EVOLUTION OF GENES AND PROTEINS

Edited by Masatoshi Nei

and Richard K. Koehn

EVOLUTION OF GENES AND PROTEINS

Edited by Masatoshi Nei

CENTER FOR DEMOGRAPHIC
AND POPULATION GENETICS
UNIVERSITY OF TEXAS, HOUSTON

and Richard K. Koehn

DEPARTMENT OF ECOLOGY
AND EVOLUTION
STATE UNIVERSITY OF NEW YORK,
STONY BROOK

SINAUER ASSOCIATES INC. • PUBLISHERS
Sunderland, Massachusetts 01375

EVOLUTION OF GENES AND PROTEINS

Copyright © 1983 by Sinauer Associates, Inc.

All rights reserved.

This book may not be reproduced in whole or in part by any means without permission from the publisher. For information address
Sinauer Associates, Inc.
Sunderland, MA 01375

Library of Congress Cataloging in Publication Data

Main entry under title:
Evolution of genes and proteins.

Bibliography: p.
Includes index.
1. Chemical evolution. 2. Deoxyribonucleic acid.
3. Proteins. 4. Genetic polymorphisms.
I. Nei, Masatoshi. II. Koehn, Richard K.
QH371.E925 1983 574.87'328 83-477
ISBN 0-87893-603-3
ISBN 0-87893-604-1 (pbk.)

Printed in U.S.A.

6 5 4 3 2 1

PREFACE

The study of evolution at the molecular level has experienced two periods of exciting development during the past two decades. The first period started when the techniques of amino acid sequencing and protein electrophoresis were introduced in evolutionary studies in the early 1960s and lasted for about 10 years. During this period the approximate constancy of the rate of amino acid substitution in evolution and the extensive protein polymorphism in natural populations were discovered. These new findings led to the proposal of various new evolutionary theories, some of which were quite controversial. Controversy was particularly heated over Kimura's neutral mutation theory, which proclaimed that most amino acid substitutions and protein polymorphisms are caused not by Darwinian selection but by neutral mutation and random genetic drift. At the same time, the discovery of approximate constancy of amino acid substitution provided new methods for dating the evolutionary history of organisms as well as methods for constructing phylogenetic trees from molecular data. In the 1970s evolutionary geneticists were extremely busy examining the validity of the new evolutionary theories and applying the new methods of constructing phylogenetic trees.

The second period started only a few years ago and is not yet over. This period was initiated by introduction of a new set of biochemical techniques: DNA sequencing, recombinant DNA, and restriction enzyme methods. These new techniques have already uncovered many unexpected properties of the structure and organization of genes (e.g., exons, introns, flanking regions, repetitive DNA, pseudogenes, gene families, and transposons) and of their evolution. It is now clear that most genes do not exist as a single copy in the genome but, rather, in clusters, and that the number of genes in a cluster varies extensively from cluster to cluster. Comparison of nucleotide sequences from diverse organisms indicates that the rate of sequence change in evolution varies considerably with the DNA region examined and that the more important the function of the DNA region the lower the rate of sequence change. Furthermore, the extent of genetic variation undetectable by protein electrophoresis is enormous. These discoveries again have led a number of evolutionists to propose new evolutionary theories such as concerted evolution and horizontal gene transfer.

Although many other exciting discoveries will undoubtedly be made in the near future, it now seems appropriate to examine and summarize in book form the general implications of these new findings, together with those in the first period of development. Such a book is needed not only for students and scholars of evolution but also for biologists in general. However, it is not an easy job for one or two persons to write such a book, because the subject is so diversified and the progress in each area is so rapid. We therefore decided to produce a book on molecular evolution with the help of experts from various specialized areas of the subject. This was facilitated by a symposium on "Evolution of Genes and Proteins" which we organized for the joint meeting of the Society for the Study of Evolution and the American Society of Naturalists held at the State University of New York at Stony Brook, June 23–24, 1982. This symposium was attended by more than 500 scientists and students from the United States and abroad. In this symposium we attempted to cover all important areas of study in molecular evolution, inviting both molecular biologists and population biologists. Each symposium speaker was then asked to write a chapter of textbook, rather than a usual symposium paper, in order to make the book understandable to a wide range of readers, from students to active researchers.

