interest is associated with sequences showing high variation. Conditions known today which can determine high sequence variation are high mutation rates and selection for diversity, and the distinction between them is not too difficult. The direct study of mutation rates in the germ line for some of these sequences will be a part of the project (see subprojects by 3), in association with the core, and 5). If there is selection for diversity, one expects that certain exons or parts of exons which are critical for certain functions of the molecule and its properties will show higher variation than the rest of the exons, or the introns, or the corresponding parts of pseudogenes if any are available (see for instance enclosed MS by Kurth et al.).

4) We want to contribute to the theory of analysing, measuring and understanding individual variation for sequences. A few methods have been suggested for estimating nucleotide diversity (Nei 1987, Weir and Basten 1990) among individuals, and among populations. The estimation of mutation rates depends heavily on assumptions made about the relative mutation rates of the various types of transitions and transversions and on the assumptions made in tree building, but theoretical and numerical work is still limited and unsatisfactory in various ways. The controversies about the interpretation of mt-DNA trees are far from settled (Gibbons 1992). There are no similar analyses for other sequenced regions. The Y chromosome could offer advantages of lack or rarity of recombination as is true of mt-DNA. Similar frequently sequenced regions (HLA) have not been considered sofar from a theretical point of view. Here, complications may arise because of classical recombination and gene conversion, and because of the possibility of selection for heterozygotes and for diversity; but observations in Drosophila and premininary ones on human immunoglobulins indicate that the analysis of sequences may nevertheless be rewarding also in such complex situations.

DESCRIPTION: State the application's broad, long-term objectives and specific aims, making reference to the health relatedness of the project. Describe concisely the research design and methods for achieving these goals. Avoid summaries of past accomplishments and the use of the first person. This abstract is meant to serve as a succinct and accurate description of the proposed work when separated from the application. **DO NOT EXCEED THE SPACE PROVIDED.**

The goal of this project has been to develop and apply mtDNA sequencing to identification of individuals and familial relationships. The project is a "hybrid," involving both technology development and ethical, legal, and social implications of human genome research. In the technology domain, new methods are created to obtain and sequence DNA from "traumatized" materials (e.g. buried teeth and burned tissue), and polymorphisms are developed and characterized that enable identification of either maternal (mtDNA) or paternal (Y chromosome) lineages of unknown individuals, even if few relatives survive. In the ELSI domain, the project applies these methods to identifying victims of human rights violations, wars, and natural disasters.

The specific aims for the next three years of the project are (1) to identify and characterize short sequence-repeat polymorphisms specific to nonrepetitive portions of the Y chromosome, (2) to extend our understanding of mtDNA sequence variation in American populations, (3) to estimate the frequency of new mutations in the mtDNA by studying maternal relatives separated by multiple generations, and (4) to develop approaches for mtDNA sequencing for difficult identification cases involving degraded DNA. Throughout the project, mtDNA sequencing and Y chromosome genotyping will continue to be applied to identification of individuals and their families.

PERSONNEL ENGAGED ON PROJECT, INCLUDING CONSULTANTS/COLLABORATORS. *Use continuation pages as needed* to provide the required information in the format shown below on ***all*** individuals participating in the project.

Name [redacted] Degree(s) BA, PhD Social Security No. [redacted]

[redacted]

Organization [redacted]

Name [redacted] Degree(s) [redacted] Role on Project [redacted]

Position Title [redacted] Date of Birth (MM/DD/YY) [redacted] Department [redacted]

Organization [redacted] Social Security No. [redacted]

Name [redacted] Degree(s) [redacted] Role on Project [redacted]

Position Title [redacted] Date of Birth (MM/DD/YY) [redacted] Department [redacted]

Organization [redacted] Social Security No. [redacted]

Role on Project [redacted]

Department [redacted]

Specific Aims

The goal of our project will continue to be the development and application of molecular genetic methods for identifying individuals and families. Specifically, we propose to:

1. Identify and characterize PCR-based, short sequence-repeat polymorphisms specific to nonrepetitive portions of the Y chromosome.

2. Extend our understanding of mtDNA sequence variation within and among American populations.

3. Estimate the frequency of new mutations in the mtDNA control region by studying maternal relatives separated by multiple generations.

4. Develop approaches for mtDNA sequencing for difficult identification cases involving degraded DNA. Continue application of mtDNA sequencing to identify maternal lineages of individuals.

Background

The goal of our project in the past two years has been to develop and apply mtDNA sequencing to identification of individuals and familial relationships. The social, ethical and legal applications of the information include testing claims of identity, providing evidence for the solution of crimes, and identifying victims of wars, natural disasters, and accidents. We have applied mtDNA sequence data to real problems of individual identification and give some examples here.

MtDNA as a genetic system for testing relationship. A wide variety of DNA sequences can be used to test whether two individuals are related. For the identification of maternal familial relationships, mtDNA is ideal for several reasons. MtDNA is haploid, maternally inherited, and homoplastic within individuals (Wilson et al. 1985; Kocher et al. 1989). Therefore, each individual has exactly the same mtDNA as his siblings, maternal aunts and uncles, maternal grandmother, cousins via the mother's sisters, and so on. Second, human mtDNA has been completely sequenced (Anderson et al. 1981). Third, human mtDNA has evolved more rapidly than nuclear DNA and is extremely variable among individuals (Horai and Hayasaka 1990; Orrego and King 1990; DiRienzo and Wilson 1991; Stoneking et al. 1991). The 1200 basepair control region of mtDNA, near the origin of replication and flanked by tRNA (Pro) and tRNA (Phe) genes, is particularly variable. The control region does not code for any genes, which perhaps has released it from strict nucleotide conservation. This diversity

Mr. Chairman and Members of the Committee, I appreciate the opportunity to appear before you today to discuss the opportunities and concerns raised by the proposed Human Genome Diversity Project. Some of these issues are related to the Office of Technology Assessment's (OTA) ongoing study, "The Human Genome Project and Patenting DNA Sequences." This assessment, which I direct, is scheduled for delivery to the Technology Assessment Board in April 1994.

I would like to emphasize at the outset, however, that my remarks will focus on the Human Genome Diversity Project, which is distinct and separate from the Human Genome Project. My comments will be a general overview of the Human Genome Diversity Project and a broad identification of some of the issues that the project might raise. Currently, OTA is not conducting a full assessment of the Human Genome Diversity Project, as it did for the Human Genome Project in its 1988 report *Mapping Our Genes -- The Genome Projects: How Big, How Fast?*. Nevertheless, I anticipate that some of our current analyses for the "DNA patents" project -- e.g., of intellectual property protection of DNA sequences and technology transfer issues in -- will be pertinent to a few aspects of the Human Genome Diversity Project. As I will elaborate further, however, some issues raised by the Human Genome Diversity Project are beyond the scope of OTA's current study.

Background

In humans, as in essentially all forms of life, deoxyribonucleic acid -- DNA -- contains the entire genetic blueprint for an individual. Today, scientists in the United States and abroad have undertaken the 15-year, $3 billion Human Genome Project. The result of this effort will be a single "reference" map of composite information that is essentially a Caucasian genome.

Yet no two individuals, except identical twins, share the same DNA sequence. Furthermore, genetic diversity clearly exists among populations around the world. The Human Genome Diversity Project proposes a systematic examination of human DNA sequence variation by sampling 20-25 unrelated individuals from each of 400-500 populations of historical interest. It would be undertaken with the expectation that fundamental questions about the origins, settlement, and migration of humans could be examined. As well, the project could elucidate why some populations are more, or less, susceptible to certain diseases.

Discussion -- and some early decisionmaking -- on *if, when,* and *how* to undertake the Human Genome Diversity Project is a matter of some urgency: Several of the proposed research populations are literally becoming extinct.

Federal Funding

The Human Genome Diversity Project is a *proposed* effort, for which its supportors currently seek funding, and I would like to emphasize the importance of a congressional role: Even at this early stage, it is obvious to OTA that the Human Genome Diversity Project is unlikely to move forward in any coordinated fashion without some U.S. Federal funding. Scientists proposing the project estimate its cost at $23-25 million over five years; OTA has not, however, analyzed nor verified this estimate.

Federal funding for the Human Genome Diversity Project might derive from three sources -- singly or in collaboration: The National Science Foundation (NSF), the National Institutes of Health (NIH), or the U.S. Department of Energy (DOE). NSF is appropriate because it is the primary Federal agency that funds anthropology research, and most -- but not all -- of the questions that might be addressed by the Human Genome Diversity Project are anthropological in nature.

NIH and DOE are appropriate because together they fund the U.S. Human Genome Project, and thus are familiar with the latest developments in genetic research. OTA has performed no analyses on funding structure or outlays for the proposed Human Genome Diversity Project.

Since each of the agencies is represented at this hearing, I will not elaborate further on individual or joint capacities to fund the Human Genome Diversity Project. I note, however, that neither NIH nor DOE has included funding for the Human Genome Diversity Project in its current 5-year plans for the Human Genome Project. Hence, Congress would need either to appropriate additional funds to NIH and/or DOE for the Human Genome Diversity Project or instruct the agencies to redirect existing funds to the diversity project -- at the expense of investigations funded under the auspices of the Human Genome Project. Further, I am not familiar with the details of funding for anthropological research through NSF and the extent to which NSF could fund this project -- again solely or in cooperation with NIH and DOE.

In sum, it appears that if the Human Genome Diversity Project is to receive Federal funds, an analysis is required to determine what level of funds is needed to conduct the project, how much is currently available from reprogramming of NSF, NIH, or DOE funds, and what new appropriations might be necessary.

Technical Issues

The Human Genome Diversity Project presents several technical issues, which other witnesses can certainly address in greater detail. Briefly, some considerations include the following.

- Although preliminary work has been undertaken to identify the populations for DNA sampling, Dr. Richard Ward, an anthropologist involved in the initial effort raises concern that the criteria for identification vary among the groups, which were

divided by geographic area. The issue is: What criteria should be employed to create a final sampling list to ensure that the project is consistently implemented? Then, does the current proposed list meet the criteria?

- The project proposes to collect blood samples from indigenous peoples. Certain blood cells will be transformed into what scientists call "immortal cell lines." This process must occur within 72 hours of collection, but offers the advantage of largely preserving the genetic heritage of an individual in perpetuity. Nevertheless, certain interesting genes -- in particular those involved in immunity to disease – will not be amenable for analysis after transformation; other DNA changes, currently unknown to scientists, might also be associated with transformation. Thus, besides collecting blood samples, what other biological material should be collected for DNA analysis (e.g., hair, cheek swabs)? From whom should this material be collected? Just the 25 individuals from whom blood has been drawn? Since interesting information can be derived from DNA samples alone, how many additional human research subjects are proposed for supplemental sampling?

- What biological material will be stored? How will it be stored? Who will manage access? Who will oversee quality control and quality assurance? Immortal cell lines require ongoing (in perpetuity) upkeep: Who will pay for these costs? What kind of databases will be used to manage the storage and retrieval of biological material? What about informatics for the information generated from the biological material? As with the biological material, who will manage the information quality, access, and upkeep?

- What minimum set of genetic markers should be analyzed for all samples? In what priority will samples be analyzed once a common set of markers has been identified?

Although these are important questions, OTA's forthcoming report on patenting DNA addresses none of them.

Ethical, Legal, and Social Issues

The proposed Human Genome Diversity Project raises several ethical, legal, and social issues, and it is impossible to fully discuss them within the confines of my testimony. It is critical, however, that these issues be addressed.

For example, many of the populations proposed for sample collection are in developing nations. What if certain data gathered under the Human Genome Diversity Project proves commercially valuable? The United States must be sensitive to the concerns of developing nations, while simultaneously preserving legitimate interests of U.S. companies to pursue commercial development and intellectual property protection of "biotechnological" products. Developing countries have already expressed concerns about the Human Genome Project and the general issues surrounding DNA patents. These concerns are similar to those that were spotlighted at the Rio de Janeiro summit last year in the context of intellectual property protection of novel plant and animal biological material and the "Biodiversity Treaty." Not surprisingly, however, they are heightened by the prospect that the substance now in question is human biological material from vulnerable populations. As Chaim Sheba of Israel has put it, "You have taken our gold and diamonds; now you are taking our genes."

Beyond legal issues of intellectual property protection are several ethical and social questions.

- Will any benefits of the research accrue to the research subjects?
- What are the risks -- particularly social risks such as stigmatization -- to research subjects who participate? To those who decline to participate?

- What about compensation -- monetary or otherwise -- for research subjects? What if a country seeks payment in return for the collection of biological samples from its citizens? What if a local leader demands compensation? If compensating individuals is an option, what happens to people not "chosen" as research subjects? What benefits realistically can be offered to participating communities, especially if they are small and isolated?

- What about confidentiality of the samples collected? What limits are necessary? Who decides? Who will be responsible for ensuring confidentiality?

- In addition to collecting samples for the Human Genome Diversity Project, what about testing blood samples for disease? In particular, special attention to HIV infection is an issue -- especially with blood samples from Africa, for example. Should samples be tested so that researchers can exercise greater caution when handling certain samples? Should the samples be tested anonymously to determine infection rates in these populations? Is it ethical to test samples anonymously? If it is unethical, then how does a demand to link a sample with an individual for purposes of disease identification balance against the importance of confidentiality for other purposes? If anonymous testing is viewed favorably and HIV testing is deemed desirable for the safety of scientists or to analyze HIV infection in indigenous populations, how will pretest counseling for HIV be handled on top of informed consent for the Human Genome Diversity Project, generally?

- The issue of informed consent raises its own questions: What constitutes meaningful informed consent in non-Western cultures? In fact, do Western notions of informed consent have any true relevance to some of the populations to be sampled? Nevertheless, informed consent, expressed in whatever form, is a minimum activity

to demonstrate respect for a culture. And while Western notions of consent might differ drastically from those of the populations to be studied, the concept of respect surely persists.

Nevertheless, any research conducted with Federal funds will need to comply with current U.S. regulations governing human subjects research, in addition to adhering to any local governmental rules. U.S. regulations lay out eight specific informed consent requirements, and I will mention at least three that may result in conflict between U.S. regulations and practices, values, or beliefs in other societies: a requirement for a written document, a clear explanation of the purposes of the research, and individual consent.

A written document is important because it endures as a record for the future -- either as an instructive tool or for auditing purposes. Yet written documents will be an anathema in some populations. How can this be reconciled with Federal regulations?

Similarly, Federal regulations require that a project's purposes be explained. One of the goals of the Human Genome Diversity Project is to elucidate information about the origin of the sample population, as well as its relationship to other populations. On face value, this might easily be put in terms understandable to all cultures. However, some cultures have deeply rooted beliefs about their origins and would find this goal of the project insulting or offensive. What if this jeopardizes efforts to obtain samples from key communities? Would it be ethical to emphasize certain goals -- e.g., identifying disease susceptibility -- over others -- e.g., examining human origins -- in order to facilitate consent and participation? Who decides?

Must informed consent be obtained from every individual in a community? Can local leaders and other central authorities decide for all members of a community? Consider for example, a situation where a local leader speaks for all members of his group. What if he agrees to the sampling, but researchers who are preparing to draw blood from a woman plainly see that the woman is distressed by the prospect? Do they proceed? Do they decide not to sample her blood? Will this now incur the wrath of her leader, who has "lost face" because he had given his word that all would participate? Will she be punished overtly? If not overtly, will this stigmatize her? Should the investigators try to gain her individual consent? Is any consent she might then give truly consent, or has it been coerced?

- If protocols vary (e.g., regarding compensation or consent) from culture to culture, as might be expected, who will arbitrate what's necessary to ensure the protection of the human research subjects? Will population-by-population approval be necessary -- especially if U.S. funds are used? If individual protocol review is deemed cumbersome by researchers, is there a reasonable expectation that all contingencies can be identified prior to embarking on sample collection?

- What about issues of genetic discrimination? What is the best mechanism to minimize misuse and misinterpretation of the gathered data? In particular, since genetic *differences* are the proposed project's focus, concerns are raised about information being used to support notions of superiority of one group over another. Dr. Diane Paul, who has studied the social history of genetics, believes that the Human Genome Diversity Project is likely to reinforce conventional views of race and ethnicity.

I have elaborated on ethical, legal, and social issues at some length, but by no means in an exhaustive fashion. I reiterate that analyzing these considerations is especially critical because many of the populations that have been proposed for sampling are groups that historically have been vulnerable or exploited. And while anthropologists have dealt with some of these questions in their research for several years, the scale of the proposed Human Genome Diversity Project changes the dynamics, as does the current international political context.

The forthcoming OTA report on intellectual property protection for human DNA sequences will examine the question of patenting DNA, per se. However, it will not directly address the issue of property rights that might be sought from samples such as those that would be gathered under the Human Genome Diversity Project. It also will not analyze the ethical issues I have just touched on.

The Outlook for the Human Genome Diversity Project

Mr. Chairman, as was my task, I have focused on identifying issues that the Human Genome Diversity Project raises and have not offered specific options to address them. Nevertheless, I think it is important in concluding my testimony to stress that, despite the many questions I have enumerated, OTA does not view the Human Genome Diversity Project either negatively or positively. Rather, I emphasize that OTA supports a thoughtful and deliberate discussion involving a broad spectrum of international perspectives, so that Congress can consider a full range of options. Clearly, a balance must be struck between our intellectual desire to pursue an interesting, exciting line of inquiry -- one that may yield information about human origins, as well as medically important data -- against both our responsibility to devote research resources in the most efficient and prioritized manner and, more importantly, our ethical obligation to respect and enhance the welfare of all peoples.

Mr. Chairman, thank you for the invitation to discuss this important topic. I will be happy to answer any questions you or Members of the Committee might have.

HUMAN GENOME DIVERSITY:

A PROPOSAL FOR TWO PLANNING WORKSHOPS AND A CONFERENCE

Human Genome Diversity:
A Proposal for Two Planning Workshops and a Conference

SUMMARY

Recent advances in molecular genetic analysis now enable us to assess human genetic diversity at a level of detail never before possible. We hope to promote an international, interdisciplinary project to link the power of human-genome analysis with anthropological wisdom, with mathematical tools of population genetics, and with other sciences. The goals of the Human Genome Diversity project are to understand who we are as a species and the processes that have led to our current genetic structure. The approach will be first to identify current representatives of descendants from ancestral human populations worldwide, with particular attention to those whose unique identity will soon be lost, and then to preserve the genetic record of those populations as data from a range of nuclear and mitochondrial DNA sequences and as immortalized cell lines. Support to collect samples and to carry out genetic analysis would be by open competition. Genetic data, genetic material (DNA), and cell lines would be universally available to the scientific community.

In order to help organize the Human Genome Diversity project, we propose two planning workshops and a conference.

Workshop 1: How to sample? A small workshop of population geneticists, anthropologists, and other scientists, to define an optimal sampling strategy for representing human diversity.

Workshop 2: Whom to sample? A workshop made up primarily but not exclusively of anthropologists, to identify important populations to be sampled.

Conference: The Human Genome Diversity project. Geneticists, anthropologists, linguists, paleontologists, archaeologists, and descendants of ancestral populations will discuss the results of the workshops, along with the ethical aspects of the project; integrate new genetic technologies, sampling strategies, and choice of populations; and set priorities and research strategies.

INTRODUCTION

The Human Genome Diversity (HGD) program is a major endeavor with two complementary aims: (1) the preservation of crucial biological information on vanishing human populations, and (2) the analysis of human genetic diversity, in conjunction with the current Human Genome project. The program was started by HUGO (the Human Genome Organization) and was set officially in motion by the president of HUGO, Sir Walter Bodmer, who named a committee for the study of human genome diversity. Members of the committee are

Julia Bodmer, Imperial Cancer Research Fund (ICRF), London;
L. Luca Cavalli-Sforza (chairman), Genetics, Stanford;
K. K. Kidd, Genetics, Yale;
M.-C. King, Epidemiology, UC Berkeley;
S. Pääbo, Genetics, U. of Munich, Germany;
A. Piazza, Human Genetics, U. of Turin, Italy;
M. Siniscalco (co-chairman), ICRF, London.

More details about the project will be found in the preliminary program prepared by the committee and enclosed as the first document of the Appendix, together with an appeal that appeared recently in *Genomics* but was widely circulated earlier. There has been a substantial, positive response in the scientific and lay communities to this proposal. Other documents in the Appendix include examples of this response in the scientific press.

In order to carry out the HGD program, it will be necessary to obtain adequate funding, which will be applied for in due time. This is a complex project, however, the results of which will be of interest to a large number of scientists, human geneticists, anthropologists, evolutionists, linguists, and archaeologists. This application aims to convene two workshops and a conference, to be held reasonably soon, which should generate information useful in the planning stage of the project. The three meetings will cover the subjects indicated below and should be held in the following sequence.

Workshop 1: How to sample? There are various questions that could be profitably answered at this stage. A small workshop of population geneticists, anthropologists, and other scientists should define optimal ways of obtaining samples of human populations to best represent human genetic diversity.

Workshop 2: Whom to sample? A workshop composed primarily (but not exclusively) of anthropologists will work to identify populations of relatively unmixed descendants of the most interesting and representative ancestral populations. Special attention should be given at first to populations with the greatest risk of vanishing.

Conference: The Human Genome Diversity project. Geneticists, anthropologists, linguists, archaeologists, paleoanthropologists, and descendants of ancestral populations will discuss the ethical aspects of the project, as well as the results of the two first workshops, considering the problems of integrating the new genetic technologies, sampling strategies, and choice of populations. The conference will set priorities, research strategies, and organizing principles of the project.

Information generated in each meeting will be made available to the meetings that follow as detailed reports to be prepared by chairpersons of the sessions. It is envisaged that the last and most comprehensive meeting will be the source of a publication, such as a symposium volume, which would also include material from the first two workshops. Information will also be made available to the scientific community at large through articles in the scientific press, and we will try to arrange for summaries to appear in widely read journals like *Science*, *Nature*, *Anthropology Newsletter*, etc., as has already happened before this application.

The dates and places for these three planning meetings are still tentative: the first workshop in mid-July 1992, at Stanford; the second, in the fall 1992, at Pennsylvania State; and the third, the conference, at the end of 1992 or in the winter or early spring 1993, in Europe.

We want the effort to be as broadly international as possible. The first two meetings are very specialized; many experts in the two fields reside in North America, but it is anticipated that many foreign participants will attend, and their number may increase if, as hoped, HUGO will provide for funding of the travel of some more foreign members. The third meeting will bring together scientists from many more disciplines; there are, therefore, fewer constraints on having a full international representation. A letter from the scientific administrator of HUGO Europe, Dr. Liz Evans, states the willingness of this organization to apply for funds adequate to invite people from countries outside North America. Currently, for statutory reasons, HUGO Europe cannot pay travel to scientists from North America, and we have therefore included, in the conference portion of this application, funds for scientists to travel only from the United States and Canada to Europe.

The following proposal was prepared by a subcommittee of HGD -- Luca Cavalli-Sforza, Marc Feldman, Mary-Claire King, and Ken Weiss -- meeting at Stanford from November 30 through December 3, 1991. Ken Kidd was also invited but was unable to participate in person because of illness; he contributed important comments. Input by other members of the HGD committee is also gratefully acknowledged. For reasons of administrative expediency, it was preferred that only one of us should be principal investigator. Marc Feldman, director of the Morrison Institute for Population and Resource Studies, was named. The Institute has staff that would help with this application and with the organization of the meetings. All the other subcommittee members appear as co-principal investigators to signify their total involvement and commitment. The subcommittee will act as local organizing committee for the first two workshops and collaborate with the HUGO committee in organizing the final conference.

BACKGROUND AND RATIONALE

The genetic diversity of living humans harbors the clues to the evolution of our species. Human genomes that exist today have been molded by the combination of evolutionary forces that have acted on populations throughout their history. Information from nuclear and mitochondrial genes from present-day populations can inform us about prehistoric migrations, natural selection, social structure, and the frequency and types of mutations our species has experienced. Genetic studies of current representatives of ancestral populations can be integrated with findings from archaeology, linguistics, and history to develop a more complete picture of our past than has heretofore been possible.

Some of the most fascinating questions of biological anthropology, epidemiology, cultural anthropology, and archaeology can be answered by applying the tools of genome analysis to human populations. The Human Genome Diversity project will contribute to the resolution of some of these questions. It will also provide resources, including universally available cell lines and data on genetic background and diversity, that will permit these issues to be addressed productively by workers from a wide variety of disciplines. Worldwide sampling and genetic documentation of human populations can offer unprecedented opportunities to address such questions as the following.

How has human phenotypic variation evolved?

Have adaptations such as skin pigmentation (or depigmentation) occurred once and spread by migration, or have there been multiple independent mutation and selection events?

In what cultures is obesity genetic? Why?

Why are pygmies short? How many different genetic ways of being short are there among ancestral populations?

What are the genetic bases of the various morphological and physiological adaptations to altitude, climates, etc.?

To what extent can distributions of diseases in modern human populations be explained by human genetic diversity?

Can the wide geographical variation in the incidence of insulin-dependent diabetes mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus erythematosis, and other autoimmune conditions be attributed to different frequencies of alleles at immune-response loci such as HLA and the immunoglobulins?

Are the high rates of hypertension and maturity-onset diabetes in some present-day populations attributable to high frequencies of alleles of critical genes that may have been under selection earlier in human history?

What genetic adaptations to infectious agents in the past may have led to differential susceptibility to present-day viruses and other pathogens?

How are human societies structured?

How do people move? How do languages and cultures move with people?

Who mates with whom? What are the relative mobilities of males and females? What is the distribution of the rates of endogamy, of inbreeding, etc.? Do tribal leaders produce more children? Which female lineages are most successful?

Where did modern human societies come from?

How did humans move out of Africa? Were there multiple emigrations? How did successive groups of emigrants mix?

How did migrations within Africa occur?

Were the Americas settled in waves or streams? How many total immigrants were there?

How and by whom and when was the Pacific settled?

What are the origins of linguistically unique populations, such as the Basques, the Hunza, the Nahali, and so on?

Urgency and Timeliness of the Project

The populations that can tell us the most about our evolutionary past, and its implications for the present, are those that have been isolated for some time, are likely to have been culturally and linguistically distinct, and are often surrounded by geographical barriers. Human populations that are currently isolated and have apparently been isolated for a long time (even if subject to high rates of drift) contain much more historically useful genetic information than more recent, urban ones. However, such isolated human populations are rapidly merging with their neighbors or emigrating in search of job opportunities, destroying the information needed to reconstruct our evolutionary history. Population growth, famine, war, and epidemics also threaten once-stable populations. Urban populations are especially problematic because admixture is so much more likely and is harder to discount for most individuals. It would be tragically ironic if, during the same decade that biological tools for understanding our species were created, major opportunities for applying them were irretrievably lost.

Our common history can be preserved by obtaining, preserving, and studying DNA samples from present-day populations that give the most insight about ancestral populations. The preservation of human genetic history requires a worldwide effort and the active involvement of anthropologists, linguists, and representatives of the populations themselves. The scientific and logistical questions raised by the Human Genome Diversity project are complex, but they can be answered by bringing together available people and resources. The two workshops and conference described below are intended to address these questions.

Confidentiality

We list below proposed participants in the two workshops and the conference. Invitations have been made subject to funding. There has not been time for everyone to be contacted or to respond, and some might decline or suggest alternates. In particular, names for the conference will to a great extent be suggested by the European organizers. Therefore, we ask that the participants' names in this application be kept confidential.

WORKSHOP 1: POPULATION GENETICS. Proposed dates: July 17-19, 1992
Proposed venue: Stanford University, Department of Biological Sciences (T-175)

The first workshop will address the mathematics and population genetics of sampling, including whether sampling units should be individuals or populations or both; the kinds of genetic material most informative for evolutionary studies; and the appropriate size and spatial structure of the sample.

Interpretation of observed genetic variation will depend on the scale at which this variation is measured. Thus, genetic differentiation of populations using point polymorphisms (e.g., restriction-fragment-length polymorphisms, or RFLPs) has usually been assessed using a chi-square test of allele frequencies in samples from different localities (see Nei 1987, p. 227). Interpretation of variation between populations at the level of DNA sequences, however, will require statistical methods that are now under development. These new methods will most likely be based on recent

mathematical evolutionary theory of molecular variation in small populations under the simultaneous influences of mutation, selection, and recombination (see the review in Hudson 1991). Parallel to these mathematical developments have come a number of new computer-based techniques for studying historical relationships between populations using phylogenetic trees (see, e.g., Li and Graur 1990).

Toward the goal of studying human genome diversity, we have identified two sets of questions whose discussion will contribute to the agenda of the project. These are questions that address the empirical and statistical bases of the project.

Question set 1: The nature of the genetic variation to be investigated

Should data be collected about frequencies of alleles at polymorphic sites, or should attention be restricted to DNA sequences?

How do logistical differences associated with each class of data affect the planned data collection?

Should individuals, mother-child pairs, or families be sampled? For questions about the relationship of genetic variation to disease susceptibility, and for construction of haplotypes, families are an important resource. For evolutionary history, however, the members of a family are not statistically independent.

Question set 2: Geographical distribution of individuals to be sampled

Should samples be taken at regular intervals in space (grid sampling), should they be taken at random with respect to geographical coordinates, or should a number of individuals be sampled from each group deemed important on the basis of anthropological, archaeological, and linguistic considerations?

To what extent should vanishing populations be selected over larger, expanding ones?

What are the best methods for estimating the extent of gene flow or admixture between groups (Weir and Cockerham 1984; Slatkin and Barton 1989; Cavalli-Sforza and Feldman 1990; Bowcock et al. 1991)?

Do theoretical and statistical properties of admixture depend on the extent to which a population is expanding?

How should information on mitochondrial DNA and nuclear DNA be combined for historical reconstruction?

What level of observed genetic polymorphism (in point polymorphisms or DNA sequences) is appropriate for informative estimates of admixture and for statistically accurate reconstructions?

What is the relationship between the evolution of mitochondrial DNA and the evolution of nuclear genes?

How can site polymorphisms, length polymorphisms (including insertions and deletions and tandem-repeat variants), and sequence variation be combined in analysis?

Preparation of Workshop Report

The organizing committee will present a report of the workshop findings to the second workshop and to the international conference. Three sessions (each subdivided into two parts) are tentatively envisaged as described below (empirical consid-

erations, modeling, statistical analysis). Ample time will be left for discussion, keeping in mind the two sets of questions listed above and others that may be suggested by the participants.

Position papers and other material will be circulated to participants in advance of the meeting. Draft reports of each session will be prepared by the chairmen, helped if necessary by other participants and distributed to participants before the end of the meeting.

A set of protocols with instructions on the collection of material will be prepared and passed on to Workshop 2 and to the conference.

There will be a final session to discuss all draft reports. It is likely that only minor polishing will be necessary after the workshop ends. The final document will be made broadly available for discussion to the interested scientific community. It will also be made available to Workshop 2 and to funding institutions. We will also endeavor to see that it is published in relevant journals, perhaps as an editorial.

