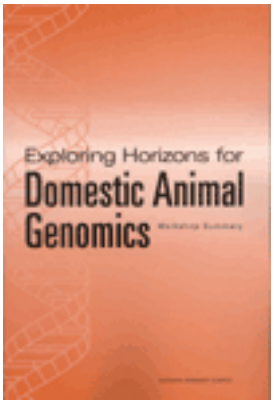


Exploring Horizons for Domestic Animal Genomics: Workshop Summary

Robert Pool, Kim Waddell, National Research Council

ISBN: 0-309-50100-8, 54 pages, 6 x 9, (2002)

Exploring Horizons for Domestic Animal Genomics

Workshop Summary

BOARD ON AGRICULTURE AND NATURAL RESOURCES

BOARD ON LIFE SCIENCES

DIVISION ON EARTH AND LIFE STUDIES

National Research Council

By Robert Pool and Kim Waddell

National Academy Press

Washington, D.C.

NATIONAL ACADEMY PRESS 2101 Constitution Avenue, NW Washington, DC 20418

NOTICE: The project that is the subject of this summary was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the planning group responsible for the planning of the workshop were chosen for their special competences and with regard for appropriate balance.

This workshop was supported by agreement numbers 59-0790-1-137, 2001-38831-11434, 263-MD-203182, DE-FG02-02ER63368, and IBN-0136019 between the National Academy of Sciences and the U.S. Department of Agriculture (ARS and CSREES), the National Institutes of Health, the Department of Energy, and the National Science Foundation (any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-08505-5

Additional copies of this summary are available from National Academy Press, 2101 Constitution Avenue, NW, Lockbox 285, Washington, DC 20055;
(800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area);
Internet, http://www.nap.edu.

Printed in the United States of America.

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

PLANNING GROUP FOR THE WORKSHOP ON EXPLORING HORIZONS FOR DOMESTIC ANIMAL GENOMICS

CLAIRE M. FRASER, Chair, The Institute for Genomic Research, Rockville, Maryland
RONALD L. PHILLIPS, University of Minnesota, St. Paul, Minnesota
ELAINE A. OSTRANDER, Fred Hutchinson Cancer Research Center, Seattle, Washington
LAWRENCE B. SCHOOK, University of Illinois, Urbana, Illinois
JAMES E. WOMACK, Texas A&M University, College Station, Texas

STAFF

KIM WADDELL, *Senior Program Officer*
JOAN ESNAYRA, *Program Officer*
MICHAEL R. KISIELEWSKI, *Research Assistant*
CINDY LOCHHEAD, *Project Assistant*
ROBERT POOL, *Science Writer*

BOARD ON AGRICULTURE AND NATURAL RESOURCES

STAFF

BOARD ON LIFE SCIENCES

STAFF

Acknowledgments

This workshop summary was enhanced by the contributions of many individuals who graciously offered their time, expertise, and knowledge. The committee thanks all who attended and/or participated in the public workshop (Appendixes B and C).

This summary has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We thank the following individuals for their review of this summary: Pieter de Jong of the Children's Hospital Oakland Research Institute, Oakland, CA; Vivek Kapur of the University of Minnesota, St. Paul, MN; and Bruce A. Roe of the University of Oklahoma, Norman, OK.

Contents

Boxes

1

Introduction

Over the past several years, scientists from the United States and around the world have been using a technique called DNA sequencing to unlock the genetic code of many different organisms. With code in hand, scientists can design sophisticated experiments that will inform our understanding of how an organism develops and functions. To date, they have carried out partial or complete DNA sequencing on human, mouse, rat, bacterial, and plant genomes. A major finding that was confirmed from these efforts is that most biologic functions are genetically conserved within and between species. This means that by studying related organisms, we acquire biologic knowledge that is broadly applicable. Sequencing the genomes of different kinds of organisms sheds more light on biologic understanding than would, for example, sequencing the genome of only a single type of organism. In essence, the more organisms that are sequenced, the greater the intellectual yield will be. Once sequenced, the DNA provides scientists with important clues about the genes and proteins that are required to create and sustain related organisms.

Why not sequence every organism scientists can get their hands on? Unfortunately, sequencing is very expensive and time consuming. At this time, obtaining a draft sequence of a mammalian genome costs as much as 100 million dollars and can take up to a year. Consequently, scientists, policymakers, and funding agencies must select carefully among the organisms they will sequence, depending on the kind of knowledge sought. Two major categories of animals that are primed for extensive genomic exploration are domesticated farm animals such as pigs, sheep, chickens, cattle, horses, and companion animals such as dogs and cats. Both groups have a number of appealing attributes that make them appropriate candidates for genomic analysis.

They offer great potential for advancing human and animal health knowledge, improving animal production practices, and they have an economic benefit. For example, scientists can attempt to locate or approximate the gene or genes that confer disease resistance, which is useful in reducing health maintenance costs in animal production operations. Moreover, in some instances these animals have a sentimental value that distinguishes them from other organisms.

Recognizing the important contributions that genomic analysis can make to agriculture, production and companion animal science, evolutionary biology, and human health with respect to the creation of models for genetic disorders, the National Academies convened a group of individuals to plan a public workshop that would: (1) assess these contributions; (2) identify potential research directions for existing genomics programs; and (3) highlight the opportunities of a coordinated, multi-species genomics effort for the science and policymaking communities. Their efforts culminated in a workshop sponsored by the U.S. Department of Agriculture, Department of Energy, National Science Foundation, and the National Institutes of Health. The workshop was convened on February 19, 2002. The goal of the workshop was to focus on domestic animal genomics and its integration with other genomics and functional genomics projects (see Box 1-1). The following is a summary and synthesis of the discussion, prepared by a science writer as a factual account of what occurred at the workshop.

Box 1-1 Goals of the Workshop

Experts from recently completed genome sequencing projects as well as those engaged in current efforts, along with policymakers and stakeholders, were brought together to participate in a workshop designed to:

1. provide a forum for exchange among the diverse communities of genomics and functional genomics experts;
2. elicit discussions of research directions in animal genomics that would benefit agriculture and society while leading to greater biologic understanding;
3. identify opportunities and obstacles that might be encountered in developing a coordinated, multi-organism functional and comparative genomics effort that would include domestic animals.

2

The Value of Sequencing Domestic Animal Genomes

Sequencing the genomes of domestic animals could be beneficial to animal production practices, animal health and welfare, and to our understanding of the genetic basis of diseases in both animals and humans. Beyond these more applied areas of study, sequencing genomes also presents opportunities for increasing our basic knowledge of the evolutionary pathways of these and related species. The precise benefits will vary somewhat from species to species, but in general they fall into three categories.

SEQUENCING FOR AGRICULTURE

The first, and most familiar category is the array of economic benefits that farmers, ranchers, and pet owners could expect from the genetic sequencing of their animals. For thousands of years, these animals have been bred for desirable traits, including disease resistance and rapid growth in farm animals, and the color of the coat or shape of the head in pets. The precision of traditional, selective breeding is low and genetic change is poorly characterized. Theoretically, with information from a sequenced genome, it will be possible to have much more precision with breeding efforts and even to genetically engineer specific traits by adding, removing, or altering individual genes.

In agriculture, the traits of interest are primarily *production* traits, noted Steven Kappes, of the U.S. Department of Agriculture's Agricultural Research Service (USDA-ARS). "We have quite a list of traits that we look at within farm animals," said Kappes, director of U.S. Meat Animal Research Center in Clay Center, Nebraska. "These include growth—both prenatal and postnatal—reproduction, egg production, and carcass traits, including fat deposition within

and between the muscles, inside organs, and layered under the skin." Other production traits are meat tenderness and palatability, as well as milk production.

Kappes pointed out that in many cases scientists already have found the general location on a chromosome for a gene that expresses a particular trait in an animal. The technical term for the general location of a gene, which affects a particular trait that is measured on a quantitative scale, is a "quantitative trait locus," or QTL. "We have in excess of 30 QTLs in cattle for these different traits," Kappes said, plus a similar number in pigs and a dozen or so in chickens. Once a particular animal genome is sequenced, it might be possible to determine which gene or genes affect a trait, and thus give breeders the information they need to enhance the production traits of that animal.

SEQUENCING FOR ENHANCED BASIC SCIENTIFIC UNDERSTANDING

Besides the agricultural benefits, genomic sequencing of domestic animals will be important in a number of areas of basic science, particularly in understanding the evolutionary relationships between species. "Nothing in biology really makes sense except in light of evolution," commented Stephen O'Brien, chief of the Laboratory of Genomic Diversity and head of the Section of Genetics at the National Cancer Institute (NCI). The genome of each species is the end result of millions of years of mutation and natural selection. The genome of every mammal alive today, for instance, can be traced back to the genome of an ancestral mammal that lived some 200 million years ago, and the genomes of the different species provide a record of how the descendants of that proto-mammal gradually diverged into many different forms, as well as a guide to how today's mammals are related to one another.

When we explore the evolution of a living species, we assume that its presence here today is the result of that species successfully adapting and negotiating the myriad of ecologic and environmental challenges over time, O'Brien explained. "Nestled in the genomes of living species are the historic footprints of the adaptive events that led them to where they are today." In other words, comparing the genomes of various mammals alive today might be the best option for understanding how the species evolved as they did.