While a book written by multiple authors has a definite advantage in covering the latest developments in diverse areas of a subject, it suffers from such deficiencies as heterogeneity of writing style, repetitions, and inconsistencies. In the present book we have worked to minimize these deficiencies by making extensive editorial changes of original manuscripts; in some cases we even rewrote parts of the original text. However, we have tried to avoid any change of an author's opinions even if they contradicted another author's views. In the forefront of research, scientists do not always agree with each other, and this disagreement often becomes an impetus for further progress. The reader will notice that there are some inconsistencies in the use of scientific terminology. For example, a noncoding DNA sequence between two coding regions of a gene is called an intervening sequence (IVS) in Chapter 1 but an intron in the other chapters. We have left such inconsistencies, because no consensus has yet been achieved in the scientific community for these terms.

The chapters of this book can be divided loosely into three groups. The first group includes the first four chapters, all of which are concerned with the long-term evolution of DNA. In Chapter 1 Edgell and his associates discuss the evolution of globin gene clusters as a model case of gene evolution. Globin genes have been very important in elucidating the evolutionary change of gene structure in the last few years. In Chapters 2 and 3 the evolutionary significance of gene duplicaton and the mechanism of concerted evolution are discussed in

此为试读,需要完整PDF请访问: www.ertongbook.com

the light of new findings at the DNA level. Chapter 4 deals with the evolutionary change of mitochondrial DNA. Mitochondrial DNA evolves much faster than nuclear DNA and thus is very useful for studying the phylogenetic relationships of closely related species.

The second group consists of Chapters 5 to 9, which are mainly concerned with the genetic variation within species. The major issue in these chapters is the maintenance of genetic polymorphism in natural populations, and data on both protein and DNA polymorphisms are examined. The controversy over the neutral mutation theory is still alive. However, unlike a decade ago, neutralists and selectionists are no longer hostile to each other, and the gap between the views of the two groups of scientists has narrowed substantially. Data on DNA polymorphism are still scanty compared with those on protein polymorphism but clearly show that the genetic variability at the DNA level is enormous. In Chapter 8 Avise and Lansman show that mitochondrial DNA is a useful genetic material for tracing back the evolutionary history of populations.

The last four chapters, which make up the third group, deal with several current evolutionary theories. (Chapter 9 can be included in this group as well as in the second group.) In Chapter 10 Jukes presents an interesting theory on the evolution of the amino acid (genetic) code, taking advantage of recent discoveries of non-universal amino acid codes in mitrochondrial genes. In Chapter 11 Kimura discusses recent developments in the neutral theory of molecular evolution. In Chapter 12 Hall describes experimental observations on the evolution of new metabolic functions in microorganisms. The last chapter is concerned with transposons, i.e., genetic elements which move within and between chromosomes. The evolutionary significance of transposons is still largely speculative, but they are potentially important in explaining the existence of repetitive DNA in higher organisms.

We would like to express our hearty thanks to the contributors of this book for writing excellent chapters and for being tolerant of our editorial suggestions and changes. We hope our joint enterprise will be successful in bringing the latest knowledge of molecular evolution to both students and scientists who are interested in the diversity and evolution of organisms.

MASATOSHI NEI
RICHARD K. KOEHN
November 17, 1982

CONTRIBUTORS

Stylianos E. Antonarakis, Department of Pediatrics, Genetics Unit, Johns Hopkins University School of Medicine, Baltimore

Norman Arnheim, Department of Biochemistry, State University of New York, Stony Brook

John C. Avise, Department of Molecular and Population Genetics, University of Georgia, Athens

Betty Brown, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Wesley M. Brown, Division of Biological Sciences, University of Michigan, Ann Arbor

Frank Burton, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Allan Campbell, Department of Biological Sciences, Stanford University, Stanford

Aravinda Chakravarti, Department of Biostatistics, University of Pittsburgh, Pittsburgh

Mary Comer, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Marshall H. Edgell, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Barry G. Hall, Department of Microbiology, University of Connecticut, Storrs

John G. Hall, Department of Ecology and Evolution, State University of New York, Stony Brook

Stephen C. Hardies, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Alison Hill, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Clyde A. Hutchison, III, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Thomas H. Jukes, Space Sciences Laboratory, University of California, Berkeley

Haig H. Kazazian, Jr., Department of Pediatrics, Genetics Unit, Johns Hopkins University School of Medicine, Baltimore

Motoo Kimura, National Institute of Genetics, Mishima, Japan

Richard K. Koehn, Department of Ecology and Evolution, State University of New York, Stony Brook

Robert A. Lansman, Department of Molecular and Population Genetics, University of Georgia, Athens

Wen-Hsiung Li, Center for Demographic and Population Genetics, University of Texas, Houston

Masatoshi Nei, Center for Demographic and Population Genetics, University of Texas, Houston

Stuart H. Orkin, Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston

Sandra Phillips, Jackson Laboratory, Bar Harbor, Maine

Robert K. Selander, Department of Biology, University of Rochester, Rochester

Charlie Voliva, Department of Bacteriology and Immunology, University of North Carolina, Chapel Hill

