THIRD EDITION

introduction to veterinary genetics

F.W. NICHOLAS



WILEY-BLACKWELL

Copyrighted material

Introduction to Veterinary Genetics

Copyrighted material

Copyrighted material

Introduction to Veterinary Genetics

Third Edition

F.W. Nicholas

Faculty of Veterinary Science,
University of Sydney, Australia

WILEY-BLACKWELL

A John Wiley & Sons, Ltd., Publication

Copyrighted material

First edition published 1996 by Oxford University Press
Second edition published 2003 by Blackwell Publishing Ltd
This edition first published 2010
© 2010 Blackwell Publishing Ltd

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing programme has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial office
2121 State Avenue, Ames, Iowa 50014-8300, USA

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Nicholas, F.W.
Introduction to veterinary genetics / F.W. Nicholas. – 3rd ed.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-1-4051-6832-8 (pbk. : alk. paper)
1. Veterinary genetics. I. Title.
[DNLM: 1. Genetics. 2. Veterinary Medicine. 3. Animal Population Groups–genetics. 4. Animals. SF 756.5 N638i 2010]

SF756.5.N52 2010
636.08'21–dc22

2009013260

A catalogue record for this book is available from the British Library.

Set in 10 on 12.5 pt Times by SNP Best-set Typesetter Ltd., Hong Kong
Printed in Singapore

1 2010

Copyrighted material

Contents

Preface x

1 Basic genetics 1
Chromosomes 1
Meiosis and mitosis 4
The biochemistry of inheritance 11
What is a gene? 18
Gene regulation 22
Mutation 23
Genes, alleles, and loci 24
Simple or Mendelian inheritance 25
Linkage 27
Inactivation 32
Types of DNA 35
Further reading 38
Appendix 1.1 Banded karyotypes of domestic species 39

2 Molecular biology 46
Restriction enzymes 46
Recombinant DNA and DNA cloning 49
Complementary DNA 51
DNA sequencing 53
Polymerase chain reaction 55
Southern analysis and related technologies 58
DNA expression microarrays 58
The detection of variation in base sequence 60
Veterinary diagnosis 63
Variable number of tandem repeats (VNTR), DNA fingerprints, and microsatellites 64
Single nucleotide polymorphisms (SNPs) 68
Copy number variation (CNV) 69
Gene mapping 70
Whole-genome sequence assembly 74
Production of polypeptide from cloned DNA 76

Copyrighted material

Transgenesis 77
Antisense technology 80
RNA interference 81
Further reading 84

3 **Single-gene disorders** 85
Inborn errors of metabolism 85
Sex-limited inheritance 87
Genetic heterogeneity of disease 88
Type of gene action and type of disease 90
Phenocopies 91
A sample of single-gene disorders 92
A revolution in mapping and identifying the causal mutation of single-gene disorders 96
Further reading 98
Appendix 3.1 A sample of single-gene traits that have been characterized at the molecular level 98

4 **Chromosomal aberrations** 103
Abnormal chromosome number 103
Abnormal chromosome structure 107
Chromosomal aberrations in cancer 112
Evolution of karyotypes 113
Interspecific hybridization 113
Fossemartins 114
Biological basis of sex 115
Classification of intersex 116
A sample of chromosomal aberrations 117
Further reading 119
Appendix 4.1 A sample of chromosomal aberrations in animals 119

5 **Single genes in populations** 121
Gene and genotype frequencies 121
Random mating 122
The Hardy–Weinberg law 123
Extensions to the Hardy–Weinberg law 125
Selection and mutation 127
Genetic drift and the founder effect 133
Extending population genetics to more than one locus 134
Further reading 137

6 **Familial disorders not due to a single gene** 139
Liability and threshold 139
The multifactorial model 139
More than one threshold 141

Copyrighted material

Some final points 143
Further reading 149

7 Is it inherited? 150
General evidence for a genetic aetiology 150
The four types of simple, Mendelian inheritance 151
Studying and analysing the data 153
Further reading 157

8 Immunogenetics 158
Antibodies 159
Red-cell antigens 160
The major histocompatibility complex (MHC) 163
Further reading 171

9 Pharmacogenetics 172
Genetic polymorphisms affecting drug metabolism 172
Genetics and anaesthesia 173
Warfarin resistance 174
Multifactorial pharmacogenetics 175
Further reading 178

10 Hosts, parasites, and pathogens 179
Host–pathogen interactions 179
Resistance in hosts 183
Resistance in parasites and pathogens 184
Control of parasites and pathogens 188
Increasing the level of resistance in hosts 190
Further reading 193

11 Single genes in animal breeding 195
Coat colour 195
Examples of coat-colour genes 195
Carpet wool 200
Prolificacy in sheep 200
Polledness 201
Muscular hypertrophy in cattle and sheep 201
Dwarf poultry 202
Genes for sexing chickens 202
Pedigree checking 204
Further reading 206

12 Relationship and inbreeding 207
The inbreeding coefficient 207
Relationship 208

Copyrighted material

The inbreeding coefficient revisited 209
A general expression for relationship and inbreeding 210
The base population 211
Inbreeding in populations 212
Inbreeding depression 212
Further reading 215

13 Quantitative variation 216
Quantitative traits 216
The performance of an individual animal 218
The differences between animals 220
Heritability 221
Correlations between traits 224
Quantitative trait loci (QTL) 224
Further reading 228

14 Selection between populations 230
Comparison between populations 230
Genotype–environment interaction 232
Further reading 234

15 Selection within populations 235
Estimated breeding values and accuracy of selection 235
Clues to a candidate's breeding value for a trait 236
Combining clues from more than one source 236
Best linear unbiased prediction (BLUP) 238
Correlated traits 239
Selection for more than one trait 240
The importance of inbreeding and genetic drift 242
Sire-reference schemes 243
Marker-assisted selection (MAS) and genome-wide selection (GWS) 244
Further reading 247

16 Breed structure 249
The traditional pyramid 249
Closed-nucleus breeding schemes 249
Open-nucleus breeding schemes 250
Information nucleus 252
Further reading 253

17 Crossing 254
Regular crossing 254
Crossing to produce a synthetic 260
Grading-up 261
Further reading 265

Copyrighted material

18 Selection and regular crossing **267**
Selection 267
Selection and regular crossing 267
Further reading 269

19 Biotechnology and the future **270**
Artificial insemination (AI) 270
Multiple ovulation and embryo transfer (MOET) 271
In vitro maturation (IVM) and *in vitro* fertilization (IVF) of ova 272
Control of sex ratio 273
Recombinant proteins 273
Transgenesis 274
Animal cloning 275
Further reading 279

20 Conservation genetics **280**
Measurement of genetic diversity within populations 280
Measurement of genetic diversity among populations 282
Importance of genetic diversity 283
Loss of genetic diversity 284
Conservation of genetic diversity 286
Further reading 289

21 Genetic and environmental control of inherited disorders **291**
Environmental control of inherited disorders 291
Genetic control of single-gene disorders 294
Gene therapy 300
Genetic control of multifactorial disorders 300
Genetic control – some final points 302
Crossing: thinking outside the square 304
Further reading 306

Glossary 307

Index 310

Copyrighted material

Preface

Can there ever have been a more exciting time to study genetics? Important new topics such as next-generation sequencing, genome sequence assemblies, SNP chips, copy number variation, non-coding RNA, genome-wide selection, RNA interference, and expression QTL have all been included, albeit briefly, in this new edition. Older topics that have been given a new lease of life by recent developments, such as linkage disequilibrium, have been somewhat expanded, but still the surface has been only skimmed. The fruits of the molecular revolution are now very much in evidence: in the six years since the previous edition, the number of single-locus traits characterized at the molecular level in at least one animal species has more than trebled from 81 to 265. And, given the amazing power of SNP chips in facilitating these discoveries (as also briefly discussed in this edition), this number is set to explode in the next few years. Even now, there are so many wonderful examples to illustrate basic principles that some very difficult decisions have had to be made in order to keep this edition to a reasonable length. As compensation, readers are invited to access Online Mendelian Inheritance in Animals (OMIA) at http://omia.angis.org.au for up-to-date information on all documented single-locus disorders and other traits; and, in contrast to the previous edition's unfulfilled promise, the OMIA web site also provides access to a home page for this book. A major innovation in this edition are dot-point summaries at the end of each chapter. It is hoped that these will help readers to grasp the big picture.

Many of the topics in this book, especially those concerning population and quantitative genetics, are covered in greater detail in *Veterinary Genetics* (Clarendon Press, Oxford, 1987). References to that publication, particularly to its derivations, are provided in the text of the present volume.

Thanks are due to many colleagues who have provided information and/or checked drafts: Leopoldo Iannuzzi, Steve O'Brien, Marilyn Raymond, Dave Burt, Max Rothschild, Darren Griffin, Brian Kinghorn, Ross Tellam, Mehar Khatkar, Claire Wade, Sasha Graphodatsky, Roscoe Stanyon, Darren Griffin, Martin Völker, Jill Maddox, Ben Hayes, John James, Hannah Nicholas, Belinda Norris, Jonathan Usmar, Bethany Wilson, Peter Thomson, Claire Wade, Noelle Cockett, Rob Banks, Mohammad Shariflou, Malcolm Ferguson-Smith, and Xuan Zhang.

Thanks are also due to the team from Wiley-Blackwell: Justinia Wood, Amy Brown, Adam Burbage, Sophie Gillanders, James Sowden, Katy Loftus, Lesley Simon and Anne Bassett. I am very grateful for your invaluable professional guidance and assistance. Finally, very special thanks are due to my wife, Jan, for understanding that retirement does not necessarily mean walking away from it all.

F. W. N
Sydney
July 2009

Copyrighted material

1 Basic genetics

This chapter provides a review of basic genetics. It concentrates on the general principles that apply to normal, healthy animals. The exceptions to these principles are often the basis of genetic diseases, which are discussed in subsequent chapters.

Chromosomes

When a culture of rapidly dividing white blood cells is treated with the alkaloid colchicine (which halts cell division), and the cells are then stained and viewed under a light microscope, structures called **chromosomes** become clearly visible. They are scattered randomly within clusters, and each cluster contains all the chromosomes from just one cell. The area of genetics concerned with chromosomes is called **cytogenetics**.

In order to study chromosomes more closely, a suitable cluster is chosen, as shown in Fig. 1.1a. Each item in the cluster consists of two rod-like structures joined together at a constricted point. Each rod-like structure is a **chromatid** and the constriction is a **centromere**. The two chromatids that are joined at the centromere have just been formed from one original chromosome. If the cell division had been allowed to proceed, the centromere would have split and each separate chromatid would then be called a new chromosome. For convenience, we talk of each pair of chromatids joined at the centromere as being just one chromosome, referring in fact to the chromosome that has just given rise to them.

All the chromosomes in the cluster are then rearranged in order of size. An arrangement such as this provides a picture of the complete set of chromosomes or **karyotype** of a cell (Fig. 1.1b). If many such arrangements are examined from normal, healthy individuals of both sexes of any species of mammal or bird, two facts become evident: each species has a characteristic karyotype and, within any species, each sex has a characteristic karyotype.

Karyotypes of different species differ in the shape, size, and number of their chromosomes. Within any species, all the chromosomes occur in pairs. In individuals of one sex, both members of each chromosome pair have the same size and shape. In the other sex, all but two chromosomes occur in such pairs, with the remaining

Copyrighted material

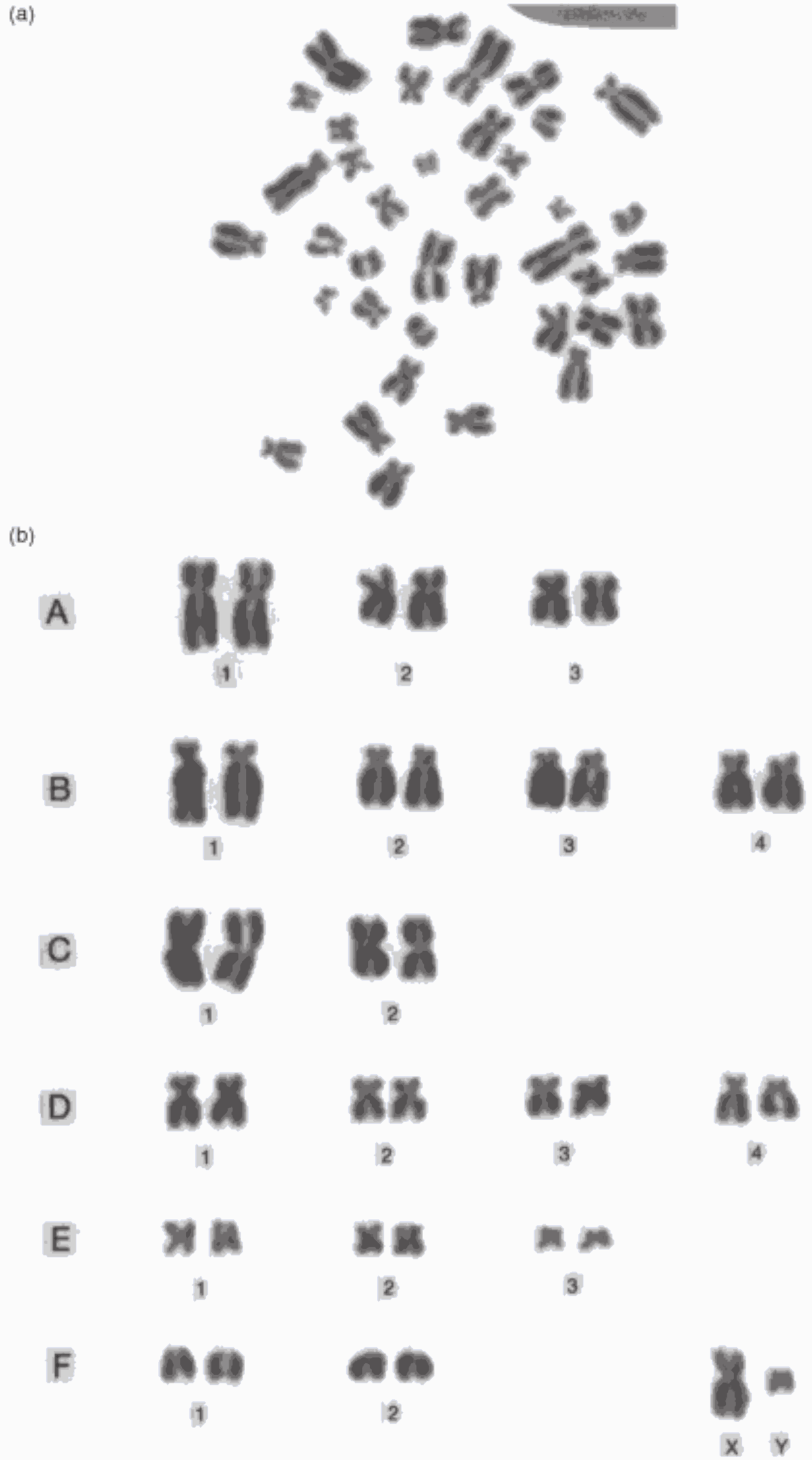


Fig. 1.1 (a) The chromosomes of a male cat, as seen through a light microscope. (b) The karyotype of a male cat, as obtained by rearranging individual chromosomes from (a). (Reproduced courtesy of P. Muir.)

pair consisting of two chromosomes of different size and shape. In this unequal pair, one chromosome has the same shape and size as members of one of the pairs in the opposite sex.

The difference in karyotype between the two sexes is the key to sex determination. In mammals, the two chromosomes that form the unequal pair occur in males, and are called the X and Y chromosomes. In female mammals, one of the pairs of chromosomes consists of two X chromosomes. Thus in mammals, males are XY and females are XX. The X and the Y chromosomes are known as **sex chromosomes**. In birds, the sex chromosomes are given different names, and their relationship to sex is the opposite of that in mammals: male birds are ZZ and female birds are ZW. For convenience, we shall refer only to mammals in the following discussion, although all statements apply equally to birds if the names of the sexes are reversed.

Chromosomes other than the sex chromosomes are called **autosomes**. Within any species, males and females have the same set of autosomes, occurring in pairs. The sex chromosomes plus the autosomes constitute a **genome**, which is the total set of chromosomes in a cell. Genomes in which chromosomes occur in pairs are said to be **diploid**, and the two members of a pair are called **homologues**. In order to emphasize that chromosomes occur in pairs, the total number of chromosomes is called the $2n$ number, where n is the number of pairs. For example, the number of chromosomes in the karyotype illustrated in Fig. 1.1 is $2n = 38$. To enable identification of each pair of chromosomes in a karyotype, the autosome pairs are labelled according to an internationally agreed convention, as shown in Fig. 1.1b. The two sex chromosomes are placed at the end.

In order to describe karyotypes more fully, chromosomes are often classified according to whether the centromere is at one end (**acrocentric**), closer to one end than the other (**sub-metacentric**) or in the middle (**metacentric**). In this book, we shall follow common practice in using metacentric to cover both metacentric and sub-metacentric. The short arm of each chromosome is designated p (think of petite = small), and the long arm is designated q. (If the centromere is in the centre of the chromosome, the designation of which arm is called p is arbitrary, but is agreed by international convention; for acrocentric chromosomes, e.g. cattle autosomes, there is only one arm which, by convention, is designated the q arm). A summary description of the karyotypes of common domestic species is given in Table 1.1. Avian karyotypes are somewhat different to mammalian karyotypes, in that their smallest chromosomes are far smaller than the smallest mammalian chromosome. Until recently, chicken chromosomes were categorized as either macrochromosomes or microchromosomes, but now, in recognition of the continuum of sizes, they are classified into four groups, from group A (the largest chromosomes) to group D (the smallest).

Banding

When karyotypes were first investigated, individual pairs of chromosomes could be identified only according to their shape and size. Since then, various methods of staining chromosomes have been developed, giving rise to alternating light and dark

Copyrighted material

Table 1.1 A summary description of the karyotypes of some domestic species

Species	Total diploid number ($2n$=)	Autosomal pairs	
		Metacentrics	Acrocentrics
Cat, *Felis catus*	38	16	2
Dog, *Canis familiaris*	78	0	38
Pig, *Sus scrofa domesticus*	38	12	6
Goat, *Capra hircus*	60	0	29
Sheep, *Ovis aries*	54	3	23
Cattle, *Bos taurus*	60	0	29
Horse, *Equus caballus*	64	13	18
Donkey, *Equus asinus*	62	24	6
Alpaca, *Lama pacos* / Llama, *Lama glama*	74	16	20
Rabbit, *Oryctolagus cuniculus*	44	19	2
Chicken, *Gallus gallus*	78	7	31

regions called **bands**. The main types of bands are broadly classified as G (stained with Giemsa), Q (stained with quinacrine), R (reverse of G-banding), C (centromeres are stained), and T (telomeres are stained). Among specific banding strategies, we have G-banding by trypsin treatment and Giemsa staining (GTG), G-banding by early bromodeoxyuridine (BrdU) incorporation and Giemsa staining (GBG), R-banding by late BrdU incorporation and acridine-orange staining (RBA), and R-banding by late BrdU incorporation and Giemsa staining (RBG).

As an example of banding, the GBG bands of cattle are illustrated in Fig. 1.2. Since the position, width, and number of bands are different for each pair of chromosomes, each chromosome pair can be identified by its banding pattern. By studying many cells treated in the same way, it is possible to draw up an **idiogram**, which is a representation of the characteristic banding pattern for each pair of chromosomes. The bands are uniquely identified according to a convention known as the International System for Cytogenetic Nomenclature of Domestic Bovids (ISCNDB). Each arm is divided into a small number of regions which are numbered sequentially starting from the centromere. Then, in each region, the bands are numbered sequentially starting nearest the centromere. For example, the second band in the third region of chromosome 1 in cattle is designated q132, while the second band in the fourth region of the long arm of the X chromosome is Xq42. The ISCNDB idiogram for cattle is illustrated in Fig. 1.3. Banded karyotypes of other domestic species are illustrated in Appendix 1.1.

Meiosis and mitosis

For many thousands of years, humans have observed two phenomena in relation to sex determination in animals: first, that there is considerable variation in the numbers of each sex among the offspring of pairs of parents; and second, that despite this

Copyrighted material

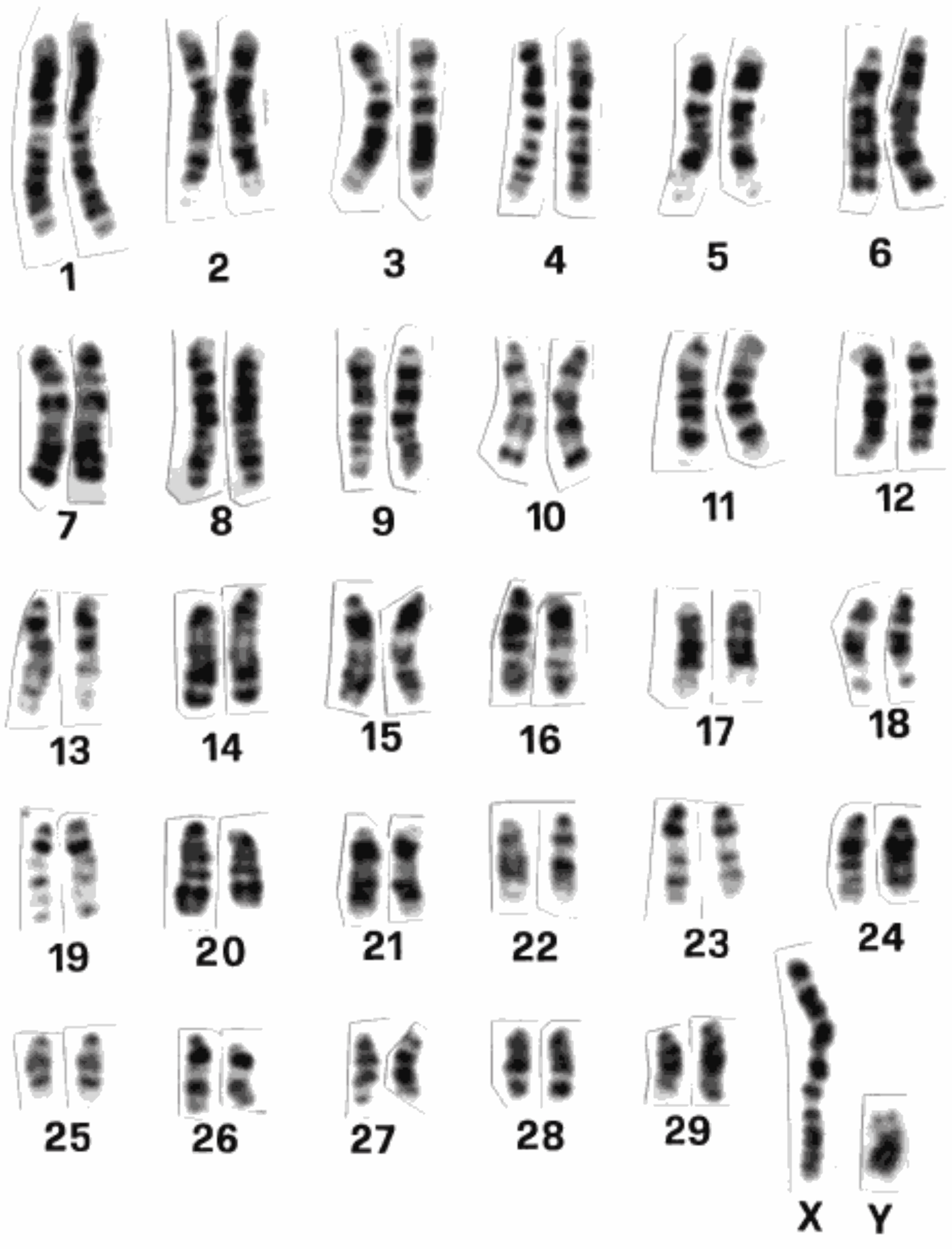


Fig. 1.2 The standard GBG-banded cattle karyotype. (Reprinted by permission from S. Karger AG, Basel: ISCNDB (2000), Di Berardino, Di Meo, Gallagher, Hayes and Iannuzzi, *Cytogenetics and Cell Genetics*, **92**, 283–99.)

variation, the overall numbers of males and females across families are approximately equal.

As noted above, the difference in sex chromosomes between the two sexes is the key to sex determination. The reason why XX individuals are females and XY individuals are males is explained in Chapter 4. For the present, we shall ask simply: why is there so much variation in the numbers of XX and XY individuals in the

Copyrighted material

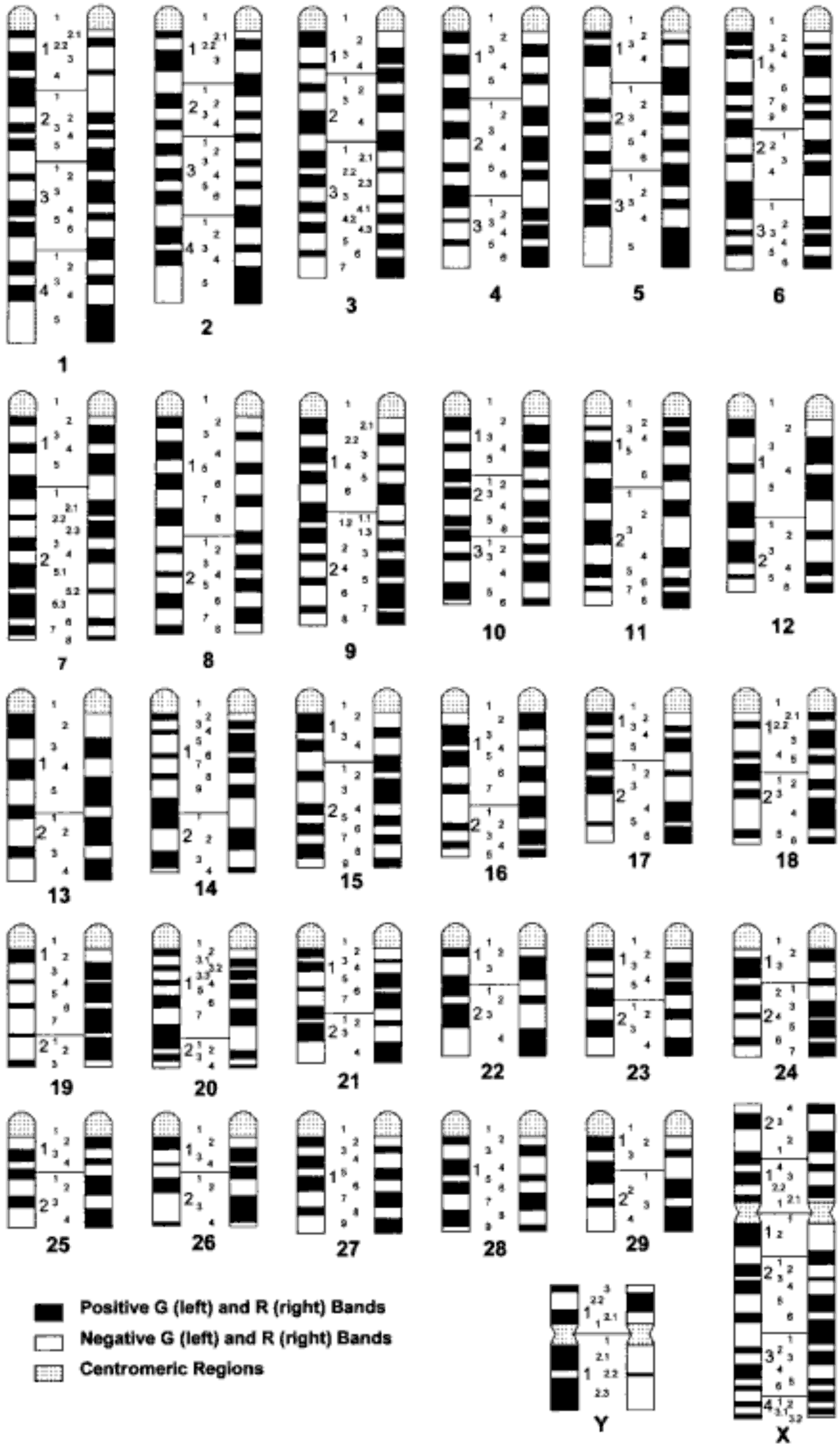


Fig. 1.3 The standard cattle idiogram, showing both G-bands (*left*) and R-bands (*right*). (Reprinted by permission from S. Karger AG, Basel: ISCNDB (2000), Di Berardino, Di Meo, Gallagher, Hayes and Iannuzzi, *Cytogenetics and Cell Genetics*, **92**, 283–99.)

Copyrighted material

offspring of pairs of parents, and yet at the same time approximately equal numbers of each sex overall? The answer lies in an understanding of gamete formation.

Meiosis

Meiosis is the process of gamete formation in which sperms are formed in testes of males and ova are formed in ovaries of females. The main result of meiosis is that each sperm and each ovum contains one member of each pair of chromosomes. Containing exactly one half of the usual diploid number of chromosomes, gametes are said to be **haploid**. The union of a sperm with an ovum at fertilization produces a **zygote** with the usual diploid number of chromosomes.

The process of meiosis commences with a normal cell containing the usual diploid set of chromosomes. To make the explanation easier, we shall consider what happens to just one pair of chromosomes (the sex chromosomes) in one sex (females), as illustrated in Fig. 1.4. In order to distinguish the two X chromosomes in females, we shall refer to them as X_p (paternal: originating from the father) and X_m (maternal: originating from the mother).

Meiosis occurs in two stages. Meiosis I begins with each chromosome duplicating itself, giving rise to two identical chromatids joined at the centromere. Then homologous chromosomes, in our case X_p and X_m, line up next to each other in the centre of the cell, in a process known as **pairing** or **synapsis**. This is facilitated by a protein structure called the **synaptonemal complex**, which 'zips' the two homologues together. The pair of homologues is called a **bivalent**. Because each chromosome has already duplicated itself into two chromatids, there are now four chromatids side by side in the cell; two X_p chromatids and two X_m chromatids. The two X_p chromatids are still joined at their centromere, as are the two X_m chromatids. At this stage, a process called **recombination** or **crossing-over** occurs, in which homologous chromatids each break at the same site and, in the process of re-uniting, exchange segments. This produces a cross-like structure called a **chiasma** (plural: **chiasmata**). In order to simplify the present discussion, we shall continue to refer to the chromatids as X_p or X_m, realizing that, as a result of crossing-over, any one chromatid may in fact consist of parts of both X_p and X_m. (A full discussion of the genetic implications of crossing-over is presented later in this chapter.) In the next stage of meiosis I, the two centromeres are pulled to opposite ends or **poles** of the cell, with the result that the two X_p chromatids move to one pole of the cell and the two X_m chromatids move to the other pole. Since this process involves the two pairs of chromatids disjoining from their previous paired arrangement, it is known as **disjunction**. In the final stage of meiosis I, the original cell divides into two cells; one contains the two X_p chromatids still joined at their centromere, and the other contains the two X_m chromatids, still joined at their centromere.

Following disjunction in females, only one cell continues to function normally; the other degenerates into a dark-staining structure known as the **first polar body**. It is entirely a matter of chance which of the cells remains functional. Consequently, there is an equal chance of either the two X_p chromatids or the two X_m chromatids ending up in the functional cell. (In Fig. 1.4, it happens to be the X_p chromatids that

Copyrighted material

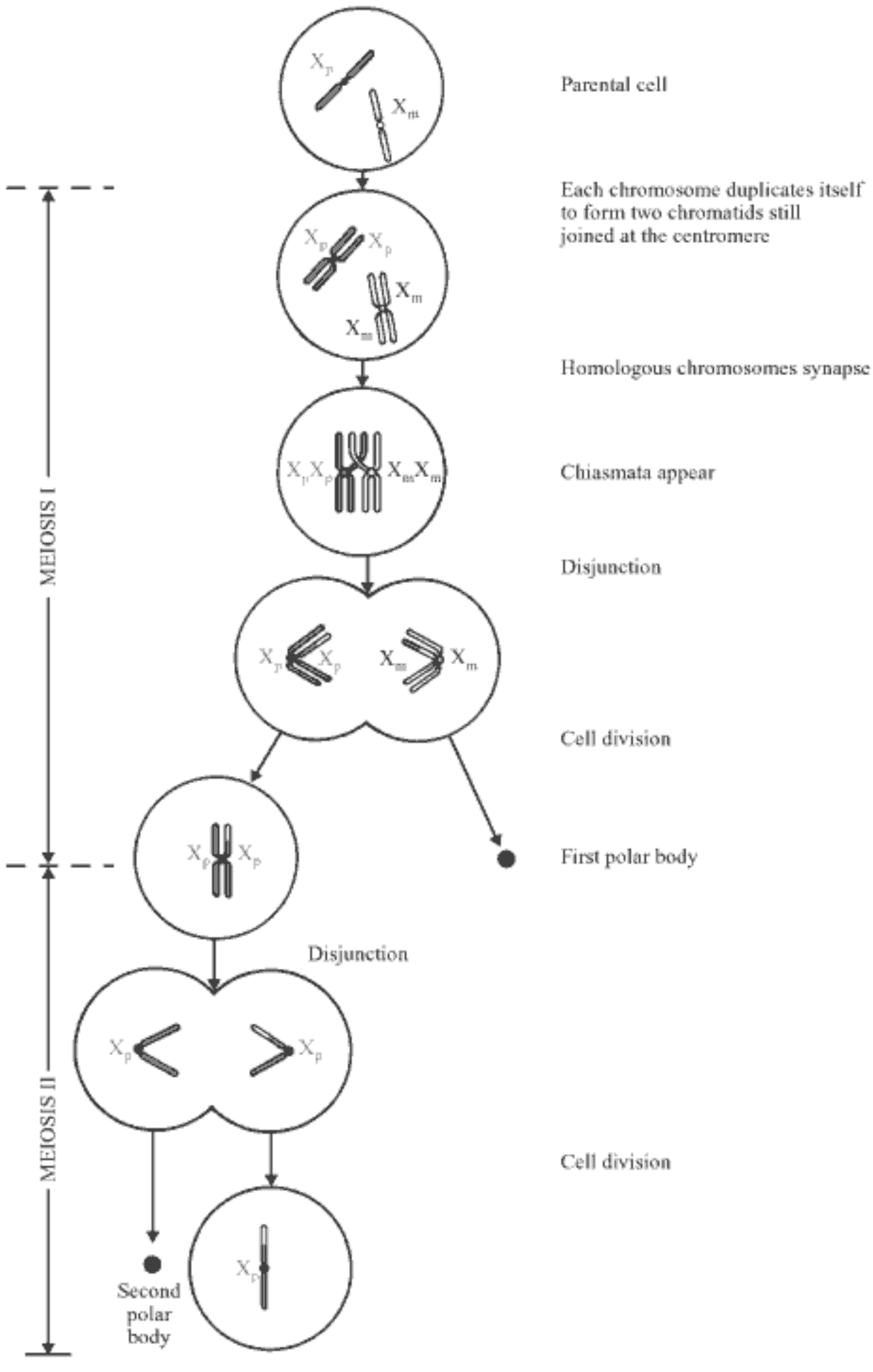


Fig. 1.4 Meiosis in a female, illustrated in terms of the sex chromosomes. A single cross-over has created two recombinant and two non-recombinant chromatids. The functional gamete in this example contains a recombinant chromosome with a paternal centromere. Exactly the same processes occur for all pairs of autosomes.

Copyrighted material

have survived.) In meiosis II in females, the two chromatids in the functional cell move apart (disjoin) and the cell divides into two cells, each containing one chromatid which is now called a chromosome. Once again, only one of the two cells remains functional; the other degenerates into the **second polar body** and, once again, it is entirely a matter of chance as to which of these two cells becomes the second polar body.

It is evident that in females, only one functional gamete results from each cell that originally underwent meiosis. It is also obvious that, irrespective of which cell ultimately remains functional, all gametes produced by females are the same in the sense that each contains one X chromosome. For this reason, females are known as the **homogametic** sex.

In males, meiosis is basically the same as described above: a disjunction followed by a cell division in meiosis I, and the same in meiosis II (Fig. 1.5). There are, however, two important differences. The first is that the X and Y chromosomes have only a small homologous region at the end of one arm (called the **pseudo-autosomal region**) where synapsis occurs; for the remainder of their length, the arms are not joined together. Despite this unusual arrangement, their subsequent disjunction is normal, and gives rise to two functional cells at the end of meiosis I: one contains two X chromatids still joined at their centromere, and the other contains two Y chromatids still joined at their centromere. The second difference between meiosis in females and in males is that polar bodies are not formed in males. Instead, both of the cells formed at the end of meiosis I undergo a cell division in meiosis II, giving rise to four functional gametes (sperms), two of which contain an X chromosome and two of which contain a Y chromosome. Since males produce two different types of gametes, they are known as the **heterogametic** sex.

Having now produced the gametes, the next stage is fertilization which, genetically speaking, is largely a matter of chance.

Chance and variation

Since all female gametes contain an X chromosome, the chance of a female gamete containing an X chromosome is one. In contrast, males produce equal numbers of X-bearing gametes and Y-bearing gametes. There is, therefore, a chance of $\frac{1}{2}$ that a particular sperm contains an X and the same chance that it contains a Y. It follows that the chance of obtaining an XY zygote is $1 \times \frac{1}{2}$ which equals $\frac{1}{2}$. Similarly, the chance of obtaining an XX zygote is $1 \times \frac{1}{2}$, or $\frac{1}{2}$. We can represent this situation by using a common genetic device called a checkerboard or **Punnett square**, in which the proportion at the head of each column is multiplied by the proportion at the head of each row, to give the expected proportions of offspring in the body of the checkerboard:

	Male gametes ½ X	Male gametes ½ Y
Female gametes all X	½ XX	½ XY

Copyrighted material

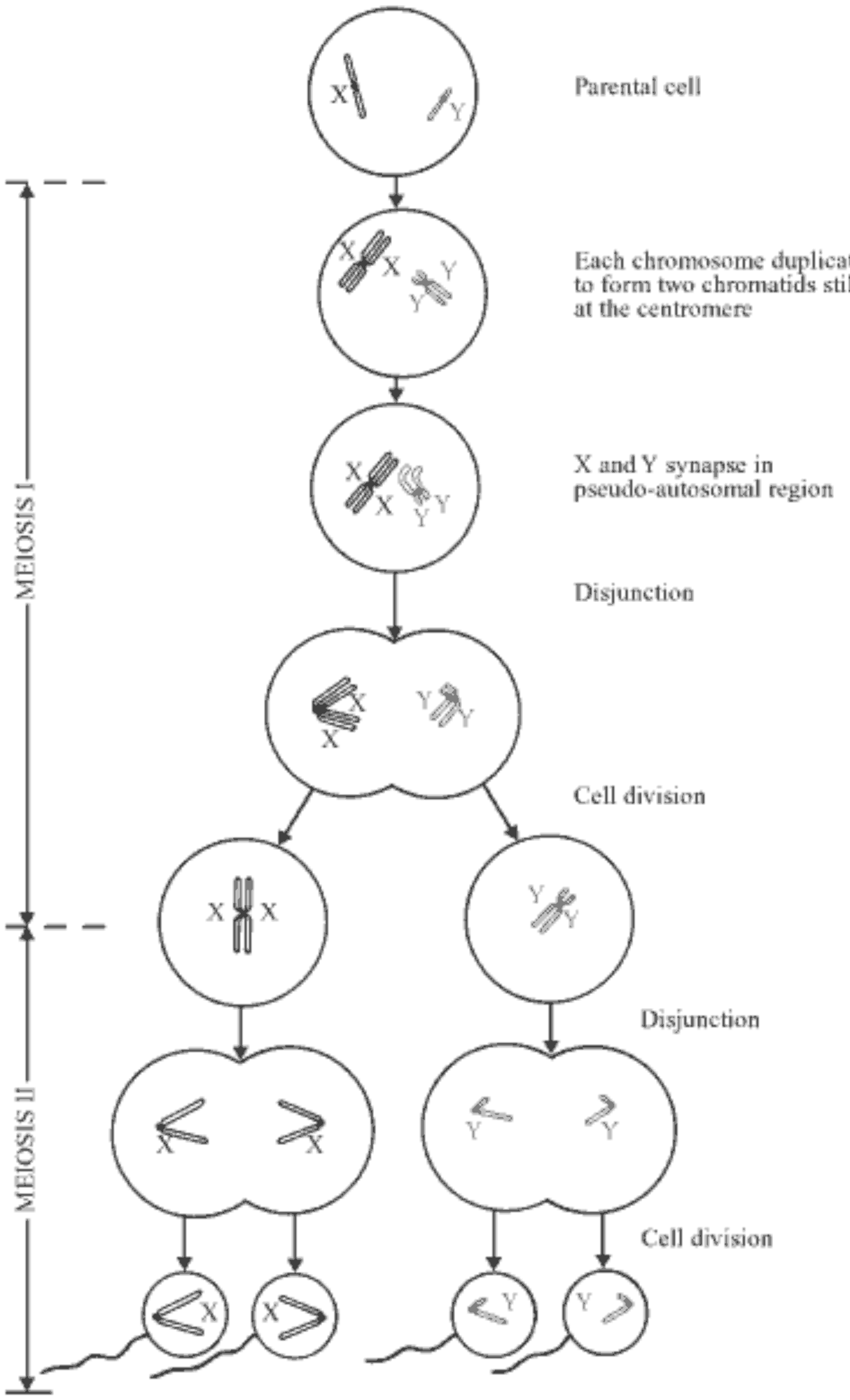


Fig. 1.5 Meiosis in a male, illustrated in terms of the sex chromosomes. With the exception of the unusual pairing in meiosis I, exactly the same processes occur for all pairs of autosomes.

We have now seen how meiosis enables the production of an expected equal proportion of each sex, which accounts for one of our original observations. How can we account for the second observation, concerning the considerable variation in numbers of each sex among the offspring of different pairs of parents? There is just one fact that enables us to explain this variation: each fertilization is an independent event. By this we mean that irrespective of whether an X-bearing or Y-bearing sperm

Copyrighted material

is successful with a particular ovum, the result of that fertilization has no bearing on subsequent fertilizations, even if they occur at the same time. For example, in a female that ovulates four ova, the chance that the last ovum is fertilized by a Y-bearing sperm is exactly $\frac{1}{2}$ irrespective of which type of sperm fertilized the other ova. In fact, any particular sequence of sexes, e.g. MMFM, is just as likely as any other sequence, e.g. FFFF.

We have now provided adequate explanations for each of the observations described earlier. In so doing, we have discussed chromosomes, simple inheritance, and chance, each of which is basic to an understanding of genetics. In order to complete the cycle of reproduction on which we embarked when discussing meiosis, we need to pass, by a process known as mitosis, from the zygote to an adult capable of producing its own gametes.

Mitosis

The growth of a single-celled zygote into a multicellular adult involves a mechanism whereby the number of cells can be expanded rapidly, while at the same time ensuring that each cell has exactly the same set of chromosomes as the original single-celled zygote. Mitosis is such a mechanism. For convenience we shall consider just two chromosomes (the sex chromosomes) in a male; but the process is exactly the same for all chromosomes in both sexes. As shown in Fig. 1.6, mitosis begins when each chromosome duplicates itself to form two chromatids still joined at their centromere. Each duplicated chromosome moves to the centre of the cell but does not, as in meiosis, synapse with its homologue. This stage, which is known as **metaphase**, is the one at which chromosomes are most readily visible. Karyotypes, therefore, consist of metaphase chromosomes. After metaphase, the centromere splits and the chromatids separate (disjoin), one going to each pole of the cell. A constriction forms in the centre of the cell and two cells are formed, each containing both an X and a Y. In this way, the two cells have exactly the same set of chromosomes as did the original cell.

In both meiosis and mitosis, chromosomes are duplicated. How does this happen?

The biochemistry of inheritance

Chemically, chromosomes consist of mostly deoxyribonucleic acid (DNA) with a small amount of a protein called histone. The latter has a binding and structural function, while the former constitutes the genetic information that is passed from 'parent' cell to 'offspring' cell during mitosis, and from one generation to the next, via meiosis.

DNA

DNA consists of two strands, each of which is a linear arrangement of **nucleotides**. All nucleotides of DNA contain an identical pentose sugar molecule (deoxyribose)

Copyrighted material

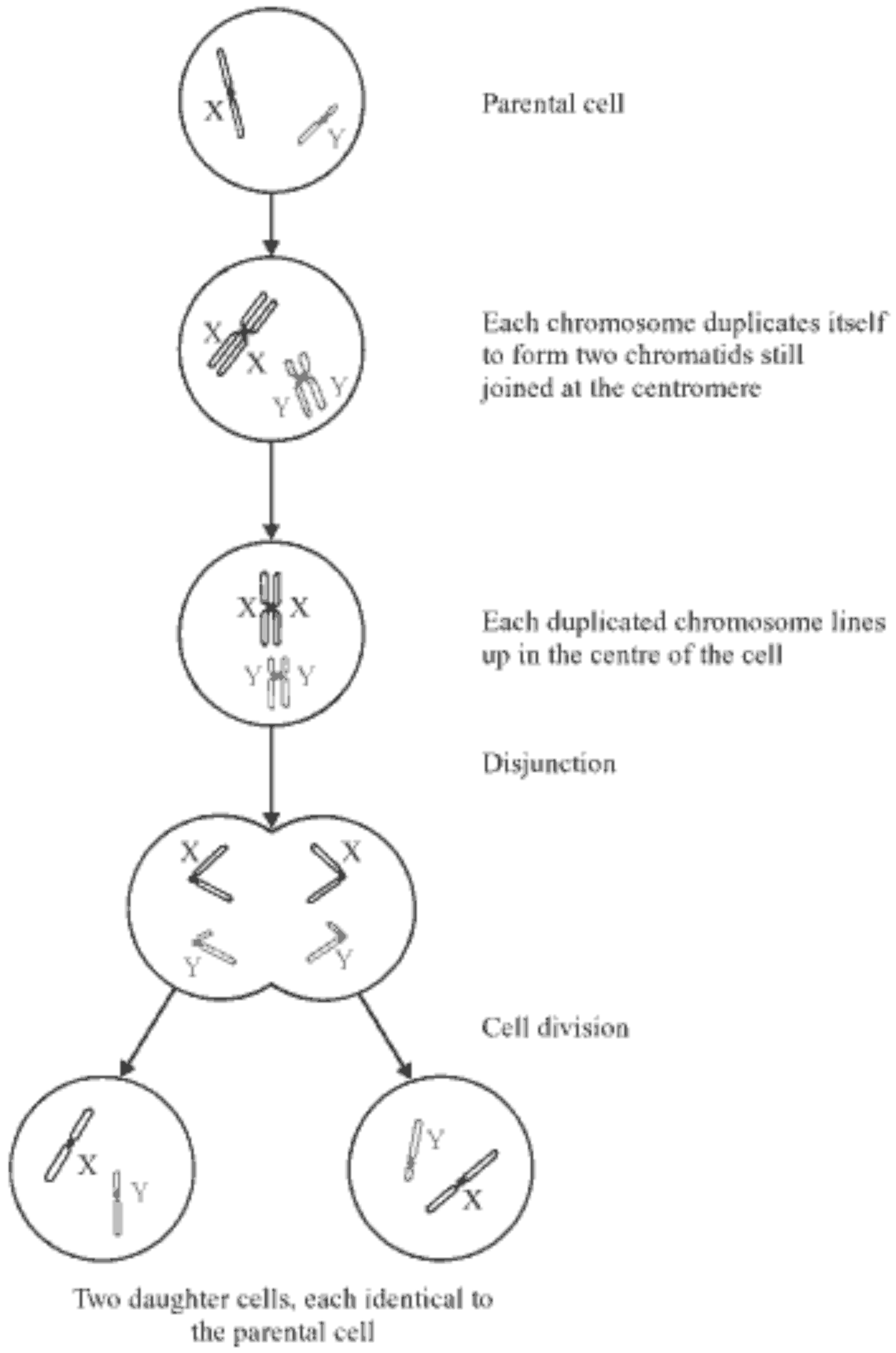


Fig. 1.6 Mitosis in a male, illustrated in terms of the sex chromosomes. The process is exactly the same for all chromosomes, and in all cells of each sex.

and an identical phosphate group. Their third component, a nitrogenous base, exists in four different forms (adenine: A; guanine: G; thymine: T; cytosine: C), giving rise to four different nucleotides, as illustrated in Fig. 1.7a. The bases A and G have a similar structure and are called **purines**; T and C have a similar structure and are called **pyrimidines**. A strand of nucleotides is held together by covalent bonding between the phosphate attached to the 5′ (pronounced 'five prime') carbon of one nucleotide and the OH attached to the 3′ carbon of the adjacent nucleotide, as shown in Fig. 1.7b. It follows that a strand of DNA has a 5′ phosphate at one end (called the **5′ end**) and a 3′ OH at the other end (called the **3′ end**).

Copyrighted material

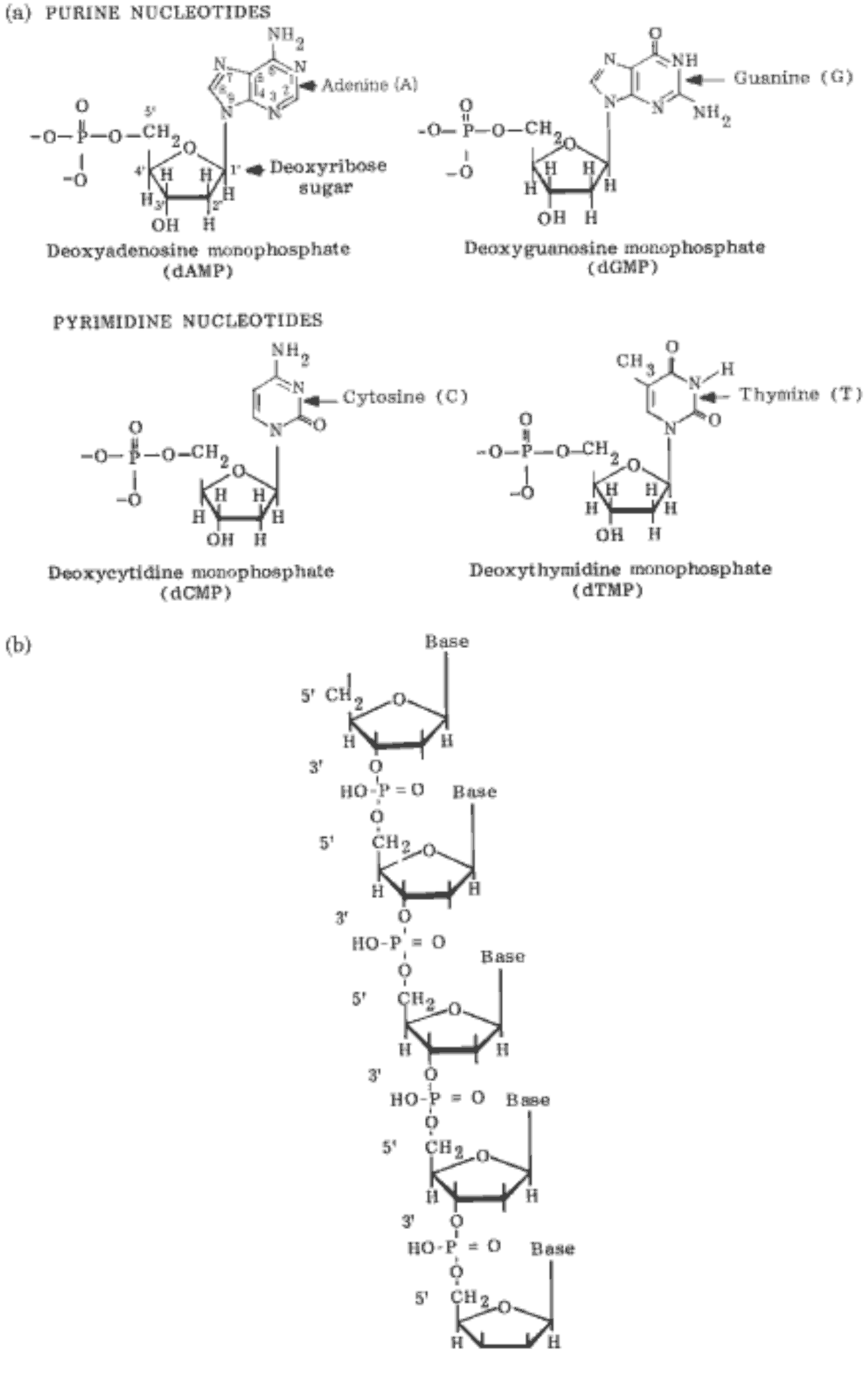


Fig. 1.7 (a) The chemical structure of the four nucleotides that are the building blocks of DNA. (b) The basic structure of a strand of DNA. ((a) Reprinted by permission from Freeman and Co.: Suzuki, Griffiths and Lewontin (1981) *An Introduction to Genetic Analysis* (2nd ed.). (b) Adapted courtesy of the estate of R.H. Symons from Symons (1981) in H. Messel (ed.) *The Biological Manipulation of Life*, Pergamon.)

Copyrighted material

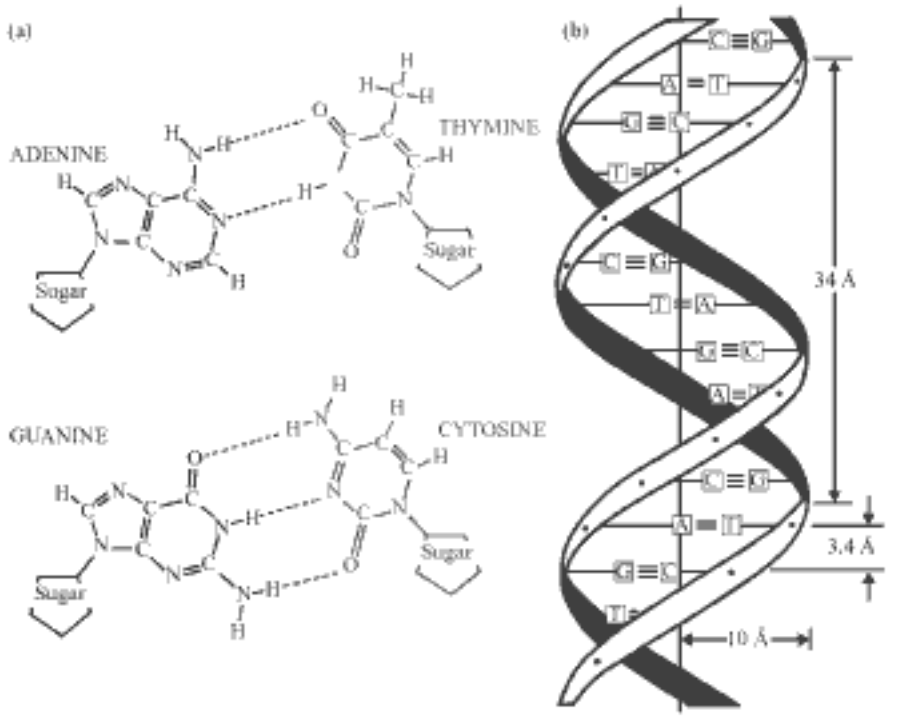


Fig. 1.8 (a) The two types of pyrimidine : purine base pairs formed by hydrogen bonding between the two strands of DNA. (b) A double helix of DNA. The two ribbons represent the sugar–phosphate 'backbones'. The structure repeats every 10 base pairs. (Adapted courtesy of the estate of R.H. Symons from Symons (1981) in H. Messel (ed.) *The Biological Manipulation of Life*, Pergamon.)

The two strands that constitute DNA are held together by very specific hydrogen bonding between purines and pyrimidines (A with T and G with C; Fig. 1.8a), giving rise to the **base pairs** A : T and G : C. Since A binds only with T and G binds only with C, one strand of DNA is complementary to the other; the sequence of bases in one strand can be predicted from the sequence in the other strand. A further consequence of the pairing arrangements is that the two strands occur together in a helix. Because two strands are involved, it is known as a **double helix** (Fig. 1.8b). The length of a short segment of DNA is usually measured in terms of the number of base pairs (bp). Longer segments are measured in terms of kilobases (1 kb = 1000 bases) or even megabases (1 Mb = 1000 kb).

