AQUACULTURE GENOME TECHNOLOGIES

ZHANJIANG (JOHN) LIU

Blackwell Publishing

Aquaculture Genome Technologies

Aquaculture Genome Technologies

Zhanjiang (John) Liu
Auburn University

Zhanjiang (John) Liu is a Distinguished Alumni Professor, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences, and Director, Aquatic Genomics Unit at Auburn University.

Blackwell Publishing Professional
2121 State Avenue, Ames, Iowa 50014, USA

Orders: 1-800-862-6657
Office: 1-515-292-0140
Fax: 1-515-292-3348
Web site: www.blackwellprofessional.com

Blackwell Publishing Ltd
9600 Garsington Road, Oxford OX4 2DQ, UK
Tel.: +44 (0)1865 776868

Blackwell Publishing Asia
550 Swanston Street, Carlton, Victoria 3053, Australia
Tel.: +61 (0)3 8359 1011

First edition, 2007

Library of Congress Cataloging-in-Publication Data
Liu, Zhanjiang.
Aquaculture genome technologies / Zhanjiang (John) Liu. — 1st ed.
p. cm.
ISBN-13: 978-0-8138-0203-9 (alk. paper)
ISBN-10: 0-8138-0203-2 (alk. paper)
1. Genomics. 2. Aquatic genetic resources. 3. Aquaculture.
I. Title.
QH447.L58 2007
572.8′6—dc22

2006038408

The last digit is the print number: 9 8 7 6 5 4 3 2 1

Contents

Foreword

The birth of livestock genomics 15 years ago was inspired by the human genome initiative and the potential for capturing both its technologies and massive comparative data sets for application to livestock species, most of which are mammals. We are currently reaping the benefits of these efforts, with sequencing projects completed or ongoing in chickens, cattle, pigs, and horses and valuable mapping resources developed for others such as sheep and turkeys. Traits of economic and physiological significance are being mapped, and underlying genes are being discovered. The biological diversity of species used in aquaculture, however, presents a unique set of problems to genomic studies, both in technology development and in the application of genomic information to food production. A book that captures the status of genomic technologies as applied to aquaculture species and the rapid state of advancement of genomics of some of the principal species is a welcome addition to the animal genomics literature.

Species used in aquaculture span both the vertebrate and invertebrate arms of the animal kingdom and incorporate a range of issues related to genome size, genome redundancy, and a variety of reproductive strategies. With the exception of the bony fishes that should benefit from the advanced genomics of zebra fish and puffer fish, aquaculture genomics will not have the direct benefit of extensive comparative genomic data sets provided by the human genome project to mammalian genetics and surprisingly, also to the chicken genome. Nonetheless, technologies for developing DNA markers, linkage and physical maps, and transcription profiling tools are universal and DNA sequencing is now being discussed in terms of a few thousand dollars per Gb in the not too distant future. Efforts to develop tools, make maps, and ultimately sequence genomes of aquaculture species will not only be rewarded by improved health and productivity of important food sources but in defining the biology underlying the genomic and physiological diversity that make these species daunting targets for genetic studies in the first place.

A surprising wealth of tools has already been generated for genome mapping and functional studies in many of the species used in aquaculture. With the potential for sequencing on the horizon, the future is bright for aquaculture genomics. As a mammalian geneticist who thoroughly enjoys a day of sport fishing or a seafood platter, I am delighted with the progress and prospects reported in this book.

James E. Womack, Distinguished Professor
Texas A&M University
College Station, TX

Preface

The completion of the Human Genome Project inspired the entire world and triggered the start of a genomics revolution. Accompanying this revolution was a complete change in the way science was conducted in the field of life sciences. Without exception, the waves produced by the genome revolution are now having a tremendous impact on aquaculture genomics and aquaculture genetics in general. As recently as 10 years ago, there were no large-scale aquaculture genome projects in the entire world! The first Aquaculture Genome Workshop held in Dartmouth, Massachusetts in the fall of 1997 could be regarded as the official start of aquaculture genomics. Today it is a reality that the entire genomes of several important aquaculture species are on the verge of being sequenced. This raises new challenges for aquaculture geneticists, breeders, and fisheries managers regarding how to best use the huge amount of genomic information now available, and how to master and apply continuously changing genome technologies to aquaculture and fisheries.