Participants in Workshop 1

Our logic in developing a list of scientists for the workshop on population genetics has been to include representative mathematical modelers, statistical geneticists, and researchers involved in collecting the data on which our current knowledge is based; some prospective participants fall into more than one category. Of course, these participants are only representative of these areas; they are by no means the only possible choices. Broad international participation is desirable, if not necessary; if additional funds can be raised by HUGO, such experts as Bill Hill from Edinburgh, Takashi Gojobori from Mishima, and Max Baur from Germany will also be invited. In addition, key personnel from funding agencies should play a role in seeing that the proposed science fits the missions of their respective institutions. For this workshop, as for the others, we anticipate involving graduate students from institutions in the area of the meeting. The HGD will generate a wide array of opportunities for graduate research.

Local organizing committee for Workshop 1
L. L. Cavalli-Sforza, M. W. Feldman, K. Kidd, M.-C. King, K. Weiss

I. Empirical considerations in sampling populations
Introduction: L. L. Cavalli-Sforza
Part 1. Chaired by L. L. Cavalli-Sforza
M. Kreitman (Princeton)
M. Stoneking (Penn. State)
Part 2. Chaired by M.-C. King (UC Berkeley)
J. Bodmer (U.K.)
A. Piazza (Turin, Italy)

II. Modeling human genetic variation
Part 1. Chaired by K. Weiss (Penn. State)
R. Hudson (UC Irvine)
J. Gillespie (UC Davis)
Part 2. Chaired by M. Feldman
H. Harpending (Penn. State)
M. Slatkin (UC Berkeley)

III. Statistical analysis
Part 1. Chaired by J. Felsenstein (U. of Washington)
W. Maddison (U. of Arizona)
M. Nei (Penn. State)
Part 2. Chaired by M. Feldman (Stanford)
B. Weir (North Carolina State U.)
L. Cavalli-Sforza (Stanford)

Other people will be asked to participate (they are classified into five scientific categories, but several people should appear in more than one):

anthropologists -- K. Aoki (Japan), Judy Kidd (Yale), A. Rogers (U. of Utah);

empirical population geneticists -- A. Baer (Oregon), J. Bodmer (U.K.), A. Di Rienzo (UCSF), L. Jorde (U. of Utah), R. Lewontin (Harvard), P. Parham (Stanford), P. Smouse (Rutgers);

statistical and mathematical scientists -- G. Bonney (Howard), F. B. Christiansen (Denmark), J. Crow (U. of Wisconsin), F. Demenais (France, now at Howard U.), B. Efron (Stanford), W. Ewens (Australia), H. Tachida (Japan), N. Takahata (Japan), S. Tavare (USC);

evolutionary linguists -- Joseph Greenberg (Stanford), K. Hale? (MIT), M. Ruhlen (Palo Alto), W. S.-Y. Wang (Berkeley);

institute people -- I. Eckstrand (GM-NIH), D. Galas (DOE), B. Graham (NCHGR-NIH), Judy Greenberg (GM-NIH), M. Teitelbaum (Sloan), M. Weiss (NSF). In addition, international agencies that may provide support for the large project, such as WHO and the EC Science Foundation, will be asked to send representatives at their own expense.

WORKSHOP 2: ANTHROPOLOGY. Proposed dates: October 9-12, 1992
Proposed venue: Pennsylvania State University (Keller Conference Center)

Many anthropological objectives can be furthered and modernized by the HGD initiative. One of these is to understand how human diversity is produced by mutation, distributed by the homogenizing effects of demic exogamy, and restricted by geographical and topographical distance. Were we a homogeneous species with a stable population structure, then isolation-by-distance models would provide a good description of human diversity, and a geographical grid sample could be used to sample that diversity. A grid that is small enough, or is concentrated in areas where ancestral peoples still live, would include many "tribal" groups in the sampling net. Human history certainly involves stochastic expansions and declines of local demes that would be sampled by a grid. However, there are important anthropological reasons to sample in other ways as well.

Major population disturbances, invasions, and absorptions have greatly distorted the pure geographical basis of ethnohistory. Examples are the movement into the Valley of Mexico from what is now southern California, politically displacing and eventually outnumbering indigenous populations there; the political turmoil in Andean South America observed at the time of contact by Europeans; the current "chaotic" distribution of language and cultural affinities in Amazonia; the presence of Athabascans in the U.S. Southwest; the Bantu expansion and San contraction in Africa; and the spread of Neolithic farmers into Europe. The genetic structure of

important human crossroads areas such as northern India are also poorly understood at present.

These events can best be represented by appropriately selecting current populations who reflect the human history of a geographical region. Biological anthropologists and ethnologists understand the current distribution of cultures in their areas of expertise; archaeologists, ethnohistorians, and anthropological linguists can contribute information on the prehistory of an area.

Adaptation, phenotypic variation, and gene function are fundamental concerns of biological anthropology. With the exception of studies of abnormal hemoglobins and malaria, these subjects have been less developed, because the necessary technology has not been available. Many more functional genes have now been identified. The organization of the human genome and the relationship between genotype and phenotype are increasingly well understood. Combined with an understanding of variation at these loci, these developments provide every prospect that gene expression and function will gain greatly in importance to anthropologists.

Data collected so far provides a sense of the DNA sequence variation in Caucasians, from whom most sequences have been sampled. But our understanding of the wealth of human diversity in other populations, indeed the majority of our species, is quite poor (Cavalli-Sforza 1990). The worldwide sampling proposed in the HGD project will provide the information that will help to program further populational studies, should they prove necessary.

Issues to Be Addressed at Workshop 2

The purpose of the proposed meeting is to establish a sampling agenda for the major geographical areas of the world. A small committee representing each area will collate and synthesize information on that area and recommend sampling priorities. The meeting as a whole will attempt to fit these recommendations into an overall plan. These recommendations will advise research initiatives for the HGD project.

A special effort will be made by workshop participants to identify populations threatened with imminent cultural or biological extinction or dissolution, which should therefore be sampled on an urgent basis. In addition, issues of sampling intensity will be discussed. Are there cultures from which entire villages should be sampled, including information on genealogy, marriage, migration, and socioeconomic status? Such data would be costly to collect but might contribute critically to our understanding of demic interactions and human behavior under tribal conditions. Or would a broader sample suffice? Should each language group be represented?

An essential issue is how to sample diverse, complex, substructured societies. Many aboriginal populations have been integrated into their surrounding populations. How can these subpopulations be sampled so as to recognize their contribution to genetic diversity? Appropriately including subpopulations has social consequences, because long-term biomedical benefits may accrue from knowledge of human variation.

An ethical issue is how subjects of the sampling, particularly isolated populations, benefit from the project. Can we ask people to contribute a blood sample, even of only 10-20 ml, without being part of the problem of cultural exploitation? Our

thinking has been guided by the example of work among the Yanomama by J. Neel, N. Chagnon, and their colleagues (Chagnon 1968; Neel 1978). Their work contributed enormously to our understanding of human genetic diversity, but it also focused scientific attention on Yanomama society. This scientific attention was turned to political advantage on behalf of the Yanomama (see Gibbons 1991), when they were threatened by the development of their region of the Amazon. In the same way that Amnesty International seeks to stop abuses of human rights of individuals by drawing the attention of governments and of the public to them, the HGD project can bring attention to the plight of threatened populations. The question is how to do so effectively and in concert with the wishes of the populations themselves.

The opportunity to incorporate human diversity into our broader picture of the human genome is one that can be of great importance to anthropology and to human sciences generally. It is unlikely that in our lifetimes any second chance will arise. Resources are limited, but the will to support this effort is there (see *Time* 1991) and the opportunity should be seized. It is important to construct the sampling design promptly and openly. We propose various mechanisms to achieve this end.

Request for Advice from Anthropologists before the Workshop

The workshop will be attended by about 50 representatives of the world's culture areas, experts chosen from those available and within the limits of the budget. However, to cast the net of opinion and information as widely as possible, we propose first an open request for advice: who should be in the sample? who should collect it? how should it be collected?

We have written an open letter to the *Anthropology Newsletter*, explaining the purposes of the HGD initiative and the structure of the planning workshops. We have tried to explain our purposes and the need for advice from many sources. We request help from the profession at large. A copy of the letter is attached; briefer solicitations will be published in *Science* and elsewhere.

Responses to these requests for advice will be distributed by workshop organizers to individual committees for each culture area. The committee for each area will include biological anthropologists, ethnologists, linguists, and/or archaeologists. Their job will be to organize, collate, synthesize, and evaluate the information and to compile a report on the cultural aspects of the sampling strategy.

Structure of the Workshop and Report of Proceedings

(1) The workshop will begin with a plenary session at which general objectives for each culture area are agreed upon (this should be short, because much work will be done in advance of the workshop). (2) The committee for each major culture area will spend one or two days discussing their area and preparing a report, using maps and identifying groups by a standard means of classification. Possible ethnological classifications include the HRAF files, ethnographic atlases, ethnographic encyclopedias, and various other linguistic compilations. We are not proposing an ethnographic conference or an ethnological study. This is primarily a project about biological, rather than cultural, diversity. It is not intended to represent all cultures or cultural traits. However, ethnological information is essential to inform us about who should represent our species, its history, and current diversity. (3) The participants will meet in plenary session, with reports given and discussed for each culture area and for ethical issues.

The workshop will produce a final report for the third, international conference. This report will be published and abstracts or summaries sent to the *AAA Newsletter*, *Science*, and the like. The report will also include another call for advice and will thus constitute a continuing open public record of the project.

Participants in Workshop 2

We are well aware of the importance of identifying specific participants. We are under some time pressure, because HUGO and other agencies interested in funding this initiative would like to get started. A number of delicate issues are involved in establishing a roster. First, time has prevented us from contacting everyone we would like to have participate (and/or from receiving an answer). Second, we do not wish to establish this list too definitively before receiving responses to our solicitations for help, since those responses may help inform our selection of participants.

The choice of 50 people to represent the world will also inevitably omit prominent persons and cannot possibly represent all points of view. We have tried to identify persons who can represent a diversity of views about their respective areas. The following list includes the world areas we have identified, along with a few clearly appropriate persons who have already been approached (an asterisk indicates that the person has been contacted and has expressed an interest in participating). Alternates will be identified for any listed person unable to attend. Several other potential invitees from outside the United States have been identified, and more are being sought. Depending upon the travel support that HUGO can raise, as many of the most important individuals as can be funded will be selected. Broad international representation and participation are crucial. Letters of agreement will be sent to NSF upon receipt.

Local organizing committee for Workshop 2
Ken Weiss, K. Kidd, M.-C. King, Luca Cavalli-Sforza, M. W. Feldman, Mark Weiss (NSF, ex-officio)

Arctic, Siberia, and northern North America
*Emoke J. E. Szathmary, U. of Western Ontario
*Michael Crawford, Department of Anthropology, U. of Kansas

North America
Chief Oren R. Lyons, SUNY Buffalo
*John H. Moore, Department of Anthropology, U. of Oklahoma
*Richard H. Ward, Department of Human Genetics, U. of Utah
Al Ortiz, Department of Anthropology, U. of New Mexico
Barbara Mills, Department of Anthropology, U. of Arizona

MesoAmerica
*William Sanders, Department of Anthropology, Penn. State
*Peter E. Smouse, Center for Theoretical and Applied Genetics, Rutgers
Ramiro Barrantes, U. of Costa Rica

South America
Napoleon Chagnon, Department of Anthropology, UC Santa Barbara
Francisco Salzano, Puerto Alegre, Brazil
*John Rick, Department of Anthropology, Stanford
*Robert Carneiro, American Museum of Natural History
Gabriel Escobar, Cusco, Peru

Europe
Luca Cavalli-Sforza, Stanford

India
*K. C. Malhotra

Middle East
*Ofar Bar-Yosef, Department of Anthropology, Harvard
B. Bonne-Tamir, Israel
Representative of Arab populations: to be determined

Southern and western Soviet Union, Afghanistan, Iran, Pakistan, etc.
*Harold C. Fleming, U. of Pittsburgh

Europe and North Africa
Representatives to be determined (plus L. L. Cavalli-Sforza)

Pacific, New Guinea, and Australia
Kuldeep Bhatia
Peter Bellwood, ANU Canberra, Australia
*Jon Friedlaender, Temple U.

Eastern Asia: Japan, Taiwan, China
T. Ishida, Tokyo
Du Ruofu, Beijing

Central, eastern, and southern Africa
*John Yellen (will attend ex-officio from NSF)
*Henry Harpending, Penn. State
Trefor Jenkins, Witwatersrand
*Peter Robertshaw, California State University, San Bernardino
Barry Hewlett, Tulane
*L. L. Cavalli-Sforza, Stanford

For each area examined, the workshop will prepare a list of populations for which sampling is recommended. Each population will be labeled on a map of the area, with a description of its history and the reasons for the choice. The degree of urgency will be classified as follows: (1) urgent, especially important and likely to become extinct soon; (2) not urgent, but sampling is desirable even if the population does not face imminent extinction; and (3) others, which are not so necessary to immortalize but which might provide useful information or substitute for others that are difficult to reach.

In addition to explaining why each group is important, the logistics of how to reach them and the names of anthropologists who might be going there should be indicated, if known, along with other practical information deemed useful.

AN INTERNATIONAL MULTIDISCIPLINARY CONFERENCE FOR THE HUMAN GENOME DIVERSITY (HGD) PROJECT

Local organizing committee
Luca Cavalli-Sforza, chairman, Department of Genetics, Stanford
Julia Bodmer, ICRF, London
Liz Evans, HUGO Europe
Marc Feldman, Department of Biological Sciences, Stanford

Ken Kidd, Genetics Department, Yale
Mary-Claire King, UC Berkeley
Svante Pääbo, U. of Munich
Alberto Piazza, Department of Genetics, U. of Turin
Marcello Siniscalco, ICRF, London
Ken Weiss, Department of Anthropology, Penn. State

The two workshops on how to sample and whom to sample described above may take place by the summer or early fall 1992. They should have discussed exhaustively many of the theoretical and practical problems. A number of other questions remain to be considered, however. Together with the reports from the two workshops, these additional issues should be the subject of an international conference to take place as soon as possible after the workshops.

Because many experts on diverse geographical areas are in North America, we plan to hold the first two workshops in the United States. It is essential that the HGD effort be truly international, however, and the same must be true of the final conference, at which all major decisions will be made. We therefore propose to hold the HGD conference in Europe, possibly in London or Alghero, Sardinia. Walter Bodmer, president of HUGO and director of the Imperial Cancer Research Fund, London, has suggested a focused meeting of 50-60 invited scientists convening for four or five days. At Alghero, a new laboratory of molecular biology is about to open, which will be partly dedicated to HGD. The director of the Alghero Institute, Marcello Siniscalco, is co-chairman of the HUGO committee for HGD and has indicated a willingness to host the conference. A letter from Liz Evans, the scientific administrator of HUGO Europe, indicating the intention of undertaking the organization of this conference, is appended.

By present rules, HUGO Europe can pay travel only for Europeans, and we therefore request support for travel for people based in the United States or Canada to go to Europe for the HGD conference. We assume that about 20 in the latter group will attend.

Subjects to be discussed, in addition to the conclusions of the two workshops, include

-- the ethical issues
-- the state of knowledge of DNA markers
-- new technologies of DNA analysis and their relevance to the project
-- new technologies in cell culture and generation of libraries from very small quantities of DNA
-- relative importance to give to vanishing populations versus large and potentially expanding ones
-- relative importance to be given to different methods of sampling DNA from individuals, and of sampling individuals versus populations
-- appropriate use of various sources of DNA, including nonrenewable ones already in existence (e.g., material collected for HLA workshops; 15 South American tribes collected by J. Neel and collaborators; >1,000 African Pygmies collected by L.C.-S.) or yet to be collected
-- all organizational and financial problems

It is too early to anticipate details of length, exact titles of sessions, or complete lists of participants. The following is a tentative list of topics with possible organizers and speakers. Some topics may be restricted to one lecture or half session or be

extended to a half day. Again, because practically none of the speakers has been contacted so far, it is asked that their names be kept confidential. We plan additional discussions with others, for example, the HUGO Ethics Committee, for further help in deciding on the specific lists of invitees.

Sessions and Topics

1. Introduction: Human Genome Diversity program
   Conference co-chairpersons: Luca Cavalli-Sforza and Walter Bodmer

2. New techniques for the study of DNA and human variation
   Chairperson: Charles Cantor (Berkeley) or Hans Lehrach (London)
   David Botstein, Stanford (techniques from the genome project)
   Alec Jeffreys, Leicester (new methods involving repeat sequences)
   Ronald Davis, Stanford (real-time sequencing, Affinax procedure)
   Henry Erlich, Cetus, Emeryville (new PCR techniques)

   Other technical topics to be addressed: ways to develop new PCR libraries from small DNA samples; how to prepare samples in the field for immortalization

3. Sampling strategies (summary of symposium and further discussion)
   Chairperson: Marc Feldman (Stanford) or Alberto Piazza (Turin, Italy)
   W. J. Ewens, Melbourne
   Bruce Weir, North Carolina State U.
   M. Nei, Penn. State

4. Choice of populations (summary of symposium and further discussion)
   Chairperson: Ken Weiss (Penn. State) or L. Cavalli-Sforza (Stanford)
   Du Ruofu, Beijing
   K. C. Malhotra or Partha Majumder, India
   I. Rychkov, Moscow

5. State of research on mtDNA from current and past peoples
   Chairperson: Mary-Claire King (UC Berkeley) or Svante Pääbo (U. of Munich)
   Mark Stoneking, Penn. State
   Doug Wallace, Emory, Atlanta, Georgia
   Anna Di Rienzo,UC San Francisco

6. State of research on immortalization and analysis of nuclear DNA, including HLA
   Chairpersons: Julia Bodmer (ICRF, London), Ken Kidd (Yale)
   Anne Bowcock, U. of Texas -- Dallas
   Andre Langaney, Paris and Geneva
   Takefumi Ishida, Tokyo
   Judy Kidd, Yale

7. Ethical aspects: Is studying human diversity the right thing to do? What are the benefits to the people studied? Who can represent the interests of indigenous people? The ethical and legal aspects of collecting blood samples from aboriginal populations and of disseminating information thus obtained will be considered.
   Chairperson: Eric Guengst (head, HUGO Ethics and Legal Office), Alain Pompidou and Nancy Wexler (co-chairpersons of HUGO's Ethics Committee)
   Federico Mayor (Director General, UNESCO)
   Representatives of aboriginal populations (Australian, Alacaluf, Brazilian, Yanomama)

Robert Cook-Deegan (Washington, D.C.)
Chief Oren Lyons, SUNY Buffalo (also North American Indian representative)
Trefor Jenkins, Johannesburg
Bankowsky, CIOMS

8. Financial, organizational, and coordinational issues: how is the program to be organized and run?
   Representatives of NSF, NIH, HUGO, DOE, WHO, UNESCO, UNIDO, ICGEB, etc., and the organizing committee

Organizational problems to be emphasized in the discussions include

(1) the creation of a permanent coordinating committee with functions similar to those of the current HGD (Human Genome Diversity) committee named by HUGO;

(2) organization of the distribution of DNA products from cell lines generated by the project (free distribution? at cost?);

(3) organization to make available a basic genetic typing of the DNA, to ensure that a minimum number of markers be analyzed on all or most populations;

(4) ways to favor the development of useful new technologies;

(5) organization of a data base available to contributors and to all scientists;

(6) organization of a data base allowing the spread of knowledge about material (blood samples, etc.) that cannot give rise to cell lines but that can form a nonrenewable source of DNA for research workers, as exemplified above;

(7) establishment of a communication network among all collaborators, through e-mail where possible; and

(8) technology transfer to the Third World, made possible and encouraged by the local collection and testing of cell lines (this program may usefully involve the UNIDO laboratories of Trieste and New Delhi [ICGEB director general A. Falaschi, Trieste, who should also be invited to the meeting] and the network that they have already established in the Third World).

ACTIVITIES TO DATE

Shared interest in these ideas led five of us -- Luca Cavalli-Sforza, the late Allan Wilson, Charles Cantor, Robert Cook-Deegan, and Mary-Claire King -- to call for a worldwide survey of human genetic diversity in a recent letter to *Genomics*. The response was extraordinary. The authors received dozens of letters and phone calls and visits from geneticists, anthropologists, linguists, and public-health workers from all parts of the world, offering suggestions and help (copies of a few of the letters are included in the Appendix). *Science* and *Nature* reported on the developing interest, leading to still more response.

In July 1991, the Human Genome Organization (HUGO) asked seven human geneticists -- Luca Cavalli-Sforza, Marcello Siniscalco, Julia Bodmer, Ken Kidd, Mary-Claire King, Svante Pääbo, and Alberto Piazza -- to serve as an *ad hoc* advisory group to investigate the possibility of a worldwide research program on Human Genome Diversity (HGD). Originally, the group included Allan Wilson, and owing to his health condition, a first meeting was delayed in hopes of including him, but this was unfortunately impossible. Members of this group met in Alghero, Sardinia, on September 20, 1991, and in Washington, D.C., on October 10, 1991. The Washington

meeting was also attended by representatives of the National Science Foundation (Mark Weiss), the National Center for Human Genome Research of NIH (Bettie Graham), the National Institute of General Medical Sciences of NIH (Judy Greenberg and Catherine Lewis), the Human Genome Program of the Department of Energy (Ben Barnhart), the International Congress of Human Genetics (Victor McKusick), and HUGO (Walter Bodmer and Liz Evans). A decision of this meeting was to apply for support for these workshops and conference.

LITERATURE CITED

Bowcock, A., J. Kidd, J. Mountain, J. Hebert, L. Carotenuto, K. Kidd, and L. L. Cavalli-Sforza. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proceedings of the National Academy of Sciences USA* 88: 839-843.

Cavalli-Sforza, L. L. 1990. Opinion: how can one study individual variation for 3 billion nucleotides of the human genome? *American Journal of Human Genetics* 46: 649-651.

Cavalli-Sforza, L. L., and M. W. Feldman. 1990. Spatial subdivision of populations and estimates of genetic variation. *Theoretical Population Biology* 37: 3-25.

Chagnon, N. 1968. *Yanomama, the Fierce People*. Holt, Rinehart & Winston, New York.

Gibbons, A. 1991. Yanomami people threatened. *Science* 252: 1616.

Hudson, R. R. 1991. Gene genealogies and the coalescent process. Pages 1-44 *in* D. J. Futuyma and J. Antonovics, eds. *Surveys in Evolutionary Biology*, Vol. 7. Oxford University Press, Oxford.

Li, W.-H., and D. Graur. 1991. *Fundamentals of Molecular Evolution*. Sinauer, Sunderland, Mass.

Neel, J. V. 1978. The population structure of an Amerindian tribe, the Yanomama. *Annual Review of Genetics* 12: 365-413.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Slatkin, M., and N. H. Barton. 1989. A comparison of three indirect methods for estimating average level of gene flow. *Evolution* 43: 1349-1368.

Time magazine. September 1991.

Weir, B. S., and C. C. Cockerham. 1984. Estimating *F* statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

BUDGET EXPLANATION

As described in the text, three meetings are planned, two workshops and one conference. The latter is to be held under the auspices of HUGO and will be in Europe, possibly London. The budgets for the meetings break down as follows.

Workshop 1, at Stanford

We expect 13 participants from the East Coast, with travel expenses at $500 per person; 7 from intermediate locations, with travel at $300 each; 8 from abroad, with travel at $1,200 each; and 2 from Davis, at $50 each. Total travel is $18,300.

For these 30 visitors, hotel expenses for four nights will be $300 each; and food is budgeted for three days at $75 per person. Subsistence is, then, $11,250 for the 30 visitors. The 9 local participants are not included in this expense estimate. It is expected that there will be considerable mailing and communications costs, for which we have budgeted $2,000.

Administrative costs are requested for the staff person, who will coordinate all aspects of the meeting. We expect to have no expense associated with lecture rooms.

Workshop 2, at Pennsylvania State University

This workshop will be held at Pennsylvania State University, but the reimbursement of expenses will be from Stanford University. Since this involves specialists in the anthropology of remote groups, more participants will come from abroad. We expect 13 from abroad, at $1,500 each for travel; 9 from the West Coast, at $500 each for travel; and 9 from intermediate locations, at $300 each for travel. The total for travel is $26,700.

Subsistence for these 31 is budgeted at $375 each, totaling $11,625.

Mailing, photocopying, and telephone and facsimile expenses, incurred at both Penn. State and Stanford, are expected to be $3,000 here. Joint administrative costs are budgeted at $3,000. Penn. State will incur additional expenses for room rental, audiovisual services, etc., estimated at $2,500, and the cost of renting 10 laptop computers for five days, which will cost $800.

Conference, in Europe

The third meeting is a conference, as described in the text. We expect that 20 people will attend from the United States or Canada, half from the West Coast and half from the East Coast. Average transportation to Europe should be $1,100 per person, totaling $22,000.

Accommodation should be $100 per day, and a food allowance of $30 is proposed. For a five-day meeting, this amounts to $650 per person for a total subsistence cost of $13,000.

Again, we expect the organization to come largely from Stanford and so are requesting the administration and supply costs shown.

HUMAN GENOME DIVERSITY WORKSHOP 1

Stanford, California, July 16-18, 1992

The genetic diversity of living humans reflects the evolution of our species, and an analysis of this diversity provides a great deal of information about who we are as a species and how we came to be. The Human Genome Diversity project (HGD) is an international interdisciplinary program whose goal is to reveal as much as possible about the current state of genetic diversity among humans and the processes that were responsible for that diversity. Classical premolecular techniques have already proved that a significant component of human genetic variability lies within populations rather than among them. New molecular techniques will permit a dramatic increase in the resolving power of genetic analysis at the population level. Recent social changes in many parts of the world threaten the identity of a number of populations that may be extremely important for understanding human evolutionary history. It is therefore urgent to conduct research on human variation in these areas, while there is still time.

The plan is to identify the most representative descendants of ancestral human populations worldwide and then to preserve genetic records of these populations. Support will be requested to collect samples and immortalize cell lines from them, so as to make available a potentially infinite supply of DNA from a number of key populations, and to carry out a genetic analysis of this material. The immortalized cell lines and genetic data would be universally available to interested research workers. Molecular analysis of individual variation in the human genome, in conjunction with knowledge from disciplines such as population genetics, anthropology, archaeology, history, and linguistics, will be used to study our past and present.

This is a report of the Population Genetics Workshop (Workshop 1), the first of three to be held to plan HGD, which was focused on sampling strategies and analytic methods from population genetics. The topics discussed were sampling and population structure; analysis of populations; drift versus natural selection; modeling migration and population subdivision; and population structure and subdivision. Workshop 2, to be held at Pennsylvania State University at the end of October, will focus on anthropological questions. Workshop 3, to be held in 1993, will review reports from the population-genetics and anthropology workshops; include discussion of genetic markers, typing methods, and social, political, and ethical issues; and design an integrated plan for the first five years of HGD.

For their essential help, support, and good-humored patience with us in organizing these

workshops, we would like to thank Mark Weiss, National Science Foundation; Irene Eckstrand, National Institute of General Medical Sciences; Bettie Graham, National Center for Human Genome Research, NIH; and Ben Barnhart, Office of Health and Environmental Research, U.S. Department of Energy.

THE PROBLEM

In recent times, a few populations -- for instance, those of Europe, China, and India -- have expanded greatly, while a much larger number have contracted and in many cases become extinct. Although extinctions, replacements, fusions, and long-distance migrations have been occurring throughout human history, the last 500 years has seen an explosive increase in the pace of such changes. One measure of the threat to many populations is the linguistic change that is being observed. Of the roughly 5,000 languages in the world, some 90 percent are expected to be lost or doomed to extinction by the twenty-first century. As languages disappear, genetically distinct populations often disappear with them, either by physical extinction or by admixture with other groups. The opportunity to study human variability as it existed in the past is rapidly disappearing. This effort should not, however, be aimed at studying vanishing populations simply because they are vanishing but, rather, should focus on those that our current knowledge suggests should be important for understanding our history.

Linguists and anthropologists are planning to videotape endangered populations for future linguistic and cultural study. Stored cells and DNA can provide material for future genetic study. One of the exciting aspects of the proposed program is that it should generate a systematic collaboration between the social and the molecular scientists of unprecedented magnitude and interest.

WHY NOW?

Anthropologists and geneticists have been concerned to study the world's populations for some time, and several important research projects were begun in the 1950's and '60's. From these have come interesting and important results. Furthermore, the studies provide important precedents, and current research can profit by the experience of these pioneers. What makes the current opportunity much greater, however, is the advent of powerful new techniques. These are of two sorts. The first is the ability to sequence large stretches of DNA from small amounts of material. The second is the possibility of immortalizing cell lines, so that they can be stored and maintained indefinitely. This means that the genetic properties can be examined not only now but in the future, using new techniques that are already available and that will certainly be developed further.

QUESTIONS

The uniqueness of the human species and our particular interest in our own population and its recent past means that pure description is important. We are properly interested in human diversity and the similarity and differences among different populations for their own sake. So one objective of the HGD is to document the kinds and frequencies of genes found in different populations and the variation among populations, along with other properties such as language, life styles, and demography being studied by other groups and with which the genetic findings can be correlated.

A second objective is to study more-general evolutionary questions, such as the relative importance of selection and random processes in human evolution and the size of the human population in the past. These questions fall roughly into two classes. The first class of problems includes those that are *locus-specific*, for instance, the association of diseases or other identified phenotypes with specific genes (here, "locus" means "gene" rather than geographical center), or are *population-specific* and address issues that pertain to only one or a few populations. These require large samples from each population. The second class is *genome-wide*. It includes problems for which it is possible to make up for deficiencies in the number of samples by studying a larger number of loci. Examples are the study of drift, migration, and in general, demographic history.

Ideally, we should have large samples from a large number of populations, but this is manifestly impossible, and some sort of compromise must be reached.

A SAMPLING PROCEDURE

A first question is whether sampling should be focused on chosen ethnic groups or should proceed by taking individuals at specific geographical intervals. Although a mix of these strategies is likely to be recommended, it seems clear that a number of important populations are likely to be distributed so irregularly that population-level sampling would be more appropriate to these and would, in addition, alleviate logistical problems connected with cell immortalization. Consideration of ethnic origin would, in any case, be essential even if a rigid geographical approach of sampling single individuals at fixed intervals were carried out.

A second question relates to the manner of DNA storage. Immortalization has the advantage of making the material available almost indefinitely. Its disadvantages are that the procedure is expensive and the cells must be immortalized within a few days of collection. Collection and storage of whole blood or DNA is simpler, but it provides only a finite supply that can eventually be used up.

There seemed to be a consensus that immortalization of cells from a certain number of groups

should be an essential part of the project. In addition, the teams that collect these samples should be able to obtain a much larger number of samples from single individuals in the same general area, including those whose cells will be immortalized. Such samples could include blood, hair roots, and cheek swabs, all of which are known to contain enough DNA for some analyses.