Several other evolutionary questions can be addressed by sequencing a variety of mammals, O'Brien said. "We don't know which of the genes make us human, as opposed to apes or as opposed to non-ape primates, or as opposed to other orders of mammals." Only by sequencing the genomes of other animals and comparing them gene by gene with the human genome will it be possible to answer this fundamental question. Some preliminary comparisons between humans and other mammals already have been made, said Harris Lewin of the

University of Illinois, and they illustrate the kinds of discoveries that could be made by comparing full genomes. "When we do the comparisons, we find things in other species that are either very rapidly evolving or are completely missing from the genomes of humans. In the mouse, for example, there are 200 to 300 genes that are not present in humans or cattle." By studying these genes, Lewin said, researchers could uncover ways in which species evolution diverged onto two or more different paths to achieve similar metabolic functions.

Furthermore, O'Brien said, comparing the ways in which evolution has structured different genomes should help researchers uncover the logic of that organization. "We don't really know," he said, "why the genes are arranged in the way that they are—why they're next to the ones that they are. We have some clues in certain cases where the genes are clustered, but by and large we don't really understand whether it was a random process or whether there was an adaptive value to it." By having a number of genomes to compare, researchers might be able to find patterns in this structure across different species of mammals and speculate as to the arrangement of genes.

SEQUENCING FOR HUMAN HEALTH AND MEDICAL RESEARCH

A third category of benefits to sequencing domestic animal genomes could have more immediate practical applications. When the human genome was sequenced, it was hailed as a major step toward finding new medical treatments and other means of benefiting human health, but it was only one step, and there is much that remains unknown about the human genome and how it structures human development. By sequencing the genomes of other mammals, biomedical researchers seek to answer more of the remaining questions about the human genome and its potential for improving human health.

What remains unknown about the human genome? First, although the sequencing of the genome allows researchers to determine the genes that are characteristic of humans, the functions of most of those genes remains unknown. According to O'Brien, "of the 30,000-odd genes that have been identified by various algorithms," only about 8,000 have been named and their functions identified. Furthermore, he noted, the genes make up only part of the genome, and the remainder is even more mysterious. "The genes are nested in a sea of non-coding regions, including cryptic regulatory elements, promoters, enhancers, silencers, transcription factor binding sites and all kinds of interesting features that have been discovered and are yet to be discovered."

Once a genome has been sequenced, there is still much more to do and many questions to address, noted O'Brien. "And one of the ways in which we are hoping to approach some of these questions is through applications of a comparative sense." That is, by studying the genomes of other species,

researchers expect to make inferences based on what is not yet understood about the human genome.

To determine the function of human genes, for example, researchers can look for similar genes in other animals whose functions are known. If scientists have identified a particular gene in the mouse and know what it does, they can search the human genome for a gene with a similar sequence and surmise that the human gene probably has a function similar to the one in the mouse. This is part of annotating—or creating a set of comments, notations, and references describing the experimental and inferred information about a gene or protein. In its most elementary form, the human genome may be described as a shorthand list of three billion "letters"—A, T, C, and G—each of them representing one of four nucleic acid *bases*: adenine, thymine, cytosine, and guanine, respectively. Bases are small, nitrogenous molecules, which in deoxyribose nucleic acid (DNA) occur in pairs (base pairs). For example, in DNA, only adenine/thymine, and cytosine/guanine, can pair together. But this base sequence information, or code, is useful only to the extent that researchers can interpret and apply it, which is why having other genomes available to study is so valuable.

Harris Lewin offered one example of how the genome of another organism can benefit the understanding of the human genome. He compared a long stretch of human DNA with the corresponding stretch of DNA in the cow, looking for similarities and differences. Because humans and cattle had a common ancestor—although it was some 60 million years ago—it is generally possible to align a stretch of DNA from one species with a stretch in the other that shares the same genes and other features, and the pattern of similarities and differences between the two stretches is very informative to the educated eye. Over those 60 million years, random mutations accumulated in both the ancestral line that led to humans and the one that led to cows, so that many of the base pairs in their DNA are no longer the same. But some pairs are much more likely to change than others. The bases that make up a gene, for instance, are relatively resistant to change because most changes in the gene (known as mutations) are detrimental to the animal—some are even fatal—and so natural selection tends to conserve the pattern of bases in a gene. By contrast, in stretches of DNA that have no apparent purpose—for instance, old genes that are no longer functional—the mutations will accumulate unabated. The result is a clear pattern of similarities and differences in the DNA of the two species. Similar patterns suggest that an important function has been conserved between the two species; different patterns suggest that function is not strongly dependent upon the particular sequence of base pairs in that segment of the DNA.

When Lewin compared the two corresponding stretches of DNA—one from humans and one from cattle—he discovered something interesting. "There is a 12-kilobase region (i.e., one that is about 12,000 base pairs long) and if you

go through and you annotate the human genome sequence, you see absolutely that there are no annotated genes in this region. So what is this region doing?" There is clearly something important about that region, something so vital to the functioning of the two species that its sequence has been mostly conserved over 60 million years, but nobody knows what it is.

Indeed, Lewin said, comparisons between the mouse and the human genome have shown that "only 56 percent of the conserved sequences between the mouse and the human could be accounted for by known features of genes." In other words, 44 percent of the DNA that has been conserved during the tens of millions of years that mice and humans have been diverging lacks the usual "landmarks" that scientists typically are able to look for. "This is an extremely important point," Lewin said. "We have absolutely no idea what the functions of these regions really are."

"We have a lot to understand about this type of conservation," Lewin continued. "Having the cattle sequence or the pig sequence or any other mammalian genome from an animal that's as distant from the primates as we can get is going to help us to annotate these very interesting and compelling regions of the genome."

Beyond such comparative genomics, said Steven Kappes, scientists will need to compare what is happening at the protein level in various species, or what he termed "comparative proteomics." "I think we're going to find that this is going to be a lot more informative than even comparative genomics—really looking at what these genes are doing in different systems. And farm animals provide a very unique perspective to identify these genes and determine gene function."

In the case of a gene called insulin-like growth factor 1 (IGF-1), for example, researchers believed they knew the function of the protein produced from the gene. But, Kappes said, "Then we started looking at more tissues and all of a sudden it turned up in a lot more tissues than we ever expected. We began to hypothesize what it was doing in these different tissues, and pretty soon the true function of that gene and its corresponding protein became cloudier and cloudier. So utilizing different animals will allow us to get at the true function of that gene, and then allow us to break down what that gene product (or protein) is doing in that biochemical pathway, and to look across different organs and tissues at different developmental moments in the organism's life, to really identify how it's regulated and how it functions."

In another case, Kappes said, a researcher discovered that a reproductive hormone was playing a key role in the development of muscle in an early embryo. "What is it doing in muscle development?" This discovery illustrated how little currently is known about the functions of many genes, but with further analyses of different animals and their traits, Kappes noted, "we will have a much better chance of truly understanding what they do."

ADVANTAGES OF DOMESTIC ANIMALS FOR COMPARATIVE GENOMICS

For researchers doing comparative genomics and comparative proteomics, domestic animals have one strong advantage over most other species: there is a long history of studying these animals. Scientists are familiar with their development, their resistance to disease, and determining how to work experimentally with them in various ways. For example, Lewin said, "Almost every human in-vitro fertilization fertility clinic employs methods that were first developed for cattle and sheep. Artificial insemination, embryo transfer, freezing semen, and sexing were all first developed for use in cattle and sheep. Furthermore, if you look at the species in which cloning has been most successful, it's actually in the ruminants [cattle, sheep, and related animals]. So application of functional genomic technology to early mammalian development using the cow and the sheep is going to be an extremely important tool to us in our understanding the early events in nuclear reprogramming and what causes embryos to live or die past a certain point, prior to and after implantation."

As a result of all the research done on domestic animals over the past several decades, decoding the genomes of cows, pigs, and others will have tremendous value for human medicine, Kappes said. "Comparative genomics will utilize a lot of the research background that we have done for the last forty or fifty years."

One area in which the genomes of domestic animals could be particularly valuable, Lewin said, is biosecurity. "There is an awareness of the problem that we're facing in risk to both human and animal health from zoonotic pathogens such as anthrax. Understanding the genes involved and creating a wider array of genomic tools is going to allow us to do the things that we need to do to protect not only the animals, but the human population as well."

Such measures will include, Kappes said, learning about how the organisms that cause disease interact with their hosts and how they are transferred from host to host. "This is important for food security, food safety, [defensive] biologic warfare and understanding the interactions of the microbe and the animal. It is another area that we have not tapped very well and we will see a tremendous amount of information come out of that."

3

Identifying Priorities

Ideally, researchers would like to have sequenced the complete genomes of every animal of interest. In practice, however, that is not possible. The worldwide sequencing capacity is, according to the workshop participants, enough to sequence a complete mammalian genome every four to eight months, given that entire full-time capacity was devoted to one species. So in theory researchers could sequence the genomes of the major domesticated animals of interest—cattle, pigs, dogs, cats, horses, sheep, chickens—within a few years. But the reality is that they likely will have to settle for much less. Moreover, it must be noted that a completely sequenced genome typically is preceded by a *draft* sequence, and draft sequences can vary in the extent of their completeness and quality.

"The problem is the price tag" explained Stephen O'Brien. It's very expensive to sequence a mammalian genome and estimated costs range from fifty to as much as one hundred million dollars. Mark Guyer of the National Institutes of Health (NIH) echoed O'Brien's point: "We know how to build sequencing capacity these days. It's not that difficult; it just takes money. The question is, if you want the genomes of domestic animals of agricultural or other importance sequenced, where's the money going to come from?"