The purpose of this book is to provide a snapshot of genome technologies from the perspectives of aquaculture and fisheries scientists, and to provide a textbook suitable for students majoring in agricultural sciences. I feel that there are several compelling reasons for producing such a book. First, while it is easy to find genomics books these days, it is rare to find books providing enough background information of the basic principles and concepts underpinning genome technologies. My background was in agriculture, but I have spent most of my recent career on basic genome research. My own experience plus that gained through teaching a graduate course on agricultural genomics suggested that in order to effectively grasp the key issues of genomics, an understanding of genome technologies is essential. Such an understanding can be gained much more effectively if the basic principles behind these technologies are clearly explained, because many students may have not systematically taken courses in molecular biology, genetics, biochemistry, bioinformatics, etc. Second, most genomics books take a pure genomics approach using classical model species examples without consideration of potential applications of genome technologies in practical settings. There is a great gap to be bridged in the understanding of how basic genomics is to be used beyond the area of human health. This book provides a thorough overview of genome technologies and their applications in aquaculture and fisheries. Third, aquaculture and fisheries species have unique biological characteristics that demand modification or adaptation of existing genome technologies. Although no chapters of this book describe novel genome technologies that have originated from or are unique to aquaculture or fisheries species, almost every chapter deals with how genome technologies can be used for aquaculture and fisheries, or for agricultural sciences in general.

This book contains 29 chapters written by well-known scientists from all over the world. Their enriched experience in both genomics and aquaculture and fisheries allowed them to provide discussions of genome technologies with unique angles that will prove to be most helpful for academic professionals, research scientists, and graduate and college students in agriculture, as well as for students of aquaculture and fisheries. In spite of its focus on aquaculture and fisheries, this book should be suitable as well for students in animal sciences, poultry science, agronomy, horticulture,

entomology, and plant pathology. I completely share the sentiments of one contributor, Dr. Eric Hallerman from Virginia Tech, as he wrote in one of his e-mails to me: "This chapter ended up being more demanding, but more rewarding to produce than I had anticipated. I ended up learning a lot, which is in part why I agreed to do the work. (Yes, teaching students was the major motivator)." Teaching students more effectively was similarly my major motivation and passion through the long process of assembling this book.

This book is divided into five parts. In Part 1, Marking Genomes, concepts, principles, and applications of various DNA marker technologies are presented. In Part 2, Mapping Genomes, various genome-mapping techniques are presented including genetic linkage mapping, QTL mapping, physical mapping, radiation hybrid mapping, and comparative mapping. In addition, the principles and applications of marker-assisted selection are presented. Topics in Part 3, Analysis of Genome Expression and Function, include EST analysis, microarrays, environmental genomics, and functional genomics. Part 4 should have been entitled Sequencing the Aquaculture Genomes, but because no genomes of aquaculture species have been sequenced, it is entitled Preparing for Genome Sequencing. This part discusses existing sequencing technologies that brought us to where we are, and the emerging sequencing technologies that will lead us into the future. Nonetheless, strategies for sequencing the genomes of aquaculture species are also discussed in this part. In the last part, Part 5, Dealing with the Daunting Genomes of Aquaculture Species, the unique biology and characteristics of aquaculture genomes are illustrated through a few examples such as the duplicated fish genomes, complexities involved in functional studies of paralogous genes, the enormously high fecundity and segregation distortion of oysters, and extremely high polymorphism in oysters as well as other bivalve species. Not only are such unique features presented in relation to genome technologies, but potential solutions are also provided, supplying researchers with potential shortcuts to avoid having to struggle through these problems again.