The most important aspect of DNA structure is that it immediately suggests a mechanism for replication. If the double helix begins to unwind and the two strands separate, free nucleotides present in the cell are able to pair with the bases of each strand, forming a new and complementary strand for each of the original strands. As the unwinding proceeds (Fig. 1.9), two double helixes are produced from one original double helix; DNA has been replicated; a chromosome has been duplicated. The formation of each new strand by the addition of nucleotides is accomplished with the aid of the enzyme **DNA polymerase**. However, this enzyme can add nucleotides only

Copyrighted material

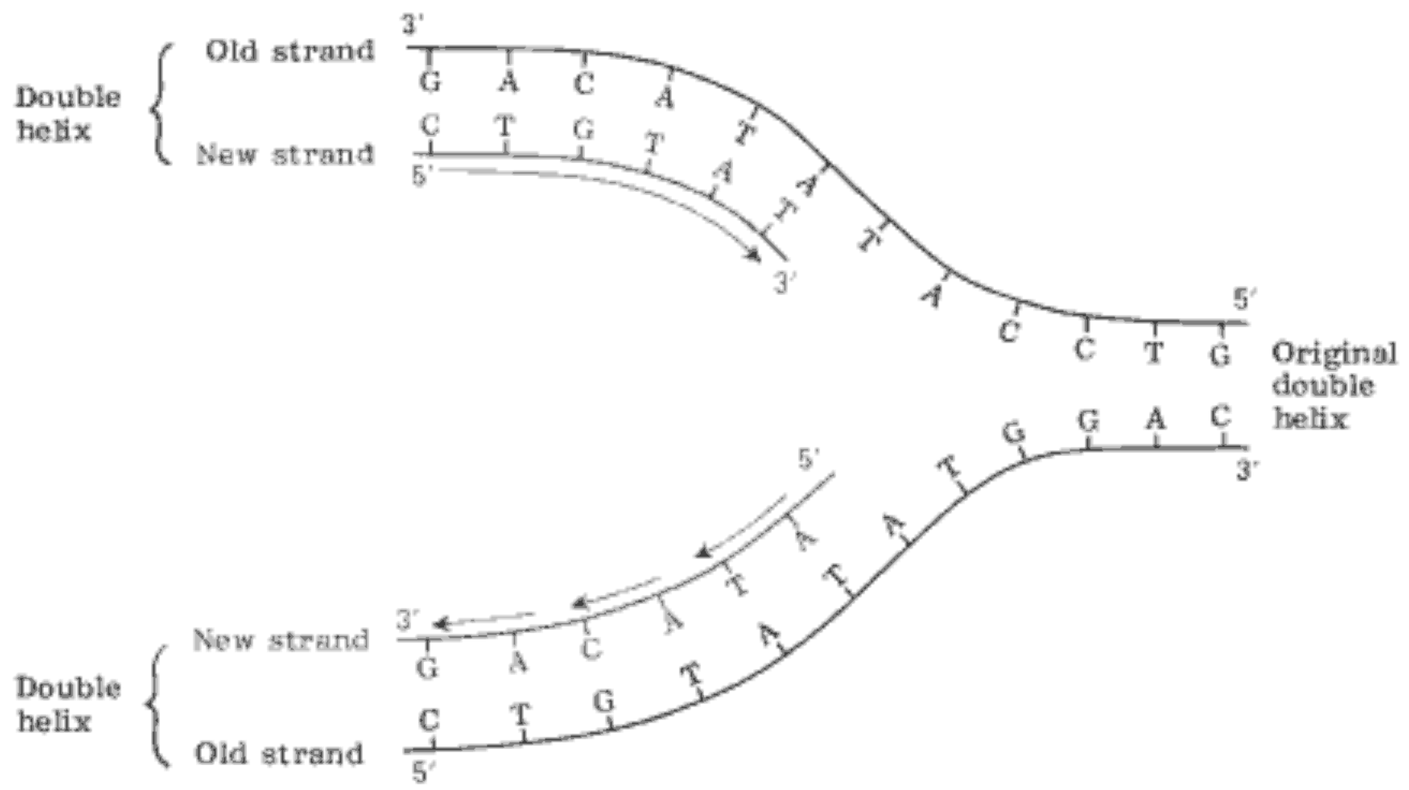


Fig. 1.9 Replication of DNA.

at the 3′ end of a growing strand, which means that replication can occur only in the 5′ to 3′ direction. Consequently, one new strand (top of Fig. 1.9) is synthesized continually, while the other new strand (bottom of Fig. 1.9) is assembled in small segments (called **Okazaki fragments**) which are each synthesized in the 5′ to 3′ direction. The Okazaki fragments are subsequently joined together by another enzyme called **DNA ligase**. The ability of these two enzymes to perform these functions has been put to good use in molecular biology, as described in Chapter 2.

It remains now to relate our knowledge of the structure of DNA to the structure of a chromosome as seen through a microscope. The total length of DNA in a mammalian cell is approximately 2 metres, which is more than 8000 times the total length of metaphase chromosomes viewed through a microscope! Obviously, therefore, a chromosome is composed of DNA that is very tightly folded or coiled. This raises the question as to how such a tight coil is unwound each time a chromosome replicates itself. The histone proteins are certainly involved in chromosome replication, but the actual mechanism has not yet been fully revealed.

The structure of DNA is the key to understanding the way in which genetic information is stored in the chromosomes, and is transmitted to the cell in such a way as to produce a particular effect. In fact, the sequence of bases in DNA has a very specific meaning recorded in the form of a code.

The genetic code

Proteins are chemical compounds with a wide range of specific roles in living organisms. Some are involved in transport (e.g. haemoglobin), support (e.g. collagen), or immunity (e.g. antibodies); some are enzymes that catalyse the innumerable biochemical reactions that occur in living cells (e.g. alcohol dehydrogenase). Some are hormones (e.g. growth hormone); some are receptors for hormones (e.g. oestrogen receptor). Some control the flow of molecules or ions in and out of cells (e.g. calcium release channel). In addition, the commercial products commonly obtained from

Copyrighted material

animals either consist almost solely of protein, e.g. meat and wool, or have protein as an important component, e.g. milk and eggs.

Proteins consist of one or more polypeptides, each of which is a chain of amino acids. There are 20 different amino acids. Each polypeptide has a specific sequence of amino acids that confers upon it a specific set of physical and chemical properties.

The information necessary for producing a specific sequence of amino acids is contained in code form within the sequence of bases in a segment of DNA. This code, which is called the **genetic code**, exists as triplets of bases (Table 1.2). With $4 \times 4 \times 4 = 64$ different possible triplets and only 20 amino acids, there is obviously some **redundancy**; in fact, the first two bases of a triplet are often sufficient to specify a particular amino acid, e.g. the triplets GTT, GTC, GTA, and GTG all specify valine. Three triplets (TAA, TAG, and TGA) do not code for any amino acid, and are known as **stop** triplets; they bring about the termination of a polypeptide chain. Another triplet (ATG) acts as a **start** signal for polypeptide synthesis. (It also codes for methionine.) The DNA between and including the start and stop triplets is called an **open reading frame** (ORF) or **coding sequence** (CDS) or **coding region**, within which the base sequence is 'read' in triplets, each of which encodes an amino acid.

Equipped with the genetic code, we can now follow the processes involved in the synthesis of proteins.

Protein synthesis

As shown in Fig. 1.10, the synthesis of polypeptides begins with the relevant segment of DNA unwinding, and the two strands separating. The sequence of DNA bases in one of the strands (called the **template** strand) acts as a template for the synthesis of a different nucleic acid (ribonucleic acid, RNA, so-called because its nucleotides contain ribose rather than deoxyribose). The synthesis is catalysed by the enzyme **RNA polymerase**, which, like DNA polymerase, adds nucleotides at the 3′ end of the growing strand, i.e. RNA is also synthesized in the 5′ to 3′ direction. Three of the bases in RNA are the same as in DNA, and the fourth, uracil (U), occurs instead of thymine. The formation of a complementary strand of RNA on the DNA template is called **transcription** (because the base sequence in DNA has been transcribed to RNA).

Before the next stage can commence, the RNA has to move from the nucleus, where the chromosomes are, to structures called **ribosomes** in the cytoplasm, where polypeptides are synthesized. (Obviously, this step is necessary only in organisms whose cells have a nucleus, i.e. eukaryotes. In prokaryotes, which have no nucleus, ribosomes become attached directly to RNA even before transcription has finished.) Because the RNA carries the code between DNA and protein, it is called **messenger RNA** or **mRNA**. Its triplets are called **codons** (shown in Table 1.2).

As also shown in Fig. 1.10, the second stage of protein synthesis involves a second type of RNA known as **transfer RNA** or **tRNA**. For each of the 20 amino acids, there is one or more specific tRNA molecules which bind to the relevant amino acid

Copyrighted material

Table 1.2 The genetic code

Triplet in DNA coding strand[1]	mRNA codon	Amino acid[2]	Triplet in DNA coding strand[1]	mRNA codon	Amino acid[2]
TTT	UUU	Phenylalanine	TAT	UAU	Tyrosine
TTC	UUC	(Phe; F)	TAC	UAC	(Tyr; Y)
TTA	UUA		TAA	UAA	STOP
TTG	UUG		TAG	UAG	
CTT	CUU	Leucine	CAT	CAU	Histidine
CTC	CUC	(Leu; L)	CAC	CAC	(His; H)
CTA	CUA		CAA	CAA	Glutamine
CTG	CUG		CAG	CAG	(Gln; Q)
ATT	AUU	Isoleucine	AAT	AAU	Asparagine
ATC	AUC	(Ile; I)	AAC	AAC	(Asn; N)
ATA	AUA		AAA	AAA	Lysine
ATG	AUG	START/ Methionine (Met; M)	AAG	AAG	(Lys; K)
GTT	GUU		GAT	GAU	Aspartic acid
GTC	GUC	Valine	GAC	GAC	(Asp; D)
GTA	GUA	(Val; V)	GAA	GAA	Glutamic acid
GTG	GUG		GAG	GAG	(Glu; E)
TCT	UCU		TGT	UGU	Cysteine
TCC	UCC	Serine	TGC	UGC	(Cys; C)
TCA	UCA	(Ser; S)	TGA	UGA	STOP
TCG	UCG		TGG	UGG	Tryptophan (Trp; W)
CCT	CCU		CGT	CGU	
CCC	CCC	Proline	CGC	CGC	Arginine
CCA	CCA	(Pro; P)	CGA	CGA	(Arg; R)
CCG	CCG		CGG	CGG	
ACT	ACU		AGT	AGU	Serine
ACC	ACC	Threonine	AGC	AGC	(Ser; S)
ACA	ACA	(Thr; T)	AGA	AGA	Arginine
ACG	ACG		AGG	AGG	(Arg; R)
GCT	GCU		GGT	GGU	
GCC	GCC	Alanine	GGC	GGC	Glycine
GCA	GCA	(Ala; A)	GGA	GGA	(Gly; G)
GCG	GCG		GGG	GGG	

[1] Following the usual convention, DNA sequences are written in the DNA-equivalent of RNA, which is the sequence in the non-template (coding) strand (see Fig. 1.10).
[2] The symbols in brackets after each amino-acid name are the standard three-letter and single-letter abbreviations.

and which have a nucleotide triplet (called an **anticodon**) that is complementary to an mRNA codon. Being complementary, a tRNA anticodon pairs with the relevant mRNA codon, bringing the correct amino acid into position in the polypeptide chain. This second and final stage of protein synthesis is called **translation**, because the

Copyrighted material

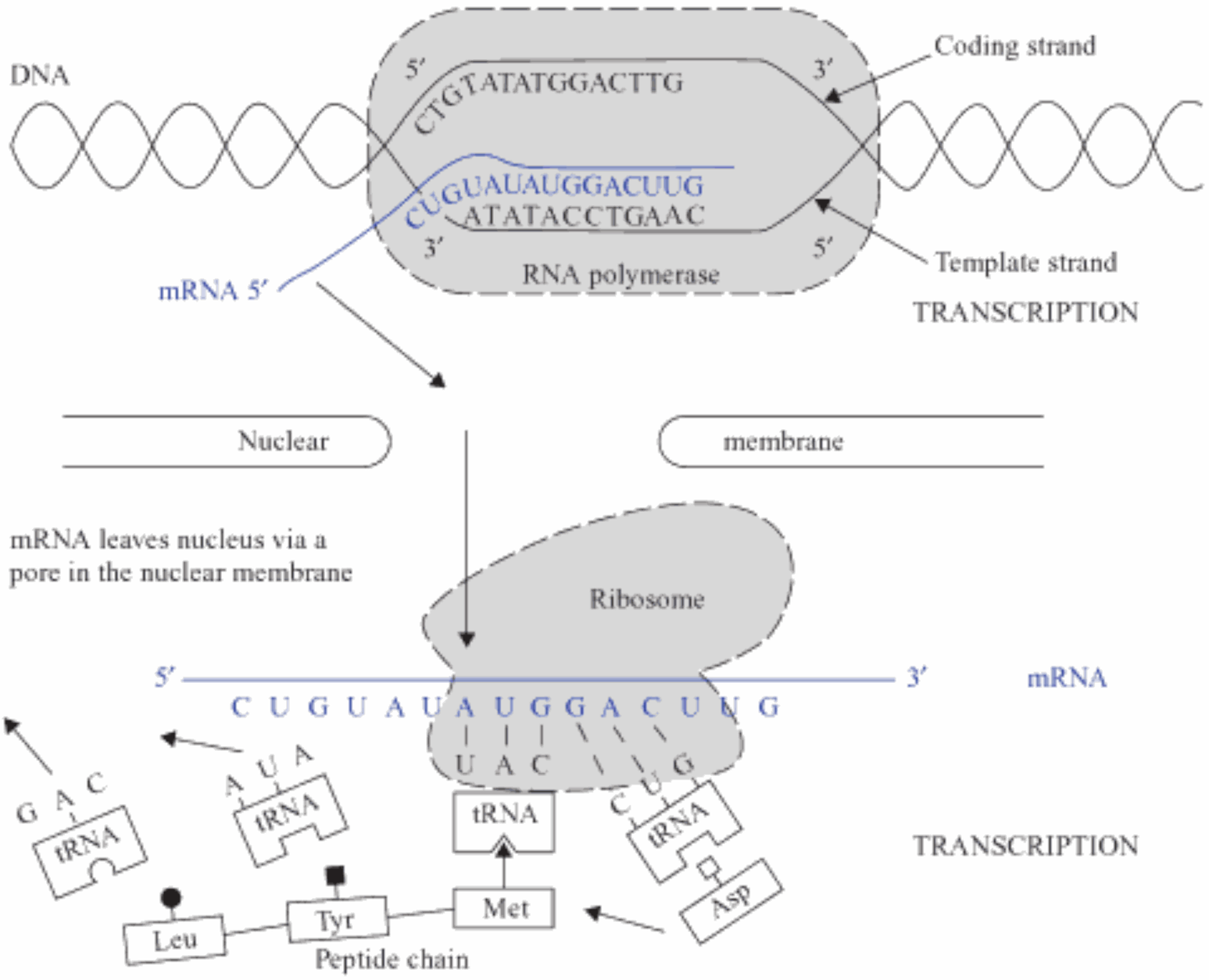


Fig. 1.10 Synthesis of polypeptide in eukaryotes, by means of transcription and translation.

base sequence has been translated (via the genetic code) to a sequence of amino acids.

What is a gene?

From the above account, it is evident that particular segments of DNA code for particular polypeptides, which are produced via mRNA. A segment of DNA including all the nucleotides that are transcribed into mRNA is called a **structural gene** (Fig. 1.11).

Since it is the mRNA sequence that is actually translated into polypeptide, and since translation starts at the 5′ end of mRNA, there is a convention that the base sequence of a gene is written as the DNA equivalent of the mRNA sequence, in the 5′ to 3′ direction. Recalling that the base sequence of mRNA is complementary to the sequence in the DNA strand from which it was transcribed, it follows that the DNA strand whose sequence is equivalent to the mRNA sequence is not the template strand, but the other strand. For this reason, the non-template strand is called the

Copyrighted material

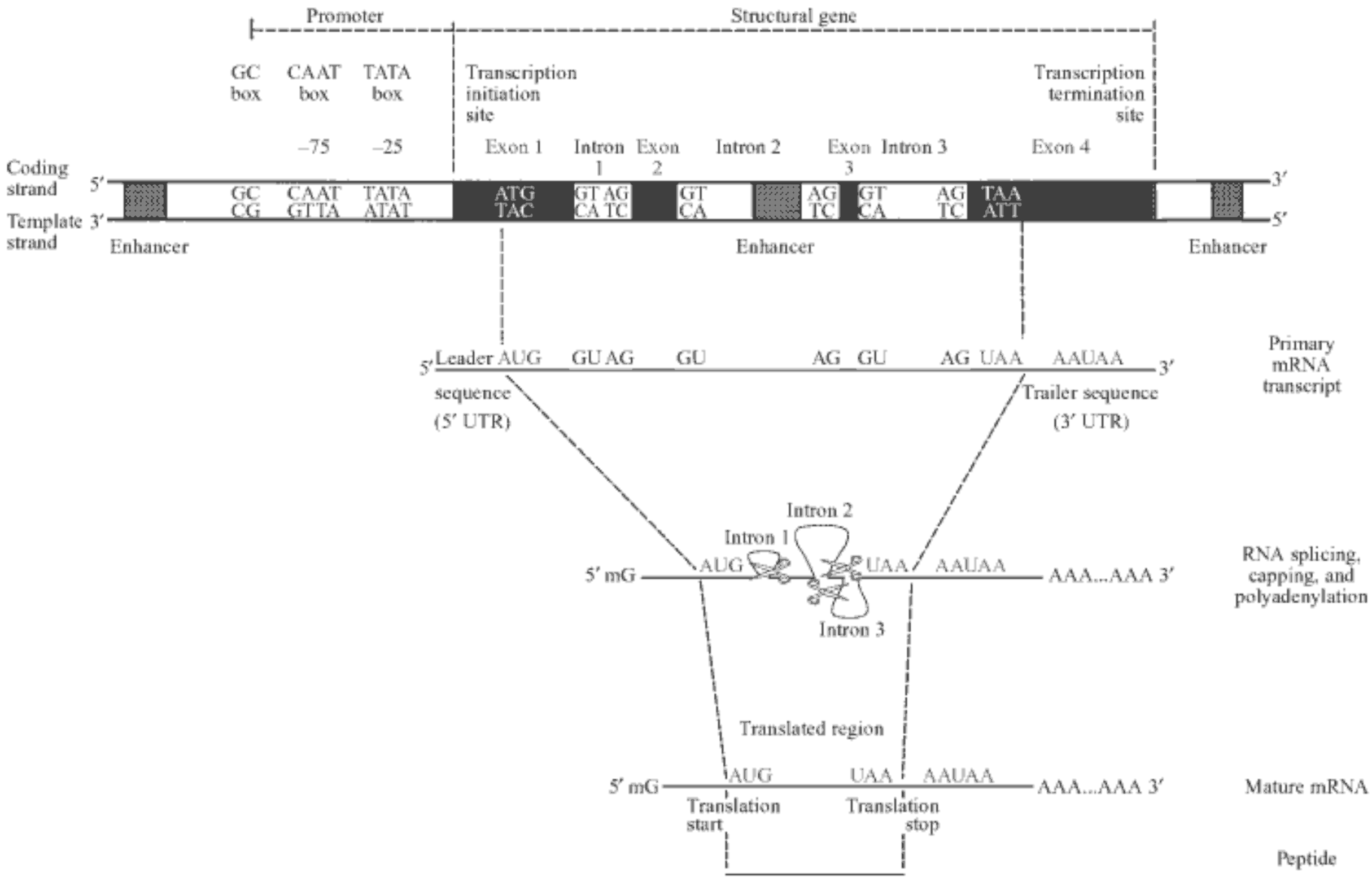


Fig. 1.11 The essential features of a gene, its messenger RNA and its peptide product.

Copyrighted material

sense or **coding** strand, while the template strand is called the **antisense** or **anticoding** strand.

Obviously a structural gene includes all of the nucleotides that are eventually translated into polypeptide, i.e. everything from the first nucleotide of the start triplet to the third nucleotide of the triplet immediately prior to the stop triplet. But it includes more than this. In fact, transcription starts before the start triplet and finishes after the stop triplet. This means that mRNA has **untranslated regions** (UTRs) at each end. The region that occurs before the start triplet (the **leader sequence** or 5′ UTR) is where ribosomes become attached. The untranslated region at the other end (the **trailer sequence** or 3′ UTR) is required for the processing of mRNA. The definition of a structural gene includes the sections of DNA that correspond to these regions. Thus, the first nucleotide of a structural gene is the nucleotide at the point where transcription actually commences, i.e. at the **transcription initiation site**. The end of a structural gene is called the **transcription termination site**. By convention, nucleotides of a structural gene are numbered from the start of the transcription initiation site, and bases preceding the site are numbered negatively, i.e. −1, −2, etc.

The region immediately preceding, i.e. **upstream** from, the transcription initiation site is particularly important, because it is the site to which RNA polymerase becomes attached prior to the initiation of transcription. This region is called the **promoter**. It contains specific sequences that are highly conserved, by which we mean that the same or very similar base sequence occurs in most genes. By studying the sequence at these conserved regions in many genes in many species, a **consensus sequence** can be deduced, consisting of the most common base at each position. The first such sequence in the promoter of eukaryotic genes is TATAAAA (called the **TATA box**), which is located approximately 25 bases upstream, i.e. at around position −25. Further upstream is the sequence GGCCAATCT (the **CAAT box**) at around position −75, and GGGCGG (the **GC box**) at around −90. These boxes are the sites for recognition and binding of regulatory proteins called **transcription factors**, which enable RNA polymerase to be positioned correctly for the initiation of transcription. In this way, they exert control over transcription. Prokaryotes have similarly conserved sequences in their promoters, namely TATAAT and TTGACA.

Termination of transcription is less well understood than initiation. It is known, however, that transcription actually extends beyond what we have called the termination site. Then the transcript is cleaved at the termination site. There is no conserved sequence corresponding to this site, but there is a highly conserved region with consensus sequence AATAAA (AAUAAA in the mRNA) located 10–30 bases before the termination site, which appears to be a recognition site in mRNA for a factor that controls the cleavage.

Split genes

Until 1977, it was thought that structural genes were just long enough to encode the mRNA that is involved in translation. In that year, it was shown that in eukaryotes, most structural genes are longer than this – they contain sections that are not represented in the mRNA by the time it undergoes translation. These sections are called

Copyrighted material

introns (because they were originally called *intr*agenic regi*ons*). Sections that are represented in the final (**mature**) version of mRNA are called **exons** (*ex*pressed regi*ons*). Genes that contain introns are said to be **split**. The original (**primary**) mRNA is a copy of the whole of a structural gene, i.e. exons and introns. Before the mRNA moves from the nucleus to the cytoplasm, the introns are removed and the exons are spliced together (in a process called **RNA splicing**), so that the mature mRNA consists solely of exons. The recognition and removal of introns is aided by the fact that in each intron, the first two bases at the 5′ end and the last two bases at the 3′ end are highly conserved – the same bases (GU and AG respectively in the mRNA) occur at these sites in most, if not all, introns. Recalling that sequences are usually expressed in terms of the coding strand of DNA, this is called the **GT–AG rule**. The base sequence immediately on either side of the highly conserved sequences, i.e. across the boundaries between introns and exons, is also conserved, but to a lesser degree. These boundaries are called **splice sites**, with the 5′ site being called the **donor site** and the 3′ site the **acceptor site**.

In the domesticated animal species for which good data exist, an average structural gene is around 43000 bases (43 kb), comprising 10 exons and nine introns; with an average mature mRNA molecule (mature transcript) of around 2000 bases (2 kb). However, there is substantial variation around these averages. Exons range in size from just a few bases to around 17000 bases (17 kb), with an average of around 200 bases. In contrast, introns range in size from just a few bases to 1000000 bases (1 Mb), with an average of around 4500 bases (4.5 kb). It is obvious from these figures that exons constitute only a small proportion of structural genes (a reasonable estimate is between 4 and 5 per cent, on average); most of the DNA in structural genes is in introns. The existence of so much 'non-functional' DNA within structural genes remains one of the great unsolved mysteries of biology. We should not be surprised, however, if the solution to the mystery is that introns really do have important functions, which are just waiting to be discovered. We already have some hints, in the form of evidence that some intronic RNA actually has a functional role in gene regulation (see next section). This is just one example of **non-protein-coding RNA** (ncRNA). In addition, introns tend to act as spacers between functional units, i.e. many exons correspond to functional units which can be assembled in different combinations to produce different polypeptides.

Apart from the excision of introns, the primary mRNA transcript is also modified in two other ways. Its 5′ (front) end is protected by the addition of a **5′ cap**, which consists of a methylated guanine nucleotide. At the other end, a **poly-A tail** is added, consisting of a variable number (typically 100–200) of adenine nucleotides. Since the cap and the tail are such important features of mature mRNA, the translation initiation and termination sites of a gene are sometimes called the **cap site** and the **poly-A site**. The AAUAAA sequence that occurs 10–30 bases upstream from the poly-A site is called the **polyadenylation signal**.

The type of gene described above is by far the most common. However, it is important to realize that there are segments of DNA whose sole function is the production, via transcription, of either tRNA or **ribosomal RNA** (**rRNA**, which is the major constituent of ribosomes). These are also called genes. Since a large

Copyrighted material

quantity of rRNA is required for the construction of sufficient ribosomes to satisfy each cell's requirements for translation, there are hundreds of genes for rRNA in the genome. They occur in several clusters of tandemly repeated rRNA genes. Each cluster produces a **nucleolus**, which is a discrete structure found within the nucleus, and which consists mainly of ribosomal RNA plus the enzymes necessary for the assembly of ribosomes. A cluster of rRNA genes is called a **nucleolar organizer region** (NOR).

In order to accommodate all the possibilities outlined above, we can redefine a gene as a stretch of DNA that produces a functional RNA molecule.

Gene regulation

It is obvious that there would be complete chaos if all genes were transcribed in all cells all at the same time. In fact, only a small proportion of genes are transcribed at any one time in any one cell. From the moment of fertilization until death, the development of each living organism is determined by genes being switched on and off at the appropriate time(s) in the appropriate cell(s). This switching is achieved by various proteins (sometimes in association with steroid hormones) becoming attached to, or being released from, specific sequences of DNA that are often highly conserved. We have already encountered three such sequences in promoters. Sequences having a similar function other than within promoters are called **enhancers**. These are located upstream, downstream, and sometimes even within structural genes (i.e. within introns). Some are located in the near vicinity of the structural gene, but others are quite some distance (20 kb or even further) away.

The regulatory proteins that exert control over transcription by binding to promoters or enhancers have in common one or more amino-acid sequences (called amino-acid **motifs**) that have been given some rather exotic names. For example, **zinc fingers** are structures which, typically, involve the repeated occurrence of a pair of cysteine molecules separated by two or three other amino acids, followed 10 or so amino acids later by a pair of histidine molecules, also separated by two or three other amino acids. In combination with a zinc atom, the two cysteines link with the two histidines, and the intervening amino acids form a finger-like loop which binds with DNA. Another example is the **leucine zipper**, which contains a periodic repeat of leucine every seventh amino acid, giving rise to a helix with the leucines aligned along one face. Two such molecules readily join together (in the manner of a zipper), creating a dimer. One end of this dimer binds to DNA.

In general, it is the binding of regulatory proteins (sometimes in conjunction with steroid hormones) to DNA that determines whether or not RNA polymerase can attach to the promoter of a gene, thereby exerting control over the first step in transcription. Regulatory proteins that increase the rate of transcription are called **activators**; those that decrease the rate of transcription are called **repressors**. In essence, these are the means by which genes are switched on and off.

Copyrighted material

For example, early embryonic development is heavily influenced by a set of **homeotic** genes (each of which determines the developmental fate of one segment of the embryo, e.g. hindbrain or spinal cord). Proteins encoded by homeotic genes have DNA-binding motifs such as zinc fingers or leucine zippers. Many of these genes also have a highly conserved 180-bp region called the **homeobox**, which encodes a DNA-binding motif called the **homeo domain** that is highly conserved throughout eukaryotes. Thus the genes that control early embryonic development encode regulatory proteins that switch genes on and off in a controlled manner.

A specific example of gene regulation later in life is provided by the way in which oestrogen controls the transcription of the ovalbumin gene in chickens. Molecules of the hormone enter a cell and bind with a protein called the oestrogen receptor. The complex of oestrogen plus receptor then binds to a region approximately 250 bases upstream from the TATA box of the gene for ovalbumin, switching it on.

Not surprisingly, groups of genes that all need to be controlled in a similar manner have a region of their promoters in common. These regions are called **response elements**. For example, all genes that need to be activated by glucocorticoids have a glucocorticoid response element, to which the glucocorticoid receptor binds, following activation by glucocorticoid. Its consensus sequence is TGGTACAAATGTTCT.

From the above discussion, it is clear that the sequences on either side of a gene are just as important as the gene itself. In fact, when the word 'gene' is used on its own, it is often taken to include the promoter and enhancer regions as well as the region in between.

Mutation

We have seen how DNA replicates, and how it gives rise to protein. Although the processes involved are remarkably elegant and usually operate faultlessly, mistakes do occur from time to time. Many mistakes have no effect at all, as they are corrected by the cell's own repair mechanisms. Uncorrected mistakes in DNA replication, however, result in an alteration in the DNA in at least one of the new cells. And because DNA replication is usually so faultless, the altered DNA is passed on unchanged to all descendent cells; until the next mistake occurs. Uncorrected mistakes in DNA replication are called **mutations**.

We shall start by considering **point mutations** (also called **gene mutations**), which involve the substitution of one nucleotide for another, or the addition or deletion of one or a few nucleotides. Other types of mutation will be considered in subsequent chapters.

There are several different possible consequences of point mutations. At one extreme, a base substitution can change a functional triplet into a stop triplet (called a **nonsense** mutation). For example, TAT codes for tyrosine; but if T in the third position is replaced with A, the resulting triplet (TAA) means stop (check this in Table 1.2). If the new stop triplet occurs before the usual stop triplet, the resultant

Copyrighted material

polypeptide is shorter than usual, and is therefore probably not functional. If a base substitution changes a triplet so as to cause an amino-acid substitution, it is called a **mis-sense** mutation. For example, substituting A for T at the third position of CAT (histidine) results in CAA (glutamine).

At the other extreme, many base substitutions have no effect on the amino acid sequence of the gene product, because the mutant triplet happens to specify the same amino acid as the original triplet. These so-called **silent** mutations are a direct consequence of the redundancy in the genetic code. For example, substituting C for T at the third position of CAT (histidine) results in CAC, which still codes for histidine (check this in Table 1.2).

Another type of point mutation involves the insertion or deletion of one or two bases, known collectively as **indels**. These are called **frameshift** mutations, because each of the triplets that occurs downstream from the site of such a mutation is shifted out of phase from the original open reading frame. At the very least, a frameshift mutation results in a completely different sequence of amino acids downstream from the site of mutation. For example, consider the following case (check it against Table 1.2):

Original coding strand TCCGAGTATCAGTCCCAG ...
Amino acid sequence Ser Glu Tyr Gln Ser Gln ...

If the second base is deleted, we have:

Mutant coding strand TCGAGTATCAGTCCCAG ...
Amino acid sequence Ser Ser Ile Ser Pro ...

The mutant amino acid sequence is obviously very different from the original sequence. In many cases, one of the 'new' triplets is a stop triplet, causing premature termination of translation. Whether or not a new stop triplet is created, it is most unlikely that the mutant polypeptide will be functional.

If a mutation occurs in cells other than those that give rise to sex cells, it is called a **somatic mutation**. The stage of development of the animal when a somatic mutation occurs determines the total number of cells that contain the altered or mutant DNA; in general, the earlier the mutation occurs, the larger the number of cells affected.

In contrast, a mutation that occurs in cells that give rise to sex cells is known as a **germ-line mutation**, which may lead to the formation of a gamete that contains the altered DNA. If this gamete is successful in fertilization, the mutation is passed on to the resultant offspring, in every cell of which it is faithfully reproduced.

Genes, alleles, and loci

The different forms of a segment of DNA that can exist at a particular site in a chromosome are called **alleles**. The particular site of a gene in a chromosome is

Copyrighted material

called the **locus** (plural **loci**). The word 'gene' is commonly used in the sense of either allele or locus. When used in this way, the appropriate meaning of the word is usually quite evident from its context.

If an offspring results from the union of a sperm with normal DNA and an ovum with an altered or mutant DNA segment in one of its chromosomes, that offspring has one normal chromosome and one mutant chromosome making up the pair of homologues. More specifically, there is one normal allele and one mutant allele at the relevant locus. In the examples below, we shall give these two alleles the symbols *B* and *b* respectively. Animals with two different alleles at a particular locus are said to be **heterozygous** at that locus. In contrast, if an animal has two copies of the same allele, the animal is **homozygous** at that locus.

Although any one animal can have a maximum of only two different alleles at a locus, the number of different alleles in a population of animals can be much greater than two. If more than two alleles exist in a population at a particular locus, that locus is said to have **multiple alleles**.

Simple or Mendelian inheritance

The passage of genes from one generation to the next is called **inheritance**. One of the major breakthroughs in science was the realization that the results of inheritance can be predicted. The first person to formulate these predictions was the Augustinian monk Gregor Mendel, who conducted his research in a monastery in Brün (now Brno in the Czech Republic) in the middle of the 19th century.

Single locus

Consider, for example, the mating of a heterozygote (*Bb*) with a homozygote (*bb*). This is exactly analogous to the situation with sex chromosomes. Consequently, the results of the mating *Bb* × *bb* can be represented in exactly the same way as that used for the inheritance of sex discussed earlier, with the aid of a checkerboard:

		Gametes from heterozygous parent	
		½ *B*	½ *b*
Gametes from homozygous parent	all *b*	½ *Bb*	½ *bb*

The result of the mating *Bb* × *bb* is expected to be an equal proportion of *Bb* and *bb* offspring. The separation of alleles at a locus during meiosis is called **segregation**, and the ratio of different types of offspring is known as the **segregation ratio**. For the mating *Bb* × *bb*, the segregation ratio is $\frac{1}{2}$ *Bb*: $\frac{1}{2}$ *bb*, which is often written as 1 *Bb*: 1 *bb*.

Copyrighted material

A checkerboard can be used to predict the outcome of any particular mating involving a single locus. The segregation ratios expected from all possible types of mating with respect to a single locus are listed in Table 1.3.

Table 1.3 Segregation ratios expected in the offspring arising from all possible types of mating in relation to a single autosomal locus, as obtained from a checkerboard

Type of mating	Segregation ratio				
	BB		*Bb*		*bb*
BB × *BB*	1	:	0	:	0
BB × *Bb*	1	:	1	:	0
BB × *bb*	0	:	1	:	0
Bb × *Bb*	1	:	2	:	1
Bb × *bb*	0	:	1	:	1
bb × *bb*	0	:	0	:	1

More than one locus

In the absence of any evidence to the contrary, it is assumed that segregation at one locus is independent of segregation at other loci. This was the assumption made by Mendel, and it is true for many situations observed in animals today. If segregation at each locus is independent of segregation at other loci, the chance of obtaining a gamete with a particular allele (say *B*) at the first locus and a particular allele (say *d*) at the second locus is simply the product of the probabilities associated with each allele independently. For example, if an individual is heterozygous at two loci (*Bb Dd*), there are four possible types of gametes, *BD*, *Bd*, *bD*, and *bd*, each of which is produced with a frequency of $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

The results of independent segregation at two loci can also be shown in a checkerboard:

		Gametes from one parent			
		¼*BD*	¼*Bd*	¼*bD*	¼*bd*
Gametes from other parent	¼*BD*	*BBDD*	*BBDd*	*BbDD*	*BbDd*
	¼*Bd*	*BBDd*	*BBdd*	*BbDd*	*Bbdd*
	¼*bD*	*BbDD*	*BbDd*	*bbDD*	*bbDd*
	¼*bd*	*BbDd*	*Bbdd*	*bbDd*	*bbdd*

Combining all cells of the checkerboard having identical offspring, and realizing that the offspring in each cell occur with a frequency of $\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$, the segregation ratio is:

1*BBDD* : 2*BBDd* : 1*BBdd* : 2*BbDD* : 4*BbDd* : 2*Bbdd* : 1*bbDD* : 2*bbDd* : 1*bbdd*

Copyrighted material

Although checkerboards quickly become rather large, in principle they can be used to derive expected segregation ratios for any type of mating involving any number of independently segregating loci.

Sex-linkage

The above patterns of inheritance illustrate simple autosomal inheritance because they describe what happens in relation to loci on autosomes. Some loci, however, are on the sex chromosomes and consequently have different patterns of inheritance. Such loci are said to be **sex-linked**. To consider the most common form of sex linkage, the inheritance patterns of **X-linked** loci can be illustrated in a checkerboard:

		Male gametes	
		$\frac{1}{2}X^H$	$\frac{1}{2}Y$
Female gametes	$\frac{1}{2}X^H$	$\frac{1}{4}X^HX^H$ (female offspring)	$\frac{1}{4}X^HY$ (male offspring)
	$\frac{1}{2}X^h$	$\frac{1}{4}X^HX^h$ (female offspring)	$\frac{1}{4}X^hY$ (male offspring)

and are summarized in Table 1.4. Very few loci have been identified on Y chromosomes (**Y-linked**).

Table 1.4 Segregation ratios expected from all possible types of matings in relation to an X-linked locus, as obtained from a checkerboard

Type of mating	Segregation ratio								
	Among females					Among males			
	X^HX^H		X^HX^h		X^hX^h	X^HY		X^hY	
$X^HX^H \times X^HY$	1	:	0	:	0	1	:	0	
$X^HX^h \times X^HY$	1	:	1	:	0	1	:	1	
$X^hX^h \times X^HY$	0	:	1	:	0	0	:	1	
$X^HX^H \times X^hY$	0	:	1	:	0	1	:	0	
$X^HX^h \times X^hY$	0	:	1	:	1	1	:	1	
$X^hX^h \times X^hY$	0	:	0	:	1	0	:	1	

At the beginning of this section it was implied that sometimes segregation at two or more loci is not completely independent. We shall now examine why this is so.

Linkage

There are probably around 22000 genes in mammals. In contrast (as we saw earlier), there are only relatively few chromosomes. Inevitably, therefore, each

Copyrighted material

chromosome consists of many different genes, each of which has a specific position (locus) on that chromosome. If chromosomes were inherited as integral units, then for all the loci on a particular chromosome, the alleles present in that chromosome would always segregate together. Consider, for example, one chromosome containing allele *B* at one locus and allele *D* at another locus, and its homologue containing alleles *b* and *d*, respectively. If chromosomes segregated as integral units, only two types of gametes would result, namely *BD* and *bd*, with equal frequency.

In practice, chromosomes are not inherited as integral units. Instead, as described earlier, recombination or crossing-over occurs when homologous chromosomes are synapsed during the first stage of meiosis. During synapsis, breakage and rejoining of chromatids occur. If the two segments of a broken chromatid rejoin, that chromatid is still inherited as an integral unit. If, however, a break occurs in the same position in two adjacent chromatids, sometimes the segments change partners, forming **recombinant** chromatids. If the two chromatids originated from just one homologue, i.e. are joined at the centromere (called **sister chromatids**), the cross-over has no effect, since sister chromatids are exact copies of each other. If, however, the two chromatids are **non-sister chromatids** (one from one homologue and one from the other), the cross-over results in the reciprocal exchange of genes between homologous chromosomes, as shown in Fig. 1.12. To the extent that breakages occur more or less randomly along the length of each chromosome, there is a direct relationship between the physical distance separating two loci on a chromosome and the chance (or probability) of a cross-over occurring between them. Unfortunately, this probability cannot be directly estimated. However, by observing the progeny of certain matings, we can calculate the **recombination fraction**, which is the proportion of gametes from one parent that can only have resulted from crossing-over during meiosis in that parent. Fig. 1.13 illustrates the concept of recombination fraction, for the two extreme cases of complete linkage and independence, and for an intermediate case.

If two loci are so close together on a chromosome that a cross-over never occurs between them (Fig. 1.13a), only two types of gametes are possible from the double-heterozygote parent, namely $\frac{1}{2}$ *EF* and $\frac{1}{2}$ *ef*. Both of these types of gametes are non-recombinants, because no recombination (crossing-over) has occurred between the two loci. Notice that non-recombinant gametes correspond to the two chromosomes of the double-heterozygote: one contains allele *E* and allele *F*, while the other contains allele *e* and allele *f*. When this double-heterozygote parent is mated with a double-homozygote parent, the checkerboard in Fig. 1.13a shows that two types of offspring (both non-recombinants) result with equal frequency. We can conclude that when loci are so close that recombination never occurs between them, the result is 100 per cent non-recombinant offspring and 0 per cent recombinant offspring.

If the two loci are some distance apart on the same chromosome (Fig. 1.13b), then in some germ cells undergoing meiosis, a cross-over will occur between the two loci. As we saw in Fig. 1.12, whenever a cross-over occurs between two loci on the same chromosome, the result is two types of recombinant gametes and two types of non-recombinant gametes. These are shown in Fig. 1.13b. For those meioses in

Copyrighted material

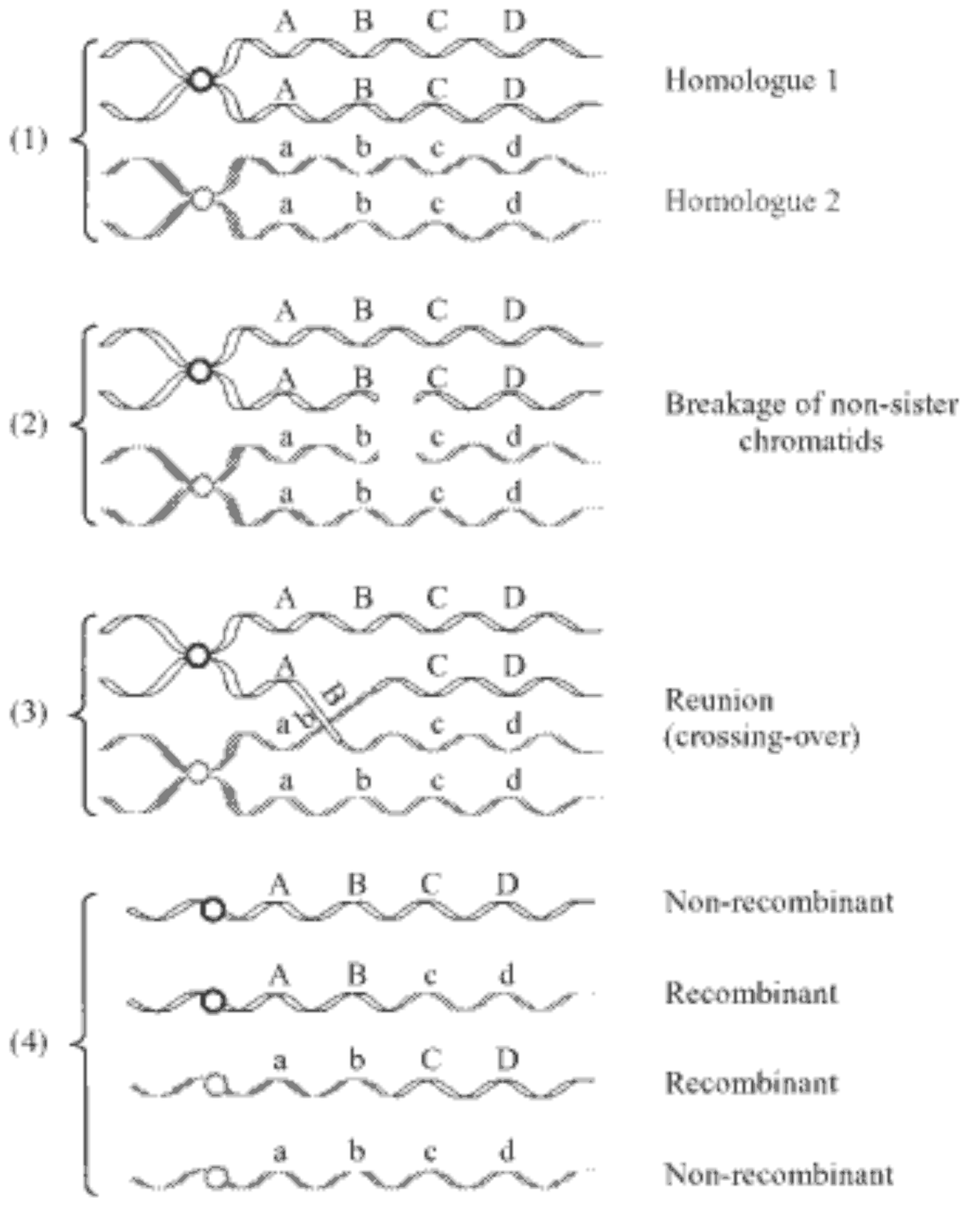
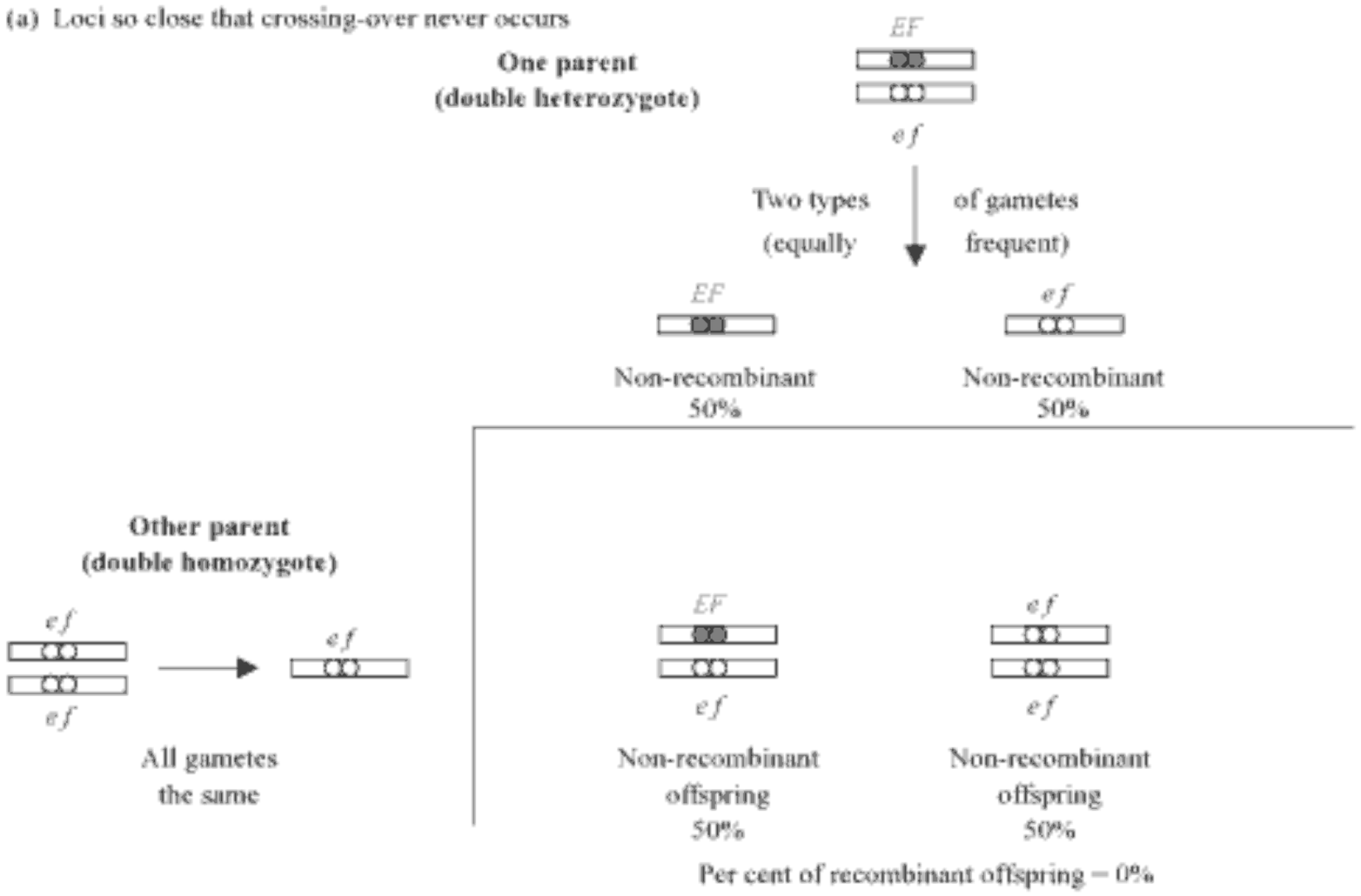


Fig. 1.12 The four stages involved in crossing-over between a pair of homologous chromosomes.

which a cross-over does not occur, only two types of gametes are produced, namely the two types of non-recombinant gametes. If the distance between the two loci is such that the chance of a cross-over between the two loci is, for example, 20 per cent, then 20 per cent of meioses in this parent will result in two recombinant gametes and two non-recombinant gametes, and the remainder (80 per cent) of meioses will result in four non-recombinant gametes. In Fig. 1.13b, we consider the results of 100 meioses (producing 400 gametes), with a 20 per cent chance of crossing-over between the two loci. Among the 400 gametes produced from this parent, the total number of non-recombinant gametes is 160 + 160 = 320 (from the meioses in which recombination did not occur) plus 20 + 20 = 40 from the meioses in which recombination did occur, making a total of 180 + 180 = 360 non-recombinant gametes. The other 40 gametes are recombinant gametes. The resultant offspring are shown in Fig. 1.13b: two types of recombinant offspring and two types of non-recombinant offspring. We can now estimate the proportion of offspring that can only have resulted from recombination between the two loci. This is the recombination fraction. In this particular example it is 40/400, i.e. 10 per cent. If the two loci had been closer together on the chromosome, the recombination fraction would

Copyrighted material



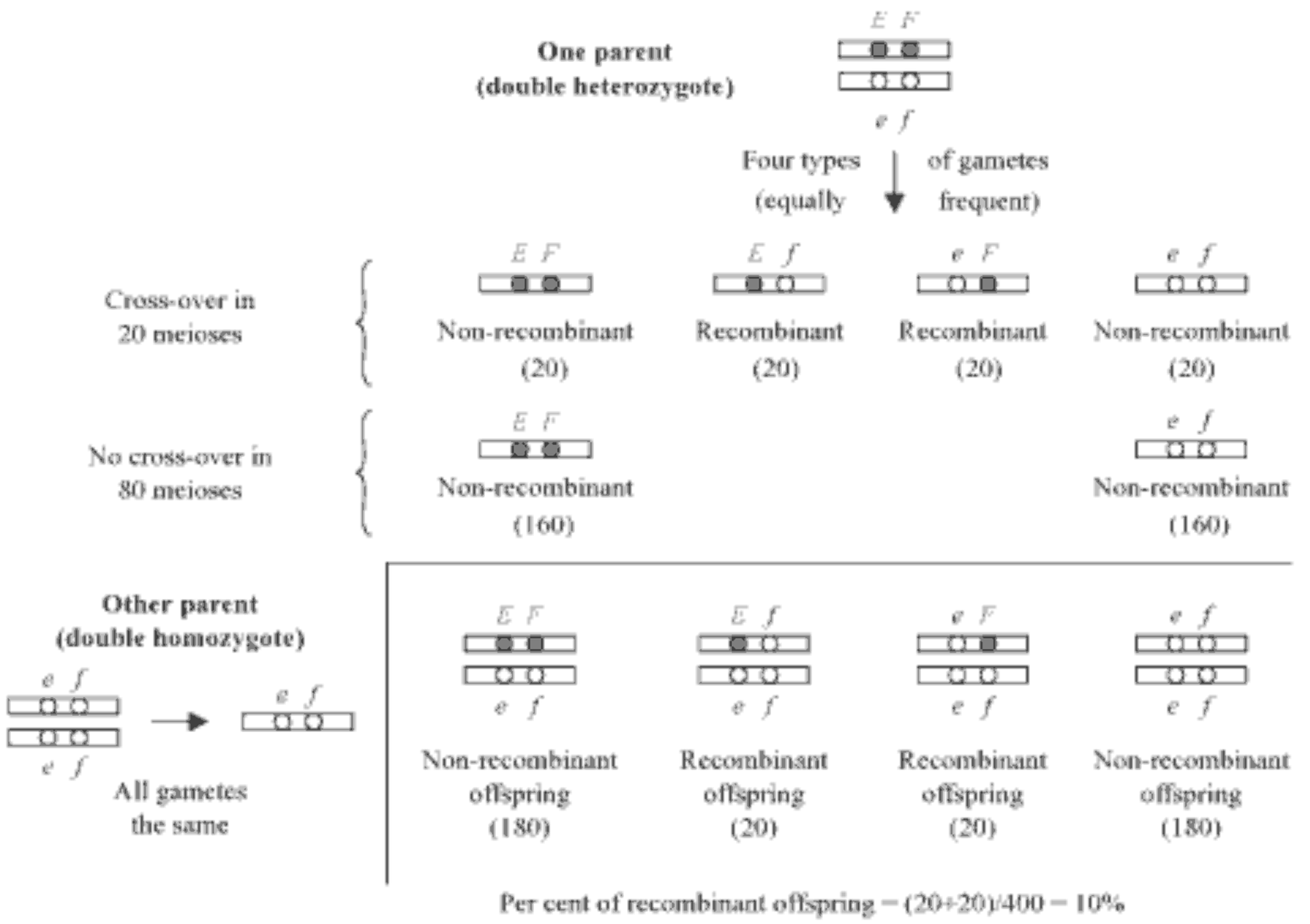


Fig. 1.13 The concept of recombination fraction, illustrated with a checkerboard. (a) Loci so close that crossing-over never occurs. (b) Loci somewhat apart; crossing-over occurs in, for example, 20 per cent of meioses. (c; next page) Loci so far apart that a cross-over occurs in every meiosis.

Copyrighted material

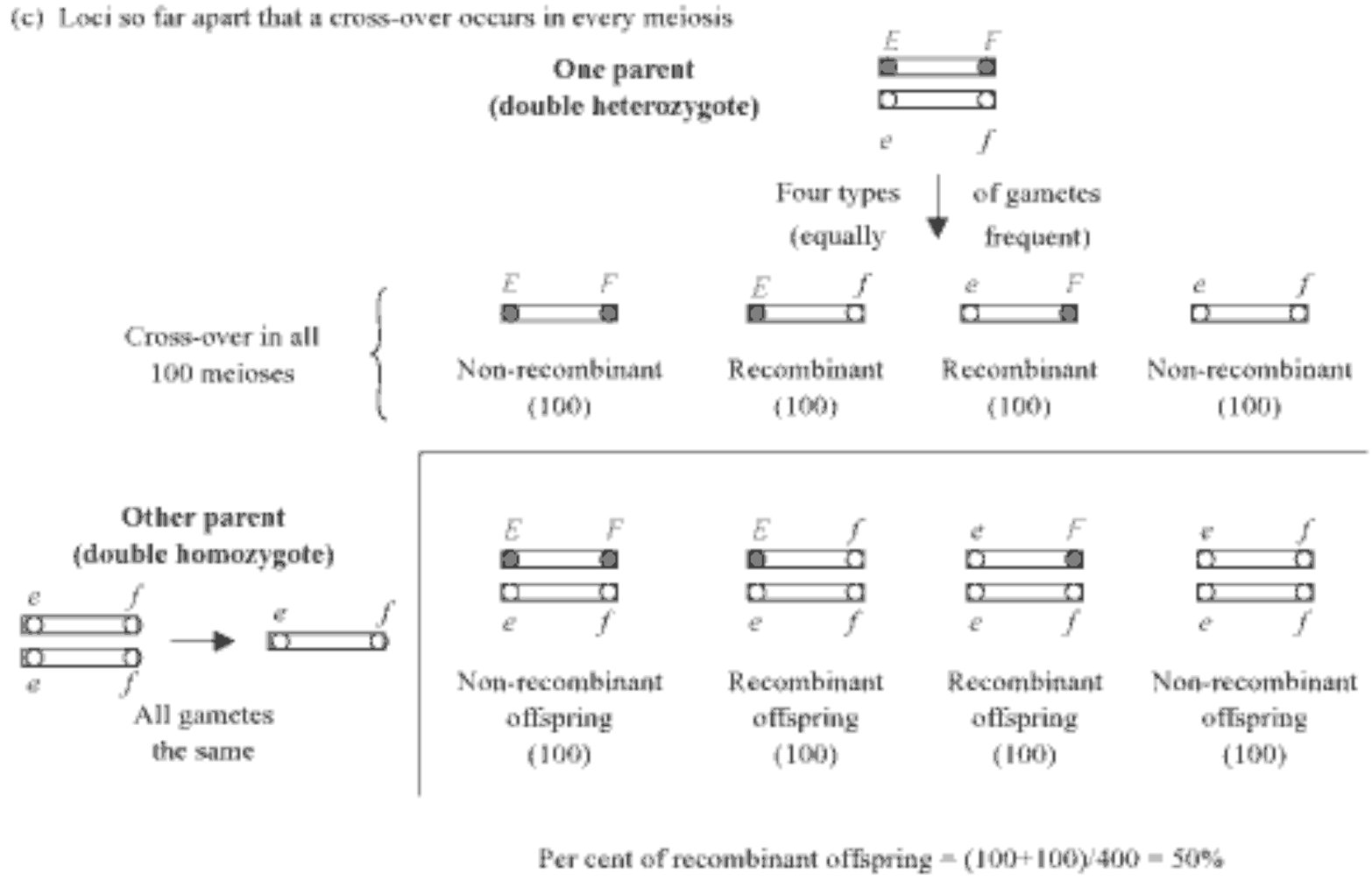


Fig. 1.13 *Continued*

have been less than 10 per cent; if the two loci were further apart on the same chromosome, the recombination fraction would have been greater than 10 per cent. From this, it can be seen that the recombination fraction is an indirect indicator of the distance between the two loci. This distance is called the **map distance**, and is measured in units of centimorgans (cM), named after Thomas Morgan, who discovered the phenomenon of linkage (for which he was awarded the Nobel prize – the first geneticist so honoured). For low recombination fractions (say, less than 10 per cent), 1 map unit (1 cM) equals 1 per cent recombination fraction. For distances greater than this, there is an appreciable chance that more than one cross-over will occur between the loci. Consequently, as the distance between a pair of loci increases, recombination fraction increasingly underestimates map distance. To cater for this, a mapping function such as that shown in Fig. 1.14 is used for estimating map distance from recombination fraction.