A major component of the discussion centered around a desirable sample size from the groups whose cells are to be immortalized. Many population geneticists suggested that samples as small as 25 unrelated individuals (corresponding to 50 autosomal genomes) per population could be extremely informative, especially for genome-wide problems. It has also been suggested that, for a selected few populations, a larger number of cell lines could be established, including not only more unrelated individuals but also some small nuclear families. The ability to study many loci, made possible by the availability of immortalized cells, more than compensates for the apparently small sample size.

In order to estimate the number of ethnic groups whose cells are to be immortalized, let us assume as a working figure that 10,000 individuals will be sampled for cell immortalization. This figure was assumed because it would generate a relatively modest total cost for the project. The final decision about the total number of samples will be made at the future workshops. With a total of 10,000 individuals, with 25 from each population, some 400 populations would be sampled, constituting about 8 percent of the number of tribes, estimated approximately from the number of languages. Analysis of world populations at the second workshop may well generate a different number of populations to be sampled. Additional individuals, for example, children of unrelated couples whose cells are to be immortalized, could also be sampled according to the other methods mentioned earlier. The total number of these non-immortalized samples could be very large.

In principle, a group of scientists studying a population can collect many more than 25 blood samples. It is necessary to establish rapport with the people and to gain their confidence. It is essential and likely that much additional information will be obtained, such as age, geographical and linguistic origin, and relationships. It is thus reasonable that the scientists be expected to collect many additional samples, from which DNA can be extracted without immortalization as described above. Such data can be used in many ways, in particular for locus- or population-specific questions. The samples can also be augmented by sample collections that are already available and whose existence should be made known to interested researchers by an appropriate organization that might be part of this project.

Samples for immortalization are to be transformed into cell lines with duplicates in two or three major cell banks. Ideally, the immortalization should take place in laboratories in the countries in which the samples are collected. Insofar as is possible, this project should encourage the development of such laboratories in countries where the facilities for immortalization and testing do

not currently exist but could be established. This would constitute valuable technology transfer to these countries.

All biological material collected is to be universally available at minimum cost, although some priority scheme will be necessary for the distribution of peripheral blood samples, etc., from which only limited DNA can be extracted.

HOW ARE POPULATIONS TO BE SELECTED?

The major objective is to obtain a worldwide sample of populations most representative of the human population and of greatest interest for the history of our species. There are many definitions of a population, but a practical and convenient one for tribal populations uses language (or, more precisely, dialect, since that is what often defines the social group) as the primary basis. Groups with distinctive language can often be regarded as a genetic population. By this criterion, there are about 5,000 populations with distinct properties and possibly distinct gene frequencies, but populations to be studied must be carefully chosen on the basis of anthropological information. Table 1 indicates how populations might be allocated to each continent. The second workshop should provide firmer indications on populations of special interest and importance to the goals outlined above.

Some general criteria include the following.

1. The populations should be mostly rural. Large urban conglomerates should be avoided, except on occasions to provide comparisons.

2. Populations should be well defined with respect to language and geography.

3. For many populations, especially the more important representatives of aboriginal peoples, there should be no recent genetic admixture. This is usually an unattainable ideal, but populations with 10 percent or more admixture can be incorporated in the analysis if enough loci are typed so as to test and correct for admixture from known populations.

4. It should be possible to collect blood from unrelated individuals (i.e., no closer than first cousins). In very small populations, however, this may not be possible.

5. Populations that are unique linguistically or in other ways should be especially sought.

6. Populations that are vanishing because of mortality, migration, admixture, etc., and that are potentially important for historical genetics should take priority, especially if they differ from surrounding populations in important ways.

7. The populations should be reasonably accessible (blood samples must reach a processing center promptly) and ideally should have been personally contacted by collaborators in the program.

8. The entire group of populations should be representative of the world, specifically, of the world before the expansion of present dominant groups.

9. It is desirable, if there are several villages or other subgroups, that more than one be sampled. If different ethnic groups overlap geographically in a chosen area, then it is desirable to sample all or the most important of them.

QUESTIONS TO BE STUDIED

Descriptive statistics. These computations characterize the different populations but are not chosen to answer specific population-genetic questions. The sample is thus relevant to the particular population sampled, so the standard error of an estimate -- for example, gene frequency -- is of the order of $1/\sqrt{n}$, where n is the number of persons or genomes sampled. In contrast, if the data are used to estimate a global parameter, such as the number of alleles, the error is often of the order of $1/\sqrt{\log n}$. There are several questions to be studied.

1. What are the frequencies of various genes in different populations?

2. What are the heterozygosities or diversity within and among populations?

3. What special phenotypes, such as diseases, characterize certain populations, and are these correlated with genes or genotypes?

4. How do within- and among-population statistics compare? What is the variation among various human populations?

Genome-wide questions. These are questions that usually permit a trade-off between sample number and gene number, such that the precision of a small sample may be greatly increased by additional loci. The following questions fall into this category.

1. What is the history of human population size? The human population has grown greatly, as is obvious, but the pattern is not known. Were there bottlenecks or sudden expansions? Pairwise comparisons of neutral alleles in mitochondrial DNA in a worldwide sample suggest a rapid expansion in the Pleistocene. This and similar hypotheses can be tested more completely with nuclear genes.

2. What is the history of migration and population subdivision?

3. What is the relative importance of random drift and selection? The debate of a few years ago between the neutralists and selectionists has been largely replaced by the realization that both drift and selection have been important, and the job is to assess their relative roles. Neutral alleles are particularly useful for phylogenetic inferences. There are a number of questions such as the importance of selection for structural or regulatory genes.

4. What is the evolutionary relationship of humans to other primates? The phylogeny of the

human species and its near relatives is now largely settled, thanks to DNA measurements.

5. Is there spatial patterning among populations? Are populations that are geographically close together more closely related? What are the correlations of genotype and distance? Are there geographical clines and, if so, why?

Locus-specific questions. These questions ordinarily require a larger sample than the 50 genomes contemplated for the immortalization procedure. Nevertheless, there will often be information from larger samples of other material collected at the same time but not immortalized. DNA from these will be available in only limited amounts. A couple of key questions can be asked in this context.

1. How is disease susceptibility distributed in and among populations? What accounts for important genotype-phenotype relationships and for genotype-by-environment interactions within and among populations?

Case-control studies of alleles and disease would require more sampling and differently targeted sampling than is anticipated for this project. However, some geographical-epidemiological surveys would be feasible. For example, it would be possible to evaluate variation at a locus of known function across populations or correlations of some easily observed phenotypes with candidate loci among populations.

2. What are the relationships among genetic, cultural, linguistic, and ecological variables?

Many analyses can be done with 25 individuals in each of many populations. These include analysis of principal components of allele frequencies and interpretation of these genetic components with respect to cultural, linguistic, and geographical variables. In places where allele frequencies for specific populations are required and larger numbers of individuals would be useful, directly extracted DNA from non-immortalized samples can be typed.

Other questions of interest are not likely to be answered from data collected in this survey alone. Yet, as ancillary information to other studies, these data can very well prove useful. Several questions can be addressed in this way. How do mutation and recombination rates vary among populations and among loci? How much linkage disequilibrium is there in the population, and is it caused by stratification or by epistasis?

CHOICE OF MARKER LOCI

If different populations are to be compared efficiently, it is necessary that an adequate collection of marker loci be assessed in all populations. Part of the difficulty in using currently available data is that different markers were assessed in different populations; and with only a partial overlap of markers studied in different populations, the analysis is correspondingly inefficient.

The choice of actual markers is a matter for the future workshops, but we recommend that the markers chosen be as unbiased as is feasible and then studied as uniformly as possible on all samples.

RESEARCH ON TECHNIQUES

Most of the innovations in the study of DNA are likely to come from other sources, such as the Human Genome Initiative, and this study will profit from them. There are, however, some kinds of techniques for which improvements made outside this program are less likely. One is the development of better and less expensive methods for immortalizing cell lines. Especially needed are easily portable equipment and other innovations permitting work in or near the field. Another area in which research would aid this project is finding methods for getting DNA from ancient material, such as currently found in archaeological or paleoanthropological samples. Related to this, research to develop enzymatic methods to amplify up the entire genome would be important for this kind of project.

Research in techniques is not expected to be a major part of this program, but it is appropriate to set aside a small fraction of the resources for this purpose.

COOPERATION

Several other programs have somewhat related goals. Efforts related to HGD of which we are aware include a project to study human genetic diversity in Europe, which has already been proposed by 14 laboratories coordinated by Alberto Piazza, and support for which has been requested from the European Economic Community. We are in close contact with the European effort. HLA workshops are held regularly, and the next one (Paris, 1995) will focus on population variability in HLA sequences. Close cooperation with this group in choosing populations to study and in the manner of collecting data should aid the efficiency of both programs.

POLICY ON OPENNESS, ETHICAL, AND LEGAL ISSUES

It is the policy of the Human Genome Diversity project that all materials and data that are not confidential should be available to all qualified investigators. Ethical, legal, and human-rights issues connected with the project will be discussed in the subsequent workshops.

TABLE 1. Basic data for sampling schemes to study human genome diversity in aboriginal populations: proportions of area, of total modern population, and of numbers of surviving languages per continent.

	Proportion (%) of World's Total			
Continent	Area	Modern Population Size	Number of Languages[a]	Number of Sampling Locations[b]
Africa	22	11	31	90 (22%)
America	31	13	13	80 (20%)
North	18			
South	13			
Asia	33	60	16	130 (33%)
Europe	8	15	1.5[c]	40 (10%)
Oceania	6	1	39	60 (15%)
Australia		.03	4	16
Pacific Islands		.5	20	24
New Guinea		.5	15	20
Totals	100	100	100	400

[a] Out of 4,799 languages (Ruhlen, 1991).

[b] The number of locations to be sampled per continent is suggested on the basis of a compromise between (a rounded-off average of) the three numbers given for each continent (area, modern population size, and number of languages).

[c] The political history of Europe has considerably reduced the number of aboriginal languages.

REPORT OF THE SECOND HUMAN GENOME DIVERSITY WORKSHOP
PENN STATE UNIVERSITY
OCTOBER 29-31, 1992

WRITTEN BY THE ORGANIZING COMMITTEE FOR THE SECOND WORKSHOP
(WEISS, CAVALLI-SFORZA, FELDMAN, KING)

INTRODUCTION

The Second Human Genome Diversity Workshop was successfully held at Penn State University from October 29-31, 1992. The Workshop was essentially organized around 7 groups, comprised of approximately 10 participants, representing the sampling issues in different regions of the world. These groups worked independently, using a common format provided by the organizers; this was adjusted as needed by the individual groups.

The Workshop began with a presentation of the mandate to the participants, and of the procedures to be followed during the workshop. Dr Feldman presented a summary of the results from the First Workshop (the report of which had previously been circulated to the participants). He and the other organizers also presented brief comments giving their perspective on the objectives of the Second Workshop.

Dr Julia Bodmer discussed the study of European genetic diversity, especially in the context of the HLA experience there, and of plans to extend such studies in the coming years. She also discussed surveys of world HLA laboratories in regard to resources related to Human Genome Diversity.

Dr Mark Weiss, who had been an important original administrative sponsor of the workshop series when he was Physical Athropology director at the US National Science Foundation, presented a few points to the group. These concerned the relevance of nonhuman primate studies for understanding how demographic processes, such as mate exchange between local groups, affected the local dispersion of genetic variation. Primate population geneticists have some relevant experience in interpreting variation at this local level, in particular, with various DNA 'fingerprinting' methods. This experience may be relevant to the Human Genome Diversity Project, in terms of practical and statistical issues.

The Workshop proceeded for most of its 3 days in individual regional groups. At the end, each group presented an overview of its written report, in a plenary session. At that time, various issues of general concern were raised. The current report includes edited, standardized versions of the 7 original regional reports. The organizers of this meeting, listed above have attempted to make the enclosed read consistently and clearly, but have attempted not to make any substantive changes to the reports themselves. Participants have been consulted to clarify details unclear in the written reports, and the regional group leaders have read the edited version to see that it is in fact representative.

Although it cannot be said that there was unanimity among the participants, and there was spirited discussion of various ways to approach the problems of a worldwide sample, and of the populations to be included, there was a remarkable degree of coherence and resolution of difference by the time the final recommendations were made.

Of the organizers of the series of workshops (i.e., those listed on the grant application which funded them), Dr Kidd was unable to participate in the Second Workshop due to illness. Drs Piazza, Paabo, and (Walter) Bodmer were unable to attend, and did not participate in the organization. Everyone, however, has been given a chance to read and

comment on this report. Feedback relevant to clarity and the proceedings has been incorporated.

The participants were limited in number by the cost constraints of the symposium grant. The Organizers did their best to invite an unquestionably distinguished worldwide panel of experts. However, funds were limited, and clearly numerous others who would have been qualified could not be invited. To augment the number of expert voices at the meetings, invitations were extended to persons who could support their own costs. A few such persons attended. Also, because they could contribute at little cost to the grant, several members of the Penn State Department of Anthropology also participated. A list of the active participants is included as an Appendix to this Report.

Before presenting the individual regional reports, we present several overall issues that were raised by the different regional groups. Generally, there was agreement among all the groups on these issues.

GENERAL ISSUES

1. The need to mix cell-line sampling with larger samples It was clear that participants felt that collecting only a limited number of samples for cell-line transformation restricted the scope of the Project in several ways. Many aspects of the Project could be improved by the addition of larger numbers of samples from the studied populations and/or from increasing the number of populations sampled. Increasing the scope by collecting samples for DNA extraction would greatly improve the overall value of the Project. This is consitent with the recommendations of the First (Statistical) Workshop.

2. The need to address important, but regionally different, disease-related problems It is important, from both an academic and a funding point of view, to note that many questions related to disease be considered in this Project, and that the sample relate to such issues. In this way, the Project can contribute to the well-being of many regions of the world in ways that go far beyond the cataloguing of variation and reconstruction of history.

Several disease-related issues were raised specifically by the regional groups. The study of malaria-related genetic variation has long been important in most of tropical Africa and Eurasia. There is a need for continued work on this problem, because new variants are still being discovered and many such variants may exist in populations not yet studied. Similarly, patterns of variation at the HLA loci suggest many aspects of genetic adaptation to disease antigens that are not yet understood. Genetic variation in the industrialized and/or non-tropical parts of the world has been studied extensively, but important problems in the rest of the world have hardly been examined.

There are special problems in regard to leukemia and AIDS viruses. One of the problems is the importance of careful collection of samples so that field and laboratory workers are protected. Equally important, individuals who are sampled and who are found to be HIV infected must have their confidentiality protected. Generally, no one who is sampled should be at risk of being discriminated against because of any aspect of his disease status (or genotype at any loci). Areas in which rates of infection are discovered to be high should be identified to relevant health authorities, and it would be important generally for the Project to involve health surveys or delivery as part of the studies in many parts of the world.

It is also now known that mutations related to susceptibility to specific diseases differ among world populations. This statement applies to severe childhood diseases like phenylketonuria and cystic fibrosis, as well as to chronic diseases such as hyperlipidemias, cancers, and perhaps chronic infectious diseases. Little is known about genetic variation in susceptibility to acute infectious or to parasitic diseases (other than malaria), and this Project represents an important opportunity to study such problems.

This can be helpful in two major ways. Specific disease-susceptibility problems can be studied as part of this Project by collecting environmental and family data in selected parts of the world (examples would be several viral diseases in Papua New Guinea, diabetes in Amerindians and various Pacific islands, various types of cancer in China and arctic populations, hepatitis in various parts of the world). Secondly, while most genetic diseases may be rare, even small samples from given areas can provide information on the chromosomal haplotypes which exist in those areas. This can be useful in searching for previously unknown mutations (i.e., that do not occur on frequently studied populations such as those of Europe and Japan).

Genetic counseling will not soon be available to many parts of the developing world, but that day will come sometime, and a large fraction of the world's urban populations (e.g., in Africa, Asia, India) have not yet been studied in this way, but could benefit in the near future as genetic diagnostic methods become less expensive and more widely distributed.

3. The need to see that sampled groups privacy and well-being is not harmed by this study

It is standard practice by careful anthropologists not to reveal anything about populations they study that can be used to harm the populations themselves. It is not always possible to prevent exploitation of remote or exotic populations by the outside world, but it is important to do whatever we can to see that the Project does no harm. Indeed, from the beginning it has been hoped that the Project may bring to public attention various issues in which populations are being seriously harmed by the outside world. Such issues are regularly raised by human rights, ecological, and other groups.

In this Project, it is important that the sampled populations understand as fully as practical the implications of the study, and its purposes. Fully informed access to the population, approved by the government of the nation in which they reside, must be provided by the investigator wishing to be sponsored by the Project. The population itself must demonstrably be provided a full level of informed consent. Religious or other cultural concerns must be protected.

This difficult subject will be considered at a future Workshop, and the results will be at the core of Project proposals for funding.

4. The need to have the actual sampling done as part of full and legitimate anthropological (linguistic, health, social, etc.) studies of the group, by investigators trusted by and familiar to the studied population

5. The nature of this report as advisory, but not fixed or proscriptive.

The Organizers asked the participants to make specific recommendations of populations that should be sampled, because it was felt that a Report that is too general will be difficult to use as the basis for a successful application for Project funds. One reason is that a specific proposal clearly shows the issues and the feasibility of the proposed worldwide study. Another is that a specific agenda is needed in order to draw up a requested budget.

It is clear, however, that it will be impossible to sample exactly all the suggested populations on any list drawn up by a general workshop in advance. Some suggested groups may not be accessible. It may be that no qualified anthropologist has access to the group, or that permission to sample it may not be available. At the same time, other investigators may wish to sample different groups, not named in this Report, but that are comparable in language, size, degree of isolation or admixture, or relevance to questions of importance to human diversity. Some regions were not adequately represented by the organizers--and for others, different investigators may differ on the Recommendations. In addition, some populations may already have been sampled for other reasons by the time the Project is implemented, and this should free up funds for alternative or additional sampling.

The mechanism for sampling that is finally established for the Project must allow for additions or subtractions, or alterations of priority, as knowledge base grows, and to be consistent with practical constraints. This Report is to be viewed as a 'living' document, that identifies major issues and some detailed specific recommendations, but that can be changed to meet the needs and practical limits of the Project.

At the same time, it is important that actual sampling NOT be done in a way unrelated to this Report or its considerations. The constraint to use this document as a framework for sampling guarantees that the Project will remain an open one, that is properly determined by sound anthropological expertise. There must be no suspicion that a small number of investigators will use the funding base of this Project to achieve their own private objectives. Anthropologists, linguists, and other qualified persons (including persons concerned with ethical issues) must be included in review panels that judge applications for funding specific populations, to ensure this openness and relevance to the issues raised in the Report.

6. The need for detailed input from specialists in areas not represented by participants at the Second Workshop

The need to continue to improve the sampling frame as knowledge accumulates, and as input from other investigators is received, has been stated. However, it is important to state clearly that the participants at the Workshop were aware that they only represented a very limited set of views on this subject. Although the participants were among the world's most prominent scholars in their respective areas, there are hundreds of other qualified investigators who could not attend (for practical and budgetary reasons, among others). Finally, many areas were not represented at the meeting. These include the region including Pakistan, the Caribbean, Australia, the Arab speaking parts of the world, West Africa, and others.

The participants did not wish to represent themselves as having expertise in these areas. The Report attempts to identify major issues in these regions, but it will be important that any final Project will include a representative series of samples as specified by help provided by experts in the regions.

Generally, the Organizers will attempt to continue to publicize the need for assistance and collaboration as widely as possible in world anthropological journals and the like. Recommendations will be changed as such information is received. Finally, wide dissemination of the availability of funding for the Project will be made so that qualified investigators, with access to the recommended populations can be supported to collect the data.

The intention of the Organizers is that the actual data collection will be done not by them, but by investigators (anthropologists, health workers, and so on) who have regular working relationships with the populations. The organizers role will be to make the availability of Project widely known so that such persons will apply for funding and will collect the data.

8. The need for joint sponsorship of the Project by many nations and funders

It is clear that US funding organizations cannot and will not be able to provide the entire amount of funding for this Project. The Project cannot succeed unless it is truly a worldwide effort. Those nations that have funding to provide to scientific projects must help to support this Project. Agencies such as the United Nations, the science or health ministries of wealthy or industrial nations worldwide must be involved. This will not only make it possible to pay the costs of such a project, but will spread the control of the Project around the world and will help to guarantee that the legitimate interests of the different regions of the world are represented fully.

9. The need for the Project to make permanent, rather than temporary technology transfer to laboratories in areas of the world that will be involved in sample processing but that do not have such capability currently.

The organizers have made it clear from the original application for funding, that an objective of the Human Genome Diversity Project would be to help establish or upgrade laboratories in many parts of the world, so that those laboratories could do modern genetics work and could participate in the Project. An immediate objective will be to enable laboratories near to Project field collection sites. Many laboratories around the world may be able, or nearly able to do cell transformation or other sample preservation procedures. Those that need it, should be provided with the equipment and the training for this Project. One series of labs that would be optimal for this would be regional HLA laboratories. The Project intends to establish a close collaborative relationship with the HLA establishment worldwide for this purpose.

In addition, it was pointed out clearly that it will not be sufficient merely to train and support local laboratories solely for a through-put processing role in the Project. The Project cannot be a worldwide success, of benefit to our species as a whole, unless permanent resources are established in all its different regions. The suggestion was made that when the Project is planned, it will be important to consider how regional laboratories, established or funded by the Project, can retain control of samples specific to their region, to serve distribution, data banking, and analytic functions. One possibility is that such laboratories retain stocks of all cell lines processed by them, as well as retaining the distribution function for all DNA samples that they process. If the latter is limited in amount, regional studies would have to be done by investigators working with the regional labs. A condition for establishing such labs, however, would have to be that they cooperate on an open basis with investigators interested in their region.

10. The need to develop a standardized "questionnaire" for ascertaining demographic, ecologic, and other information about individuals and populations included in the Project sample.

The next Workshop, or some other effort, should develop a standard set of information that would always be collected in association with studies funded by the Project. Other studies around the world that might contribute to the Project, even if not funded directly by it, should be asked to collect and provide the same kind of ancillary data. Age, sex, social group affiliation, parental and grandparental birthplace, genealogical relationships, language spoken, geographic location and ecotype, climate, and other variables should be collected. Dr Baer submitted some suggested formats, which are retained by the Organizers for future use; a copy is available to any interested party who requests it.

11. Sensitivities must be recognized to the maximum possible extent.

Many issues of sensitivity to various points of view need to be considered. For example, the establishment of permanent cell lines needs to be explained in terms that are understandable, but that do not mislead subjects in any population. English terms such as 'immortalization' of cell lines can be badly misunderstood, and have been avoided in this Report. Similarly, there is no fully acceptable way to refer to populations that are in danger of physical extinction or of disruption as integral genetic units (gene pools); some existing terms such as 'endangered' populations can have various connotations.Many populations around the world, especially isolates living traditional lifestyles, will soon disappear as independent units, because of disease, economic or physical deprivation, genetic admixture or cultural assimilation. In this Report, we refer to such groups as "Isolates of Historic Interest" (IHI's), because they represent groups that should be sampled before they disappear as integral units so that their role in human history can be preserved.

The Organizers have attempted to use terminology in this Report that is as sensitive as possible in these regards. Undoubtedly, errors have been made. There is no attempt to

make judgments about the inherent values of any people or populations, cultures, languages, and the like. The ethics review mechanisms of the Project will be designed to protect, to the extent that it is possible, against any such insensitivities.

==========
THINGS ADDED FROM OTHER FILES AND NEEDING TO BE FOLDED IN
==========

The elucidation of genetic variation in underserved populations, including most African populations, will provide important baseline data which will be of value in studying these genetic diseases. The markers known to be linked to disease loci have been almost exclusively defined in Caucasoid peoples and they may not, therefore, be informative in the peoples of Africa.

The role of genetic factors in conferring on the individual protection against infectious (including parasitic) diseases is likely to be best understood by studying individuals living in ecological areas where these diseases are still prevalent, i.e. in regions of developing countries. This has been strikingly true for the malaria-protective traits, like sickle cell haemoglobin, the thalassaemias and G6PD deficiency, and may well be the case for many other genes.

The study of the human genome, including elucidating its diversity, should not detract, in any way, from the need to address the health problems of the Third World, the bulk of which could be solved by the wide-scale application of knowledge already available; what is needed is the will to do so and the committment of adequate resources.

HUMAN GENOME DIVERSITY WORKSHOP ON ETHICAL AND HUMAN RIGHTS ISSUES
FEBRUARY 17, 1993

This document contains a schedule for the workshop, a non-exclusive list of issues that may be discussed, and a listing of the workshop participants.

SCHEDULE

Within each roundtable discussion, the participants are listed in the order in which they are currently expected to speak, although those orders are subject to change. Each participant should take, at most, 10 minutes for opening comments. Most of the time during each roundtable should be used for discussion of the issues among the participants and the audience.

9:15 Opening -- Prof. Hank Greely, Stanford Law School (moderator)
Summary of HGD-- Dr. Marcus Feldman, Stanford

9:30 Roundtable Discussion of Issues Raised Issues Raised by the Possible Commercial Value of the Project's Samples, Data, or Findings
- Dr. Val Giddings, USDA
- Dr. Walter Reid, World Resources Institute
- Dr. Jason Clay, Cultural Survival Inc.

10:50 Break

11:00 Roundtable Discussion of Issues Raised by the Sample Collection Process
- Dr. Ken Weiss, Pennsylvania State University
- Dr. Gary Ellis, NIH Office of Protection from Research Risks
- Dr. Rachelle Hollander, NSF
- Dr. Joan Porter, NIH Office of Protection from Research Risks

12:30 Lunch

1:30 Roundtable Discussion of the Possible Misuse of Project Data or Results
- Dr. Diane Paul, University of Massachusetts at Boston
- Dr. William Schneider, Indiana University-Purdue University at Indianapolis
- Dr. Robert Murray, Howard University
- Dr. Eric Juengst, NIH

3:00 Break

3:10 General Discussion of Other Issues, Including Organization
Participants and audience

4:00 Scheduled Conclusion

ISSUES

The following pages contain a list of some of the issues that may be discussed at the workshop. This list, in somewhat earlier form, was circulated to the participants. The list is not intended to preclude discussion of other issues, although issues that do not relate to one of the three roundtables may be referred, at the moderator's discretion, to the afternoon's general discussion period.

I. Issues Raised by the Possible Commercial Value of the Project's Samples, Data, or Findings

This area includes a number of issues related to property, financial arrangements, and technology transfer. These issues are ethical and political. I believe we should deal with them first, because they have some implications for the second set of issues

A. Payment Arrangements in the Event the Project Has Commercial Value

1. What is the potential for the commercial use of the project's data or results?
2. Is it ethically permissible to pay, even in a contingent manner, for a human's DNA, or does that impermissibly "commodify" that person and demean his or her humanity?
3. Will some governments demand a system of payment for any commercial value from the gene samples collected by their citizens?
4. If so, how should the project respond?
5. Should the project volunteer a uniform payment system for all countries, regardless of whether they request one?
6. If the project were to agree to a kind of "royalty" arrangement, how should it be handled?
   a. Should the payment be based made on behalf of all the populations sampled or just to those whose samples contributed to the commercial value?
   b. Should the payments be made to the governments of the relevant populations, or to the populations directly?
   c. If payments were to be made to populations directly, how should that be accomplished?

B. Payment Arrangements and the Sampling Process

1. Should the payment arrangements reached, if any, or the absence of such arrangements, be part of the informed consent to individuals whose DNA is sampled?
2. Should those arrangements be subject to renegotiation by each sampled population?
3. Should the informed consent ask the sampled individuals expressly to release any property rights they may have in the samples or in the use of their DNA?
4. Is it appropriate to provide any gifts or any payment to people who agree to participate?

II. Issues Raised by the Sample Collection Process

This area includes a number of issues revolving around how the samples are taken and how information obtained from the samples is used with respect to those people sampled.

A. Who Should Be Sampled

1. Should the project ever sample children?
2. If the project will sample children, what special protections should apply?
3. How should adulthood be defined for these purposes -- by American standards, by those of the sampled population, or by both?
4. Are there other classes of potential donors who deserve special consideration?

B. Informed Consent

1. Should the project require a uniform manner of informed consent among all sampled populations?
2. If variable methods of informed consent (written, oral, etc.) are permissible, considerations should be relevant in deciding the method to use with a particular population?
3. Should the project ever accept the informed consent of one person as binding other persons (local leader and population, husbands and wives, parents and children, etc.)?
4. Should the project ever permit someone other than the person to be sampled (a local leader, spouse, parent, etc.) to veto that person's participation?

C. Privacy

1. What level of privacy should the project seek to ensure in the sampling process, with respect to the informed consent and the actual decision to donate or not?
2. Should the samples, once taken and processed, continue to be identified with individual donors?

D. Handling the Samples

1. Should the blood samples be screened for any disease organisms?
2. If so, is the screening primarily to protect people who may come in contact with the samples in the future or to help the individuals who gave the samples?
3. If any samples are to be screened for any disease organisms, who will decide what screening will be done and what considerations should that decision-maker apply?
4. If a sample is screened and is found to be contaminated with a pathogen, what should be done with the sample?
5. If a contaminated sample is found, should the person who donated the sample be told? Under all circumstances? Under some circumstances? Under no circumstances?
6. Does the project have an obligation to provide medical assistance to individual donors who are found, through the project's work, to have medical conditions?

III. Issues Raised by the Possible Misuse of Project Data or Results

This area involves mainly the possibility that the project's existence, data, or findings will be used to claim that human populations are more different than the science supports and that these differences have social or political implications. The concern about such misuse stems from two reasonable fears. First, some people may try to use the project to support racist or nationalist claims of the "genetic superiority" of particular populations. Second, well-intentioned laymen may misconstrue some of the project's findings to similar ends.

1. Based in part on the history of cultural use of genetic information, how likely is such misuse of the project?
2. Are the concerns about such misuse sufficiently great that the project should not be undertaken?
3. If this research is not undertaken through the HGD project, will it be undertaken by others in any event?
4. Should access to samples or data from the project be widely available, or should they be limited to researchers with appropriate academic credentials?
5. Should access to samples or data from the project be limited on political grounds?
6. Does the project have an ethical obligation to try to educate the public about the meaning of its results? If so, how should that obligation be fulfilled?

IV. Other Issues

This area is a catch-all for other ideas. We will discuss these are the end of the workshop to the extent we have time (and energy) left. This list currently contains only two areas of issues, but more are expected to surface during the day's discussion.

A. Technology Transfer

1. How should the project respond to a request from a host country for technology transfer?
2. Even without such a request, should the project feel an ethical obligation to attempt to transfer relevant technologies to the countries or the populations in which it samples?
3. If the project should commit to technology transfer, how can such transfer best be accomplished?

B. Organization

1. Should the project provide for some further consideration of ethical and human rights issues that it raises?
2. How should that consideration be achieved -- through an ethics advisory committee? through more workshops? through some other method?

PARTICIPANTS

The following people are participating in the workshop discussions. The extremely short biographical material has not been cleared with all of them; I regret any errors.