I would like to thank all of the chapter contributors who are truly experts in aquaculture genomics. Their willingness to share their knowledge and expertise made this book possible. I am honored to have one of the most prestigious genome scientists in the world working in the area of livestock genomics, Dr. James Womack, a member of the National Academy of Sciences USA from Texas A&M University, to write the Foreword for this book. I am grateful to my students Eric Peatman, Peng Xu, Shaolin Wang, and Jason Abernathy, and my colleague Dr. Huseyin Kucuktas who helped in proofreading some of the chapters. I have had a year of pleasant experience interacting with Erica Judisch, Editorial Assistant for Blackwell Publishing Professional, and Justin Jeffryes, Commissioning Editor for Plant Science, Agriculture, and Aquaculture with Blackwell Publishing Professional. Finally, I must thank the two most important women in my life, my mother Youzhen Wang and my wife Dongya Gao; the former inspires me to succeed, while the latter makes sure I do succeed.

Zhanjiang (John) Liu

List of Contributors

Jonas S. Almeida
Department of Biostatistics and
Applied Mathematics
University of Texas
MD Anderson Cancer Center, Unit 447
1515 Holcombe Boulevard
Houston, TX 77030 USA

Louis Bernatchez
Réseau Aquaculture Québec (RAQ)
Pavillon C-H Marchand
Université Laval
Québec, QC
Canada G1K 7P4

Robert W. Chapman
Marine Resources Research Institute
South Carolina Department of
Natural Resources
Charleston, SC 29412 USA

Alan Christoffels
Computational Biology Group
Temasek Life Sciences Laboratory
1 Research Link
National University of Singapore
Singapore

Karl J. Clark
Department of Animal Sciences
University of Minnesota
St. Paul, MN 55108 USA

Avner Cnaani
Hubbard Center for Genome Studies
University of New Hampshire
Suite 400, Gregg Hall
35 Colovos Road
Durham, NH 03824 USA

Glenn A. Cooper
Centre for Biomedical Research
University of Victoria
Victoria, British Columbia
Canada V8W 3N5

Jason P. Curole
Department of Biological Sciences
University of Southern California
3616 Trousdale Parkway, AHF 107
Los Angeles, CA 90089-0371 USA

Roy G. Danzmann
Department of Integrative Biology
University of Guelph
Guelph, Ontario
Canada N1G 2W1

William S. Davidson
Department of Molecular Biology
and Biochemistry
Simon Fraser University
8888 University Drive
Burnaby, British Columbia
Canada V5A 1S6

Chunguang Du
Department of Biology and
Molecular Biology
Montclair State University
Montclair, NJ 07043 USA

Pierre Duchesne
Réseau Aquaculture Québec (RAQ)
Pavillon C-H Marchand
Université Laval
Québec, QC
Canada G1K 7P4

Karim Gharbi
Division of Environmental and
Evolutionary Biology Institute of
Biomedical and Life Sciences
University of Glasgow
Glasgow, Scotland UK G12 8QQ

Paul J. Grobler
Faculty of Natural and
Agricultural Sciences
University of the Free State
P.O. Box 339, Bloemfontein 9300
South Africa

Ximing Guo
Haskin Shellfish Research Laboratory
Institute of Marine and Coastal Studies
Rutgers University
6959 Miller Avenue
Port Norris, NJ 08349 USA

Perry B. Hackett
Department of Genetics, Cell Biology
and Development
Arnold and Mabel Beckman Center for
Transposon Research
6-106 Jackson Hall
University of Minnesota
Minneapolis, MN 55455 USA

Eric M. Hallerman
Department of Fisheries and
Wildlife Sciences
Virginia Polytechnic Institute and State
University Blacksburg, VA 20461-0321
USA

Limei He
Department of Soil and Crop Sciences
Texas A&M University
College Station, TX 77843 USA

Dennis Hedgecock
Department of Biological Sciences
University of Southern California
3616 Trousdale Pkwy, AHF 107
Los Angeles, CA 90089-0371 USA