The final situation to consider is that shown in Fig. 1.13c, in which the two loci are so far apart that a cross-over always occurs between them. Since every meiosis in this situation involves a recombination between the two loci, each meiosis produces equal numbers of recombinant and non-recombinant gametes: all four types of gametes occur with an equal frequency (of 25 per cent), and, as shown in Fig. 1.13c, the recombination fraction is 50 per cent. You may recognize that this result is exactly the same as that seen for the segregation of two loci that are located on different chromosomes, as shown in the two-locus checkerboard on page 26. In other words, loci that are far apart on the same chromosome segregate independently, just as if they were located on different chromosomes. It should also now be clear that

Copyrighted material

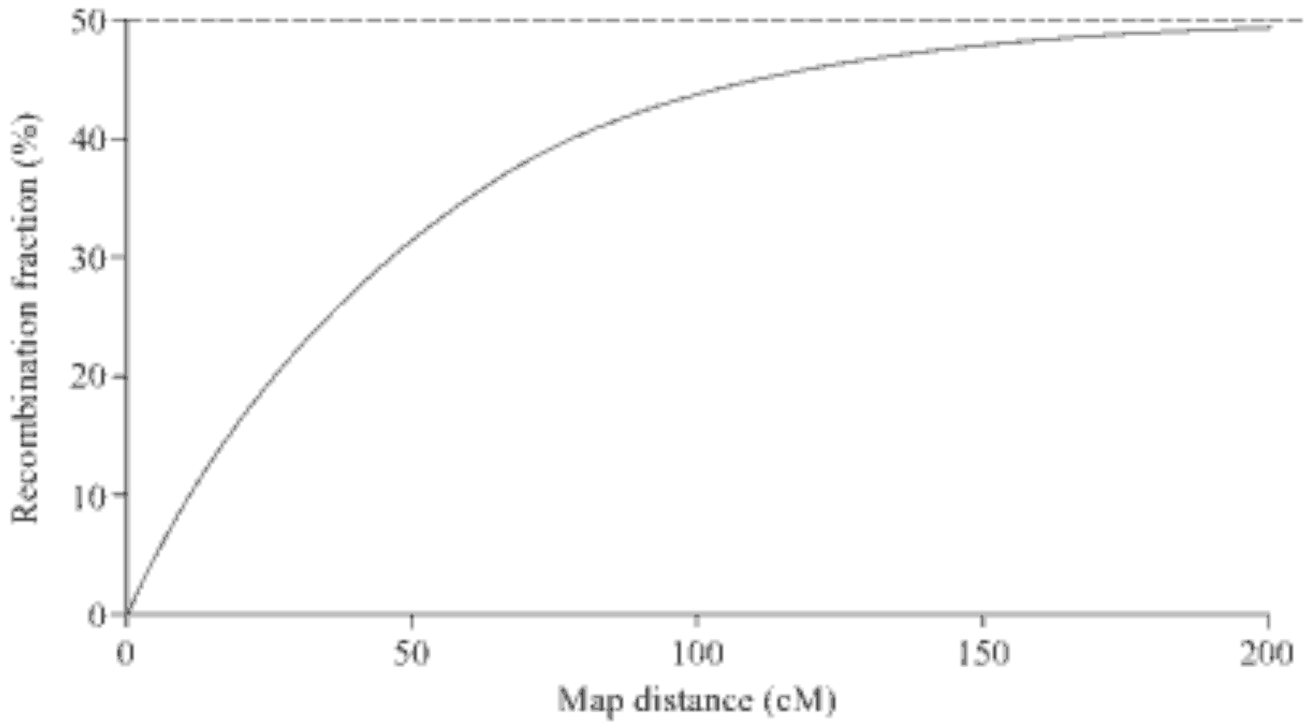


Fig. 1.14 A mapping function, showing the relationship between recombination fraction and map distance.

the maximum recombination fraction is 50 per cent, as shown in the mapping function in Fig. 1.14.

The relationship between the recombination fraction and the distance between loci enables the construction of linkage maps, in which loci are positioned according to the map distance between them. The construction of linkage maps, which have very important practical applications, is discussed in Chapter 2.

Inactivation

X-inactivation and dosage compensation

Among the many coat colours seen in cats, the mosaic of orange and non-orange, which is known as tortoiseshell (Fig. 1.15a) is one of the most attractive. The orange hairs result from the action of an X-linked allele *O*, which prevents the production of dark pigment (black and brown), but enables the production of yellow pigment. The non-orange hairs are due to the normal (**wild-type**) allele at the same locus, *o*, which enables the production of dark pigment, in whatever manner is determined by alleles at other coat-colour loci. (See Chapter 11 for more information on the genetics of coat colour.) Since obviously both alleles must be present to produce the mosaic of orange and non-orange, tortoiseshell cats must be heterozygous, $X^{O}X^{o}$, at this X-linked locus. But why do some parts of the body express the effect of the orange allele, while other parts express that of the non-orange allele? And why are the patches of non-orange and orange approximately equal in total area, and why are they scattered more or less randomly throughout the coat?

The answers to these questions lie partly in another observation, first made in cats, by Barr and Bertram, who in 1949 reported that the nucleus of non-dividing nerve cells in females usually contains a small dark-staining body, whereas that in males

Copyrighted material

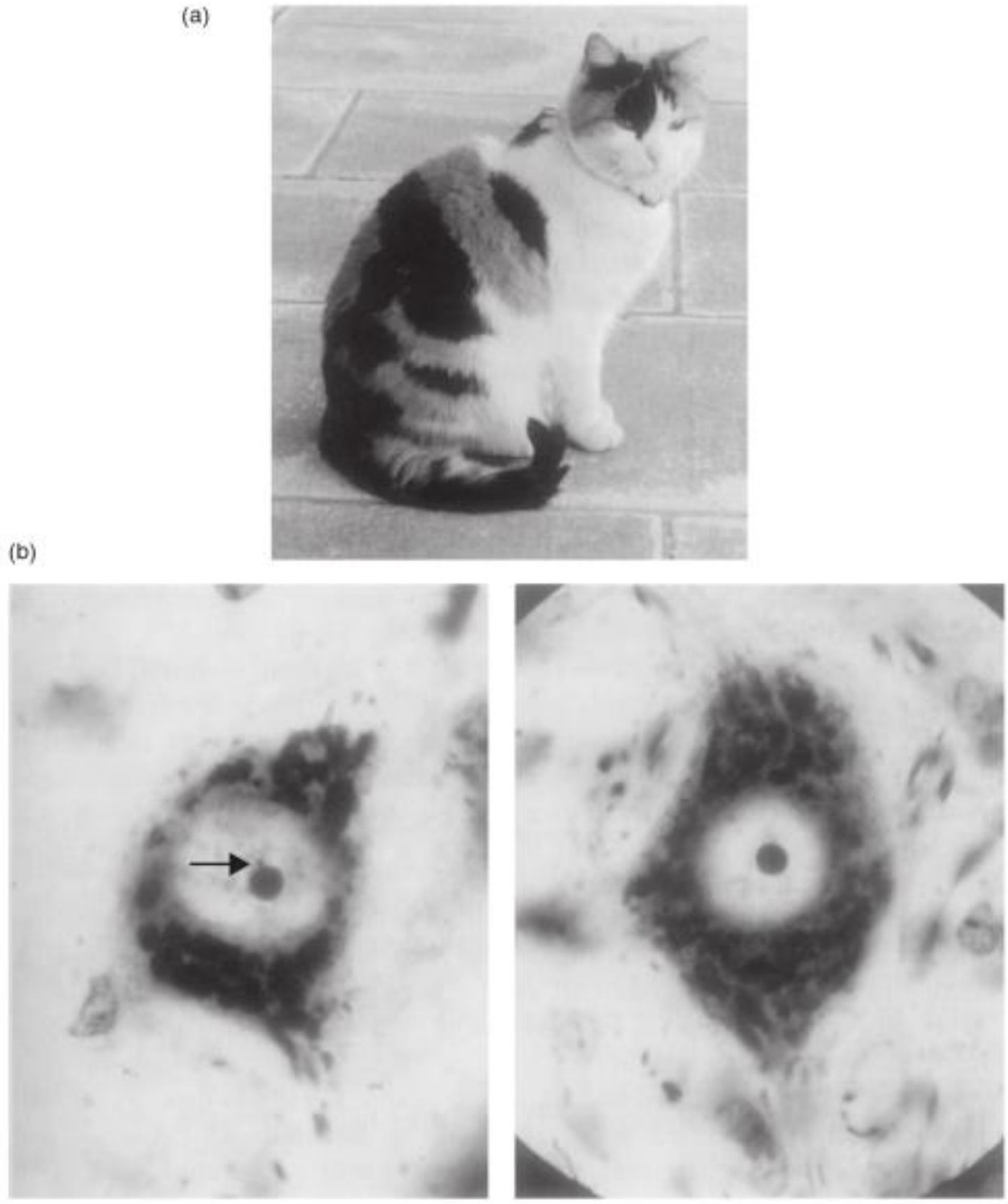


Fig. 1.15 (a) A tortoiseshell cat with white spotting. The white spotting is due to an allele at an autosomal locus (see Chapter 11).
(b) Nucleus of a hypoglossal nerve cell from a mature female cat (left) and a mature male cat (right). The large dark-staining body in each nucleus is a nucleolus. The small dark-staining body (arrowed) in the female cell is a Barr body. ((b) Reprinted by permission from Macmillan Publishers Ltd: *Nature*, (1949) **163**, pp 676–7.)

does not (Fig. 1.15b). This dark-staining body is now called a **Barr body** or **sex chromatin**. Although it had been observed by many previous researchers, Barr and Bertram were the first to note that the dark-staining body occurs in only female cells. In an attempt to explain their observations, they speculated that it might be an X chromosome which had become very highly condensed. Other researchers showed

Copyrighted material

that they were correct; a Barr body is, in fact, an X chromosome that is late in replicating during mitosis.

Drawing on similar observations in mice, Mary Lyon suggested in 1961 that the highly condensed X chromosome seen in female cells is the result of one of the X chromosomes (chosen at random) becoming inactive in each cell of all female embryos at an early stage of development. This is the **Lyon hypothesis**. (In fact, it is now known that not all genes on the inactivated X chromosome are inactivated; those in and near the pseudo-autosomal region remain functional in both X chromosomes.)

Since the Lyon hypothesis postulates that the choice of X for inactivation is entirely random, it follows that each of the X chromosomes in normal females is active in approximately one-half of that female's cells.

The process of **random X-inactivation** provides an adequate explanation for tortoiseshell coat colour in cats; each patch of orange represents the cells that descended from a cell in which the non-orange allele was inactivated, and *vice versa*. In addition, the apparently random distribution of patches and the approximately equal total area of orange and non-orange are to be expected if the inactivated X is chosen at random.

In passing, it should be noted that since tortoiseshell cats are heterozygous at an X-linked locus, they must have two X chromosomes, in which case they should be female. Normal male cats, having only one X chromosome, can be either orange (X^OY) or non-orange (X^oY), but not tortoiseshell. It is therefore a fairly safe bet that any tortoiseshell cat is a female. Occasionally male tortoiseshells are reported, but they often turn out to be abnormal males having an extra X chromosome, as described in Chapter 4.

The result of random X-inactivation is that each female is a **mosaic**, consisting of two distinct populations of cells derived from a common source; in one population of cells the maternal X chromosome is inactive, and in the other cell population, the paternal X is inactive.

The only well-documented exception to random X-inactivation occurs in marsupials, where it is the paternal X chromosome that is inactivated, and often only incompletely so. The reason for this is not known.

Obviously the result of X-inactivation in females is that each female cell has the same quantity of gene product from X-linked genes as do males. Thus, X-inactivation is a mechanism that compensates for the difference in gene 'dosage' between males and females in relation to X-linked genes. This effect of X-inactivation is called **dosage compensation**.

Finally, an important difference between mammals and birds must be noted: while X-inactivation appears to occur in all mammals, Z-inactivation does not occur in birds. The reasons for this are not known.

Imprinting

Inactivation is not confined to the X chromosome. At certain loci on other chromosomes, the extent to which an allele is expressed (or even whether it is expressed at

Copyrighted material

all) depends on the parent from which it came. This differential expression of alleles is called **genomic imprinting**. As might be imagined, this can be a source of frustration in attempts to determine the mode of inheritance of disorders, because imprinting can result in atypical inheritance patterns.

Inactivation results from methylation

At the molecular level, inactivation is associated with the addition of a methyl group (CH_3) to cytosine molecules that occur immediately on the 5′ side of guanine molecules, i.e. inactivation is associated with methylation of cytosines in so-called **CpG islands**, where p stands for the phosphate link between the two adjacent bases. Within an individual animal, all descendants of each cell in which inactivation first occurred have the same inactive gene or chromosome, because after each replication of a methylated strand of DNA, the new strand is automatically methylated at the same CpG sites as in the original strand. In meiosis, however, or in early embryonic development, the methylation patterns are reset.

Not all methylation patterns are set for the lifespan of the animal. In fact, in regions that are not subjected to X-inactivation or imprinting, methylation of CpG islands in promoters is a common attribute of inactive genes, and demethylation is a prerequisite for transcription of many genes. Thus, methylation is another means by which genes are regulated.

Inactivation and imprinting are just two examples of **epigenetics** – the inheritance between cell generations and sometimes between animal generations of changes in 'phenotype' that are not the result of changes in DNA sequence. As we shall see in later chapters, there are some fascinating biological stories emerging from epigenetics research. However, as the name implies, these are additions to, rather than (as some people tend to think) replacements of, the principles of Mendelian inheritance.

Types of DNA

How can we categorize total DNA, and where do structural genes fit into the picture?

The most common category of DNA consists of **unique** or **single-copy** sequences, which account for 60–70 per cent of the genome of mammals. These single-copy sequences are dispersed throughout the genome. A small proportion of this DNA accounts for most structural genes. Evidence emerging from genome sequence assemblies (see Chapter 2) suggests that the total number of structural genes in animals is around 22 000.

Some structural genes occur in **multigene families**, which consist of sets of identical or very similar genes, whose individual members are usually scattered around the genome, or, in some cases, occur as sets of adjacent genes. Not surprisingly, the structural genes that occur in multigene families are those whose products are required in relatively large quantities, e.g. histone, keratin, collagen, ribosomal RNA, and transfer RNA.

Copyrighted material

The third major category of DNA is **repetitive** DNA, which consists of multiple copies of particular sequences called **repeat units**, which range in size from a single base to several thousand bases. Repetitive DNA has assumed increasing importance in recent years, with the realization that it holds the key to some important inherited diseases, and provides a major tool for the practical application of molecular biology to measurement of genetic diversity within and among animal populations and, to a lesser extent in recent years, animal health and animal improvement. We shall discuss these aspects of repetitive DNA in following chapters.

There is one other category of chromosomal DNA that should be mentioned. Scattered throughout the genome are small DNA fragments called **transposable genetic elements** (**TGEs**) or **jumping genes**. A notable property of a TGE is that the nucleotide sequence at one end is an **inverted repeat** (or occasionally a **direct repeat**) of the sequence at the other end. In cattle, for example, there is a TGE which is 611 bases long. Its terminal sequences are:

5′ GCCGGGGA … TCCCCGGC 3′
3′ CGGCCCCT … AGGGGCCG 5′

Notice that the sequence TCCCCGGC at the 3′ end of the top strand is really a repeat of the GCCGGGGA sequence at the other end of the same strand, only in an inverted (i.e. reverse) and complementary form. Another way to look at this is to note that reading the sequence in the top strand from 5′ to 3′ is exactly the same as reading the bottom sequence from 5′ to 3′. Because they contain exactly the same message when read in either direction, inverted repeats are said to be **palindromes** (by analogy with palindromic sentences such as ABLE WAS I ERE I SAW ELBA).

The repeats are homologous, and are therefore able to pair with each other, just as homologous chromatids pair during meiosis I. In the case of TGEs, the pairing of the repeats results in the TGE itself forming a loop. Often when this occurs, the whole of the TGE excises itself from wherever it happens to be, and moves to another section of the same or a different chromosome, into which it then inserts itself by the reverse process. Sometimes, the TGE is replicated, and the resultant copy moves elsewhere, leaving the original one in its original position. As we shall see in Chapter 10, TGEs are extremely important in relation to the rapid spread of multiple antibiotic resistance in bacteria. In eukaryotes, TGEs are surprisingly ubiquitous: there are hundreds of thousands of them in the mammalian genome, making up from 1 to 5 per cent of total DNA. In cattle, for example, the TGE described above occurs 35 000 times. Many of the copies of this and other TGEs have lost the ability to move from one site to another (to transpose). However, those that are able to transpose themselves are extremely important, because if a TGE inserts itself into a structural gene, it will most likely inactivate that gene. Alternatively, if a TGE inserts itself in the control region of a gene, it may interfere with normal control, resulting in the gene being expressed at inappropriate times and places, or not being expressed when it should be. Because of these effects, TGEs are important sources of mutation. The mutations they create are called **insertion mutations**. TGEs are also important causes of cancer, as we shall see in Chapter 11.

Copyrighted material

Summary

- A chromosome is, in essence, a double helix of DNA
- Each species has a characteristic set of chromosomes (karyotype)
- Typically, an animal karyotype comprises two subsets of chromosomes (diploid): one inherited from the mother and one from the father, with each maternal chromosome having a matching paternal chromosome (pairs of homologues), except for the sex chromosomes
- The first stage of meiosis and mitosis involves the duplication of each chromosome (i.e. the replication of a double helix), resulting in two chromatids (two double helixes) joined at the centromere
- In mitosis, this is the stage at which karyotypes become visible
- Meiosis results in gametes having one set of chromosomes (haploid), comprising one member (chosen at random – maternal or paternal) of each homologous pair
- Recombination (crossing-over) during meiosis results in some chromosomes having a portion of maternal and paternal double helix
- The sequence of bases in some segments of DNA acts as a template for the production of a specific sequence of amino acids which form a particular polypeptide
- This process involves transcription of the DNA sequence into mRNA, followed by translation of the mRNA into peptide
- A segment of DNA including all the nucleotides that are transcribed into mRNA is called a structural gene; more generally, a gene is a stretch of DNA that produces a functional RNA molecule
- From the moment of fertilization until death, the development of each living organism is determined by genes being switched on and off at the appropriate time(s) in the appropriate cell(s), by means of regulatory proteins (activators and repressors)
- Uncorrected mistakes in DNA replication are called mutations
- Each mutation creates an altered DNA sequence (allele) at a particular location (locus) on a chromosome
- An inevitable consequence of meiosis is that the alleles at any locus segregate into gametes in a predictable manner (Mendelian inheritance)
- Alleles at different loci typically segregate independently
- If two loci are close together on the same chromosome, their alleles do not segregate independently – they are said to be linked
- In mammals, one of the X chromosomes (chosen at random soon after fertilization) in females is inactivated by methylation of cytosine molecules in its DNA, resulting in dosage compensation for genes on the X chromosome (in relation to males, which have only one X chromosome)
- Some sections of autosomes are also inactivated by methylation (imprinting)
- Inactivation and imprinting are just two examples of epigenetics – the inheritance between cell generations and sometimes between animal generations of changes in 'phenotype' that are not the result of changes in DNA sequence.
- Only a few per cent of DNA accounts for all structural genes; the remainder includes vast quantities of repetitive DNA and transposable genetic elements (jumping genes)

Copyrighted material

Further reading

Ellegren, H., Hultin-Rosenberg, L., Brunstrom, B., Dencker, L., Kultima, K. and Scholz, B. (2007) Faced with inequality: chickens do not have a general dosage compensation of sex-linked genes. *BMC Biology*, **5**, 40.

Feschotte, C. and Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, **41**, 331–68.

Hartl, D.L. and Jones, E.W. (2009) *Genetics: analysis of genes and genomes*, 7th edn. Jones and Bartlett, Boston.

King, R.C., Stansfield, W.D. and Mulligan, P.K.(2006) *A dictionary of genetics*, 7th edn. Oxford University Press, New York.

Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*, **10**, 155–9.

O'Brien, S.J., Menninger, J.C. and Nash, W.G. (eds) (2006) *Atlas of mammalian chromosomes*. Wiley, Hoboken, NJ.

Payer, B. and Lee, J.T. (2008) X chromosome dosage compensation: how mammals keep the balance. *Annual Review of Genetics*, **42**, 733–72.

Presole, G. (2008) What is a gene? An updated operational definition *Gene*, **417**, 1–4.

Robinson, T.R. (2005) *Genetics for dummies*. Wiley, Hoboken, NJ.

Snustad, D.P. (2009) *Principles of genetics*, 5th edn. Wiley, Hoboken, NJ.

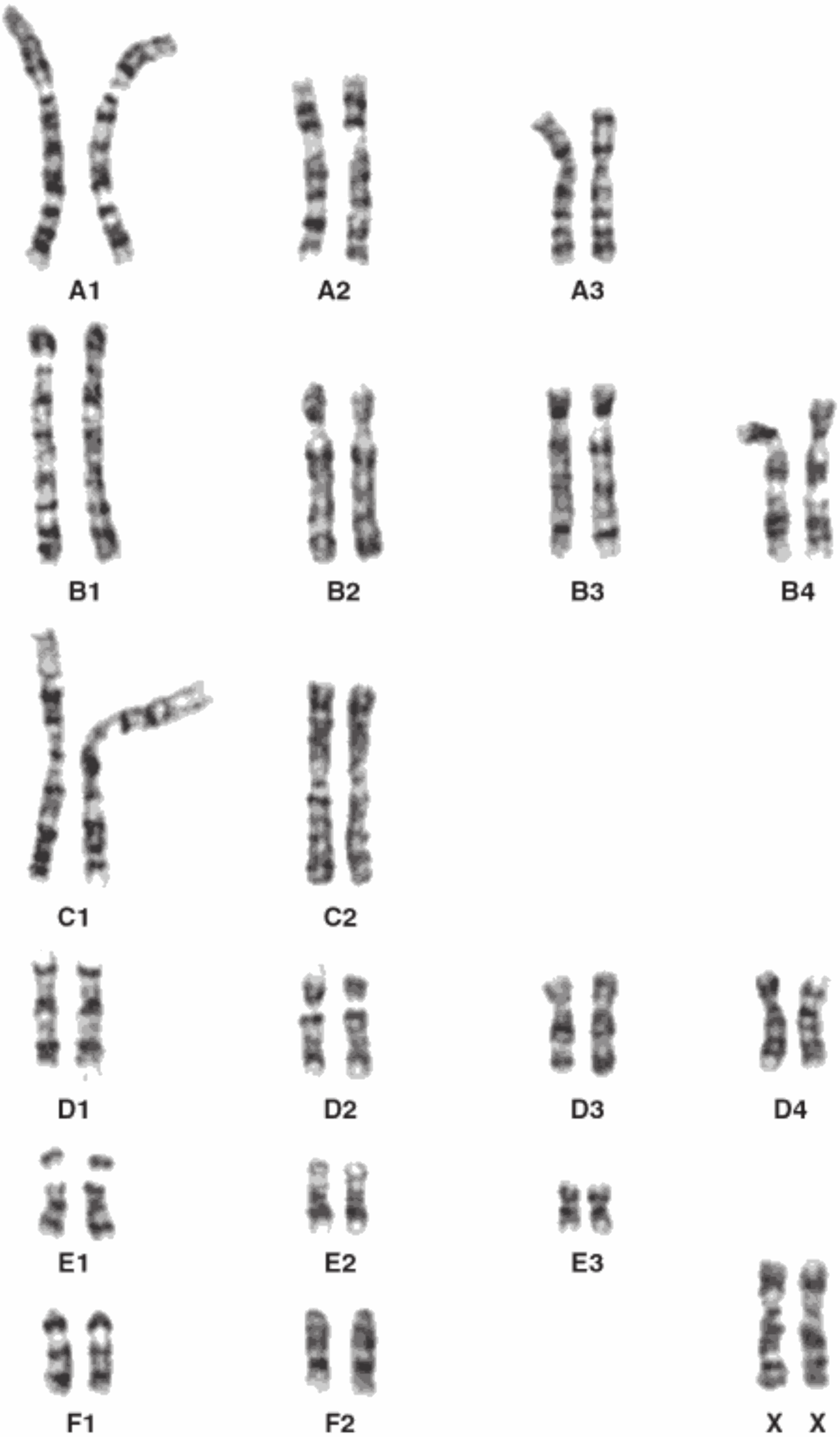
Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, **9**, 465–76.

Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M. and Losick, R. (2008). *Molecular biology of the gene*, 6th edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor & Benjamin Cummins, San Francisco.

Yang, F. and Wang, P.J. (2009) The mammalian synaptonemal complex: a scaffold and beyond. *Genome Dynamics*. **5**, 69–80.

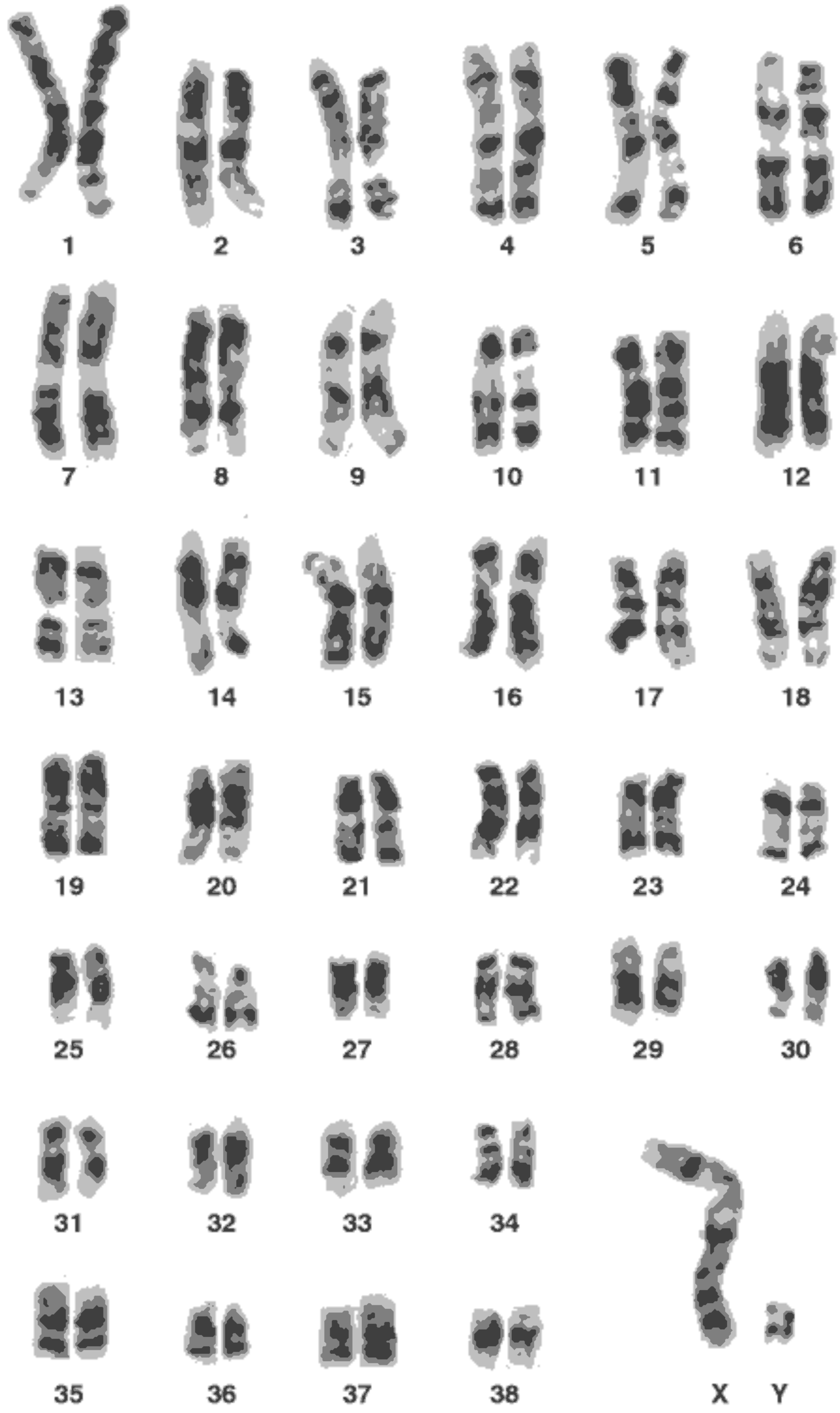
Copyrighted material

Appendix 1.1 Banded karyotypes of domestic species



Cat (G-banding). (Reproduced by permission of W. Nash from O'Brien, S.J., Menniger, J.C. and Nash, W.G. (eds) (2006) *Atlas of Mammalian Chromosomes*, Wiley-Liss, p. 514.)

Copyrighted material



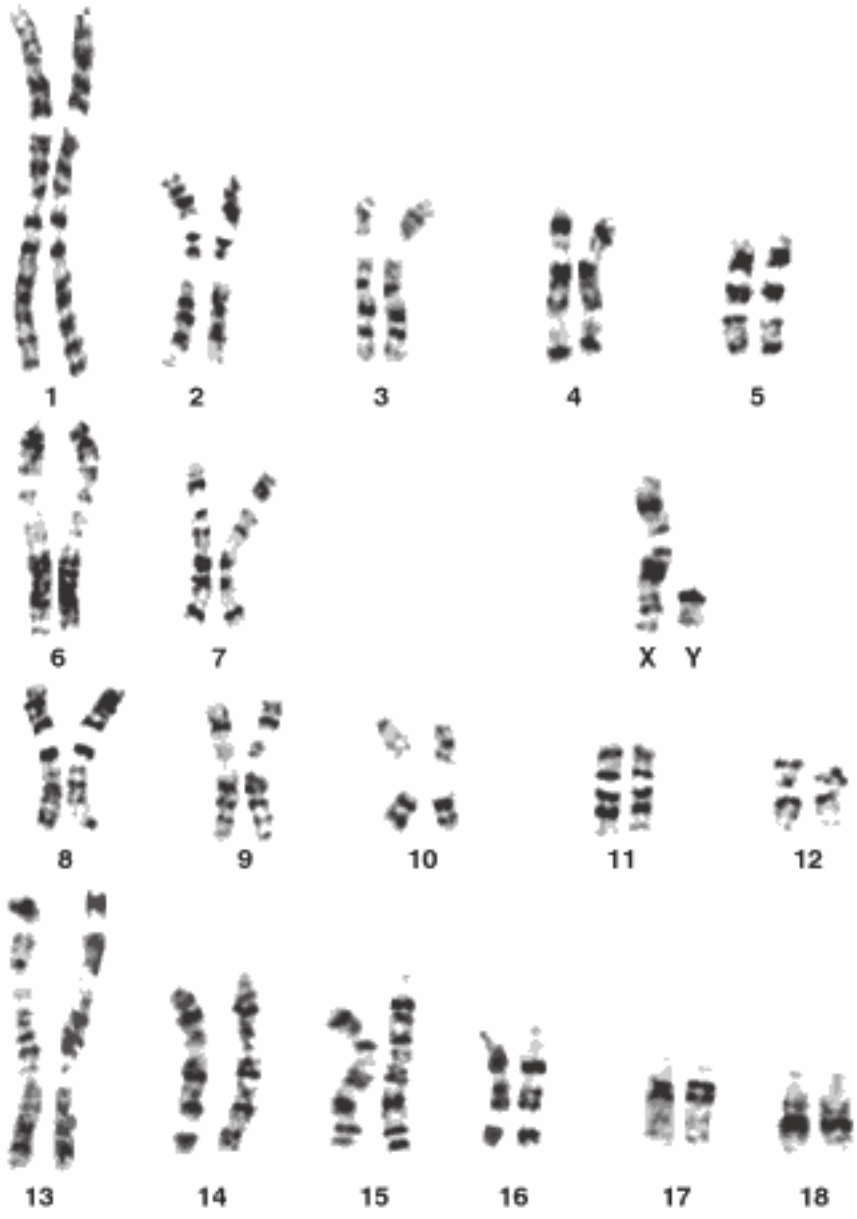
Dog (GTG-banding). (Reproduced by permission of A. Graphodatsky from O'Brien, S.J., Menninger, J.C. and Nash, W.G. (eds) (2006) *Atlas of Mammalian Chromosomes*, Wiley-Liss, p. 467.)

Copyrighted material



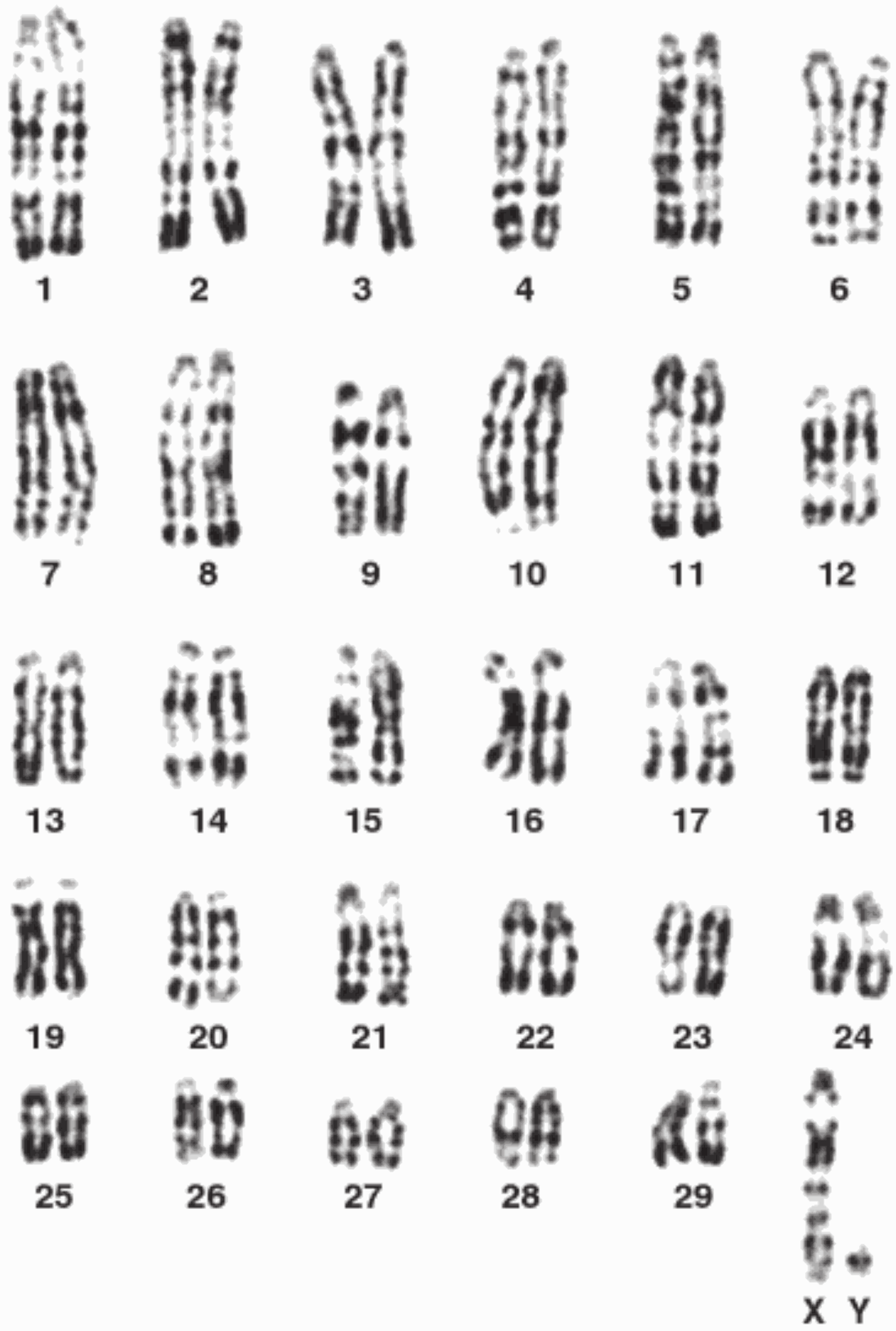
Chicken: (above) the original chromosome spread; (below) the eight largest autosomes and the sex chromosomes. (DAPI/AMD banding; producing the equivalent of G bands). (Reproduced courtesy of M. Völker and D.K. Griffin.)

Copyrighted material



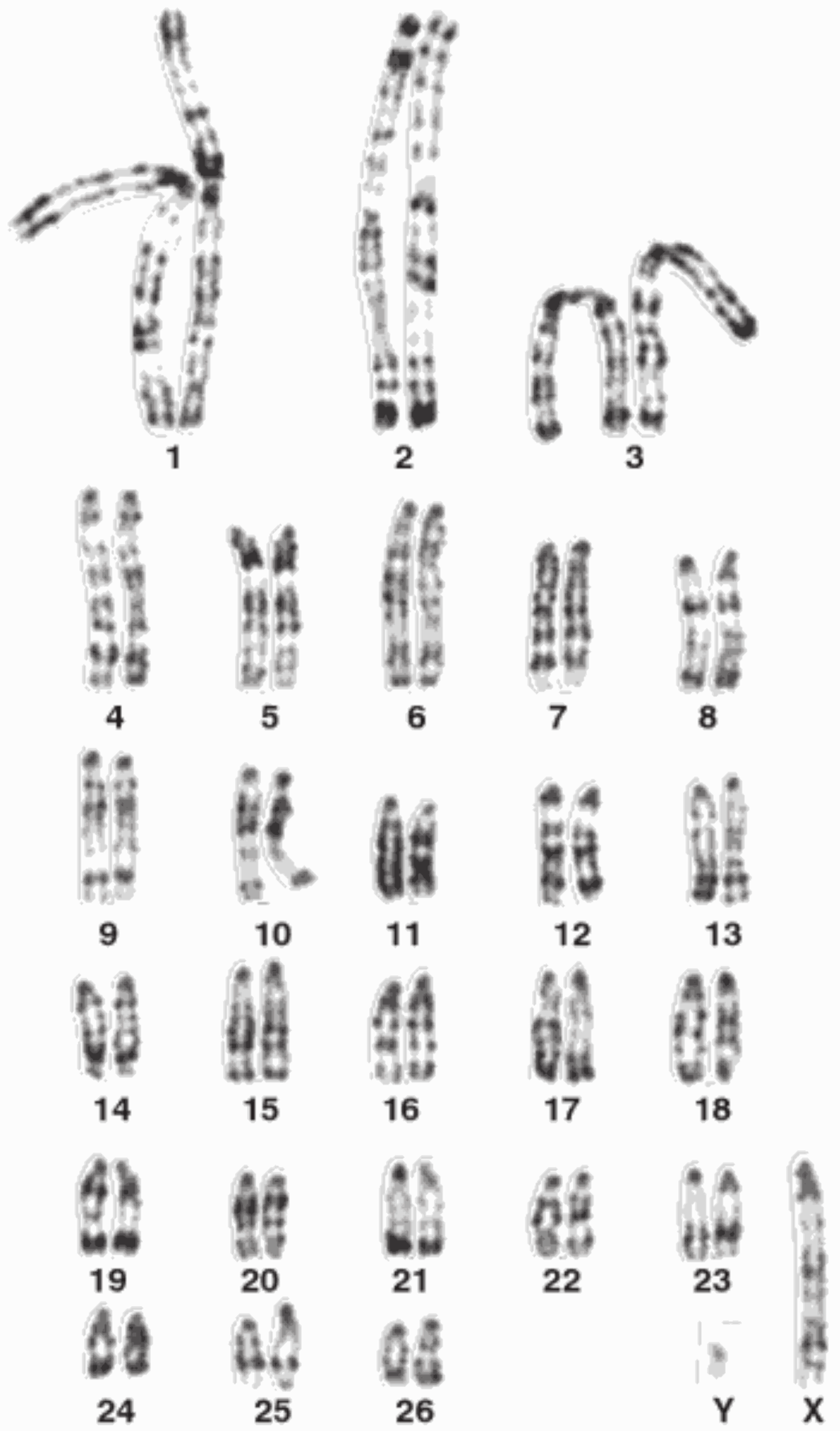
Pig (GTG-banding). (Reproduced by permission of A. Graphodatsky from O'Brien, S.J., Menninger, J.C. and Nash, W.G. (eds) (2006) *Atlas of Mammalian Chromosomes*, Wiley-Liss, p. 566.)

Copyrighted material



Goat (RBG-banding). (Reprinted by permission from S. Karger AG, Basel: ISCNDB (2000), Di Berardino, Di Meo, Gallagher, Hayes and Iannuzzi, *Cytogenetics and Cell Genetics*, **92**, 283–99.)

Copyrighted material



Sheep (RBG-banding). (Reprinted by permission from S. Karger AG, Basel: ISCNDB (2000), Di Berardino, Di Meo, Gallagher, Hayes and Iannuzzi, *Cytogenetics and Cell Genetics*, **92**, 283–99.)

Copyrighted material



Horse (GTG banding). (Reproduced by permission of R. Stanyon from O'Brien, S.J., Menninger, J.C. and Nash, W.G. (eds) (2006) *Atlas of Mammalian Chromosomes*, Wiley-Liss, p. 669.)

Copyrighted material

2 Molecular biology

Molecular biology is a general term encompassing the manipulation of DNA and RNA in the laboratory. As well as providing powerful research tools, molecular biology is already being applied in vaccine production, veterinary diagnosis, manufacture of therapeutic proteins, detection of carriers of harmful genes, pedigree varification, genome-wide selection, and in gene therapy. In the future, it will be increasingly difficult to find areas of animal health and production that are not substantially affected by this technology. In the following account, we shall review the most important features of this rapidly expanding field.

Restriction enzymes

In 1970, it was discovered that bacteria produce enzymes that are able to degrade foreign DNA that enters the bacterial cell. Because these enzymes play a key role in a phenomenon known as host restriction (whereby a bacterium protects itself from viruses by cutting the viral DNA into segments), they are called **restriction enzymes**. Hundreds of restriction enzymes are now known, and a standard method of naming them has been developed: the first letter of the genus name, followed by the first two letters of the species name, followed by letters indicating strain, order of discovery, etc. For example, the enzyme *Bam*HI is obtained from ***Bacillus amyloliquefaciens***, with H indicating a particular strain, and I indicating that it is the first enzyme obtained from this strain. Restriction enzymes bind to specific sequences of bases called **recognition sequences**, and each enzyme cuts the DNA at a specific **cleavage site** that is usually located within the recognition sequence. In most cases, recognition sequences are palindromes, which, as we saw in Chapter 1, contain exactly the same sequence when read in either direction. A list of commonly used restriction enzymes, together with their recognition sequences and cleavage sites, is given in Table 2.1. Check the table for palindromes: the sequence in the top strand when read from 5′ to 3′ is exactly the same as the sequence in the bottom strand when read from 5′ to 3′.

Some restriction enzymes, such as *Sma*I, cut both strands of the double helix at the same place, leaving two **blunt ends**. Others, such as *Eco*RI, cut each strand at a different position, leaving some nucleotides unpaired. Because complementary

Copyrighted material

Table 2.1 Some common restriction enzymes

Source	Enzyme	Recognition sequence, and cleavage site (↑↓)
Arthrobacter luteus	*Alu*I	↓ 5′… -A-G-C-T- … 3′ 3′… -T-C-G-A- … 5′ ↑
Bacillus amyloliquefaciens H	*Bam*HI	↓ 5′… -G-G-A-T-C-C- … 3′ 3′… -C-C-T-A-G-G- … 5′ ↑
Escherichia coli R	*Eco*RI	↓ 5′… -G-A-A-T-T-C- … 3′ 3′… -C-T-T-A-A-G- … 5′ ↑
Haemophilus influenzae R_d	*Hin*dIII	↓ 5′… -A-A-G-C-T-T- … 3′ 3′… -T-T-C-G-A-A- … 5′ ↑
Haemophilus aegyptius	*Hae*III	↓ 5′… -G-G-C-C- … 3′ 3′… -C-C-G-G- … 5′ ↑
Providencia stuartii	*Pst*I	↓ 5′… -C-T-G-C-A-G- … 3′ 3′… -G-A-C-G-T-C- … 5′ ↑
Serratia marcescens	*Sma*I	↓ 5′… -C-C-C-G-G-G- … 3′ 3′… -G-G-G-C-C-C- … 5′ ↑

nucleotides readily bind to these unpaired nucleotides, enzymes such as *Eco*RI are said to create **sticky ends**.

When a restriction enzyme is added to DNA, the DNA is cleaved in as many positions as there are cleavage sites for that enzyme; the DNA is said to be **digested** into fragments. Since cleavage sites are not positioned at regular intervals along a DNA molecule, the fragments are of different lengths and hence different weights. If the mixture of fragments is placed at one end of a gel and subjected to electrophoresis, the fragments migrate along the gel at a rate inversely proportional to the logarithm of their weights. At any given time after the electric field was first applied, there are as many different bands of DNA fragments in the gel as there were

Copyrighted material

different-sized fragments in the original mixture, and the position of each band indicates the size of each fragment.

Consider, for example, a relatively small DNA molecule such as that of the bacterial virus called bacteriophage lambda (abbreviated to phage λ), whose total genome consists of a single DNA molecule 49 kilobases (kb) long (Fig. 2.1a). If the total DNA from this virus is digested with the enzyme *Eco*RI, and if the resultant fragments are subjected to electrophoresis and stained with ethidium bromide (which binds to DNA and fluoresces under ultraviolet light), a picture like that in Fig. 2.1b emerges. Each band corresponds to a different fragment, whose size in kb is given beside the gel. The existence of six fragments indicates that there are five *Eco*RI

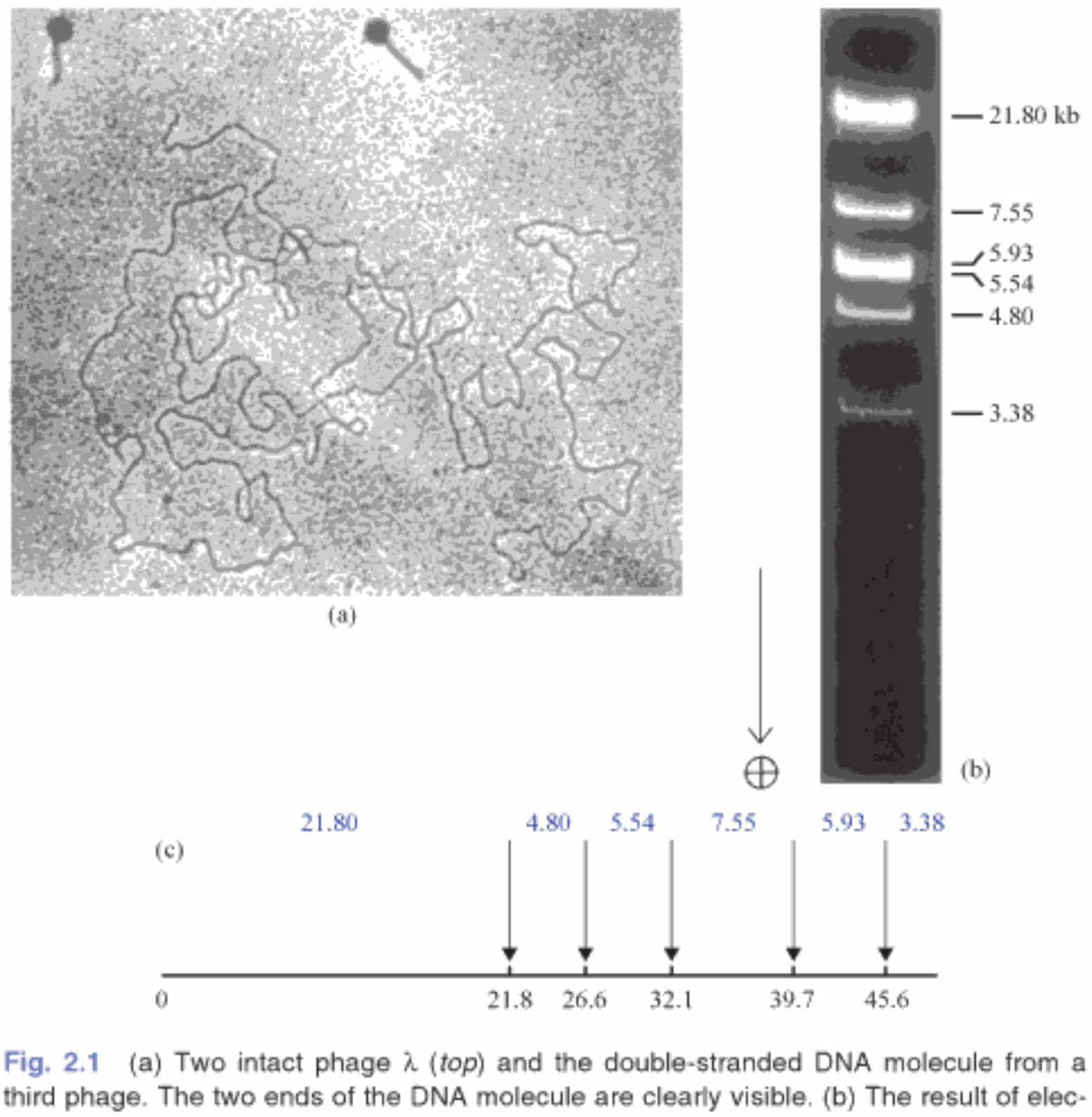


Fig. 2.1 (a) Two intact phage λ (*top*) and the double-stranded DNA molecule from a third phage. The two ends of the DNA molecule are clearly visible. (b) The result of electrophoresis of DNA from phage λ, after digestion with *Eco*RI. (c) An *Eco*RI restriction map of DNA from phage λ. Cleavage sites are indicated by arrows. The numbers under each arrow indicate the location of each site in terms of the number of kilobases (kb) from the left-hand end of the molecule. Numbers above the line (coloured) indicate the size of each fragment in kb, and correspond to the numbers in (b). ((a) Reprinted by permission from J.D. Griffith from Grobstein (1977) *Scientific American* 237(1): 22–33. (b) and (c) Reprinted by permission from Blackwell Publishing Ltd. Old and Primrose (1985) *Principles of Gene Manipulation* 3rd edn.)

Copyrighted material

cleavage sites in the DNA of phage λ. To determine the order of the *Eco*RI fragments in the phage λ genome, the DNA has to be subjected to three separate digestions: (a) with that enzyme, (b) with one other enzyme, and (c) with a mixture of the two enzymes. By comparing fragments resulting from each of the three digestions, the order of the fragments for each enzyme can be determined. This enables a **restriction map** to be drawn (Fig. 2.1c).

Recombinant DNA and DNA cloning

One of the main processes in molecular biology is the production of an unlimited number of copies of a particular segment of DNA. This mass production of a DNA segment is a necessary prerequisite for determining the segment's base sequence (which provides the key to its function) and for many other molecular techniques. It can be achieved in two ways: the first is called gene cloning or DNA cloning, and the second is called the polymerase chain reaction (PCR). We shall start with DNA cloning; PCR will be considered in a later section.

DNA cloning is achieved by joining the DNA that is to be cloned (called **foreign** or **insert** DNA) to a **vector** that is capable of replication within a **host**. The initial requirement for this process is that the insert and vector DNA must have compatible ends (generated by an appropriate restriction enzyme). When insert and vector DNA are mixed together with the enzyme DNA ligase (described in Chapter 1), the end of the insert DNA binds with the end of the vector DNA, in a process called **splicing** or **ligation**. The resultant DNA molecule is one example of **recombinant DNA**. (The term applies to the result of ligating any two segments of DNA from different sources.)

One type of commonly used vector is a **plasmid**, which is a circle of double-stranded DNA, typically 1–3 kb in length, that exists in certain bacterial cells (hosts) independently of the main bacterial chromosome. Plasmids are very useful for cloning relatively small fragments of DNA – up to around 10 kb. A general outline of DNA cloning using plasmids is given in Fig. 2.2.

Another popular vector is the phage λ, which we have already encountered. Up to around 20 kb of the central region of its 49-kb DNA molecule can be removed without interfering with the phage's ability to infect a bacterial cell. If this central region is replaced with an insert of approximately the same size, the resultant recombinant DNA molecule is sufficiently close in total size to a normal phage λ DNA molecule for it to be 'packaged' *in vitro* into phage particles which then obligingly follow their usual practice of attaching to bacterial cell walls and injecting their DNA (in this case, recombinant DNA) into the bacterial (host) cell. Once inside the host, the two ends of the phage DNA join together to form a circle, which replicates in the host. In practice, phage vectors can accommodate inserts in the range of around 10–20 kb.

Larger fragments can be cloned if just the two ends of phage λ DNA (called **cos** ends because they are cohesive, i.e. they join together in the host) are added to a plasmid, forming a **cosmid** vector. Since the phage normally packages a DNA

Copyrighted material

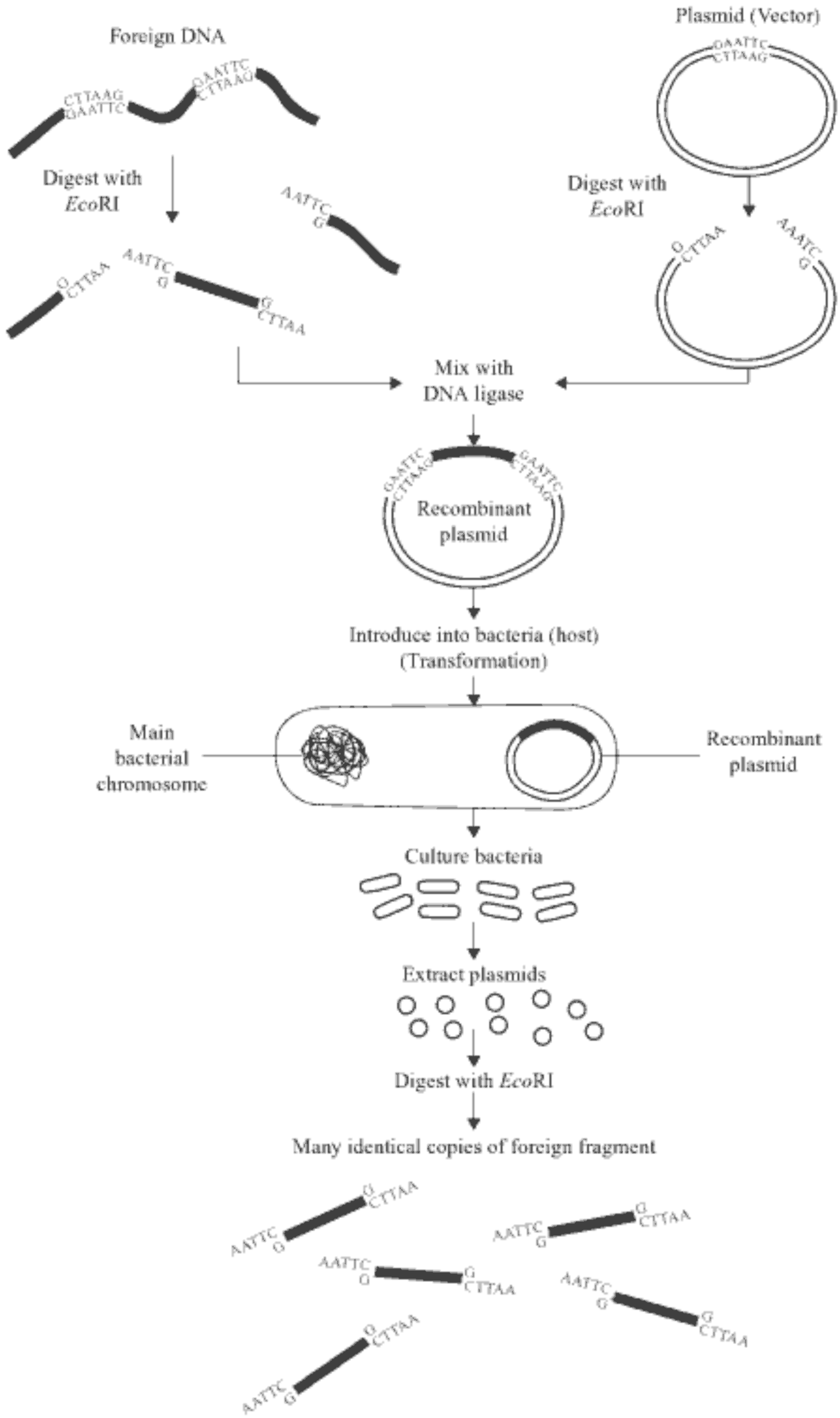


Fig. 2.2 Cloning an *Eco*RI fragment of DNA in a plasmid.