Dr. Jason Clay
Cultural Survival, Inc., Boston, Massachusetts. Dr. Clay has a Ph.D. in anthropology. From 1980 to 1989, he was director of research for Cultural Survival, Inc., a nonprofit human rights organization working with indigenous peoples around the world. In 1989 he founded Cultural Survival Enterprises, which he directs. In that capacity he works to develop economic resources for indigenous peoples.

Dr. Gary Ellis
Office of Protection from Research Risks, National Institutes of Health. Dr. Ellis has a Ph.D. in Biological Sciences. He recently assumed the post of director of the NIH Office of Protection from Research Risks after several years as director of the Division of Health Promotion and Disease Prevention at the Institute of Medicine.

Dr. Val Giddings
Biotechnology, Biologics, and Environmental Protection Division, Animal and Plant Health Inspection Service, U.S. Department of Agriculture. Dr. Giddings has a Ph.D. in genetics. He is Chief of Science and Policy Coordination for the branch of the USDA that regulates the use of biotechnology products in the environment. Dr. Giddings was a member of the United States delegation that negotiated the Biodiversity Treaty, and he was particularly involved in the Treaty's technology transfer and royalty provisions.

Dr. Marcus Feldman
Stanford University, Department of Biological Sciences. Dr. Feldman has a Ph.D. in mathematical biology. A leading population geneticist, he is a member of the HGD Committee.

Prof. Hank Greely(chair)
Stanford Law School. I am a law professor specializing in health policy. I have written on aspects of the Human Genome Project and on some aspects of bioethics.

Dr. Rachelle Hollander
National Science Foundation. Dr. Hollander has a Ph.D. in philosophy. She is director of the NSF Ethics and Values Studies Program.

Dr. Eric Juengst
National Center for Human Genome Research, National Institutes of Health. Dr. Juengst has a Ph.D. in philosophy. He is the principal ethicist with the NCHGR and directs its program on the ethical, legal, and social implications of the Human Genome project.

Dr. Robert Murray
Howard University, College of Medicine, Department of Genetics and Human Genetics (Chair). Dr. Murray is an M.D. He has written on genetic diversity among peoples of African origin and has long been interested in ethical issues concerning genetics.

Dr. Diane Paul
University of Massachusetts at Boston, Department of Political Science. Dr. Paul has a Ph.D. in political science and is the director of her university's Program in Science, Technology, and Values. She has written on the historical use of human diversity in the debates over eugenics and other social issues.

Dr. Joan Porter
Office of Protection from Research Risks, National Institutes of Health. Dr. Porter has a doctorate in public administration. She has served as Special Assistant to the Director in the Office of Protection from Research Risks for the past 11 years and has written on protection of human research subjects.

Dr. Walter Reid
Vice President for Program, World Resources Institute. Dr. Reid has a Ph.D. in zoology with a specialization in population and community ecology. His work has focused on world biodiversity.

Dr. William Schneider
Indiana University-Purdue University at Indianapolis, Department of History (Chair). Dr. Schneider has a Ph.D. in history. He has written on the historical uses of ethnic diversity in blood types in debates over eugenics and other social issues.

Dr. Kenneth Weiss
Pennsylvania State University, Department of Anthropology (chair). Dr. Weiss has a Ph.D. in anthropology and works in biological anthropology, epidemiology, and genetics, with a special interest in native American populations. Dr. Weiss is a member of the Human Genome Diversity committee.

HUMAN GENOME DIVERSITY PROJECT
SUMMARY OF PLANNING WORKSHOP 3(B):
ETHICAL AND HUMAN RIGHTS IMPLICATIONS

TABLE OF CONTENTS

HUMAN GENOME DIVERSITY PROJECT
SUMMARY OF PLANNING WORKSHOP 3(B): ETHICAL AND HUMAN RIGHTS IMPLICATIONS

The third planning workshop of the Human Genome Diversity Project was held on the campus of the U.S. National Institutes of Health in Bethesda, Maryland, from February 16 through February 18, 1993. The second day of the workshop was devoted to an exploration of the ethical and human rights implications of the Project. This open meeting centered on three roundtables, involving 12 invited participants, and the resulting discussions among all those present. Attendees and their affiliations are listed in the attached Appendix A. The discussion was guided by a schedule and list of possible issues, distributed to all present and attached as Appendix B. The meeting was organized and chaired by Professor Henry T. Greely, the principal author of this summary.

This is a relatively complete, and thus lengthy, summary of the comments at the meeting. The beginning of the summary sets out as conclusions some issues on which there appeared to be widespread agreement, but those conclusions not intended to serve as a set of detailed recommendations. The meeting organizer will distribute his recommendations in a separate memorandum; recommendations from others who attended the meeting are welcome and will be distributed by the meeting organizer to the participants and to the Project committee.

Conclusions

These 11 conclusions represent the views of most, if not all, of the individuals invited to participate in the meeting. These are the minimum conclusions to be drawn from the workshop and should not preclude any participants from setting forth additional views.

1. The ethical concerns raised by this Project are both real and significant. Although no participant believes those concerns mean that the Project should not go forward, the participants think careful attention must be paid in designing and executing the Project in order to minimize the risks of harm.

2. Consideration of ethical issues needs to be integrated into the Project's decisionmaking, both in the planning stages and, on a continuing basis, during the life of the Project.

3. The Project should be designed and executed with help from the populations to be sampled as far as is feasible, although the participants realize that there will often be enormous logistical barriers to such assistance.

4. If any funding is obtained from federal agencies, American law concerning informed consent must be followed, but applying it in a manner that provides useful information to the populations to be sampled will sometimes prove difficult. No one method of providing informed consent will be appropriate for every

population. This kind of sampling has been done in the past in projects financed by federal agencies, so it seems likely that these difficulties can be surmounted. The Project should collect samples of informed consent protocols for previously approved research. It may want to create several model protocols or to review itself the informed consent protocols of researchers who seek to collect samples for it. The structure for central review, in any, of informed consent protocols requires thought.

5. The Project should consider beginning its sampling with populations that raise the fewest ethical and political problems. It should give special consideration to beginning with populations in countries that sponsor the Project. Experience with those populations, and a record of success in dealing with them, may be very helpful in sampling more vulnerable groups.

6. It is not clear whether any governments of populations to be sampled will seek payments in return for the collection of DNA samples within their borders. The Project should consider what approach it would take to such requests. Because of the many sensitive issues involved in patenting genes, human or otherwise, the Project may want to agree that no genes may be patented based on samples collected through the project. It may also want to agree that populations or countries will receive some form of payment in the unlooked for event that samples collected for the Project lead to products of commercial value. The form of such payments requires further thought.

7. Some people will almost certainly attempt to misuse the Project's data and findings to support of racist or nationalist ends or what Dr. Juengst has termed "demic discrimination." Whether that misuse would have any significant consequences is unclear, but the participants believe the Project has a duty to try to minimize the effects of such misuse.

8. A program of public education would probably be a useful, and perhaps an essential, element in efforts to limit the misuse of Project data and findings.

9. As part of such an education and information program, it is important that the Project define itself, its goals, and its limitations, to the public, rather than allow it to be defined by others.

10. The Project should ensure that it is informed about the uses to which its samples and data are being put and the conclusions that are being drawn from them. It may want to be able to respond quickly to published work or press inquiries in order to ensure that claims made on the basis of the Project's work are put into their proper scientific context.

11. There is no reason to believe that the ethical concerns raised by this Project are insurmountable. The participants believe that, with appropriate safeguards, the Project should proceed. The Project should, however, evaluate throughout its life the ethical consequences of its work. If unexpectedly strong and negative effects appear, the Project should be willing, if necessary, to bring itself to an early end. The participants believe that the Project should provide

data that will be invaluable in answering important questions in a wide variety of fields, including genetics, anthropology, history, linguistics, and others that cannot yet be guessed. The value of this research and the urgency caused by the continuing disappearance of isolated human populations makes the ethical concerns all the more important. If the Project does not proceed carefully and properly, it could spoil the last good opportunity to obtain some of this data.

Workshop Summary

This workshop session convened at 9:30 a.m., February 17. Prof. Greely welcomed those in attendance, introduced the roundtable participants and the present members of the Human Genome Diversity Project committee, and made a variety of administrative announcements. He laid out his goal for the meeting -- not the resolution of many, or perhaps any, of the ethical and human rights questions raised by the Project, but instead a narrowing and focusing of the issues involved. Dr. Marcus Feldman, a population biologist at Stanford and a member of the Human Genome Diversity Committee, then briefly explained the Project and the described the planning workshops held thus far.

First Roundtable Discussion – Collection Issues

The first roundtable discussion covered the issues involved in collecting blood, DNA samples, and information from populations of interest to the Project. Prof. Greely noted some of those issues briefly, then turned the meeting over to the four speakers on that issue: Dr. Joan Porter, Dr. Kenneth Weiss, Dr. Rachelle Hollander, and Dr. Gary Ellis.

Dr. Joan Porter

Dr. Porter has long been affiliated with the Office of Protection from Research Risks (OPRR) of the Department of Health and Human Services (HHS). The OPRR implements HHS regulations that protect human subjects. She began by reviewing the historical development of ethical strictures on human experimentation, starting with the Nuremberg Code and the Declaration of Helsinki and drawing special attention to the Council for International Organizations for Medical Sciences (CIOMS), which also provides guidance on ethical issues.

In the United States, federal regulatory law is based on the Belmont Report, prepared by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979. That Report focuses on three principles:

- Respect for Persons:
  This principle requires researchers to seek informed consent from subjects and to provide special protections for people with diminished ability to consent.

- Beneficence, which includes the principle of non-maleficence:

Non-maleficence means doing no harm; beneficence means doing good. This principle is realized largely through careful prospective and ongoing risk benefit analysis by an independent committee.

- Justice:
  This principle means many things, including treating people fairly. Equitable selection of subjects is one way to help ensure justice.

Most of the principles focus on individuals, but justice also concerns populations. The Project lends itself to a "macro-ethical" framework, focusing on communities rather than only on individuals. In considering macro-ethical issues, she has found particularly useful the International Guidelines for the Ethical Review of Epidemiological Studies, prepared by CIOMS in 1991, although there are some differences between those guidelines and American law, particularly on informed consent. She believes the Project raises several particularly interesting macro-ethical questions:

Will the studies be able to maximize benefits to the community by communicating the results?

Will there be a way to minimize harm by avoiding stigmatization or prejudice or loss of self-esteem or economic loss?

Can harmful publicity be avoided?

Can some kind of confidentiality be retained for groups?

Dr. Porter then set out ten other areas of concern that she believed this group and the Project's organizers must address.

1. What are the real risks and benefits to the individuals and communities? Physical risks are probably minimal, but it is important to anticipate any social risks.
2. Will genetic information be individually identified?
3. Will the sampling process lead to any information of use to the individuals sampled? For example, if a subject's blood had a genetic marker for breast cancer, could (and should) any useful information be conveyed to that individual? What would happen if everyone in the sample had the marker?
4. Will information about disease susceptibility have any possibly subtle effects on employment, immigration, or access to health care for the subjects?
5. Will samples be taken from families and, if so, what concerns are there about pressures to participate, recruitment issues, and problems of confidentiality that have been recognized in large pedigree studies?
6. Will any inferences be made about the relationships between non-human primates and certain groups?
7. Are the groups of most interest to the Project among those most

disenfranchised in their societies and, if so, what are the implications of this status? Who can give permission for such groups?

8. Will DNA from non-living persons be collected? How can community sensitivities about that kind of collection be taken into consideration?
9. What kind of local independent assessment is important? If there are 200 sample areas, will there be 200 institutional review boards (IRBs)? HHS generally prefers local IRBs but perhaps there is an equally effective and more efficient method to do this.
10. How can we ensure appropriate recognition of collaborators, particularly those from other countries?

The Project presents an unprecedented opportunity to advance some areas of human knowledge, making it crucial that we think through the ethical and legal issues carefully.

Dr. Kenneth Weiss

Dr. Weiss is a professor of anthropology at Pennsylvania State University and a member of the Human Genome Diversity Committee. He started by noting that many of the populations to be sampled will be in urban areas. For them, many of the standard answers to ethical issues will apply. Dr. Weiss focused his presentation on the special problems of sampling populations that are living in isolated circumstances and under relatively traditional conditions. He stressed that, because of differing conditions, no one approach would work for every population.

The first challenge will be to explain the Project's goal to these populations. Explaining that the Project wants blood isn't difficult (though it may lead to rejection), but explaining the goals can be harder. For example, he noted that describing one of the goals of the research as discovering information about the origin of a population would be meaningless, or perhaps insulting, to many populations that have their own deeply held beliefs about their origins. Such an explanation could do cultural harm as well as jeopardize the efforts to get permission. In general, explanations must to be made by people who are very familiar with the populations, whether they call themselves anthropologists or not.

All communications must occur in the local language and each must to be tested to make sure it is correctly understood. Although that requirement is not very different from how any kind of epidemiological survey is done, it is more challenging in uncommon languages.

Whether the identity of a studied population should be publicly revealed through the research is an interesting point. Anthropologists have long been deeply divided over the degree to which they should identify specific populations with which they work. This would have to be done differently in different contexts. Identifying isolated communities can sometimes harm those communities and sometimes work to their benefit. This is a real issue, although one without a blanket answer.

The disease questions are interesting, although breast cancer is a bad example for most of these populations. Something like HIV infection is more sensitive because it might have political ramifications. In some places, disclosure of infection could bring help and in other places, it could bring harmful attention. Someone familiar with the area and the population needs to decide how to deal with samples that show signs of disease before any collections are authorized. No one standard can apply everywhere, but wherever possible, people with current disease should be referred to the national health systems when possible. In many countries, referrals to a national health care system can make a difference in the health care provided to individual subjects, which may lead to at least a temporary quid pro quo. For a few populations, there will be no health care system to which they can be referred.

In some populations, payments are required for the research, either to local leaders or to individuals in the studied group. In some of those populations, payments will have to be made to a local leader because paying individuals directly rather than the group leader would completely disrupt the social system. A person who has experience with that group must decide these issues.

Coercion means different things in different situations. Bravado or lack of bravado, for example, may have a major effect on who volunteers. If coercion happens, it will be from within the population. Sometimes a person who works within the group will have to be persuasive in order to get consent. In many populations, getting a signed consent form will be impossible. People will not sign any form, no matter what it says, because of their intense suspicion that any written form will be used against them in land disputes.

"Immortalization" can be a very sensitive term and should be avoided when talking about the intended creation of cell-lines. (Someone suggested using "transformation," the standard European practice.) Whether to tell people what you intend to do, as a technical matter, is a difficult question. Translating the concepts will be very hard.

As to anonymous samples from dead people, the National Graves Protection and Repatriation Act in the United States provides rules for the return of skeletal remains. If you can identify an existing successor group to which the remains are related, that group's permission is necessary; if you cannot identify an existing group to which the remains are related, he believes they may be used without permission.

The use of fetal and placental tissue will, of course, be governed by local rules and morals. There is no blanket answer to whether you should be able to take samples from people who appear at isolated hospitals outside their home.

Getting permission is complicated and, again, differs in different areas. In working with Native American tribes in the United States, for example, you get permission from the tribal group and go around to the populations with representatives of that group. In some parts of the world, approval must also be sought from a government from outside the population to be studied. In many of

these populations, anthropologists or linguists are already working with the population and can facilitate obtaining permission. Before any collection activity would be funded, its leaders would have to document to the satisfaction of a reviewing group that they understand the permission process in that region and are complying with it. How the review group could check on that information is not clear, but it could not demand an identical process for every population. The Project must always work with people who are familiar with and trusted by the groups to be sampled.

Dr. Gary Ellis

Dr. Ellis has recently become director of the Office of Protection from Research Risk. His office regularly hears from researchers who say that the rules don't fit into their research. His reaction is "let's find a way." Informed consent is a process, not just a form, with the prospective research subject's ability to make a voluntary decision as the key requirement. The procedures should be designed to educate the subjects in language and terms they can understand. The document should be a teaching tool, not a legal document. Lay language, understandable to the potential subjects, must be used even though this is difficult in other cultures. It is important to have a written document so the subject can refer to it in the future.

U.S. federal government regulations lay out eight elements of informed consent.

1. The informed consent must state that the study involves research, must explain the purposes of the research and the expected duration of the subject's participation, and must describe the procedures to be followed with a special description of any procedures to be followed that are experimental.
2. The informed consent must describe any reasonably foreseeable risks or discomforts to the subject.
3. The informed consent must describe any reasonably foreseeable benefits to the subject or others.
4. The informed consent must disclose any appropriate alternative procedures or courses of treatment that may be advantageous to the subject.
5. The informed consent must state the extent that confidentiality of records identifying the subject will be maintained. If the Project found that someone had a disease and went back and referred them for treatment, obviously the samples would have to be linked to the donors, which may not be the best approach.
6. In research involving no more than minimal risk, the informed consent must explain whether any compensation or medical treatments would be available if injury occurs, what the compensation or medical treatments would be, and where further information about those matters can be obtained.
7. The informed consent must tell prospective subject whom to contact for answers to questions concerning the research, the research subject's rights, and who to contact about any research-related injuries. These

questions ordinarily cannot all be answered by the same person for reasons of limited knowledge or conflict of interest.

8. The informed consent must state that participation is voluntary and can be declined without any effect on eligibility to benefits and that participation can be discontinued at any time without any penalty or loss of benefits to which the subject is otherwise entitled. The Project needs to consider what discontinuation means in a context where there will be ongoing cell-lines and databases.

Dr. Rachelle Hollander

Dr. Hollander is Director of the Ethics and Values Studies program of the National Science Foundation. She began by noting that the rhetoric of science romanticizes or mythologizes the science. Research subjects, on the other hand, tend to de-mythologize science. Both the public and geneticists need to be educated about the Project. Genetics has a mixed record in its effects on people. The urgency of this Project needs to be moderated by the recognition that it is important to go slow enough for the public and the scientists to understand the consequences of the science.

The Human Genome Diversity Project was mentioned at the AAAS meeting and was the subject of some misunderstanding. For example, some wondered why we should study diversity when geneticists can create diversity. Others wondered why people were concerned about preserving DNA but not ways of life. There are concerns about evolution within humans because the public and scientists often equate "early" to "primitive" to "less valued".

The need for the involvement and approval of affected communities, from local level to the nation-state, is crucial. What counts as "owning", or "having access", or "benefit" will differ in different cultures. Does this Project involve questions of common human heritage? Will there be questions of ownership or disputes over property rights? How will that be managed? Property rights will make a difference in obtaining informed consent because it affects what you are seeking consent to do. Reciprocity is fundamental and so the Project will have to think about the interactions between the researchers, the sponsoring governments, the populations, and the nation-states that include the populations.

The process of answering these questions must empower individuals from affected populations. They should be involved in developing the Project as well as in allowing access to populations. Informed consent as a mark of respect is very important; respect is an idea shared by many cultures, but sometimes expressed in different ways. Does this consent make sense in terms of the community or person giving the consent? Is there reciprocity? What do *they* think is important? That kind of evaluation is important and it is important that it be ongoing, throughout the Project.

The group might want to devise ethical guidelines for the research, but do so with the participation of the groups involved. This will also lead to the development of ideas about what community payback might be appropriate. The process should also incorporate ongoing evaluation of those ethical guidelines.

The document summarizing the second workshop ends by saying that all materials and data not confidential should be available to all. What that means requires considerable thought.

Discussion

Prof. Greely noted the tension between centralized review for western standards and federal rules, on the one hand, and making allowances for local conditions and cultures. Devising a process that does both is not a trivial issue.

Dr. Paul said it is useful to note what we know about informed consent in American medicine. We know a lot about how it works and it doesn't work as a teaching tool. For American researchers and doctors, the phrase is "getting consent" and it is viewed as a defense against malpractice suits. Nancy Press's recent research on California alpha-fetoprotein consents in California showed that most of the women signing the form either didn't know that they had had the test or thought the test was some form of therapy for the fetus. The true significance of informed consent differs from the ideal even in the United States. (Prof. Greely noted that "consent" has become different kind of verb in American medicine -- rather than patients consenting to treatment, doctors "consent" the patients.)

Dr. Giddings seconded those remarks. He thinks the Project runs a risk of getting into a political morass. The Project will inevitably have to deal with local political structures and many people in those structures will not believe whatever the Project says because its representatives will be from the developed world. He hesitates to suggest getting involved with the United Nations bureaucracy, but it might be a good idea. The political problems are so great that the Project either must keep a very low profile or seek political cover, and a low profile is impossible. The Project might want to get involved with the World Health Organization. It might also want to divide the populations of interest into low, medium, and high political risk. It could sample the low risk populations first, as well as populations from the sponsoring nations first. Pursuing the Project poorly might poison the well forever for future scientists. Lots of people will be very suspicious about the Project. Even though those suspicions may be inappropriate, you have to deal with them.

Dr. Siniscalco pointed out that he has a lot of experience on these issues from his work with the European Commission. The study of human genome diversity has been done for a long time -- it constitutes the core of human genetics. Various groups have been doing this kind of research and have been interacting with local governments as well as funding organizations. The main effort should be to advertise the importance of the Project. He suggests something like the DNA Learning Center at Cold Spring Harbor in an effort to educate the public. He exhibited a copy of the Cold Spring Harbor exhibit in a museum in Sardinia and found that the public was extremely interested. He believes the Project also needs to emphasize the crucial importance of environment in its interaction with genes. The Project should not need to keep information about individual samples that identifies their donors, because individuals will need to be tested only if a medically important gene is found.

There are rules and regulations for this kind of research, but the Project should focus on inventing a system to uphold those standards. We have an enormous opportunity; but we need to explain the Project to people in order to gain their trust.

Dr. Reid pointed out that although this kind of work has been going on, this Project will face new obstacles. One reason is the scale of the Project. It will attract attention from people who wouldn't otherwise know that this kind of research is going on. The international political context will also be disturbing. The combination of the Rio discussions and the ethical dimensions raised by dealing with human genes will attract an enormous amount of attention. Education about the motivations for the work won't be enough; it comes down to a question of creating a process that makes this workable. He shares some of Dr. Giddings skepticism about the United Nations, but believes it deserves a very hard look because of the certifying role a U.N. connection might play. He also thinks it would be very valuable to involve the subjects of the research in the planning, although he recognizes the tremendous logistical problems. Finally, it will require tremendous anthropological expertise in the populations to be sampled; maybe that should be factored into the considerations for choosing populations. (Prof. Greely pointed out that the existence of such expertise had been one of the consideration the second workshop used for suggesting populations to be studied.)

Dr. Juengst agreed that the Project will have a high visibility and that will change the dynamics. That has happened with the Human Genome Project; mapping and sequencing studies had been going on for a long time, but a coordinated Project with capitalized name becomes a lightning rod. As a result, many of the rules for various studies are being reinvented, at least to provide coordination within the Genome Project. It may be the case that all is needed here is to codify existing rules for pursuing this kind of work.

Dr. Cavalli-Sforza stated that we may be making the problems seem more difficult than they actually are. He has collected samples from some of the most isolated and difficult locations in the world. You must have the permission of the local authorities, or they may put you in jail, and you must have the permission of the local people, or they won't participate. The permission issues are therefore self-policing. The important issue is the information. Right now, we cannot name any immediate benefits, except possibly with respect to vaccination, which is itself controversial among many people. It is very difficult to explain the science. A traveling Cold Spring Harbor show is a good idea and may work in the cities, but won't work in the villages where most of the sampling will be done. It may be very difficult to make a truly informed consent as the American regulations seek, but we should go as far as we can.

He has been trying to get the involvement of UNESCO. UNESCO is preferable to WHO, because the Project has no immediate medical purpose. He had a conversation with the Director General of UNESCO, who was very supportive. He agreed with Dr. Weiss that it was impossible to have general rules; we need to stick to the general rule, "do no harm."

Dr. Schneider was reminded of the fact that the largest samples of blood for blood grouping came from troops on the Salonika front during the First World War. Dr. Hirschfeld, who, with his wife, collected the blood, said in his memoirs that how he collected blood depended on the nationality of the troops. The British troops would do it if they were told it was "for science"; the African troops would do it if told that they could get out of military service; and the French troops would do it enthusiastically if told it would help them know who they could make love to with impunity. Consent issues go back a long way; scientists are pragmatic and do what they have to do to get the material. He suggested that the proper model might not be medical studies, but traditional anthropological studies because this is not, basically, medical research. Are consent forms used by anthropologists? Another important question is what will be done with the drawn blood. Will this be used as a storehouse for future work, or just to be analyzed once? If used as in a long run, you won't be sure what the uses will be.

Dr. Weiss said that anthropologists don't use forms, but do get the permission of the groups they study. The Human Genome Diversity committee has known from the beginning that local government approval will be essential and has thought about technology transfer within that context. He thinks we are flattering ourselves to think that the Project will have a high profile. Most of the time, the permission is local, from government officials who want to know that you are there and that you are not going to stir up trouble. Anthropologists get approvals in most parts of the world without much trouble, although there are occasional, usually temporary, exclusions. Some groups like being studied. Groups are very different; confidentiality is meaningless in some contexts. In some groups, doing something in someone's house, away from the group's sight, would be deeply offensive. You more often have to include people you don't want rather than exclude people you do want. To the Committee, from the point of view of funding the Project, it is very important that the Project promises not to reveal individual's names to outsiders.

Dr. Ellis asked who is sponsoring the Project; the response was that we don't know yet. The workshops are planning workshops to put together the framework for writing proposals. Prof. Greely pointed out that if the Project wants U.S. funding, the U.S. regulations will have to be followed for legal reasons, as well as for ethical concerns. He agreed with Dr. Weiss that the application of the principles would necessarily differ from culture to culture. He agreed with Dr. Siniscalco that this kind of work has been going on, so protocols have been approved. Finally, he agreed with Dr. Juengst that, perhaps, "codification" of protocols that have already been used would be sufficient to deal with these issues.

Dr. Ellis added some political advice, saying that the Project should define itself to the public in simple terms first, before someone else does it. It would be a shame, for example, if someone said this Project was intended to find the genetic basis for violence around the world and the Washington Post reported that.

Dr. Giddings said there is a substantial amount of evidence to indicate that this Project is being prosecuted by some guiding intelligence. There will be no way to go forward without a lot of local involvement -- there may be no feasible alternative to funding from something other than a multinational consortium. For political and financial success, this will have to involve multiple governments or international agencies.

Dr. Bodmer pointed out that Europe is involved. The European Community has provided some funding and is interested in providing more. European populations are to be sampled as part of the Project, with at least some European funding. There is interest in making genetic diversity part of the next European funding framework for the biological sciences.

Dr. Schneider asked for clarification about the boundaries of "Europe" for work funded by the European Community. Dr. Bodmer replied that the European side of the Project was able to act outside the boundaries of the European Community, including doing work in Eastern Europe and the former Soviet Union. Dr. Siniscalco noted that the European group is exploring the possibility of Soros Foundation grants for working in Eastern Europe and the former Soviet Union.

Dr. Piazza asked whether it would change things if we collected mouthwash rather than blood. Prof. Greely said it would probably change only the details of the consent, but wouldn't make major changes in the process.

Dr. Feldman noted that throughout the Committee's endeavors, it has had close contact with both China and Japan. The vice president of the Chinese Academy of Sciences was present yesterday and will be tomorrow.

Second Roundtable Discussion -- Payment and Property Issues

The second roundtable discussion explored issues concerning the possible commercial value of the Project's samples, cell-lines, or findings. Prof. Greely explained that he had included these issues in the workshop because he believed the recent Biodiversity Treaty would lead some countries with populations of the interest to the Project to seek royalties or other payment for sampling the genes of populations within their borders. He then turned the meeting over to the two speakers on that issue: Dr. Val Giddings and Dr. Walter Reid. (The third scheduled speaker, Dr. Jason Clay from Cultural Survival, Inc., was not able to attend because bad weather grounded his flights from Boston. Dr. Clay is receiving copies of this summary and other correspondence concerning the Workshop and has been invited to submit a statement of his views for distribution to the group.)

Dr. Val Giddings

Dr. Giddings, whose trainin was in human genetics, participated in the United States government's negotiating team for the Biodiversity Convention. He read excerpts from the Convention's provisions concerning technology transfer and payment. *[The tape did not record this part of Dr. Giddings's comments and*

our notes are not very detailed. I hope he will let me know exactly what passages he read.]

He said that the Convention and the negotiations from which it emerged will have important implications for this Project. It is a new era for genetic resources. The Food and Agriculture Organization's principle of "free access" to genetic resources is not going to survive. Movements of genetic resources across international boundaries in the future will be subject to a variety of constraints, particularly if there is any reasonable inference that the resources could result in commercial products. He urged that the Project must anticipate those constraints and build in some kind of agreement for royalty partitioning or some other provision for distribution of fair and reasonable benefits. Otherwise, it won't be able to get permission to take samples.

Dr. Walter Reid

Dr. Reid is associated with the World Resources Institute and has a special interest in issues of global biodiversity. At Rio, Dr. Reid participated with other organizations in developing a global biodiversity conservation strategy. That process was not as polarized as the Convention negotiations, but it also reached the conclusion that the era of treating biodiversity as a common heritage or open access resource is over. Arguably, the distribution of benefits from the old regime was not equitable and developing countries see opportunities to get more benefits in the future, through both financial arrangements and technology transfer.

In the Convention negotiations, intellectual property rights and technology transfer were most contentious. This Project will not have a profile as high as the Convention's, but he suspect the Project could not go into countries to get permission for sampling without encountering officials who were familiar with the Convention. Those officials might draw on the Convention's terms to deal with the Project's request, particularly in countries that signed the Convention.

He drew a parallel to the existing seed banks, which were collected under the common heritage regime. For example, there is a bank of 60,00 or 70,000 varieties of rice in the Philippines, with copies in the United States. The Convention negotiators decided they couldn't deal with the existing collections and so excluded that material collected in trust for humanity. How to deal with those materials remains a subject of discussion. Some seed banks are talking about patenting their materials to make them available; others prefer a situation where the seed banks transfer material only subject to a materials transfer agreement that included a provision that no one seek a patent on the material.

Because this Project seeks discoveries for all humanity and commercial profit is not its main concern, he suggested that the Project sign material transfer agreements with the host country and with anyone who took samples from the Project. Those agreements should provide that no genes in the materials would be patented. Although such a ban on patents might not have much practical importance, the issue of patenting genes is politically very sensitive in many countries. A stipulation that no one would patent genes taken from cell-lines from

the Project, combined with an undertaking to share profits if the Project's cell-lines led to products with commercial value, would be very useful. He argued that unless the Project made a preemptive strike on the patent issue, it would open a Pandora's box.

Discussion

After an inquiry from Prof. Greely, Drs. Giddings and Reid said that the Biodiversity Convention neither expressly includes nor expressly excludes coverage of human genetic resources. Dr. Giddings agreed with Dr. Reid on the importance of a preemptive position on this issue, though he would allow more scope for intellectual property protection for materials that had been improved in accordance with international common law or U.S. patent law.