Jess W. Jones
U.S. Fish and Wildlife Service
Blacksburg, VA 24061-0321 USA

Thomas D. Kocher
Hubbard Center for Genome Studies
University of New Hampshire
Suite 400, Gregg Hall
35 Colovos Road
Durham, NH 03824 USA

Ben F. Koop
Centre for Biomedical Research
University of Victoria
Victoria, British Columbia
Canada V8W 3N5

Abraham Korol
Institute for Evolution
Haifa University
Haifa, Israel 31905

Huseyin Kucuktas
Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Bo-Young Lee
Hubbard Center for Genome Studies
University of New Hampshire
Suite 400, Gregg Hall
35 Colovos Road
Durham, NH 03824 USA

Yaning Li
Department of Plant Pathology
Agricultural University of Hebei
Biological Control Center of Plant
Disease and Plant Pests of Hebei
Province
Baoding, China 071001

Lei Liu
W.M. Keck Center for Comparative
and Functional Genomics
University of Illinois at
Urbana-Champaign
330 Edward R. Madigan Laboratory
1201 W. Gregory Dr.
Urbana, IL 61801 USA

Zhanjiang Liu
Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Eric Peatman
Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Caird E. Rexroad III
USDA/ARS National Center for Cool
and Cold Water Aquaculture
11861 Leetown Road
Kearneysville, WV 25430 USA

Matthew L. Rise
The Ocean Sciences Centre
Memorial University of Newfoundland,
1 Marine Lab Road
St. John's, NL
Canada A1C 5S7

Max F. Rothschild
Department of Animal Science and the
Center for Integrated Animal Genomics
2255 Kildee Hall
Iowa State University
Ames, IA 50011 USA

Anatoly Ruvinsky
The Institute for Genetics and
Bioinformatics
University of New England
Armidale, Australia NSW 2351

Chantel Scheuring
Department of Soil and
Crop Sciences
Texas A&M University
College Station, TX 77843 USA

Andrey Shirak
Agricultural Research Organization
Institute of Animal Science
Bet Dagan, Israel 50250

Kristian R. von Schalburg
Centre for Biomedical Research
University of Victoria
Victoria, British Columbia
Canada V8W 3N5

Shaolin Wang
Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Yongping Wang
Haskin Shellfish Research Laboratory
Institute of Marine and Coastal Studies
Rutgers University
6959 Miller Avenue
Port Norris, NJ 08349 USA

Gregory W. Warr
Department of Biochemistry and
Molecular Biology
Marine Biomedicine and Environmental Sciences Center
Hollings Marine Laboratory
Medical University of South Carolina
Charleston, SC 29412 USA

Peng Xu
Department of Fisheries and
Allied Aquacultures
Auburn University
Auburn, AL 36849 USA

Zhe Xu
Haskin Shellfish Research Laboratory
Institute of Marine and
Coastal Studies
Rutgers University
6959 Miller Avenue
Port Norris, NJ 08349 USA

Hong-Bin Zhang
Department of Soil and Crop Sciences
Texas A&M University
College Station, TX 77843 USA

Aquaculture Genome Technologies

Aquaculture Genome Technologies
Zhanjiang (John) Liu

Chapter 1
Concept of Genomes and Genomics

Zhanjiang Liu

When searching for the basic concept of genomics, one may find numerous definitions such as:

- The study of genes and their functions
- The study of the genome
- The molecular characterization of all the genes in a species
- The comprehensive study of the genetic information of a cell or organism
- The study of the structure and function of large numbers of genes simultaneously
- etc., etc.

In order to have a good concept of genomics, let us first explore the concept of genome, and its relationship to genome expression and genome functions.

The Concept of Genome and Genomics

The term genome is used to refer to the complete genetic material of an organism. Strictly speaking, the genetic material of an organism includes the nuclear and mitochondrial genomes for plants and animals, and also chloroplast genomes for plants. Since the mitochondrial and chloroplast genomes are small and contain only a limited number of genes, the focus of genome research is on the nuclear genome. Hence, I will limit this chapter largely to the nuclear genome.