Steven Weaver, Department of Biological Sciences, University of Illinois, Chicago

Thomas S. Whittam, Department of Biology, University of Rochester, Rochester

Anthony J. Zera, Department of Ecology and Evolution, State University of New York, Stony Brook

ACKNOWLEDGMENTS

This volume is based on a symposium, "Evolution of Genes and Proteins," organized by R. K. Koehn and M. Nei, and sponsored by The Society for the Study of Evolution at Stony Brook, New York, June 23–24, 1982. We gratefully acknowledge financial sponsorship of the symposium and this volume by the Offices of the Vice Provost for Research and Graduate Studies and the Dean of Biological Sciences of the State University of New York, Stony Brook, the Stony Brook Foundation, and Grant DEB 8118404 from the National Science Foundation. Kathleen Ward, Center for Demographic and Population Genetics, University of Texas at Houston, compiled and checked the Bibliography. Her careful and diligent work is greatly appreciated.

Chapter 2/Wen-Hsiung Li
The author thanks Takashi Gojobori for his help in the preparation of the manuscript and Gregory Whitt for discussions. This study was supported by grants from the National Science Foundation and the National Institutes of Health.

Chapter 6/Richard Koehn, Anthony Zera, and John Hall
Preparation of the manuscript was supported by USPHS Grant GM 21131 and NSF Grant DEB 7908802. This is contribution 431 from Ecology and Evolution, State University of New York, Stony Brook.

Chapter 8/John C. Avise and Robert A. Lansman
This work has been supported by NSF Grants DEB 7814195 and DEB 8022135. The authors thank Charles Aquadro and Berry Greenberg for supplying unpublished data. Charles Aquadro also critically reviewed the manuscript.

Chapter 9/Masatoshi Nei
The assistance from Takashi Gojobori and Dan Graur in the preparation of the manuscript is acknowledged. The author's research was supported by grants from the National Institutes of Health and the National Science Foundation.

Chapter 10/Thomas Jukes
Support from NASA Grant NGR 05-003-460 and the assistance of Carol Fegté are acknowledged.

This book was set in Linotron 202 Century Schoolbook at DEKR Corporation. The production team consisted of Jodi Simpson, copy editor, Joseph J. Vesely, designer and production coordinator, and Fredrick J. Schoenborn, illustrator. The book was manufactured by R. R. Donnelley & Sons.

CONTENTS

此为试读,需要完整PDF请访问: www.ertongbook.com

CHAPTER 1

EVOLUTION OF THE MOUSE β GLOBIN COMPLEX LOCUS

M. H. Edgell, Stephen C. Hardies, Betty Brown, Charlie Voliva, Alison Hill, Sandra Phillips, Mary Comer, Frank Burton, Steven Weaver, and Clyde A. Hutchison III

The globins are an exceptionally well studied gene system and hence represent an excellent molecular data base with which we can articulate and challenge assumptions concerning the evolution of DNA sequences. Our interests, as molecular geneticists, have been primarily to recognize the regulatory elements within the gene system and secondarily to understand the biological mechanisms that control genome organization. As such, we are newcomers to evolutionary analysis, and hence what we would like to do in this opening chapter is to use the mouse β globin genes as an opportunity to raise issues. Those issues will be addressed more fully either elsewhere in this volume or in the technical literature.

One can imagine using comparative biochemical genetics as a method of recognizing important components of a complex gene system like the β globins. Utilizing such a strategy, one might compare the β globin gene clusters of various species such as the mouse, rabbit, goat, and human in order to identify conserved features in the loci. In such a manner, assuming that the sequence features important to globin metabolism will have changed less than other features, one would expect to find the important regulatory elements relatively conserved. This approach seems threatened by the considerable divergence in gene organization actually found when one compares the