Prof. Greely, seeking clarification, posed a hypothetical question. If the Project discovered that an isolated population had a gene that protected it against an infectious disease and a researcher using the Project's cell-lines isolated the gene, found the protein for which it coded, and patented not the gene but the protein as a basis for a drug, would Dr. Reid expect problems? Dr. Reid said that the major political concern in most countries was the actual patenting of the gene. Protection for the protein would be appropriate, although he thought it should come with some sort of the return to the host country. Dr. Reid thought the discovery of a human gene that a commercial firm wanted to add to a crop's genome would be a much trickier issue, although one unlikely to arise. In such a case, the firm would want to patent the gene itself. It would be useful to the Project to say, from the beginning, that no patents will be sought for genes taken from the Project's samples. A firm interested in patenting the gene could always go back independently to the country where the gene was found and make whatever financial arrangements were appropriate, but that would not implicate the Project.

Dr. Giddings said limiting a ban on patenting to the gene itself would eliminate many, if not all, of the problems he had seen with Dr. Reid's position. Even so, he believes Dr. Reid's proposition would meet with opposition from some developing country governments. The developing countries often have a completely erroneous view of the process of product development and one that will often hurt innovation. Questions of intellectual property rights should not be involved significantly in this Project. This Project is a research project, not intended to have commercial value, and should deal with these issues only to the extent necessary to avoid real political problems.

Dr. Kenneth Weiss said the Committee wanted to make sure that anyone, anywhere in the world, can get access to the samples for research. He also pointed out that none of the property rights, or benefits, involved here are likely to get back to the villages. He argued that we aren't really talking about benefits to the people who would be sampled. Prof. Greely said he hoped they would discuss how to try to get the funds back down to the sampled population. Dr. Giddings said the Food and Agriculture Organization had been dealing with this at some length and found that there was no good solution. He believes that the Project would be better off avoiding this very tricky issue.

Dr. Paul noted her disagreement with Dr. Giddings's view of intellectual property and the developing world and then asked for clarification of the positions of Drs. Giddings and Reid. Dr. Reid said that the issue of patenting genes may be relatively trivial as a practical matter, but that possibility would raise a lot of political concerns. Unlike Dr. Giddings, Dr. Reid thinks developing countries, such as India, might well accept his formulation. For one thing, it would allow them to continue to sidestep the tricky issue of the patentability of human genetic material.

Dr. Schneider asked whether the Project could proceed by excluding any possible commercial use. Prof. Greely said it would be hard to exclude commercial use while keeping open access to the cell-lines. And he added that the distinction between research use and the subsequent commercial use would be difficult. Dr. Schneider said perhaps you could prohibit the commercial use of the sample; it was pointed out that copying the materials would be easy.

Dr. Siniscalco raised the possibility of having any royalties or profits derived from the Project going to an international body to be used for general human benefit. Prof. Greely replied that the question was really one of what host governments would accept. Dr. Feldman pointed out that if part of the funding came from U.S. government sources, researchers would have to sign a standard patent form. He was not sure what effects, if any, that form might have on obligations to the host countries. No one present knew the answer to that question, although Dr. Eckstrand *(I think it was Irene, but I can't tell from the tape)* pointed out that almost every bilateral agreement involving the NIH contains annexes dealing with the division of property rights in different countries. These generally reserve intellectual property rights in different parts of the world to participants from different countries.

Dr. Reid noted that one could have any royalties go to a UNESCO fund to be used to promote technology transfer to developing countries. UNESCO management would alleviate many political concerns; technology transfer was a major concern in the Biodiversity Convention negotiations. Prof. Greely noted that this still would not directly help the sampled populations, although Dr. Weiss pointed out that about three quarters of the populations to be sampled would be groups that were integrated into their country's economy and culture.

Dr. Reid said there had been a great increase in awareness among ethnopharmacologists of the need to help the communities involved in their research. There are not formal contracts, yet, but there has been a great increase in the benefits going back to the affected communities. The New York Botanical Gardens has been funding health centers and other benefits the populations want. This has been going on without the attention of national governments.

Dr. Evans pointed out that the significance of a gene can only be seen from the entire samples; one cannot know that one population's allele is important without comparing it to the entire sample. The complexity of

compensating the populations from all over the world is daunting. She also pointed out that this Project should not be viewed in isolation, but as an offshoot, to some extent, of the Human Genome Project. In that Project, the thrust of the discussion is against patenting genes but in favor of patenting their useful applications. Commercial interests will come in at that stage. If the information is available to industrialists, you cannot stop them from trying to use it or to protect any valuable applications they make. She pointed out that patenting genes is different from patenting useful knowledge derived from them.

Prof. Greely said he thought Dr. Evans's point on the importance of all populations was very important. He recalled a suggestion Dr. Eric Lander had made to him in a telephone conversation that any royalties should be viewed as a "mutual fund" for the benefit of all the sampled populations. This, he felt, plays back into the idea of UNESCO or another international organization as fundholder to promote the interests of indigenous peoples or, perhaps, those who contributed samples to the Project.

Dr. Paul noted that the discussion was confusing two separate questions -- who is the appropriate fundholder and what is generating the funds. There will be disagreements between proponents of intellectual property rights and proponents of the third world's interests. The Project will have to confront those broader issues and not just questions of the patenting regime.

Prof. Greely urged that the Project needed a consistent approach to avoid having different arrangements with different countries.

Dr. Siniscalco pointed out that patenting a gene, gained through one method, could be circumvented easily. Those who gain will be those who build the product most efficiently; another slightly different version of the gene could always be found for use. Prof. Greely argued that, even if the practical implications of intellectual property rights in these genes were meaningless, it may be necessary for political reasons for the Project to have a position on this point. Dr. Giddings suggested that it might be easy for the Project to offer high royalties, because the Project would not, in fact, be conceding much of value.

Several people disagreed over what legal or ethical entitlements might be owed to those people or populations who provided samples that led to commercial products. The *Moore* case in California was evidence that at least one donor thought he had property rights to his cells and that issue has been settled, as far as those present knew, only in California. Dr. Giddings urged that the Project need not answer all these issues before going forward; he supported Dr. Reid's general approach as a good way to deal with the issues.

Prof. Greely noted that some ethicists felt that human genes, and humans, could not ethically be termed into market commodities. He did not know whether anyone at the workshop took that position, but sought comment on it. Dr. Evans said it was very important and noted the different European laws on payment for blood donation. Prof. Greely noted American laws banning payment for most organs and Dr. Evans said the same was true for many countries in Europe.

Dr. Evans noted that the European Project was considering seriously advertising for donors so that everyone who was sampled would have made an affirmative decision to come forward and participate in this Project. She recognized that this method could not be extended to every population, but thought it should be kept in mind. The information seeking participation could include language dealing with property issues.

Dr. Weiss pointed out that this discussion fed back into informed consent. It would be very difficult to explain to people that their blood could be "sold" or, having explained it, to get their permission to take samples. Distributing the subjects' genetic material as a business would raise great concerns. Dr. Feldman pointed out that one goal was to have several cell banks, located in various regions. The laws of distribution of the cells are likely to differ in major ways among these countries; whatever the Project wants to do may be overriden by those local laws.

Dr. Schneider noted that he personally would be much less likely to donate blood if he were told it would be used to make a profit. He then raised the question of whether any governments were very eager to participate in the Project and even wanted to do a more thorough study. Dr. Evans said that France had already done a province-by-province study of its populations and was, as a result, more eager to participate in the overall Project.

Dr. Weiss noted that some countries might demand, as a short term quid pro quo, laboratory facilities in return for giving permission to sample.

The U.S. government is interested in looking at variation within its borders for forensic purposes. That raises concern among some groups. Dr. Cavalli-Sforza pointed out that the Project was not likely to have major value for forensic purposes because most of the populations it will sample are not heavily represented in the U.S. population.

Dr. Bodmer noted that there had been a Japanese study of 128 different populations. The cell-lines from that study were considered public domain, with no identification of individuals. There was no government involvement in this because the people who collected the data were from the countries where the sampled populations were located.

Prof. Greely noted again his belief that the higher profile of the Project and the negotiation of the Biodiversity Convention would make past experience a poor guide to predicting host governments' demands in the future.

Third Roundtable Discussion -- Racism and Other Possible Misuses of the Project's Data

The third roundtable discussion concerned the possible misuse of the Project's samples or results. Prof. Greely started by saying that he thought the issues in this roundtable were the most important to be discussed today.

Although a variety of misuses were possible, Prof. Greely noted that, as an American familiar with America's historical problems, he considered white/black racism the greatest concern. He then turned the meeting over to the speakers on that issue: Dr. Diane Paul, Dr. William Schneider, Dr. Eric Juengst, Dr. Luca Cavalli-Sforza, and Dr. Robert Murray. (Dr. Murray was detained by an emergency and arrived late in the discussion of this issue.)

Dr. Diane Paul

Dr. Diane Paul directs the Program in Science, Technology, and Values at the University of Massachusetts at Boston and has studied the social history of genetics. She started by saying she thought the issue of racism would be the least important or difficult ethical problem with the Project. Everyone now says they are against racism, including, these days, every racist. We know where we come out on general principles. Difficult issues remain, but they are of a different nature from the discussions we just had over intellectual property. She urged that the Project consider intellectual property issues from a variety of perspectives, including the perspectives of the developing world and its advocates.

She said the Project is likely to reinforce conventional understandings of race and ethnicity. For example, in current genetic screening programs, although geneticists and anthropologists emphasize that traditional race is arbitrary, screening programs use traditional ethnic and racial categories for the screening. This tends to make those categories *seem* more real. She could propose no solution, but is one way in which the Project will have social implications.

On the other hand, she does not think the specific results of the Project will either reinforce or undermine racism, but instead are likely to have very little effect on such attitudes. Dr. Mary-Claire King has said that the Project will undermine racism by showing how much alike we are; Dr. Paul said this is naive because the findings will not be clear-cut. The findings will be a set of statistics that will have to be interpreted to the general population. The discussion will be in terms of "more or less" -- is this is a substantial or insubstantial difference? It is like the nature/nurture controversy. Even if scientists put a number on the hereditability of a trait, the number will not settle anything because some people will interpret the same result as large or small depending on their social interests. If the average human being is heterozygous at 10 percent of loci or if a population is polymorphic at one third of loci, is that a lot or a little? It depends on the background assumptions brought to the question. She would not put too much emphasis on the consequences of the Project's results because, whatever they are, they will be controversial and will be invoked on behalf of every possible claim.

She pointed out that it also is not easy to say what findings support the "progressive" side. In the classical balance controversy, which was a controversy about the extent of human diversity, H.J. Muller argued in the 1950s and 1960s that we are all alike. At the same time, Theodosius Dobzhansky argued that we were all very different, that genetic diversity was good for individuals and was good for populations, and therefore that we should expect to find diversity

maintained. For Muller it made sense to talk about a normal evolved type of a gene and for Dobzhansky it was nonsense.

Today we think that if the Project shows that people are all basically alike, that will be good and will undermine racism. But much of the bitter dispute between Muller and Dobzhansky flowed from the fact that they thought different social policies flowed from their views. In that dispute, Muller was the eugenicist and Dobzhansky was the anti-eugenicist. Dobzhansky wanted to show we were different in order to undermine Muller's eugenics policies. Muller's policies made sense if one thought there was genetic uniformity, because then one could conclude nature was striving for a best type. Dobzhansky wanted to say that nature loves diversity and people should not reduce diversity through eugenics. These policy implications are tremendously plastic and the particular connections people make are highly contingent. As a result, Dr. Paul would not put much hope in the Project undermining racism. Look at the former Yugoslavia; the Human Genome Diversity Project will probably not have much effect on that situation.

Dr. William Schneider

Dr. Schneider is chair of the Department of History at Indiana University-Purdue University at Indianapolis and has studied the history of research examining blood groups and populations. He began by saying that the Project's results will be misused, but it was not clear by whom. He noted that his analysis is based on his study of the history of similar efforts, and not a deep knowledge of current science. He did note, however, that the Project probably will not produce the results that are anticipated, because, historically, scientific projects rarely do produce the expected results.

Dr. Schneider has studied the use of blood groups from 1900 to 1950 and in particular between the two world wars. He has looked at the work done by scientists from the discovery of blood groups and their inheritance by Mendelian laws and primarily the discovery, during World War I, that blood groups were differently distributed in different ethnic groups. Hundreds of researchers published over 1200 articles in dozens of journals based on tests of several hundred thousands of subjects in scores of countries and colonies around the world. As a result, this effort may have been larger than the proposed Human Genome Diversity Project. The most frequently asked question was whether blood groups could provide a different definition of a race. Researchers were also interested in questions of links to disease, insanity, criminality, and so on. The working title of Dr. Schneider's project is "The First Genetic Marker;" the research he examines has both similarities to and differences from the Human Genome Diversity Project.

The first similarity is that the blood group effort was based on a new scientific discovery, in its time probably as revolutionary as the genetics revolution. Second, blood grouping had immediate practical applications in blood transfusions. Third, the blood group research wanted to explore differences between peoples. There had been earlier efforts, but, like the Project, the blood

group research used new technology. The blood group researchers were also interested in the origins, movements, and mixings of populations.

The blood group research began with individual researchers finding samples from whatever groups they could, taking their samples, and publishing their results. Hospitals, soldiers, prisons, and mental institutions provided convenient sources for samples, as did the patients of doctors or missionaries stationed in remote parts of the world. The general model was that individual researchers did their work and published it in journals. Periodically, someone would publish compilations and analysis of that data. Some countries put forth systematic proposals similar to the Project. One of the most ambitious proposals was made in Germany in 1926, by the German Society of Blood Group Research. This Society, founded by a particularly "volkish" anthropologist named Otto Rechte *(spelling?)* and a Navy doctor named Paul Stefan, divided the German world (including Austria) into 900 districts and called for the testing for 500 subjects in each district. It appealed for the cooperation of local doctors through publications in medical journals and in general interest magazines. After about 12 studies, some based on school children, the Society asked the Prussian state welfare ministry to perform blood testing in all school districts. After the lengthy hearings, the council recommended only limited testing. The Society did not get what it wanted but it did get some legitimacy. It went on to create a journal to publish the results of the work of others.

This interest was not limited to Germany. The Soviet Union created two research centers, one in Leningrad, which urged doctors to send samples to a central location for analysis under controlled conditions, and one in the Ukraine that followed the German model of asking doctors to do their testing on site. Two smaller countries, Holland and Denmark, persuaded their governments to support systematic studies of blood groups within their countries. Both of them used centralized analysis. Between 1919 and 1939, 75 countries and colonies were sampled with over 200,000 subjects.

Was this used for racist purposes? Yes, in two ways. The first came from Hirschfeld's findings that a higher percentage of Type A blood was in northern and western Europe with increasing proportions of Type B blood in central and southern Europe and moving into Asia. He devised something he called the biochemical index of race, which was simply a population's percentage of Type A blood divided by its percentage of Type B blood. Consciously or subconsciously, this meant that the highest numbers applied to northern and western Europe. It was fairly quickly recognized that the index did not work as a guide to conventional views of race, but studies through the late 1930s continued to use the Hirschfeld index to put the populations of the world into a hierarchy.

In Germany, especially, the Aryan ideologists liked to show charts of the world based on the Hirschfeld index, interpreting the results as an invasion of Europe by Type B blood with Germany as a bulwark against this invasion.

Similar kinds of studies were done in other countries as well. Gypsies, Jews, the Lapps, and Native Americans elicited particular interest.

Dr. Schneider asked whether this story has a moral. The research into blood groups and ethnicity did not represent a totally unmitigated disaster. The studies showed that the simplistic idea that a single number could define race was untenable. The blood groups ultimately provided a body of information that became a fundamental part for the whole field of human population genetics after World War II. The concept of race was different in 1949 than in 1919, at least in part because of this research. Thus, the research did make a difference, at least among scientists. This underscores the importance of access to information. It is a fundamental tenet of science that open access helps determine whether the truth will come out. You need to think long and hard before restricting any access to data.

Dr. Eric Juengst

Dr. Eric Juengst has a Ph.D. in philosophy and directs the NIH program on the Ethical, Legal, and Social Implications of the Human Genome Project. He began by saying that he is an optimist about the ability of the scientific and genetic community to anticipate and to manage the impact of its research. He sees as success stories for such active management both the recombinant DNA debate in the 1970s and the procedures that came out of it and the discussion of gene therapy and the procedures that were developed for its assessment. He hopes that in the 1990s the Human Genome Project and its efforts to build in assessment will be a success story. He got involved in the Project to because he thought we could anticipate the issues and prevent overly deterministic and reductionistic interpretations of personal genetic information and thus prevent "genetic discrimination."

He believes the Diversity Project will take us to the next level of difficulty. The Diversity Project stands to inherit from the Human Genome Project the role of the lightning rod of genetics. The main social risk of the Human Genome Diversity Project is what he would call "demic" discrimination. This is not necessarily racism or racial discrimination, because "racial" categories look quite outdated, but discrimination against particular demes or subpopulations of the human community as a result of the misinterpretation or the misuse of the conclusions of studies done with the data collected by the Diversity Project. He sees three reasons to think the Project should anticipate such demic discrimination.

First, unless it is very careful with its educational efforts, the Project is likely to be perceived by the public as an effort designed to establish a taxonomy of human types and categories. That will inevitably bolster contemporary notions of race and ethnicity. The statement from the first workshop report that the plan is to identify "the most representative descendants of ancestral human populations" worldwide and then preserve the genetic sequences suggests a typology. The counterargument is that the Project will find more similarity that difference but this is called the Human Genome *Diversity* Project, not the Human Kinship Project or the Human Family Project, because the scientifically interesting parts are the differences.

At the very least, a massive educational Project like the one planned for Europe is probably a prerequisite for the Project as a whole. That means the Project will necessarily have a high profile; there will be no chance to do this quietly. How will the Project be perceived and interpreted and whether it be perceived as a typology of human types that begins to look like a search for the "pure strains" of various types of humanity are vital issues.

Second, the findings the Project is expecting could be used to fuel existing human antagonisms. At the top of the list of the scientific questions are questions about relations to neighboring groups and questions of migrations -- when groups arrived in particular areas. But, he noted, lineage and land tenure are probably humanity's favorite excuses for making claims about social privilege. New scientific evidence bolstering the claims of one party or another could easily fuel existing fights. One can envision two scenarios. In one, the dominant group in a society is included in the Project, has its DNA hallmarked, and uses those hallmarks as inclusion criteria -- if you want to be part of the ruling group, you need to have the right DNA marking. The other side could be the flip side -- a dominant group uses DNA marking to identify groups for oppression. These groups would have to have access to appropriate DNA types or hallmarks to plug into a system of forensic DNA typing, but that will probably coming anyway, given police interest in this methodology.

Third, one reason this Project is urgent is that many of the populations are vanishing. Why are they vanishing? Some are being assimilated into larger populations for no particularly nefarious reasons, but others are vanishing because they are at political or social or economic disadvantage and can no longer maintain the cohesion of the communities that once sustained them. That suggests that some of these demes already face political risks. Dr. Juengst likes the approach of starting with the least politically risky groups, in order to learn from experience what problems to anticipate and how to avoid them.

GAP IN THE TAPE -- the following section is reconstructed from notes and not from the audiotape. Those of you who spoke during this section should examine it with special care.

What can we do to minimize these dangers? Dr. Juengst had two suggestions. First, the Project should be preceded by an educational campaign to define the Project to the public. It would be very dangerous to allow others to say what the Project is about; it would be much better for the Project to take an active role in presenting itself to the public.

Second, he suggested that the Project should include a standing advisory group, akin to an IRB, to review requests for access to information. This group would ask researchers why they wanted the data or samples and what they intended to do with them. Such a group could serve to limit the misuse of the Project's data.

Dr. Luca Cavalli-Sforza

Dr. Luca Cavalli-Sforza is a geneticist at Stanford and a founding member of the Human Genome Diversity Committee. He began by defining racism as a belief in the biological superiority of some group. He noted that there is no biological basis for this belief, either in terms of genotype or in terms of phenotype. To the extent arguments are made for one group's superiority, they are always based on phenotype and mainly on behavioral traits about which we have very little genetic data.

Questions of the genetic basis of IQ became controversial about at the same time Dr. Cavalli-Sforza arrived in the United States in 1971. Arthur Jensen, encouraged by physicist William Shockley, argued that IQ had a substantial genetic basis and that the American black population had a lower IQs than the white population for genetic reasons. In fact, this argument was based on bad science in a variety of ways and scientists, including himself, exposed its shortcomings.

The traditional concept of race is not biologically meaningful. There is usually as much or more genetic diversity among the inhabitants of one isolated village as there is between population groups. If we want to understand something about human anthropology or history, we need to look at hundreds of traits. This DNA work will only confirm what we already know to be true about the meaningless of race.

He agreed that an educational effort is a good idea. He noted that the Project will not add much to the forensic uses of DNA technology because the groups to be sampled are not generally of forensic importance in the countries that are pursuing forensic technology.

Discussion

Prof. Greely opened the discussion by asking how can we minimize the misuse of the Project and whether the concerns are so great that the Project should not proceed.

Dr. Paul said that she had little confidence in the ability of educational campaigns to improve public understanding. Dr. Siniscalco stressed the importance of the Project for "euphenics," improving the phenotype of humans through a better understanding of the interactions of genes and environments.

Dr. Kenneth Weiss pointed out that racism is much older than science and is based in differentiating one group from another. Racists will always find data to misuse, whether or not it comes from the Human Genome Diversity Project.

Dr. Bodmer asked what group would be appropriate to provide oversight of the use of the Project's data.

Dr. Mark Weiss noted that there is plenty of data in the published literature that could be used for racist views. If someone wanted to use the Project's cell-lines to generate new data, he would need a fairly complicated lab. Dr. Weiss also noted that there are plenty of examples of racism in modern America,

sometimes from surprising directions, and pointed some aspects of the Afrocentric movement as an example.

Dr. Feldman noted the distinction between misusing bad data, as Jensen did, or misusing techniques in manipulating good data. The markers that geneticists use today are racially biased because they are drawn from white populations, thereby making them, for many purposes, incomplete. He also noted that data can be abused by omitting some data, as in one study that purported to show a high degree of hereditability based on studies of 44 pairs of twins, while omitting data on another 111 twin pairs.

Dr. Hollander stated that the Project does have an ethical obligation to seek to limit misuse. *At this point, the tape resumes* She noted that reactions to the Project vary in part with individuals' different levels of hope and fear. She would make a very different prediction from Dr. Cavalli-Sforza, but we cannot know in advance what the effects will be. How can we address problems when we can't know their full dimensions? She noted that Dr. Juengst had suggested the Genome Project as a model, but she argued that it is a flawed model. With the Genome Project, the scientific work goes forward and the Ethical, Legal, and Social Implications program goes forward, but the conclusions of the ethical, legal, and social implications program affect the science, if at all, in ways that are highly mediated and perhaps not very effective. If the Project is going to take these issues seriously, it needs to set up some kind of forum that is involved in making decisions on how to proceed on the technical side. If the ethical concerns are not integrated with the technical decision, they will not be effectively addressed.

Dr. Hollander noted that Dr. Juengst had earlier argued for involving the affected communities. This involvement needs to be addressed not only in the design of the project, but in some kind of continuing oversight, monitoring, and reconnaissance effort. If for no other reason, having community representatives involved will prepare the Project to have these conversations again and again, as it explains itself to the world.

Dr. Paul said the Project needs to leave open the question of calling a halt in case the problems turn out to be worse than anticipated. The second question in section three of the agenda for this meeting was whether the concerns about such misuse are so great that the Project should not be undertaken. She said that the discussion had have for granted that the answer to that is "no," but that she believes it should be an open question. Prof. Greely noted that he had put the question on the table and that there had not been any effort to "hide" that question; calling an open workshop at NIH is not consistent with trying to avoid the question.

Dr. Eckstrand pointed out that NIH has a policy of open sharing of resources regardless of what those resources are. The example she knows best is Genbank, the DNA sequence database. That is a database of all DNA sequences that have ever been published. It is available to anyone who wants to tap into it. Genbank is maintained by the scientific community and people have

the right to use or misuse the data. By and large, the misuses have been honest scientific ones. She cannot imagine NIH saying it should have a resource of this nature and not make it available to anyone who wants to use it.

Dr. Kenneth Weiss said first that he knows people from American minority groups who are offended that the Human Genome Project is using their tax dollars but is not including them. African-Americans know that a lot of medical genetics for them cannot be done because we don't know enough about the genotypes. If it were decided that it was not important to sample the rest of the world, it would be saying, in effect, that nothing out there is different or important. We know, he said, that this is not true.

Second, before the second planning workshop, which he had organized, he had expected the Project to produce a great deal of controversy among anthropologists, who, as a group are very sensitive to these problems. Instead, there was almost universal enthusiasm for this. And for biomedical reasons, given the acceptance of the infinite alleles model that most mutations are unique, there is no excuse for not sampling everyone. He said we know mutations are going to be different in different peoples. Some of the populations the Project should sample represent tens of millions of people in the U.S. alone.

Dr.Juengst said he thought for the Project to have this conversation in public, with the press in the room, was a great beginning. More generally, he noted that no one present had criticized or dismissed the reasons to do this Project. Instead, the discussion has centered on how to execute the Project without causing more trouble than necessary.

Dr. Cavalli-Sforza said he was glad that the NIH policy is to have everything open. That is how science should proceed. He then noted that data on some 215 genes have already been collected. That information is not as good as the data at the DNA level sought by the Project, but it has been collected from a great number of populations. Nothing bad has come out of those efforts. What harm can come out of data of the new type? A trait like IQ is a product of perhaps 200 genes. Any ability to collect and analyze data on the frequency of those and other genes for behavioral traits in populations is so far in the future that not even the youngest person in this room is going to see.

Dr. Murray then arrived and, after a break, spoke.

Dr. Robert Murray

[The tape did not record the very beginning of Dr. Murray's comments -- I hope he will fill in anything I missed.]

Dr. Robert Murray is a physician and a faculty member in clinical genetics at Howard University. He has long worked with issues of genetic diversity and genetic screening in African-American populations. He began by noting that he had been involved in these kinds of issues for some time. Howard University has been collecting samples and information on genetic diversity among African-Americans. His experience has made him concerned about the possible misuse of genetic information concerning populations with respect to behavioral or social

issues. Thus, for example, people might use information about a genetic propensity to low birth weight in one African population to argue that the problem with low birth weight African-American infants was "genetic" and should not be used to justify improving living conditions or medical care.

Dr. Murray said he had just come from the AAAS meeting earlier in the week. At that meeting, the media was interested only in the question of crime and heredity, even though there is no data to support a true genetic relationship. He had been interviewed three times, not to discuss genetic screening, which he knows something about, or genetic engineering, for which there is some information, but about crime and heredity, where there is nothing to discuss. A person who misuses the Project's data will have a ready platform unless there is a quick response from someone the press can turn to for information.

Concerns about the social uses of information on genetic diversity arose because of the pressure in the black community to do a better job of collecting bone marrow and kidneys for transplantation. Many people expressed the worry that the existence of different markers would be used to suggest that "these people" should be kept separate.

Dr. Murray said the risk of harm comes from when studies are released and how they are reported. It is important that data be reported only when researchers have a substantial amount of information is also important. For example, the past few years highly publicized studies have reported genetic linkages for manic-depressive syndrome, alcoholism, and other traits. Those linkages were then disputed. Word of the initial studies is widely distributed, but the follow-up that the linkages do exist gets much less publicity. As a result, people continue to believe, erroneously, that these linkages have been proven to exist.

Our society talks about populations with different backgrounds and characterize certain life styles as "primitive" or "underdeveloped" when those populations live lives very different from ours. If significant genetic differences turn up in markers for these populations, some will attempt to attribute their cultural status to a genetic basis. Wilson's views of sociobiology are believed by many people who think much of our behavior is genetically controlled.

Dr. Murray said he hoped that, in the design of the Project and whatever standards are set for the reporting of the data, significant thought will be given to providing responses or rebuttals to people who draw unjustified conclusions from the Project's data.

Discussion

Prof. Greely noted that Dr. Juengst had earlier suggested something like an IRB to see who should have access to data. He noted that Dr. Murray's concerns might be allayed by an entity that would have advance warning of what was being explored and perhaps even what results were to be published. This entity could constitute a "ready response team" to try to put findings into perspective.

Dr. Murray said that this had been tried with gene therapy. The human gene therapy subcommittee spent a lot of time drawing up guidelines, including some concerning dealing with the media. They were primarily concerned about early report giving false hope if, for example, someone gets evidence of gene transfer in one particular case.

Dr. Siniscalco reminded the group that this Project is the best opportunity to collect this kind of data. Knowledge may be misused but we know that ignorance is at the basis of every misconception. The Project, he said, will give us a chance to understand not just history but the relationship between genetic makeup and environmental influences. In Sardinia, within the space of a few hundred miles there were major ecological differences and human populations with substantial inbreeding. This has led to huge genetic differences based on the degree of exposure to malaria. The misconceptions of people like Dr. Jensen will draw forth responses from people like Dr. Cavalli-Sforza to correct them. We need to record these differences before human intermingling makes it too late. We must get the data; issues of the possible misuse of the information will be dealt with as it is interpreted. Otherwise we will not be able to accumulate the knowledge that could benefit everyone.

Dr. Porter noted that the discussion has brought out issues of how much should be centralized and decentralized. Dr. Juengst's proposed oversight body would be a centralized effort to ensure the highest standards of scientific review of the work that will be done on a decentralized basis. Centralization also will help in informing the press and in dealing with the political process. There will necessarily be some disadvantages to becoming too centralized, including possible stifling of academic freedom or new uses for the data, but the question of centralization or decentralization is one useful way to think about these issues. What kind of oversight body would the Project want to construct to get the work going and to keep it going?

Dr. Kenneth Weiss said that all that the Committee is proposing so far is to organize the collection of data, not to do any kind of analysis of that data, except to run a standard set of markers on all samples collected worldwide. The Committee has not proposed that the Project should be responsible for administering grant applications for analysis. What happens after sampling is a question the Committee has not considered, although it is a valid one.

Dr. Murray said the problem is that collecting the samples will provide a source for abusive kinds of analysis. He had not thought there was any question about the Project going forward. He thinks it should go forward -- on the whole it will be beneficial rather than harmful. However, just one or two instances of misuse could destroy all the good the Project will do. The University of Maryland debacle undercuts all the good work geneticists have tried to do. He does not want that to happen with this Project. The discussions today are about trying to design the Project in a way to minimize the possibility of misuse of the data. The Project cannot divorce itself from responsibility for possible misuse, although it might want to.

Dr. Kenneth Weiss replied that the Committee was very concerned that the data be open and not be seen as the private domain of two or three labs. Once it is open, it becomes like Genbank. Genbank is only statistical; CEPH (the Center for the Study of Polymorphisms in Paris) might be a better example because it provides physical samples and not just data. He asked whether anyone could get samples from CEPH?