Let us define genomics in its narrowest sense using the genetic central dogma (Figure 1.1) where in most cases, deoxyribonucleic acid (DNA) is transcribed into ribonucleic acid (RNA), and RNA is translated into proteins. Although genetic information is stored in DNA, it cannot be realized without being transcribed into the intermediate molecules RNA, which with a few exceptions, must be translated into proteins in order to have biological functions. Thus, the entire DNA content of an organism is called the genome; the entire RNA world of an organism is called its transcriptome, and the entire protein content of the organism is called its proteome. The science of studying the genome is called genomics; the science of studying the transcriptome is called transcriptomics; and the science of studying the proteome is called proteomics. In spite of such divisions, the term genomics often is used to cover not only this narrow sense of genomics, but also transcriptomics, and in some cases proteomics as well.

Genomics can be divided into structural genomics, which studies the structures, organization, and evolution of genomes, and functional genomics, which studies expression and functions of the genomes. Since genome functions are reflected in the transcripts and proteins that the transcripts encode, genomics must also study the transcriptome and the proteome.

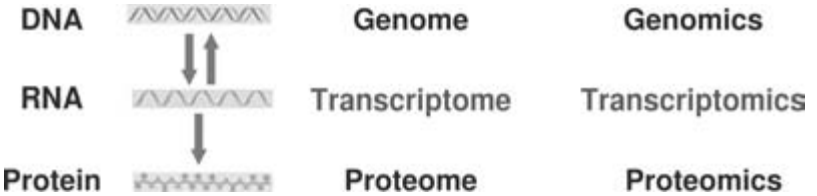

Figure 1.1. The concept of genome and genomics in relation to the genetic central dogma. The entire DNA content of an organism (the genome) is transcribed into RNA (the entire RNA content of the organism is called the transcriptome), and the RNA is translated into proteins (the proteome). Genomics, transcriptomics, and proteomics are sciences that study the genome, transcriptome, and proteome, respectively.

It must be pointed out that while the genome is relatively stable in an organism in most cell types with the exception of gene rearrangements in immune-related cell types, the transcriptome is highly dynamic. The types of transcripts and their relative levels of expression are highly regulated by tissue specificity, developmental stage, physiological state, and the environment. For instance, if an organism has 25,000 genes, not all genes are expressed in every type of cell. Those genes required for the basic cell structure and functions are probably expressed in all tissues, organs, and cell types; whereas each cell type expresses a subset of the genes specific for that cell type. Many genes are expressed throughout development, but certain genes are expressed only at a specific developmental stage. Physiological state can affect gene expression in a fundamental and dramatic way. For instance, gonadotropin genes are expressed only in the pituitary and gonad, and expressed highly during spawning seasons of the reproductive cycle in fish. The environment can insert its effect on gene expression in multiple dimensions. Temperature, pH, water quality, stress, dissolved oxygen, and many other environmental factors can induce or suppress expression of a large number of genes.

In addition to the dynamic nature of the transcriptome, variation of the transcriptome can also be brought about by production of alternative transcripts by the same set of genes. It is now widely believed that the complexity of the transcriptome is much larger than the genome because of alternative transcripts. The largest proportion of alternative transcripts is produced by alternative splicing where a single gene is transcribed into heterogeneous nuclear mRNA (hnRNA); through splicing, more than one mRNA molecule is produced, leading to the phenomenon that introns of one transcript may be exons of another. The second mechanism for the generation of alternative transcripts is through the use of alternative promoters. In a single gene, more than one promoter can be functional leading to the generation of different, but related transcripts. In addition, use of differential polyadenylation sites can also lead to the generation of alternative transcripts. Therefore, it is widely believed that the information stored in the genome is amplified and diversified at the transcriptome level. The genetic information is further amplified and diversified at the protein level. Though each transcript may only encode one protein, the primary protein may be differentially processed to produce more than one active polypeptide; posttranslational glycosylation, acetylation, phosphorylation, and other modifications can result in a much larger complexity leading to drastically different biological functions. Even highly related gene products may encode proteins leading to absolutely opposite biological functions. For instance, an interleukin-1 Type II receptor is a decoy target for