Copyrighted material

molecule of 49 kb with cos ends, and since the total length of the cos ends plus the plasmid is only relatively small, cosmids can be used for cloning fragments in the range 40–50 kb.

Even larger fragments of foreign DNA can be cloned, using vectors such as **Bacterial Artificial Chromosomes** (BACs; 100–300 kb inserts) or **Yeast Artificial Chromosomes** (YACs; 300 kb to 1 Mb inserts).

In many cases, the aim of cloning may simply be to generate an unlimited number of copies of a particular segment of DNA, in preparation for sequencing or some other molecular procedure.

But there are other uses for DNA cloning. For example, if all the DNA from a sample of cells from a cow is digested with one or more restriction enzymes, all of the resultant fragments can be cloned, resulting in a **genomic library** (Fig. 2.3a). In principle, such a library contains all the DNA for that particular species. (In practice, some small segments may be missing.) If more than one restriction enzyme is used, or if a single restriction enzyme is used at a sufficiently low concentration such that not all cleavage sites are actually cleaved (a so-called **partial** digestion), some of the fragments will have segments of DNA in common; the fragments are then said to be **overlapping**.

Complementary DNA

There is a class of viruses whose entire genome consists only of RNA. Because RNA is not capable of replication, the only way for these viruses to propagate is for their RNA to be transcribed 'backwards' into DNA. Because of this, they are called **retroviruses**. The DNA then replicates and acts as a template for the synthesis of more RNA. The enzyme that catalyses the reverse transcription of RNA into DNA is called **reverse transcriptase**. It is extremely important in molecular biology because, among other things, it enables rapid isolation of the coding regions of a gene.

A tissue that is producing polypeptides contains relatively high concentrations of mature messenger RNA (mRNA) for those polypeptides. If mRNA is extracted from that tissue, and if nucleotides and the enzyme reverse transcriptase are added to the mRNA, a complementary strand of DNA (called **complementary DNA** or **copy DNA** or **cDNA**) can be synthesized. Then, by adding more nucleotides and the enzyme DNA polymerase, the single-stranded cDNA can be replicated into double-stranded cDNA.

This cDNA corresponds to the genes that were active in the tissue at the time the mRNA was extracted. Recalling that mature mRNA consists solely of exons, it is evident that cDNA is equivalent to genes from which the introns have been removed.

Fragments of cDNA can be cloned, creating a **cDNA library** (Fig. 2.3b). Such libraries contain cDNA copies of all the genes that were functional in the tissue at the time the mRNA was extracted. Note, however, that whereas a genomic library contains all the DNA of a species (including exons and introns of all structural genes,

Copyrighted material

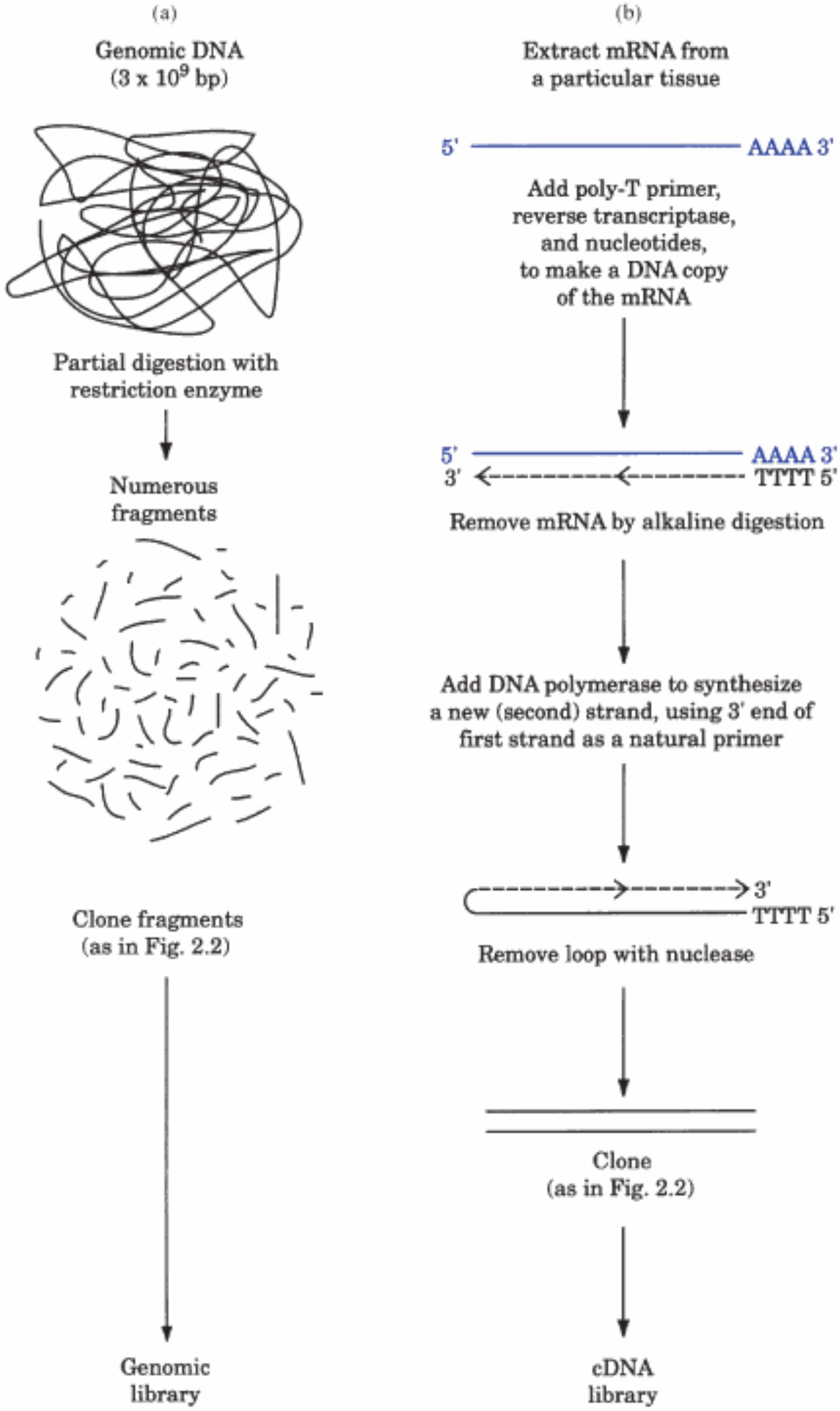


Fig. 2.3 Construction of (a) genomic and (b) cDNA plasmid libraries. RNA is coloured.

Copyrighted material

plus all of the other DNA), a cDNA library contains only the exons of those structural genes that happened to be functioning (i.e. producing protein) at the time when the tissue was sampled.

DNA sequencing

One of the major breakthroughs in molecular biology has been the development of methods for rapidly sequencing fragments of DNA. In fact, DNA sequencing has become so commonplace that it is often one of the first steps taken in studying a segment of DNA. Increasingly, it is also being used for diagnostic purposes.

For many years the method of choice was the **Sanger** or **dideoxy** or **chain-terminating** method (Fig. 2.4). It involves the creation of a series of labelled single strands of varying length, starting from one end of the fragment being sequenced. Electrophoresis of the strands in a polymer gel or through a capillary separates them according to size, creating a 'ladder' of labelled bands, with each band corresponding to an occurrence of a base. In the early days of the technology, bands were labelled radioactively and were visualized by autoradiography, the sequence being read from four adjacent lanes (each lane containing all the bands generated by the sequencing reaction for one of the four dideoxy nucleotides), as shown in Figs 2.4 and 3.3. Increasingly, the use of a different-coloured fluorescent label for each dideoxy nucleotide enables all four sequencing reactions to be pooled into a single lane (or a capillary gel), with the sequence being read directly as a series of peaks, as shown in Fig. 2.5. The size of fragment that can be sequenced by this technology is up to approximately 1 kb. Longer sequences are assembled by combining the sequences obtained from overlapping fragments.

Until recently, large-scale sequencing was also conducted on cDNA libraries. Typically, a single sequencing run was performed on each clone in a cDNA library, resulting in partial sequence of each clone. Called **expressed sequence tags** (**ESTs**), these sequences provide a snapshot of genes that were being expressed in a particular tissue at the time the mRNA was collected from that tissue. If a certain gene was being highly expressed at that time, it will be represented by many clones in the library, and each of these will be represented by a similar EST sequence. By clustering similar sequences together, it was possible to get a rough estimate of the number of different genes being expressed (the number of clusters) and the extent to which each gene was being expressed (the number of ESTs within a cluster). As we shall see later, this strategy for investigating gene expression has been largely overtaken by the use of so-called gene-expression 'chips'.

Recent years have witnessed the development of **pryosequencing**, which has the distinct advantage of not requiring the time-consuming and expensive electrophoresis step. Starting with a single-strand template, pyrosequencing involves synthesizing a complementary strand by attempting to add each of the four nucleotides in turn at each position in the sequence. A successful addition stimulates an enzymatic reaction that results in the generation of visible light that is automatically recorded.

Copyrighted material

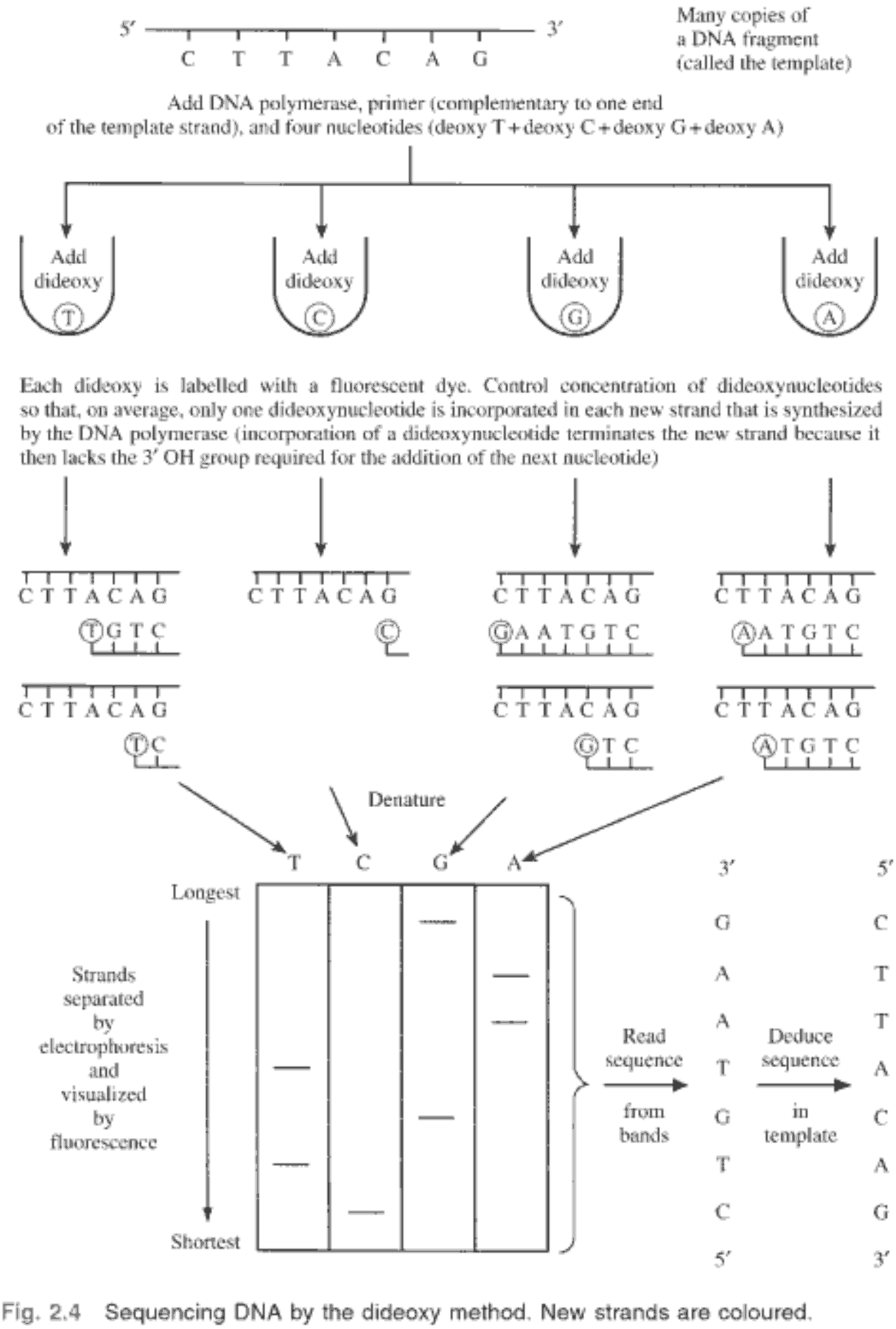


Fig. 2.4 Sequencing DNA by the dideoxy method. New strands are coloured.

Copyrighted material

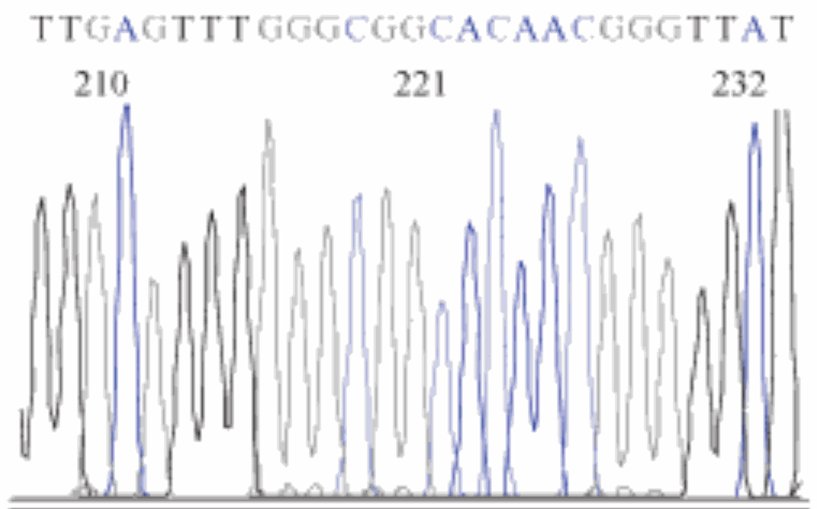


Fig. 2.5 Output from a sequencer in which all four reactions have been pooled into a single lane. Each peak represents a base, whose identity is determined by the colour of the peak. (Adapted from an image provided by R. Harrison, Applied Biosystems.)

The sequence can be read by noting the nucleotide associated with each consecutive emission of light. This **sequencing-by-synthesis** strategy has been extended to enable the simultaneous sequencing of vast numbers of templates, each attached to a microscopic bead. The resultant **massively-parallel** or **next-generation sequencing** has revolutionized the cost-effectiveness of sequencing: prices have plummeted and volume has skyrocketed. In 2008, the whole diploid genome of an individual human (James Watson) was sequenced in just two months at around 1% of the cost of Sanger sequencing. At present, the major limitation of this new sequencing strategy is that the fragments of sequence generated (the **reads**) are far shorter than the approximately 1 kb reads generated by Sanger sequencing. In many circumstances, this is not a major disadvantage. However, as we shall see below, it does restrict the extent to which the resultant sequence can be combined into larger fragments for purposes such as genome assembly.

How do we keep track of, let alone make sense of, sequence data that are being accumulated at an ever-increasing rate? The International Nucleotide Sequence Database Collaboration (INSDC) has been established specifically to coordinate the recording of such data, and free on-line access is available throughout the world (www.ncbi.nih.gov/Genbank; www.ddbj.nig.ac.jp; www.ebi.ac.uk/embl). Most research journals now insist that sequence information be submitted to one of the INSDC collaborators, as a prerequisite for publication. With each mammalian genome containing approximately 3×10^9 bp (i.e. three million kb or 3000 Mb), the rapid expansion of sequencing is posing substantial challenges to information scientists. However, much progress has been made, especially in terms of the development of powerful software for searching and analysing sequences.

Polymerase chain reaction

One of the most powerful tools in molecular biology is the **polymerase chain reaction** (**PCR**). Developed in 1985 by Kary Mullis and colleagues, it is now used

Copyrighted material

extensively in research and is also being applied in diagnosis and in the detection of particular genes (both favourable and harmful) in domestic animals. Increasingly, it is widely used in many areas of veterinary practice and in animal improvement programmes.

The DNA on which PCR is performed (the **template** DNA) can be genomic DNA (extracted from white blood cells or a sample of spleen or other tissue) or a fragment of DNA from any source. PCR creates approximately a million copies of a small segment of the template DNA; sufficient to enable that segment to be cloned or sequenced or analysed in other ways described below. The basic requirement is to know the sequence of the segment, or, at the very least, the sequence at each end of the segment, that we wish to amplify. Using this sequence information, two strings of nucleotides (called **oligonucleotides**, or **oligos** for short), each around 20 bases in length, are artificially synthesized: one is complementary to the 3′ end of one of the strands of the segment to be amplified; and the other is complementary to the 3′ end of the other strand of the segment. Because the role of these oligonucleotides is to prime the synthesis of new strands of DNA, they are called **primers**.

The essential features of PCR are illustrated in Fig. 2.6. The template DNA is placed in a tube, together with the primers, a quantity of deoxynucleotides, and some DNA polymerase. The mixture is heated to around 95 °C to separate (**denature**) the two strands of template DNA. (The DNA polymerase has to be heat-stable so that it can withstand the high temperature of the denaturation phase. A commonly used

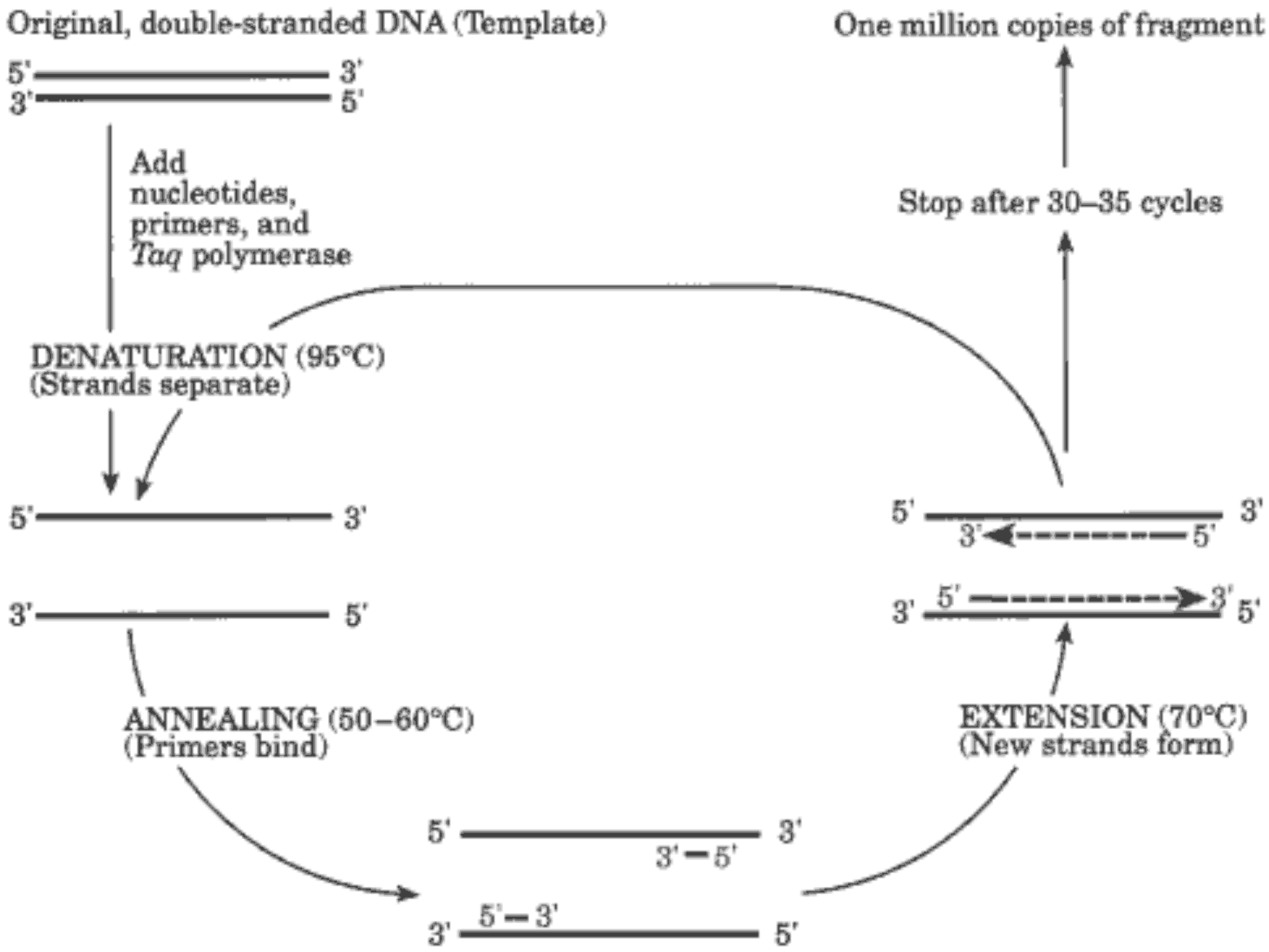


Fig. 2.6 The polymerase chain reaction (PCR). Primers and new strands are coloured.

Copyrighted material

form is ***Taq* polymerase**, so called because it comes from the thermophilic bacterium *Thermus aquaticus*.) The temperature is then decreased to around 50–60 °C, to allow the primers to bind (**anneal**) to their complementary sequences in the template DNA (i.e. to either end of the segment that we wish to amplify). Once the primers have bound, the temperature is increased to around 70 °C, this being optimum for the DNA polymerase, which adds the appropriate nucleotides at the 3′ end of each primer, thereby **extending** the new DNA strands. After allowing sufficient time for the new strands to be synthesized, the temperature is increased to 95 °C again, thereby commencing a new cycle of denaturation, annealing, and extension. The length of one cycle (called an **amplification cycle**) is typically about 5 minutes, consisting of 15 seconds for denaturation, 30 seconds for annealing, 90 seconds for extension, plus the minimum time required to change temperatures between phases (30–60 seconds for each change).

The amplification cycle is usually performed between 30 and 35 times in custom-built PCR machines (thermal cyclers), which provide exact, programmable control over the length and temperature of each step in each cycle. In theory, the result is between 2^{30} and 2^{35} times the original number of copies of the segment – equivalent to between one billion and 35 billion identical segments (called the **PCR product**) in each tube. In practice, the exponential amplification is not perfect. It is still quite possible, however, to achieve at least a one-million-fold amplification, and to obtain 1 µg of PCR product in each tube.

The range and power of PCR is remarkable. For example, PCR products can be obtained from DNA in living organisms, in animal products such as meat and eggs, or in museum or archaeological specimens; and the sample can be as small as a single sperm or a single hair follicle. PCR products can even be obtained from milk and urine, because both contain somatic cells, albeit at a low concentration.

Many variations on the basic PCR theme are now in use. For example, several pairs of primers can be included in the one **multiplex** PCR reaction. The main requirement is that the primers be designed so that the amplified fragments differ sufficiently in size to be distinguishable in a single lane of a gel. In another variation, primers with a random sequence are constructed, giving rise to **random amplified polymorphic DNA** or **RAPD**. (The word 'polymorphic' refers to the existence of more than one type or form.) In some species, RAPDs are a useful source of random DNA markers for genetic maps or (in the case of parasites) for species identification. Another variation involves reverse transcription of a sample of RNA to produce cDNA, which is then amplified by PCR. Known as **reverse transcription PCR** (**RT-PCR**), this is a powerful means of identifying and studying genes without having to go through the cloning process. Another very powerful development has been **real-time quantitative PCR** (**Q-PCR**), in which the quantity of PCR product is measured after every cycle, allowing (by extrapolation) accurate estimation of the quantity of original template. Using the same approach, **quantitative RT-PCR** (also called real-time RT-PCR) enables estimation of the quantity of template mRNA, and hence enables quantification of gene expression. This latter technology is proving to be a very popular diagnostic tool, because it enables rapid and accurate identification of pathogens in clinical specimens.

Copyrighted material

An important application of PCR has been to combine it with dideoxy DNA sequencing, giving rise to a procedure called **cycle sequencing**. Just one primer is used, so that only one strand is amplified. In addition to the usual deoxynucleotides, the PCR mixture also contains some dideoxynucleotides, which, as already explained in Fig. 2.4, terminate the extension of any new strand into which they are incorporated. The result is a graded series of PCR products (exactly the same as those produced in dideoxy sequencing), which can be electrophoresed and visualized as bands from which the sequence can be read, in the usual manner. The advantages of cycle sequencing (and other combinations of PCR and sequencing) are that it can be performed on very small quantities of DNA, and there is no need to clone the DNA before it is sequenced.

Southern analysis and related technologies

Another important molecular technique is Southern analysis, which is named after Ed Southern who developed its central component in 1975. As with PCR, we start with genomic DNA or DNA from some other source. The first step involves digesting this DNA with a restriction enzyme, producing thousands of fragments of various lengths. The fragments are then separated according to their size, by means of gel electrophoresis (as described earlier in this chapter), denatured into single strands, and transferred (blotted) onto a membrane of nitrocellulose or nylon, either by capillary action or under vacuum. (This last step is **Southern blotting**.) The membrane is then baked in an oven at 80 °C (to firmly attach the DNA fragments), and bathed in a solution containing labelled denatured DNA. The bathing is called **probing**, and the labelled DNA is called a **probe**, which is typically derived from a cloned DNA fragment, either genomic or cDNA. Being single-stranded, the probe binds strongly to any cDNA fragment on the membrane. Unattached probe is then washed off, and the fragments to which the probe has bound are visualized by autoradiography or fluorescence (see Fig. 2.7).

The same basic steps have since been adapted to a wide range of circumstances: electrophoresis of RNA followed by probing with DNA is called **northern blotting**; electrophoresis of protein followed by probing with labelled antibody is called **western blotting**; and a more complex procedure in which protein is probed with DNA is called **south-western blotting**.

DNA expression microarrays

As mentioned in Chapter 1, the entire life of all organisms involves sets of genes being turned on and off as required. At the risk of being too reductionist, life can be regarded as essentially a succession of transcriptions and translations. It follows that in order to gain an understanding of many important aspects of living organisms, we need some way to study simultaneously the actions of large numbers of genes.

Copyrighted material

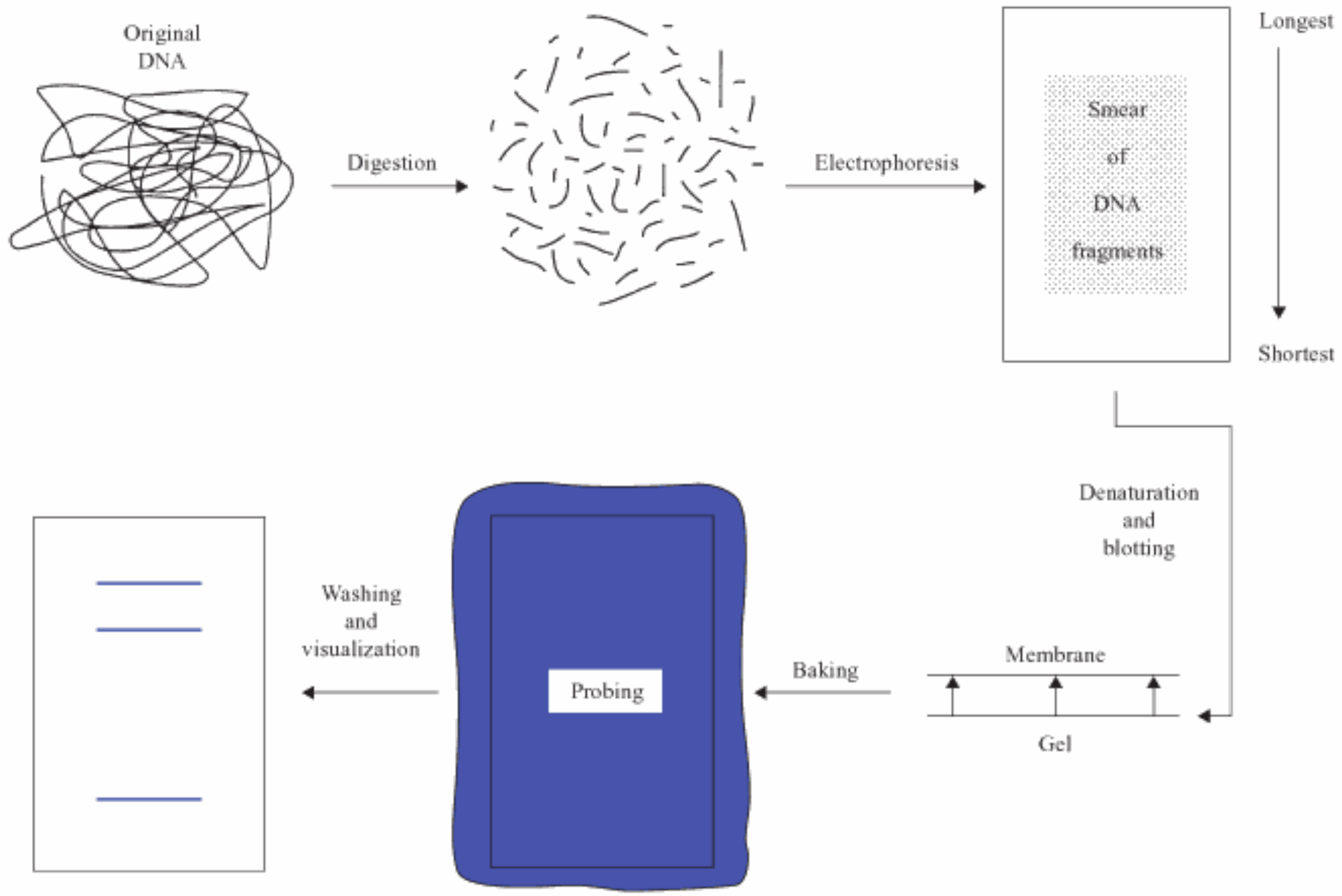


Fig. 2.7 Southern analysis.

Copyrighted material

With so much sequence information having been generated, both genomic and cDNA, and with some extraordinary developments in automation and miniaturization of molecular biological technologies, it has become possible to commence investigating what is called the **transcriptome**, i.e. the entire picture of sequences being transcribed in particular tissues at various stages of development, or under various threats such as infectious disease. The tool by which this is achieved is called a **DNA expression microarray** or **DNA expression chip** or gene expression array or gene expression chip. This comprises a solid surface (typically a glass or silicon chip, or a set of microscopic beads), onto which has been placed tens of thousands of microscopic dots of segments of DNA. This DNA, called the probe, comprises either cDNA or synthesized segments (called oligos) of interest (often representing ESTs), spotted in a grid onto the glass or nylon surface by a robot. Arrays representing most or all of the known coding sequences in some organisms are available commercially, and many more have been produced by individual researchers and commercially. The utility of such a tool will be obvious to readers: for example, if one collects RNA from a tissue infected with a particular pathogen and from a control (uninfected) tissue, reverse-transcribes this into cDNA using fluorescently labelled nucleotides (e.g. green for infected and red for control), and then hybridizes the labelled cDNA (the target) onto an array (analogous to the hybridization step of Southern analysis), the extent to which each of the genes represented in the array is activated in the infected tissue can be automatically measured by estimating the relative quantities of red and green fluorescence emitted from each spot. Green spots are genes activated in infected tissue but not in the control, red spots are the opposite, and yellow spots are genes equally expressed in both samples. In a single small experiment, one obtains information on the simultaneous action of an enormous number of genes.

There are numerous ways in which this technology can be used, all of them involving the comparison of transcription (gene expression) among tissue samples within or between individuals. A far-from-exhaustive list includes comparisons of different tissues, e.g. heart muscle and liver; comparisons of the same tissue at different stages of development or under different circumstances, e.g. infected versus non-infected, treated with a particular drug versus untreated, exposed to a particular environmental factor versus not exposed; and comparison of relevant tissues from animals that are known to carry alleles conferring high or low pathogen/parasite resistance or productivity for a particular trait.

The detection of variation in base sequence

Molecular technology can distinguish between alleles that differ by as little as a single base substitution. In order to illustrate how this is done, we shall consider a practical example (Fig. 2.8).

In a herd of cattle, one animal is homozygous for the sequence AGCT at a particular site in its genome, another animal is homozygous for the sequence AGTT at

Copyrighted material

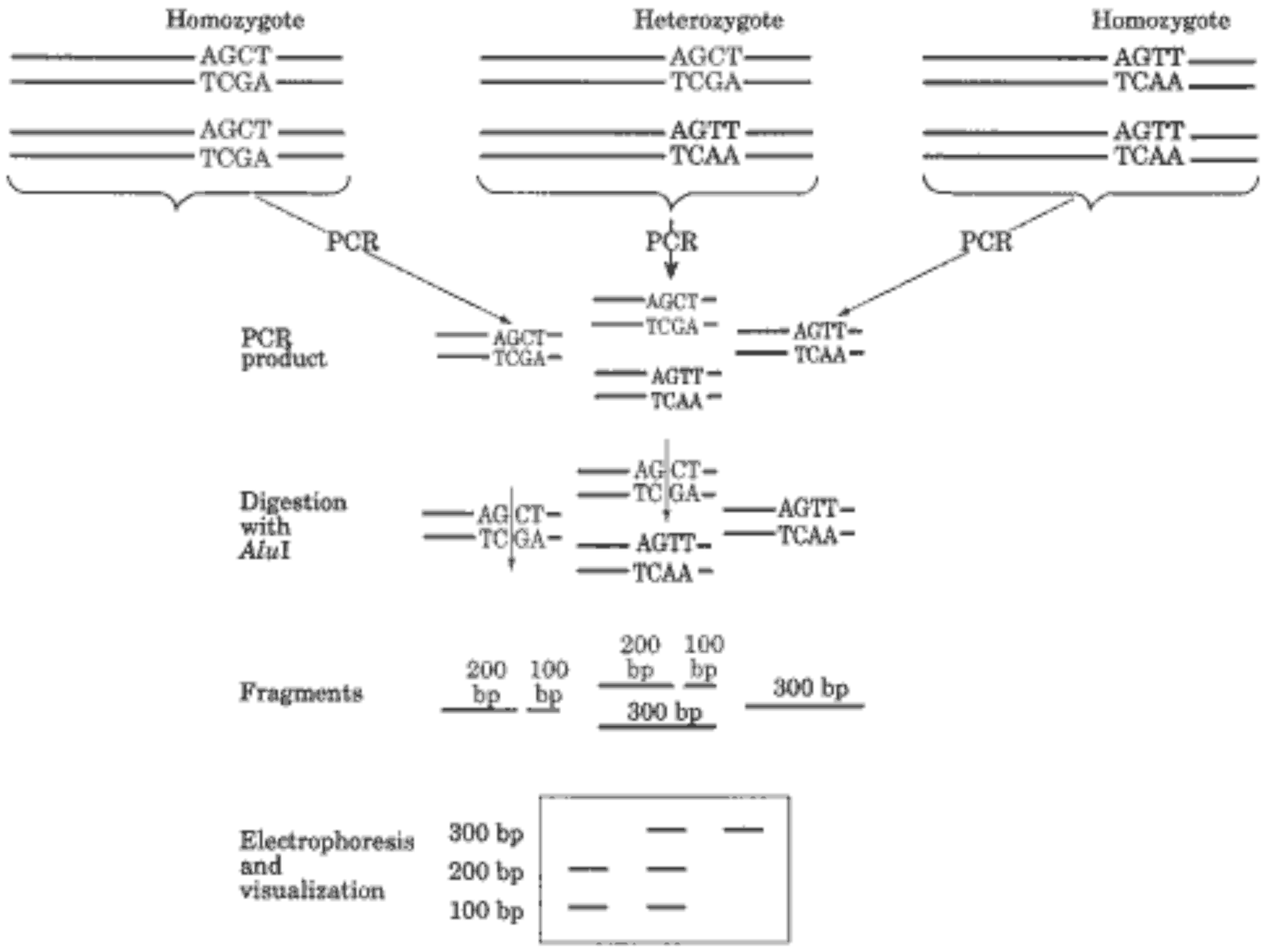


Fig. 2.8 Detection of restriction fragment length polymorphism (RFLP) by polymerase chain reaction (not drawn to scale). Short vertical arrows represent cleavage by the enzyme *Alu*I. The allele with the restriction site is coloured.

the same site, and a third animal is heterozygous at this site. By referring to Table 2.1, you will see that AGCT is the recognition sequence for the restriction enzyme *Alu*I.

PCR is an extremely powerful means of differentiating between these three animals. As mentioned earlier, to use PCR it is necessary to know the base sequence in the immediate vicinity of the region of interest. Using this information, a pair of primers is synthesized so as to amplify a fragment that includes the site containing the base at which the animals differ. In this example, the primers are chosen so that the fragment is 300 bp long and the site is located 100 bases from one end of the fragment.

After amplification, the PCR product from each animal is digested with *Alu*I. All of the PCR products from the first animal contain the sequence AGCT, and are therefore cut by the enzyme at this site, yielding fragments of 100 bp and 200 bp. The second animal produces two types of PCR product: the first is amplified from the chromosome in which the sequence is AGCT, and the other is amplified from the chromosome in which the sequence is AGTT. The former is cut by *Alu*I, but the latter is not, since AGTT is not a recognition sequence for this enzyme. Digestion of the PCR product from the second animal therefore produces fragments of 100 bp, 200 bp, and 300 bp. By similar reasoning, none of the PCR products from the third animal is cut, which means that it has only 300-bp fragments.

Copyrighted material

After digestion, the fragments are electrophoresed and then stained in the gel with either ethidium bromide or silver nitrate. As shown in Fig. 2.8, the three animals can be differentiated readily.

Restriction fragment length polymorphism (RFLP)

The readily visible difference in band patterns between animals in Fig. 2.8 arises from the smallest of all possible genetic differences: the two alleles differ by just a single base substitution. This illustrates the power of this technology.

The difference in band pattern between the three animals is called a **restriction fragment length polymorphism** or **RFLP**. Recalling that the word 'polymorphism' refers to the existence of more than one form or type, the origin of the term is obvious: the gel reveals polymorphism in the length of DNA fragments resulting from digestion with a restriction enzyme.

Although the above example concerns a single base substitution, other types of genetic differences can be detected as RFLPs. In principle, any mutation that destroys or creates a recognition sequence for a restriction enzyme can be detected as an RFLP. For example, insertions, deletions, or base substitutions within a recognition sequence alter that sequence so that it is no longer recognizable by the enzyme.

What about base substitutions that neither destroy nor create a recognition sequence? Fortunately, there are methods for their detection as well. In fact, there are methods which, in principle, enable the detection of any difference in DNA sequence.

Other methods

In one method, the PCR product is denatured into single strands, which are then electrophoresed in a gel. Strands which differ by as little as a single base substitution adopt different shapes (and hence travel at different speeds) when moving through the gel. When the electrophoresis is stopped, each different strand has travelled a different distance, and can be visualized as a different band after adding a DNA stain to the gel. The polymorphism revealed by this technique is called **single-stranded conformational polymorphism** or **SSCP**.

In another method, the double-stranded PCR product is electrophoresed in a gel, through a gradient of some factor that is capable of denaturing double-stranded DNA. (The method is therefore called **denaturation gradient gel electrophoresis** or **DGGE**.) One of the most practical forms of denaturing gradient is temperature, giving rise to **temperature gradient gel electrophoresis** or **TGGE**. As soon as the strands start to separate, they become clogged in the gel. Two fragments of double-stranded DNA that differ by as little as a single base substitution start to separate at different temperatures, and therefore give rise to different bands that are readily visible after staining. A powerful variation on the TGGE theme is to utilize the temperature-changing capability of a PCR machine to expose the PCR product to a temperature gradient in the PCR tube, immediately after amplification, in the presence of a dye that binds only to double-stranded DNA: as the temperature gradually

Copyrighted material

rises, the double-stranded PCR product dissociates (melts) in a manner determined by its sequence, and the melting 'profile' of the PCR product can be quantified as the decrease in fluorescence with increasing temperature. By removing the need for running DNA in a gel, this technology – called High-Resolution Melt (HRM) analysis – lends itself to high-throughput automation, enabling the processing of large numbers of samples.

A variation on Southern blotting is the **dot blot**. There are various types of dot blot, but one of its most powerful forms is where PCR is used to amplify a segment of DNA from each of a group of animals or pathogens. The segment is chosen so as to include one or more sites at which allelic variation is known to occur. The amplified segment from each animal is placed as a small dot on a membrane, which is probed with a labelled synthetic oligonucleotide whose sequence corresponds to a particular allele at the locus. It is possible to control the conditions such that the labelled probe will hybridize only if its sequence is exactly the same as in the amplified segment. By preparing several replicate membranes, each with the same set of dots (from the same group of animals or pathogens), and probing each membrane with a different labelled oligonucleotide corresponding to a different allele, it is possible to determine which alleles exist at that locus in each animal or pathogen. An oligonucleotide used in this type of analysis is called an **allele-specific oligo** or **ASO**. If there are many ASOs, often the procedure is reversed: all ASOs are dotted onto a single membrane, which is then screened with the labelled amplified DNA from a single animal or pathogen. This is called a **reverse dot blot**. Yet another variation is to design primers that are unique to each known allele at a locus, and then to run the PCR under conditions such that amplification of a particular allele occurs only if a perfectly matched primer is present. This is called **allele-specific amplification** (**ASA**).

Increasingly, many of these variations are being replaced by automated cycle sequencing, which can detect *all* base variation in an amplified fragment, i.e. alleles are identified according to their base sequence. If a different fluorescent label is used for each of the four sequencing reactions (A, C, T, G), the products of the four reactions can be placed in a single lane, as shown in Fig. 2.5.

Veterinary diagnosis

The techniques described in the previous section are creating a revolution in veterinary diagnosis: time-consuming and inexact diagnostic tests are being replaced by DNA-based tests which are far quicker and much more sensitive. Most of these tests rely on PCR. The original sample can be from living animals (solid tissue, blood, swabs, semen, faeces, urine), from animal products (eggs, milk, meat), from post-mortem tissue (either fresh or archival), or from water suspected of being contaminated. By appropriate choice of primers and/or probes, the tests can identify the presence of particular species/strains/serogroups of viruses, retroviruses, bacteria, fungi, protozoa, roundworms, or tapeworms. They can also be used to distinguish

Copyrighted material

between meat from different species of animals. The simplest tests involve amplification of a particular segment of DNA, followed by electrophoresis and staining of the gel, which results in a diagnostic band. For RNA viruses, RT-PCR is tailor-made.

The specificity of these tests depends on the region that is amplified (i.e. depends on the choice of primers): the less conserved the region, the more specific is the test. The whole range of possibilities has been exploited, from primers that amplify a fragment that is found in only one strain or serogroup, to primers that amplify a fragment that is common to several genera. In the latter case, digestion of the PCR product with a restriction enzyme often produces a banding pattern that is diagnostic of one species or strain or serogroup. Alternatively, the undigested PCR product can be dot-blotted with oligonucleotides that are diagnostic of species, strain, or serogroup. The possibilities are limitless, and the power of the overall approach ensures that PCR will be an extremely important diagnostic tool for many years to come.

Increasingly, these tests are becoming available via 'chips', on which DNA amplification and sequence detection are performed simultaneously on many samples in so-called massively-parallel processing, thereby enormously speeding up the whole procedure, and decreasing its cost.

Variable number of tandem repeats (VNTR), DNA fingerprints, and microsatellites

While substitutions, deletions, and insertions are very important sources of genetic differences between alleles and between animals, they are not the only type of genetic difference. Another important category involves variation in the number of tandem repeats in repetitive DNA. Before discussing how repetitive DNA can be detected using molecular techniques, we must first expand upon the information provided in Chapter 1.

Repetitive DNA

There are two types of repetitive DNA: **dispersed** and **tandemly repeated**.

Dispersed repetitive DNA is classified according to the size of the repeat unit: the longer repeat units are called **long interspersed elements** (**LINEs**), being typically more than 1000 bases long. The shorter forms are called **short interspersed elements** (**SINEs**), which are usually shorter than 500 bases. Typically, LINEs occur about 10000 times in the genome, while SINEs occur about 100000 times. Considered together, LINEs and SINEs account for about 20 per cent of mammalian DNA. In essence, LINEs are cDNA copies of functional genes present in the same genome. Known as **processed pseudogenes**, these non-functional segments of DNA most probably arose from the unintended action of reverse transcriptase during a retroviral infection. Interestingly, some LINEs appear also to have a degenerate open reading frame for the enzyme reverse transcriptase, suggesting that at some time in

Copyrighted material

the past, they had the capacity for their own transposition, i.e. they were transposable genetic elements (TGEs – explained in Chapter 1). This would help to explain the occurrence of so many copies. Most SINEs also appear to have arisen from the unintended action of reverse transcriptase, but in this case, the DNA is a copy of certain tRNA molecules. For example, a typical SINE is a 73-bp sequence, which occurs in cattle and goats, and which is a DNA copy of the tRNA for glycine.

In tandemly repeated DNA, the repeat unit occurs end-on-end. If there are many copies (typically thousands) of the repeat unit in tandem at one site, the DNA is called **satellite** DNA. The size of the repeat unit in satellite DNA typically is in the range from 5 to 500 bp, with the total number of repeat units at any one site ranging between 1000 and 50 000. In some cases, the longer repeat units include many imperfect repeats of a smaller repeat unit. An example of satellite DNA is a 23-bp repeat unit which occurs at several locations in the genome of cattle, with a total of approximately 7 000 000 copies. Another example is a 483-bp repeat unit in cats, which includes about 25 imperfect (i.e. slightly variable) copies of a 6-bp repeat unit (TAACCC), and which comprises 1–2 per cent of the cat genome. A similar, slightly longer repeat unit exists in dogs.

As the size and number of tandem repeats at any one site decreases, the terms **minisatellite** and **microsatellite** are used to describe them.

Minisatellite repeat units typically range in size from 10 to 100 bases. They are more widespread in the genome than satellite DNA, but their distribution still tends to be concentrated in certain regions such as telomeres and sites with an unusually high frequency of recombination (**recombination hotspots**). Indeed, it is thought that minisatellite DNA may actually be involved in the initiation of recombination.

The term microsatellite refers to the smallest repeat units (around five bases or less, e.g. ACCGG, ATTT, GGC, AC, T, etc.), with the number of repeat units at any one site typically varying from just a few up to around 30. Microsatellites occur throughout the genome. A typical example of a microsatellite is the repeat unit AC, which occurs at approximately 100 000 different sites in a typical mammalian genome.

VNTR and DNA fingerprints

A major characteristic of repetitive DNA is that there is considerable variation in the number of repeat units at any one site: a site on one chromosome might have, say, 24 tandem copies of a certain repeat, while at the same site on the homologous chromosome in the same animal, there may be 21 tandem copies. In another animal, there might be 23 copies at the same site on one chromosome, and 15 copies on the homologous chromosome. With some important and notable exceptions, which are discussed in Chapter 3, this variation in the number of repeat units does not seem to have any adverse effect.

***V*ariable *n*umber of *t*andem *r*epeats** (abbreviated to **VNTR**) can be detected by Southern analysis, using a restriction enzyme that does *not* cut within the repeat unit, and probing with a tandem set of the repeat unit (Fig. 2.9). The size of repeat unit

Copyrighted material

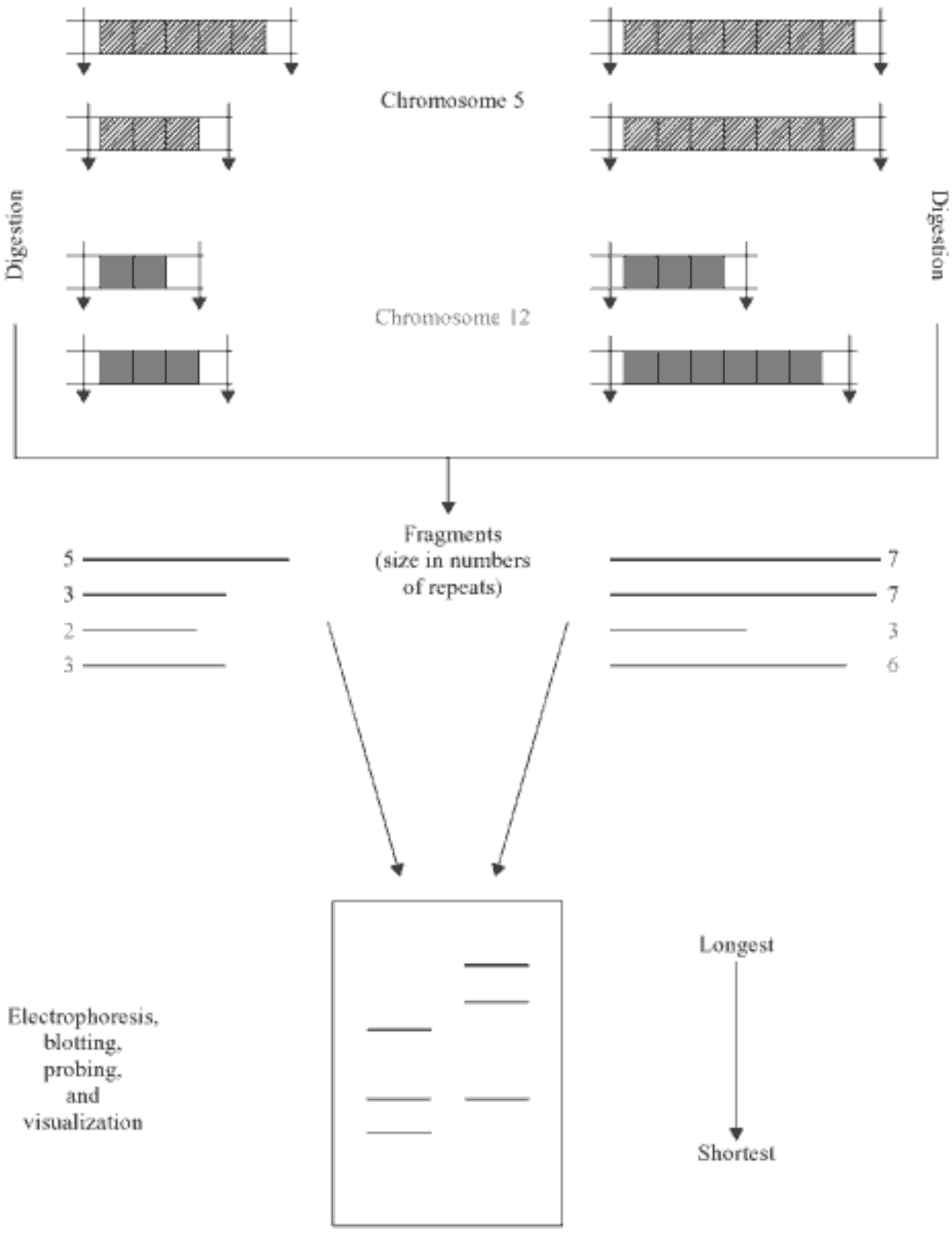


Fig. 2.9 Detection of variable number of tandem repeats (VNTR) in two animals, using Southern analysis. Short vertical arrows indicate cleavage sites for a restriction enzyme that does not cut the repeat sequence. In this example, the repeat sequence occurs at two different sites (loci) in the genome; the site on chromosome 12 is coloured, as are the bands arising from this chromosome. With the exception of the locus on chromosome 5 in animal B, each animal is heterozygous for the number of repeat units (shown alongside the fragment) at each locus. The ladder of bands is called a DNA fingerprint. In practice, there are usually more than two loci, and hence more bands than shown in this example. Note that the middle band from animal A should be both coloured and black, because it actually comprises two bands of the same size, one from each locus.

Copyrighted material

most suitable for this type of analysis is in the minisatellite range of 10–20 bases. Since these repeat units tend to be located at several sites in different chromosomes, and since there is a different number of repeats at almost every site, the repeat-unit probe detects many bands (in theory, two bands at each site, assuming a detectably different number of repeats on each homologous chromosome at each site). The result is a relatively large number of bands. Since there is so much variation in the number of repeat units at any one site, each animal has an almost unique set of bands – a **DNA fingerprint**. Although each band is inherited from a parent, it is not easy to follow the inheritance of DNA fingerprint patterns, because they are jointly determined by so many loci. They were, however, of substantial practical benefit in providing the first DNA-based means of checking pedigrees, as discussed in Chapter 11. An alternative way to detect minisatellite VNTR is to use a probe corresponding to unique-sequence DNA adjacent to a particular minisatellite site. The result is a far simpler band pattern which shows single-locus inheritance.

Notice that even though we are not dealing here with functional genes, we can still use the terms locus and allele: a locus is a site where the repeat units occur, and each different number of repeat units can be regarded as a different allele at that locus.

Microsatellites

As we saw above, microsatellites involve tandem repeat units at the smallest end of the size range, typically from one base to five bases. They are detected using PCR, with primers corresponding to the unique-sequence DNA that flanks the tandem repeats (Fig. 2.10). Also, electrophoresis has to be conducted in circumstances that allow bands to be distinguished which differ by as little as just one base. Since the electrophoresis conditions used for sequencing DNA fulfil this requirement, microsatellite PCR products are electrophoresed in sequencing gels. In fact, the whole procedure is very similar to that used for sequencing: labelled primers are used, and bands are visualized either by autoradiography or fluorescence. As with sequencing, the detection of microsatellites is becoming increasingly automated.

As its name suggests, the unique-sequence DNA to which the primers correspond occurs only once in the entire genome. This means that even if a particular microsatellite repeat unit occurs at many different sites in the genome, the PCR will amplify just one site, namely the site whose flanking sequence corresponds to the sequence in the pair of primers. Such sites are called **sequence-tagged sites** or **STS**. The length of the PCR product varies according to the number of repeat units at that site.

As with any PCR product amplified from unique-sequence primers, only one or two bands are produced from an animal; one band if the animal is homozygous, and two bands if it is heterozygous. This means that microsatellite PCR products do not give rise to the large number of bands per animal that are characteristic of DNA fingerprints. As we shall see below, microsatellite PCR products are extremely useful, because each pair of primers corresponds to a specific site on a chromosome, and because there are many different alleles at each site, i.e. they are extremely polymorphic.