Dr. Cavalli-Sforza replied that after a request, CEPH writes to members of CEPH asking for information about the scientific validity of the planned research. CEPH makes an investment in sending DNA to people for free. Its data bank is accessible to all members, but there is a screening process for getting cell-lines. CEPH can turn people down. Dr. Feldman noted that CEPH is more restrictive than the Human Genome Diversity Committee had intended the Project to be. Dr. Cavalli-Sforza was concerned about whether access should be limited or whether, if NIH funds were used, it could legally be limited. Prof. Greely pointed out that there may not be a true dichotomy between limited and open access. One could have open access for researchers to cell-lines, but might require information about the planned studies and advance notice of the results.

Dr. Evans noted that access involves both the cell-lines and the data generated from that material, which will be in the data bank. The Project might want to have the data freely available, while demanding information about a study before giving access to samples. Someone noted that this would not require a screening committee, but perhaps just an organization that watched how the information was being used and what issues might be arising.

Dr. Murray said the discussion of open access reminded him of the early days of sickle-cell testing, before there were any requirements for laboratory or professional certification. Many people were performing sickle-cell testing without any background and they did lots of harm. He doesn't think the Project wants this material to be accessible to people without scientific background. If it does, one of his nightmares will come true. Someone with mischief on their mind will ask for the materials for the purpose of doing something harmful with them.

Dr. Mark Weiss said he had just been looking at the catalog of cell-lines from National Institute of General Medical Science. That catalog, he noted, qualifies the meaning of "open." Cell cultures and DNA samples are distributed only to qualified professional persons who are associated with recognized research, medical, or education institutions, so it will not be sent to absolutely anyone. Dr. Eckstrand *(?)* agreed that a research could not just write in from her home address, particularly because that material is limited. If the material is unlimited, such as the DNA database, then NIH has made it open to anyone. She was not sure whether cell-lines should be classified as limited or unlimited. Dr. Feldman asked whether NIH imposes any restrictions. If someone writes in and says they have money from a right wing foundation and he wants a sample, would NIH provide it? Dr. Eckstrand replied that if he writes in from a recognized institution, the bank might check with its project officer and would probably be advised to send the cell-lines.

Dr. Giddings noted that the U.S. Department of Agriculture had had similar problems. He thinks the Project needs to be aware of the potential for misuse and to have contingency plans to deal with it. It should not restrict access, but it should be prepared to respond to abuses.

Prof. Greely gave as an example the possibility that the Project had learned a researcher was about to claim that, based on data from the Human Genome Diversity Project, all Irishmen (his own ethnic background) were inherently prone to alcoholism. In that case, the Project could have a spokesman prepared to put that claim into perspective for the press.

General Closing Discussion

Prof. Greely noted that the meeting's time was nearly up. He invited people to discuss any other issues they thought were particularly important (while adding that he would be happy to hear comments on any of those issues later).

Dr. Murray said that education about the goals and limits of the Project could help prevent a lot of problems. It was noted that there was fairly general agreement about the value of education.

Dr. Siniscalco said that classifying populations is an old habit of people. We may now be able to classify people more scientifically, but that is not necessarily a bad thing. We have moved from eugenics to euphenics, where we can try to make the life of everyone equal regardless of what genes they are born with. To do that, we need to understand the relationship between genes and environment, which is what this Project is about.

Prof. Greely and Dr. Feldman both thanked the participants sincerely for their useful discussion of these issues and, at 4:00, the workshop concluded.

March 5, 1993

To: Participants in the Human Genome Diversity Workshop 3(B) on Ethical and Human Rights Issues
From: Henry T. Greely, Workshop Chair
Subject: Draft of My Recommendations to the Human Genome Diversity Committee

Attached is an eight page memo setting out my recommendations to the Human Genome Diversity Committee. I am circulating it in draft form for your comments, if any, before sending to the Committee. This is **not** a Workshop document and you should not feel any obligation to read it, let alone give me comments on it (or agree with it). But I learned a lot from our discussions in Bethesda and I believe that, at this point, you are more knowledgable about these issues than anyone else. I would really welcome your comments or criticism.

As with the summary, I would like to finish this document fairly soon. It would be great if you could get me comments at the same time you comment on the Workshop Summary (Friday, March 12), but, in reality, I probably won't put this into final form until Friday, March 19. You can send your comments by telephone at (415) 723-2517 (office) or (415) 961-2304 (home, where I am often working this term); by fax at (415) 725-0253; and by e mail at henry.greely@forsythe.stanford.edu.

Once again, thanks for your help.

March _, 1993

To: Human Genome Diversity Committee
Participants in the Workshop on Ethical and Human Rights Issues Raised by the Human Genome Diversity Project
Other Interested Persons (list attached)
From: Henry T. Greely
Subject: Recommendations Concerning the Human Genome Diversity Project and Ethical and Human Rights Issues

The Human Genome Diversity Project Workshop on Ethical and Human Rights Issues, held on February 17, 1993, was very valuable in raising and refining issues the Project should consider. Based on those discussions, and relying heavily on ideas expressed by participants in that Workshop, I want to recommend some specific actions to the Committee for its consideration. Although these recommendations are based in large part on the discussions at the Workshop and have been circulated to Workshop participants for comments, I want it to be clear that they are my personal recommendations and not the recommendations of the Workshop. Workshop participants may well have different views.

As the holding of the Workshop indicates, the Committee has been sensitive to these issues thus far. You have approached these issues appropriately; my most important recommendation is that you continue to take them seriously, for two different reasons.

First, the Project has the potential to be wonderfully beneficial, but there are real possibilities that it could have significant negative effects -- among the sampled populations or elsewhere. Concerns about those effects are legitimate and need to be addressed. I believe Project has a moral obligation to strive to minimize any harmful results from its work.

Second, these ethical issues are also political issues. Unless they are treated appropriately, the Project could face political opposition similar to the storm invoked by the proposed University of Maryland seminar on genetics and crime. The scientific issues are different, the nature (and significance) of the ethical concerns are different, but the political effects could be similar. And even if political opposition does not emerge, the risk of such opposition could easily frighten possible sources of funding. Political disputes arising out of these ethical issues can destroy the Project more quickly than anything else and, as the Committee has pointed out, science is running out of time to sample some populations.

My recommendations are broken into six categories: structural, collecting issues, patent and royalty issues, misuse issues, education, and other issues. Many of those recommendations deal with setting up organizations to accomplish particular functions, leading to a total of one committee and four subcommittees. I know this may make my recommendations look like a bureaucrat's dream. I believe, however, that in the long run organizational structure is the best assurance (albeit not perfect assurance) that these functions will be performed. I strongly feel the Project should want someone to perform the functions this memorandum identifies. The details of my

proposals may turn out to be more or less relevant, depending in part on the future structure of the Project, but I think you will need to consider their substance.

Recommendations

Structure

1. The Project should have an Ethics Committee at a high level in its organization. Whether it should report to the international committee or to a specifically United States committee will depend on issues that depend on future contingencies and will not be discussed in this memo.[1] In either event, I will refer to the high-level committee to which it will report as "the Project Committee." In order to ensure that the findings of the Ethics Committee are considered in the operation of the Project, the Ethics Committee should both make regular reports to the Project Committee and share some members with the Project Committee.

2. The Ethics Committee should not include all members of the Project Committee, but only those members who are willing to put time and effort into the Ethics Committee. I do not think it would be useful to have all principal investigators on the Ethics Committee, as was suggested at the February 18 workshop. The same level of impact could be achieved by having several members of the Project Committee on the Ethics Committee without diluting the intensity of the Ethics Committee.

3. A majority of the members of the Ethics Committee should not be members of the Project Committee and probably should not be involved in the scientific work of the Project. It would, of course, be important to choose members with a decent understanding of the scientific work.

4. The Ethics Committee should include representatives from different ethnic backgrounds and, if possible, representatives from some of the populations to be sampled. It should also include at least one representative from advocacy groups for indigenous populations.

5. The Ethics Committee should be chaired by one of the members not directly involved in the scientific research. An independent chair provides both more oversight and more credibility.

6. The Ethics Committee should be encouraged to create subcommittees or other independent operating bodies to perform particular tasks. (I would expect subcommittee members usually to be members of the Ethics Committee itself, but that is not essential.) These recommendations contemplate four and perhaps five such subcommittees: an Executive Subcommittee, a Collection Subcommittee, a Social Implications Subcommittee, a Review Subcommittee, and possibly an Education Subcommittee. Others may prove useful. The Collection, the Social Implications, and possible Education Subcommittees are described in the sections below on collection

---

[1] Some of the issues involve the actual organization of the Project and others turn on whether the federal government would give any formal status to the Ethics Committee, such as allowing it to serve as the equivalent of an Institutional Review Board.

issues, misuse, and education issues, respectively. (I have attached a possible organization chart.)

7. The Executive Subcommittee would be a small group charged with handling any urgent duties that might arise.

8. The Review Subcommittee would be, in some ways, similar to review committees in medical clinical trials. It would be charged with examining the actual ethical impact of the Project, both on sampled populations and on society at large. It would collect information on that impact and report at regular intervals to the overall Ethics Committee. The Review Subcommittee would have as part of its charge the responsibility to recommend halting or making major changes in how the Project operates if it believed those steps were necessary. Dr. Paul pointed out the importance of this kind of check on the on-going effect of the Project and I strongly agree with her. It is important that the Review Subcommittee be (and be perceived as) substantially independent of the Project.

9. Some of the work of the Ethics Committee and its subcommittees might be able to obtain direct funding. As two examples, creation of sample collection protocols or the preparation of an educational program for the Project might attract government or foundation funds. The NIH program on the Ethical, Legal, and Social Implications of the Human Genome Project is an obvious place to start looking. I expect that the other functions would require funding through the Project, but with appropriate use of telecommunications for meetings, their costs should not be large. I think one would want to have semi-annual meetings of the whole ethics committee and few, if any, physical meetings of the subcommittees. I have not tried to prepare an estimated budget.

Collection Issues

1. The Ethics Committee should have a subcommittee on issues arising from collection of samples. This Collection Subcommittee should represent a number of different interested groups, but should definitely have the active participation of both anthropologists and members of or advocates for indigenous peoples.

2. There is an active and interesting on-going debate about the degree to which "western" ethical standards should be applied to research conducted in other cultures. This is a fascinating and important debate, but, as long as the Project wants funding from the United States government, as a practical matter it is largely irrelevant. Federal regulations require the application of "western" standards. Those have been successfully applied in similar work in the past and I see no reason why they cannot be successfully applied to the Project. Fighting them will be much more difficult than working with them

3. The Collection Subcommittee should begin by collecting protocols from similar sampling projects that have already received federal funding. It should analyze those protocols to see how they dealt with the ethical issues involved.

4. One issue that came up at the Workshop but was barely discussed involves subjects who are, or seem likely to become, ill. Issues include what illnesses to screen for, what to tell the subjects, what to do for the subjects, and what to do with infected blood samples. In spite of their limited discussion at the Workshop (where Dr. Weiss may have been the only person to raise them), these issues are **very** important, both in themselves and because their resolution necessarily affects the degree of confidentiality for individual samples. The Collection Subcommittee should examine this set of issues and quickly report its conclusions to the Ethics Committee.

5. The Collection Subcommittee should set as an early goal the drafting of one or more sample protocols for researchers involved in collecting for the Project. These sample protocols would discuss and describe many of the difficult issues and provide options that might be appropriate under differing local conditions. They would be used not as requirements for protocols for this research, but as guides to writing a protocol that would be appropriate in an individual situation. If possible, the Office for Protection from Research Risks should be involved in the process of preparing these guidelines. The resulting sample protocols would be reported to the Ethics Committee for adoption.

This seems to me to be potentially a very exciting and useful task, akin to the CIOMS guidelines recently prepared for epidemiological work. I do not know whether anthropologists have similar kinds of written guidelines (and would be interested in hearing Ken Weiss on that point). The Project could provide a real opportunity to pull together thinking about scientific work in different cultures in the practical context of sample protocols. I think this may be the most important role for the Collection Subcommittee and, if done well, could be useful for the conduct of biomedical and anthropological research far beyond the Project itself.

6. For many issues, the sample protocols should provide different permissible ways to proceed and should discuss why a researcher might want to proceed one way rather than another. These issues include at least the following:

- Ways to seek informed consent
- Descriptions of the purposes of the research
- How to deal with the confidentiality of individual subjects
- How to deal with the confidentiality of sampled populations
- How to deal with the confidentiality of collected samples
- Methods of providing reciprocal benefits to the sampled populations
- Different ways to deal with subjects that the researchers recognize are, or are likely to become, ill.

7. The Collection Subcommittee should review the collection protocols of researchers seeking Project support to collect samples. Depending on the structure of the Project and the requirements of the federal regulators, this review might have the formal effect of an Institutional Review Board and thus require Collection Subcommittee approval as a condition for proceeding with the work. Even if it did not have that effect, it could still serve two important functions. First, it could help improve the protocols submitted by researchers. Second, it could keep the Project from being

involved in research that did have great ethical problems and thus protect the Project from the ethical and political implications of such involvement.

8. The Collection Subcommittee should also examine the circumstances surrounding the collection of materials previously collected that are sought to be added to the Project's cell-lines. Those collectors of those materials should not be held to the same standards as current and future researchers and I do not anticipate that serious questions would often be raised. Nonetheless, I think it is important to have some kind of check to avoid any Project involvement with horrendously unethical research. (This is akin to the unending debate about the ethics of using data from unethical Nazi-era German medical experiments.)

Patent and Royalty Issues

1. These issues may not, in fact, arise, but it is important that the Project be prepared for them. On the other hand, the issues appear to be more political than ethical. The Ethics Committee might be a useful place for the Project Committee to seek some guidance, but I do not think its expertise will be crucial in this set of issues.

2. On the merits of these issues, I believe the suggestions by Dr. Giddings and Dr. Reid could form a strong basis for the Project's position. The Project should agree with host governments that any user of samples collected by the Project must stipulate that no genes derived from that sample will be patented. The Project should also agree to require any user to provide some kind of payment (probably a royalty) to the Project in the event any commercial products developed from the sample. The Project would then use those royalties in some way for the benefit of the sampled populations. The exact manner of using those royalties requires additional thought and depends, in large part, on what host governments would find acceptable. Possibilities include donating them to an existing organization (like UNESCO, for example), returning them to the host governments, or applying them directly to the sampled populations.

Misuse Issues

1. In the long run, I believe the misuse issues are the most worrisome, both ethically and politically. I know it can be hard to see how, scientifically, harm can come from the Project's data or findings, but scientific validity and political effectiveness are different issues. I recognize that any data can be misused, that good data may be less subject to misuse than bad data, and that racism and "demic discrimination" have been around for a long time and are likely to continue without much effect from the Project. Nonetheless, I believe the Project has a moral obligation to try to anticipate the problems of misuse and to limit them, as far as it reasonably can. Otherwise, it may end up like Dr. Hirschfeld, the leading researcher on blood groups and populations, who, as Dr. Schneider pointed out, was trapped in the Warsaw ghetto by the Nazi regime that used his work as one bit of legitimation for its policies.[2]

2. The Ethics Committee should have a Social Implications Subcommittee to deal with the problems of misuse. I see this Subcommittee performing several functions.

---

2 Hirschfeld survived.

3. Along the lines suggested by Dr. Juengst, researchers requesting samples (and perhaps data) collected through the Project should be required to make a written request for the samples. That request should describe briefly the purposes and methods of the proposed research. Neither the Social Implications Subcommittee nor any part of the Project should have the power to **deny** a qualified researcher access to material. Whether some academic credentials should be required to make a researcher "qualified" is a different question and one on which I do not have a view. The Social Implications Subcommittee or the Ethics Committee should address that issue quickly. ??

The Social Implications Subcommittee would review the requests for several purposes. First, that review will give it the ability to keep track of the uses to which the Project's samples are being put. Second, it will give the Subcommittee the opportunity to provide helpful suggestions to researchers whose proposals seem flawed. Third, it will provide the Project with advance notice of at least some research that could lead to ethical concerns.

4. The written requests for samples (or perhaps data) should contain the stipulation that the researchers will give the Social Implications Subcommittee advance notice of any publication or press release concerning research using materials from the Project. Once again, this stipulation would **not** give the Subcommittee or the Project any power to block release of information. It would give the Project warning of the imminent release of any ethically troubling research. I do not know what advance notice period would be reasonable, but would think that the Subcommittee would want at least three working days.

5. When the Social Implications Subcommittee learns, either through the advance notice or retrospectively, of the release of research that has used Project materials and that raises ethical concerns, it should make experts available to the news media to discuss the research. These experts would try to place the research into its proper scientific context. I think much of the working press understands that it has problems in evaluating scientific research and it would probably welcome this function. The Subcommittee would serve as a clearing house that the media would consult in order to get the kind of perspective they lack. Such a "ready response" function could both limit the damage done by incomplete or questionable research and limit the damage done by unintentionally misleading reporting of good research.

Education

1. The Project should work quickly to prepare a campaign of public education about the goals of the Project and the meaning of its expected results. It will be important, as many Workshop participants stressed, for the Project to define itself rather than let it be defined by others. The Project may want to put primary responsibility for this campaign on an Education Subcommittee of the Ethics Committee. On the other hand, the Project Committee may want to keep primary responsibility for the educational program.

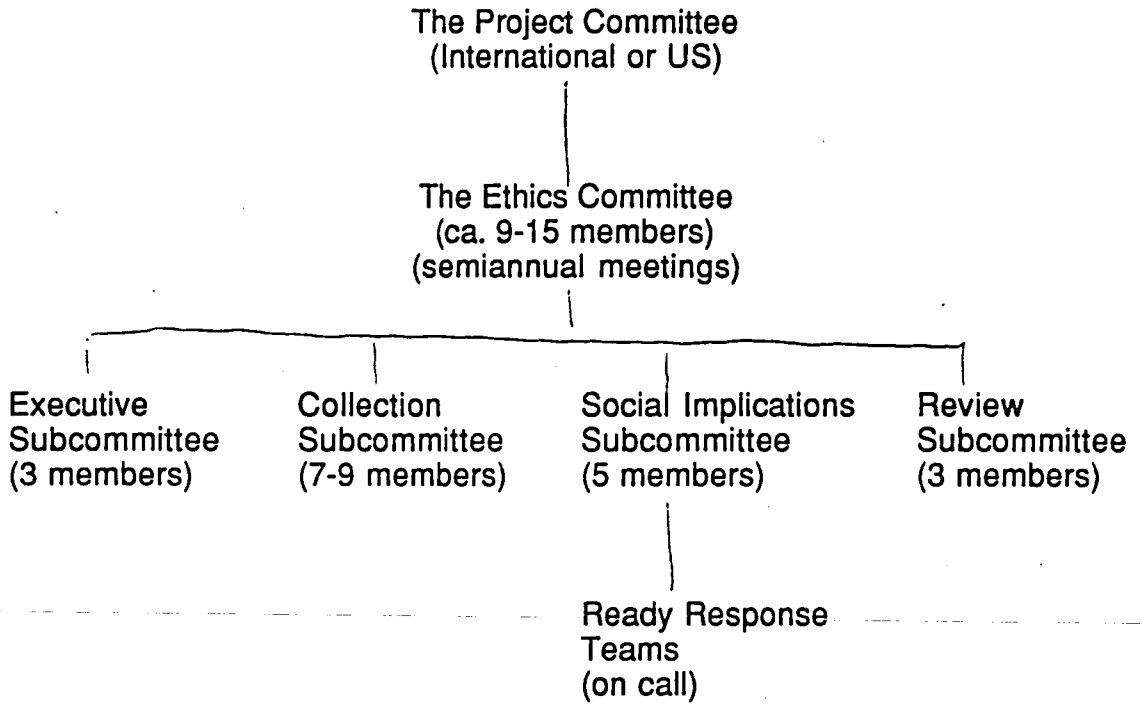
2. Methods for providing this kind of education may differ. I can imagine the Project wanting to do anything from a pamphlet to a museum exhibit to a videotape. Independent funding and connections with other organizations (such as public television producers) will be important for this mission.

3. At least some kind of educational material needs to be prepared quickly so that the public can get careful and consistent answers from the Project about its goals and methods. Something along the lines of a short memorandum, press release, or pamphlet, explaining the Project and its goals in simple terms, should be prepared very quickly.

Other Issues

1. I agree with the suggestion several participants made that the Project may want to affiliate itself with a United Nations body, probably UNESCO. I know Dr. Cavalli-Sforza has been working along these lines and I encourage him to continue.

2. As Dr. Giddings suggested, the Project should consider beginning its sampling with populations that raise the fewest ethical and political problems. It should give special consideration to beginning with populations in countries that sponsor the Project. Experience with those populations, and a record of success in dealing with them, may be very helpful in sampling more vulnerable groups. Of course, in individual cases this general preference for politically safer populations will have to be weighed against the risk that a population will disappear.



The size of the committees and subcommittees are my early guesses, based in part on the amount of work I anticipate and in part on the maximum effective size for carrying out some of these functions. The Project might also want to have an Education Subcommittee.

Memorandum of Agreement
Between the
National Center for Human Genome Research/NIH
and the
National Science Foundation

NCHGR Agreement No.: 1 YO1 HG20006-00

I. Purpose and Scope

The purpose of this Memorandum of Agreement is to enable the National Center for Human Genome Research (NCHGR), National Institutes of Health (NIH) to contribute to the National Science Foundation (NSF) conference grant entitled "Human Genome Diversity" (DBS-9209745). The NCHGR will provide partial support for this project, which will support two workshops and one conference to outline the framework for asking and ultimately answering questions related to human genome diversity.

II. Services to be Performed

This support will allow the principal investigator, Dr. Marcus W. Feldman and his co-investigators, Drs. Luigi L. Cavalli-Sforza, Mary-Claire King, Kenneth Kidd and Kenneth Weiss, to convene two workshops and one conference to develop a strategic plan for the Human Genome Diversity Program. This Program has two complementary aims, the preservation of crucial biological information on vanishing human populations and the analysis of human genetic diversity. The specific aims of these planning meetings are: (1) Workshop 1 will define optimal ways of obtaining samples of human populations to best represent human genetic diversity; (2) Workshop 2 will identify populations of relatively unmixed descendants of the most interesting and representative ancestral populations for sampling; and (3) The conference will set priorities, research strategies, and organizing principles of the Program and will identify ethical concerns.

Experts from North America, Europe, Australia, Japan, the former Soviet Union, and developing countries will be invited to participate. The two workshops will be held in the United States and the conference will be held in Western Europe. Information will be made available to the larger scientific community through articles in the scientific press.

III. Reporting and Other Requirements

The grant award document will contain a statement specifying that the NCHGR is providing funds toward the support of the two workshops and the conference.

The NCHGR will receive a final report summarizing the results of each workshop and conference and a copy of any publications that result.

Any publications resulting from this project that is supported by NCHGR will acknowledge such support.

Since the Human Genome Diversity Program is an international undertaking, the NCHGR encourages the NSF to ensure that other countries and international agencies support fiscally this endeavor.

IV. Period of Agreement

This agreement shall be in effect from the date of execution through December 1993.

V. Revision, Modification, or Cancellation of Agreement

This agreement may be reviewed or modified by means of an amendment in writing only by the signature of the parties signatory to this agreement or by their designees. Neither agency may cancel the agreement.

VI. Funding

The NCHGR/NIH agrees to transfer FY 1992 funds in the amount of $38,000 to NSF to be used to fund Grant No. DBS-9209745 (Feldman/Stanford University). These funds are to partially support two workshops and one conference; approximately one-third of these funds should be used for each activity.

A. Method of Payment:

Payment shall be made up to the total amount upon receipt of Standard Form 1081. Funds not obligated by September 30, 1992, must be returned to NCHGR/NIH for deobligation.

B. Accounting Data:

Appropriation	CAN#	Amount
7520891	8427235	$38,000

C. Billing Address:

National Institutes of Health
Division of Financial Management
Federal Assistance Accounting Branch
9000 Rockville Pike
Building 31, Room B1B04
Bethesda, MD 20892

VI. Liaison Officers

VII. Authority

This agreement is made under the authority of the Economy Act of 1932, as amended (32 U.S.C. 1535).

VIII. Acceptance and Approval

National Center for Human Genome Research

By: Elke Jordan
Elke Jordan, Ph.D.
Deputy Director, NCHGR, NIH

Date: 4/15/92

By: James C. Vennetti
James C. Vennetti
Executive Officer, NCHGR

Date: 4/15/92

National Science Foundation

By: _____
Mark Weiss, Ph.D.
Program Director, NSF

Date:

By: _____
Karen Sandberg
Grants and Contract Officer, NSF

Date:

Tracking the Parade of Mankind Via Clues in the Genetic Code

2/22/93, page A3

By Boyce Rensberger
Washington Post Staff Writer

Hidden within the DNA of each human being is a record of that person's ethnic history. It is a chronicle that begins in humanity's dim evolutionary past and traces the ancient migrations and ancestral intermixings that have shaped every tribe and culture on Earth.

Encoded in the genes, scientists say, are the answers to such questions as:

- Did today's Europeans arise in Europe or did they migrate from a homeland elsewhere? How are they related to the people of India?
- From what tribes in Africa are black Americans descended? Can individual African Americans be linked to specific tribes?
- Are Iraqis in the city of Basra the direct descendants of the ancient Sumerians, as other evidence suggests?
- Did all the Indian tribes of North and South America descend from a single group that migrated from Asia or were there separate migrations from different parts of Asia at different times?
- Was much of Africa once populated by the small, light-skinned people known as "Bushmen," who were then pushed into a small enclave in southern Africa by the Bantu people migrating out of west and central Africa?
- Did modern human beings evolve in Africa and migrate to other continents, following more primitive hominids, or did they evolve on the other continents from those more primitive forms?

Such longstanding mysteries and countless others have tantalized historians, anthropologists and armchair explorers for centuries. Now, according to a group of scientists who met at the National Institutes of Health in Bethesda last week, it is possible to answer those questions with several new techniques of molecular biology similar to DNA "fingerprinting." Taken together, the answers could help anthropologists reconstruct human history from its evolutionary origins to its present diversity.

But it can be done only if DNA samples can be collected from hundreds of ethnic groups, including many small tribes living in remote locations, and the individual genomes tested for hundreds of different "markers."

Collecting 'Mud' Samples

This is the genetic equivalent of Sherlock Holmes's trick of checking the mud on a suspect's boots. Just as a person carries evidence of his travels in the various muds on his shoes, an ethnic group carries odd bits of DNA codes picked up from other groups with which it has intermarried, sometimes including ethnic mixings that occurred thousands of years in the past.

With large samples from many populations, geneticists can check the DNA for various markers and, often, determine how long each has been present in the group, based on estimated rates of evolutionary change.

"It's a big undertaking. But if we can do it, there is every reason to think that we can address some of the most fundamental questions about who we are as a species and where we came from," said Marcus Feldman, a Stanford University geneticist who organized the three-day meeting of anthropologists, molecular biologists, ethicists and officials of funding agencies.

The group's central committee has been working out a plan for the project and preparing to seek funding. A tentative cost estimate is $23 million over five years.

The money would pay for anthropologists and technicians to visit 400 ethnic groups (a preliminary list has been drawn up), draw blood samples from 25 unrelated individuals in each group and rush the blood to a laboratory where cells from each of the 10,000 individuals can be processed into a form that will grow indefinitely in laboratory dishes. The money would also cover costs of examining each "cell line" for a standard set of markers and for a system of making the data available to any researcher who wanted to analyze the collection.

Anthropologists say the project has great urgency. Many of what

PLEASE SEE NEXT PAGE

CONTINUED FROM PREVIOUS PAGE

one project leader calls "the hippest people"—for "Historically Interesting Populations"—are vanishing. Whole cultures are becoming extinct at an increasing pace. If not literally dying out, their members are intermarrying with other groups. The nearly extinct Ubykh people of Turkey, for example, are thought to number no more than three. Dozens of other groups are down to a few hundred members.

The 'Reference' Genome

The effort is called the Human Genome Diversity Project—similar to, but emphatically different from, the Human Genome Project, the $3 billion effort to decipher all the 50,000 or so human genes and map their positions on the chromosomes. The larger project, which is in progress under the sponsorship of NIH and the Department of Energy, seeks to establish a single "reference" genome.

"That project will give a completely misleading impression," said Luca Cavalli-Sforza, a Stanford geneticist and one of the first to propose the diversity project. "There isn't just one human genome. Each group has its differences and each person has differences. If we don't understand that diversity, we're missing a lot that's important."

Along with anthropological questions, said Mary-Claire King, a geneticist at the University of California at Berkeley and another of the group's leaders, the project would also have great payoffs in the study of human disease, much of which results from the interaction of genes with environmental factors. She said the project could study why genetically susceptible people develop certain diseases in some locales but not in others. One mystery, for example, is why Americans of African ancestry have a high risk of hypertension but native Africans do not.

One of the thorniest issues discussed at the meeting was whether discoveries of genetic differences among the world's peoples would or could be misused as support for belief in the superiority of one race or ethnic group over another. The scientists said the findings could not logically lead to such conclusions and noted that even the concept of race had long ago lost its scientific validity. King predicted the project would undermine popular notions of race by showing that the genetic differences within groups now thought to belong to one race are more numerous than the differences used to define races.

But a panel of ethicists and historians specializing in debates over genetics emphasized that misunderstanding and misuse is inevitable.

"The project is likely to reinforce conventional views of race and ethnicity," said Diane Paul, who heads a program in science and values at the University of Massachusetts in Boston. "Whatever the findings are, what matters is how they are interpreted." Paul said the simple fact that genetic diversity is being studied will reinforce the idea that groups are different and that the differences are significant.

Eric Juengst, an ethicist with the Human Genome Project, predicted that if some studies showed certain groups to have occupied a given territory before others, the finding would be used to support the increasingly incendiary claims of land tenure in ethnic disputes.

"Questions about who arrived where first are probably humanity's main excuse for asserting privilege," Juengst said.

Even the term "diversity" was seen to have such negative connotations in Europe that a separate but linked genetic effort to chart 35 to 50 European ethnic groups substitutes the word "variation." For some Europeans, a British scientist said, "diversity" evokes images of Nazi rationales for eliminating it.

Nonetheless, the European Community's science agency is expected to grant money this year to begin the European genetic survey. The rest of the world, to be covered by the U.S-based project, may have to wait. Officials of the NIH and the National Science Foundation cited President Clinton's budget-cutting goals and told the scientists their chances were not good.

MOLECULAR ANTHROPOLOGY

How to Sample the World's Genetic Diversity

Allan Wilson wasn't at the meeting at Stanford last month—he died last July—but, without invoking the supernatural, it's safe to say his presence was strongly felt. The meeting was the first in a series to plot the course of an ambitious project that Wilson, a molecular anthropologist at the University of California, Berkeley, and Luca Cavalli-Sforza, a population geneticist at Stanford, conceived last year: to survey the genetic diversity of humanity. By analyzing how people vary one from another, researchers hope to glean clues into human origins, evolution, migration patterns, adaptation, and disease. The goal, says Berkeley geneticist Mary-Claire King, one of the organizers of the meeting, is nothing short of understanding "who we are as a species and how we came to be."