So it is necessary to identify priorities, regarding which domestic animal genomes should be sequenced first, and how well each should be sequenced. Is it always important to sequence the complete genome, for instance, or is it possible with some species to get by with a partial genome sequence, choosing certain parts of the DNA and ignoring others? To obtain the greatest accuracy, it is necessary to repeat the sequencing as many as six or eight times, and each replication adds to the overall cost. The workshop participants were asked to consider how such priorities might be set, taking into account not

just scientific factors but also the practical aspects, such as how likely it is that funding can be secured for sequencing a genome.

FINDING THE BALANCE BETWEEN SCIENTIFIC INTEREST AND PRACTICAL NEEDS

One of the most important things in identifying which genomes to sequence, O'Brien said, is to maintain a balance between the purely scientific interest in various genomes and the practical benefits that can be gained from sequencing them. "If we're going to get the resources for sequencing, we cannot be so academic and ideal as to ignore the fact that it is taxpayers or pharmaceutical companies who will have to write tens of millions of dollars in checks. We need to have benefits that will pay them back for their investment. So there's always going to have to be a balance between scientific relevance and things that have a payoff in other ways."

Medical Relevance

In terms of funding potential, the most important practical criterion is medical relevance, O'Brien noted, since that is what the NIH is most interested in, and it is the NIH that to date has been the major source of funds for genome sequencing. To appeal to the NIH, researchers interested in sequencing the genomes of domestic animals will have to consider which work will address issues of human health.

"The most important aspect, then," O'Brien said, "is what can we do with a species?" O'Brien noted how the mouse, for example, was selected in part for its versatility for genetic manipulation. Scientists can develop lines or families of mice by inducing mutations by "knocking out" genes (Knockouts are the deactivation of specific genes, and are often created in laboratory organisms such as yeast or mice so that scientists can study the knockout organism as a model for a particular disease). They also can be used to develop stem cells to be delivered for medical research. Due to their small size and fast generation time, mice are easy to breed and sustain in captivity and are inexpensive to maintain in laboratories (compared to other mammals). These features allow researchers to derive inbred lines for studies of genetic disorders. Mice also are used for drug and vaccine trials. More recently, researchers have been able to develop transgenic versions for even more research and investigation.

"Which of those things can we say about cattle?" asked O'Brien. Which of those things can we say about the elephant? Which of those things can we say about the other species we nominate?"

"In these terms, the pig genome is a natural choice for sequencing because growth and development in the pig follows a very similar path to

growth and development in a human. It is standard to dissect a fetal pig in high school and college biology courses, for instance, because the organs of the fetal pig are arranged in a way that is anatomically similar to those of a human. There also are good arguments for sequencing the chicken. A great deal of classic embryology has been done on chicken embryos, for example, so that there is a large body of knowledge available for combining with knowledge about the chicken genome." O'Brien continued. A second reason for choosing the chicken, O'Brien said, concerns the major histocompatibility complex (MHC), an important family of genes involved in the body's immune system. "In terms of comparative biology, we've learned a lot already about the MHC because the MHC of the bird is a minimal form," he said. "It's something like 19 genes compared to 250 in the humans."

Agricultural Relevance

After medical relevance, a second practical consideration in choosing which genome to sequence is the agricultural value of the animal. "Clearly," O'Brien noted, "the things we eat are important to humans, and we need to have better knowledge of some of these species, including the cattle, pigs and sheep." If one focuses strictly on agricultural value, a different ordering of priorities emerges. In purely economic terms, cattle, pigs, and chickens are the most important species to sequence, followed by horses and sheep. Cats and dogs also must be taken into account because of the amount spent by their owners on keeping them healthy.

Basic Scientific and Evolutionary Considerations

As discussed by O'Brien, evolutionary considerations form the third set of criteria. They have implications both for the task of annotating the human genome, which will have many direct medical benefits, and for a better understanding of how species evolved over tens of millions of years, which, as an issue of basic science, will have more indirect benefits in the future.

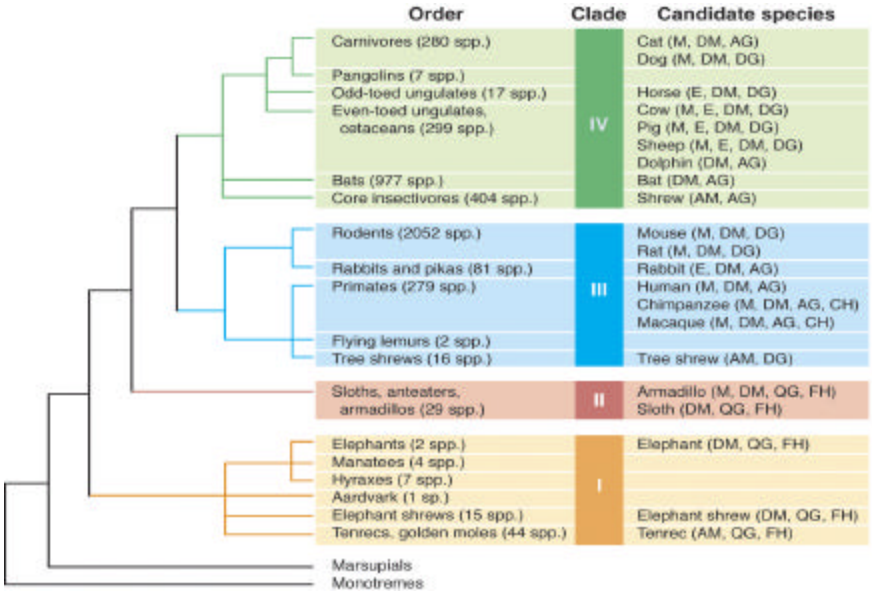
One key evolutionary factor to consider in choosing which genomes to sequence is how closely related a species is to other species that have been or will be sequenced. "The evolutionary aspect makes it important to cross a range of vertebrates," said one of the participants, "because we don't know which ones are going to be important, which ones are going to be most informative."

O'Brien showed a diagram of the mammalian family tree, as determined by a comparison of corresponding stretches of DNA among seventy different mammalian species, work done by researchers at the University of California, Riverside. He said, "there are four major mammalian radiations that have happened since the divergence of the placental mammal away from the

marsupials on the order of a hundred million years ago. Thus the primary placental mammals are sorted into four major clades, or groups" (see Box 3-1).

Box 3-1 Phylogenetic Relationships Among Modern Orders of Placental Mammals

Cladistics is a system of arranging taxa by the analysis of primitive and derived characteristics, so that the arrangement will reflect a pattern of descent among the species in question. Cladistics attempt to determine which characteristics of the organisms are specialized, derived ones that truly reflect recent common descent and it emphasizes such features, which are called "shared derived characters", in classification. Following are the four clades with examples of the animals found within them.



Reprinted from O'Brien, S. J., E. Eizirik, and W. J. Murphy. 2001. On Choosing Mammalian Genomes for Sequencing. Science. 292:2264–2266.

Presently, even-toed ungulates and cetaceans are being categorized into a fifth clade, Cetartiodactyla, on the premise that they are closely related. See J. G. M. Thewissen, E. M. Williams, L. J. Roe, and S. T. Hussain, 2001. Skeletons of Terrestrial Cetaceans and the Relationship of Whales to Artiodactyls. Nature 413:281; and *Kimball's Biology Pages*, available online at *www.ultranet.com/~jkimball/BiologyPages/V/Vertebrates.html*.

"The first is a group called Afrotheria, which consists of the elephants, the manatees and the elephant shrews. This was an African group of species."

"The second was a South American group, Xenarthra, which includes the sloths, anteaters and armadillos."

"The third major clade, Euarchontoglires, includes the rodents, rabbits, primates, tree shrews and flying lemurs."

"The fourth group has the rest of the species. It contains all your favorite species, such as the whales and the even-toed ungulates, the horses and the carnivores, as well as the primitive tree shrews and the bats. This fourth group, called Laurasiatheria, is a widely dispersed group and includes all the barnyard animals and the carnivores."

To understand the evolution of mammals, O'Brien said, researchers would like to have the genome of at least one representative from each of the four clades. That has not yet happened. "The three species that have already been nominated for full genome sequencing—human, mouse and rat—are all nested in a single clade," he noted. "That means that three of the four mammalian major clades are unrepresented entirely."

That lack, O'Brien argued, offers a strong argument for sequencing at least one or two domestic animals as representatives of the fourth clade, such as cattle and perhaps one of the members of the order Carnivora —either a dog or a cat. Evolutionary biologists also would like the sequences of representatives from the first two clades—say, the elephant and the armadillo—but that interest does not help researchers choose among domestic animals, since all of them sit within the fourth clade.

There are other evolutionary considerations that do distinguish among domestic animals, however. "Many species have a slow or conserved rate of evolution," O'Brien said. That is, the overall structure of their genomes has changed relatively little from their distant ancestors. This is true, for instance, of cats and humans. "But there are other species that have a three- or four-fold reorganization relative to the primitive mammalian genotype." This is true of mice and rats, dogs, and gibbons. "That shuffling of the genome just seems to happen once in a while in a backdrop of very slow genome evolution. Why it happens is an open question, but the point is, some species have a conservative genome and other species have a very shuffled, derived genome that is punctuated by global reorganization." It would be valuable to sequence genomes from both types of species.

"In addition to that," he said, "some species have highly derived morphometric (body-proportion) characteristics, such as the shrews. The primitive mammals looked much like today's insectivores. It looked like a little rodent." Most other mammals look little like their distant ancestors. In

choosing genomes to sequence, researchers might want to consider how primitive a species' characteristics are.