Copyrighted material

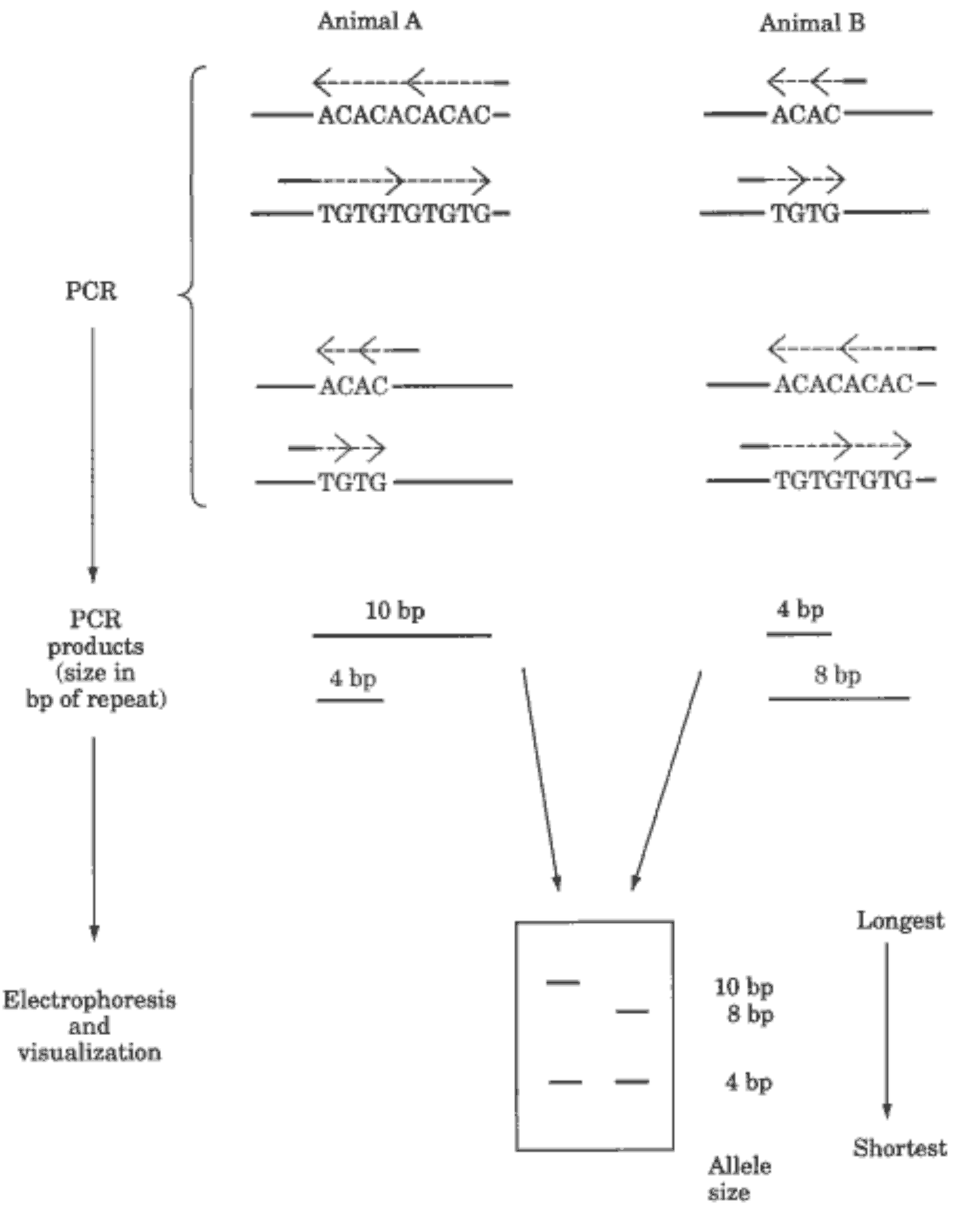


Fig. 2.10 Detection of a microsatellite in two animals, using a single pair of primers, i.e. at a sequence-tagged site. The repeat unit is AC. Polymerase chain reaction products are identified according to the size of the fragment. Fragments from the second chromosome in each animal are coloured.

Single nucleotide polymorphisms (SNPs)

In Chapter 1 we saw that mutation occurs all the time (albeit at a low rate), and that the simplest type of mutation is the substitution of one base for another. Most of these substitutions occur in the vast regions between structural genes, and presumably most have little or no effect. Some occur within coding regions, where they can be either silent, mis-sense, or nonsense. When the sequence of a segment of

Copyrighted material

DNA is compared among a group of animals, on average one finds a base substitution every 500 bases or so. In other words, one finds two alleles at this particular site (locus), where the alleles differ by a base substitution. As mentioned earlier, the existence of two or more alleles at a locus is called **polymorphism**. Where the polymorphism results from a single base substitution, it is called a **single nucleotide polymorphism** or **SNP** (pronounced 'snip'), and the location of each SNP can be called a **locus**. From our earlier discussion of RFLPs, it is evident that RFLPs are a subset of SNPs; in particular, RFLPs are SNPs that create or destroy the recognition site for a restriction enzyme. At some SNP loci, one of the alleles is very rare. But at many SNP loci, both alleles have sufficient frequency to be useful as DNA markers (as we will see below). Judging from humans, in which more than 4 million SNPs have been identified, it seems that a SNP with alleles at a frequency of no less than 5 per cent occurs every 1 kb or so throughout the entire genome.

There is now considerable effort being devoted to the discovery of SNPs in domestic animals. There are several important reasons for this. First, they are truly ubiquitous, being located throughout the genome. Second, they account for much of the genetic variation in traits of interest, including inherited disorders, conformation traits, and production traits (as discussed in many of the remaining chapters of this book). Third, the genotype at SNP loci can be determined without recourse to a gel for separation of DNA fragments, as is required for microsatellites and for the detection of RFLPs. Since the use of gels is a relatively expensive business, SNPs have opened up the prospect of large-scale genotyping at a cost that is orders of magnitude cheaper than microsatellites and RFLPs.

One of the promising methods for large-scale gel-free genotyping of SNPs involves the use of microarrays ('chips'). For example, it is possible to synthesize a segment of DNA (an oligo) that spans a known SNP. If four different versions of this segment are prepared – one for each of the possible variants at the site of the SNP (with A, T, C, or G at that site) – and placed in four different spots on an array, it is possible to perform hybridizations with genomic DNA (from animals to be genotyped) under conditions such that hybridization occurs only if there is an exact match.

SNP chips (containing tens of thousands or even hundreds of thousands of SNPs) are now available, or will soon be available, for the major domestic animal species. With a single chip (costing several hundred dollars, and likely to decrease in cost markedly) it is possible to determine the genotype of an individual animal at each of tens of thousands or hundreds of thousands of loci covering all regions of all chromosomes. This is opening up a whole new set of possibilities for the large-scale practical application of DNA markers. We shall investigate these possibilities throughout the remaining chapters of this book.

Copy number variation (CNV)

In recent years, as researchers have looked more closely at sequence variation between individuals in a population, it has become evident that throughout the

Copyrighted material

genome, individuals differ substantially in the number of copies of chromosomal segments, with the segments ranging in size from around 1 kb up to a staggering 5000 kb, with an average of around 250 kb. This variation, which appears to occur in more than 10 per cent of the genome, is known as **copy number variation** (**CNV**) or **copy number polymorphism** (**CNP**). Not surprisingly, some of the larger regions include functional genes, which means that individuals differ in the number of functional copies of certain genes. Given this, we should not be surprised at tantalizing evidence suggesting that this variation could be a major determinant of genetic variation in so-called multifactorial traits (see Chapters 6 and 13), including many diseases and defects. In some cases, CNV can be detected with SNP chips. As described in Chapter 11, a particularly interesting example of CNV concerns the alleles responsible for white wool in sheep.

Gene mapping

As mentioned earlier, there are probably around 22000 genes in each species of mammal and bird. Ultimately, we would like to be able to identify each of these genes, to ascertain their function, and to document the extent of their allelic variation. A major first step along this path of discovery is the creation of a genetic map for each species.

There are two complementary forms of a genetic map. The first is a **linkage map**, which consists of lists of linked genes (linkage groups) arranged in linear order according to the map distance between them. The second is a **physical map**, which indicates the physical location of each gene on its chromosome.

Both forms of genetic map have existed for many years in a very incomplete form for several domestic species, and in a quite substantial form for chickens. Until recently, the expansion of these maps was limited by the relatively low levels of polymorphism detectable at most loci by conventional means such as visual assessment of traits or biochemical assays. All this was changed with the development and widespread use of molecular technology, and in particular, with microsatellites and SNPs, which, among other things, enable the detection of extensive DNA polymorphism at thousands (or tens of thousands in the case of SNPs) of locations (loci) throughout the genome.

The creation of a linkage map involves estimating the map distance between all pairwise combinations of as many loci as can possibly be genotyped in a group of related individuals (called a **linkage-mapping resource family**). A popular way to do this is to cross two populations that are as genetically different as possible (to create a first-cross or **F1** generation). The F1s are then mated either with other F1s, to produce an **F2** generation, or with members of one of the parental populations, producing a **backcross** (**BC**) generation. The reason for choosing parental populations as different as possible is to maximize the chance of them having different alleles at each locus. This in turn maximizes the chance of the F1s being heterozygous at each locus, which consequently maximizes the chance of being able to

Copyrighted material

estimate the recombination fraction (and hence map distance) for each pair of loci. As an indication of how such an estimate is obtained, the mating between an F1 animal and a member of one of the parental populations is equivalent to the mating shown in Fig. 1.13.

Another method of linkage mapping involves looking at gametes directly, and thereby obtaining direct estimates of recombination frequency. Known as **single-sperm typing**, the method involves flow-sorting sperm cells by **fluorescence-activated cell sorting** (**FACS**), which results in a single sperm cell being placed in each well of a microtitre plate. Each sperm cell is then lysed to expose its DNA, and is subjected to PCR for each of the loci of interest. The identity of the allele at each locus in each sperm is then determined by appropriate analysis of the PCR product. In contrast to traditional linkage analysis, which requires the genotyping of parents and offspring from particular matings, single-sperm typing requires no matings; all that is required is a sample of sperm cells from a single male that is heterozygous at the loci of interest.

While all of the methods of detecting DNA polymorphisms have been tried in gene mapping, it was not until the advent of microsatellites that there were sufficient readily-available polymorphic markers to enable the creation of maps that covered most regions of most chromosomes: in the 1990s, hundreds of microsatellite markers were mapped in each of the major domestic animal species. In most cases, the base sequence of the primers for each microsatellite marker was publicly available, which meant that anyone could generate their own set of mapped markers, simply by preparing the appropriate PCR primers. The first decade of the 21st century witnessed the creation of even denser linkage maps, with hundreds (or even thousands) of SNPs added to the microsatellites. Very quickly, the number of markers available for mapping outstripped the capacity of the available linkage-mapping resource families to map them, the main limitation being that the number of recombinations (cross-overs) in a resource family was too few to separate loci located close together on a chromosome.

This caused attention to switch increasingly towards the other type of map – the physical map. An early form of physical map involves the use of a panel of **somatic-cell hybrids**, where each hybrid is a cell line with a different mixture of chromosomes from two species, e.g. hamster and cattle, derived from an original fusion of cells from the two species. The presence of any particular locus in a particular cell line can be detected by various means, typically PCR. DNA markers or genes that are always **concordant** (all present, or all absent) in each cell line must be located on the same chromosome; the presence or absence of the concordant group of markers and/or genes corresponds to the presence or absence of their chromosome. Loci that are shown in this manner to be located on the same chromosome are said to be **syntenic**, and the group to which they belong is called a **syntenic group**. In the end, each syntenic group should correspond exactly to a linkage group. Even finer location of loci is possible if cell lines contain only portions of chromosomes. For example, by exposing cells from the species of interest to radiation prior to fusion with the hamster cell line, the chromosomes from the species of interest are broken into small fragments. The resultant **radiation-hybrid** (**RH**) panel is a very

Copyrighted material

powerful mapping resource. In cattle, for example, by 2005 there were three RH maps, with the number of loci mapped ranging from 2000 to around 5500.

Another method of physical mapping involves *in situ* hybridization of cloned DNA to its corresponding site on the chromosome, creating a cytogenetic map. Initially, this involved radioactive labelling of the cloned DNA, but now a non-radioactive method known as **fluorescence *in situ* hybridization** (**FISH**) is standard.

A development that shows much promise is the preparation of chromosome-specific DNA libraries, each library consisting of DNA from just one chromosome. These can be obtained from chromosomes separated by FACS or by **microdissection**, which involves using a micromanipulator or laser to pick up just a single chromosome from a spread of chromosomes like that shown in Fig. 1.1. In either case, all of the DNA from that chromosome can be amplified by PCR and then cloned in the usual manner. This is called **microcloning**. When the entire DNA from a chromosome in one species is fluorescently labelled and hybridized (called **chromosome painting**) to a metaphase spread of another species, the homologous chromosomal regions become immediately obvious (Fig. 2.11). Not surprisingly, this process is called **zoo-FISH**, and it has proved to be exceedingly powerful in identifying chromosome homologies across a wide range of species. The results of much of this research can be seen at http://www.chromhome.org/chromhome/. The most striking result is the extent to which whole chromosomes have been conserved throughout mammalian evolution.



Fig. 2.11 A zoo-FISH diagram showing a set of cattle chromosomes that has been 'painted' with DNA from human chromosome 2. It can be seen that this human chromosome shares homology with cattle chromosomes 2 and 11. (Reproduced courtesy of L. Hirst and M. Ferguson-Smith.)

Copyrighted material

Another type of physical map involves digesting the hundreds of thousands of clones from a BAC library (described earlier in this chapter) with a restriction enzyme, and then identifying restriction fragments of the same size from different clones. If two clones each contain a fragment of a particular length, the most likely explanation is that they overlap at this fragment, in which case they can be combined at this fragment, creating a single entity called a **contig** (from contiguous, i.e. adjoining), comprising two overlapping BAC clones. This process of identifying overlapping clones, based on sharing fragments of the same size, is repeated over and over again, creating larger and larger contigs. The resultant **BAC fingerprint map** is really just an extension of the restriction map mentioned at the beginning of this chapter, only now it applies to BAC contigs which can constitute relatively large segments (tens of Mb) of chromosomes.

With very useful and complementary information being generated in linkage maps and in a range of physical maps, an important need emerged for a means to integrate all this information into a single composite map for each species. The most comprehensive strategy involves projecting each separate map (linkage or physical) onto the idiogram of each chromosome, ending up with an estimated location for each locus in terms of kb from one end of the chromosome. In effect, this provides a prediction of where each locus will be located when there is a complete genome assembly for that species. Because the composite locations and the locations from each separate map are all placed in a database, this approach to map integration is known as the Location DataBase (LDB) strategy. An example of an LDB integrated map is shown in Table 2.2.

One of the most striking features to emerge from the creation of genetic maps is the remarkable conservation of groups of loci across widely divergent species, i.e. there is substantial **conservation of synteny** across species, consistent with the conservation of large segments or even whole chromosomes evident from zoo-FISH studies (noted above). For example, many of the genes located on cattle chromosome 9 are on human chromosome 6 (Fig. 2.12) and the X chromosome shows even higher conservation. In fact, if a gene is X-linked in one species of mammal, it is almost certainly X-linked in every other species of mammal. This high degree of conservation has spurred considerable interest in **comparative gene mapping** – the comparison of genetic maps between species. Apart from its evolutionary interest, comparative gene mapping also has substantial practical importance, because the high degree of conserved synteny across species means that a well-mapped region in one species can provide very useful information about the location of genes in other species, long before those genes have actually been mapped in the other species. Comparative mapping has proved to be very useful in tracking down disease loci in domestic animals, based on information from mice and humans.

Keeping track of the rapidly expanding genetic maps, and of comparative mapping information, is proving to be quite a challenge. Fortunately, much of the information is available on the Internet, e.g. http://www.ncbi.nlm.nih.gov/Genomes/; http://www.ensembl.org; http://oxgrid.angis.org.au/; www.thearkdb.org; http://www.informatics.jax.org/orthology.shtml.

Copyrighted material

Table 2.2 Creation of an integrated map from a physical map and a linkage map. In this particular example (the end of sheep chromosome 18 nearest the centromere – called the **proximal** end), nine loci have been physically mapped; six to a single band and three between a pair of bands. The linkage map comprises 20 loci. Two loci occur in both maps, and hence provide valuable links. The final column presents the integrated map, showing the predicted composite location of each locus, in kb from the end of the chromosome

Locus symbol	Contributing maps			Composite location (kb)
	Physical map		Linkage map	
	Left-most band	Right-most band	cM	
DIK4593			0.0	0
MCM216			5.9	3 077
IGF1R			8.2	4 277
BMS1117			11.7	6 103
MCM131			16.8	8 763
SRCRSP5	q15	q15	20.8	10 849
BM3413			22.3	11 632
UBE3A	q12	q12		14 911
NDN	q12	q12		14 911
GABRB3	q12	q13		15 747
OCA2	q12	q13		15 747
BMS1494			35.5	18 517
ILSTS052			35.5	18 517
OARCP134			41.9	21 855
OARVH54			41.9	21 855
BP33			43.3	22 585
EAM			43.3	22 585
IDVGA45			44.2	23 055
MCM148			46.5	24 254
CSRD275H			52.0	27 123
AVEN	q15	q15		27 452
IDVGA79	q15	q15		27 452
D21S45	q15	q15		27 452
CSRD223			53.2	27 749
UWCA4			56.0	29 209
BMS2815	q17	q21	57.9	30 200
BM7243			60.0	31 296

Whole-genome sequence assembly

In recent years, much sequencing effort has been devoted to the creation of an 'assembly' of sequence representing the entire genome (typically around 3 Gb) of a species. The initial challenge in such an exercise is to combine as many overlapping fragments of sequence as possible into a single larger fragment called a **contig** (from contiguous, meaning continuing). Usually, the result is a frustratingly large number

Copyrighted material

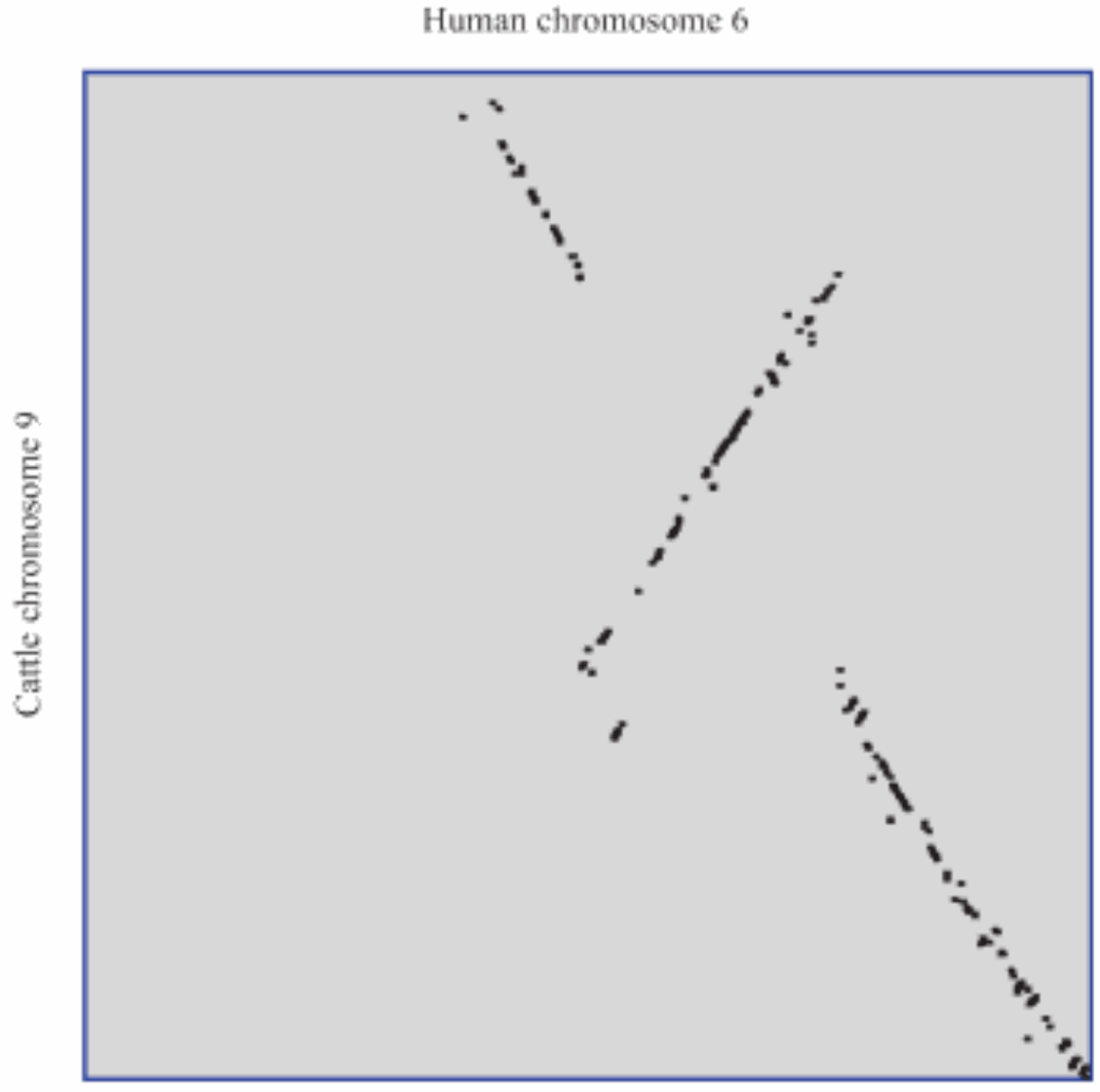


Fig. 2.12 A comparative map showing that almost all of cattle chromosome 9 is comprised of the last two-thirds of human chromosome 6. Each dot represents the location of a locus that has been mapped in both species. The central segment represents an inversion (see Chapter 4). This type of comparative map is called an Oxford grid, after the university in which it was conceived. (Reproduced courtesy of M. Hobbs.)

of non-overlapping contigs, or sets of contigs for which the extent of overlap cannot be determined because of repetitive sequence. The problem is exacerbated with the shorter reads generated by next-generation sequencing. The solution to this problem is to place the contigs on the best available map of that species, preferably a map that integrates all available linkage and physical mapping information. As more sequence becomes available and/or the integrated map is improved, better versions of the assembly (called **builds**) are released.

At the time of publication, 2105 entire genomes have been sequenced, including 35 vertebrates (including cat, dog, pig, cattle, horse, rabbit, opossum, platypus, chicken), 22 insects (including honey bee and the model organism *Drosophila melanogaster*), nine roundworms (including the model organism *Caenorhabditis elegans*), 13 plants, and 1907 microorganisms (http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html).

There are many uses to which sequence information can be put. For example, it is possible to search sequence data, looking for long open reading frames (explained in Chapter 1), which indicate structural genes. This is one way of discovering hitherto unknown genes, and it is the basis of the automated annotation software that is regularly run on each new genome sequence that is assembled. Researchers then have the sequence of putative genes without knowing anything about their polypep-

Copyrighted material

tide. By comparing the sequence in their newly discovered gene with sequences from genes of known function, it is often possible to determine the type and function of the polypeptide, long before it has ever been isolated. For example, if a newly discovered gene contains sequences that give rise to zinc fingers (described in Chapter 1), the newly discovered gene is likely to encode a regulatory protein.

What has been learned to date from the genome-sequencing projects? One of the most startling findings, as mentioned in the previous chapter, is that the total number of genes in a mammal is around 22000. This is considerably less than the best previous guesses of 50000–100000, and is not that much greater than the number of genes in apparently simpler organisms such as flies and worms. There is, however, evidence that many mammalian genes are subjected to alternative splicing, by which means a single gene can give rise to several different peptides.

Production of polypeptide from cloned DNA

When foreign DNA is cloned in bacteria or some other microorganism, the sole aim is to produce unlimited copies of that foreign DNA. If the foreign DNA is a gene, it is possible to go one step further than this; it is possible to arrange for that foreign DNA to produce its corresponding polypeptide in that host. To achieve this, it is necessary to incorporate an appropriate promoter, and possibly some other control regions.

There is one important difference between prokaryotic and eukaryotic control arising from the absence of split genes in the former. Having no split genes, prokaryotes have no need for the elaborate splicing mechanisms required by eukaryotes for converting a primary mRNA transcript into mature mRNA. Because of this, the only eukaryotic genes that can be transcribed and translated in prokaryotes are those consisting solely of exons.

Also, because there are substantial differences in control mechanisms between species, the control regions used should be those of the host cell.

In practice, this is usually achieved by ensuring that the foreign DNA is the coding sequence only, and by inserting this immediately downstream from a leader sequence and promoter in the vector DNA. This 'tricks' the host into thinking that the foreign DNA is, in fact, part of its own DNA. In some cases, it is easier to insert the foreign DNA into the middle of a vector gene, which results in the production of a hybrid (**fusion**) polypeptide from which the foreign polypeptide can be released by appropriate chemical treatment. Vectors that enable the transcription and translation of a foreign gene are called **expression vectors**, and libraries containing such vectors are called **expression libraries**.

Many different polypeptides from prokaryotes and eukaryotes are being mass-produced in the manner described above, using a wide range of microorganisms (both prokaryotes and eukaryotes) and mammalian cell lines as hosts. Examples include recombinant vaccines (protective polypeptides from pathogens or parasites); therapeutic proteins such as insulin, blood-clotting factors, interferons, immuno-

Copyrighted material

globulins, growth hormone; and the many enzymes used in molecular biology. In addition, the technology is not limited to existing genes: it is possible, for example, to design and produce novel polypeptides for specific purposes, either as variants of existing polypeptides, or as completely original molecules.

From using microorganisms to produce foreign polypeptide in the laboratory, it is only a small extra step (at least in principle) to arrange for the microorganism to produce its foreign polypeptide where it is actually needed, such as inside an animal's body. This has given rise to the concept of **live recombinant vaccines**, typically consisting of a harmless virus or bacterium (the vaccine vector), some of whose genes have been replaced by a functional gene for the protective polypeptide from a pathogen or parasite. Rather than having to provide a means of delivering the protective polypeptide by injection or some other labour-intensive process, the vaccine vector quite naturally infects the animal and automatically manufactures the polypeptide where it is needed – inside the animal. The same approach can be used as a means of combating insect pests. For example, if an insect gene for juvenile hormone esterase is inserted into a virus that is a natural pathogen of insects, and if the recombinant virus infects the insects, the esterase inactivates the insects' juvenile hormone, which causes a substantial reduction in larval feeding, which is the most damaging aspect of many insect pests.

Transgenesis

We have seen that it is quite possible to insert a gene from one species, e.g. a pig, into the DNA of a microorganism or mammalian cell line. Furthermore, because of the universality of the genetic code, the same polypeptide will be produced by any host. If this can be achieved with microorganisms and mammalian cell lines, in principle it should be possible to achieve the same result by inserting a foreign gene into the DNA of a living animal. The name given to this process is **transgenesis**.

The first step in transgenesis is to combine the cloned gene with an appropriate promoter, producing what is called a **gene construct**. The second step is to introduce the construct into one or more chromosomes of an animal. At present, this is most commonly achieved by microinjection of many copies of the construct into either the male or the female **pronucleus** of a newly fertilized ovum (the male pronucleus is the haploid nucleus of the sperm cell, and the female pronucleus is the haploid nucleus of the ovum), as shown in Fig. 2.13. After microinjection, the fertilized ovum is transferred into a suitably prepared recipient female, which carries the resultant embryo through to birth in the usual manner. In a very small proportion of cases (currently less than 1 per cent in domestic animals), the genome of the resultant animal contains one or more copies of the construct. Such animals are said to be **transgenic**, and the inserted construct is called a **transgene**. About one-half of transgenic animals actually express (i.e. produce polypeptide from) the transgene, but almost all of them pass the transgene on to their descendants in a normal manner.

Copyrighted material

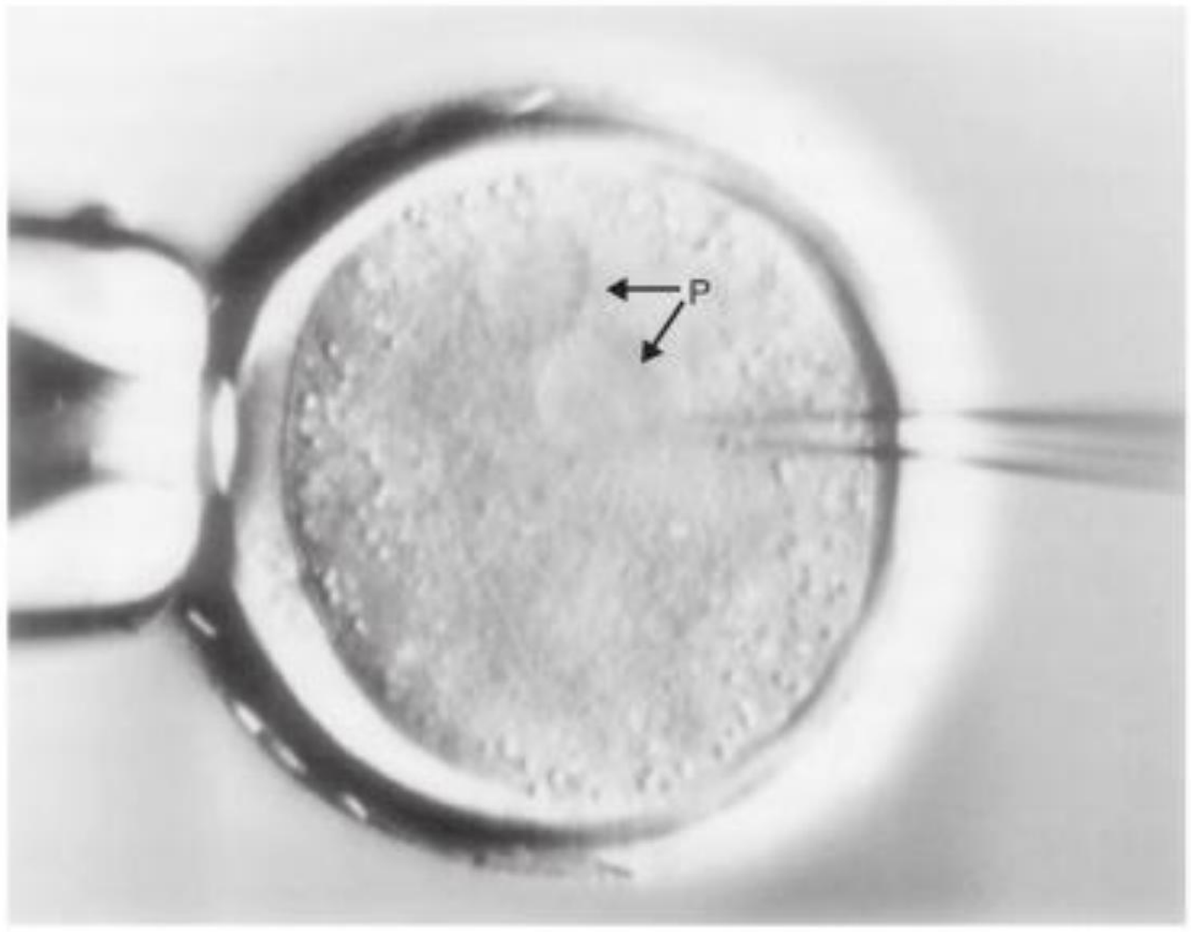


Fig. 2.13 Producing a transgenic sheep by microinjection of a cloned transgene into one of the two pronuclei (P) of a fertilized ovum. (Reproduced courtesy of J. Murray.)

The first transgenic animals were mice produced in 1980. The first application of this technology to farm animals was reported in 1985, when pigs and sheep were made transgenic for the human growth hormone gene. Since then, many different transgenes have been used, from species as diverse as bacteria and humans; and the range of domestic species made transgenic has been expanded to include fish, chickens, goats, and cattle. The microinjection technique has several disadvantages, e.g. there is no control over where the transgene is inserted, nor is there any control over how many copies have been inserted. In the future, we will see more controlled methods used for creating transgenic animals. In mice, for example, undifferentiated and totipotent embryonic cells (called **embryonic stem cells**) can be made transgenic, and are then tested in the laboratory for expression of the transgene. In this way, only those cells that actually express the transgene in the desired manner are transferred to the recipient mouse. This overcomes the problems and expense associated with the very low success rate of microinjection into fertilized ova. Furthermore, by a process called **targeted gene replacement**, it is possible to replace a normal gene in embryonic stem cells with a version of the same gene that has been mutated in a specific manner. By appropriate matings of the resultant transgenic mouse and its descendants, it is possible to establish a strain of transgenic mouse that is homozygous for the specific mutation.

What is the purpose of making transgenic animals? By far the most extensive use of transgenesis is in basic research using laboratory animals, especially mice. The technology has proved to be extremely useful in investigating the regulation of gene expression and in testing methods of gene therapy. When used as a means of replac-

Copyrighted material

ing a functional gene with a non-functional one (producing **knockout** mice), it is also a powerful way to investigate gene function and inherited diseases. Also, transgenic mice whose own antibody genes have been inactivated by targeted gene replacement, but which are transgenic for antibody genes from another species, can be used for the production of species-specific monoclonal antibodies. In relation to farm animals, the uses of transgenesis can be considered from two points of view, associated with the two main parts of a transgene, namely the promoter and the coding sequence.

The appropriate choice of a promoter brings transgene expression under human control. For example, one of the promoters commonly used in early transgenesis was from a metallothionein gene. Metallothionein is a protein that occurs naturally in many species; its role is to bind to heavy metals. In normal (non-transgenic) animals, if the concentration of heavy metals becomes too high, the metallothionein gene begins to produce metallothionein which binds to the excess heavy metal, thereby removing the threat of toxicity. If the promoter from a metallothionein gene is attached to the coding sequence of, say, a growth hormone gene, the resultant transgene can be switched on simply by adding heavy metal to the drinking water of the transgenic animal. This means that the production of growth hormone from the transgene can be turned on and off by a simple manipulation of the drinking water. Alternatively, if the aim is to have a transgene expressed in just one tissue, a promoter is chosen from a gene that is naturally expressed only in that tissue. For example, a transgene containing the promoter from a milk-protein gene will be expressed only in the mammary gland of the transgenic animal.

In relation to the coding sequence, transgenesis offers several possibilities. First, it can be used to introduce a novel allele of a gene that already exists in the species. For example, the transgene for an enzyme involved in lactogenesis could be designed and constructed specifically to produce milk that is low in lactose, thereby giving rise to an improved form of animal product. Second, transgenesis offers the possibility of introducing genes that are completely novel to the species. For example, there is a plant enzyme called chitinase which degrades chitin (a major component of the cuticle and digestive tract of insects). If a chitinase gene could be expressed in the skin of sheep or cattle, it could provide them with a non-toxic form of natural resistance to insects. Another example of this approach comes from chickens, in which natural resistance to viral disease has been produced by transgenes that express polypeptide from the virus that causes the disease. Third, transgenesis can be used to produce human proteins required for therapeutic purposes, such as α1-antitrypsin or blood clotting factors. For example, sheep have been made transgenic for the coding sequence of a human gene (e.g. blood clotting factor IX) attached to the promoter from a gene for a protein that occurs naturally in sheep's milk, e.g. β-lactoglobulin. These sheep produce the human polypeptide in their milk, from which it can be extracted readily. (One of the major advantages of transgenesis as a means of producing therapeutic proteins is that eukaryote hosts are much more likely to be able to process the polypeptide into its appropriate tertiary structure than are the alternative hosts, namely prokaryotes such as bacteria.) Fourth, transgenesis can be used to create animals whose organs or tissues are likely to be suitable for trans-

Copyrighted material

plantation into humans. For example, pigs have been made transgenic for a gene that expresses a human polypeptide on the surface of the cells of the organ or tissue to be transplanted, with the aim of decreasing the problem of rejection. This is seen as an important means of overcoming the critical shortage of human organs for transplantation.

The likely future uses of transgenic animals, and the ethical controversies that surround them, are discussed in Chapter 19.

Antisense technology

In Chapter 1 we saw how genes are regulated by proteins. We also saw the tantalizing evidence that much non-protein-coding RNA is also involved in gene regulation.

This raises some intriguing possibilities for artificial control of gene expression. If the promoter of a gene is placed at the other end of that gene, the non-template strand will be transcribed, producing an mRNA whose sequence is the same as that of the template or antisense strand, and which is therefore complementary to the normal or sense mRNA sequence for that gene. This **antisense RNA** will bind to the sense RNA, preventing translation.

There are many potential ways in which antisense RNA could be utilized. For example, bacteria responsible for the fermentation reactions that result in dairy products such as cheese and yoghurt, could be made resistant to the ever-troublesome phages that are a major cause of fermentation failure. This could be achieved by inserting an essential phage-coding sequence into a bacterial plasmid, ensuring that the coding sequence is in reverse orientation, adjacent to a bacterial promoter. When the recombinant plasmid is inserted into the fermentative bacteria, the promoter will produce antisense RNA which could inactivate the phage RNA, and hence combat the phage infection. Another possibility is to make animals transgenic for one of their own genes, but with its promoter at the wrong end (an antisense transgene). If the gene's normal product acts as an inhibitor of other genes, the antisense RNA produced by the transgene could remove the inhibition, and thus possibly increase the animal's performance in say, reproduction. Alternatively, if the coding sequence in the transgene were from a pathogenic virus or bacteria, the antisense RNA produced by the transgene would be directed against the RNA from the pathogen, thereby controlling the infection in the transgenic animal. Another possibility is to design and construct an antisense oligonucleotide directed against, for example, a key region of the genome of a virus, or even against an aberrantly expressed mutant gene in an animal (e.g. a cancer-causing gene), and then to administer the oligonucleotide as a therapeutic drug.

Apart from its substantial use as a research tool, antisense technology is currently showing more promise than practical results. If it can be properly harnessed, however, it may result in some quite important methods of controlling host genes and combating infections.

Copyrighted material

RNA interference

Double-stranded RNA (**dsRNA**; comprising one sense strand and one antisense strand) can also interfere with translation, by activating a process that leads to the destruction of mRNA or, at the very least, suppression of translation. This is just one example of a more general phenomenon called **RNA interference** (abbreviated to **RNAi**). Two natural sources of dsRNA are viral RNA from an infection or from transposons located within the host genome. The dsRNA is first cut by an enzyme called Dicer into small fragments of around 25 nucleotides (called **small interfering RNA**; abbreviated to **siRNA**). The antisense strand of an siRNA molecule then becomes attached to an RNA-induced silencing complex (**RISC**). Any mRNA with a sequence complementary to the antisense siRNA then becomes attached to the RISC by base-pairing, and is cleaved and subsequently degraded. These forms of RNAi represent important lines of host defence against RNA viruses and the activation of transposons. Another very important example of RNAi involves small segments of host DNA that are transcribed into **microRNA** (**miRNA**) molecules approximately the size of siRNA. The miRNA molecules attach directly to RISC, leading to the cleavage and degradation of any mRNA that includes complementary sequence or to the suppression of translation of that mRNA. This form of RNAi is now thought to be a major means of gene regulation. Also, since injected dsRNA has exactly the same effect as dsRNA from the various natural sources mentioned above, RNAi obviously has enormous potential for controlling the expression of host genes and for combating infection.

Summary

- Restriction enzymes cut (digest) DNA within specific nucleotide sequences, producing either blunt or sticky ends
- DNA cloning is achieved by joining the DNA that is to be cloned (called foreign or insert DNA) to a vector (e.g. a bacterial artificial chromosome, BAC) that is capable of replication within a host (e.g. bacteria)
- A genomic library comprises all the DNA of a species, stored as clones of small fragments created by digestion with one or more restriction enzymes
- The enzyme reverse transcriptase enables a strand of complementary DNA (cDNA) to be created from mRNA (typically mature mRNA, from which the introns have been excised)
- A cDNA library comprises the exons of those structural genes that happened to be functioning (i.e. producing protein) at the time of sampling of the tissue from which the RNA was extracted
- A new generation of sequencing technologies (sequencing-by-synthesis) is based on the synthesis of a new DNA strand against a single strand of DNA whose sequence is unknown (the template), with automatic detection of each nucleotide as it is added during the synthesis

Continued

Copyrighted material

- This sequencing-by-synthesis strategy has been extended to enable the simultaneous sequencing of vast numbers of templates, each attached to a microscopic bead
- The resultant massively-parallel (next-generation) sequencing has revolutionized the cost-effectiveness of sequencing
- The polymerase chain reaction (PCR) involves a chain reaction of replications (2, 4, 8, 16, etc) of a small fragment of DNA (the template) via the enzyme DNA polymerase, resulting in a million or so copies (the PCR product) of the template
- PCR products can be obtained from DNA in living organisms (hosts, parasites, pathogens); in animal products such as meat, eggs, and even the somatic cells in milk and urine; in faeces; in a single sperm; in a single hair follicle; and in museum or archaeological specimens
- A DNA expression microarray comprises a solid surface (a chip or a set of beads) onto which have been arrayed tens of thousands of microscopic dots of segments of DNA corresponding to coding sequences (the probes)
- By measuring the extent of hybridization to the DNA expression array of cDNA from a tissue of interest sampled at a time of interest (the target), it is possible to determine which genes (of those on the array) are being expressed and the extent of their expression in that tissue at that time
- Often, the sequence difference between two alleles at a locus corresponds to the presence or absence of a recognition sequence for a particular restriction enzyme
- When the DNA from animals with different genotypes at such a locus is amplified via PCR and digested with the relevant restriction enzyme, the resultant restriction fragment length polymorphism (RFLP) can be detected by running the digested PCR products on a gel
- Alternatively, PCR primers can be designed that are specific to each allele
- In many different ways, PCR has become a powerful diagnostic tool for determining host genotype at particular loci and for diagnosis of pathogens
- A microsatellite is a locus that includes a variable number of a repeated base sequence, typically one to five bases in length
- Microsatellites are detected by PCR, using primers of unique sequence on either side of the repeats, followed by electrophoresis of the PCR product, which will vary in size (length) according to the number of repeats in that allele
- Having many alleles per locus (high level of polymorphism), and being located in virtually all regions of the genome of most animal species, microsatellites are a valuable form of DNA marker
- A single base substitution results in a single nucleotide polymorphism (SNP)
- SNPs are proving to be exceedingly useful DNA markers, because a SNP with alleles at a frequency of 5 per cent or more occurs roughly every 1 kb or so throughout the entire genome, and also because they can be detected in a massively-parallel manner (on a chip), without the need for running gels

Continued

Copyrighted material

- SNP chips (containing tens of thousands or even hundreds of thousands of SNPs) are now available, or will soon be available, for the major domestic animal species
- With a single SNP chip (costing several hundred dollars but this is likely to decrease markedly) it is possible to determine the SNP genotype of an individual animal at each of tens of thousands or hundreds of thousands of loci covering all regions of all chromosomes
- By estimating recombination fraction between all pairwise combinations of a set of loci, it is possible to identify groups of loci that are linked to each other (linkage groups)
- If this estimation is continued for more and more loci, the end result is a linkage map in which the number of linkage groups equals the haploid number of chromosomes in that species
- A physical map indicates the physical location of a set of loci on their chromosomes
- The most useful map for a species is one that integrates all available linkage and physical maps for that species
- When the maps of different species are compared, it becomes evident that large sections of chromosomes (and sometimes even whole chomosomes) have been conserved among species
- This conservation of synteny among species provides a powerful resource for comparative mapping, whereby locus location and order in a poorly-mapped species can be inferred from location and order in a well-mapped species
- With sequencing becoming ever more cost-effective, the genomes of more and more animal species, and of their pests, parasites, and pathogens, are being sequenced in small fragments
- Combining these small fragments into a complete genome assembly is a daunting task, typically relying on placing the fragments onto a backbone of an integrated map of that species, sometimes using comparative mapping information as well
- The success achieved in coaxing bacteria and other microorganisms to mass-produce polypeptides from a diversity of eukaryotes is remarkable testimony to the universality of the genetic code
- Transgenesis involves inserting foreign DNA into a living animal, with the aim of controlling the expression of a gene in a particular tissue
- Antisense RNA has a sequence complementary to the normal (sense) RNA, and hence has the ability to bind to normal RNA, blocking translation
- RNA interference involves double-stranded RNA (from viruses or from transposons or from an external artificial source) entering the cell and being digested by an enzyme called Dicer into fragments of siRNA; the antisense strand of an siRNA molecule then becomes attached to an RNA-induced silencing complex (RISC) and binds with complementary host mRNA which is cleaved and degraded
- A particularly important form of RNA interference involves small microRNA molecules (miRNA) encoded by host DNA, which appear to be a major means of natural gene regulation

Copyrighted material

Further reading

Beaudet, A.L. and Belmont, J.W. (2008) Array-based DNA diagnostics: let the revolution begin. *Annual Review of Medicine*, **59**, 113–29.

Beckmann, J.S., Estivill, X. and Antonarakis, S.E. (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nature Reviews Genetics*, **8**, 639–46.

Chowdhary, B.P., Paria, N. and Raudsepp, T. (2008) Potential applications of equine genomics in dissecting diseases and fertility. *Animal Reproduction Science*, **107**, 208–18.

Graphodatsky, A.S., Perelman, P.L., Sokolovskaya, N.V., Beklemisheva, V.R., Serdukova, N.A., Dobigny, G., O'Brien, S.J., Ferguson-Smith, M.A. and Yang, F. (2008) Phylogenomics of the dog and fox family (Canidae, Carnivora) revealed by chromosome painting. *Chromosome Research*, **16**, 129–43.

Holt, R.A. and Jones, S.J.M. (2008) The new paradigm of flow cell sequencing. *Genome Research*, **18**, 839–46.

Gupta, S.K., Thangaraj, K. and Singh, L. (2006) A simple and inexpensive molecular method for sexing and identification of the forensic samples of elephant origin. *Journal of Forensic Sciences*, **51**, 805–7.

Liao, W., Collins, A., Hobbs, M., Khatkar, M.S., Luo, J. and Nicholas, F.W. (2007) A comparative location database (CompLDB): map integration within and between species. *Mammalian Genome*, **18**, 287–99.

Mardis, E.R. (2008) Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, **9**, 387–402.

Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, **24**, 133–41.

O'Brien, S.J., Johnson, W., Driscoll, C., Pontius, J., Pecon-Slattery, J. and Menotti-Raymond, M. (2008) State of cat genomics *Trends in Genetics*, **24**, 268–79.

Ostrander, E.A., Giger, U. and Lindblad-Toh, K. (eds) (2006) *The dog and its genome*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Pinard-van Der Laan, M.H. and Gay, C.G. (2008) The future impact of animal genomics in animal health research. *Developmental Biology*, **132**, 3–11.

Primrose, S. and Twyman, R. (2009) *Principles of gene manipulation and genomics*, 8th edn. Blackwell Science, Oxford.

Reale, S., Campanella, A., Merigioli, A. and Pilla, F. (2008) A novel method for species identification in milk and milk-based products. *Journal of Dairy Research*, **75**, 107–12.

Voelkerding, K.V., Dames, S.A. and Durtschi, J.D. (2009) Next-generation sequencing: from basic research to diagnostics. *Clinical Chemistry*, **55**, 641–58.

Zaid, A., Hughes, H.G., Porceddu, F. and Nicholas, F.W. (2001) *Glossary of biotechnology for food and agriculture*. FAO, Rome. (freely available from http://www.fao.org/DOCREP/004/Y2775E/Y2775E00.HTM)

Copyrighted material

3 Single-gene disorders

Inborn errors of metabolism

If a polypeptide acts as an enzyme or is part of an enzyme, a mutation in its gene sometimes results in a deficiency of that enzyme, with a consequent blockage in the biochemical pathway where that enzyme is required. Diseases that result from such blockages are called **inborn errors of metabolism**. A good example of such a disease in animals is citrullinaemia, which occurs in Holstein–Friesian cattle.

Calves with citrullinaemia appear normal at birth but show signs of depression within just a few hours. By the next day, they exhibit considerable depression, tongue protrusion, and an unsteady gait. Deterioration continues during the next few days, with the calves wandering around aimlessly, frothing at the mouth, and pressing their heads against solid objects such as fences. By days 3–5, they collapse and die. The cause of these profound clinical signs is ammonia poisoning caused by a breakdown in the urea cycle, which is the biochemical process by which potentially toxic ammonia (derived from catabolism of proteins) is converted to urea, which is excreted in urine (Fig. 3.1). Calves with citrullinaemia are deficient in one of the enzymes involved in the urea cycle, namely argininosuccinate synthetase (ASS). As can be deduced from Fig. 3.1, the absence of this enzyme leads to a build-up of citrulline and, more seriously, of ammonia, with the latter causing the clinical signs.

What is the cause of the deficiency of ASS? As a result of comparisons between the complementary DNA (cDNA) sequence of the ASS gene obtained from affected and normal animals, it is now known that the cause of citrullinaemia in Holstein–Friesians is a single base-substitution at the first position in the 86th triplet of the gene for ASS. In the normal allele, this triplet is CGA, which encodes the amino acid arginine. In the mutant allele, the C has been replaced by a T, giving rise to a triplet TGA, which (as can be deduced from the genetic code in Table 1.2) is a stop triplet. Thus, the profound clinical signs and eventual death of calves affected with citrullinaemia are the result of the simplest of all possible mutations, namely the substitution of one nucleotide by another.

If tests for ASS activity are conducted among normal cattle within a population known to have produced calves with citrullinaemia, it is found that in some normal animals, including all those that are parents of affected animals, the level of ASS is approximately 50 per cent of that observed in the remainder of normal animals. Not surprisingly, all calves affected by the disease have zero levels of ASS.

Copyrighted material

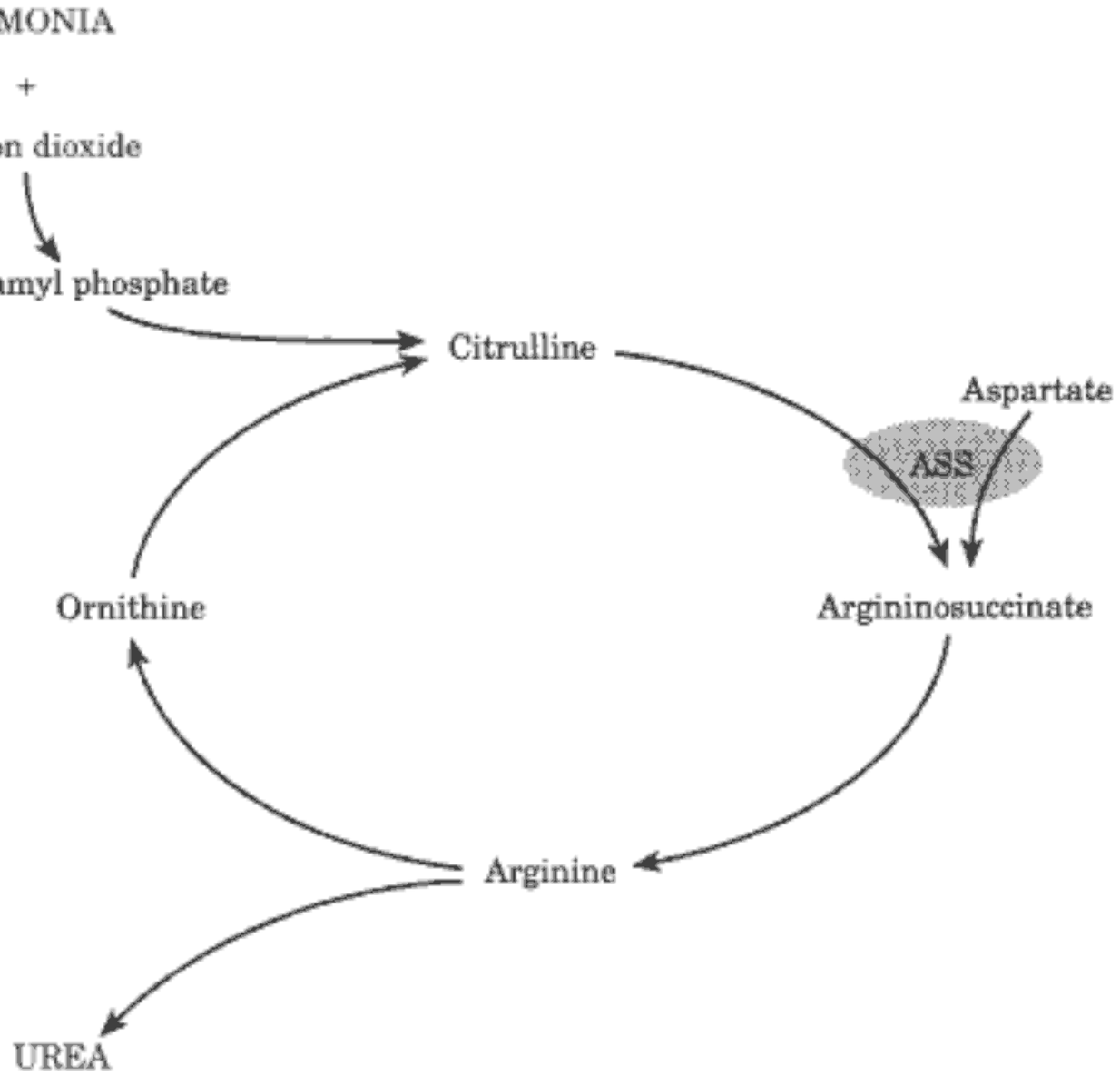


Fig. 3.1 The urea cycle, showing the role of the enzyme argininosuccinate synthetase (ASS) in converting citrulline plus aspartate to argininosuccinate.

The explanation for these observations is that in affected individuals, both copies of the gene responsible for the production of ASS are mutants: affected calves are homozygous for the defective allele. The normal animals with normal ASS activity have two normal ASS alleles: they are homozygous for the normal allele. The normal animals with approximately 50 per cent of normal ASS activity are heterozygotes (often called **carriers**), having one normal allele (producing normal ASS) and one defective allele (producing no ASS). Thus the level of ASS activity is directly proportional to the number of 'doses' of the normal allele.

As well as illustrating carrier detection by biochemical means, inborn errors of metabolism provide good illustrations of other important genetic concepts.

Type of gene action

In the case of citrullinaemia discussed above, and many other inborn errors of metabolism, the normal allele is said to be completely **dominant** to the defective allele with respect to clinical signs. Another way of expressing this is to say that the defective allele is **recessive** to the normal allele, or that citrullinaemia is a recessive disease.

More generally, an allele is recessive for any trait if its effect with respect to that trait is not evident in the heterozygote. Likewise, an allele is dominant with respect to a particular trait if its effect is the same in heterozygotes as in homozygotes.

Copyrighted material

Suppose that the trait with which we are concerned is the level of ASS activity, rather than clinical signs. In this case, the normal allele and the defective allele are said to be each **co-dominant** or **incompletely dominant**, because the heterozygote exhibits the effect of both alleles. The terms recessive, dominant, co-dominant, and incompletely dominant describe relationships between alleles, or **types of gene action**. From the above example, it is evident that two alleles at one locus can exhibit more than one type of gene action, depending on the trait being considered.

Genotype and phenotype

It is convenient to introduce two more terms at this stage. The first is **genotype**, which refers to the genetic constitution of an individual at one or more loci. Using the symbol *D* for the normal allele and *d* for the defective allele, the genotype for carriers of citrullinaemia is *Dd*, while affected calves have the genotype *dd*. The other term is **phenotype**, which is an observable trait of an individual. In relation to ASS activity, there are three different phenotypes (normal activity, 50 per cent activity, and zero activity) corresponding exactly to the three respective genotypes (*DD*, *Dd*, and *dd*). In relation to clinical signs, however, a one-to-one relationship between genotype and phenotype does not exist: the same three genotypes give rise to only two phenotypes. *DD* and *Dd* animals are normal, while *dd* animals are affected. Situations in which there is a one-to-one relationship between phenotype and genotype are the exception rather than the rule.

Sex-limited inheritance

The typical feathering of hens is part of their secondary sexual characteristics, produced by the action of oestrogen. Much of this oestrogen is produced from androgen in the ovaries by the enzyme aromatase. In certain strains of two breeds of chicken, the Sebright Bantam and the Golden Campine, roosters have the same feathering as hens, rather than the typical male form of feathering. This disorder is called henny feathering. It results from a mutation in the aromatase gene, causing the gene to be expressed in the skin of both sexes. In males, this leads to abnormally high levels of oestrogen, which in turn produces henny feathering.

The aromatase gene is autosomal, but the form of inheritance of henny feathering is not what is normally seen with an autosomal mutation because the phenotype associated with the mutation (female feathering in males) can be seen only in males. This is an example of a **sex-limited** disorder.

Henny feathering is interesting on three counts. First, it illustrates that not all mutations result in loss of activity; some can cause a gene to be switched on in cells in which it is normally inactive. Second, it provides a different illustration of the ideas expressed in the previous section. With respect to aromatase activity in skin, gene action is co-dominant, i.e. heterozygotes have an enzyme activity mid-way between that of the two homozygotes. In this case, however, the activity in the skin

Copyrighted material

of the normal homozygote is zero. With respect to feathering, the mutant is dominant, because heterozygotes produce sufficient enzyme in the skin, and hence sufficient oestrogen, to cause henny feathering. Third, while the mutation gives rise to a phenotype that is seen only in males, the normal expression of the gene is seen only in females. Thus, two alleles at the one locus give rise to the two possible forms of sex-limited inheritance.

Genetic heterogeneity of disease

Some animals are born with easily extendible and very fragile skin, a condition that is known as Ehlers–Danlos syndrome, dermatosparaxis, or cutaneous asthenia. In animals suffering from this disease, severe lacerations result from the slightest scratch that in normal animals would cause only trivial damage. The basic cause of these severe clinical signs is the presence of abnormal type-I collagen in the skin.