There was little time to waste if those questions were to be answered, Wilson and Cavalli-Sforza realized, as many of the aboriginal tribes most essential for reconstructing human history are disappearing—some literally dying out, but most lost to acculturation, intermarriage, and the influx of modern society. Wilson, Cavalli-Sforza, and colleagues issued a plea for help last year in the journal *Genomics*, calling on scientists worldwide to help them collect DNA samples from hundreds of indigenous populations and preserve them for future study, creating a huge resource and database for the scientific community (*Science*, 21 June 1991, p. 1615).

But there was a problem: Though Wilson and Cavalli-Sforza were united on the project's goal, they were deeply divided on how to achieve it. Cavalli-Sforza wanted to use the traditional approach of sampling well-defined populations, while Wilson, eschewing all the assumptions inherent in identifying populations in the first place, wanted to sample along a geographic grid, collecting DNA from aboriginal peoples at more or less evenly spaced locations around the world. They reached no agreement before Wilson's death, though that has not dampened enthusiasm for the project, which has garnered an outpouring of support from anthropologists, geneticists, and linguists worldwide and, as of this summer, some modest funds for three planning workshops.

At the first of these workshops, held at Stanford in July, 40 top population geneticists and molecular anthropologists found themselves dealing with the same issue that had divided Cavalli-Sforza and Wilson, albeit framed more broadly—namely, how best to sample the world, and do so without bias. The problem, as the workshop participants were well aware, is that the way in which they collect data today will determine what answers scientists can tease from them later. And everyone wants something different. Anthropologists are seeking clues about the migration of early humans out of Africa and the settlement of the Americas. Linguists want to look at how languages and cultures move with people, and population geneticists want to evaluate the relative importance of drift versus natural selection, among other things, in human evolution.

Thus the challenge to the group, as stated by Pennsylvania State University anthropologist Kenneth Weiss, another of the meeting organizers: "Can you help us design a sample to answer different questions, and to be sure that what we propose now will be useful later?" That set off 3 days of intense debate not just on population versus geographic sampling but also on sample size, the relative merits of blood samples versus cell lines, even the goal of the project: whether it is to survey human biodiversity or reconstruct human history. (The answer, it seems, is both.)

"Can you help us design a sample to answer different questions, and to be sure that what we propose now will be useful later?"

–Kenneth Weiss

Populations versus geography. Cavalli-Sforza gamely took a first crack at outlining a strategy, lobbying hard for the population-based approach he has followed for years and dismissing Wilson's geographic grid as "impractical." The chief difference between the two approaches, he says, is that Wilson's looks at individuals in specific locations; whereas his looks at populations, defined by some ethnic identifier like language or culture, in regions. Both are useful for reconstructing human evolution and migration, and each has its strengths, said Cavalli-Sforza. A grid approach, for instance, is particularly useful for tracking the spatial advance of new mutations. But the problem with a grid on a worldwide scale, he said, is that there simply aren't representatives of aboriginal populations scattered every 50 or 100 miles across the globe. Some places may have no inhabitants at all, while others may have several tribes. "We have to consider the ethnic origin of people. It is not enough to go by a lattice," or grid-like approach, asserted Cavalli-Sforza, adding, "Unfortunately, Allan is not here to defend his view." "He'll try," quipped King, alluding both to Wilson's doggedness and the fact that many of Wilson's "descendants"—including herself—were in attendance.

Instead of using a grid, Cavalli-Sforza wants to select a minimum of 200 isolated aboriginal tribes that best represent ancestral populations. A scientific/medical team would then collect blood samples from 50 individuals in each and rush them off to a few centralized labs, where white blood cells from each person would be transformed into permanent cell lines. While cell lines are a financial and logistical nightmare, he conceded—at the worst case, they might cost $500 each, and blood samples have to be at a lab within 48 hours—they are the only way to ensure an inexhaustible supply of DNA from members of vanishing populations.

Cavalli-Sforza tried to accommodate Wilson's preferences as well, proposing that each team would also collect hair roots or cell scrapings from the inner cheek—which provide a simple and cheap, though limited, supply of DNA—or perhaps blood samples not for immortalization from additional people in the surrounding region. Even though it wouldn't establish a full-fledged grid, this strategy would allow some of the wider sampling Wilson had advocated.

But Cavalli-Sforza instantly found his compromise challenged, not just by the Wilson camp but also from others, like Weiss, who were disturbed that the number of populations was so small—a mere 200 out of the 5000 to 7000 believed to exist. "It greatly restricted the kinds of questions we could look at," said Weiss. (In all fairness, Cavalli-Sforza saw 200 as the "absolute minimum," chosen with an eye on the budget, tentatively estimated at about $23 million for the first 5 years of the project.)

The arguments did not begin in earnest, however, until Cavalli-Sforza took a stab at spelling out his criteria for selecting populations. The groups should be aboriginal, he said, which in the New World means those that were in place as of 1492, and well-defined, either by language, geography, or endogamy (the habit of marrying within one's own tribe). In addition, there should have been little recent interbreeding with other groups. Practical considerations enter in as well: A team should be able to reach the group "without spending $1 billion," said Cavalli-Sforza, which would rule out certain populations in Borneo that take 2 to 3

3

weeks to reach. And to ensure access, an anthropologist must already be working with the tribe.

"I am very troubled by the list [of criteria]" said Mark Stoneking of Penn State, who, like his mentor Wilson, believes the population approach is predicated on too many assumptions. "It just focuses on well-defined ethnic and linguistic groups. And when you are done with your survey you will find the human species is made up of well-defined ethnic and linguistic groups. By sampling that way you bias the results."

The crux of the issue, said Svante Paabo of the University of Munich, formerly of the Wilson lab, is how to define a population. Cavalli-Sforza believes, as do numerous other population geneticists and linguists, that language is a useful if imperfect approximation for identifying a population. True enough, said Paabo, but he worries about those instances when genetic and linguistic boundaries do not coincide. If you sample only on the basis of linguistically defined populations, he cautioned, "you may miss something interesting." An alternative would be the type of experiment he and former Wilson lab colleague Anna Di Rienzo, now at the University of California, San Francisco, are contemplating for the Nile Valley. They plan to collect DNA samples every 50 kilometers to see whether genetic variation does in fact correlate with linguistic variation, among other things.

Others grumbled about Cavalli-Sforza's admittedly symbolic focus on 1492 as the time when aboriginal groups began to be displaced. Plenty of "jostling went on before then that could have had an effect on gene flow," said Weiss. "Agriculture, which arose 10,000 years ago, had a big effect. The Romans expanded long before 1492, and those people expanded at the expense of someone."

L.L. CAVALLI-SFORZA/STANFORD UNIVERSITY

Guides to the past. The genes of indigenous people, like these Andaman Islands aborigines, can help us understand the evolutionary history of the human race.

But it was Stoneking who laid out the first major challenge when he questioned the very underpinning of the survey. We don't need immortalized cell lines, he blithely told the group. Not only are they too expensive, but the logistics preclude you from getting samples from places far from airports. Instead, he said, they could get nearly as much information by extracting DNA directly from blood samples. With existing technology, one sample would supply enough DNA for 1000 PCR procedures (a very sensitive method of analyzing DNA variation)—"basically an infinite amount." What's more, he said, whole blood is stable one week on ice, and the cost is $5 to $10 a sample, instead of $500 or so per cell line. The upshot, he said, is that they could survey 500,000 to 1 million people for the same price as Cavalli-Sforza's 10,000. That would enable scientists to recreate later whatever sample they like by selecting data from the database—for example, reconstructing a geographic grid to look at Bantu expansion, or focusing on an entire village to look at evolution on a microscale.

Stoneking's idea of "saturation" sampling caught people's fancy, but no one was willing to budge on cell lines. Says Weiss: "We are conservative enough as scientists that no one wants to give up something that we know works." And so the group remained at an impasse, chafing at the limits of such a small sample but unwilling to abandon Cavalli-Sforza's focus on well-defined groups and cell lines.

Finally, the theoreticians offered a way out when they challenged another tenet of the population-based approach: that DNA from 50 individuals per group is needed to tackle most questions about human diversity. "There has been one big sleight of hand here. Where did 50 come from? Why not 25 or 100?" asked Charles Langley, a *Drosophila* geneticist at the University of California, Davis, who spearheaded the charge.

The answer, said Cavalli-Sforza, is pragmatic: "One person can bleed 50 people and get to the airport in 1 day." That's not the only reason, adds Robert Sokal of the State University of New York at Stony Brook, who notes that a sample size of 50 is needed to give an accurate estimation of gene frequencies in a population.

Times are changing, responded Langley, who was supported by Montgomery Slatkin of Berkeley, Joe Felsenstein of the University of Washington, and others. Until recently, Langley explained, few markers existed for studying human genetic variation, so geneticists had to compensate with a large sample size. But now, he said, with an essentially unlimited supply of new DNA markers—coupled with sophisticated new statistical tools—most questions can be answered with a smaller sample by simply increasing the number of the gene loci examined, provided the genes are independent. Explained Felsenstein: "If you double the number of loci, you can halve the number of people." For most questions, in fact, a sample size of 25 or even 10 per population would be sufficient, this group argued.

Compromise reached. But there are tradeoffs. While the approach is extremely powerful for addressing global questions of evolutionary history, such as migration patterns, it is not good for questions that involve specific loci, such as how disease susceptibility is distributed among populations. "Anthropologists will have to face reality—some questions of interest just can't be answered," said Weiss, who was clearly taken with the idea of spending less on cell lines.

After some haggling, that argument carried the day, and the group settled on collecting samples from 25 individuals in each population for immortalization rather than the original 50. And that meant they could survey 400 populations instead of 200, to the delight of everyone there. The meeting was breaking up when Felsenstein, one of the advocates for small samples, got cold feet. "We think we are correct," he later explained to *Science*, "but it is a huge project. What if they go out and use our number and we are wrong?" As a "fall back," he urged them to collect many extra blood samples, as Stoneking recommended.

"I always intended to," answered Cavalli-Sforza, who made good on his word. In the meeting summary, the organizing committee agreed that in addition to establishing cell lines from a core group of well-defined populations, the goal should be to collect blood samples from many individuals in each region where the teams are working. And with that, the group crafted a compromise that even Wilson, who had vowed to dig in his heels, might have accepted. At least his former postdocs, like Di Rienzo, did. "I am happy to live with that," she said. "It provides a control to see if the definition of a population is real or in your head."

At the next workshop, however, consensus may prove more elusive. Several dozen leading anthropologists will meet at Penn State in October with their lists of populations that must be surveyed—a list that seems certain to exceed the allotted 400 populations. Setting priorities may make the sampling issues seem like a piece of cake.

—Leslie Roberts

GENOME DIVERSITY PROJECT

Anthropologists Climb (Gingerly) on Board

Since a group of geneticists first called for a massive survey of humanity's genetic diversity a year and a half ago, anthropologists have been dying to get their two cents in. In late October they got their chance—perhaps even more than they bargained for—at a grueling 3-day workshop at Pennsylvania State University. The organizers of this effort called together about 50 of the world's leading anthropologists, archeologists, and linguists and gave them a tough challenge: to identify the 500 or so indigenous populations most worthy of genetic study, out of the roughly 7000 believed to exist worldwide. For the assembled anthropologists, that was rather like trying to put together a sparse meal from a smorgasbord groaning with delights.

Adding to the immensity of the task before the anthropologists was the fact that the participants, selected for their expertise on specific regions of the world, came with their own perspectives, biases, loyalties, and research agendas. And then there was the tension, at least in some eyes, between the twin goals of the Human Genome Diversity Project: to get a snapshot of genetic diversity and how populations are related, and to probe human evolutionary history—human origins, migrations, and expansions. A few of the anthropologists also brought some skepticism about the design of the project and even the value of genetic data in elucidating human history.

Yet to the great surprise of nearly everyone involved, the anthropologists put aside their doubts and differences—albeit after some grumbling—and plunged in. They divided up the world into six regions, each of which was assigned to a working group. With overworked graduate students manning the word processors, the groups hammered out a several hundred page report in just 3 days, with details on some 500 populations across the globe. What pulled them together, several members of the group told *Science*, was their sense that however imperfect the survey might be, it is, as South African anthropologist Trefor Jenkins put it, "impossible to resist."

The anthropologists arrived at Penn State to find that the organizers of the project—who include geneticists Luca Cavalli-Sforza and Marcus Feldman of Stanford University, Mary-Claire King of the University of California, Berkeley, Kenneth Kidd of Yale University, and genetic anthropologist Kenneth Weiss of Penn State—had already laid the groundwork (*Science*, 28 August, p. 1204). At an earlier planning meeting, the group had settled upon the somewhat arbitrary target of collecting DNA samples from a core of 400 or 500 populations worldwide—in addition to Europe, which will be handled separately. And they had tentatively agreed on the procedure: taking blood samples from at least 25 individuals in each group. The samples will then be preserved in permanent cell lines to provide reservoirs of DNA for analysis.

Open arms to anthropology. Luca Cavalli-Sforza.

The organizers had also settled on two overall criteria to guide the anthropologists' choices: to strive for a representative sample of human diversity but also to choose populations that are essential for answering major historical questions. The anthropologists were given free rein to identify the most interesting questions—for instance, how many expansions occurred across Beringia (now the Bering Strait) into the New World, or the relations among the many small populations in the Amazon. The problem, though, as the groups quickly realized, is that the two criteria don't necessarily result in the same populations.

For some populations there was no contest. Everyone agreed the highest priority should go to unique, historically vital populations that are in danger of dying out or being assimilated (see box on next page). But selecting the others was not so easy, as the deliberations of the sub-Saharan Africa group, chaired by John Yellen, archeology program director at the National Science Foundation, made clear. Most of the anthropologists in the group were much less interested in a broad

...ey of genetic diversity than in designing ...mple that would bring genetics to bear on ...the questions such as the origins and na... of the Bantu expansion, when the first ...riculturists swept across much of Africa ...me 2000 years ago. They quickly realized, however, that they could use up all of the 100 populations that were allotted for study in Africa on just one question if they were to probe it in detail.

In the end, after railing against the limitations, the anthropologists settled on a methodology that, as Yellen said, gave a little bit to everyone. First, they decided they had to ignore certain classes of questions altogether. Then they designed a sample to capture overall diversity and to look at some of the broadest questions in Africa, such as the origins of modern humans and the dynamics of the important population expansions—the Bantu explosion, the climate-driven migrations in and out of the Sahara, and others.

To do so they first selected isolated populations believed to be relatively unmixed descendants of ancestral populations, like the !Kung of Botswana and Namibia and the Hadza of Tanzania. Next they selected a nearest neighbor or two, to determine if these isolates are as distinct genetically as they are culturally or linguistically. Then, at Cavalli-Sforza's urging, they tried to get a representative sample of the entire continent by using the 1500 or ...major language groups as a guide. Finally, ...e group plotted all these populations on a ...ap to reveal geographic holes and selected 25 additional populations to fill them in.

Other groups ran through a similar exercise, but each made different calls about which questions to pursue and thus how and whom to sample. The North American group, for instance, used a scarcity of native-language speakers—a good indicator of assimilation pressure—as one criterion for selecting populations. The Indo-Pacific group selected some populations specifically to look at phenotypic adaptation. Differences aside, all of the working groups agreed on the value of sampling populations that have already been well studied, since genetic data are far less useful in a vacuum than when culture and history are understood.

After 3 exhausting days, each group ended up with a list of populations and a report describing the status of each population, why it was selected, which anthropologists to contact, who has already collected blood samples, and what ancient specimens in the region—skeletons or mummies—might be available for additional genetic study. The draft, however, is laden with caveats pointing out that it is just a first cut. In addition, the anthropologists caution, no experts were at the ...eeting for some regions of the world—West ...frica and the Caribbean, to name just two—so any suggestions there are especially tenuous.

Weiss calls the list a "living document" that will need to be revised. "This is just 50 people trying to represent the world. We recognize this is not the Encyclopedia Humana but an abstract of what we want to do. There needs to be input from people who may disagree," says Weiss, who plans to publish the report or otherwise make it widely available to the anthropology community. Ultimately, which populations are sampled will depend on finding anthropologists who want to study them—and on funding for the project, which has received only planning money so far. In that respect, the hastily assembled Penn State report will play a key role; after refining and condensing it, the organizers plan to use it to appeal to numerous agencies here and abroad.

Some of the participants, though, would like another crack at the list before it is circulated. Richard Ward, a University of Utah population geneticist who chaired the South American group, for example, worries that the working groups' criteria varied so much that there is little hope of implementing the project consistently. He is pushing for another meeting of anthropologists: "I think we need to sit down and define a consistent approach to decide which should have the highest priority." But the prospects for another workshop anytime soon are dim, says Weiss; funds are limited and the program is already planning two more workshops on other topics. Still, there should be time between now and 1992, when the organizers hope to have funds to begin the project, for the anthropologists' "living document" to evolve.

–Leslie Roberts

A Few of the Chosen

When anthropologists met last month at Pennsylvania State University to draw up a list of populations for DNA sampling in a planned massive study of human genetic diversity, they faced some bewildering choices. The task of winnowing 7000 or so populations down to about 500 was only the beginning of the challenge. They also had to justify their choices, based on specific questions in human prehistory, the possibility that a group is a relatively unmixed "remnant" of a much larger ancestral population, or signs it is rapidly losing its genetic or cultural identity. Here is a sampling of the groups that made the admittedly rough first cut and the puzzles their DNA may solve.

NAPOLEON CHAGNON

BORIS MORTON

They made the list. Yanomami *(top)* and the Chukchi, of Siberia.

■ **Hadza.** These vanishing people—about 200 are left in Tanzania—speak a language like that of the Bushmen of southern Africa thousands of miles away, but morphologically and genetically they resemble the East Africans. Who are they? Did they borrow only the Bushmen's language or do they share their genes?

■ **!Kung.** About 15,000 remain in the Kalahari Desert. Genetic evidence indicates they were once far more widespread, but it is not clear who they are or where they came from.

■ **Plains Apache.** A distinctive group among the Apache, they number about 1000 in Oklahoma. One of few Athabaskan-speaking groups in the Southwest—most are in Canada and Alaska—do they represent a separate migration from the North, perhaps even Siberia?

■ **Yanomami.** The 20,000 individuals living along the border of Venezuela and Brazil speak an isolated language and more closely resemble Central American groups than their immediate neighbors. In addition to the mystery of their origins, the heavily studied Yanomami are a model system for tribal-scale demography.

■ **Yukaghir.** Though fewer than 100 are left in Siberia, these reindeer hunters once dominated the Arctic and Sub-Arctic. Did their ancestors cross over Beringia and contribute to the peopling of the New World?

■ **Chukchi.** About 10,000 of these Paleo-Asiatic speakers live in the Chukchi Peninsula of northeastern Siberia, a relatively short hop to the New World. Genetic studies of them could play a key role in reconstructing the origins of Native Americans.

■ **Onge and Greater Andamanese.** These distinctive populations in the Andaman Islands off Malaysia number fewer than 100 each and are disappearing rapidly. Are they descendants of the tribes that migrated from Africa to Oceania thousands of years ago?

–L.R.

A Genetic Survey of Vanishing Peoples

Racing the clock, two leaders in genetics and evolution are calling for an urgent effort to collect DNA from rapidly disappearing indigenous populations

INDIGENOUS PEOPLES ARE DISAPPEARING across the globe—victims of war, famine, disease, or simply what Cole Porter called the "urge to merge." As they vanish, they are taking with them a wealth of information buried in their genes about human origins, evolution, and diversity.

Now a group of scientists, including population geneticist Luigi Luca Cavalli-Sforza of Stanford University and molecular anthropologist Allan Wilson of the University of California, Berkeley, is calling for an urgent, last-ditch effort—involving geneticists, anthropologists, and medical researchers worldwide—to collect, analyze, and preserve for future study DNA from these populations as part of a massive survey of human genetic diversity. The group, which also includes geneticist Mary-Claire King of the University of California, Berkeley, Charles Cantor of Lawrence Berkeley Laboratory, and Robert [redacted] Deegan of the Institute of Medicine, [redacted] its impassioned plea for funds and [redacted]ance for this new survey in an article to appear this summer in *Genomics*.

They want to study, among others, the Bushmen of South Africa, the Hill People of New Guinea, the African Pygmies, the Etas of Japan, the Basques of Spain, the unique peoples of the Andaman Islands, and the Yanomami Indians of the Amazon rain forest, who are literally becoming extinct (see box on p. 1616). What these populations have in common is that each has been isolated and has only rarely—if ever—intermixed with its neighbors. Consequently, each offers "a window into the past," explains population geneticist Kenneth Kidd of Yale—a unique glimpse into the gene pool of our ancestors who lived thousands of years ago. In an accompanying article in *Genomics*, Cavalli-Sforza and Anne Bowcock of the University of Texas Southwestern Medical Center put it this way: "It is only from the knowledge of the gene pools of these populations that we can hope to reconstruct the history of the human past."

And time is of the essence, according to Cavalli-Sforza, who views humans as an en[redacted]d species in terms of genetic diver[redacted] says this survey should be done [redacted]he next 5 years, or 10 at the outside. For some groups it may already be too late, he says, citing the peoples of Basra, Iraq, who are believed to be the descendants of the Sumerians, as well as the Kurds of eastern Turkey.

In this survey of human diversity, the group plans to exploit the new molecular genetic techniques emerging from the Human Genome Project and, with any luck, some of the project's money as well. Both Cavalli-Sforza and Wilson see this survey as a fitting, indeed long-overdue, expansion of the 15-year, $3-billion effort to map and sequence the human genome, which in their view has been too narrowly defined. For reasons of expediency, the human genome being mapped and sequenced is essentially a Caucasian one, a composite of the genomes already present in DNA banks around the world. Sequencing this "reference" genome will provide a wealth of information, both scientists agree, but they argue that what is sorely lacking is a study of genetic variation, both among individuals and among populations. It is that variation, the minute differences in DNA from one person to another, that researchers want to use to trace the movements of ancient tribes and build what Wilson describes as "a genealogic tree of the peoples of the world."

Francis Black

Vanishing resource. *Geneticists want to collect DNA from such groups as the Arawete. Just 130 members of this tribe remain on the middle Xingu River in Brazil.*

What's more, says Cavalli-Sforza, it can be done for a mere pittance, 1% or less than the total tab for the genome project. "How can we afford to ignore diversity and not study it when we have a chance?" asks Wilson. "It's an insult to a lot of people."

Already the idea has garnered a groundswell of support, from genome project researchers to anthropologists and other scientists who want to use the data. "Intellectually, it is one of the most interesting things to come out of the genome project," says Cantor, who is principal scientist for the Department of Energy's (DOE) genome project.

Kenneth Weiss, a molecular anthropologist at Pennsylvania State University, is equally enthusiastic. "This is the future of the whole area. Even archeologists are trying to get genes out of bones. I am not saying genes are everything but [these techniques] are a major tool for reconstructing history, taxonomy, and phylogeny."

Nor are the benefits strictly intellectual, Weiss adds. Both he and Yale epidemiologist Frank Black have a very practical interest in the project. Weiss is studying diabetes in Native Americans; Black is looking at infectious disease and isolation. Both think that data on genetic variation may shed light on why some groups are more susceptible to certain diseases than others—insights they would be hard-pressed to garner from the Human Genome Project.

Indeed, the only question seems to be who will pay for the project, which falls between the cracks of the federal funding agencies. By Cavalli-Sforza's admittedly rough estimate, the project might cost $10 million over 5 years—far too much for the National Science Foundation's (NSF) anthropology program, which has an annual budget of $8.5 million. And while $10 million or even $20 million is a small chunk of the genome project's total budget, the survey is clearly a bit far afield from what either the National Institutes of Health (NIH) or the DOE typically supports.

However, Elke Jordan of the NIH genome center suggests that the center might be willing to contribute to an international

Scientific Split Over Sampling Strategy

Allan Wilson and Luigi Luca Cavalli-Sforza, who were long at intellectual loggerheads in the debate over human origins, have now joined forces to push for an ambitious new project to survey human genetic diversity (see story). Indeed, a big part of that project's appeal is that it unites these two leaders in the study of genetics and evolution. But while both men agree on the goal—to build a repository of genetic diversity by collecting and preserving DNA samples from vanishing indigenous peoples—they have very different ideas about how to do it.

Both would probably agree on a list of "must sample" populations—say, the African Pygmies or the Yanomami of the Amazon rain forest. But beyond that, arguments begin over how to define an "indigenous" population and where, exactly, to sample. Cavalli-Sforza wants to sample in depth the truly isolated peoples—those who have been living in geographic pockets for hundreds if not thousands of years. Wilson would cast a wider net, setting up a grid across the globe and sampling representatives of indigenous populations wherever he can find them, every 50 or 100 miles or so. Their ancestry would be more mixed up, he concedes, but their genes nonetheless harbor clues to human evolution.

Photo Researchers

African Pygmies. *Cells from the two Pygmy populations in the Central African Republic* (above) *and Zaire have already been preserved for future study.*

The differences between these two researchers are rooted in the technologies they have used for years. Indeed, their split comes as no surprise to their colleagues. "Each has his own unique perspective on the world," says geneticist Mary-Claire King of the University of California, Berkeley. In his efforts to develop a phylogenetic tree of man, Cavalli-Sforza has mainly studied the variation contained within nuclear genes. In the nuclear DNA, any particular gene variant, or allele, is likely to be geographically widespread. In other words, an allele found in Africa will be found in the rest of the world. What differs is the frequency with which they occur in the various populations. Using the nuclear approach, the unit of measurement is a population, not an individual. On this basis, Cavalli-Sforza uses cultural and linguistic criteria to identify well-defined populations, and preferably those that have been geographically and thus genetically isolated for many years. He then samples DNA from members of a series of families—ideally, both parents and their children.

Wilson, on the other hand, has pioneered the study of mitochondrial DNA—a separate genome inherited only from the mother. Says Wilson: "There is a big potential conflict between the nuclear way of looking at things and the mitochondrial way." Because the mitochondrial DNA accumulates mutations much faster than does nuclear DNA, it is far more variable, yielding a stronger genetic fingerprint. And because it does not recombine, each person Wilson samples will be informative, which eliminates the need to measure gene frequencies in populations.

What's more, in his work in Africa and elsewhere, Wilson has found that, unlike the nuclear DNA, the mitochondrial DNA bears "a strong geographic imprint." By that he means that a distinctive mitochondrial allele found in an indigenous person will occur within a fairly small radius, perhaps 50 or 100 miles. Beyond that, you virtually never see it again. And that leads him to his current conviction: "If you want to understand the geography of the human gene pool you need to sample at least every 50 miles all over the world, and I think every 10 miles, really." He concedes, however, that his way might cost a bit more.

To Wilson, Cavalli-Sforza's approach is too full of presuppositions. Instead, he asserts, we should "abandon previous concepts of what populations are and go by geography. We need to be explorers, finding out what is there, rather than presuming we know what a population is."

"We seem to start from very different perspectives," sighs Cavalli-Sforza, who adds that Wilson's type of grid survey would be "very difficult. You can't just sample whomever you come across." It might work in Europe, where people haven't moved around as much as they have, say, in southern Africa, he concedes, but in most places you simply won't find indigenous people every 50 miles. "Certainly we can find a compromise that takes into account both of our needs," Cavalli-Sforza adds diplomatically.

But Wilson says he is "digging in his heels" and will lobby for the grid strategy when he, Cavalli-Sforza, and others members of a new committee set up by the Human Genome Organisation meet later this summer to flesh out plans for the survey. King, for one, does not see the differences between the two as irreconcilable. "The goals are agreed on, and they arrived at them independently," she says. As for the strategy, "we just need to sort out which to use where. I think we need some of Wilson's grid, as well as Cavalli's populations in depth." Ultimately, she acknowledges, the amount of money the group raises will determine just how ambitious the survey will be. ■ L.R.

effort, though it would not pick up the entire tab. And that would be just fine with Cavalli-Sforza, Wilson, and their colleagues. They already have plans to persuade a consortium of agencies, including NIH, DOE, NSF, the equivalent agencies in other countries, as well as international organizations, like UNESCO, to each kick in some money.

Already, the Human Genome Organisation, or HUGO, has offered to help, urged on by its president, Sir Walter Bodmer, who calls the new survey "a cultural obligation of the genome project." Bodmer's enthusiasm is perhaps not surprising, as he wrote a classic text on population genetics with Cavalli-Sforza in the 1970s. While HUGO is not a funding agency, it will act as a broker to help raise funds, pledges Bodmer. Meanwhile, HUGO has set up a committee on human diversity, cochaired by Cavalli-Sforza and Marcello Siniscalco of the University of Sassari, Italy. The committee, which includes Wilson, Kidd of Yale, King of Berkeley, and others, will meet this summer to hammer out a firmer budget and gameplan.

If the project does get off the ground, it will bring to fruition a 7-year quest by Cavalli-Sforza and longtime collaborator Kidd, who first came up with this idea in 1984 while trapped on a transatlantic flight.

For years the two population geneticists had been studying the genetics of aboriginal populations in an attempt to reconstruct the movements of early peoples, whom they [illegible] with, and how they were related. But their research was handicapped because the "old style" genetic markers, such as the ABO blood group antigens, were "cumbersome and awkward to work with," says Kidd. These markers are used to compare how people vary at specific spots along their chromosomes, thereby creating genotypes or distinctive genetic "fingerprints."

What's more, there simply weren't very many of these markers around. Says King: "Ten years ago, you could determine the genotype of perhaps 50 loci [along the chromosomes]. That was the end of the story." Indeed, for that reason, Cavalli-Sforza and Kidd essentially stopped their phylogenetic studies in the early 1970s. "We ran out of data," says Kidd.

But by 1984, recalls Kidd, the new molecular tools coming on line promised to transform the field. He and Cavalli-Sforza were particularly excited by the new genetic markers, or DNA polymorphisms, that were being developed at the University of Utah and elsewhere. Not only did they provide a much more direct measure of genetic variation than the older methods, but they would soon be available in unprecedented numbers. "What's changed," says Kidd, "is the availability of thousands of genetic markers, scattered around the genome."

The irony was that just as the new markers and other techniques were becoming available, the populations Cavalli-Sforza and Kidd wanted to study were disappearing. What they needed to do, the two decided, was to collect DNA samples from members of indigenous groups immediately and preserve them. That could be done by collecting blood and then inducing the white cells to grow permanently in culture. Cavalli-Sforza and Anne Bowcock at Stanford and Kidd's team at Yale began that year, establishing cell lines from two groups of African Pygmies, whom Cavalli-Sforza had studied since the early 1960s, and others.

They have continued ever since, working on a shoe string. Says Kidd: "Both Luca and I have been frustrated by the difficulty of getting grants for this kind of work." Indeed, they have not mounted any special expeditions, which would be costly, but rather have persuaded anthropologist friends to collect blood samples for them, often meeting them at Kennedy Airport and rushing the samples off to the lab. So far, the Stanford and Yale teams have established permanent cell lines from individuals in 13 of the 250 populations the researchers identified on their 1984 flight. Clearly, they would need international help to complete the project.