Other Criteria

In addition to the criteria of medical relevance, agricultural value, and evolutionary significance, the workshop participants offered a variety of other criteria that could be considered in deciding which genomes to sequence.

"Genome size is an issue," O'Brien said, "in a sense that it's a little bit cheaper to do a bat, which is on the order of 1.72 billion base pairs, which is a little bit over half the size of the human genome."

"It's important for this information to be useful," added Ernest Bailey of the University of Kentucky. "You need to have a community of scientists that is prepared to use it. The elephant would be fascinating to do, but I don't know how many scientists will use that information."

Daphne Preuss of the University of Chicago suggested that species be chosen based on how easy it would be to use their genomes to trace the causes of various genetic diseases. "In the human genetics community," she said, "gene discovery has been fueled by isolated populations that have discrete genetic disorders. That's really been a key to driving gene discovery forward. I think the species chosen should have genetic diversity as well as inbred populations that reveal diseases. Unlike a wild species like the elephant, where the identifiable disease states would be very limited, domesticated animals are really valuable in that way."

Another consideration, Preuss suggested, should be the value of different genomes in helping researchers to understand gene expression and regulation. "In the human genome project," she said, "it was surprising to everyone that there were so few genes, and so a lot of people are now focusing on gene regulation. We've got to understand these regulatory sequences to understand the array of gene expression. If you go too far away in evolution, you start to lose the ability to compare regulatory sequences. But there is also value in going further away. So, in evolutionary terms, we need some species that are close and some that are farther apart."

Joachim Messing, director of the Institute of Microbiology at Rutgers University, added that researchers should keep in mind how far along genomics research already has come for various species. "We should think a little bit about entry points," he said, "that is, with what information is available for a particular genome. Is there already a genetic map? Are there Expressed Sequence Tags (ESTs – stretches of DNA used to identify functional genes)? And so on." Studies of ESTs, for example, can be done with relative ease at a fairly low cost, and they can provide valuable information when annotating genomic-based sequencing.

There indeed is a difference among domestic animals in how far along genetic mapping and sequencing work has come, said Steven Kappes. For example, the number of ESTs varies from species to species. "Within GenBank (the major repository of genetic sequences in the United States), cattle have the most among farm animals, with 230,000 sequences. This is fifth in GenBank behind humans, mouse, rat, and *Drosophila* (the fruit fly widely used in genetic research). Pigs are about half of that. Chickens have only 44,000 (that had been discovered in earlier research), but a United Kingdom effort has nearly finished with sequencing 300,000, and those will become public, so that will dramatically increase the numbers for the chicken. On sheep it's relatively few, and only a little more on horse."

The story is similar for other resources used in genetic mapping and sequencing; cattle, pigs, and chickens are most advanced, while horses and sheep lag behind. Thus mapping the sequences of the first three species could be finished more quickly.

It might not be necessary to do a complete genome for each species, Messing said, and so researchers who are prioritizing genome projects should consider whether to sequence the entire genome for a particular animal. "We should look at the extent of sequence coverage that we want to allocate to a particular project," he said, "either a complete sequence or to go for targeted regions, which I think also has great value in terms of comparative genomics" (see Box 3-2).

Box 3-2 Targeted Sequencing

Sequencing an entire mammalian genome is very expensive. Thus far, the quality of a *draft* sequenced genome—by using the "whole-genome shotgun approach"—has depended upon costs. Because basic "1X" coverage of a genome can cost roughly $15 million, and because 6X coverage typically is preferred, a draft sequence alone can cost nearly $100 million, and an additional $90–125 million would be required to improve the draft to a high quality finished version. Eric Green of the National Institutes of Health, however, suggested an alternative: targeted sequencing, or sequencing only certain portions of a genome that are of particular interest.

"A 1X shotgun sequence of a mammalian genome costs something on the order of $10 to $15 million," he said. "I think that that is a fairly accurate number. So if you're thinking about 6X or 8X coverage, you quickly approach $100 million per genome, at least by current technologies. And if they go for $100 million a crack, there is going to be a limited set of organisms that can be subjected to global sequencing."

"I think it is really important to recognize," he continued, "that so far the discussion has been a little bit, all, or none. It's either you're going to get your organism up on that list and get it sequenced or else it was going to forever

be lost. I don't think that's the case at all. And so I want to tell you there is a great, great value and great future in targeted sequencing efforts."

Green described work done in his laboratory that involved comparing corresponding stretches of DNA from the mouse and ten other species. The technique demands the use of bacterial artificial chromosomes, or BACs. A BAC is a long stretch of DNA from a human or another organism that is put into a bacterium in the form of an artificial chromosome, so that when the bacterium makes copies of itself, each copy has its own identical BAC, complete with the DNA of interest. In this way, researchers easily can work with these long stretches of DNA, making copies and comparing them with other bits of DNA. If the entire genome is thought of as an encyclopedia containing the information necessary to build an organism, Green said, each BAC corresponds to a single page in the encyclopedia.

Working with these BACs, Green and coworkers were able to do a comparative genome analysis on five different stretches of DNA taken from eleven species. In doing so, he derived a great deal of information about the relationships among the species while needing only small pieces of their genomes, not the entire thing. "Certainly this provides a greater potential for exploring a wider array of genomes," he said. "You don't have to invest $100 million to get a little bit of sequence information about a particular region of a particular organism."

In the case of domestic animals, he suggested that targeted sequencing would make it possible to investigate stretches of DNA that contain genes of interest without sequencing the entire genome. "In these cases you start with some chromosomal region of interest, for example, some region of the livestock genome that may have a quantitative trait locus (QTL) that you're interested in studying. You're going to want to isolate and map that in overlapping BAC clones and go through and systematically sequence each of those individual BACs."

Such a strategy demands that collections of BACs be available for the species under study, but, said Green, that does not appear to be a problem with domestic animals. "The good news here is that both the NIH and the National Science Foundation (NSF) have realized the importance of this, and, as a result, they are now either currently funding or soon will be funding major efforts to generate dozens and dozens of new BAC libraries in the coming years."

One drawback associated with targeted sequencing is that it assumes beforehand which genes are deemed most important. Hence, the criteria for selecting genes remain subjective. Although some suggestions for criteria were discussed during the workshop, the participants did not discuss a uniform set of standards.

In addition, Kappes noted, it will be necessary to decide how much redundancy is needed in the genome sequence for each animal. ("Fold

redundancy" reflects the average number of times each base pair has been sequenced from independent DNA clones of the bacterium *Escherichia coli*.) Will six- or eight-fold coverage be needed, or will it be possible in some cases to get by on much less? (The number of 'folds' [i.e., 6-fold or 6X] refers to the number of bases sequenced relative to the genome size [in base pairs] of an organism. Depending on the number of genes estimated to be in the genome of interest, researchers have to decide how much coverage they need to be statistically secure in their efforts to identify genes.) "I think 6X hits it about right for eutharian primate non-rodent species. The other species within that clade we do not need to do as much, so we could back off on that. I am a little bit careful in saying that because Claire Fraser's group has shown very well that you miss a lot of things if you only do a rough draft so I think the jury is still out. But we obviously have to be realistic in what it costs and how many genomes we can do. I think we should start out with a moderate coverage in some of these species and then back off as we go down the line." (Lower fold redundancy eventually might be possible due to the increased bioinformatics tools and the increased number of comparative data sets from many species. Hence, one might expect that new tools and increased data would allow one to rely upon less sequence data.)

WHICH ANIMAL GENOMES SHOULD BE CONSIDERED FOR SEQUENCING?

The workshop participants were not asked to rank the genomes of domestic animals from most important to least, but they were asked which genomes should be put under consideration. A number of participants expressed the view that cattle, pig, and, perhaps, chicken genomes should be put at the top of the list for a variety of reasons, including agricultural value and helpfulness in understanding the human genome better.

"As far as farm animals are concerned," O'Brien said, "it would be hard not to put the cow up on the top of the list with the people that are here at this meeting, but the people interested in the pig genome make a pretty strong argument, too. They have good opinions. Those are the two front-runners that I see…. the second tier would certainly be chickens, sheep, perhaps even horses."

Max Rothschild, of Iowa State University, on the other hand, argued for putting chickens in the top group. "It seems to me that pigs, cattle and chickens are the three highest for domestic species," he said. "But you can argue on two grounds why the cow ought to be third on that list. It's only first on the list because it's better organized but from a financial standpoint, most of the meat consumed in the world is either poultry or pork. From a health standpoint, chickens and pigs are more important. Certainly, pigs from a

xenotransplantation perspective and certainly chickens from what we could learn about immunology and some other things."

There also was some disagreement about horses. "I'm a horse guy," Douglas Antczak of Cornell University's College of Veterinary Medicine said, "and I can't justify working on horses in genetics at all. They have only one offspring per year, you can't superovulate them, they're very large, they fight, and they kick."

But Ernest Bailey offered a different perspective: "The horse is economically important in the United States, although not because it produces food. It's important from a recreational standpoint. Many people regard it as a companion animal, but I think the racing industry is quite large. There's a lot of money there and a lot of money spent on health and animals.

"Furthermore, the horse is a separate family. It's a member of a family that has ten different species, all with different chromosome numbers. There's been a rapid chromosome evolution over a period of about two million years. One of the experiments that will be interesting in the long run is to look at the reasons for the chromosome evolution in the horse. There are ideas that gene duplication is responsible. Do these kinds of gene duplications exist in the different species?"