Like all collagens, type-I collagen is a protein and must therefore be the product of a gene. In fact, it is the product of two genes, because a type-I collagen molecule consists of a triple helix of two α-1 chains (the product of one gene) and one α-2 chain (the product of a separate gene), as shown in Fig. 3.2. As with many polypeptides, the products of the type-I collagen genes are precursors: they contain additional amino acids at each end of the chain that are not present in the mature molecule. When the two α-1 and one α-2 chains initially form a triple helix, the resultant molecule (called type-I procollagen) is not functional and cannot form itself into the cylindrical fibrils that are characteristic of normal connective tissue. In order for this precursor molecule to become fully functional, the surplus amino acids at each end of each chain must be removed, thereby creating a type-I collagen molecule. Not surprisingly, this removal of terminal amino acids is performed by specific enzymes, namely two endopeptidases called procollagen I carboxy-proteinase (PCP-C-I) and procollagen I amino-proteinase (PCP-N-I), as shown in Fig. 3.2. Each of these enzymes is also the product of a gene, so we now have at least four genes involved in the manufacture of mature type-I collagen molecules: two coding for procollagen chains, and at least one for each of the enzymes. (In fact, the enzymes themselves consist of at least two different polypeptides, and hence are the product of at least two genes, but this need not concern us here.)

Obviously, mutations can occur in any of these genes. For example, in cattle and sheep, mutations have occurred in one of the genes coding for one of the chains of the enzyme PCP-N-I. Animals that are homozygous for such a mutation are deficient in the enzyme and therefore experience a build-up of partially processed collagen molecules, still having the surplus amino acids at their N-terminal ends. The skin of these animals is easily extendible and very fragile. In contrast, heterozygotes for this mutation do not show any clinical signs because one-half of the normal concentration of enzyme is sufficient to process all the procollagen molecules. Since heterozygotes are normal, this defect is recessive.

Exactly the same type of abnormal collagen molecule (and hence exactly the same clinical signs) results from a mutation in either one of the procollagen genes, which

Copyrighted material

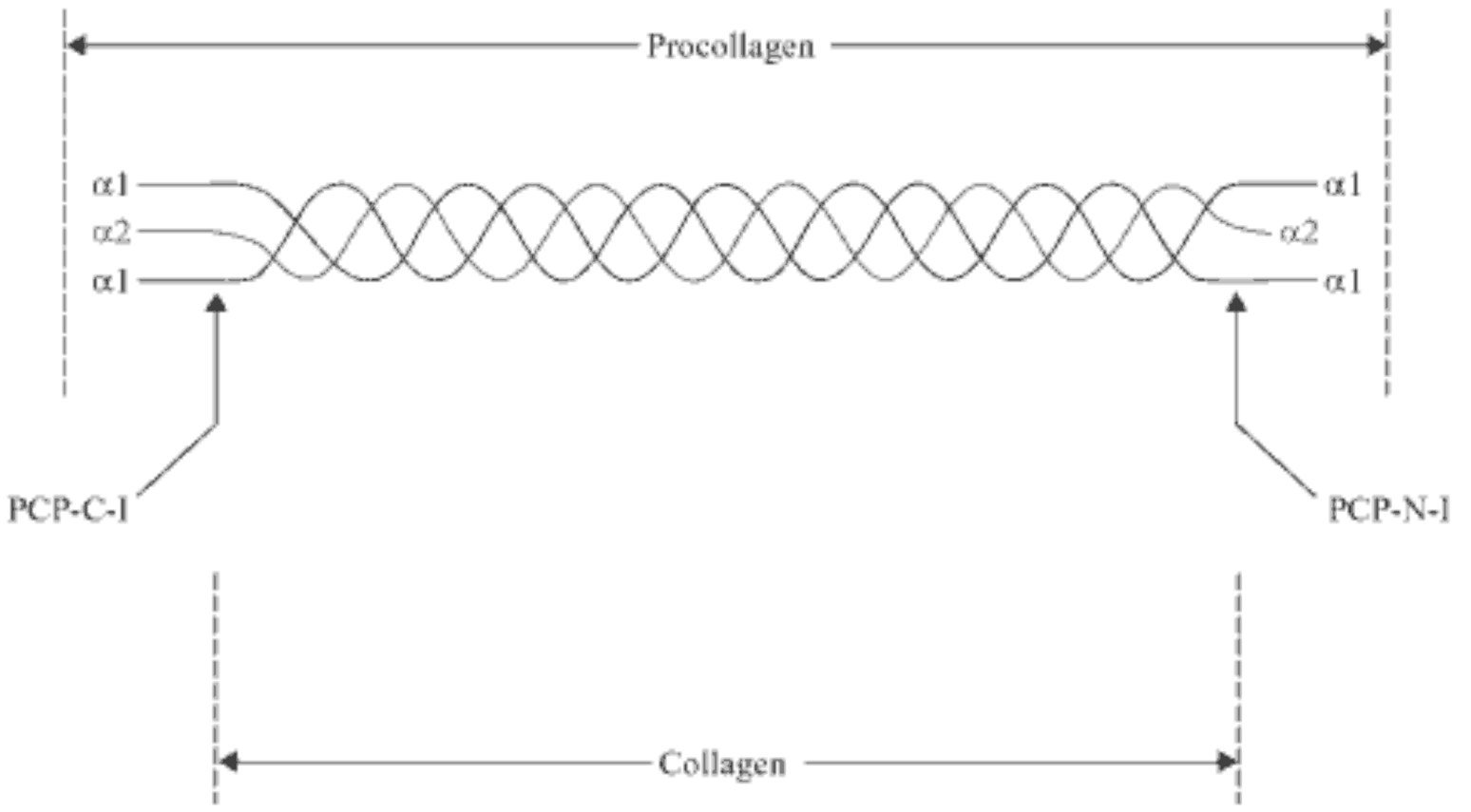


Fig. 3.2 A procollagen type-I molecule, showing the cleavage sites of the endopeptidases PCP-N-I and PCP-C-I.

alters the procollagen molecule so that the cleavage site for the enzyme PCP-N-I is suppressed. In heterozygotes for such a mutation, one-half of all type-I collagen molecules are only partially processed; despite both enzymes being present at normal concentration, only one of them (PCP-C-I) can actually cleave the procollagen. Since heterozygotes for this mutation have only 50 per cent normal type-I collagen molecules, they show the clinical signs of the disorder. In other words, this form of the disorder is dominant.

If a specific set of clinical signs arises from more than one mutation, there is **genetic heterogeneity** for that disorder.

From the somewhat limited evidence available, it appears that the dominant form of the disorder occurs in horses, cats, dogs, mink, and rabbits. The recessive form definitely occurs in sheep and cattle.

This is an example of genetic heterogeneity *between* species. Although such heterogeneity can cause confusion from time to time, a far greater source of confusion is genetic heterogeneity *within* a species. For example, both forms of the above connective tissue disorder are well documented in humans, and it would be most surprising if both forms of the disorder do not occur within particular species of animals as well.

Genetic heterogeneity of disease is not confined to situations where more than one gene is involved; it also exists if there are two or more different mutations in the one gene. In some human genes, hundreds of different mutations have been detected. Not surprisingly, some of the mutations have a greater clinical effect than others; on average, for example, nonsense mutations and deletions in coding sequences produce more serious clinical signs than mis-sense mutations. In some cases, different mutations within the one gene give rise to different clinical entities. In animals, there is little comparable information. However, in a few cases, there is

Copyrighted material

evidence that the same clinical signs in different populations within a species are the result of different mutations. In broiler chickens, for example, sex-linked dwarfism in several populations is caused by a deletion in the growth-hormone receptor gene, but in the Cornell sex-linked dwarfism strain, the same disorder is the result of a mis-sense mutation in the same gene.

The above discussion has highlighted two important points, namely that mutations in different genes can give rise to exactly the same clinical entity, and different mutations within the same gene can produce different or similar clinical entities. All of these situations represent genetic heterogeneity.

If a set of clinical signs appears to have a genetic basis, but if genetic heterogeneity within a species remains undetected, it may be difficult or even impossible to establish the form of inheritance. Although it is easy to be wise after the event, it is very difficult in practice to ensure that the available data on any syndrome all belong to the one disorder entity and hence are homogeneous. Data are more likely to represent genetic homogeneity if they have been obtained from animals that not only present with the same clinical signs but which also have the same histopathological and biochemical lesions. In addition, data obtained from animals that are related (have at least one ancestor in common) are more likely to be homogeneous than data collected from several unrelated families.

Type of gene action and type of disease

In the connective tissue disorders described above, two types of gene action have been observed. One disorder is inherited as an autosomal recessive condition and is associated with an enzyme deficiency. The other is inherited as an autosomal dominant condition and is the result of a mutation in one of the genes for type-I collagen.

In general, recessive disorders tend to be caused by enzyme deficiencies, while dominant or co-dominant disorders are more likely to be caused by defects in non-enzymatic polypeptides. The reason for this is that enzymes are required in such small quantities that 50 per cent of normal activity (as in heterozygotes) is sufficient for normal functioning. In contrast, if the mutant polypeptide has, for example, a structural role, the structures incorporating that polypeptide will be defective, and the heterozygote may well show some form of the disorder. Similarly, if the polypeptide is a substrate in a particular process or is involved in transport, a decrease by one-half in the production of normal polypeptide, as occurs in heterozygotes, might be expected to cause some clinical signs in the heterozygote, because such polypeptides are also often required in relatively large quantities.

A special comment must be made concerning X-linked diseases, because random X-inactivation has important practical implications. In females that are heterozygous for X-linked disorders, one-half of their cells are expected to express the mutant allele. If the gene product normally functions outside the cell, e.g. in the blood or some other body fluid, the result is the same as with heterozygotes for autosomal

Copyrighted material

disorders: the gene product is at one-half the normal concentration. However, if the gene product's normal role is inside the cell, all of those cells expressing only the mutant allele are deficient in the gene product and the animal may show some clinical signs. In addition, because inactivation is a random process, not all females have exactly one-half of each type of cell. In fact, a wide range of proportions of the two cell types can occur, from females having most cells with the normal allele active, to females having most cells with the mutant allele active. Females in the former category may be indistinguishable from females that are homozygous for the normal allele, in terms of concentration of gene product and/or lack of clinical signs. In contrast, those with a high proportion of cells expressing the mutant allele may actually exhibit quite severe clinical signs, resembling males with a mutant X chromosome. Finally, even in cases where the mutant allele is active in only a small proportion of cells, if these cells happen to be the only ones in which the X-linked gene is transcribed and translated, the female may show clinical signs. Obviously, X-inactivation can create substantial complications in inheritance patterns for X-linked disorders, and for biochemical detection of carriers of such disorders.

Phenocopies

Sometimes the phenotype of a single-gene disorder has an environmental cause. Such an occurrence is called a **phenocopy**. A good example of a phenocopy is provided by α-mannosidosis, which is a lysosomal storage disease in cats and cattle.

Lysosomes are small membrane-bound organelles found in the cytoplasm. They are the digestive system of the cell: they contain many enzymes that act in a stepwise manner to break down complex molecules into monomeric units of simple lipids, amino acids, monosaccharides, and nucleotides. If a particular lysosomal enzyme is absent or inactive, the stepwise degradation is halted, with a consequent build-up ('storage') of the material that is normally broken down by that enzyme. In other words, an inborn error of lysosomal catabolism produces a **lysosomal storage disease**. For most lysosomal storage diseases, affected animals are usually normal at birth but fail to grow as rapidly as their contemporaries; in most cases there is a progressive neurological deterioration, caused by the build-up of storage products in nerve cells, producing incoordination and aggression which eventually leads to death, usually before sexual maturity.

In α-mannosidosis, the deficient enzyme is α-mannosidase. In most cases, the deficiency is caused by a mutation in the gene for α-mannosidase, and α-mannosidosis is therefore a single-gene recessive disorder. However, if cattle graze on pasture that contains the legume Darling Pea (*Swainsona* spp.), they often develop α-mannosidosis. The reason for this is that plants in the *Swainsona* genus produce a trihydroxylated indolizidine alkaloid called swainsonine, which inhibits α-mannosidase by binding to it, thereby producing a build-up of exactly the same mannose-rich oligosaccharides that are characteristic of the inherited form of the disorder. A similar compound is produced by plants in the *Astragalus* and *Oxytropis*

Copyrighted material

genera. Indeed, plants within these genera are called locoweed, because of the clinical signs they induce. Not surprisingly, these clinical signs can be induced in any animal species that grazes on the plants.

By their very nature, it is difficult to distinguish phenocopies from inherited forms of a disorder. All we can do is be aware of the possibility that an individual showing the clinical signs of an inherited disorder may not, in fact, have the genotype usually associated with those clinical signs. Phenocopies, therefore, are another potential cause of complications in inheritance patterns.

A sample of single-gene disorders

The above discussion has covered the important principles concerning single-gene disorders, using specific examples from domestic animals. Obviously there are many more single-gene disorders than those mentioned above. In fact, the potential total number of single-gene disorders is far greater than the total number of genes: a deficiency in almost any polypeptide is likely to give rise to either embryonic lethality (spontaneous abortion) or clinical signs, and (as we have already seen) different mutations in the same gene can give rise to different sets of clinical signs. Of course, mutations producing embryonic lethals are difficult to detect. However, recalling that the total number of genes is probably around 22000, it is evident that the ultimate list of single-gene disorders that will be detected is likely to be very long! At present, only a small fraction of all potential single-gene disorders have been documented in domestic animals.

The increasing use of molecular technology in the investigation of single-gene disorders in animals has already uncovered several interesting examples of different types of mutations.

An example of a mis-sense mutation is provided by canine haemophilia B. This disorder is caused by a substitution of A for G at nucleotide 1477 in the gene for canine factor IX, resulting in the substitution of glutamic acid for glycine at position 379 in the factor-IX molecule (Fig. 3.3). This particular site has been highly conserved throughout evolution – there is a glycine present at this position in factor IX from humans, pigs, and cattle. Not surprisingly, therefore, this single amino-acid substitution profoundly alters the tertiary structure of the factor-IX molecule, to the extent that no functional factor IX can be detected. Mis-sense mutations are also the cause of bovine leucocyte adhesion deficiency (BLAD), periodic paralysis II in horses, malignant hyperthermia syndrome in pigs, and a behavioural defect called tremor, X-linked, or shaking in dogs (Appendix 3.1).

Nonsense mutations have also occurred. In Afrikander cattle, for example, inherited goitre is the result of a nonsense mutation in the thyroglobulin gene. The gene itself is approximately 250 kb long, while the mRNA consists of 8431 bases. The disease is the result of a C → T substitution in exon 9, changing a CGA triplet (arginine) to TGA (stop). In Dutch goats, a nonsense mutation in exon 8 of the thyroglobulin gene, changing TAC (tyrosine) to TAG (stop), causes the same disorder. Other disorders caused by nonsense mutations are citrullinaemia in cattle

Copyrighted material

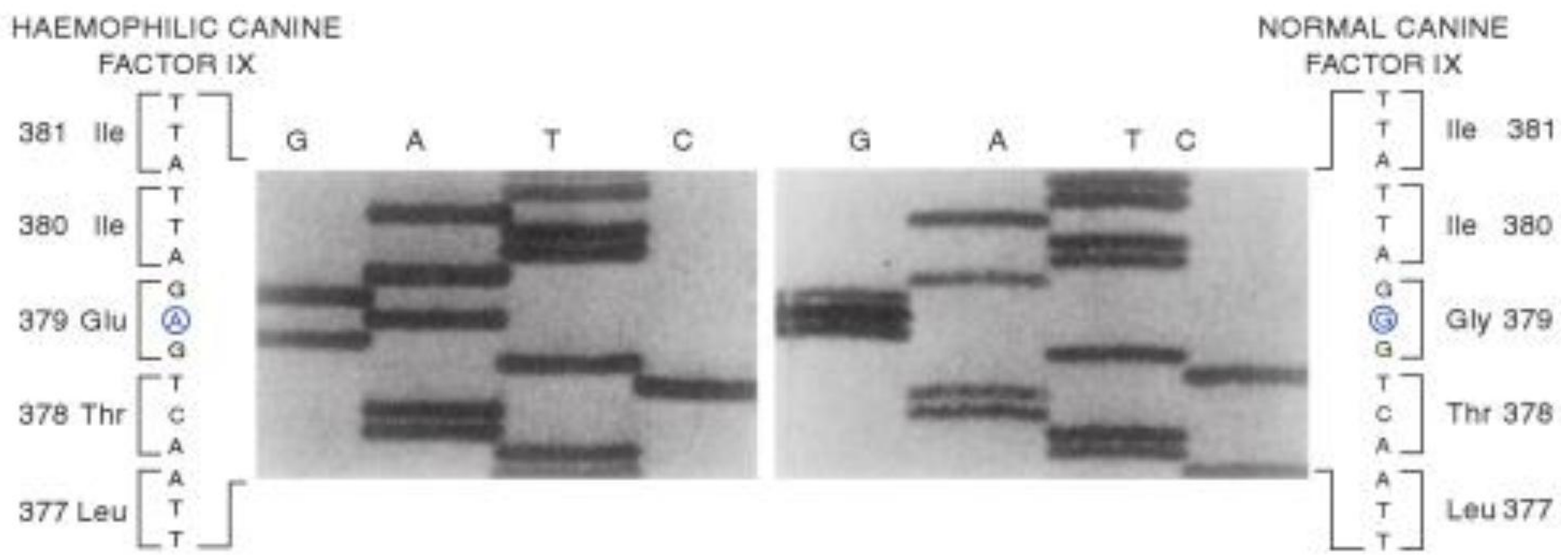


Fig. 3.3 A small section of sequencing gels (see Fig. 2.4) showing the substitution of A for G in the gene for factor IX in dogs. The consequent substitution of glutamic acid for glycine results in haemophilia B. When checking the sequences against the genetic code in Table 1.2, note that in this figure they are read from bottom to top. (Reproduced courtesy of J.P. Evans from Evans *et al.* (1989) *Proceedings of the National Academy of Sciences*, USA, 86, 10095–99.)

(mentioned earlier), deficiency of uridine monophosphate synthase (DUMPS) in cattle, hypotrophic axonopathy in quail, maple syrup urine disease in cattle, nanomelia in chickens, nephritis, X-linked in dogs, and PRA rod–cone dystrophy type 1 in dogs (Appendix 3.1).

DUMPS is particularly interesting, because it is an example of an embryonic lethal that has been detected. The enzyme that is deficient in this disorder, namely uridine monophosphate (UMP) synthase, is responsible for converting orotic acid to UMP, which is an essential component of pyrimidine nucleotides. Because nucleotides are required in such vast quantities during embryonic growth, it is not surprising that homozygosity for the nonsense mutation in the UMP synthase gene results in embryonic death around 40 days *in utero*. The practical effect of this disorder is that carrier cows show a higher rate of return to service, because some of their pregnancies end in early natural abortion. Given that returns to service can have so many different causes, it would not have been possible to identify the mutation from reproductive records. How, then, was this embryonic lethal mutation detected initially? The answer is: by chance! As part of a nutrition study at the University of Illinois, the level of orotic acid was determined in the milk of cows in the university herd. Some cows had exceptionally high levels of this acid, and one possible explanation was that they were deficient in UMP synthase. Subsequent biochemical tests showed that these cows had only 50 per cent of the normal activity of this enzyme. Inheritance and molecular studies finally brought the whole story to light.

All of the above point mutations occur in exons (more specifically, in coding sequences within exons). In contrast, it might seem that point mutations occurring in introns do no damage because introns are excised before translation. However, if a point mutation occurs in the first few bases or in the last few bases of an intron, it is very likely to disrupt the RNA-splicing mechanism. Muscular dystrophy in Golden Retriever dogs, for example, is caused by a base substitution in the 3′ splice-acceptor site of intron 6 of the dystrophin gene (Fig. 3.4). This changes the sequence

Copyrighted material

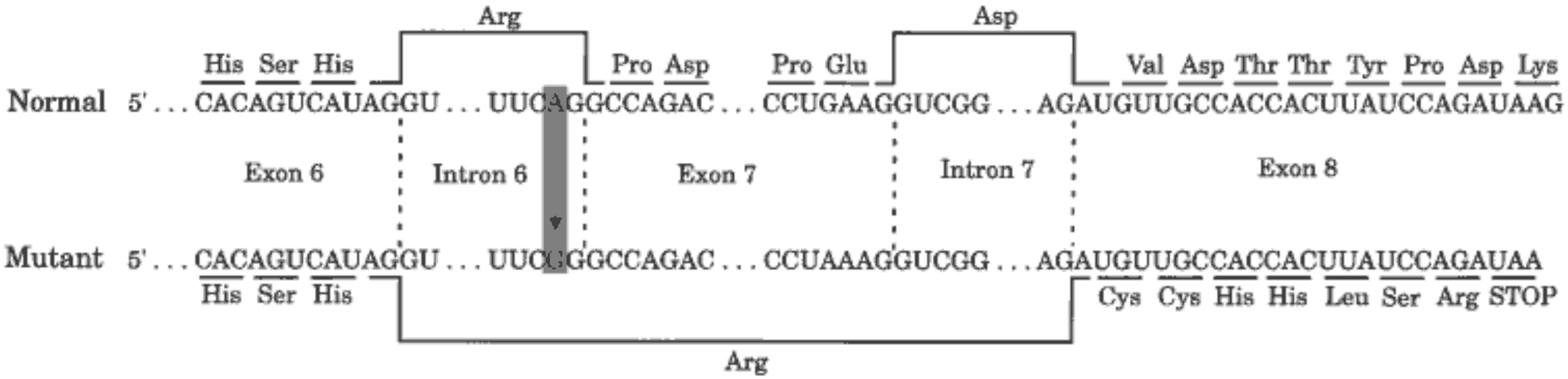


Fig. 3.4 Part of the base sequence of the normal and mutant dystrophin mRNA transcripts in Golden Retrievers, showing the A → G mutation in the 3′ splice-acceptor site of intron 6. The consequent excision of exon 7 from the mRNA creates a frameshift mutation, which produces a stop codon in exon 8. Codons in each mRNA are indicated by lines spanning three adjacent bases. (Drawn from *Genomics*, Vol **13**(1), pp. 115–21., Sharp *et al.* (1992) with permission from Elsevier.)

Copyrighted material

in the mRNA transcript from the usual AG (see Chapter 1) to GG, which means that the following exon 7 is not recognized as such, and is therefore excised along with intron 6 and intron 7 (Fig. 3.4). The dystrophin gene is one of the largest known; it is approximately 2000 kb long, and its mature mRNA is 14 kb long. The mutation in Golden Retrievers occurs at base 736 of the mRNA, i.e. very early in the gene. Exon 7 actually starts at the third base of a codon, and finishes 119 bases later, at the first base of a codon (Fig. 3.4). In effect, therefore, the deletion of exon 7 introduces a frameshift mutation, so that the remainder of the mRNA transcript contains a garbled message. In this particular case, the frameshift introduces a stop codon in the following exon (Fig. 3.4). The result is a truncated transcript, and no functional polypeptide.

An example of a disorder caused by a mutation at the other end of an intron is riboflavinuria in chickens. In this case, the 5′ splice-donor sequence GT (see Chapter 1) in intron 2 of the gene for riboflavin-binding protein becomes AT, resulting in exon 2 being spliced out of the mRNA transcript. In normal chickens, riboflavin-binding protein is synthesized in the liver, and is released into the bloodstream, where it binds riboflavin derived from the diet. Riboflavin-binding protein transports its bound riboflavin from the serum to the yolk of the developing egg, and from the oviduct to the egg white. The deletion of exon 2 causes a deficiency of riboflavin-binding protein, which in turn leads to a deficiency of riboflavin in the developing embryo. This causes embryonic death around day 13, because riboflavin is essential for embryonic development.

Yet another type of mutation is illustrated by henny feathering, which was described earlier in this chapter. This disorder is an example of an insertion mutation. In this particular case, it appears that the terminal repeat sequence (described in Chapter 1) of a retrovirus has been inserted into the 5′ promoter region of the aromatase gene. This terminal repeat has a promoter of its own, which causes the aromatase gene to be switched on in atypical places, such as the skin of both sexes, giving rise to the henny-feathering trait in males. The actual peptide produced by the mutant allele is exactly the same as that produced by the normal allele, as we would expect for such a mutation.

Another source of single-gene disorders has recently come to light in humans, and is likely to be discovered in animals in the future. Known as **unstable trinucleotide repeats**, they arise in some genes that contain microsatellites (see Chapter 2) in which the repeat unit is a triplet. Small numbers of the repeat unit appear to be a normal part of the gene, and appear to do no harm, being usually located in an untranslated region. However, for reasons that are not yet understood, these repeats can sometimes rapidly expand in number between generations, and as the number of repeat units increases, clinical signs begin to appear, associated with a deficiency of the polypeptide product of the gene. Apparently the expanded number of repeats results in the surrounding DNA becoming methylated, which inactivates the gene (see Chapter 1). As the number of repeat units continues to increase during subsequent generations, the clinical signs become increasingly severe. This increase in the severity of the disease over generations is called **genetic anticipation**. Many such diseases occur in humans. The first report of an unstable repeat in animals was

Copyrighted material

a form of epilepsy in dogs (Myoclonus epilepsy of Lafora; see Appendix 3.1), associated with 19 to 26 copies of a 12 bp sequence, compared with only two to three copies in healthy dogs.

A revolution in mapping and identifying the causal mutation of single-gene disorders

The ability to perform genome scans with SNP chips is creating a revolution in mapping single-gene disorders. Consider autosomal-recessive disorders – the most common type of single-locus disorder. When a causative mutation first occurs, it is located in a chromosome with whatever alleles happened to be located at all the other loci on that mutated chromosome. When this chromosome passes to the next generation, a cross-over will occur somewhere along its length, breaking the original mutated chromosome into two segments: one containing the causative mutation and its original neighbouring alleles at surrounding loci; and a second fragment containing the rest of the chromosome comprising alleles from the homologous non-mutated chromosome. Over subsequent generations, recombination will gradually reduce the size of the segment containing the causal mutation, and will create a range of segments in which a copy of the causal mutation is surrounded by different lengths of copies of the original neighbouring alleles. When two copies of the original mutated segment occur together in an embryo, that embryo is homozygous for the causal mutation, and we have our first case of the disorder. Notice, however, that this affected animal will also be homozygous at neighbouring loci on either side of the causal mutation. The extent to which this homozygosity extends either side of the causal mutation will most likely vary in the two chromosomes of the affected animal, reflecting the random nature of recombination in previous generations. If, as is typical, the causal mutation has occurred only once, it is evident that all affected animals will be homozygous for the smallest set of neighbouring alleles that have 'survived' recombination. In contrast, animals heterozygous for the causal mutation, or homozygous for the normal allele, will not be homozygous at surrounding loci.

The key to mapping the causal mutation should now be evident: 'affecteds' and 'normals' (controls) are genotyped at the tens of thousands or hundreds of thousands of loci on the SNP chip, covering all regions of all chromosomes. It is then a matter of looking for a region of a chromosome in which all 'affecteds' are homozygous for overlapping sets of SNPs, but 'normals' are not homozygous. As the disorder is the result of a single causal mutation, there will be only one region in the entire genome where this condition is fulfilled and that region has to be the region containing the causal mutation. Experience in cattle has shown that by genotyping as few as 10–20 affected individuals and a similar number of normal cattle, it is possible to map a causal mutation to a region varying in size from 1 Mb to 10 Mb, which is likely to contain roughly 10 to 100 genes, respectively. If the species has an annotated genome sequence assembly, it is then a matter of searching through the list of genes located in that region, looking for a likely candidate, i.e. a gene whose product,

Copyrighted material

when deficient, is likely to produce the clinical signs of the disorder. If the species does not have an annotated sequence assembly, comparative mapping can readily provide a list of so-called comparative positional candidates. The final step then involves sequencing the most likely candidate genes in affected animals and controls, looking for the actual causal mutation. As it is most likely that the causal mutation will be located within the coding sequence, the total sequencing effort can be substantially reduced by collecting mRNA from relevant tissue, and sequencing cDNA of the candidate gene.

In an early illustration of this strategy, Charlier *et al.* (2008) mapped each of five autosomal-recessive disorders in cattle to a region ranging in size from 0.87 Mb to 11.78 Mb, and readily identified the causal mutation in three of the five disorders. A similar effort involving around 10 'affecteds' and 10 'normals' for each of two single-gene disorders in dogs (Karlsson *et al.*, 2007), using a somewhat different strategy, was successful in mapping the two disorders to regions of less than 1 Mb; and subsequently identifying the causal mutation in one.

With so few affected and control animals being required to map accurately a single-gene disorder, there is likely to be an explosion of knowledge concerning the causal mutations of such disorders. As we shall see in Chapter 21, this will be of great benefit for the control of inherited disorders.

Summary

- If a polypeptide acts as an enzyme or is part of an enzyme, a mutation in its gene sometimes results in a deficiency of that enzyme, with a consequent blockage in the biochemical pathway where that enzyme is required
- Diseases that result from such blockages are called inborn errors of metabolism
- Two alleles at one locus can exhibit more than one type of gene action, depending on the trait being considered
- Genetic heterogeneity of disease occurs because mutations in different genes can give rise to exactly the same clinical entity, and different mutations within the same gene can produce different clinical entities
- Recessive disorders tend to result from enzyme deficiencies, whereas dominant or co-dominant disorders are more likely to be caused by defects in non-enzymatic polypeptides
- Sometimes an environmental cause produces the same phenotype as a single-gene disorder; such an occurrence is called a phenocopy
- With SNP chips comprising tens or even hundreds of thousands of DNA markers now available, in many situations it is possible to map and identify the causal mutation of a single-locus disorder with as few as 10–20 affected animals and the same number of normals (controls)

Copyrighted material

Further reading

Castelhano, M.G., Acland, G.M., Ciccone, P.A., Corey, E.E., Mezey, J.G., Schimenti, J.C. and Todhunter, R.J. (2009) Development and use of DNA archives at veterinary teaching hospitals to investigate the genetic basis of disease in dogs. *Journal of the American Veterinary Medical Association*, **234**, 75–80.

Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B.R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R.J. and Georges, M. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics*, **40**, 449–54.

Gough, A. and Thomas, A. (2009) *Breed Dispositions to Disease in Dogs and Cats*, 2nd edn. Blackwell Science, Oxford.

Ibeagha-Awemu, E.M., Kgwatalala, P., Ibeagha, A.E. and Xin Zhao, X. (2008) A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mammalian Genome*, **19**, 226–45.

Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon Hillbertz, N.H., Zody, M.C., Anderson, N., Biagi, T.M., Patterson, N., Pielberg, G.R., Kulbokas, E.J. 3rd, Comstock, K.E., Keller, E.T., Mesirov, J.P., von Euler, H., Kampe, O., Hedhammar, A., Lander, E.S., Andersson, G., Andersson, L. and Lindblad-Toh, K. (2007) Efficient mapping of Mendelian traits in dogs through genome-wide association. *Nature Genetics*, **39**, 1321–8.

Karlsson, E.K. and Lindblad-Toh, K. (2008) Leader of the pack: gene mapping in dogs and other model organisms. *Nature Reviews Genetics*, **9**, 713–25.

Sargan, D.R. (2001–present) *Inherited diseases in dogs*, http://www.vet.cam.ac.uk/idid.

Appendix 3.1 A sample of single-gene traits that have been characterized at the molecular level

A continually updated and expanded version of this table (with references), called Online Mendelian Inheritance in Animals (OMIA), can be accessed on the Internet through the Australian National Genomic Information Service (ANGIS) at the University of Sydney. The URL is http://omia.angis.org.au.

Name of trait	MIA number	Species
Achromatopsia-3	001365	dog
Adrenal hyperplasia, congenital	000017	rabbit
Alopoecia, colour mutant	000031	dog
Analphalipoproteinaemia	000042	chicken
Anhidrotic ectodermal dysplasia	000543	cow, dog
Atrichia with papular lesions	001348	rhesus monkey
Beta-lactoglobulin, aberrant low expression	001437	cow
Blood group system AB	000119	cat
Blood group system ABO	001089	pig
C3 deficiency	000155	dog

Continued

Copyrighted material

Name of trait	MIA number	Species
C8 deficiency	000156	rabbit
Cardiomyopathy, dilated	000162	turkey, golden hamster
Cardiomyopathy, hypertrophic	000515	cat, golden hamster
Cardiomyopathy and woolly haircoat syndrome	000161	cow
Casein, beta, null allele	001424	goat
Chediak–Higashi syndrome	000185	cow
Chondrodysplasia	000187	cow, sheep
Chronic interstitial nephritis with diffuse zonal fibrosis	001332	cow
Citrullinaemia	000194	cow
Coat colour, agouti	000201	horse, Norway rat, red fox, pig, sheep, cat
Coat colour, albinism	000202	cat, cow, Norway rat, Japanese medaka, *Astyanax*, American mink
Coat colour, albinism, oculocutaneous type I, temperature-sensitive	000205	cat
Coat colour, brown	001249	cow, horse, dog, cat, sheep
Coat colour, champagne	001263	horse
Coat colour, cream dilution	001344	horse
Coat colour, dilute	000206	cat
Coat colour, dominant black	001416	dog
Coat colour, dominant white	000209	pig, horse
Coat colour, extension	001199	cat, cow, dog, horse, red fox, pig, rabbit, sheep, jaguar, jaguarundi, American black bear, woolly mammoth
Coat colour, golden	001433	zebrafish
Coat colour, grey	001356	horse
Coat colour, merle	000211	dog
Coat colour, roan	001216	cow
Coat colour, silver	001438	horse, zebrafish
Coat colour, white spotting	000214	dog
Collie eye anomaly	000218	dog
Complex vertebral malformation	001340	cow
Cone–rod dystrophy 1	001432	dog
Cone–rod dystrophy 2	001455	dog
Congenital muscular dystonia 1	001450	cow
Congenital muscular dystonia 2	001451	cow
Cystinuria	000256	dog
Deficiency of uridine monophosphate synthase (DUMPS)	000262	cow
Degenerative myelopathy	000263	dog
Dermoid sinus	000272	dog
Dwarfism, Dexter	001271	cow
Dwarfism, growth-hormone deficiency	001473	cow
Dwarfism, sex-linked	000309	chicken
Ectodermal dysplasia	000323	dog
Ehlers–Danlos syndrome	000327	cow, horse
Ehlers–Danlos syndrome, type VII	000328	cow
Elliptocytosis	001318	dog

Continued

Copyrighted material

Name of trait	MIA number	Species
Epidermolysis bullosa	000340	cow
Epidermolysis bullosa, dystrophic	000341	dog
Epidermolysis bullosa, junctionalis	000342	dog, horse
Exercise–induced collapse	001466	dog
Factor VII deficiency	000361	dog
Factor XI deficiency	000363	cow, dog
Feather colour, albinism	000369	chicken
Feather colour, albinism, sex-linked, imperfect	000370	chicken, Japanese quail
Feather colour, dominant white	000373	chicken
Feather colour, extended black	000374	chicken, red-footed booby
Feather colour, extended brown	001435	Japanese quail
Feather colour, lavender	001445	chicken
Feather colour, panda/dotted white	000375	Japanese quail
Feather colour, recessive white	001252	chicken
Feather colour, roux	001322	Japanese quail
Feathering, Z-linked	000380	chicken
Fecundity, Booroola	000383	sheep
Fecundity, Cambridge	000384	sheep
Fecundity, Hanna	001350	sheep
Fecundity, Inverdale	000386	sheep
Fecundity, small tail Han	001423	sheep
Fucosidosis, alpha	000396	dog
Gangliosidosis, GM1	000402	cat, dog
Gangliosidosis, GM2	000403	cat
Gangliosidosis, GM2, GM2A deficiency	001427	cat
Glycogen storage disease I	000418	dog
Glycogen storage disease II	000419	cow
Glycogen storage disease IV	000420	cat, horse
Glycogen storage disease V	001139	cow, sheep
Glycogen storage disease VII	000421	dog
Goitre, familial	000424	cow, goat
Haemophilia A	000437	cow, dog
Haemophilia B	000438	cat, dog
Hair length	000439	cat
Henny feathering	000452	chicken
Horns	000483	goat
Hypercholesterolaemia	000499	pig
Hyperkeratosis, epidermolytic	001415	dog
Hyperlipoproteinaemia	001210	cat
Hypocatalasia	001138	dog
Hypothyroidism	000536	dog
Hypotrichosis	000540	sheep
Hypotrophic axonopathy	000545	Japanese quail
Ichthyosis congenita	000547	cow
Krabbe disease	000578	dog, rhesus monkey
Lethal trait A46	000593	cow
Leucodystrophy	001130	dog
Leucocyte adhesion deficiency (LAD)	000595	cow (BLAD), dog (CLAD)
Macrostomus	000618	rhesus monkey
Malignant hyperthermia	000621	dog, horse, pig
Mannosidosis, alpha	000625	cow, cat, domestic guinea pig

Continued

Copyrighted material

Name of trait	MIA number	Species
Mannosidosis, beta	000626	cow, goat
Maple syrup urine disease	000627	cow
Marfan syndrome	000628	cow
Meat quality	001085	pig
Megacolon	000629	horse
Melanoma, Sinclair swine cutaneous malignant	000633	pig
Membranoproliferative glomerulonephritis type II	000636	pig
Metabolizer of a cognitive enhancer	001405	dog
Mucolipidosis II	001248	cat
Mucopolysaccharidosis I	000664	dog, cat
Mucopolysaccharidosis IIIA	001309	dog
Mucopolysaccharidosis IIIB	001342	cow, emu, dog
Mucopolysaccharidosis IIID	000665	goat
Mucopolysaccharidosis VI	000666	cat, dog
Mucopolysaccharidosis VII	000667	cat, dog
Multidrug resistance 1	001402	dog
Muscular dystrophy, Duchenne and Becker types	001081	cat, dog
Muscular hypertrophy	000683	cow
Muscular hypertrophy, Callipyge	001354	sheep
Muscular hypertrophy, Texel	001426	sheep
Myasthenic syndrome, congenital	000685	dog, cow
Myoclonus	000689	cow
Myoclonus epilepsy of Lafora	000690	dog
Myopathy of the diaphragmatic muscles	001319	cow
Myotonia	000698	goat, dog
Nanomelia	000702	chicken
Narcolepsy	000703	dog
Neonatal encephalopathy with seizures	001471	dog
Nephritis, X-linked	001112	dog
Neuronal ceroid lipofuscinosis, 2	001472	dog
Neuronal ceroid lipofuscinosis, 6	001443	sheep
Neuronal ceroid lipofuscinosis	000181	cow, dog, sheep
Neutropenia, cyclic	000248	dog
Niemann–Pick disease, type C	000725	cat
Non-shivering thermiogenesis, absence of	001436	pig
Ocular defects	000733	cow
Osteogenesis imperfecta	000754	dog
Osteopetrosis	000755	Japanese quail
Oxalosis II	000821	cat
Pea comb	000782	chicken
Periodic paralysis II	000785	horse
Persistent Mullerian duct syndrome	000791	dog
Polycystic kidney disease	000807	cat
Polydactyly	000810	chicken
Porphyria cutanea tarda	001176	sheep
Progressive retinal atrophy (PRA) autosomal dominant	001346	dog
PRA rod–cone dystrophy type 1	000882	dog

Continued

Copyrighted material

Name of trait	MIA number	Species
PRA rod–cone dystrophy type 2	001260	dog
PRA rod–cone dystrophy type 3	001314	dog
Protoporphyria	000836	cow
Pseudomytonia, congenital	001464	cow
Pyruvate kinase deficiency of erythrocyte	000844	dog, cat
Reduced casein concentration	000852	goat
Renal cystadenocarcinoma and nodular dermatofibrosis	001335	dog
Renal dysplasia	001135	cow
Resistance to avian sarcoma and leucosis viruses, subgroup B	001302	chicken
Resistance to oedema disease	000862	pig
Restricted ovulator	000865	chicken
Retinal degeneration I	001245	chicken
Retinal degeneration II	001244	cat
Retinal pigment epithelial dystrophy	001222	dog
Retinitis pigmentasa	001413	dog
Retinopathy globe enlarged	001368	chicken
Retinopathy, multifocal	001444	dog
Rhabdomyolysis	001158	horse
Riboflavinuria	000876	chicken
Ribosomal DNA deficiency	001101	chicken
Rod–cone degeneration, progressive	001298	dog
Sensony ataxic neuropathy	001467	dog
Severe combined immunodeficiency disease, autosomal	000220	horse, dog
Severe combined immunodeficiency disease, X-linked	000899	dog
Sex reversal: XY female	001230	horse, cow
Skin colour, yellow	001449	chicken
Sperm, short tail	001334	pig
Spherocytosis	001228	cow, zebrafish
Spinal muscular atrophy	000939	cat, cow
Spongiform encephalopathy	000944	sheep
Syndactyly	000963	cow
Tail, short	000975	dog
Talpid-3	000979	chicken
Thrombasthenia	001000	dog
Thrombopathia	001003	dog
Tremor, X-linked (shaking)	000770	rabbit, dog
Trimethylaminuria	001360	cow, chicken
Vitamin D-deficiency rickets, type I	000837	pig
Von Willebrand disease I	001057	dog
Von Willebrand disease II	001339	dog
Von Willebrand disease III	001058	dog
Waardenburg syndrome, type IIA	001401	golden hamster
Wilson disease	001071	dog

Copyrighted material

4 Chromosomal aberrations

From time to time, mistakes occur in mitosis, in meiosis, or in fertilization, and sometimes these mistakes give rise to aberrant karyotypes. The aim of this chapter is to describe these aberrations, and to discuss the role they play in determining abnormalities in animals.

In many cases, abnormal karyotypes have been observed in healthy and/or highly productive animals. Likewise, many diseased and/or unproductive animals have normal karyotypes. It follows that a normal karyotype is not itself a guarantee of high production and freedom from disease, and neither is an abnormal karyotype certain to indicate a diseased or unproductive animal.

Abnormal chromosome number

One of the most common abnormalities is the absence of an X chromosome. The lack of just one chromosome is termed **monosomy**, and animals lacking a chromosome are said to be **monosomic** for that chromosome. In the case of females that are monosomic for the X chromosome, their karyotype is written as XO, with the O indicating absence of an X chromosome.

XO individuals have been reported in many domestic species, and also in humans where the condition is called Turner's syndrome. XO individuals have a more-or-less normal female external phenotype but are usually sterile.

Another abnormality is the presence of three X chromosomes rather than the normal two. This is an example of **trisomy**, and individuals having an additional chromosome are said to be **trisomic**. In the case of the X chromosome, their karyotype is written as XXX. When examined cytologically, non-dividing cells of X trisomics are seen to contain two Barr bodies, indicating that only one X chromosome in each cell remains active. Consequently, XXX individuals are less abnormal than XO individuals; they are often fertile. Some of their offspring have normal karyotypes, while others are trisomic.

XO and XXX individuals arise from failure of the chromosomes or chromatids to disjoin during meiosis, which is called **non-disjunction**. Since there is normally one disjunction in each stage of meiosis (one during meiosis I and one during meiosis II), as described in Chapter 1, there are two opportunities for non-disjunction

Copyrighted material

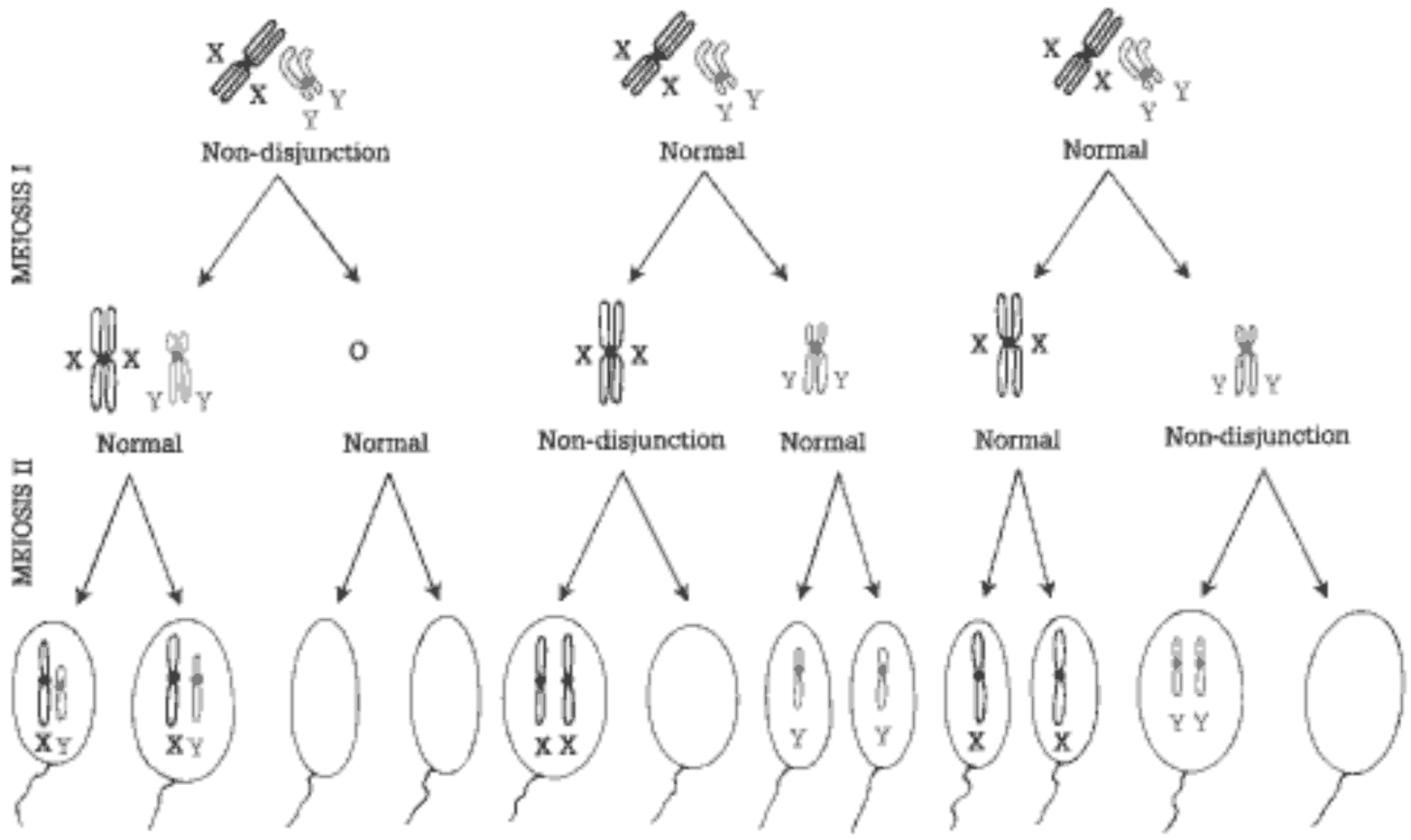


Fig. 4.1 Non-disjunction of sex chromosomes during meiosis in an XY male. 'Normal' refers to normal disjunction, as illustrated in Fig. 1.5.

during the formation of a germ cell. Non-disjunction occurs in approximately 5 per cent of all meioses, and can involve any of the autosomes or the sex chromosomes. In the latter case, the results differ depending on the sex of the individual in which meiosis is occurring. In females, the results are straightforward: irrespective of when non-disjunction occurs, some of the ova contain two X chromosomes while others contain none. With males, however, four different types of unbalanced sperms can result, depending on whether non-disjunction occurs during meiosis I or meiosis II (Fig. 4.1).

The unbalanced gametes that result from non-disjunction are usually still able to be involved in fertilization, and the possible outcomes are illustrated in Table 4.1. It is evident that XO and XXX individuals can arise from non-disjunction in either sex. It is also evident that many other types of unbalanced karyotypes can result from non-disjunction in one or both parents. In general, karyotypes with a small number of extra chromosomes, and karyotypes that lack a small number of chromosomes, are said to be **aneuploid**, in comparison with those having the normal number, which are said to be **euploid**. Thus monosomics and trisomics are both aneuploids. Some of the other aneuploids shown in Table 4.1 such as O, YO, and YYO, have never been observed, and are assumed to be lethal at an early stage of embryonic development. All the remaining aneuploids in Table 4.1 have been recorded in at least one species of mammal. Among those reported in domestic animals are XXY and XXXY, both of which have a predominantly male phenotype, often associated with underdeveloped male sexual behaviour. Among domestic animals, XXY individuals have been reported in cats, cattle, dogs, horses, pigs, and sheep (Appendix 4.1). They are usually sterile. In humans, the condition is known as Klinefelter's syndrome.

Copyrighted material

Table 4.1 The zygotes that can theoretically result from all possible combinations of abnormal sperms and eggs, in relation to sex chromosomes (colour indicates chromosomes from sperms)

		Sperm					
		Normal		Abnormal			
		X	Y	XY	XX	YY	O
Eggs	Normal X	XX	XY	XXY	XXX	XYY	XO
	Abnormal XX	XXX	XXY	XXXY	XXXX	XXYY	XX
	Abnormal O	XO	YO	XY	XX	YYO	O

As with XO and XXX females, XXY males are generally not recognizable in domestic animals until they are karyotyped. One notable exception is male tortoiseshell cats. In Chapter 1, we saw that tortoiseshell cats are heterozygous at an X-linked coat-colour locus, and because males have only one X chromosome, we would expect all tortoiseshells to be female. However, male tortoiseshells do occur from time to time. They are usually sterile, and when examined cytogenetically, a majority of them are found to be XXY in at least a proportion of their cells. The most interesting tortoiseshell males are those that have a normal XY karyotype. Being tortoiseshell, they must have two different types of X chromosome, but each cell appears to have only one X chromosome. The most likely explanation is that XY tortoiseshell males result from the fusion of two male embryos, one carrying an orange allele on its X chromosome, and the other carrying a non-orange allele. Having a normal genome in every cell, these males are fertile. It is possible, therefore, to come across a fertile male tortoiseshell cat but they constitute only a minority of tortoiseshell males, and are therefore very rare.

One of the most common forms of sex-chromosomal aneuploidy is the occurrence of more than one cell line in an individual, e.g. XY/XXY or XX/XXY/XXYY. The usual cause of these mixtures is non-disjunction during mitosis at an early stage of embryo development. The result is almost invariably some form of sexual abnormality. Because each of the cell lines has come from a single source (the fertilized ovum), the individuals are called **mosaics**.

The general conclusion we can draw from the above discussion is that aneuploidy of sex chromosomes is usually associated with sexual abnormalities.

We have concentrated on sex chromosomes when discussing aneuploidy because most cases of aneuploidy involve these chromosomes. However, aneuploidy of autosomes can also arise from non-disjunction, and some cases have been reported. The best known example in animals is trisomy 18 in cattle, which is associated with extreme brachygnathia in calves which die soon after birth (known as lethal brachygnathia trisomy syndrome or LBTS). Trisomies of several other small autosomes have been reported in cattle, and similar cases have been observed in horses (Appendix 4.1). Autosomal trisomics often have severe abnormalities, resulting in early death. However, some cases have survived to adulthood.

Copyrighted material

All of the occurrences of autosomal trisomy in domestic animals (and in humans) involve small chromosomes; trisomics for large autosomes, and monosomics for any autosome have not been observed in living animals or humans. The absence of monosomics is somewhat surprising, because non-disjunction gives rise to equal numbers of gametes with an extra chromosome and gametes lacking a chromosome. Evidence in mice indicates that equal numbers of trisomic and monosomic embryos are produced at fertilization, and remain alive at least until the late morula or early blastocyst stage (day 3 in the mouse). Thereafter, the frequency of monosomics relative to trisomics decreases, and no monosomics remain alive beyond days 12–13.

This introduces us to the second major effect of chromosomal abnormalities, namely the loss of embryos before birth. Surveys of embryos in animals have shown that approximately 7 per cent of all embryos have chromosomal abnormalities which are rarely, if ever, seen in animals after birth or hatching. Noting that approximately 30 per cent of all conceptions in animals are naturally aborted, we can conclude that around one-quarter of embryonic loss is the result of chromosomal abnormalities.

Common among the chromosomal abnormalities seen in early-stage embryos is the presence of just one set of chromosomes (**haploidy**) or the presence of more than two sets (**triploidy**, **tetraploidy**, etc.; collectively called **polyploidy** if there is just one type of ploidy, or **mixoploidy** if there are mixtures of different ploidies). While most of these abnormalities are never seen in living animals after birth or hatching, there are reports of triploid adult chickens. Indeed, in Australia there is a line of chickens which regularly produces around 12 per cent live triploid chickens that survive to adulthood. However, the triploids that survive are only about 50 per cent of those that are hatched. In other words, even in this unusual case where triploids survive to adulthood, polyploidy decreases life expectancy.

The triploid individuals in the Australian line arise from a high frequency of non-disjunction in both meiosis I and meiosis II in the normal (fertile) diploid females of this line. It is interesting to consider the different combinations of sex chromosomes that could occur in triploids. WWW chickens are never seen, even as embryos, because the gametic contribution from any male parent has to contain a Z chromosome. ZWW embryos are seen, but none of these survive to hatching. ZZW chickens are particularly unusual. At hatching, their left gonad is an ovary containing oocytes, while their right gonad is a testis. Soon after hatching, the left gonad begins to develop testicular-like tissue. Externally, these chickens have a normal female appearance until they reach puberty, at which stage they undergo an apparent sex reversal, ending up looking like a normal mature male! By then, both gonads are producing abnormal sperm cells. These sex-reversed birds are still viable, but are sterile. The final category, namely ZZZ, has a normal male phenotype but produces abnormal sperm cells and is sterile.

Summarizing the results described above, we can conclude that abnormal chromosomal numbers have a relatively high incidence among newly conceived zygotes, but have a relatively low incidence among individuals sampled after birth. Chromosomal abnormalities, therefore, are an important cause of embryonic loss.

Copyrighted material

Induced polyploidy in fish

Unlike mammals, many species of fish can happily exist as polyploids. If normal diploid female fish are subjected to short treatments of heat or cold or pressure, some of their developing ova fail to undergo meiosis II, and therefore remain in a diploid state. When such ova are fertilized with normal (haploid) sperm, triploid offspring result. Because of their unusual chromosome complement, triploid fish are sterile. Their viability, however, is quite good. Their sterility is one of their main attractions; they can devote all of their energies to growth, without being sidetracked by reproduction, which not only reduces growth but also increases mortality in males. There is also some evidence that the extra level of ploidy in itself results in faster growth. However, being sterile, each new generation of triploids has to be produced from diploids, as described above. If triploids really are advantageous from the commercial fishery point of view, it would be of great benefit to be able to produce them without having to go to the trouble of the treatments described above. This can be achieved by first producing tetraploid stock, as follows.

If normal fertilized ova are subjected to the same short treatment as described above, one round of mitosis is suspended, resulting in tetraploid offspring, which are viable and fertile. (The fertility presumably arises from the fact that every chromosome can find a homologue with which to pair.) The attraction of fertile tetraploids is that their 'natural' gamete is diploid, which means that a mating between a tetraploid and a normal diploid produces triploid offspring. And because the tetraploids themselves are fertile, they are self-reproducing. In fish, therefore, polyploidy can be exploited as a means of increasing food production.

Abnormal chromosome structure

Reciprocal translocations

In Sweden in 1963, a Swedish Landrace boar served 21 sows and produced an average litter size of 5.6 piglets. In previous pregnancies when mated to other boars, the same sows had an average litter size of 12.7 piglets. The karyotype of this boar was examined, and, as shown in Fig. 4.2, it was found that part of one chromosome had been interchanged with part of another: a process called **reciprocal translocation**. This involves each of two non-homologous chromosomes breaking into two segments, followed by an exchange of segments between the two chromosomes.

In the translocation shown in Fig. 4.2, one chromosome 15 lacks most of its long arm and is therefore written as 15q−. As the short arm of one chromosome 11 has been extended by the addition of the portion from 15, it is designated 11p+. The translocation is written as t(11p+; 15q−) or rcp(11p+; 15q−). Because only one member of each pair of chromosomes is affected, the individual in Fig. 4.2 is said to be a **translocation heterozygote** or **translocation carrier**.

Many reciprocal translocations have been reported in domestic animals, including some that involve exchanges between three different chromosomes (Appendix 4.1).