IL-1, whose binding to interleukin 1 intercepts the function of interleukin 1. Therefore, the genetic central dogma is correct in terms of the basic flow of genetic information, and the capacities of the primary functions of transcription and translation, while much larger complexities result from amplification and diversification of the same set of genetic material, lead to the generation of biologically different molecules. Such differences in biological molecules, when considered for the various combinations of many genes, can result in numerous biological outcomes.

As a new branch of science, genomics has its own defined scope of study, its own box of tool kits, and its own unique set of approaches. It is different from traditional molecular genetics which looks at single genes, one or a few genes at a time. Genomics is trying to look at all of the genes as a dynamic system, over time, to determine how they interact and influence biological pathways, networks, physiology, and systems in a global sense. Genome technologies, the focus of this book, have been developed to cope with the global scope of tens of thousands of genes as a snapshot. Much like dealing with a globe, landmarks (or as we have called them molecular markers) are needed to mark the position within the huge genome. Genetic and physical maps have been developed to understand the structure and organization of genomes, and to understand genomic environs and genome evolution in relation to genome expression and function. Specific approaches have been developed to cope with the large number of genes, regardless if it is for gene discovery, cloning and characterization, or for analysis of gene expression. Thus large-scale analysis of expressed sequence tags using highly normalized complementary DNA (cDNA) libraries allows rapid gene discovery and cloning in the scale of tens of thousand of genes. Such operations have also been supported by other genome technologies such as powerful automated sequencing to allow gene discovery and identification in a streamlined industrial fashion. Expression of genes is determined in an entire genome scale, or sometimes referred to as genome expression, to relate complex regulation of genes to their functions in terms of systems biology. Expression of tens of thousands of genes can be monitored simultaneously and continuously, allowing their interactions and networking to be detected. Signal transduction is no longer "behind the scene" molecular events, but can be observed with clustering of co-regulated gene expression under specific development, physiology, or environmental conditions. Genes and their functions are studied much in terms of their sociology, networking, and interactions, rather than looking at one or a few genes at a time, as conducted by traditional molecular biology. Such operations demand the development of very powerful gene expression analysis such as microarray technologies. Such technological advances allow the generation of tremendously large data sets that have been beyond the comprehension capacities of biologists. Assistance is needed from all areas of biology, and more so from disciplines outside biology that can handle large amounts of information. Computer sciences and mathematics are among the first disciplines genomics has demanded cooperation from. While handling large data sets from the genome, genome expression, and genome function, much confusion has emerged regarding whether the observed phenomenon is real or if it is just a fluctuation of the systems biology. As such, statisticians are also called upon to join computer scientists, mathematicians, and the biologists. Because these scientists speak different languages (e.g., English for one group, French for the second, Chinese for the next, and so on), understanding all of the languages and being able to function among these different disciplines is becoming the goal of a large group of scientists who define themselves as bioinformaticians working in the new area of bioinformatics. It is clear that genomics cannot be a science without

bioinformatics. Clearly, the definition of genomics is becoming more complex with this discussion. Now, you can certainly come up with your own definitions.