Assuming that the cattle genome is sequenced, Kappes said, that would take much of the pressure off the need to sequence the sheep genome. "The sheep genome is very similar to cattle genome. There are only three different changes in chromosomal organization between cattle and sheep and, basically, when we find a gene in sheep on this particular location, we find the similar gene in cattle."

As for domestic companion animals, the two natural choices are dogs and cats, but there was a difference of opinion as to which genome would be more useful to sequence. "The cat has been a favorite in my laboratory for almost twenty years," O'Brien said, suggesting that it would be the better choice for a number of reasons. "It's a model for many human hereditary diseases, such as hemophilia, as well as several infectious diseases, such as leukemia. There's an acquired immune deficiency syndrome (AIDS) virus in cats, feline immunodeficiency virus (FIV). There's extensive medical surveillance and literature. Finally, the human genome and the cat genome are both very primitive for their respective orders. That is to say, the ancestor of carnivores look a lot like a cat, and the ancestor of primates looks a lot like a human."

An audience member echoed O'Brien's arguments: "One of my preferences is the cat, and that's because of it being such a good model for FIV and human immunodefiency virus (HIV) and because it is an animal we can do drug therapies on." Vivek Kapur, University of Minnesota, noted, "Cats share a very large number of common pathogens with humans, as do pigs."

But Anczak offered a counterpoint. "I'd like to speak for the dog over the cat," he said. "A dog has all the advantages that have already been

mentioned in the cat and, in addition, it has behavioral and morphologic traits of interest, which the cats don't have."

Finally, Harris Lewin proposed a couple of longshot candidates for sequencing. "I will put a plug in here for fish, because fish really are an incredibly powerful tool for genetic mapping. You can get down to a half centimorgan (unit for measuring the recombination frequency in DNA) resolution and there are several groups around the world that are interested in the fine mapping (high resolution genome mapping) of traits with fish."

"I'll add that the honeybee is an incredibly interesting model for fine mapping, because it has a very high recombination rate. One centimorgan is about 50 kilobases (50,000 nucleotides), which is an incredible tool. It has a very short generation interval as well, which is a powerful tool for fine mapping of quantitative trait loci. I think that we may see a big push to sequence the honeybee genome as well in the next few years."

4

Roles of Public, Private, and Nongovernmental Organizations in Advancing Genomics Research

In addition to the scientific issues surrounding domestic animal genomics, workshop participants were asked to discuss more practical matters relating to the roles of public, private, and nongovernmental organizations (NGO) in coordinating domestic animal genomics research projects. How should the sequencing work be divided up among the different types of institutions? How should the information that comes from this work be organized and managed?

PARTNERSHIP IS ESSENTIAL FOR ADVANCING ANIMAL GENOMIC RESEARCH

For Roger Wyse, Managing Director of Burrill & Company, and Chairman of the Alliance for Animal Genome Research, the strongest approach is to involve a number of different entities in domestic animal genomics research programs. "It's quite clear that what we're talking about here is not a stand-alone initiative, but rather one that needs to be integrated across agencies to take advantage of their structure and strengths." Although it will be more complicated to coordinate efforts in such a broad coalition ranging from government to the private sector, it is in practice the only viable option. For example, scientists then can consider their research priorities and decide which governmental agency is most appropriate for working with them, instead of concentrating on an agency first and trying to present their research in such a way that it fits the agency's agenda.

Presently, the role of the federal government in genomics research remains important with respect to the opportunities that it offers for competitive

research grants. Most of the support for basic research into domestic animal genomes, such as determining the billions of individual base pairs that make up the genetic sequence of each mammal, is likely to come from research grants, said Wyse, and most of those grants will come from the federal government. This type of funding has two important functions, Wyse said.

"We think it is very important to have the funding to increase the knowledge base," he said. So determining the cattle genome, for instance, is valuable in itself. However, Wyse also noted the importance of securing human resources, stating that "…it is equally important to get a competitive grants program to allow us to build the human resources that are necessary for not only doing the science, but commercializing the science." The benefits of increasing human resources in this field will be broad ranging, and spread across different types of institutions, Wyse said. "If you are a small company now in the animal genome space and you are looking to hire really good people, there aren't very many out there. It is going to be important that we populate our universities with research funds so that we can train that next generation to do research in this area."

THE ROLE OF FEDERAL AGENCIES

The issue of which federal agencies most likely would provide the basis for research on domestic animal genomes generated a great deal of discussion at the workshop. Most of the federal funding for sequencing the human genome came from two agencies, the National Institutes of Health (NIH) and the Department of Energy (DOE). The NIH was interested in the human genome because of its importance to medicine and improving human health, while the DOE's genome effort stemmed from its long interest in the threat that radiation poses to human health—particularly its ability to cause genetic mutations. From one perspective, the NIH and DOE are well positioned to oversee research into domestic animal genomes, for they already have strong support for genomics, some of which could be applied to domestic animal research. On the other hand, NIH and DOE traditionally have focused on humans and on animals, such as the mouse, that serve as traditional laboratory models for disease, and domestic animals traditionally have not been part of their domains.

By contrast, the U.S. Department of Agriculture (USDA), with its mission for improving agricultural productivity in the United States, would seem to be appropriate for initiating further sequencing of domestic animal genomes, but its budget requests and allocations traditionally have not accounted for increased genomics research. Thus, workshop participants discussed the best ways to match their research objectives to the goals and missions of these various government agencies.

Wyse suggested, for instance, that the USDA has a number of existing programs that could be oriented toward research on domestic animal genomes. However, " we have not really used the competitive advantage of the structure of an organization like USDA. For example, you have in it the Agricultural Research Service, which is an in-house research unit and has the necessary structure and budgets, has the ability to assign people to projects, and, most importantly, has a rewards system that allows it to do directed kinds of activities that can be both long-term high-risk or programs like managing germplasm. I think we ought to take advantage of some of the structures that we've got in place…Also at USDA is a competitive grants program that could be used to move into (research determining) gene functions that are important in applications in agriculture."

On the other hand, as Jerry Dodgson of Michigan State University pointed out, the USDA budget cannot accommodate genome programs for farm animals. "We are talking about three mammals and the chicken, so we are talking $340 million by today's dollars. Unless we take a chunk out of the farm subsidy pot, that becomes a big constraint."

For that reason, one audience member suggested NIH as the most appropriate starting point. "In light of NIH calling for proposals, I think what we do is work very hard to get NIH to see the value in sequencing farm animal genomes."

As for companion animals, such as dogs and cats, the USDA Agricultural Research Service is not authorized by law to fund research on them. In this case, there is no choice but to consider other non-agricultural agencies such as NIH.

"As we think about the sequencing that needs to be done," Wyse said, "we need to establish the criteria and the rationale" and then engage the "various agencies, whether that's NIH, DOE, National Science Foundation (NSF), or USDA. So an important part of this is thinking through what we want to do and then using the competitive strengths and the missions of the various agencies to appropriately support it."

THE ROLE OF THE PRIVATE SECTOR

In order to address the later stages of basic research achieved through competitive grants, Wyse said that participation by the private sector must be considered as the research moves closer to commercial applications. He then mentioned the role of venture capital and funding from corporations interested in agriculture and human health.

"A lot of the fundamental knowledge that is going to be developed is really to develop the knowledge on which you can form small companies that people like us (venture capitalists) can invest in." Wyse noted that venture

capital is fairly new to agriculture, but made it clear that it is important because these small companies become the source of moving basic knowledge towards commercialization. Agricultural biotechnology represents a small fraction of the amount invested by venture capitalists in U.S. companies. Still, Wyse said, venture capitalists are just discovering agricultural biotechnology, and the amount they invest in it should grow. "I would argue that it's going to be critically important in the future."

In addition to competitive grants from the federal government, and venture capital, Wyse discussed the role of corporations interested in agriculture and animal health. To date, he said, these companies have done relatively little. The animal health industry, he said, "has been the weak sister of the big pharmaceutical companies, and there hasn't been a lot of funding or initiative on their side to develop new things on their own."

According to Wyse, this is about to change. For example, "There is a whole group of companies in the germplasm development area. Those folks are probably the furthest along in (terms of) thinking about genomics and how it can be used to help them select superior animals. But the issues for them are the profit margins that they have and the cost of doing both the basic work and some of the assays that would be used in marker assisted breeding or gene-expression profiling."

THE ROLE OF A PUBLIC-PRIVATE PARTNERSHIP

To help define the roles of the public and private sector in domestic animal genomics research, Wyse suggested a public–private consortium in which private companies and research universities worked together. If the consortium had well-defined goals and limited itself to precompetitive research (work where companies or institutions are not averse to their competitors having equal access to the results of their efforts), it would have a good chance of success. Wyse believes that food and agricultural industries are prepared to be strong partners in an alliance, because the companies are interested in developing the technologies that they can apply. In particular, Wyse noted, the companies that process food animals are "poised and ready to participate. I would think that if we constructed the right kind of consortium, there might be an opportunity for a public partnership across that sector." It also is likely, he said, that various special-interest groups would support or help fund such a consortium. Wyse identified the Canine Health Foundation as a prime example of a special interest group that has accomplished a great deal with a relatively small amount of resources and would be quite amenable to forming a consortium.

Kevin Schultz of Merial supported the idea of such a consortium. "One of the things that we've done as a company is to make a decision to do a lot

more external interactions, rather than do everything internally. In fact, the shareholder companies Aventis and Merck, who own Merial, have made similar kinds of decisions. And so I think that there will be partners. It's clearly an expensive venue and we've got discussions ongoing with certain companies along those lines."