Copyrighted material

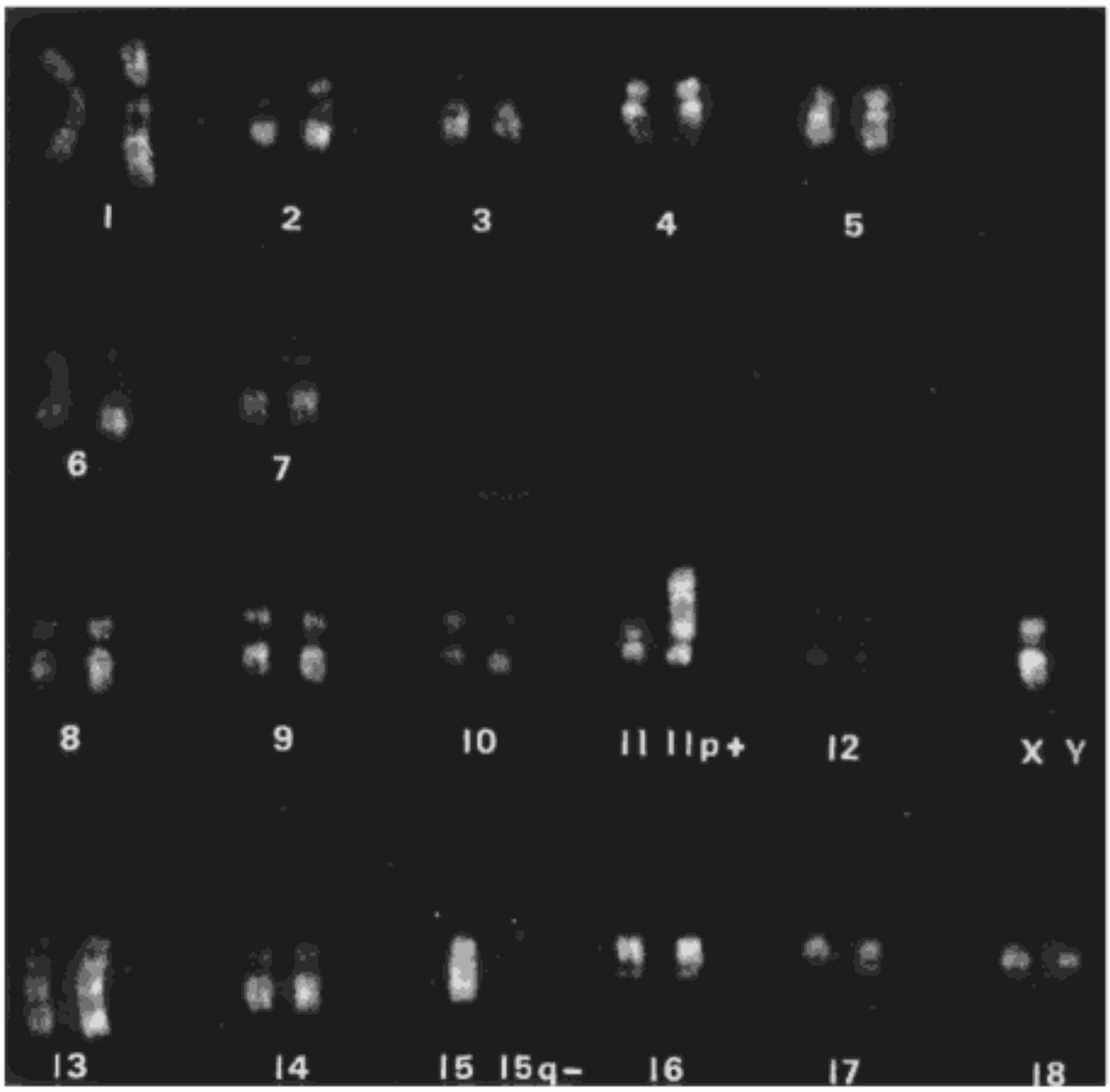


Fig. 4.2 The Q-banded karyotype of a Swedish Landrace boar, showing a reciprocal translocation between chromosomes 11 and 15, which caused the boar to produce substantially smaller litters than other boars. (Reprinted courtesy of I. Gustavsson from Hageltorn *et al.* (1973) *Hereditas* 75, 147–51.)

Animals carrying a reciprocal translocation have a normal phenotype, but show a significant reduction in fertility. The reason for the normal phenotype is that animals with a reciprocal translocation still have a normal complement of chromosomal material; it just happens to have been rearranged. The decrease in fertility arises from the problem faced during meiosis I in these animals. In normal circumstances, as we saw in Chapter 1, pairs of homologous chromosomes synapse to form bivalents. But in animals heterozygous for a reciprocal translocation, as shown in Fig. 4.3, four chromosomal structures have to come together in order to enable all homologous regions to synapse. This combination of four chromosomes is called a **quadrivalent**. There are many possible outcomes of disjunction from a quadrivalent, depending on how close the breakpoints are to the centromeres, and on the type of disjunction. Some of the outcomes are shown in Fig. 4.3; full details are given in Section 4.3.2.1 of *Veterinary Genetics*. In summary, we can say that in many cases, at least some unbalanced gametes are produced, containing either one or more additional chromosome segments, or lacking some segments. Most unbalanced gametes

Copyrighted material

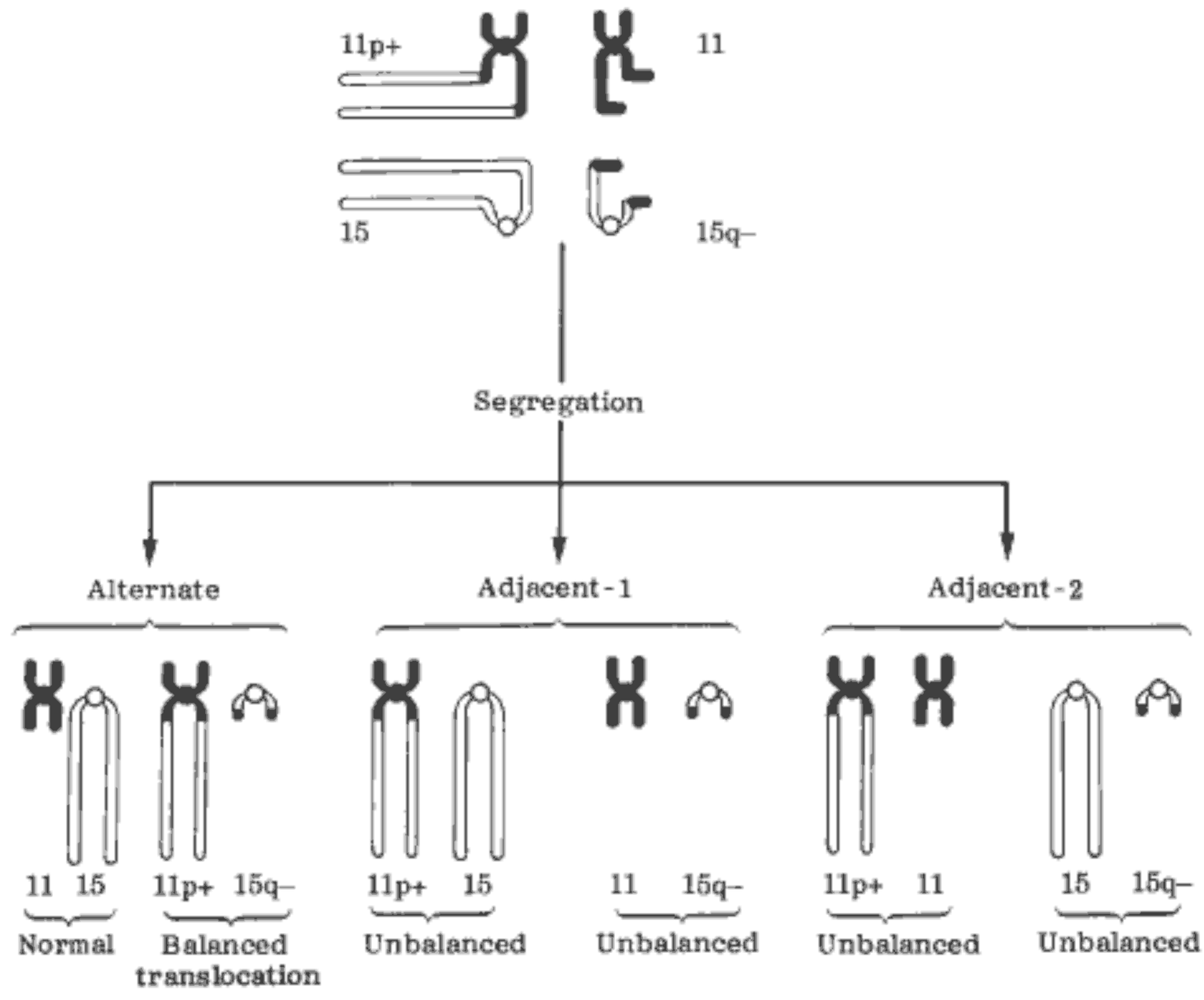


Fig. 4.3 Quadrivalent formed by the synapsis of homologous segments of chromosomes 11 and 15 in an animal heterozygous for the reciprocal translocation t(11p+; 15q–) (*top*). The six possible outcomes from the quadrivalent, assuming there is no crossing-over between the centromere and the breakpoint, i.e. in interstitial segments (*bottom*).

are able to function quite well as ova or sperm cells, but when an unbalanced gamete combines with another gamete to form a zygote, that zygote is unbalanced and dies before term. Thus, unbalanced gametes result in embryonic death. In theory, the resultant reduction in fecundity can range from zero to 100 per cent, depending on the factors mentioned above. In practice, there is always a substantial reduction in fecundity.

The importance of this reduction in fecundity is that it is an inherited defect: it is passed on from parents to offspring. This is because one-half of the balanced gametes resulting from meiosis contain translocation chromosomes, which, when combined with a normal gamete, result in another translocation heterozygote. Among the offspring produced by a translocation heterozygote, therefore, one-half are translocation heterozygotes.

Finally, if two gametes carrying balanced translocation chromosomes unite, a translocation homozygote results, with pairs of chromosomes that are exactly homologous and which consequently form only bivalents during meiosis. Thus, animals that are homozygous for reciprocal translocations produce only balanced gametes and therefore show normal fecundity.

Copyrighted material

Tandem translocations

Occasionally there are reports of part of the arm of one chromosome breaking off and joining onto the end of another chromosome. This is called a **tandem translocation**. In domestic animals, this type of translocation is quite a rarity compared with reciprocal translocations. Those that have been reported are listed in Appendix 4.1.

Centric fusions and fissions

A type of translocation that has generated considerable interest is one in which two acrocentric chromosomes fuse to produce one metacentric chromosome. This is known as a **centric fusion** or **Robertsonian translocation** (Fig. 4.4). It contains all the genetic material previously present in the two separate chromosomes.

Centric fusions involve the replacement of two chromosomes by one. It therefore follows that the total number of separate chromosomes in individuals heterozygous for a centric fusion is one fewer than the normal number (Fig. 4.4). However, unlike monosomics, which also have one fewer than the normal number of chromosomes, individuals that are heterozygous for a centric fusion have a complete genome and hence have normal phenotypes. Individuals that are homozygous for a centric fusion have two fewer than the normal number of chromosomes, but they too have normal phenotypes because they still have a complete genome.

What is the effect of centric fusions on reproductive ability? Meiosis in individuals heterozygous for a centric fusion is certainly unusual, because three chromo-

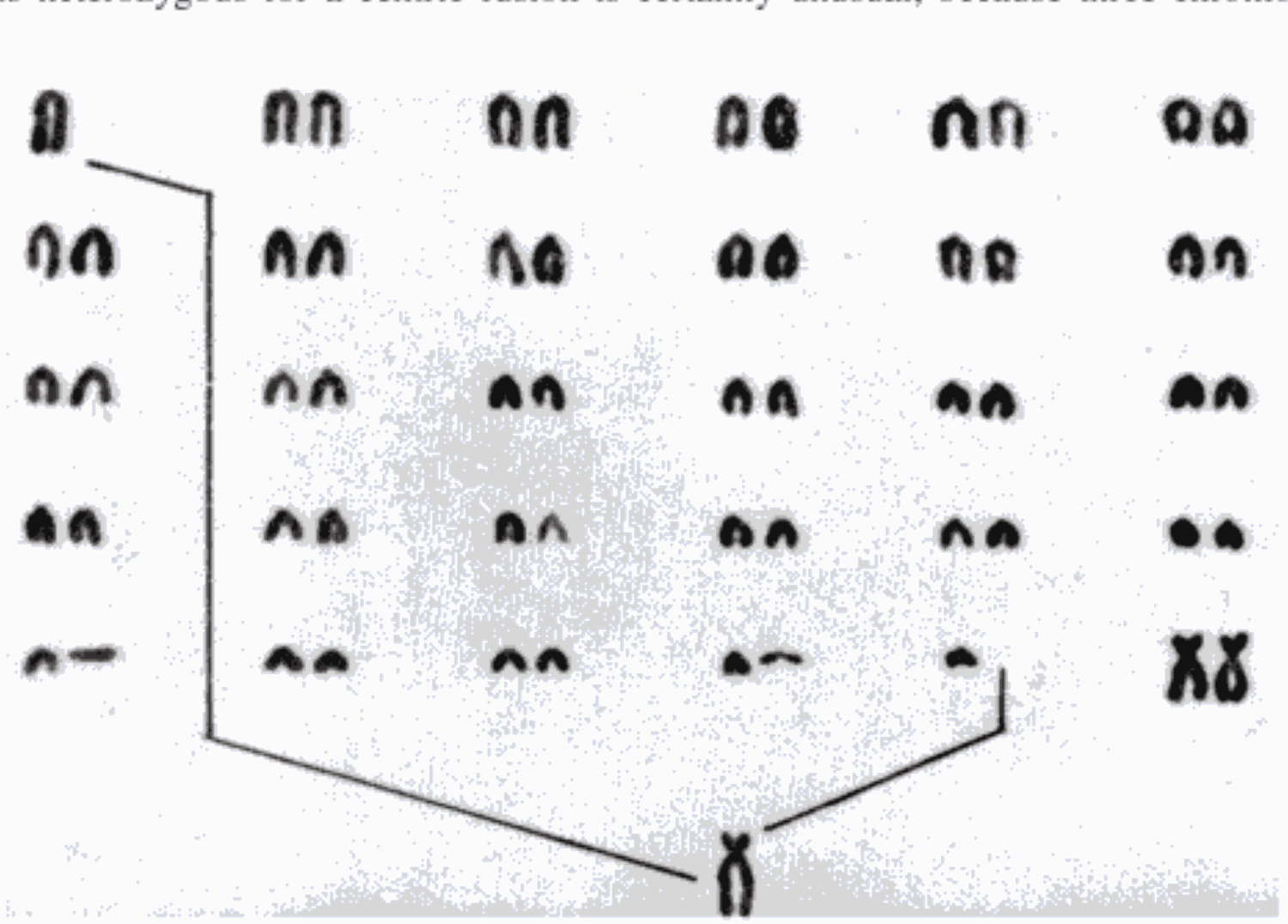


Fig. 4.4 An unbanded karyotype of a Swedish Red and White cow that is heterozygous for a centric fusion translocation between chromosomes 1 and 29. (Reprinted by permission from Macmillan Publishers Ltd: *Nature*, (1964) Vol **203**, pp. 990, Gustavsson, I. and Rockborn, G.; courtesy of I. Gustavsson.)

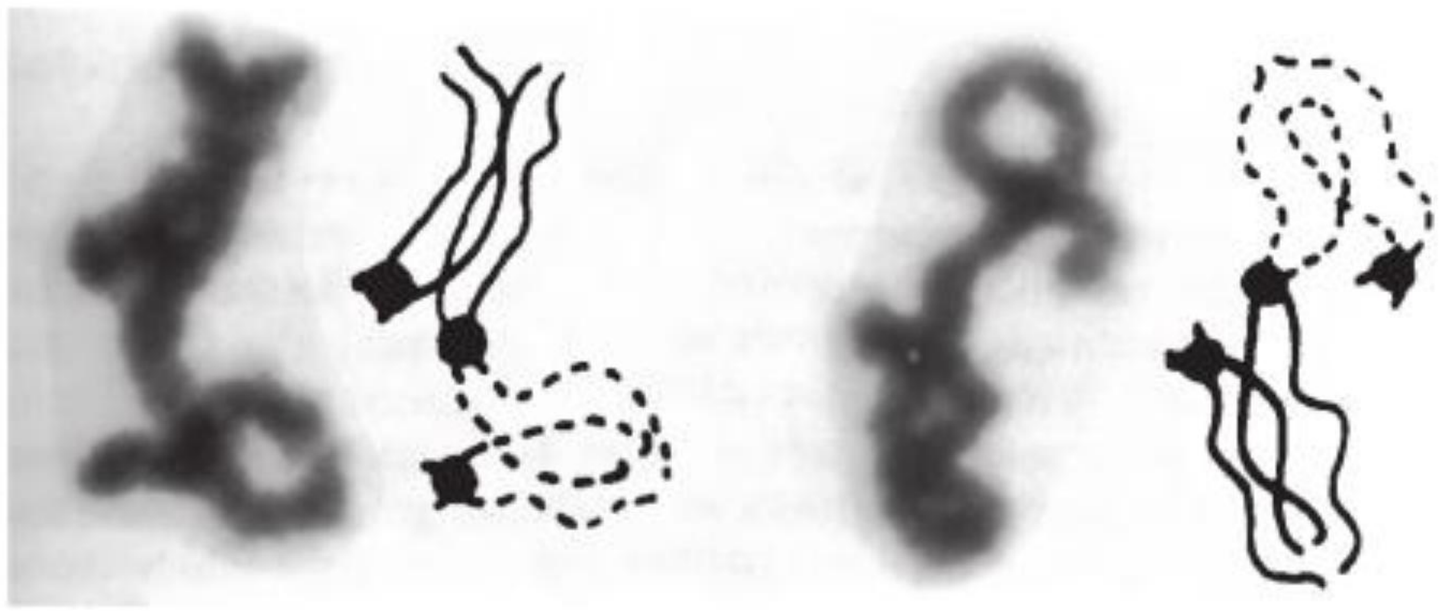


Fig. 4.5 Two examples of trivalents in meiosis I in rams heterozygous for a centric fusion between chromosomes 9 and 12. (Reprinted by permission from National Research Council, Canada: Chapman and Bruere (1977), *Canadian Journal of Genetics and Cytology* 19: 93–102; courtesy of N. Bruere.)

somes have to synapse, forming a **trivalent**, as shown in Fig. 4.5. If the centric-fusion chromosome disjoins from the other two, balanced gametes result. Any other type of disjunction leads to unbalanced gametes which would be expected to lead to aneuploid embryos. In practice, results vary substantially. In cattle, most centric fusions have an adverse effect, giving rise to a varying proportion of monosomic and trisomic embryos that fail to survive to term. In sheep, there seems to be no adverse effect. The most likely explanation for this difference is that unbalanced gametes fail to function in sheep, whereas they do function in cattle. It is still an open question as to why this should be so.

Occasionally, a metacentric chromosome splits into two acrocentrics, i.e. there is a **centric fission**, producing an animal that has one additional chromosome but is not trisomic. This has been reported in the donkey (Appendix 4.1).

Inversions and deletions

As their name suggests, inversions arise when a segment of chromosome becomes inverted following breakage of a chromosome in two positions. If the segment includes the centromere, the inversion is said to be **pericentric**; if the centromere is not included, the inversion is **paracentric**. Deletions arise if, following the breakage of a chromosome in two positions, the segment between the two breakpoints is lost. It is obvious that inversions give rise to the realignment of genes on a chromosome, and deletions result in the loss of genetic material.

Inversions and deletions are the least frequently observed of chromosomal abnormalities in domestic animals, and only isolated cases have been reported (Appendix 4.1). Deletions usually produce serious abnormalities, as the individual is monosomic for the segment of chromosome deleted. In contrast, because inversions involve the rearrangement of existing genes without any loss or addition, individuals carrying inversions usually have a normal phenotype. Furthermore, inversions are not

Copyrighted material

always easy to detect in karyotypes. A chromosome with a paracentric inversion, for example, has exactly the same shape and size as a normal chromosome. It may, however, differ in banding pattern, and as banding techniques become more sophisticated, so too will more inversions be detected.

Isochromosomes

This type of chromosomal abnormality arises when the centromere of a metacentric chromosome splits transversely (i.e. perpendicular to the chromatids) rather than longitudinally (i.e. parallel to the chromatids), during mitosis or meiosis. This leaves two identical chromatid arms joined to the one centromere, so that each arm contains exactly the same set of genes, in the same centromere-to-telomere order. The resultant structure is called an **isochromosome**. Each transverse split of a centromere gives rise to two isochromosomes (one corresponding to the p arm of the original chromosome, and the other corresponding to the q arm). However, sometimes one of the products of the split is lost, especially if the centromere splits unequally. There have been isolated reports of isochromosomes in animals. For example, an infertile heifer which was a mosaic for sex-chromosome aneuploidy, also had an isochromosome Y in a proportion of her cells.

Fragile sites

If certain substances, such as caffeine or aphidicolin, are added to the cell culture from which karyotypes are being prepared, partial gaps appear at particular sites on some chromosomes. Known as **fragile sites**, these gaps are inherited in a Mendelian manner. In dogs, for example, there are three fragile sites on the X chromosome, at Xp22, Xq21, and Xq27. Interestingly, each of these sites corresponds to a fragile site on the human X chromosome. For many years, fragile sites in human chromosomes have been known as sites where chromosomal rearrangements often occur in certain types of cancer, or have been associated with various inherited defects. Some of the fragile sites in pigs correspond to sites where breakage and reunion has occurred in reciprocal translocations. The Xq27.2 fragile site in humans (which is associated with mental retardation) is actually an unstable trinucleotide repeat (described in Chapter 3). It is quite likely that fragile sites in animals will be shown to be the result of unstable trinucleotide repeats, and to be associated with various chromosomal aberrations and inherited diseases.

Chromosomal aberrations in cancer

A wide spectrum of chromosomal aberrations has been recorded in cancerous tissue, including aneuploidy, polyploidy, translocations, centric fusions, deletions, and isochromosomes. In most cases, the abnormalities are a consequence of the lack of control of cell division that is characteristic of cancer cells.

Copyrighted material

Evolution of karyotypes

Despite the wide range of chromosome numbers in different species, the total quantity of DNA in each cell in all mammals is very similar. In fact, the evolution of modern mammals has been associated with much rearrangement and shuffling of the same total quantity of DNA. One of the most common forms of shuffling appears to have been centric fusion. For example, the three metacentric chromosomes (1, 2, and 3) in domestic sheep each correspond to a pair of acrocentric chromosomes in cattle and goats. It is an open question as to which came first, i.e. whether there was centric fusion or fission. However, because sheep are more closely related to goats than to cattle, it is likely that the acrocentric form is the most ancestral, in which case the three sheep metacentric chromosomes would be the result of centric fusion after the split between cattle and the common ancestor of sheep and goats.

In this context, it is interesting to compare the total number of chromosome arms in different species. If chromosomal evolution has proceeded largely by the accumulation of centric fusions, we would expect to find that the total number of major chromosome arms (called the **Nombre Fondamental** or **NF**) is fairly constant across species. In many situations, this is indeed so. Within the superfamily Bovoidea, for example, the diploid number varies from 30 to 60, but the NF in all but three cases varies only from 58 to 62.

Furthermore, there is extremely close resemblance in banding patterns within chromosome arms of sheep, goats, and cattle. We have already seen in Chapter 2 that there is considerable similarity in the gene maps between species. In other words, the chromosome structure within arms is also conserved.

Interspecific hybridization

An appreciation of the differences in karyotype among species is necessary for an understanding of the results of matings between members of different species. These matings are known as **interspecific hybridizations**. At one extreme, the result is an unsuccessful fertilization which is reflected in a failure of the female parent to conceive. At the other extreme is the production of viable, fertile offspring. In between are all possible combinations of success and failure. This is exactly what we would expect from the fact, first emphasized by Charles Darwin, that the concept of species is an artificial human contrivance by which we attempt to categorize into discrete units what is in reality a continuum. In most cases, the result of interspecific hybridization is primarily determined by the degree of similarity between the chromosomes of the two species, rather than the degree of physical resemblance.

This last point can sometimes be the source of considerable frustration in attempts to save a species by bringing together animals that come from different areas. Even if the animals closely resemble each other, and so appear to belong to exactly the same species, it is possible that they differ sufficiently at the chromosome level for there to be only a low chance of a viable, fertile offspring being produced. Other aspects of captive breeding programmes are discussed in Chapter 20.

Copyrighted material

Freemartins

As long ago as the first century BC, it was known that most female calves born co-twin to a male are sterile. Such females are called **freemartins**. The same condition has since been recognized in other species, and the term freemartin is now used to describe sterile females born co-twin to a male in any species. In the following discussion, we shall concentrate mainly on cattle, because freemartins occur much more commonly in cattle than in any other species.

If an individual receives cells from another individual, it has two populations of cells, each derived from a different source. Such individuals are said to be **chimeric**. In the case of freemartins, there is a fusion of chorions of the two embryos, and anastomosis of their blood vessels (called **vascular anastomosis**). The result is an exchange of haemopoietic cells which remain active for the remainder of each animal's life. Thus, each member of the twin pair is chimeric for erythrocytes and for leucocytes. This in turn means that each member exhibits its own blood groups (see Chapter 8) plus those of its co-twin. Since chromosomes are readily visible in leucocytes (as described in Chapter 1), it follows that in unlike-sex twins, the two populations of leucocytes are readily distinguishable according to their sex chromosomes: those derived originally from male haemopoietic cells are XY, and those derived originally from female haemopoietic cells are XX. The existence of these two readily observable different populations of leucocytes in unlike-sex twins is called **XX/XY chimerism**. A wide range of proportions of XX leucocytes in unlike-sex twins has been reported, ranging from 1 per cent to 100 per cent, with an average of approximately 50 per cent.

We are accustomed to thinking of the karyotype of leucocytes as indicating the karyotype of all other cells in an individual. But in twins whose chorions were fused, this is not so. Indeed, erythrocytes and leucocytes are the only cells for which there is convincing evidence of chimerism following vascular anastomosis; all other cells in female twins are XX, and all other cells in male twins are XY. Similarly, blood-cell chimerism can cause confusion when genotyping individuals using either biochemical or DNA technologies, because the blood cells may show the genotype of the individual's twin, as well as its own genotype (see Chapter 21). Obviously, if the twins are identical, this presents no problem. But if the twins are not identical (i.e. if they are dizygotic), considerable confusion can result, especially if it is not realized that the individual was born as a twin. Of course, the problem can be avoided by using some tissue other than blood cells.

The final result of vascular anastomosis is homograft tolerance, which is the ability of a co-twin to accept a graft of skin or other tissue from its fellow co-twin without showing any signs of rejection. The practical implication of this phenomenon is that in cattle, where vascular anastomosis is common, skin grafting cannot be used to distinguish between monozygous and dizygous twins.

In females that are born co-twin to a male, the major effects are seen in the gonads and in the reproductive tract. Until day 60 of fetal life in cattle, the female gonads appear to be developing normally. Thereafter, development is 'masculinized' to an extent that varies considerably from female to female. At one extreme, the gonads sometimes develop into apparently normal ovaries that are capable of ovulation. At

Copyrighted material

the other extreme, the gonads sometimes develop into miniature testes. In most cases, the result is that one or both gonads are classified as ovatestes, containing both ovarian and testicular tissue. The external genitalia are usually the same as those for normal females, except for the clitoris, which is often enlarged. Internally, there tends to be repression of the Müllerian duct derivatives (Fallopian tubes, uterus, cervix, and upper portion of the vagina) and over-development of Wolffian duct derivatives (epididymis, vas deferens, and seminal vesicle). Depending on the extent of alteration from normal development, the internal reproductive tract varies from more-or-less normal female to more-or-less normal male; some females have a normal vagina, cervix, and uterus, while at the other extreme, some have a blind vagina together with vasa deferentia and seminal vesicles instead of a cervix and uterus. All intermediate combinations are possible. Unfortunately, the extent of masculinization is not correlated with the proportion of XY leucocytes in the female co-twin.

Vascular anastomosis between unlike-sex twins has little, if any, effect on the structure of male gonads and reproductive tracts; their structure is essentially normal. There is, however, an effect on reproductive ability, with decreased sperm motility and sperm concentration leading to a 10-fold increase in the chance of a bull born co-twin to a female being culled for poor reproductive performance.

The probability of vascular anastomosis occurring between cattle twins is very high, being approximately 90 per cent. Because vascular anastomosis almost invariably leads to sterility of the female in unlike-sex twins, it follows that approximately 90 per cent of all female calves born co-twin to a male are freemartins.

Freemartins have been reported in red deer, goats, pigs, sheep, and Rocky Mountain bighorn sheep but in each case, the incidence is much lower than in cattle. In sheep, for example, exchange of cells occurs in only about 1 per cent of twins. In other species such as horses, exchange of cells between embryos is common, but the resultant chimerism does not lead to infertility. The reasons for these differences are not known.

The most effective way to check whether a young female calf born co-twin to a male is likely to be a freemartin, is to search for XX/XY chimerism. Since some freemartins have only a small proportion of XY leucocytes, it is obvious that the efficiency of this diagnosis increases with the number of cells scored from each suspect female. Taking account of the observed distribution of XY leucocytes in freemartins, the general recommendation is to score around 200 cells. Of course, as soon as the first XY cell is spotted, the counting can stop. More recently, molecular techniques have been developed for diagnosis of freemartins. For example, PCR amplification of a segment of the Y chromosome can be used to detect XY cells.

Biological basis of sex

We saw in Chapter 1 that sex is inherited in a simple Mendelian manner, determined by the presence or absence of a Y chromosome. In 1990, Andrew Sinclair and colleagues showed that the male-producing effect of the Y chromosome is the result of a gene which they called *SRY*, for sex-determining region of the Y chromosome. *SRY* is located on the short arm of the Y chromosome, quite near to the

Copyrighted material

pseudo-autosomal region (described in Chapter 1). The importance of the *SRY* gene is that it appears to provide the initial signal which causes the undifferentiated gonad in an embryo to develop into a testis rather than an ovary. The exact nature of this signal has still to be worked out, but the polypeptide product of the *SRY* gene is a DNA-binding protein, consistent with its apparent role as a regulator. Current knowledge suggests a somewhat complex form of regulation, in which the *SRY* peptide inhibits the action of another, related gene that encodes a testis inhibitor. In the absence of *SRY*, the inhibitor performs its normal role of preventing the undifferentiated gonad from developing as a testis, instead enabling it to proceed down the ovarian path. Whichever path is followed, all other aspects of sexual development result from the action of hormones produced by the differentiated gonad.

XX males and XY females

Occasionally there are reports of phenotypic males with an XX karyotype and phenotypic females with an XY karyotype. In most cases, molecular inspection shows that one of the X chromosomes of XX males has a very small insertion from the Y chromosome (which includes the *SRY* gene). XY females have several possible causes. In some cases, their Y chromosome has a small deletion, including the *SRY* gene. In other cases, they could have a mutant *SRY* allele that results in a deficiency of SRY polypeptide. In either case, in the absence of the message from the *SRY* gene, these individuals develop ovaries, which (as described above) then give rise to female secondary sexual characteristics. Some XY females, however, have a normal *SRY* gene, but lack the receptor for the male hormones (androgens) produced by the testes. This means that the only hormones that can exert any influence on secondary sexual development are the small quantities of female hormones produced by the testes. The result is that these XY individuals develop all the secondary sexual characteristics of females, despite having testes rather than ovaries. This particular defect is called testicular feminization. The gene for the androgen receptor is located on the X chromosome and for this reason testicular feminization is an X-linked defect. It has been reported in cats, cattle, horses, and dogs, as well as in humans.

Classification of intersex

It is evident from this chapter that there is a continuing occurrence of **intersex** individuals, i.e. individuals having various mixtures of maleness and femaleness. Several terms are used to classify intersexes. One is **hermaphrodite**. In its broadest sense, this term is synonymous with intersex as just defined. However, it is sometimes used in a narrower sense to indicate the presence of both ovarian and testicular tissue. Such individuals are often called **true hermaphrodites**, to distinguish them from **pseudohermaphrodites**, which are intersexes having either ovarian or testicular tissue, but not both. If only ovarian tissue is present, the intersex is called a **female pseudohermaphrodite**, whereas if only testicular tissue is present, the term **male pseudohermaphrodite** is used.

Copyrighted material

A very useful way to classify intersex is according to the stage at which the abnormality of sexual development occurs. This gives rise to the following three categories:

(1) Chromosomal intersex: animals whose abnormal sexual development is caused by abnormalities in sex chromosomes. It includes all the cases of sex-chromosome abnormalities discussed in this chapter.
(2) Gonadal intersex: individuals having either a normal male or a normal female karyotype, but with gonads that do not correspond to the chromosomal sex. Included in this category are XX and XY individuals with ovatestes, XX individuals with testes only, and XY individuals with ovaries only. Freemartins fit into this category because the vast majority of their cells are XX. Also included in this category are cases of XX and XY individuals in which the gonads have failed to develop, a condition called gonadal dysgenesis.
(3) Phenotypic intersex: individuals with normal chromosomal and gonadal sex, but with abnormalities in some or all of their reproductive tract and in other sexual characteristics. Testicular feminization fits into this category.

The final word on intersexuality must be left for the case of an intersex rabbit, which started life as a male, siring more than 250 offspring. It then lost interest in mating with females, and instead became pregnant to itself, giving birth to a litter of three males and four females, which it raised quite successfully on its own milk. We can conclude that just about anything is possible in the realm of intersexuality.

A sample of chromosomal aberrations

Just as with single-gene disorders in Chapter 3, it is possible to draw up a sample list of chromosomal aberrations that have been reported in animals. Such a list is presented in Appendix 4.1. It is not complete, but it does provide an overview of the range of different abnormalities that occur.

Summary

- A normal karyotype is not itself a guarantee of high production and freedom from disease, and neither is an abnormal karyotype certain to indicate a diseased or unproductive animal
- Abnormalities of chromosome number include the lack of one chromosome (monosomy) and the presence of an extra chromosome (trisomy), both of which arise from non-disjunction during meiosis
- The most common forms of these abnormalities in living animals involve the sex chromosomes, with the outcome often being decreased reproductive performance

Continued

Copyrighted material

- Trisomics for large autosomes and monosomics for any autosome have not been observed in living animals, presumably because they lead to embryonic death
- The presence of only one set of chromosomes (haploidy) and the presence of more than two sets of chromosomes (polyploidy) are common chromosomal abnormalities among young embryos but are very rare among living animals, again because they lead to embryonic death
- Abnormalities of chromosome structure include translocations, inversions, and deletions
- Animals heterozygous or homozygous for reciprocal translocations have a normal complement of genetic material and hence have normal phenotypes
- Translocation heterozygotes give rise to unbalanced gametes and hence show reduced fertility
- Centric fusions (Robertsonian translocations) involve the joining of two acrocentric chromosomes into one metacentric chromosome, but involve no loss or gain of genetic material and hence give rise to normal phenotypes
- However, meiosis in animals heterozygous for centric fusions involves segregation from trivalents, which results in reduced fertility in cattle but not sheep
- The effect of eradicating a centric fusion in cattle will depend on its frequency; only if it is relatively common will its eradication have a noticeable effect on fertility
- Inversions arise when a segment of chromosome becomes inverted following breakage of a chromosome in two positions; if the segment includes the centromere, the inversion is said to be pericentric; if the centromere is not included, the inversion is paracentric
- A wide spectrum of chromosomal aberrations has been recorded in cancerous tissue, including aneuploidy, polyploidy, translocations, centric fusions, deletions, and isochromosomes
- One of the most common forms of karyotype evolution has involved centric fusion or centric fission; for example, the main difference between sheep and cattle karyotypes is that the former has three large metacentric chromosomes, each comprised of two cattle acrocentric chromosomes
- A freemartin is a sterile female born co-twin to a male, characterized by masculinized gonads and internal genitalia, erythrocyte and leucocyte chimerism, and tolerance of reciprocal skin grafts
- The male-producing effect of the Y chromosome in mammals is the result of a gene called *SRY*, for sex-determining region of the Y chromosome, which provides the initial signal that causes the undifferentiated gonad in an embryo to develop into a testis rather than an ovary, most likely by inhibiting the action of another, related gene that encodes a testis inhibitor
- Intersexes are individuals with a mixture of male and female characteristics, resulting from various abnormalities of sexual development
- Intersexes can be classified according to whether the abnormality of sexual development occurred in the sex chromosomes (chromosomal intersex), in the gonads (gonadal intersex), or at a later stage of development (phenotypic intersex)

Copyrighted material

Further reading

Freeman, G. (2007) Explaining the freemartin: Tandler and Keller vs. Lillie and the question of priority. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **308**, 105–12.

King, W.A., Iannuzzi, L. and Villagómez, D.A.F. (Eds) (2008) Veterinary cytogenetics. *Cytogenetic and Genome Research*, **120**, 5–187.

Marshall Graves, J.A. (2008) Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. *Annual Review of Genetics*, **42**, 565–86.

Sekido, R. and Lovell-Badge, R. (2008) Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*, **453**, 930–4.

Appendix 4.1 A sample of chromosomal aberrations in animals

Aneuploidy of autosomes	
Trisomy 12, trisomy 16, trisomy 17, trisomy 18, trisomy 20, trisomy 22, trisomy 23, trisomy 24	cattle
Trisomy 14	pig
Trisomy 15	chicken
Trisomy 23, trisomy 26, trisomy 28, trisomy 30	horse
Tetrasomy 15	chicken
Aneuploidy of sex chromosomes	
XO	buffalo, cat, dog, horse, llama, opossum, pig, sheep
XXX	buffalo, cattle, dog, horse
XXY	cat, cattle, dog, horse, pig, sheep
XXXY	horse, pig
XXXXY	horse
ZZW	chicken
Centric fission	
3	donkey
Deletion	
X	horse
Fragile site	
Fragile site	buffalo, cat, cattle, dog, goat, horse, pig, rabbit, sheep
Insertion	
16	cattle
19	sheep

Continued

Copyrighted material

Inversion, paracentric	
8	pig
Inversion, pericentric	
2	chicken
14, X	cattle
Isochromosome	
X	cattle
Y	horse
Polyploidy	
Tetraploidy	chicken, rainbow trout
Triploidy	chicken, macaw, oyster, rainbow trout
Sex reversal	
XX male	dog, goat, horse, llama, pig
XY female	cat, cattle, horse, sheep, wood lemming
Translocation, centric fusion	
1/20, 6/26, 7/25, 8/11	sheep
1/4, 1/21, 1/23, 1/26, 1/27, 1/28, 1/29, 2/4, 3/4, 3/27, 4/4, 4/8, 4/10, 4/21, 5/18, 5/23, 5/26, 6/28, 7/12, 7/21, 8/9, 9/23, 11/21, 11/22, 13/21, 14/20, 14/21, 14/28, 15/25, 16/18, 21/27, 27/29	cattle
13/17, 15/17	pig
17/19	Arabian oryx
1/31, 13/23, 21/33	dog
23/24	blue fox
5/15, 6/15	goat
Translocation, reciprocal	
1/6, 1/7, 1/8, 1/11, 1/14, 1/15, 1/16, 1/17, 2/4, 2/14, 2/15, 2/4/15, 3/7, 4/13, 4/14, 5/8, 5/14, 5/15, 6/14, 6/15, 7/11, 7/12, 7/13, 7/15, 7/17, 8/14, 9/11, 11/15, 12/13, 13/14, 13/17, 14/15, 15/16, 16/17, X/13, X/14	pig
1/20, 13/20, 23/24	sheep
1/3	horse
1/4, Z/1, Z/microchromosome	chicken
1/8/9, 8/13, 8/15, 10/11, 20/24, X/23, Y/17	cattle
Translocation, tandem	
1/16, 1/18, X/23	cattle
1/30, X/15	horse

Copyrighted material

5 Single genes in populations

Most single-gene disorders are rare and consequently are not cause for common concern. Occasionally, however, a disorder caused by a single gene reaches a high frequency among the animals belonging to one or a few breeders, or sometimes within a breed as a whole. The economic consequences of such an increase in frequency are sometimes quite severe, and breeders often ask how the disorder can be decreased in frequency, if not eliminated. In order to give useful advice, we need to know more than just simple Mendelian genetics. We must also understand the way in which genes behave within a herd or flock or kennel or cattery, or within a breed as a whole. In other words, we need to understand the basic principles of population genetics. The aim of this chapter is to explain those basic principles.

Gene and genotype frequencies

Haemoglobin in sheep exists in two different forms (HbA and HbB), which are the products of two different alleles, *A* and *B*, at an autosomal locus. With two different haemoglobin alleles, there are three different possible genotypes (*AA*, *AB*, and *BB*), each of which produces a distinctive protein electrophoretic pattern, as shown in Fig. 5.1. Each electrophoretic pattern can be regarded as a distinguishable phenotype, i.e. each genotype has a distinguishable phenotype. Blood samples were taken from a flock of 175 sheep, and after electrophoresis to determine haemoglobin type for each sheep, it was found that the numbers of the genotypes were 91, 28, and 56 for *AA*, *AB*, and *BB*, respectively. Given this information, it is possible to calculate the proportions of each genotype in the sample. These proportions are called **genotype frequencies**. For this particular example, they turn out to be $\frac{91}{175} = 0.52$ for *AA*, $\frac{28}{175} = 0.16$ for *AB*, and $\frac{56}{175} = 0.32$ for *BB*.

It is also possible to calculate the proportions, i.e. frequencies, of each allele in the sample. By convention, the term 'gene' is used instead of 'allele' in population genetics, when referring to frequencies. The proportions of each allele are therefore called **gene frequencies**.

All of the genes in the *AA* animals are *A*, and half the genes in *AB* animals are *A*; it therefore follows that the gene frequency of *A* is the frequency of the *AA* genotype plus half the frequency of the *AB* genotype, which equals $0.52 + \frac{1}{2}(0.16) = 0.6$. Similarly, the gene frequency of *B* is $0.32 + \frac{1}{2}(0.16) = 0.4$.

Copyrighted material

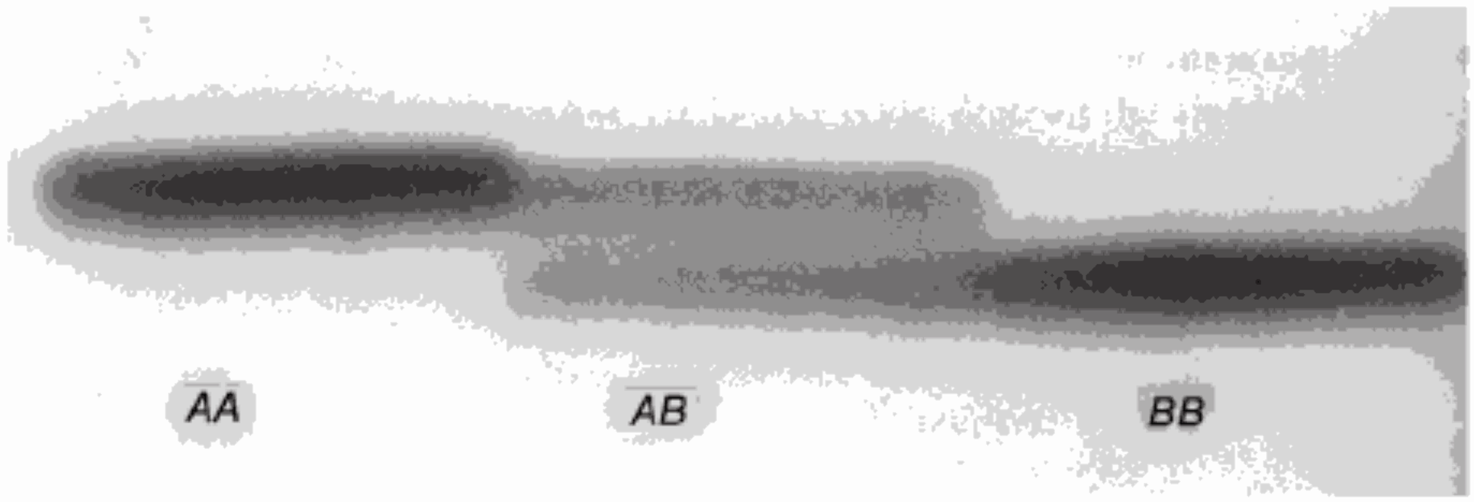


Fig. 5.1 The electrophoretic patterns, or phenotypes, corresponding to the three different haemoglobin genotypes in Merino sheep. (Reproduced courtesy of K. Bell.)

If there are only two alleles at a locus (as in the above example), then it is becoming increasingly common to refer to the less frequent of the two alleles as the minor allele, and hence to call its frequency the **minor allele frequency** (**MAF**). This concept is typically applied to SNPs, almost all of which are bi-allelic. The MAF in the example above is 0.4.

Random mating

Random mating for a particular trait or locus occurs when the choice of a mating partner is independent of the genotype or phenotype for that trait or locus. Because the breeding of domestic animals is largely under the control of humans, who often decide, for example, to select only a small proportion of individuals to be parents of the next generation, it might be thought that the concept of random mating is largely irrelevant to domestic animals. However, even though selection can be quite intense, mating is usually at random for all traits and loci, *among those animals that have been selected*. In fact, in production species such as cattle, sheep, and pigs, it is usually impractical to arrange any other type of mating. Even with companion animals, in deciding which animals to mate to which, humans base their decision on only a relatively small number of traits such as coat colour or various aspects of conformation.

For those traits that are taken into account by humans when matings are being planned, mating is often not random. For most other traits, however, mating is usually at random.

In most situations, for example, mating is at random with respect to blood groups, because these are usually not known. For many single-gene disorders, too, mating among animals that survive to reproductive age is random, because for many such disorders, the different genotypes among survivors all give rise to the same phenotype. For this reason, the concept of random mating is very important in populations of domestic animals.

Copyrighted material

The Hardy–Weinberg law

The Hardy–Weinberg law is a remarkable theory that is extremely useful in enabling us to understand what happens to gene frequencies and genotype frequencies in real populations. The law can be derived by observing what happens to gene and genotype frequencies in any population of animals that goes through one generation of random mating (see *Veterinary Genetics*, Section 5.4.1 and Appendix A5.1). When this is done, the following conclusions become evident:

In a random mating population in which there is no selection, mutation, migration, or genetic drift

(1) *genotype frequencies in offspring are determined solely by gene frequencies in parents, such that*
   (a) *the frequency of homozygotes equals the square of the relevant gene frequency;*
   (b) *the frequency of heterozygotes equals twice the product of the relevant gene frequencies;*
(2) *gene frequencies and genotype frequencies remain constant from one generation to the next.*

These statements are known as the **Hardy–Weinberg law**. The practical implications of the law are best seen if we express its main conclusions in simple algebra. We can do this by considering just two genes at a locus (e.g. *A* and *B* from the sheep haemoglobin example), and writing their frequencies as p and q. These frequencies can each have any value from zero to one, providing that the combined frequency of the two genes is one, i.e. $p + q = 1$. The first part of the law says that irrespective of the genotype frequencies in the parents, one generation of random mating results in genotype frequencies of p^2, $2pq$, and q^2 for *AA*, *AB*, and *BB*, respectively. The second part of the law says that if we worked through another generation (and another and another, etc.), we would discover that p and q remain the same, and that the genotype frequencies remain the same, namely p^2, $2pq$, and q^2. These are known as the **Hardy–Weinberg frequencies**.

Some of the assumptions require explanation. The assumption of no selection means that each genotype has an equal opportunity to contribute offspring, and each offspring has an equal opportunity to survive until it has an opportunity to mate. The assumption of no migration means that genes do not enter the population from outside – the population is closed. Genetic drift refers to changes in gene frequency that are the result of chance (see later in this chapter). The assumption of no genetic drift means that the number of parents is sufficiently large, and the number of offspring is sufficiently large, that chance fluctuations in gene frequency are negligible.

The part of the Hardy–Weinberg law that makes a specific prediction about the relationship between gene and genotype frequencies is often tested by comparing the observed numbers of each genotype in a population, with the numbers expected if the genotypes occur in Hardy–Weinberg frequencies. Despite the large number of

Copyrighted material

assumptions in the law, it turns out that most populations have Hardy–Weinberg frequencies of genotypes. In other words, despite the fact that selection, migration, mutation, and genetic drift are known to occur in almost all populations, Hardy–Weinberg frequencies of genotypes are usually observed. How can this be? The answer is that the test for Hardy–Weinberg frequencies cannot detect mutation, and it fails to detect many cases of selection, migration, and genetic drift. As we shall see below, this is a very fortunate situation, because it means that for most populations, we can safely assume that Hardy–Weinberg frequencies exist, which in turn enables us to draw conclusions that have very important practical implications.

The special case for recessives

The Hardy–Weinberg law applies irrespective of the type of gene action at the locus in question. In fact, so long as there is no selection, the type of gene action at a locus has no effect on the frequency of genes or genotypes. For example, it is clear from the second part of the Hardy–Weinberg law that recessive traits will neither decrease nor increase in frequency from one generation to the next, unless selection, mutation, migration, or genetic drift acts in such a way as to alter the frequency of the recessive gene. In fact, recessive traits can have a frequency anywhere in the range zero to one, depending solely on the frequency of the recessive gene.

Let us now examine the recessive case more closely, using coat colour in Angus cattle as an example. The typical black coat colour seen in Angus cattle is caused by a dominant allele *B*, while the red coat colour seen occasionally in the same breed is because of the homozygosity for a recessive allele *b*. Because there are two alleles at this locus in Angus cattle, there are three genotypes (*BB*, *Bb*, and *bb*). However, because *B* is completely dominant to *b*, the genotypes *BB* and *Bb* have exactly the same phenotype, namely black. In general, for traits determined by dominant and recessive alleles, it is not possible to deduce the genotype of all animals simply from their phenotype. We cannot, therefore, calculate gene frequencies as in the case where all genotypes are identifiable. We can, however, distinguish red (*bb*) from black (*B*–, where the dash indicates either *B* or *b*). In order to estimate gene frequencies, we make use of the general principle discussed in the previous section, namely that for the majority of loci, most populations have Hardy–Weinberg frequencies of genotypes, which means that the genotypes at a single locus with two alleles are in the proportions p^2, $2pq$, and q^2. Applying this to the case of red versus black in Angus cattle, and letting q be the frequency of *b*, we have the following situation:

Genotype	*BB*	*Bb*	*bb*
Phenotype	Black	Black	Red
Frequency	p^2	$2pq$	q^2

It is obvious that the frequency of red Angus equals q^2, which is the square of the gene frequency of *b*. The gene frequency of *b* can therefore be estimated as the square root of the frequency of red calves.

Copyrighted material

For example, the frequency of red calves in pedigree Angus herds in the USA is approximately 5 per 1000. Assuming Hardy–Weinberg frequencies, the estimate of the gene frequency of b is $\sqrt{(0.005)} = 0.07$. Then, because the only other allele at this locus is B, its frequency must be $1 - 0.07 = 0.93$.

We can now make one more interesting calculation. Since $p = 0.93$ and $q = 0.07$, it follows that the frequency of heterozygotes or carriers, which is $2pq$, equals $2 \times 0.93 \times 0.07 = 0.13$. Thus 13 per cent of Angus calves born in the USA are carriers of red. What proportion of black Angus are carriers? The answer is $2pq/(p^2 + 2pq)$, which equals $0.13/(0.86 + 0.13) = 0.13/0.99$, which equals 0.13.

This is a surprisingly high figure, but it is typical of the situation for all rare recessive traits, namely that the frequency of carriers is much higher than the frequency of the recessive trait itself.

Another example of a simple recessive trait is yellow coat colour in Labrador dogs, which is recessive to black. Many dog breeders prefer yellow to black, so the frequency of the yellow genotype (ee) is quite high in many populations. In one Australian population, for example, the frequency of yellow is approximately 64 per cent. Assuming Hardy–Weinberg frequencies, the gene frequency of e in Australian Labradors must be $\sqrt{(0.64)} = 0.8$, which means that the frequency of the dominant allele must be only 0.2. This is a good illustration of the fact that recessive genes and recessive traits can be much higher in frequency than dominant genes and dominant traits.

Extensions to the Hardy–Weinberg law

Multiple alleles

There are three different forms of glucose 6-phosphate dehydrogenase (G6PD) in horses, corresponding to three alleles, D, F, and S at the G6PD locus. Of course, any one horse can have at the most only two different alleles. In a large sample of horses, however, all three alleles are likely to be found, and the frequency of each allele can be estimated.

If p, q, and r represent the frequencies of the three alleles, the expected frequencies of the respective homozygotes are p^2, q^2, and r^2, and the respective heterozygotes have expected frequencies of $2pq$, $2qr$, and $2pr$. Using exactly the same principle, this prediction can be extended to any number of alleles at a locus.

X-linked genes

As we saw in Chapter 1, X-linked genes have a different pattern of inheritance compared with autosomal genes. To understand the implications of this, consider the X-linked coat colour locus in cats, for which each genotype in each sex has a distinguishable phenotype. Many population surveys have been conducted on cats in many countries, and the results for X-linked coat colour from two such surveys are shown in Table 5.1.

Because males have only one X chromosome, they have only one gene (i.e. they are **hemizygous**) at all X-linked loci. This means that the frequency of each pheno-

Copyrighted material

Table 5.1 Combined results of two surveys taken in Iceland of X-linked coat colour in cats, together with calculations of gene frequencies in males and females

Phenotype	Female				Male		
	Non-orange	Tortoiseshell	Orange	Total	Non-orange	Orange	Total
Genotype	*oo*	*Oo*	*OO*		*o*	*O*	
Numbers	117	53	3	173	149	28	177

Gene frequency of *o* in females = $(2 \times 117 + 53)/(2 \times 173) = 0.83$
Gene frequency of *O* in females = $(2 \times 3 + 53)/(2 \times 173) = 0.17$
Gene frequency of *o* in males = $149/177 = 0.84$
Gene frequency of *O* in males = $28/177 = 0.16$

type in males equals the frequency of the respective gene, so that the calculation of gene frequencies in males is very straightforward. In females, where there are three genotypes, gene frequencies can be calculated from first principles. It is evident from Table 5.1 that the gene frequencies in males and females are essentially the same, with average values of 0.835 for *o* and 0.165 for *O*. As the existence of the three genotypes in females is analogous to the general case for autosomal genes discussed earlier, it is tempting to see whether the female genotypes occur in Hardy–Weinberg frequencies. Using the average gene frequencies, expected genotype frequencies are $(0.835)^2$, $2 \times 0.835 \times 0.165$, and $(0.165)^2$ for *oo*, *Oo*, and *OO*, respectively. With a total of 173 females in the sample in Table 5.1, this gives expected numbers of 120.6, 47.7, and 4.7, which agree very closely with the observed numbers of 117, 53, and 3, respectively. Thus the female genotypes occur in Hardy–Weinberg frequencies.

These results are typical of those obtained with X-linked genes, which means that it is usually safe to assume that X-linked gene frequencies are the same in males and females, and that X-linked female genotypes occur in Hardy–Weinberg frequencies. What are the practical implications of this conclusion? The most important implication is that X-linked traits are expected to occur with different frequencies in males and females. This is most relevant to X-linked recessive traits, for which the frequency of the trait in males (q) is expected to be much higher than the frequency of the trait in females (q^2). Notice that the square of q is much smaller than q because q is always less than one. For example, if an X-linked recessive condition occurs with a frequency of 10 per cent in males ($q = 0.1$), its expected frequency in females is 1 per cent [$q^2 = (0.1)^2 = 0.01$].

Similar types of predictions can be made for any trait that is seen only in animals that are homozygous or hemizygous for an X-linked allele. For all such traits, the frequency in males (either p or q) is greater than the frequency of the same trait in females (either p^2 or q^2, respectively). This simple observation can be used to great effect to make some money from a small bet. For example, the frequency of the *O* gene in Australian cats is around 0.2. It follows that the expected frequency of orange males is 0.2, whereas the expected frequency of orange females is $(0.2)^2 = 0.04$. Thus the odds are 5 : 1 that the next orange cat you see in Australia will be a male. It is left to the reader to verify that the lower the gene frequency, the greater the odds.

Copyrighted material

Selection and mutation

Selection acts on phenotypes, and occurs whenever some phenotypes have a greater opportunity to contribute offspring to the next generation than do other phenotypes. Selection may act at any stage during the life cycle of an animal from conception to mating.