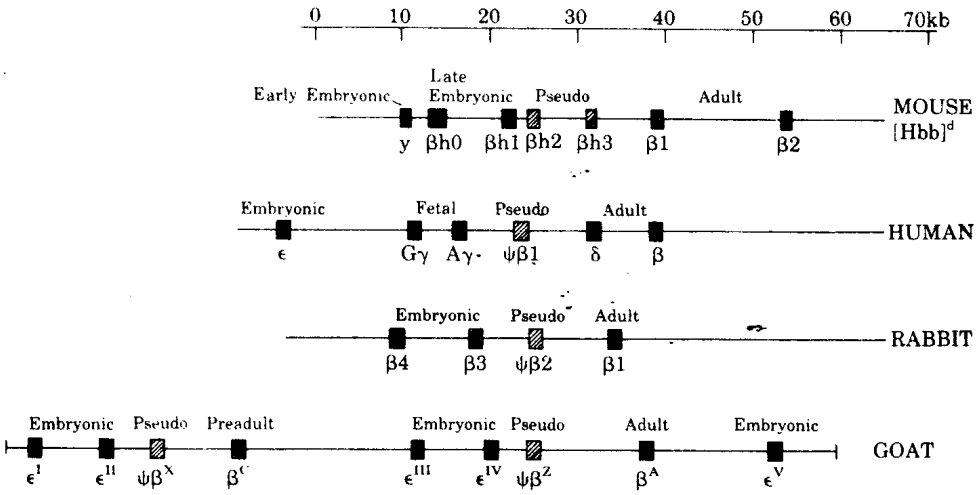

FIGURE 1. Mammalian β globin complex loci.

characterized complex globin gene loci (Figure 1). The gene clusters do not in fact share the same number of gene-like structures, and therefore it is not a trivial matter to decide which genes are to be compared to which (i.e., which are the evolutionarily homologous or "orthologous" sequences in the various species). That there has been considerable divergence in the organizational features of the gene family when one compares different species seems to be a general property of the genome and is not just a special property of the globins. Hence, the identification across species of the gene pairs that are truly orthologous becomes a serious issue. Generally, we feel that the regulation of a gene system is intimately tied into the structural features of the gene cluster. This unexpected degree of structural divergence must, therefore, cause us to at least consider the possibility that the regulatory features of the various globin clusters may not, in fact, be identical.

MOUSE β GLOBINS

The mouse β globin haplotype, [*Hbb*]^{d}, specifies four β-like proteins: two adult β globins and two nonadult β globins (Russell and McFarland, 1974). However, the 65 kilobases (kb) of DNA cloned from this locus contains seven gene-like structures (Tilghman et al., 1977; Jahn et al., 1980; Edgell et al., 1981) with sequence homology to the adult β globin genes (Figure 1). Three of these were shown by sequence analysis (Jahn et al., 1980; Konkel et al., 1978, 1979; Hansen et al., 1982) to correspond to known β globin proteins: the adult proteins d-major (dmaj) and d-minor (dmin) and the nonadult protein y. In order

to have unique gene names and to retain the traditional gene/allele nomenclature, we have renamed the adult genes β1 and β2 to replace the β previously used for both adult genes (β^{maj} and β^{min}, respectively). The four additional genes were given the designation βh, which refers to "beta homologous."

We have done extensive cross-hybridization analyses to define the locations of repetitive sequences within the complex locus (Figure 2). At least six repetitive DNA families have been defined in the locus. The fraction of sequence that is repetitive is not uniform within the β globin locus. At the resolution of *Hae*III fragments probed with labeled genomic DNA, the embryonic region contains only 15% repetitive DNA as compared to 50% in the adult region (F. Burton, pers. comm.).

We have examined the other β globin haplotype prevalent in *Mus musculus*, $[Hbb]^s$, by library construction, cloning and characterization (Weaver et al., 1981). Given the considerable degree of organizational divergence between species, we were interested in determining the degree of homology existing between two haplotypes. At the level of electron-micrographic heteroduplex analysis, we have found very few differences between the haplotypes (Figure 3). There is an interesting pair of insertions in $[Hbb]^d$ near the β2 gene, but in general the haplotypes are quite homologous. However, not all of the structures within the locus are evolving at the same rate. For example, as we will demonstrate later, βh2 and βh3 are evolving much more rapidly than the functional genes. Therefore, the observed extent of homology implies either that the two haplotypes have only recently become distinct or that there is nonreciprocal sequence exchange between the two haplotypes on a quite large scale.

The $[Hbb]^s$ haplotype gives rise to only a single adult β globin. However, both β1 and β2 are transcribed in this haplotype (S. Weaver

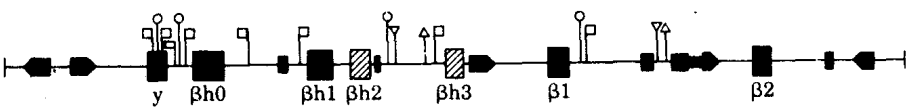

FIGURE 2. The β globin complex locus in *Mus musculus* haplotype $[Hbb]^d$. Functional genes are indicated by the large filled-in blocks and pseudogenes by the hatched blocks. The smaller filled-in blocks of various shapes mark the location and size of the repetitive sequence family which we have most extensively characterized. The stick/flag symbols tag the locations of at least five other repetitive sequences. These latter sequences are probably quite short.