Albert Paszek of Cargill showed support for the idea of a consortium as well. "When it comes to animal genomics activity, everything is outsourced anyway, and the major strategy that we operate under is massive collaboration and sponsorship of projects at public institutions, including universities."

Max Rothschild of Iowa State University added that, historically, consortia have been successful in the agricultural research area. "There's some good evidence in the swine business that there have been good consortiums if it's precompetitive. There was one in Europe at the start of the 1990s. I helped organize one for QTLs in the mid-1990s."

Finally, Wyse said, those interested in furthering domestic animal genomics should consider highlighting the ultimate value of this research to drugs and other human medical products. " Part of the strategy will be playing up the fact that the biotechnology companies that take the information you generate and apply it to agriculture are actually going to build value in the human health care market." Because returns are higher in human than in animal healthcare, and because the human healthcare market is much larger than its animal counterpart, the applications that animal genomics research has to human health will be an important consideration.

ALLOCATING WORK IN ANIMAL GENOMICS RESEARCH

In reflecting on the roles of government and the private sector, workshop participants focused on one major question with respect to the role of academia and its importance in creating a strong genomics research community: How much of the sequencing and other routine work should be conducted outside of an academic setting?

Over the past several years, for instance, several private companies have begun to offer sequencing services. One such example is The Institute for Genomic Research (TIGR), which is a nonprofit research institution located in Rockville, Maryland. With banks of automated sequencing machines, such centers can determine DNA sequences more quickly and cheaply than most university sequencing projects. A number of workshop participants found it sensible to use these specialized centers performing parts of the genomic research that mostly are a matter of technique.

Several workshop participants agreed that a number of the routine aspects of genomic research such as constructing BAC libraries or sequencing should be contracted out to these specialized facilities while researchers and

students spend more of their time investigating "more interesting things" such as looking at gene function and gene expression and how those relate to the important traits, phenotypes, or diseases that we are interested in, be they in agriculture or in human health.

Others argued, however, that a university setting could be beneficial for carrying out sequencing. "One of the downfalls of outsourcing all sequencing activity," said one audience member "is the fact that you cannot encourage small laboratories to train the next generation of scientists. You also discourage people from using those data because they don't feel ownership or a tie to it. I understand the efficiencies of outsourcing, but there needs to be a happy medium."

A second audience member concurred. "Some of these species that we are talking about—cattle and pigs included—don't really have a community of people built up who are working with these things. There needs to be a balance between outsourcing and how much you give to academic institutions to encourage that community development in graduate students, post-doctoral fellows, and the like. There needs to be some recognition of the fact that the community is not widely developed, and it's in everyone's best interests that the community be developed."

But, said Claire Fraser, president and director of TIGR, there does not necessarily have to be a choice between high efficiency and training new scientists. "What's been set up at Baylor (College of Medicine) and what's been set up at Washington University represent excellent examples of how you can create high-throughput facilities to get this work done at the most efficient cost yet, at the same time, train students and train post-doctoral fellows. I don't think it's an either/or situation."

Ultimately, Wyse concluded, those calling for a centrist approach seem to have a valid argument. "There is a balance between doing routine sequencing in an academic setting with faculty and graduate students and post-docs, versus contracting it out to TIGR or someone else."

But for the parts of the project with more intellectual content, most seemed to agree that it makes sense to use university researchers. "It's been my philosophy," Wyse said, "that universities are better positioned to do the competitive-grant functional-genomics work and the like, as opposed to the basic sequencing. It seems to be a better fit with the university environment as well as its reward system."

5

Data Access

The final issue tackled by the participants was how best to work with the tremendous amount of data that will be generated by domestic animal genome projects. The data create a number of challenges, said Daniel Drell of the U.S. Department of Energy (DOE). "These have to do with the interoperability of data, the sharing of data in some cases, but, principally, organizing it in such a way that others can come along and add value to it in some efficient ways." So far, he said, "the genome projects have been largely unsuccessful at dealing with many of these."

APPROPRIATE TOOLS AND THE IMPORTANCE OF DATA ACCESS

One can frame the issue in terms of access to data, said Claire Fraser. "When it comes to data access," she said, "there are two ways to think about it. One, are the data accessible in GenBank or someplace else? And the answer is yes. But individual sequence reads or assembled data are only so useful. What we really need in terms of data access, in order to empower all of the users that are interested in getting a hold of these data, are far better databases and tools to really exploit the information. And I think this is an area that so far has been more of an afterthought with these projects than it should have been."

The result, she said, is that some genomics researchers end up having easier access to the data than others. "We are seeing a bit of a genomics-divide being created between those groups that are involved in generating the data and have been forced to build the tools in order to manipulate it, and the more typical user who doesn't necessarily have access to the same tools, (and) who

doesn't have bioinformatics expertise at his or her university. I think that's one of the real problems that we need to address."

The other problem, Fraser added, is that the various genome projects generally make no allowance for taking care of the data they generate once the project is finished. "For the most part, even for sequencing projects with bioinformatics support during the term of the project, that support ends when the sequence is completed. There's been no plan put in place for how to maintain and update all of this information."

"That problem is going to get even worse as we begin to accumulate more data. There have been all sorts of models proposed, from letting people in the community who are passionately interested in an organism do it on an ad hoc basis, to having this done in a more centralized facility, to having this done in a distributed way but with clear rules for interoperability. I've even heard some people go so far as to suggest that perhaps we need to come up with some sort of tax on genome projects that goes to fund a bioinformatics trust managed by an inter-agency group responsible for maintaining these databases."

Several participants pointed out that in order to maximize the value of the information generated by domestic animal genome projects, researchers and information technology specialists will have to pay more attention to data handling. In particular, programs need to be designed not only to maintain the data and make it accessible to any researcher who needs it but also to make sure the information can be integrated with new data and new understandings as they appear.

THE CHALLENGE OF SCALING UP IN RESPONSE TO INCREASES IN DATA

The biggest difficulty is the problem of scaling: A database must be designed so that it continues to work, and work well, when the amount of data in it is doubled or increased by a factor of ten or twenty. That will be a challenging job, Fraser noted.

"I'm not convinced," she said, "that any of the existing databases that have been built so far to handle sequence information are robust enough to scale to the level that we know we are going to need in going forward." The databases built to handle the sequence information are actually the easy part, she said. "We would like to begin to add in functional information, either directly or through links, to all of the existing gene and protein databases. When you start thinking about doing that, the challenge goes up by several orders of magnitude."

Owen White, of The Institute for Genomic Research (TIGR), made a similar point. "The National Center for Bioinformatics (NCBI) is doing a heroic job," he said. "They are doing an amazing job managing sequence data and

publication data. That's a specific data type, and they have a fighting chance of scaling up for just the raw sequence information.

"But there's another data type that a lot of us are familiar with, which is annotation. Annotation is kind of a generic term, but I usually mean identification of all the genes and trying to give functional assignments to those genes and trying to represent them well in a structured database. So if you've got 500 microbial genomes and people want to come in and work with the data, I would argue that we don't really have representation systems for that type of thing."

While the problem of scaling up the databases that hold basic information, such as sequences of base pairs, is challenging but seemingly solvable, no one yet has constructed databases that will be able to handle the amount of annotation that likely will proliferate in the years to come.

STRUCTURING GENOME DATABASES

Workshop participants had various perspectives on how a system of genome databases should be structured. White, for instance, offered a vision of large central repositories that would handle all the data of one particular type—say, information on how genes are expressed—for many different species. He warned that it would not be feasible to have one mega-center handle all different types of data for every type of organism, but he argued that if each center focused on one type of data, it would work quite well.

"There are a number of reasons why I think this is a much more attractive model," he said. "Training becomes much easier, and there is reduced reinvention of the wheel. Once you instantiate those infrastructures, they are easy to apply to new organisms."

Furthermore, he added, these data-specific centers should be able to expand easily enough to accommodate ever-growing amounts of data. "I think they are the only things that had a chance of scaling." Suppose, he said, that some individual research center had developed a good way to represent expression information for the particular organism studied at that center. "Hopefully they generalize their services enough so they can apply them to another organism. Then if they instantiate what the standard operational procedures are, they develop a relatively good training program, and they have a robust representation system going on in the database. That's the hard part. That is the energy of activation, so to speak. Then adding another organism is actually much simpler."

A member of the audience disagreed with White's suggestion, however. For him, it made more sense to keep smaller, individualized databases and develop standards so that the various databases could exchange information and work with each other almost as if they had a single database. "You don't

have to bring things into gigantic warehouses' or try to federate databases. You try to create a level of information that can be exchanged among databases. In part, this goes along the lines of the discussions about whether you sequence in a center only or distribute the work in order to create local communities of scientists and train graduate students. This is particularly true in bioinformatics. If you have only centers for collecting information, you develop no local skills and no local students to use that information."

"Centers like NCBI do an extraordinary job of archiving low-level information," he continued. "But in the plant community, for instance, there is an immense difference in the interests of, say, the cereal genomicists versus just the legume folks. The legume folks have a high interest in secondary metabolism, symbiosis, and nitrogen fixation. Those are all functions that fit within community exploration of data and creation of data models and data-mining mechanisms appropriate to those. But they don't map onto cereals, and if you try to force these into a one-size-fits-all model, you come down to a lowest common denominator of things that are done well." In short, having different centers for different organisms allows each to specialize and take into account the areas of interest for that particular organism. It might make sense to accumulate certain types of information—generally the very basic, low-level information—in one, large central repository, but the higher-level information, with its sensitivity to the type of genome being considered, is better handled at individual centers.