Meanwhile, Wilson at Berkeley was pioneering the use of a new technique—the direct sequencing of mitochondrial, as opposed to nuclear, DNA—to study aboriginal populations in Africa and elsewhere. This led him to, among other things, his controversial theory about mitochondrial Eve. While pursuing these studies, Wilson also came up with essentially the same idea that Cavalli-Sforza and Kidd had. And by the late 1980s, Cavalli-Sforza and Wilson, who had earlier been on opposite sides of a heated debate on human origins, had begun

Yanomami People Threatened

Just 3 years ago, the Yanomami people were the largest group of native Amazonians still living in relative isolation in the jungles of Brazil. Almost 10,000 of them inhabited 125 villages, spread throughout a 94,191-square-kilometer region near the border of Venezuela. There they hunted and thrived in the tropical forest much as their ancestors had for thousands of years before them. But today, the Yanomami are a threatened people. According to anthropologists who have just returned from a fact-finding mission to Brazil, they will become extinct in the next decade if the Brazilian government does not move to protect them.

The warning is part of a hard-hitting report released this week—timed to coincide with Brazilian President Fernando Collor's visit to Washington—by the American Anthropological Association. Earlier this year, the association had taken the unprecedented step of sending a special commission of anthropologists to check out reports of the devastation of the Yanomami. What they found shocked even those prepared for the worst: Malaria and other diseases are killing the Yanomami at a rate of 13% per year and have thinned their ranks to 8000 already. As a result of the rampant malaria, fertility is near zero, and those people who have survived are sick and starving. "It's a desperate situation. A lot of villages have no more children and old people," says University of Chicago anthropologist Terence Turner, chair of the special commission.

What has visited this plague upon the Yanomami? The report lays the blame squarely at the feet of the Brazilian government, saying that it has failed to honor the Brazilian constitution, which guarantees Yanomami land rights, as well as its own promises to protect the indigenous people of the Amazon. Although the government had taken steps in the mid-1980s to turn the Yanomami territory into a special refuge, it later reneged on that promise and opened much of the region to miners of gold and the tin ore cassiterite.

In the gold rush that followed, as many as 40,000 miners invaded the region, using high-powered hoses and mercury to blast gold out of the soil. The mercury and soil runoff polluted the rivers and streams where the Yanomami fish, while the noise scared off the wildlife they hunted. Worse yet, the pools of stagnant water from the mining operations became breeding grounds for malaria-carrying mosquitoes. "Brazilian policies and economic activities have turned the land of the Yanomami into a death camp for its own people," says the report.

But there are some signs of hope. The report notes that President Collor, who assumed office in 1990, announced in April that he would revoke earlier decrees that had expropriated most of the land of the Yanomami. But he stopped short of giving the Yanomami legal title to the land, in favor of allowing the government to study the issue for 6 months. The anthropologists question the need for further study, as does the Environmental Defense Fund, which released a report of its own this week decrying the Brazilian government's policies on the Yanomami and other environmental issues. But Jose Goldemberg, Brazilian secretary for science and technology, said that the extra time was needed to figure out how much land to include in the reserve. The original Yanomami territory comprises 40% of the state of Roraima, which is home to about 500,000 other people, many of whom would have to be displaced for the reserve.

At press time, the anthropologists and environmentalists were hoping to meet with Collor during his Washington visit to discuss their report, which calls for the immediate return of the land to the Yanomami, the expulsion of miners from the region, and the provision of adequate medical supplies. No less is at stake than the future of one of the oldest Native American cultures.

■ ANN GIBBONS

talking and exchanging DNA samples, if not collaborating. Thus, it was only logical that they should team up on this new venture.

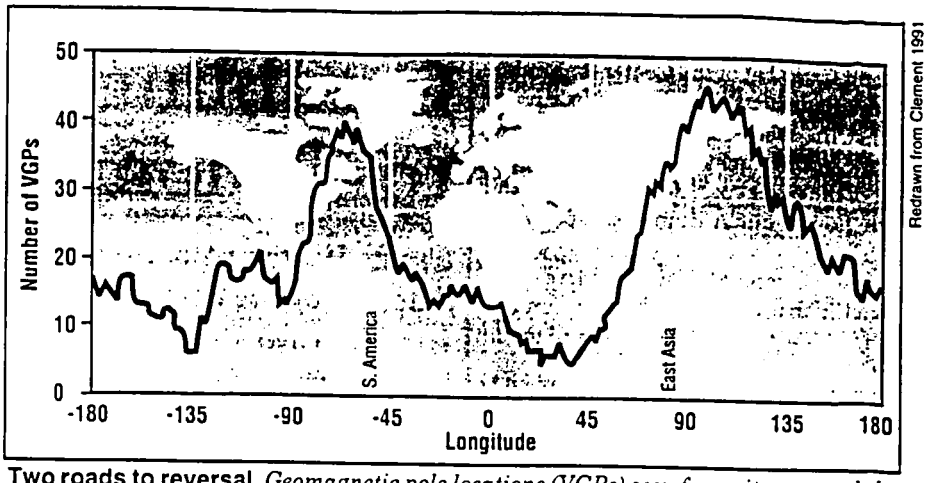
While the exact plan is still a matter of considerable debate (see box), the general idea is to collect blood and other tissue samples from 100 or so indigenous populations—in, for example, the Amazon, sub-Saharan Africa, and across southern Asia. "One hundred is a minimum," says Cavalli-Sforza, adding that the number could climb to 500, if money allowed. "Even at 100, you are leaving out some important populations." Within each group, the researchers would sample perhaps 100 individuals.

What they definitely don't envision is a planeload of Western geneticists descending on the jungle, collecting blood, and then disappearing. Rather, they hope to enlist the help of anthropologists, medical researchers, and local scientists who already have access to the more isolated groups. Already, says King, the group is thinking of what it can offer to the populations in return, such as medical supplies.

The bulk of the cost and the major technical obstacles will come from setting up the cell lines. The procedure itself is not complicated, he says, but the samples must be transported to a lab within a day or two—no small feat if they come from the upper reaches of the Amazon. The hope, he says, is to establish regional collection centers and train local workers in the technique, if need be. Eventually, all the cell lines would be collected in central repository, though the identify of each donor would be kept private.

Once the long-sought DNA is in hand, the group wants to determine genotypes of each individual for the same basic set of genes or DNA markers. This work might take several years and would probably be shared among numerous collaborators, says Cavalli-Sforza. After that, the anonymous DNA would be available to researchers around the world who wanted to use it, whether for investigating the population distribution of disease genes or for doing basic studies of evolution or human diversity.

Already, there are indications of the wealth of information harbored in the DNA of aboriginal peoples. Both Wilson and Cavalli-Sforza's data indicate that the prevailing view of race, which divides the world into blacks, whites, and so forth, is outmoded and mistaken, says Wilson. They and others have found that the genetic variation within a race is far greater than the variation between races. Says Berkeley colleague King: "The concept of race in America has a social meaning that does not correspond to its scientific meaning." She, like Wilson, predicts more surprises will emerge from their study of human variation. ■ LESLIE ROBERTS



Two roads to reversal. *Geomagnetic pole locations (VGPs) seen from sites around the globe cluster on two paths during a reversal some 730,000 years ago.*

A Core-Mantle Link?

Records of magnetic-field reversals point to a connection between the mantle and the underlying molten core

AT INTERVALS OF HUNDREDS OF THOUSANDS of years, Earth's magnetic field flip-flops, the north and south magnetic poles trading places. Geologists struggling to understand these reversals have concentrated on the churning liquid metal of the outer core, where the field is generated in the first place. They have paid little attention to the solid, rocky mantle encasing the core. Mantle and core, it seemed, were like ice floating on water—in contact but unlikely to intluence each other.

Now researchers at a variety of institutions around the globe have stumbled on a hint that the mantle does leave its mark on magnetic reversals. They have found that as the poles wander from south to north or vice versa during successive reversals, they show a startling tendency to trace out the same paths across the surface of the planet. Something about the processes that generate or modulate the field must be persisting for millions of years, through reversal after reversal, the researchers realized. And that could only happen, they feel, with the help of the mantle, a far less mercurial layer of the deep interior than the core. Of course, the precise connection between mantle and core is far from clear, but if proven and understood it might shed light on the larger mystery of why the reversals happen in the first place.

The intriguing discovery of repetitive behavior during magnetic reversals emerged as paleomagneticians collected more and more geologic records that caught the field in the act of flipping. Ocean sediments and lava flows, accumulating layer by layer, capture snapshots of the field's orientation as they form. Most such records show the field in one of two orientations—either the present one, with magnetic field lines looping out of Earth's south geographic pole and into the north pole, or the reverse. But every few hundred thousand years or so, sediments or lavas record the 4000- to 8000-year period during which the field switches orientation. And these rarer specimens told scientists an interesting tale.

"We all started to see this startling pattern," says Bradford Clement of Florida International University (FIU). The presumption had been that there would be no consistency at all. The core churns too rapidly, researchers had thought, for any memory of a reversal to linger until the next one. During each reversal, the north magnetic pole as viewed from a given site should appear to follow a random path from one geographic pole to the other.

To Clement and colleagues it was an "astonishing thing" to find that the poles actually tend to follow one of two well-worn routes. One leads through North and South America; the other, less heavily traveled, path stretches across eastern Asia and Australia. Eric Tric of the Center for Studies in Weak Radioactivity in Gif-sur-Yvette, France, recently pointed out that fully two-thirds of 48 reversal records from the past 3 million years show the poles following one of these two paths. Despite its short memory, the core seems to recall where its magnetic poles should go during reversals hundreds of thousands or even millions of years apart.

The paleomagneticians suspect it is the mantle that reminds the core where the preferred pole paths are. In independent papers appearing this month, Clement of

...all for a Worldwide Survey of ...uman Genetic Diversity: A Vanishing Opportunity for the Human Genome Project

L. L. CAVALLI-SFORZA,*,[1] A. C. WILSON,† C. R. CANTOR,† R. M. COOK-DEEGAN,‡ AND M.-C. KING†

Stanford University, Stanford, California; †University of California, Berkeley, California; and ‡National Institute of Medicine, National Academy of Sciences, Washington, D.C.

The Human Genome Project can now grasp a vanishing opportunity to preserve the record of our genetic heritage. A major goal of the Human Genome Project is to create biological tools that will permit access to any region of a human genome. One of the more important reasons to do this is to understand human diversity, both normal variation and that responsible for inherited diseases. The genetic diversity of people now living harbors the clues to the evolution of our species, but the gate to preserve these clues is closing rapidly. We call upon geneticists and public and ...ate agencies to collaborate now in collecting sufficient ...erial to record human ethnic and geographic diversity ...fore this possibility is irretrievably lost.

Human genomes that exist today have been determined by historical population structure and dynamics (Cavalli-Sforza *et al.*, 1992; Wilson, 1990). Hence, information from nuclear and mitochondrial genes from present-day populations worldwide can document prehistoric migrations, natural selection, the social structure of populations, and the frequency and types of mutations our species has experienced (Bowcock *et al.*, 1991). The novel perspective of genetics can supplement and strengthen findings from archeology, linguistics, and history (Cann *et al.*, 1987; Cavalli-Sforza *et al.*, 1988, 1992; DiRienzo and Wilson, 1991).

The populations that can tell us the most about our evolutionary past are those that have been isolated for some time, are likely to be linguistically and culturally distinct, and are often surrounded by geographic barriers (Cavalli-Sforza *et al.*, 1992). Isolated human populations contain much more informative genetic records than more recent, urban ones. Such isolated human populations are being rapidly merged with their neighbors, however, destroying irrevocably the information needed to reconstruct our evolutionary history. Population growth, famine, war, and improvements in transportation and communication are encroaching on once stable populations. It would be tragically ironic if, during the same decade that biological tools for understanding our species were created, major opportunities for applying them were squandered.

We must act now to preserve our common heritage. Preserving this historic record will entail a systematic, international effort to select populations of special interest throughout the world, to obtain samples, to analyze DNA with current technologies, and to preserve samples for analysis in the future. Recent advances in techniques for establishing permanent cell lines and for obtaining DNA by amplification of very small samples of blood, hair, or other tissue make a collection program feasible. Cell lines have already been established for 20 to 80 individuals from each of 13 populations from all parts of the world (Bowcock *et al.*, 1991; Cavalli-Sforza *et al.*, 1986).

Logistical obstacles to collecting representative samples from the world are daunting, but could be surmounted with sufficient planning and financial commitment. The most expensive portion of the project will be sample collection, processing, and long-term storage. Because the populations of greatest interest are far from airports and modern laboratories, regional facilities to process samples will be necessary. This problem could be solved by providing equipment and training to centers in appropriate locales. The final collection of samples could be distributed among facilities such as the Coriell Institute in the United States, the UNIDO-supported biotechnology research centers in New Delhi and Trieste, and designated facilities in Latin America and the Middle East. The need for dedicated storage facilities is already apparent (Bowcock *et al.*, 1991).

Once collected, DNA from cell nuclei and from mitochondria can be analyzed to illuminate variation, selection, population structure, migration, mutation frequency, mechanisms of mutation, and other genetic events of our past (Bowcock *et al.*, 1991; DiRienzo and Wilson, 1991; Cavalli-Sforza *et al.*, 1986, 1992). These processes can be examined quite inexpensively by determining the genotype of each individual at perhaps 100 loci selected for different mutation rates (Cavalli-Sforza, 1991).

A complete survey would be worldwide and geographically comprehensive. Among indigenous populations of great interest are peoples of the Sahara, of eastern, western, and southern Africa (Vigilant *et al.*, 1989), the Etas of Japan, insular populations in Malaysia and southeastern Asia (Stoneking and Wilson 1989; Stoneking *et al.*, 1990), ethnic minorities of China, Polynesians, aboriginal populations of Australia and Melanesia, the Kurds of eastern Turkey, peoples of the Caucasus, the Lapps, the Basques, other peoples in the Pyrenees, Appenines, Carpathians, and Alps, and the many indigenous American populations. Among these very informative groups have been many peoples historically vulnerable to exploitation by outsiders. Hence, asking for samples alone, without consideration of a population's needs for medical treatment and other benefits, will inevitably lead to the same sense of exploitation and abandonment experienced by the survivors of Hiroshima and Nagasaki (Sekimori, 1988). It will be essential to integrate the study of peoples with response to their related needs.

We call for a concerted effort to obtain and store samples

[1] To whom correspondence should be addressed.

...88-7543/91 $3.00

m diverse populations in order to understand human iation. We urge national funding agencies to gaze favorly on proposals to collect and preserve DNA samples from human populations worldwide. In the United States, the relevant agencies include the National Institute of General Medical Sciences, the National Center for Human Genome Research, the National Science Foundation, and the genome program at the Department of Energy. Medical and biological research funding agencies in other countries can also assist in the effort. We urge that international organizations such as UNESCO, WHO, and UNIDO consider supporting such efforts and that the Human Genome Organization (HUGO) assist in promoting and coordinating a worldwide program. HUGO has asked us to prepare a proposal for a worldwide genetic survey. We hope colleagues interested in joining the effort will be in touch with us. The potential medical gains from the Human Genome Project are immense, and the benefits these will bring are enormous. The potential intellectual benefits of understanding human diversity and its origins are equally striking. By an intense scrutiny of human diversity, we will make enormous leaps in our grasp of human origins, evolution, prehistory, and potential.

REFERENCES

1. BOWCOCK, A. M., KIDD, J. R., MOUNTAIN, J. R., KIDD, K. K., AND CAVALLI-SFORZA, L. L. (1991). Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* **88**: 839–843.
2. CANN, R. L., STONEKING, M., AND WILSON, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* **325**: 31–36.
3. CAVALLI-SFORZA, L. L. (1991). How can one study individual variation for 3 billion nucleotides of the human genome? *Am. J. Hum. Genet.* **46**: 649–651.

3a. CAVALLI-SFORZA, L. L., KIDD, J. R., KIDD, K. K., BUCCI, C., BOWCOCK, A. M., HEWLETT, B. S., AND FRIEDLAENDER, J. S. (1986). DNA markers and genetic variation in the human species. *Cold Spring Harbor Symp. Quant. Biol.* **51**: 411–417.

4. CAVALLI-SFORZA, L. L., PIAZZA, A., AND MENOZZI, P. (1992). "History and Geography of Human Genes," Princeton Univ. Press, in press.
5. CAVALLI-SFORZA, L. L., PIAZZA, A., MENOZZI, P., AND MOUNTAIN, J. (1988). Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* **85**: 6002–6006.
6. DIRIENZO, A., AND WILSON, A. C. (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **88**: 1597–1601.
7. SEKIMORI, G. (1988). "Hibakusha: Survivors of Hiroshima and Nagasaki," Kosei, Tokyo.
8. STONEKING, M., JORDE, L. B., BHATIA, K., AND WILSON, A. C. (1990). Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* **124**: 717–733.
9. STONEKING, M., AND WILSON, A. C. (1989). Mitochondrial DNA. *In* "Colonization of the Pacific: A Genetic Trail." (A. Hill and S. Serjeantson, Eds.), pp. 215–249, Oxford Univ. Press, Oxford.
10. VIGILANT, L., PENNINGTON, R., HARPENDING, H., KOCHER, T. D., AND WILSON, A. C. (1990). Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci. USA* **86**: 9350–9354.
11. WILSON, A. C. (1990). Will sequencing the human genome revolutionize biology? *New Biologist* **2**: 585–586.

The Study of Variation in the Human Genome

ANNE BOWCOCK* AND LUCA CAVALLI-SFORZA†

*Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas 75235-9063; and †Department of Genetics, Stanford University, Stanford, California 94305

The human genome project is under way, with its final goal the determination of the sequence of approximately 3 $\times$ 10^9 bp that constitute the haploid DNA component of every human somatic cell. The goal is immense and eventually every human gene and regulatory sequence will be identified. What will not necessarily emerge from the sequence is which regions of the genome vary, both between individuals and between populations. At present, over 2300 DNA polymorphisms in the human genome have been described (Williamson *et al.*, 1990), but this is likely to be less than one-thousandth of the variation that exists. Knowledge of the variability within these known polymorphic regions will provide some insight into other potentially variable regions, but in many instances the variation within much of the genome will be hidden and unpredictable.

A study of human variation provides information on our evolutionary history and on fundamental genetic mechanisms such as selection, mutation, and genetic drift and is essential for analyses involving individual DNA fingerprints such as in forensic science.

DNA Sequence Variation in Different Human Populations

The study of variation is the basis of the study of evolution, one of the principal aims of genetics. Studies of genome diversity can also provide information on phenomena such as gene conversion and allow one to identify regions that may have functional importance.

Most studies of the diversity of RFLPs have been made on Caucasoid samples for obvious reasons of expediency. These have been limited to the determination of gene frequencies, heterozygosities, and PICs (Botstein *et al.*, 1980). The use of an appropriate sample of individuals drawn from a worldwide net would increase the variation available for study.

So far, most of the work on human evolution performed with studies on populations from different parts of the world has been carried out on non-DNA ("classical" or protein) polymorphisms. Essentially, these behave in the same fashion as DNA polymorphisms, but are much less numer-

0888-7543/91 $3.00

recorded mammalian mutation rate is in an allele of a class I gene from the [redacted] MHC (ref. 15), probably [redacted] by a specialized mutational [redacted]m? He did not, however, refer to the ultimate extension of his position, namely, that although infectious disease may be the principal cause of polymorphism, it would be virtually ineffective in the absence of segregation and, for multiple loci, independent assortment. In other words, the sexual mechanism itself is an essential ingredient of the resistance system. Which came first? Most of us have been brought up with the idea that sex causes disease. A longer term view, and one for which there is increasing support, is that the boot belongs on the other foot[16,17]. □

J. C. Howard is in the Department of Immunology, AFRC Institute of Animal Physiology and Genetics, Babraham, Cambridge CB2 4AT, UK.

POPULATION GENETICS

A way to world knowledge

Jared M. Diamond

THINK of what we could deduce if we knew the DNA base-pair sequence of every human now alive. We could reconstruct much of the history of modern human populations in terms of prehistoric migrations and founder effects. We could study the social structure of human populations, the frequencies and types of human mutations, and natural selection in humans along environmental gradients. We could understand our differing genetic susceptibility to many diseases. Although that goal of completely sequencing everyone is unfortunately not [redacted] a flurry of papers shows that a [redacted]ser enterprise could yield a [redacted]al of information[1–3].

The immediate impetus comes, of course, from the Human Genome Project, which seeks to sequence completely what amounts to one Caucasian's genome. That endeavour does not in itself address the questions outlined above, but Cavalli-Sforza *et al.*[1] argue that with a mere one per cent of extra effort and expense it could. Two practical issues arise: how to allocate resources among populations (for example do we learn more by sequencing the genomes of 500 individuals from each of 20 populations, or those of 20 from each of 500?); and how large a fraction, and which specific fractions, of each individual's genome to sequence. If we had already sequenced everybody, we could then with hindsight design an efficient sampling strategy. The dilemma is that we don't know the results, yet we have to guess at them to design the strategy.

An essential consideration that Cavalli-Sforza *et al.* point out stems from the main pattern of human history for the past 10,000 years. By virtue of acquiring plants and animals suitable for domestication, certain groups that [redacted] constituted only tiny fractions [redacted]orld's human population have [redacted] much of the globe. Prime examples are the expansions of Anatolian farmers and proto-Indo-European speakers over Eurasia, of Austronesians over Indonesia and the Australian realm, of Bantus over sub-Saharan Africa, and of modern Europeans over Australia and the Americas. Conversely, other populations that once accounted for much of human diversity and population numbers are now close to vanishing, overrun by the expansions of others. Cavalli-Sforza *et al.* answer the first of the two practical issues by proposing a crash programme for collecting samples (permanent cell lines, DNA or both) from these vanishing populations. With that material secured, the matter of how to study it could then be addressed at leisure.

Any geneticist or anthropologist could quickly come up with a wish-list of declining human populations to sample. High on anyone's list would come the !Kung of Namibia and Botswana, remnants of the population that occupied most of southern Africa until a few millennia ago, and located at the deepest branches in the world tree of mitochondrial DNA (mtDNA)[4]; scattered tribal peoples of southeast Asia and Indonesia, probably remnants of that area's former population related to native Australians and Melanesians and swamped by the Austronesian expansion; and Japan's Ainus, remnants of who-knows-what. Other relict populations persist in regions protected by mountainous terrain such as the Pyrenees and Caucasus. Still other priorities would be areas of high genetic diversity such as New Guinea, home to one-fifth of all the world's languages and (especially in the highlands) to an unknown fraction of the world's genetic diversity.

This wish-list begs the question of whether to sample a few dozen areas in detail or else hundreds in less detail. Here, too, the best strategy depends on the incompletely known patterns of genetic diversity that we seek to discover. For example, mtDNA sequencing detected no geographical variation within the peoples of the Middle East, but abundant variation in those of the New Guinea highlands and among African pygmies[2,4,5]. Similarly, the number of people to sample from each population depends on intrapopulation variability, which surely differs not only among populations but also among the genetic loci studied.

Hence one's sampling scheme will depend on the particular problem to be tackled. For instance, at one extreme mtDNA is especially variable, the control region of mtDNA is hypervariable, and the first 400 base pairs of that region contains most of its hypervariability. Of 88 mtDNA types from the control region derived from 117 Caucasian people, only 12 were shared by two or more individuals, and 10 of those 12 shared types were shared between Sardinian individuals or between Middle Eastern individuals[2]. Again, identical mtDNA types were shared between individual !Kung or Biaka Pygmies or Mbuti Pygmies, but not between two or more of those populations[4]. At the opposite extreme, a study of nuclear DNA in Pygmies, Europeans, Chinese and Melanesians identified a polymorphism (*HP*/*Bam*HI) whose gene frequency ranged only from 0.35 (in Chinese) to 0.46 (in Zaire pygmies)[3]. A larger population would be required to study variation in *HP*/*Bam*HI than in the mtDNA control region.

By analogy, imagine trying to reconstruct the evolution of the animal kingdom if some species such as pigeons and house mice had undergone a population explosion, but if most terrestrial animals weighing over ten kilograms and most organisms inhabiting tropical rainforest had just vanished (which indeed may come to be the case before too long; the argument for sampling the genomes of such species is also compelling). The human problem is partly similar, but only partly. Although some threatened human populations may vanish because of the deaths of most members, more will vanish as their members become absorbed into other populations[6]. But attempts to sample the genomes of animals and humans have in common a need to reconcile our intellectual desire to know with our ethical desire to preserve. Above all, both share a sense of urgency, because so little time remains before it will be too late. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

1. Cavalli-Sforza, L. L., Wilson, A. C., Cantor, C. R., Cook-Deegan, R. & King, M. C. *Genomics* (in the press).
2. Di Rienzo, A. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1597–1601 (1991).
3. Bowcock, A. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 839–843 (1991).
4. Vigilant, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 9350–9354 (1989).
5. Stoneking, M. *et al. Genetics* **124**, 717–733 (1990).
6. Cavalli-Sforza, L. L. *et al. Cold Spring Harbor Symp. quant. Biol.* **51**, 411–417 (1986).

heavier atoms of oxygen in the impacting cluster. During the impact of a cluster, Vandenbosch suggests, deuterium atoms might backscatter only to be hit like a base- [redacted] by relatively massive oxygen atoms still [redacted]eening forward. Several whaps like that [redacted]ight boost the deuterium's energy enough to cause fusion with another nearby deuterium nucleus, he adds. That mechanism, Friedman notes, could explain why the all-deuterium clusters in the Lyons experiments, which lacked heavier atoms such as oxygen, yielded no evidence of fusion.

Bae, Yeong Kim of Purdue University, and colleagues at SRI and the Electric Power Research Institute have developed a different scenario, which relies on shock waves to concentrate the collision's energy. They propose that the shock waves heat nanometer-sized regions of the target to temperatures of stellar interiors—hundreds of millions of degrees. Fusion would occur in the tiny compressed pockets of plasma that result, Kim says. "It's like having lots of tiny microsuns," he muses.

Beuhler, Friedman, and Friedlander conjure up another analogy, proposing that the process might be more like the one at work in an ore-blasting explosive charge or an armor-piercing shell. Such "shaped charges" can channel energy so that particles emerge [redacted]om the explosions in jets traveling many [redacted]s faster than the detonation wave. Like[redacted], the tiny cavities chiseled into the target [redacted] the impacting clusters might serve to confine atoms under huge compressions while amplifying their energies to fusion-triggering levels.

The Brookhaven trio and their growing ranks of allies freely admit that they're speculating, but they think it won't be long before experiment catches up with them. The ingredients of a bona fide area of fusion research are now in place, after all. Experiments continue at Brookhaven. More are under way, or in the works, at other labs in the United States, France, the Netherlands, and perhaps Japan. Theorists are theorizing. And skeptics are keeping everyone on their toes.

At the moment, most people in the field are pursuing little more than the thrill of basic scientific discovery. But press them a little, and they will admit that somewhere in the back of their minds lurks the possibility that their research could someday harbor payoff to people who have never heard of accelerators, deuterium, and fusion.

In that vein, Friedlander, Beuhler, and Friedman closed their first *Physical Review Letters* paper with a remark as gingerly [redacted]lated as it was bold: "The high fusion [redacted]and the sensitivity to projectile energy [redacted]gest the possibility of a possible new path to fusion power." ■ IVAN AMATO

Genetic Survey Gains Momentum

Last summer population geneticist Luca Cavalli-Sforza of Stanford University, molecular anthropologist Allan Wilson of the University of California, Berkeley, and others issued a call to action: an urgent plea for help—and money—to collect DNA samples from aboriginal populations around the world before those groups vanish. Now, just a few months later, even the proponents of this bold new plan seem amazed at the response.

As word gets out, numerous anthropologists are offering to help collect samples from the isolated tribes they study. And in an unexpected twist, several federal agencies have approached the scientists—unsolicited—to see how they can help. Indeed, the agencies are already talking about picking up at least part of the tab, which could run to $20 million or more over the next 5 years.

The basic plan is to collect blood samples from members of at least 100 indigenous populations, such as the Bushmen of southern Africa and the Hill People of New Guinea. Such populations, isolated for hundreds or thousands of years, contain in their genes clues to human evolution, migration, and diversity. But the opportunity to analyze those genes is rapidly vanishing as society encroaches upon these once-distinct peoples. Once the samples are collected—probably from about 50 individuals in each group—the researchers would establish permanent cell lines to preserve the DNA in perpetuity, allowing it to be studied even after the tribes have disappeared (*Science*, 21 June, p. 1614).

Walter Bodmer, president of the Human Genome Organization (HUGO), was keen on the idea as soon as he learned of it, setting up a committee headed by Cavalli-Sforza and Marcello Siniscalco of the University of Sassari, Italy, to firm up the scientific strategy and the budget. The group was dealt a tragic blow last July, when Wilson died of leukemia following a bone marrow transplant. Shortly thereafter, Cavalli-Sforza was taken ill. As he recuperates, the dispersed committee has been doing its best to cobble together a proposal for both national and international funding agencies.

But with the proposal still incomplete, Cavalli-Sforza received a letter from Mark Weiss, who runs the physical anthropology program at the National Science Foundation (NSF). Weiss, who had read about the plan in *Science*, said that although his own research budget is too small to make much of a dent in the total cost, he thought his and other NSF programs could provide at least partial support. In late September Weiss brought together representatives from other potential funding sources as well: the genome projects at both the National Institutes of Health and the Department of Energy, and the National Institute of General Medical Sciences. "Everyone is excited," says Weiss. "No one said in stone that they would fund the project, but the general consensus is we are looking forward to receiving a formal proposal."

In response, Cavalli-Sforza and the HUGO committee are furiously revising and fleshing out their proposal into what they call a "grand vision" of the project. As they do so, both scope and cost are growing. The group is now talking about collecting DNA from 200 to 500 populations at a cost of several million dollars a year, double what they were thinking just last summer.

Before they start sampling, though, they'll have to resolve some strategic questions. Cavalli-Sforza and Wilson were deeply divided on the sampling strategy, with Cavalli-Sforza advocating sampling populations that have been isolated in geographic pockets, and Wilson proposing instead setting up a grid and sampling every 50 or 100 miles (*Science*, 21 June, p. 1615). "What is the best way to sample the world? Allan and I had different views," says Cavalli-Sforza. "I have been thinking a lot about a compromise, but I want to hear opinions of theoreticians."

He plans to bring together statisticians, mathematicians, geneticists, and anthropologists to tackle that issue in a workshop, perhaps as early as this winter. A second workshop will bring in physical and cultural anthropologists to help identify which populations to study, and which ones should come first. That's an urgent question, because for some groups, it is almost too late already.

Weiss thinks funding for those workshops is likely to be forthcoming from U.S. agencies. If enough money for the rest of the project materializes from U.S. and international sources, Cavalli-Sforza and his colleagues think they can collect all the samples within 2 to 3 years and establish the cell lines within 5. Then would begin the long-term analysis to tease out the DNA's secrets. ■ LESLIE ROBERTS