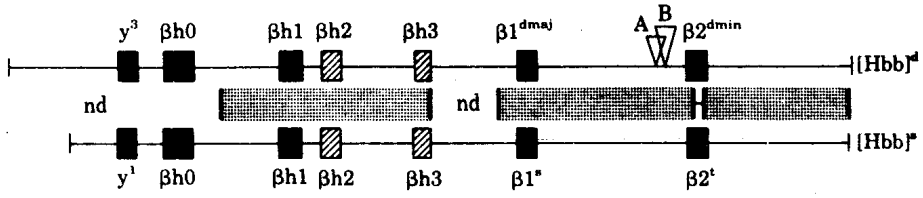

FIGURE 3. Homology relations between two globin haplotypes of the mouse (*M. musculus*). The stippled regions between the two maps indicate where "perfect" heteroduplexes form when examined by electron microscopy. The A and B inserts are each approximately 1.5 kilobases in length.

and B. Brown, pers. comm.). Sequence data from the first coding block of the two genes (S. Weaver, pers. comm.) suggest that they have identical coding sequences. Presumably, these adult genes have been subjected to a recent gene conversion event. The *s* allele of β1 is very homologous to the dmaj sequence, but the $\beta 2^{dmin}$ gene is quite different from the other three adult genes.

HOMOLOGY WITHIN THE COMPLEX β GLOBIN LOCUS

Globin genes consist of three coding blocks (exons) and two intervening sequences (introns). A comparison of the large intervening sequences (IVS2s) of $\beta 1^{dmaj}$ and $\beta 2^{dmin}$ indicates considerable divergence (Figure 4). It is usually concluded that because the intervening sequences (IVSs) are more divergent than their associated coding sequences, they must be under less selective constraint than the coding blocks (Chapter 11 by Kimura). This interpretation is quite consistent with our observation that the nucleotide sequences of the IVSs are more divergent than those found in coding blocks. It has been known for a long time that mutations do not distribute themselves uniformly within a sequence (Benzer and Champe, 1961; Drake, 1970). Mutational hotspots and differences in rates for transversions and transitions have been identified in many different systems. Although some of this may be due to selection, these effects are usually attributed to the nature of the nucleotide sequences and mutational processes. Clearly the number of observed differences we see between two sequences is a complex function of nucleotide sequence and depends on both mutational susceptibility and selective constraints.

Generally, the number of observed mutations is considered to be equal to the intrinsic mutation rate times fixation processes. In the absence of selection, the observed mutation rate within a sequence is dependent on both sequence susceptibility and repair. For example, a poly(T) sequence will accumulate more changes due to ultraviolet

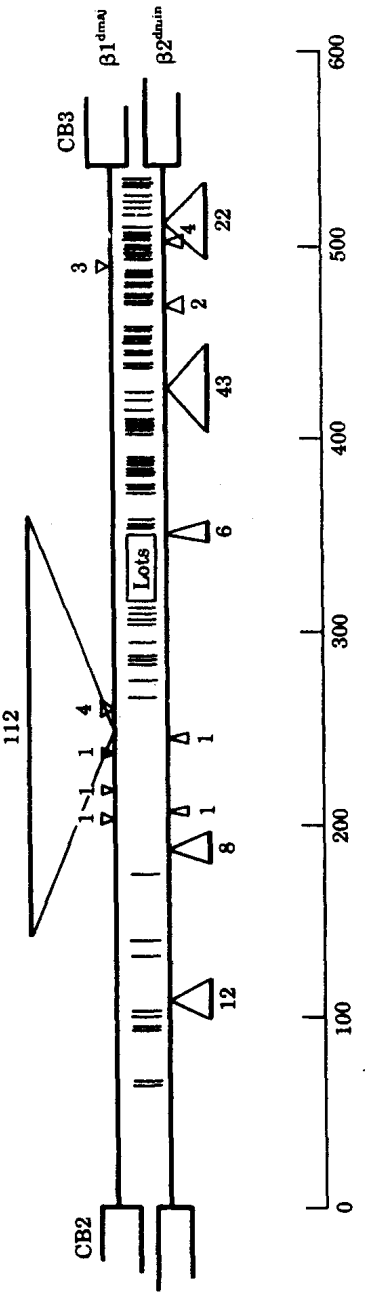

FIGURE 4. A comparison of the large intervening sequences (IVS2s) from two adult β globin genes. CB2 and CB3 refer to the second and third coding blocks, respectively. The vertical lines indicate point mutations and the triangles insertions of the indicated length in nucleotides.

此为试读,需要完整PDF请访问: www.ertongbook.com