ALLOCATION OF RESOURCES FOR BIOINFORMATICS

No matter how the centers ultimately are organized, several participants expressed the view that more resources must be allocated toward bioinformatics if researchers are to be able to work with all the data that is being accumulated. "If you want a system," White said, "that can dynamically manage data that's coming in from several projects in parallel and have version dates and a help desk and just a well-engineered system, we are talking about a completely different magnitude of budget that's required to do that."

6

Looking Forward

Caird Rexroad of the U.S. Department of Agriculture's Agricultural Research Service commented, "It's well recognized within USDA, maybe because of what's been done in the plant sciences with the plant genomics initiative, that it is a very good time to be going forward with this (domestic animal genome projects)." Still, there are no guarantees of success, and workshop participants offered a variety of advice for how researchers could improve the odds that domestic animal genome programs would grow. "What is it," asked Richard Gibbs of the Baylor College of Medicine, "that makes for a genome to get all the way through these hoops and hurdles to get to the point where it is going to be sequenced? So far the real issue has been advocacy. By advocacy I mean someone who is really pushing for it and is working with everybody else who is interested in the organism."

A STRATEGY FOR AN ANIMAL GENOME INITIATIVE

Ronald Phillips of the University of Minnesota offered seven factors that he believed were instrumental in getting support for the Plant Genome Initiative—a program that was developed in part from discussions among officials from the USDA, NSF, and the National Academy of Sciences in 1997 (see Box 6-1).

The first point that Phillips presented was posed as a question, "Has the science and technology matured to a point where you really could make a good argument [to policymakers]? I think that's certainly the case with all of the work reported here today." While he noted the impressive advances in animal genomics, he also acknowledged that there was much left to do. The challenges

of functional genomics and proteomics are just beginning to emerge. Phillips questioned whether some of the current techniques that are in use, such as knockouts in mice, are adequate for the challenges ahead. "Are they adequate

Box 6-1 Factors That Contributed to the Establishment of the Plant Genome Initiative

- Science and technology that had matured enough to deserve serious consideration
- An interagency approach
- A scientific foundation
- Support of stakeholders
- Support of Congress
- Key agency leadership
- Input from a broad range of scientific experts from throughout the world

Source: Ronald Phillips, University of Minnesota.

for what researchers are trying to do, or should they devise a system of knockouts in a species that is important, in terms of your future lists of candidate species for sequencing?"

"My second point that I thought was important for success of plant genome initiative was the interagency approach." By bringing together various federal agencies such as the Department of Energy (DOE), USDA and National Institutes of Health (NIH), identifying their common interests within animal and microbial genomics, and preparing an interagency agenda for a research initiative, Phillips suggested, discussion in the U.S. Congress will be stimulated.

Given the diversity of interests and the resources involved for genomic research, politics are an inevitable component of the development of an initiative. Phillips noted that the researchers and advocates should focus their efforts on what they know and do best—the science. "The third point was it should be science-based. Make your arguments based on science and let that carry the day."

"The fourth point was commodity support. The corn growers brought this to the fore. They went to Congress and Congress asked them, where do you place this in your order of priorities? And they said, number one. After that was said, that was the end of the argument in many ways. It was a matter of how do you get it done. They were convinced when the commodity groups, particularly corn growers, said that was important to them."

"The fifth important thing was having key congressional support. If you have good goals and can get someone to articulate that, that's helpful." Phillips also said that it would be key to show Congress that the work planned under the initiative was not being duplicated elsewhere, for example, in the private sector.

"Sixth is having key agency leadership, and I assume that will happen with the interagency working group (representatives of the USDA, DOE, and NIH have formed a working group that is focusing on animal genomics research program). The Office of Science and Technology Policy (OSTP) was extremely important in our case. We had people who understood and really worked on our behalf both here and abroad."

"Finally, one of the important aspects was that we had discussions with respected science sounding boards. The first thing we did was to have this kind of meeting at the National Academy of Sciences. That was followed up with a colloquium of a broader set of scientists held at National Academies facilities in Irvine, California. We discussed it at a Gordon Conference (Gordon Research Conferences provide an international forum for the presentation and discussion of frontier research in the sciences), and particularly brought in the international dimension there. We had discussions with the panel that was reviewing the *Arabidopsis* situation. And they told us it could be speeded up by several years with more funding. So, that became one of our priorities. I think Congress respected the fact that we had talked to the best scientists in the world to design this program. Finally, be sure you keep the international community involved, not just in terms of some interactions, but, actually helping you work through your goals." By adopting a similar set of approaches, Phillips said, genome researchers interested in domestic animals could improve their chances of establishing and developing their own genome programs.

Appendix A

Workshop Agenda

Exploring Horizons for Domestic Animal Genomics:

A Public Workshop

The National Academies
Lecture Room
2101 Constitution Avenue NW
Washington, DC 20418

Agenda
February 19, 2002

8:30 AM	Welcome and Introduction **Kim Waddell,** Board on Agriculture and Natural Resources
8:40 AM	The Landscape of Comparative Genomics in Mammals **Stephen O'Brien**, The National Cancer Institute, National Institutes of Health
9:10 AM	Animal Genomics Research in the U.S.–Where We Are and Where We're Going **Steve Kappes**, U.S. Department of Agriculture

9:40 AM	Livestock Genome Sequencing Initiative: Status and Importance **Harris Lewin**, University of Illinois
10:10–10:30 AM	Break
10:30 AM	Multi-Species Comparative Sequencing of Targeted Genomic Regions **Eric Green**, National Institutes of Health
11:00 AM	The Rat Genome Sequencing Project **Richard Gibbs,** Baylor College of Medicine
11:30 AM	A Private Sector Perspective: Financing Innovation **Roger Wyse,** Burrill & Company
12:00–1:00 PM	Lunch
1:00–2:00 PM	Priorities for Genome Sequencing: Which Species? **Group Discussion**
2:00–3:00 PM	What Are the Roles of the Public, Private, and NGO Sectors for Advancing Genomics Research? **Group Discussion**
3:00-–3:15 PM	Break
3:15–4:15 PM	How Can We Facilitate Data Sharing and Access? **Group Discussion**
4:15–5:00 PM	Summary and Wrap-up
5:00 PM	Adjourn

Appendix B

Workshop Participant List

Douglas Antczak
Professor
College of Veterinary Medicine
Cornell University
Ithaca, NY

Ernest F. Bailey
Professor
Department of Veterinary Science
M. H. Gluck Equine Research Center
University of Kentucky
Lexington, KY

Guy F. Barbato
Associate Professor of Physiological Genetics
Poultry Science Department
Penn State University
University Park, PA

Leah M. Becker
Government Relations Representative
National Pork Producers Council
Washington, DC

Jonathan E. Beever
Assistant Professor
Department of Animal Sciences
University of Illinois
Urbana, IL

Peter Brayton
Program Director
U.S. Department of Agriculture,
National Research Initiative
Competitive Grants Program
Animal Genome and Genetic Mechanisms Program
Washington, DC

John Byatt
Genomics Technical Leader
Monsanto
St. Louis, MO

Anthony V. Capuco
Research Physiologist
U.S. Department of Agriculture,
Agricultural Research Service
Gene Evaluation and Mapping Lab
Beltsville, MD

Bhanu P. Chowdhary
Associate Professor
College of Veterinary Medicine
Texas A&M University
College Station, TX

Noelle Cockett
Vice Provost for Academic Affairs
Utah State University
Logan, UT

Erin E. Connor
Research Molecular Biologist
U.S. Department of Agriculture
Gene Evaluation and Mapping Laboratory
Beltsville, MD

Neal G. Copeland
Chief
Mouse Cancer Genetics Program
National Cancer Institute
Center for Cancer Research
Frederick, MD

Pieter J. de Jong
Director
Oakland Research Institute
Oakland, CA

Gregory Dilworth
Team Leader
U.S. Department of Energy
Energy Biosciences Program
Germantown, MD

Jerry B. Dodgson
Professor and Chair
Michigan State University
East Lansing, MI

Daniel Drell
Program Manager
U.S. Department of Energy
Life Sciences Division
Germantown, MD

Leland Ellis, Jr.
National Program Leader
U.S. Department of Agriculture,
Agricultural Research Service
Animal Production, Product Value and Safety
Beltsville, MD

Kellye A. Eversole
President
Eversole Associates
Chevy Chase, MD

Richard R. Frahm
National Program Leader, Animal Genetics
U.S. Department of Agriculture,
Cooperative State Research, Education, and Extension Service
Washington, DC

Claire Fraser
President and Director
The Institute for Genomic Research
Rockville, MD

Richard Gibbs
Professor and Director
Baylor College of Medicine-
Human Genome Sequencing Center
Houston, TX

Peter Good
Program Director
National Human Genome Research Institute
National Institutes of Health
Bethesda, MD

Eric Green
Senior Investigator and Chief, Genome Technology Branch, and Director, Intramural Sequencing Center
National Human Genome Research Institute
National Institutes of Health
Bethesda, MD

Frank Greene
Division Director
Division of Integrative Biology and Neuroscience
National Science Foundation
Arlington, VA

Mark Guyer
Assistant Director for Scientific Coordination
National Human Genome Research Institute National Institutes of Health
Bethesda, MD

Aziz Jamai
Researcher
Plant and Soil Science and General Agriculture
Southern Illinois University at Carbondale
Carbondale, IL

Nancy Jenkins
Principal Investigator
Mouse Cancer Genetics Program
National Cancer Institute
Center for Cancer Research
Frederick, MD