The excitement and success of genomics has brought the emergence of numerous '–omics' sciences (http://genomicglossaries.com/content/genomics_glossary.asp). Subbranches of genomics are emerging in large numbers. The following list includes some of those subbranches:

- agricultural genomics
- applied genomics
- behavior genomics
- biochemical genomics
- chemogenomics
- clinical genomics
- combinatorial genomics
- comparative genomics
- computational genomics
- deductive genomics
- ecotoxicogenomics
- environmental genomics
- evolutionary genomics
- forward genomics
- functional genomics
- immunogenomics
- industrial genomics
- intergenomics
- inverse genomics
- lateral genomics
- nanogenomics
- network genomics
- oncogenomics
- pharmacogenomics
- phylogenomics
- physiological genomics
- population genomics
- predictive genomics
- reverse genomics
- structural genomics
- toxicogenomics
- translational genomics
- and so on

Cells, Nucleus, Chromosomes, Genomes, and Genomic DNA

Genomes can exist in various forms. A genome can be either RNA or DNA, single-stranded or double-stranded. For example, the human immunodeficiency virus (HIV) is a retrovirus whose genome contains a single-stranded RNA molecule. However, such unusual genomes are mostly found within viruses and bacteriophages. In prokaryotes such as bacteria, by definition they do not have a nucleus; the genomes are made up with double-stranded DNA in either circular or linear forms. For instance, the *Escherichia coli* genome is made of a single circular DNA molecule, whereas the genome of *Borrelia burgdorferi* is composed of a linear chromosome approximately one megabase (million base) in size. Eukaryotic genomes contain two or more linear molecules of double-stranded DNA in the form of chromosomes.

Within each eukaryotic cell, there is a nucleus in which chromosomes are located. Individual species harbor a fixed number of chromosome pairs (2n) with fixed shapes, sizes, and centromere location. These chromosome morphologies are commonly known as the karyotypes. All somatic cells in a diploid organism harbor identical chromosome pairs that are randomly shared into a single chromosome set during meiosis to produce eggs and sperms. Upon fertilization of an egg (n) by a sperm (n), the embryo recovers the diploid state with two sets of chromosomes.

Chromosomes are threadlike structures containing genes and other DNA in the nucleus of a cell. Different kinds of organisms have different numbers of chromosomes.

Humans have 46 chromosomes—44 autosomes and 2 sex chromosomes. Each parent contributes one chromosome to each pair, so children get half of their chromosomes from their mothers and half from their fathers. This is important in sexual reproduction where the gametes (i.e., sperms and eggs) are haploid cells, and upon fertilization of an egg by a sperm, the embryo recovers the diploid state. The number of chromosomes is usually constant for each organism, but may vary greatly from species to species. For instance, the fruit fly *Drosophila melanogaster* has four chromosomes whereas *Ophioglossum reticulatum*, a species of fern, has the largest number of chromosomes with more than 1,260 (630 pairs). The minimum number of chromosomes found in a species occurs in a species of ant, *Myrmecia pilosula*, in which females have one pair of chromosomes and males have just a single chromosome. This species reproduces through a process called haplodiploidy, in which fertilized eggs (diploid) become females, while unfertilized eggs (haploid) develop into males.

Each chromosome is a portion of the genome and all the chromosomes compose the entire genome. Although all chromosomes maintain their own integrity, they each can be viewed as a segment of the genome. The total length of genomic DNA thus is equal to the sum of all chromosomal DNA. In their natural existence, the physical pieces of DNA in each cell are equal to the number of chromosomes. It must be emphasized that such entire chromosomal DNA is essentially impossible to obtain for routine molecular analysis. Chromosomal DNA is randomly broken during genomic DNA extraction even under the most sophisticated preparation by the most skilled researchers. Most often, millions of cells are used in a single DNA extraction. Therefore, genomic DNA used in molecular analysis represents multiple copies of the genome with multiple overlapping segments, simply because the breakage points are random and different in each cell genome.

Genome Sizes

Genome sizes of organisms vary greatly, spanning a range of almost 100,000 fold. The bacterial genomes are commonly at the range of a million base pairs (Mbp), while the largest animal genome reported to date is 133 picograms (pg) (or about 1.3×10^{11} base pairs; 1 pg DNA = 1×10^{-12} g = 978 Mbps) for a species of lungfish, *Protopterus aethiopicus*, which is some 40 times larger than the human genome, followed by a number of amphibians, *Necturus lewisi* and *N. punctatus* at 120 pg, *Necturus maculosus* and *Amphiuma means* and the lungfish *Lepidosiren paradoxa*, all at roughly 80 pg. In general, the genome size is correlated with biological complexities, but many exceptions exist. For instance, some plant species and amphibians can have very large genomes, dozens of times larger than the human genome.

The largest teleost genome size is 4.4 pg in the masked *Corydoras metae*, and the smallest teleost genome size is approximately 0.4 pg in several puffer fish of the family *Tetraodontidae*. Fish as a whole have the largest ranges for genome sizes.

Crustaceans also have a wide range of genome sizes from 0.16 pg to 38 pg with an average of 3.15 pg. The smallest crustacean genome size (0.16 pg) is in a water flea, *Scapholeberis kingii*, and the largest crustacean genome size (38 pg) is in *Hymenodora* sp., a deep-sea shrimp. The most important crustacean species for aquaculture involves several major species of the shrimps. Their genome sizes are approximately 2.5 pg.

The molluscan genome sizes are more uniform ranging from the smallest molluscan genome size of 0.4 pg in the owl limpet *Lottia gigantean*, to the largest molluscan genome size of 5.9 pg in the Antarctic whelk *Neobuccinum eatoni*. Many aquacultured shellfish belong to the molluscans. The most important of these species in aquaculture include the oysters, such as the Pacific oyster (genome size 0.91 pg), the eastern oyster (genome size 0.69 pg), and the scallops (genome size between 0.95 to 2.1 pg).

The size of the genome of an organism is a constant. However, the ploidy of organisms varies. For instance, channel catfish are believed to be a diploid organism, whereas most salmonid fish used in aquaculture are believed to be tetraploid. In cultivated wheat plants, various ploidies exist including diploid, tetraploid, and hexaploid. In order to standardize the genome size so that they can be compared, genome sizes are presented in C-values, which is the haploid genome size in picograms.

Several excellent databases exist for genome sizes. The Animal Genome Database (http://genomesize.com/) is a comprehensive catalogue of animal genome size data. It includes haploid genome sizes for more than 4,000 species including approximately 2,750 vertebrates and 1,315 invertebrates compiled from 5,400 records from more than 425 published sources (Gregory 2005; Animal Genome Size Database, http://www.genomesize.com/). The Database Of Genome Sizes (DOGS) (http://www.cbs.dtu.dk/databases/DOGS/) is also a very useful database that includes a number of links to genome size and genome research resources such as the following:

- the Plant DNA C-Value Database (http://www.rbgkew.org.uk/cval/)
- the Genome Atlases for Sequenced Genomes (http://www.cbs.dtu.dk/services/GenomeAtlas/)
- the DBA mammalian genome size database (http://www.unipv.it/webbio/dbagsdb.htm)
- several other useful databases and resources.

Knowledge of genome size is not only important for genome studies in relation to genome structure, organization, and evolution, but also for a number of practical reasons such as genome mapping, physical mapping, and genome sequencing. As listed in Table 1.1, the primary methods for the determination of genome sizes are Feulgen densitometry (Hardie et al. 2002), flow cytometry, and Feulgen image analysis densitometry (Lamatsch et al. 2000). These three methods account for over 81% of all methods used for the estimation of genome sizes (Table 1.1). Readers with an interest in methodologies for the determination of genome size are referred to the literature list of the Animal Genome Size Database (http://www.genomesize.com/).

Number of Genes

The number of genes in a given organism is fixed, but discovering it is a daunting task. For the best characterized human genome, the number of genes now is believed to be approximately 25,000. In the 1980s, the number of human genes was believed to be 100,000 to 125,000. In the early 1990s, the human genome was believed to include 80,000 genes. Although the final completion of the Human Genome Project was celebrated in April 2003 and sequencing of the human chromosomes is essentially "finished," the exact number of genes encoded by the genome is still unknown. In