Peter Johnson
Division Director
Animals Division, and the Nutrition, Food Safety, and Health Division
U.S. Department of Agriculture, National Research Initiative
Washington, DC

Elke Jordan
Director of Extramural Research
National Human Genome Research Institute
National Institutes of Health
Bethesda, MD

Ed Kaleikau
Divison Director for Plant Genomics and Plant Sciences
U.S. Department of Agriculture, National Research Initiative
Washington, DC

Sagarika Kanjilal
Assistant Professor
Department of Veterinary Pathobiology
University of Minnesota
St. Paul, MN

Steven Kappes
National Program Leader, Food Safety and Health, and National Program Leader, Animal Nutrition
U.S. Department of Agriculture, Agricultural Research Service
Clay Center, NE

John Kirby
Director, Program in Cell and Molecular Biology
University of Arkansas
Fayetteville, AR

Susan J. Lamont
Professor and Head, Animal Science Department
Iowa State University
Ames, IA

Yuandan Lee
Post-Doctoral Fellow
The Institute for Genomic Research
Rockville, MD

Harris Lewin
Professor
Biotechnology Center, University of Illinois at Urbana-Champaign
Urbana, IL

Joan K. Lunney
Research Leader
Immunology and Disease Resistance Lab
U.S. Department of Agriculture, Agricultural Research Service
Beltsville, MD

Leslie A. Lyons
Assistant Professor, and Staff Scientist
Department of Population Health and Reproduction
School of Veterinary Medicine
University of California, Davis
Davis, CA

Khalid Meksem
Assistant Professor
Plant and Soil Science, and General Agriculture
Southern Illinois University at Carbondale
Carbondale, IL

Laurie McGinley
Staff Reporter
Wall Street Journal
New York, NY

Joachim Messing
Director
Institute of Microbiology
Rutgers University
Piscataway, NJ

Stephen Moore
Chair, Beef Genomics, Agriculture, Food, and Nutritional Science
Canada Alberta Beef Industry Development Fund
University of Alberta
Edmonton, AB

Adilson Mota
Visiting Scientist
U.S. Department of Agriculture, Agricultural Research Service
Beltsville, MD

James D. Murray
Professor
Department of Population Health and Reproduction, and Department of Animal Science
University of California, Davis
Davis, CA

Scott Newman
Customer Programs Leader
Sygen International
Franklin, KY

Stephen James O'Brien
Sylvio Conte Senior Biomedical
Research Service Fellow
Laboratory of Genomic Diversity
National Cancer Institute
Frederick, MD

Terezinha Padilha
Research Scientist
U.S. Department of Agriculture,
Agricultural Research Service,
Immunology and Disease
Resistance Laboratory
Beltsville, MD

Albert Paszek
Cargill, Inc.
Minnetonka, MN

Aristides Patrinos
Office of Biological and
Environmental Research
Life Sciences Division
U.S. Department of Energy
Germantown, MD

Ronald Phillips
Regent's Professor
Department of Agronomy and Plant
Genetics
University of Minnesota
St. Paul, MN

Daphne Preuss
Assistant Professor
Department of Molecular Genetics
and Cell Biology
University of Chicago
Chicago, IL

Steven Pueppke
Associate Dean for Research
Consumer and Environmental
Sciences
College of Agriculture
University of Illinois
Urbana, IL

Ernest F. Retzel
Director
Center for Computational
Genomics and Bioinformatics
University of Minnesota
Minneapolis, MN

Caird Rexroad, Jr.
Associate Deputy Administrator,
Animal Production, Product Value,
and Safety
U.S. Department of Agriculture
Beltsville, MD

Max F. Rothschild
C. F. Curtiss Distinguished
Professor in Agriculture
Iowa State University
Ames, IA

James F. Schneider
Babcock Genetics, Inc.
Holmen, WI

Lawrence Schook
Professor of Comparative
Genomics
University of Illinois
Urbana, IL

Kevin T. Schultz
Member of the Management
Committee, Head of Research and
Development
Merial Limited
Duluth, GA

John Shadduck
Chief Executive Officer
Optibrand Ltd., LLC
Ft. Collins, CO

Furman G. Sizemore
Department of Microbiology and Molecular Genetics
Michigan State University
East Lansing, MI

Loren C. Skow
Professor
Department of Veterinary Anatomy and Public Health
College of Veterinary Medicine
Texas A&M University
College Station, TX

Ed J. Smith
Associate Professor
Department of Animal and Poultry Sciences
Virginia Polytechnic and State University
Blacksburg, VA

Tad S. Sonstegard
Research Geneticist
Genetic Evaluation and Mapping Laboratory
U.S. Department of Agriculture, Agricultural Research Service
Beltsville, MD

Curtis Van Tassell
Research Geneticist
Genetic Evaluation and Mapping Laboratory
U.S. Department of Agriculture, Agricultural Research Service
Beltsville, MD

Anne K. Vidaver
Chief Scientist
U.S. Department of Agriculture, Cooperative State Research, Education, and Extension Service Programs
Washington, DC

George Weinstock
Professor and Co-Director
Human Genome Sequencing Center
Baylor College of Medicine
Houston, TX

James Womack
Professor
Texas A&M University
Department of Veterinary Pathobiology
College Station, TX

Roger Wyse
Managing Director
Burrill & Company
San Francisco, CA

Nieves M. Zaldivar
Eversole Associates
Chevy Chase, MD

Appendix C

Workshop Speaker Biographies

ERIC GREEN, is Senior Investigator and Chief, Genome Technology Branch, and Director, Intramural Sequencing Center, of the National Institutes of Health (NIH) National Human Genome Research Institute (NHGRI). Dr. Green's laboratory focuses on the mapping and sequencing of eukaryotic genomes and the development of efficient approaches for utilizing the resulting information to study important biologic problems. One of his major projects currently involves mapping and sequencing genomic regions orthologous to human chromosome 7 in other animals, initially in mouse but more recently in other distantly related vertebrates. This project is being performed in close collaboration with the NIH Intramural Sequencing Center. Dr. Green's Laboratory also is interested in identifying and characterizing genes associated with human disease such as those associated with hereditary deafness (Pendred syndrome), cancer, neurologic disease, mental disorders, vascular disease, and others. He received M.D. and Ph.D. degrees from Washington University, St. Louis.

RICHARD GIBBS, is Director of the Human Genome Sequencing Center, and Professor, in the Department of Molecular and Human Genetics, at the Baylor College of Medicine. Dr. Gibbs' current focus is on producing a draft sequence of the rat genome, in collaboration with Celera Genomics Corporation and Genome Therapeutics Corporation. He received a B.Sc. (Hons) in 1979 and a Ph.D. in genetics and radiation biology in 1986 at the University of Melbourne, Melbourne, Australia. He moved to Houston as a postdoctoral fellow at the Baylor College of Medicine to examine the molecular basis of human-linked diseases, and to develop technologies used for rapid genetic analysis. In 2000, Dr. Gibbs received the annual Michael E. DeBakey, M.D., Excellence in Research Award.

STEVEN KAPPES , is Center Director of the Roman L. Hruska U.S. Meat Animal Research Center (MARC) of the U.S. Department of Agriculture-Agricultural Research Service (ARS), in Clay Center, Nebraska. Before receiving his Ph.D. in Animal Science (Molecular Biology and Reproduction) from the University of Missouri-Columbia in 1992, he worked for ARS at MARC for six years as a Cattle Operations Assistant Manager. After earning his Ph.D., Dr. Kappes started his professional career at MARC as a Research Physiologist. In 1999, he joined the ARS Animal Production, Product Value, and Safety unit of the National Program Staff in Beltsville, Maryland.

HARRIS LEWIN, is a Professor in the Department of Animal Sciences at the University of Illinois, Urbana-Champaign. Dr. Lewin's laboratory currently is focusing on three research areas, including the molecular characterization of the bovine major histocompatibility complex (MHC), the genetic and immunologic mechanisms of resistance and susceptibility to bovine leukemia virus (BLV) infection and disease progression, and gene mapping in cattle, with an emphasis on comparative mapping and mapping genes controlling economically important traits. Over the past five years his laboratory has developed a 300 marker linkage map of the bovine genome. Presently, Dr. Lewin and his colleagues are beginning to expand the map, with "expressed sequence tags," or ESTs.

STEPHEN O'BRIEN, is Chief of the Laboratory of Genomic Diversity and head of the Section of Genetics at the National Cancer Institute (NCI). Dr. O'Brien came to National Institutes of Health as a postdoctoral fellow and designed a program based on mammalian somatic cell genetics. He is co-Chair of the International Committee on Comparative Gene Mapping and Editor of *Genetic Maps.* The central focus of Dr. O'Brien's laboratory research concerns the collaborative interaction of mammalian cellular genes operative in concordant evolutionary descent of the immune system, retroviruses, and cancer onset. The purpose of these investigations is to determine the comparative mammalian genetic principles that participate in these processes.

ROGER WYSE, is Managing Director of Burrill & Company, which sponsors and manages a family of venture capital funds. Dr. Wyse has more than 27 years of experience as an internationally recognized scientist and was previously an administrator at Rutgers University and the University of Wisconsin, Madison, where he was Dean of the College of Agricultural and Life Sciences. Dr. Wyse received a Ph.D. in plant physiology from Michigan State University and a B.S. in agronomy from The Ohio State University. He served as co-CEO of Third Wave AgBio, which offers animal and plant genetics assays, was founding President and CEO of AniGenics, a fully integrated animal genomics company, and serves on the Boards of Directors of AniGenics and E-Markets.