Selection most commonly occurs through differential viability and/or differential reproductive ability, with reproductive ability including factors such as mating ability, fecundity, and fertility. For convenience we shall refer to the combined effect of viability and reproductive ability as **fitness**. If selection occurs as a result of decisions made by humans it is called **artificial selection**, while in all other situations it is called **natural selection**. In either case, the principles by which selection operates are exactly the same. Although selection acts on phenotypes, we are mainly interested in its effect on genotypes and through them, on gene frequencies. Because of this interest, we often talk about selection acting on genotypes and on genes. Whenever we do this, however, we must remember that selection really acts only on individual animals according to their phenotype. The extent to which this affects genotypes and hence genes, depends on the extent to which particular phenotypes are associated with particular genotypes. In order to understand how selection works in practice, we shall start by considering the population of Labrador dogs mentioned earlier. Let us write the frequencies of the E and e genes as p and q, respectively.

Selection against a dominant

We have already noted that many Labrador breeders tend to favour yellow coat colour. This means that there is selection in favour of the recessive phenotype, which is the same thing as saying there is selection against the dominant phenotype. Let us represent this selection by saying that, relative to the fitness of genotype ee, the fitness of genotypes EE and Ee is reduced by a proportion s, where s is called the **selection coefficient**. The so-called **relative fitnesses** of the three genotypes are then $1 - s$, $1 - s$, and 1 for EE, Ee, and ee, respectively. To obtain the genotype contributions after selection, we multiply the genotype frequency before selection (which will be the Hardy–Weinberg frequency) by the relative fitness:

	Genotype			
	EE	Ee	ee	Total
Frequency prior to selection	p^2	$2pq$	q^2	1
Relative fitness	$1 - s$	$1 - s$	1	
Proportion after selection	$p^2(1 - s)$	$2pq(1 - s)$	q^2	$1 - sp(2 - p)$

What is the frequency of the E gene after selection? Using exactly the same approach as earlier, this is the frequency of the EE genotype plus $\frac{1}{2}$ the frequency of the Ee

Copyrighted material

genotype, which is $p^2(1 - s) + \frac{1}{2} \times 2pq(1 - s)$, all divided by the new total, namely $1 - sp(2 - p)$.

Now that we have an expression for the gene frequency after selection, we can calculate the change in gene frequency due to selection, Δp, which must equal the new gene frequency minus the previous gene frequency, which was simply p. When this expression is written out and simplified (as shown in *Veterinary Genetics*, Appendix A5.4.1), we end up with $\Delta p = -sp(1 - p)^2/[1 - sp(2 - p)]$. Among other things, this expression shows us that the change in gene frequency resulting from selection against a dominant phenotype depends on just two factors: the strength of selection (measured as s) and the gene frequency before selection. By substituting various values for these two parameters into the above equation, it can be seen that selection against a dominant phenotype can lead to quite substantial decreases in the frequency of the dominant gene. With strong selection, the gene is soon eliminated from the population. In the extreme case where selection is complete, i.e. no *EE* or *Ee* animals are able to contribute genes to the next generation, then $s = 1$, and the above expression reduces to $\Delta p = -p$, which means that the gene frequency after just one generation of selection will be zero. In other words, the *E* gene will be eliminated from the population in just one generation, which is, of course, exactly what we would expect because, in this extreme situation, no *E* genes are passed on to the next generation.

Selection/mutation balance for a dominant

With the black gene removed from the population, we might expect that black dogs would never appear again, unless the gene for black was introduced from another population by migration. However, mutation occurs from time to time at all loci, and the effect of mutation in the present context is to alter occasionally a yellow gene, *e*, to a black gene, *E*.

We therefore have two forces opposing each other: mutation is occasionally introducing dominant genes into the population, and selection is removing them. The result of these two opposing forces is that an equilibrium is reached, at which stage the number of mutant genes entering the population is the same as the number removed by selection, and the frequency of the dominant gene remains stable from generation to generation. This is called a **selection/mutation balance**. The smaller the effect of the dominant gene on fitness (the weaker selection is against it), the higher is the equilibrium frequency. Similarly, the greater the mutation rate, the greater is the equilibrium frequency. In fact, if μ is the mutation rate, then, as shown in Appendix A5.4.2 of *Veterinary Genetics*, the frequency of a dominant gene at a mutation/selection balance is μ/s, and the corresponding frequency of the dominant phenotype is $2\mu/s$.

To get some feel for what these expressions mean, consider a dominant trait that is subjected to only weak selection, say, $s = 0.05$; and let us take a common value for the mutation rate, say $\mu = 10^{-6} = 0.000001$. This gives the equilibrium gene frequency as 0.00002, and the corresponding phenotypic frequency as 0.00004, or 4 per 100 000, which is quite a low frequency. We can conclude that only very weak selection against a dominant gene is required to keep that gene at a very low frequency.

Copyrighted material

Selection against a recessive

Suppose now that Labrador breeders tend to favour black instead of yellow. In this case we have selection against the recessive gene for yellow, and selection in favour of the dominant gene for black.

Let us examine the effects of this type of selection for the general case, again using symbols.

Our approach is exactly the same as that taken in the previous section, except that now the relative fitnesses are 1, 1, and $1 - s$ for the three genotypes *EE*, *Ee*, and *ee*, respectively. Readers are urged to work through one generation of this type of selection for themselves, and to check their results against those given in Appendix A5.4.3 of *Veterinary Genetics*. It should be possible to show that the change of frequency of a recessive gene following one generation of selection against a recessive trait is $\Delta q = -sq^2(1 - q)/(1 - sq^2)$. Study of this equation shows that while selection against a recessive is effective at reducing gene frequency if the initial frequency is quite high, this type of selection becomes very ineffective at lower frequencies. In other words, selection against a recessive gene is a very inefficient means of removing that gene from a population. The reason for the decreasing effectiveness of selection against a recessive gene is that as the frequency of the gene decreases, an increasing proportion of recessive genes are 'hidden' from the effects of selection by occurring in heterozygotes. It follows, therefore, that selection against a recessive gene would be much more effective if heterozygotes could be detected.

If all heterozygotes were detected and then not used for breeding, all recessive genes would be removed from the population at once, and the frequency of the recessive gene would fall to zero. However, in practice there is no need to go to this extreme. As we shall see in Chapter 21, in order to eliminate recessive disorders from a population, all that has to be done is to prevent heterozygotes mating with other heterozygotes. This is the reason why so much research effort is being devoted to detection of heterozygotes in relation to recessive disorders. Notice that there is no requirement to destroy heterozygotes once detected. In fact, they could still be used for breeding, provided that no two heterozygotes were ever mated together. In this way, although the recessive gene remains in the population, the abnormality or disease that it causes never occurs, because homozygotes for the recessive gene never occur.

Selection/mutation balance for a recessive

Even if we are successful in removing a recessive gene from the population or at least preventing its appearance in homozygotes, we still have to contend with the effect of mutation, which, slowly but consistently, is adding new recessive genes to the population.

If we are detecting heterozygotes and not using them for breeding, we are in essentially the same situation as with complete selection against a dominant gene ($s = 1$): each new recessive gene entering the population is immediately removed, and a selection/mutation balance is reached at which the frequency of the recessive gene equals the mutation rate. However, because homozygotes never occur in this

Copyrighted material

situation, the frequency of the recessive trait at this selection/mutation balance is zero: affected animals never appear.

If we are detecting heterozygotes and then using them for breeding but avoiding matings between heterozygotes, there is no selection operating against the recessive gene. In this case, the frequency of the recessive gene will increase very gradually at a rate determined solely by the mutation rate, which is usually so low that we would be unlikely to detect an appreciable increase in frequency during a period of, say, 100 years, even in species with short generation length. The final case that we must consider is the conventional scheme of selection against a recessive gene, where heterozygotes are not detected. As shown in Appendix A5.4.4 of *Veterinary Genetics*, it turns out that the equilibrium frequency of a recessive gene under these circumstances is $\sqrt{(\mu/s)}$, and the corresponding equilibrium frequency of the recessive phenotype is μ/s. It is left to the reader to verify that even with weak selection against a recessive gene, the equilibrium frequency resulting from a selection/mutation balance is very low.

A very important practical implication of selection/mutation balance

Let us consider for a moment the most extreme form of autosomal recessive alleles – those that are lethal ($s = 1$). From above, the expected equilibrium frequency of such alleles is simply $\sqrt{\mu}$. A ball-park estimate of μ is 1 in a million (10^{-6}), which gives the estimated equilibrium frequency of a lethal recessive allele as $\sqrt{(10^{-6})}$, which equals 10^{-3} or 1 in a thousand (i.e. 0.001). The frequency of the normal allele is therefore 0.999, and hence (from Hardy–Weinberg) the expected frequency of homozygote normals is $(0.999)^2 = 0.998$. For two loci, each having a lethal allele, the chance of an individual being homozygous for the normal allele at both loci is $(0.998)^2$. For any given number (n) of such loci, the chance of an individual being homozygous for the normal allele at each locus is $(0.998)^n$. What is a realistic value for n? In chapter 1, we saw that the total number of structural genes could be around 22000. To be really conservative, let us assume that, as a result of mutation, there are lethal alleles at a small proportion (say 5000) of the structural genes in a particular population, with each lethal allele maintained at a frequency of around 0.001 by selection/mutation balance. In this case, $n = 5000$, which gives the chance of a member of that population being a non-carrier at all 5000 loci as $(0.998)^{5000} = 0.00004$, which is, for all practical purposes, zero. In other words, almost every animal (and human) that has ever lived has been carrying at least one lethal allele.

We can take these calculations one step further: we can predict the typical number of lethal alleles carried by any individual. This number is the product of the number of loci with lethal alleles and the chance of being a carrier at any locus ($2pq$). Using the above numbers, the expected number of lethal alleles in an individual is $5000 \times 2(0.999)(0.001)$, which is approximately 10. We do not yet have any estimates of this number in animals, but we do have one estimate from humans: in 2008, the entire diploid genome sequence of James Watson (the co-discoverer of the structure of DNA) was published, as mentioned in Chapter 2; and he was shown to be carrying 11 autosomal recessive alleles for inherited disorders!

Copyrighted material

The practical implications of this conclusion should be immediately obvious: it is simply not practicable to think in terms of eliminating all deleterious alleles. In other words, we have to learn to live with harmful alleles. We shall return to this issue in Chapter 21.

Other interesting statistics from the Watson sequence are that he is heterozygous for around 3.3 million SNPs, including around 11 000 that result in an amino-acid substitution in a coding sequence. In other words, Watson has a non-synonymous SNP in approximately half of his structural genes. These statistics serve to reinforce the incredible level of polymorphism that exists in diploid organisms.

Selection favouring heterozygotes

We saw in previous sections that the effect of selection against a dominant gene or against a recessive gene is to reduce the frequency of that gene to a relatively low level, at which stage a balance between mutation and selection maintains an equilibrium. Selection favouring heterozygotes also results in an equilibrium, but of a rather different type. In the extreme case, where no homozygotes pass on their genes, the only parents are heterozygotes, in which case both genes are maintained in the population at a frequency of 0.5. In less extreme circumstances, where both homozygotes have only a partially reduced fitness, the results are exactly the same so long as both homozygotes have equal fitness. Thus, even if both homozygotes have only a 1 per cent reduction of fitness, selection is still acting equally against each gene and maintains an equilibrium gene frequency of 0.5. However, if one homozygote has a lower fitness than the other, selection is less intense against the gene whose homozygote has the highest fitness.

It turns out, as shown in Appendix A5.4.5 of *Veterinary Genetics*, that if s_1 is the selection coefficient against one homozygote, say A_1A_1, and s_2 is the selection coefficient against the other homozygote, say A_2A_2, the equilibrium frequency of the A_2 gene is $s_1/(s_1 + s_2)$.

If for example, one homozygote is only 10 per cent less fit than the heterozygote ($s_1 = 0.10$), while the other one is lethal (100 per cent less fit; $s_2 = 1.00$), the equilibrium frequency of the lethal gene is $0.10/(0.10 + 1.00) = 0.09$. Thus, if heterozygotes for a lethal gene are slightly more fit than homozygotes for the normal gene, the result is an equilibrium frequency much higher than that expected under a selection/mutation balance with a recessive lethal gene. Although definite proof is usually lacking, selection favouring heterozygotes is often suggested as a reason for recessive lethal traits reaching unusually high frequencies. In many situations, it is certainly the most plausible explanation available, and sometimes data are available to back it up. Syndactyly or fusion of digits in Holstein–Friesian cattle, for example, is, in effect, a lethal recessive trait for which there is some evidence of selection favouring heterozygotes. In this case, there is a suggestion that heterozygotes produce higher milk and butterfat yields and thus have a higher fitness as a result of artificial selection. In a similar manner, it appears that carriers of the lethal recessive genes for DUMPS and for weaver syndrome in cattle may be favoured by artificial selection for milk traits. In pigs, carriers of the recessive gene for malignant

Copyrighted material

hyperthermia (see Chapter 9) were favoured by selection for leanness; and there is some evidence that horses carrying the dominant gene for hyperkalaemic periodic paralysis tend to be favoured by judges in the show ring.

When carriers of defective genes are consistently favoured by artificial selection, one explanation is that the gene has a favourable effect on the trait that is being selected, in addition to its negative effect on viability or reproductive ability. The term **pleiotropy** is used to describe situations where a gene affects two or more seemingly unconnected traits. Alternatively, the defective gene could be closely linked to a gene that affects the trait being selected. In either case, the defective gene provides an opportunity for investigating the genetic basis of the desirable trait.

Selection against heterozygotes

Neonatal diarrhoea in piglets is of considerable economic importance. It is often caused by strains of *Escherichia coli* bacteria having a cell-surface antigen called F4, which combines with a receptor on the wall of a piglet's intestines, enabling the bacteria to attach themselves to the intestine. Once attached, they proliferate, releasing enterotoxins and thus producing diarrhoea which can lead to very high mortality. Some strains of *E. coli* lack the F4 antigen (they are said to be F4-negative) and cannot therefore attach themselves to the intestinal mucosa. Being thus unable to proliferate and release enterotoxins, such strains are non-virulent. Certain piglets, however, are not susceptible even to F4-positive bacteria, and it has been found that they lack the appropriate receptor for F4, thus preventing the attachment and subsequent proliferation of bacteria. Lack of the F4 receptor (and hence resistance to the F4-positive bacteria) is the result of homozygosity for a recessive gene called *s*. Piglets having the receptor, and therefore being susceptible, are either homozygous for a dominant gene *S* or are heterozygous.

Because F4-positive *E. coli* are fairly common, and because the susceptibility gene is dominant, we might expect that the susceptibility gene would be maintained at a very low frequency by a selection/mutation balance. However, in four English herds examined in one survey, the frequency of the dominant gene in three herds was much greater than 0.5, and in the other herd it was around 0.4. The most satisfactory explanation for these unexpectedly high frequencies involves a fascinating combination of basic principles from immunology and population genetics.

Consider a population into which the resistance gene *s* has been introduced only recently by mutation and/or migration, and suppose that a build-up of F4-positive bacteria occurs. Because nearly all animals are susceptible, there is very strong selection against the dominant susceptibility gene, i.e. in favour of the recessive resistance gene. At the same time, however, the susceptible sows mount an immune response to the F4-positive bacteria, supplying antibodies to F4 via their colostrum to all their piglets. These antibodies are sufficient to prevent diarrhoea, and so selection against the dominant susceptibility gene becomes less intense not long after it commenced. Thus, the previously expected rapid elimination of the dominant gene does not occur. Suppose that it has, however, been reduced in frequency from 1.0 to, say, 0.7, which means that $(1 - 0.7)^2 = (0.3)^2 = 0.09$ of all sows are now homo-

Copyrighted material

zygous for the resistance gene. These sows certainly carry the F4-positive bacteria, but because attachment does not occur, they never develop antibodies to F4 and hence they provide no protection to their piglets. This is of no consequence to their *ss* piglets because they are naturally resistant anyway. However, not all offspring of *ss* sows are *ss*; some of the *ss* sows mate with *Ss* or *SS* boars and consequently produce some or all *Ss* piglets, all of which are susceptible. These heterozygous piglets get the worst of both worlds; they have a receptor for F4, and they fail to receive antibody from their *ss* female parent.

Thus we have selection against heterozygotes, with both homozygotes having equal and normal fitness, in the case of *SS* because of antibodies received from the sow, and in the case of *ss* because of lack of a receptor for F4. Selection against heterozygotes in this situation is only partial, because only those born to *ss* sows are likely to be affected by diarrhoea. What are the likely effects of selection against heterozygotes?

A moment's reflection indicates that for every heterozygote eliminated from a population, an equal number of both genes is removed, which has a greater effect on the less common gene. Consider, for example, a herd of 100 pigs which has Hardy–Weinberg genotype frequencies corresponding to gene frequencies of 0.7 and 0.3 before selection. The numbers of the genotypes are 49 *SS*, 42 *Ss*, and 9 *ss*. Suppose that 10 heterozygotes die from neonatal *E. coli* diarrhoea, leaving 49 *SS*, 32 *Ss*, and 9 *ss* in a total of 90 pigs after selection. The gene frequencies are now $[(2 \times 49) + 32]/(2 \times 90) = 0.72$ for *S* and $1 - 0.72 = 0.28$ for *s*, which represents a reduction in frequency of the less common gene.

In general, if homozygotes have equal fitness, selection against heterozygotes results in a decrease in frequency of the less frequent gene. Thus, the gene that was at a frequency of less than 0.5 when selection against heterozygotes commenced, gradually decreases in frequency so long as such selection continues.

In the piglet case, selection against heterozygotes ceases when the pathogenic *E. coli* disappear, leaving the frequency of the *s* gene at whatever level it had reached at that time. With the next outbreak of neonatal *E. coli* diarrhoea, the cycle of events described above recommences. If the *s* gene had increased in frequency to more than 0.5 before selection against heterozygotes became effective, such selection would increase its frequency even further. If it were less than 0.5, it would decrease in frequency towards its former low level. If a large number of herds were surveyed, it might be expected that the *s* gene would be at a low frequency in some and at a high frequency in others. Results of the very limited surveys reported to date are consistent with this prediction.

Genetic drift and the founder effect

There is another reason why deleterious and even lethal conditions can reach quite high frequencies in certain populations. It is called **genetic drift**, which, as we saw earlier in this chapter, refers to changes in gene frequency that are entirely caused by chance. These changes result from the sampling of finite numbers of genes that is inevitable in all finite populations. Because these changes in gene frequency are

Copyrighted material

entirely the result of chance, their direction is random and is completely outside the control of humans. All real populations are finite, so it follows that genetic drift occurs in all real populations. However, the larger the population size, the smaller is the magnitude of genetic drift.

The most extreme case of genetic drift in domestic animals can be illustrated by imagining that only one male and one female are chosen to be parents of the next generation. Consider a single locus with two alleles *A* and *B*. Irrespective of what the frequency of *A* was in the generation from which the parents were chosen, its frequency in the actual parents must be either 0.00 (if both parents happen to be homozygous for *B*), 0.25 (if one parent is *AB* and the other is *BB*), 0.50 (if the parents are *AB* and *AB*, or *AA* and *BB*), 0.75 (*AA* and *AB*), or 1.00 (both parents homozygous for *A*). Suppose that the frequency of *A* was 0.10 in the population from which the parents were chosen, and suppose that the parents were chosen at random, and happened to be *AA* and *AB*. In this case, the gene frequency has changed from 0.10 to 0.75, a change of 0.65, which is very large and which is entirely the result of chance or sampling. As the gene frequency in the parents is now 0.75, it follows that the gene frequency in their offspring is expected to be 0.75, unless selection, mutation, or migration causes a further change. If only one male and one female are chosen at random from the offspring generation to be parents of the next generation, exactly the same situation applies: the gene frequency in the next generation will be either zero, 0.25, 0.50, 0.75, or one. If it reaches zero, the gene is **lost** forever, or until mutation recreates it. If the frequency reaches one, the gene is said to have been **fixed** in the population, and all other genes at that locus have been lost.

The **founder effect** is a special case of genetic drift, referring to situations where a small number of individuals move to a new site, and commence (found) a new population. Even if the new population rapidly expands in numbers, and soon has a large number of parents in each generation, the gene frequencies will be a reflection of the frequencies in the founders, rather than in the population from which the founders came. The reason for this is that, in the absence of selection, mutation, migration, and further genetic drift, the gene frequencies remain at whatever level they were when the population was started. Thus a population that is founded by just one male and one female can have a very different set of gene frequencies from that in the population from which the founding parents were chosen.

The founder effect is an example of a **population bottleneck**, which is a situation in which the number of parents in a population becomes very small for one or more generations. As illustrated above, a population emerging from a bottleneck may be very different from the population that existed before the bottleneck.

Extending population genetics to more than one locus

With populations of animals now being genotyped for tens and even hundreds of thousands of DNA markers, we have to extend our understanding of population genetics to include the joint consideration of more than one locus at a time. In fact,

Copyrighted material

we have to start thinking in terms of the set of alleles that reside in a gamete. Such a set is called a **haplotype** (for *hap*loid geno*type*). Consider two loci, each with two alleles in a population: locus A with alleles A_1 and A_2 at frequencies p_A and q_A; and locus B with alleles B_1 and B_2 at frequencies p_B and q_B. It should be evident that there are four possible gametes or haplotypes: A_1B_1, A_1B_2, A_2B_1, and A_2B_2. What are the expected frequencies of these four haplotypes in the population? One possible situation would be where there is a random association between the alleles at the two loci, so that the frequency of each haplotype equals the product of the relevant allele frequencies, e.g. the frequency of haplotype $A_1B_2 = p_Aq_B$ etc. Where this occurs, we say there is **gametic equilibrium** (sometimes called **linkage equilibrium**). Another way to look at this is to say that with gametic equilibrium, the A_1 allele occurs with the same frequency in haplotypes containing the B_1 allele as in haplotypes carrying the B_2 allele.

The alternative situation is for there to be a non-random association between the alleles at the two loci. This is called **gametic disequilibrium** (or **linkage disequilibrium**, often abbreviated to **LD**). For example, allele A_1 could be associated with allele B_1 more frequently than expected on the basis of the product of their allele frequencies. In extreme cases of disequilibrium, each allele at one locus occurs only with one of the alleles at the other locus, and never with the other allele, e.g. A_1 occurs only with B_1.

Disequilibrium is quantified as the difference, D, between the observed frequency of a haplotype and its expected frequency at equilibrium, i.e. $D = \text{freq}(A_1B_1) - p_Ap_B$. However, because D is heavily dependent on allele frequencies, it is not readily comparable between different pairs of loci or between different populations. To address this limitation, D is often divided by the maximum value (D_{max}) that it can take for a given set of allele frequencies, to give D' (called D prime), i.e. $D' = D/D_{max}$. Another way to address the dependence of D on allele frequencies is to 'standardize' D to $r^2 = D^2/(p_Aq_Ap_Bq_B)$. Thus, D' and r^2 are the two standard parameters of disequilibrium.

Disequilibrium can arise if selection favours particular combinations of alleles at different loci, or it can be generated by mutation, migration, or genetic drift. Once it has arisen, D decreases over generations at a rate determined by the recombination fraction (c) between the two loci. Specifically, D declines by a proportion $(1 - c)$ every generation. Readers will recall from Chapter 1 that the maximum value of c is 0.5, even for loci located on separate chromosomes. This means that if two loci are completely unlinked, the magnitude of disequilibrium between them decreases by one-half per generation. Consequently, we expect very little disequilibrium between loci on different chromosomes. For pairs of loci on the same chromosome, the more closely they are linked, the slower will be the decrease of disequilibrium over generations.

The question then arises: in typical populations of animals, for what distance along a chromosome does disequilibrium extend? A popular criterion is $r^2 = 0.2$, i.e. how far apart do two loci have to be before r^2 falls below 0.2? There is much variation in this distance at different locations within the genome, and among breeds. To generalize: within cattle and pig breeds, preliminary estimates place the distance at around 50 kb; in chickens, the distance is shorter, around 10 kb. When measured

Copyrighted material

across a mixture of many dog breeds, the distance is somewhat similar to that in chickens – around 10 kb. However, within dog breeds, this distance can extend much further – to around 2 Mb in some breeds. This high level of disequilibrium within dog breeds reflects the relatively recent origin of most breeds and the relatively small number of animals on which most breeds were founded (i.e. many dog breeds were founded via a population bottleneck).

Why are researchers interested in disequilibrium? The simplest way to answer is to consider a situation where one locus is a DNA marker and another locus is a gene of economic importance for which a genotype test is not available. If there is disequilibrium between the two loci, then it will be possible to apply indirect selection on the gene of economic importance by selecting on the DNA marker. Recalling that disequilibrium decreases at $1 - c$ per generation, the utility of a DNA marker depends on how close it is to the gene of economic importance: the closer the marker, the greater its utility. Viewed more broadly, the greater the extent of disequilibrium, the larger are the haplotypes that are passed on from generation to generation, and hence the smaller is the number of DNA markers that are needed to 'tag' all the halotypes in a genome. As we shall see in later chapters, this information is central to the application of SNP chips in the discovery of the molecular basis of single-locus disorders, and in their use as aids to selection for productive traits in livestock and for inherited disorders in all animals.

Summary

- For most loci in most populations, mating is at random
- One generation of random mating results in genotype frequencies that are a simple function of gene frequencies: p^2 *AA*, $2pq$ *AB*, q^2 *BB*; the so-called Hardy–Weinberg frequencies
- This natural phenomenon (called the Hardy–Weinberg law) provides predictions of genotype frequencies and hence of phenotype frequencies that often have considerable practical importance
- The law shows, for example, that the frequency of carriers of genes for rare recessive diseases is much higher than the frequency of the disease itself; and it explains why the frequency of traits caused by X-linked recessive genes is much higher in males than in females
- Selection acts on phenotypes and occurs whenever some phenotypes contribute more offspring to the next generation than do other phenotypes; to the extent that phenotypes result from the action of genes, selection may alter gene frequency
- The immediate effect of complete selection against a dominant gene is to remove it from the population altogether; in contrast, selection against a recessive gene is much less effective
- The elimination of genes from a population as a result of selection is often opposed by mutation, which continually introduces new versions of the same

Continued

Copyrighted material

gene; a selection/mutation balance usually results, in which defective genes are permanently maintained at relatively low frequencies
- An important practical implication of the selection/mutation balance is that almost every animal is carrying at least one lethal allele
- If selection favours heterozygotes, a stable equilibrium also results, but usually with gene frequencies less extreme than in the case of a selection/mutation balance; selection favouring heterozygotes is a possible explanation for the rapid spread of an undesirable recessive gene
- The existence of two or more alleles at a locus in a population is called polymorphism
- The smaller the number of gametes passed from one generation to the next, the greater is the effect of random sampling (or chance) on the frequency of genes in the resultant offspring
- Change in gene frequency due to chance is called genetic drift
- A particular case of genetic drift is the founder effect, the undesirable effects of which are seen when a new population is founded with a small number of individuals, some or most of which by chance carry undesirable genes that are relatively rare in the population from which they came
- When two loci are considered together, it is often observed that an allele at one locus is associated with an allele at the other locus more or less often than would be predicted by the product of their frequencies; such loci are said to be in gametic or linkage disequilibrium (LD)
- Unlike the situation for a single locus considered on its own, one generation of random mating does NOT create linkage equilibrium
- In fact, under random mating, disequilibrium decreases by $1 - c$ per generation, where c is the recombination fraction between the loci
- If one locus is a DNA marker and the other locus is a gene of economic importance, the existence of disequilibrium between the two loci enables indirect selection on the gene of economic importance by selection on the DNA marker
- Because disequilibrium decreases at the rate of $1 - c$ per generation, the utility of a DNA marker depends on how close it is to the gene of economic importance: the closer the marker, the greater its utility

Further reading

Amaral, A.J., Megens, H.-J., Crooijmans, R.P.M.A., Heuven, H.C.M. and Groenen, M.A.M. (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics*, **179**, 569–79.

Andreescu, C., Andndano, S., Brown, S.R., Hassen, A., Lamont, S.J. and Dekkers, J.C.M. (2007) Linkage disequilibrium in related breeding lines of chickens. *Genetics*, **177**, 2161–9.

Copyrighted material

Charlesworth, B. (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Hartl, D.L. and Clark, A.G. (2007) *Principles of population genetics*, 4th edn. Sinauer Associates, Sunderland, Massachusetts.

Hedrick, P.W. (2010) *Genetics of populations*, 4th edn. Jones and Bartlett, Boston.

Hurst, L.D. (2009) Fundamental concepts in genetics: genetics and the understanding of selection. *Nature Reviews Genetics*, **10**, 83–93.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A., Barris, W., Schnabel, R.D., Taylor, J.F. and Raadsma, H.W. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein–Friesian cattle based on a high-density SNP panel. *BMC Genomics*, **9**, 187.

Slatkin, M. (2008) Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, **9**, 477–85.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J. R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A. and Rothberg, J.M. (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–6.

Copyrighted material

6 Familial disorders not due to a single gene

Many important disorders are familial – they 'run in families', which means that the incidence among relatives of affected animals is greater than the incidence in the population to which they belong. If a disorder is familial, it may be the result of shared environment or shared genes, or a combination of shared environment and shared genes.

With single-gene disorders of the type discussed in previous chapters, there is no difficulty in determining why they are familial: it is because of shared genes. There are, however, many familial disorders that do not fit into any of the categories of single-gene inheritance.

The aim of this chapter is to explain what is known about such disorders.

Liability and threshold

For reasons that will soon become evident, it is convenient to think of each animal as having a certain **liability** to a disorder, where liability refers to the combined effect of all factors, both environmental and genetic, that render an animal more or less likely to develop that disorder.

It must be clearly understood that the term 'environment' is used here and throughout this book in its widest possible sense, literally meaning non-genetic. In fact, an environmental factor is any factor that cannot be attributed to the Mendelian aspects of genes in the nucleus.

Liability is a continuous variable that could, in principle, be measured on a continuous scale, in the same manner as we measure a trait such as body weight. In this chapter, however, we shall introduce the concept of liability in the context of animals being grouped into one of two classes: affected and normal. We can accommodate this so-called all-or-none classification on a continuous scale by using the concept of **threshold**, which is a certain level of liability above which animals develop the disorder, and below which animals are normal.

The multifactorial model

We shall start this section by considering one of the best known familial disorders, namely hip dysplasia in dogs.

Copyrighted material

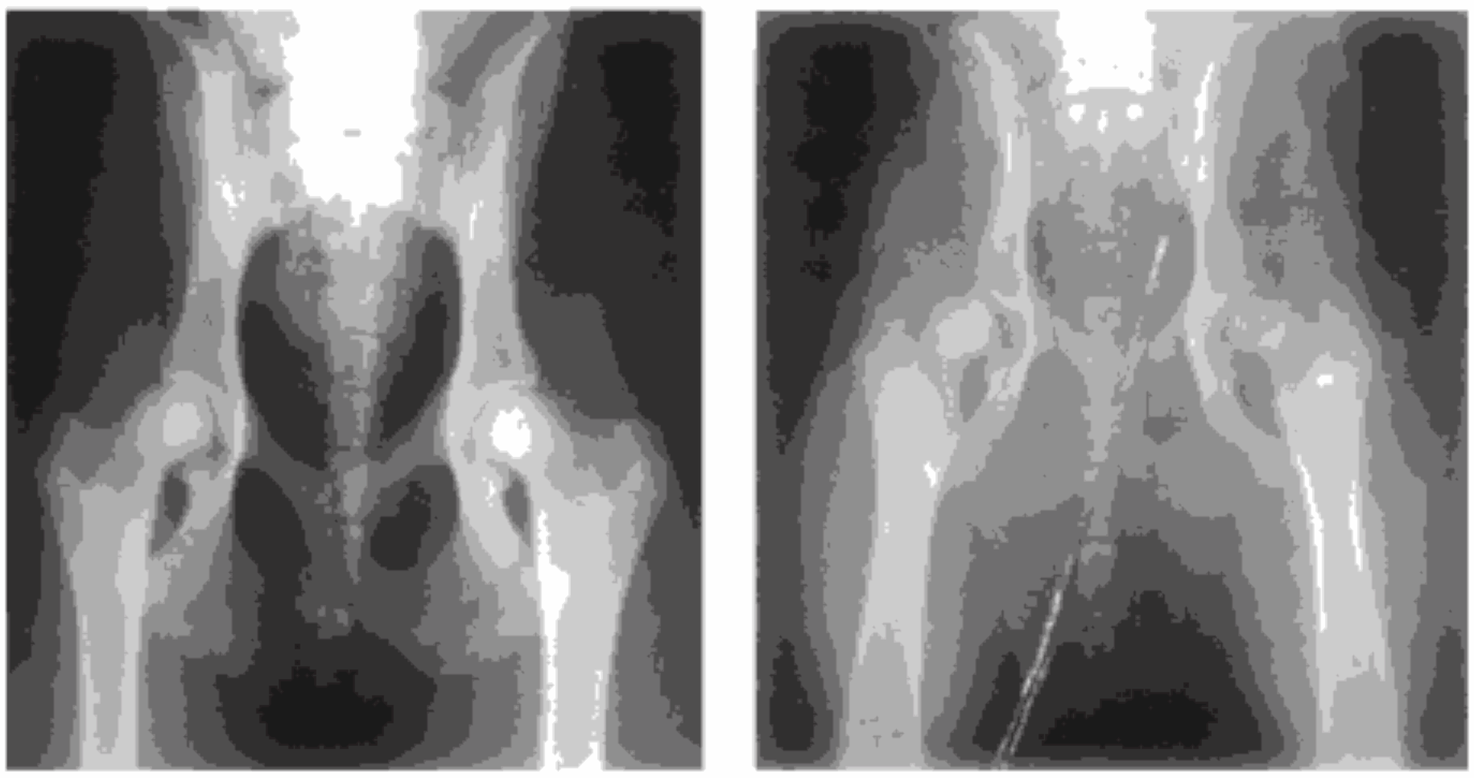


Fig. 6.1 Radiographs of (*left*) normal canine hip joints and (*right*) hips showing bilateral coxofemoral subluxation due to joint instability or laxity, which is the earliest sign of canine hip dysplasia. (Reproduced courtesy of G. Allan and R. Zammit.)

Hip dysplasia is usually diagnosed by subjective evaluation of a radiograph (Fig. 6.1) or by palpation, so misclassification is one obvious non-genetic factor that will contribute to liability. In addition, level of feeding, dietary electrolyte balance, and extent of exercise during early growth have been shown to affect an animal's liability to hip dysplasia. Liability could also be partly the result of the action of alleles at any locus whose gene product is involved in the development of the hip joint.

The situation is now like that shown in Fig. 6.2, where there is a bell-shaped, or normal, distribution of liability. Animals that are over-fed and over-exercised, and happen to have alleles tending to produce ill-fitting hip joints, have a relatively high liability. Animals with only some of these predisposing factors are more common than those with all the predisposing factors, which accounts for the increasing frequency of animals with liabilities closer to the mid-point of the distribution. A similar trend is expected on the other side of the mid-point: animals given the most favourable diet and most appropriate exercise regime, and who also happen to have alleles that tend to produce sound hip joints, have much lower liabilities. But such animals are less common than those with only some of these factors operating to decrease liability.

Despite the continuous nature of liability to hip dysplasia, historically dogs were often classified as either normal or affected, by thinking in terms of a threshold like that in Fig. 6.2.

When a trait is determined by the combined effect of many factors, both environmental and genetic, it is said to be **multifactorial**.

One virtue of thinking of such disorders as being multifactorial is that we can estimate the relative importance of genetic and environmental factors as contributors to the disorder's aetiology. This is done by estimating a parameter called **heritability**, which for present purposes is the proportion of total variation in a trait that can be attributed to variation in genetic factors.

Copyrighted material

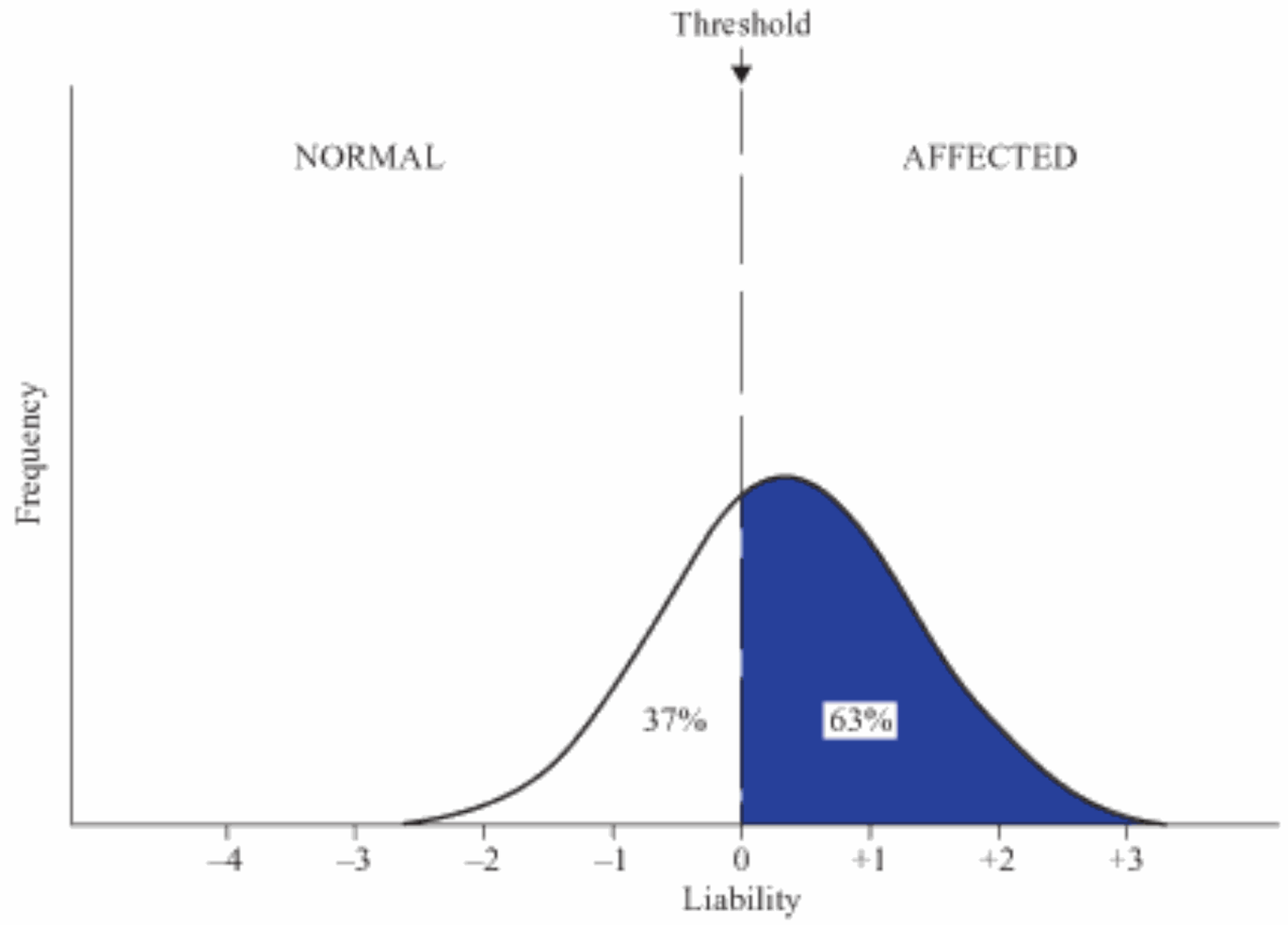


Fig. 6.2 The concepts of liability and threshold for a multifactorial disorder with an incidence of 63 per cent.

In the present case, the trait is liability. Variation in liability refers to the differences among animals in liability to a particular disorder. Heritability of liability, then, is the proportion of differences in liability that are the result of genetic differences among animals.

The majority of familial disorders have an intermediate heritability, which indicates that both environmental and genetic factors contribute to their aetiology.

More than one threshold

Congenital heart disease occurs in many forms, the most important of which are familial in humans and in animals. In an extensive and very thorough set of analyses conducted by Don Patterson and colleagues, the inheritance of congenital heart disease in dogs has been clearly demonstrated to be compatible with a multifactorial model.

The most common congenital heart disease is patent ductus arteriosus (PDA), resulting from defective closure of the ductus arteriosus.

Like many other multifactorial disorders, defective closure of the ductus arteriosus is a graded phenomenon, with increasing severity corresponding to increasing liability.

This is clearly indicated in the results of various matings among Poodles, which produced some normal offspring, others with partial closure (called ductus diverticulum or DD), and others with PDA. The presence of these three grades can be represented by two thresholds, as illustrated in Fig. 6.3a. In this case, the PDA × PDA

Copyrighted material

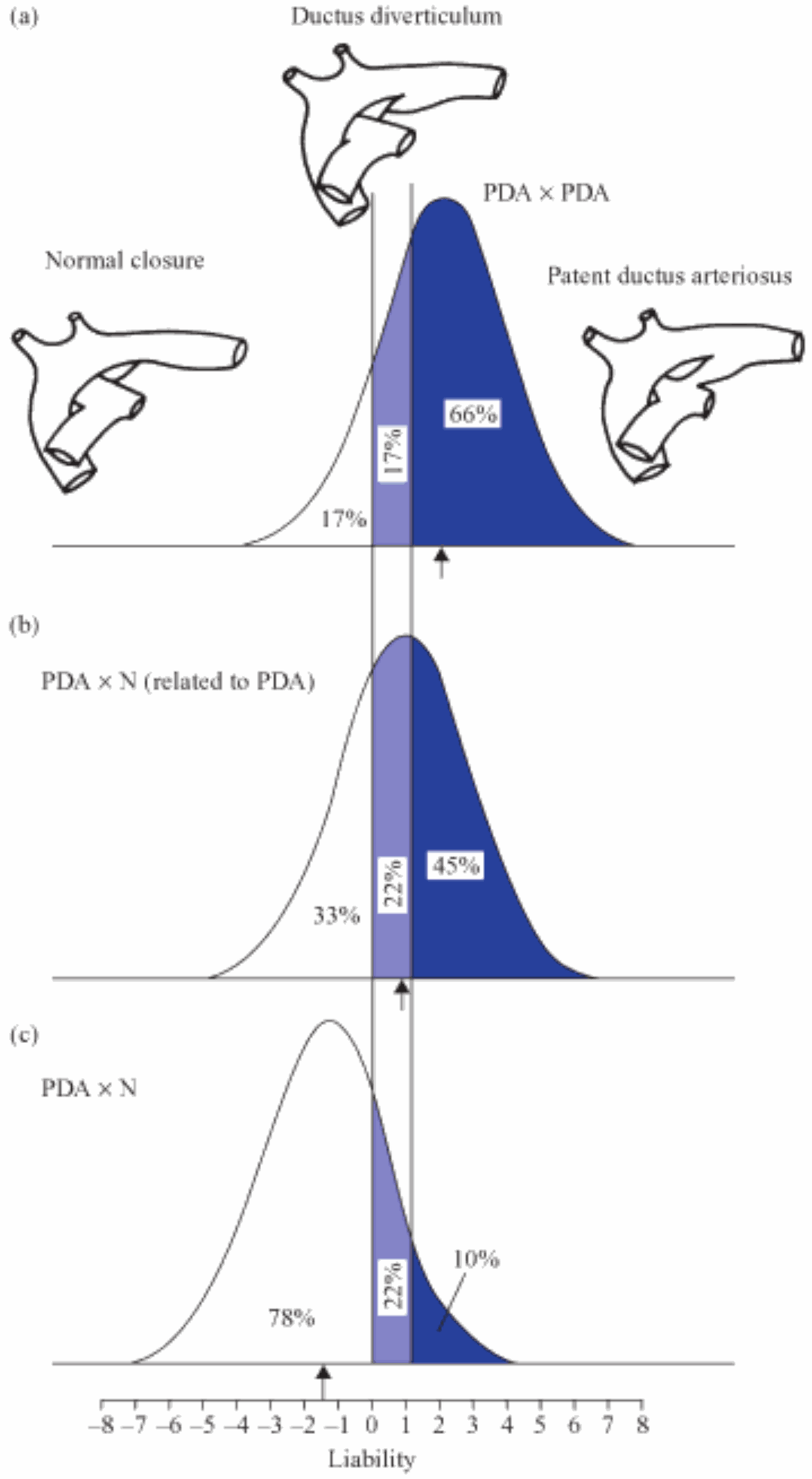


Fig. 6.3 Graphical representation of three populations of Poodles for patent ductus arteriosus (PDA): (a) from PDA × PDA matings; (b) from PDA × normal matings, where the normal parent has a close relative that is affected, e.g. parent or offspring or full-brother or full-sister; (c) from PDA × normal matings, where the normal parent has no close relatives that are affected. The position of the distributions relative to the thresholds is determined solely by the incidence of each grade of defect. Vertical arrowheads indicate means of distributions. (Adapted with permission from Patterson *et al.* (1972), in D. Bergsma (ed.) *The Cardiovascular System*, part XV, Williams & Wilkins, Baltimore, for the National Foundation – March of Dimes, BD:OAS, 8(5), 160–74.)

Copyrighted material

matings produced 66 per cent of offspring with PDA, 17 per cent with DD, and 17 per cent normal offspring.

An even better understanding can be obtained from other types of matings, the results of which are also given in Fig. 6.3.

From a practical point of view, the difference in incidence of affected offspring from the two types of matings involving normal dogs is particularly important: normal dogs that are closely related to affected dogs (Fig. 6.3b) produce more than three times the incidence of affected offspring, and more than four times the incidence of severely affected offspring, when compared with normal dogs that are unrelated to affected dogs (Fig. 6.3c).

In general, the tendency for normal animals to throw affected offspring and the severity of the disorder among their affected offspring depend on how closely the normal animal is genetically related to an affected animal.

Another relatively common form of congenital heart disease, conotruncal septum defects (CSD) in Keeshond dogs, clearly illustrates another important implication, namely that the frequency and severity of the disorder will be greatest among relatives of more severely affected animals. Conotruncal septum defects are particularly useful in this context, as they can be divided into four different grades of increasing severity, as illustrated in Fig. 6.4.

From an extensive set of matings among Keeshonds with various grades of CSD, the overall incidence of CSD and the incidence of each grade of CSD in offspring were determined for various mating types, according to the average CSD grade of the parents. The results for overall CSD incidence and for incidence of the most severe form (grade 3) are shown in Fig. 6.5. It is evident that the overall frequency of CSD and the frequency of grade 3 are almost directly proportional to the average severity in parents. These results are a good example of what is commonly seen with multifactorial disorders, namely that the frequency and severity in relatives of affected animals are expected to increase as liability increases.

Some final points

Heritability

The concept of heritability was introduced with the concept of multifactorial traits, and for this reason it might be thought that heritability is a valid concept only in relation to multifactorial traits. But this is not so: it is, in fact, an equally valid parameter for any trait (including disorders) determined by a single locus.

As introduced earlier in this chapter, heritability is 100 per cent for all traits determined by a single locus, because all the differences in phenotype between animals in relation to those traits are the result of different genotypes at the locus in question. When heritability is defined in this sense, we refer to it as **heritability in the broad sense** or the **degree of genetic determination**.

There is, however, another sense in which the term heritability is used. Known as **heritability in the narrow sense**, it expresses the extent to which phenotypes

Copyrighted material

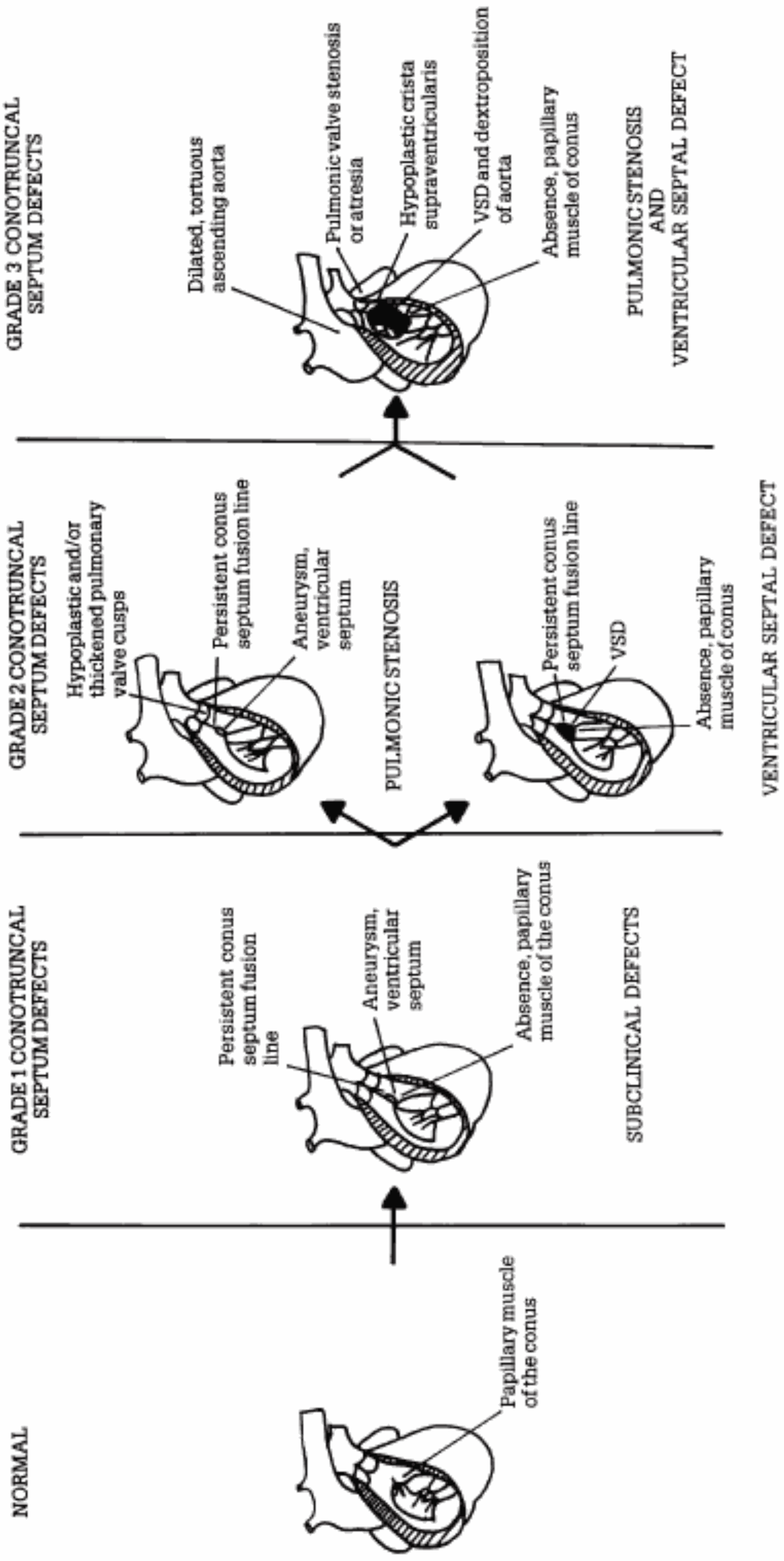


Fig. 6.4 A graded series of defects of the conotruncal septum in Keeshond dogs. (Reprinted by permission from Elsevier: *American Journal of Cardiology*, Patterson *et al.*, 1974, **34**(2), 187–205.)

Copyrighted material

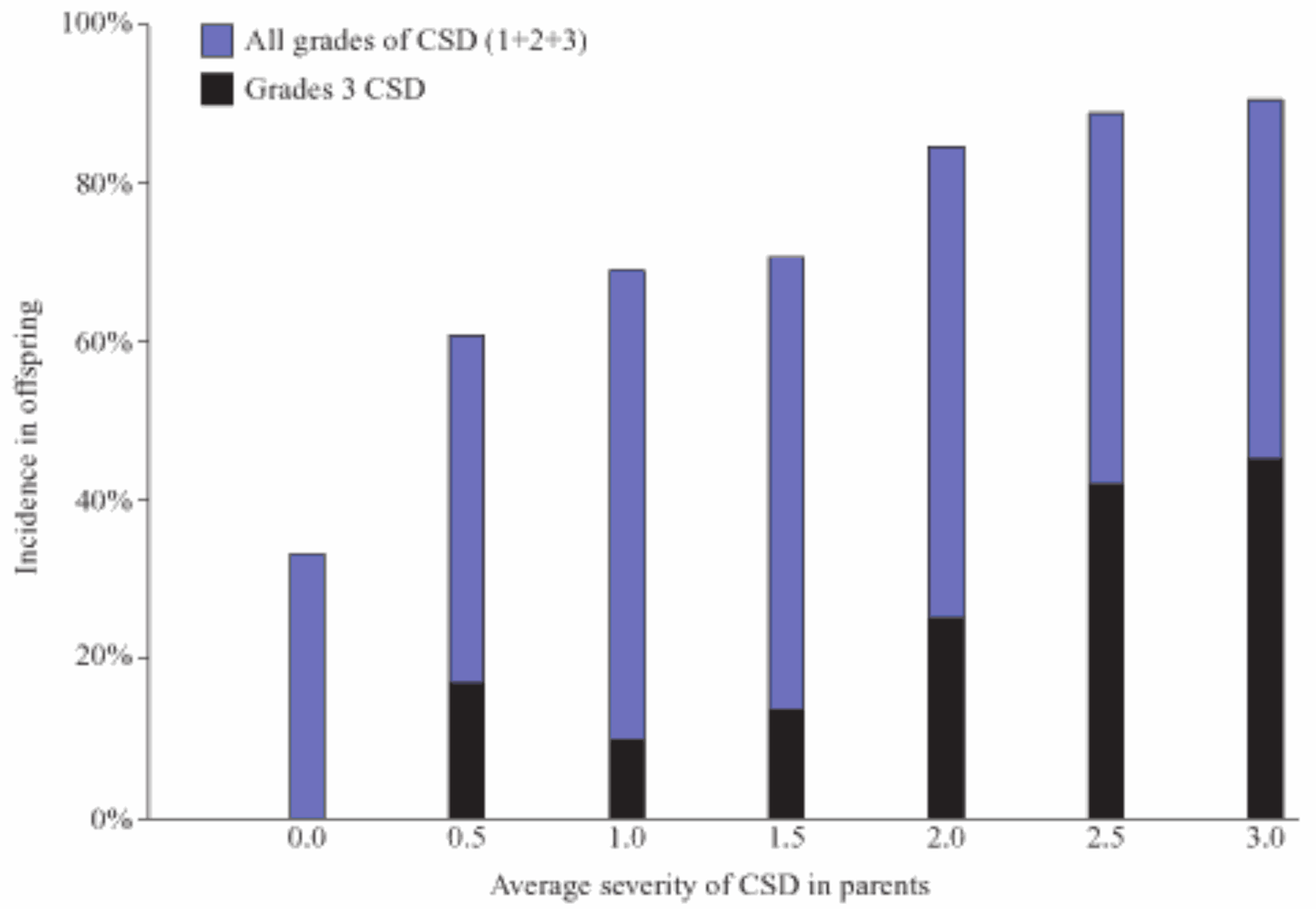


Fig. 6.5 Overall incidence of conotruncal septum defects or CSD (grades 1, 2, and 3) and incidence of grade 3 CSD in offspring, as a function of average severity of CSD in parents.

are transmitted from parents to offspring. In other words, it expresses the extent to which offspring resemble their parents, or in more general terms, the extent to which relatives resemble each other.

Because we are most interested in understanding and predicting how traits are transmitted from one generation to the next, narrow-sense heritability is much more useful than broad-sense heritability. In fact, whenever we see the word 'heritability' on its own, we can take it to mean narrow-sense heritability, unless otherwise stated. The concept of heritability is explained fully in Chapter 13.

Estimating heritability involves using statistical techniques to estimate the extent to which relatives resemble each other for the trait of interest, compared with unrelated animals. Table 6.1 lists some representative values for heritability of liability for a range of disorders in domestic animals. The practical importance of this table is that if the heritability of liability is greater than zero, it is possible to decrease the incidence of the disorder by selection; and the higher the heritability, the greater the response to selection. The use of selection to control inherited disorders is discussed in Chapter 21.

Penetrance and expressivity

From time to time, multifactorial disorders such as hip dysplasia have been claimed to be single-locus disorders. In order to accommodate this claim with the fact that

Copyrighted material

Table 6.1 Representative values of heritability of liability, expressed as a percentage. Since there is considerable variation in actual estimates, the values have been rounded to the nearest 5 per cent

Cattle	
Bloat	10
Cystic ovaries	10
Displaced abomasum	10
Foot problems	10
Ketosis	20
Leg problems	10
Metritis	5
Milk fever	40
Retained placenta	5
Chicken	
Tibial dyschondroplasia	45
Dog	
Calcified intervertebral discs	20
Elbow osteochondrosis	50
Hip dysplasia	35
Duck	
Chronic interstitial nephropathy	45
Horse	
Osteochondrosis in the tibiotarsal joint	50
Bony fragments in the palmar plantar portion of the metacarpophalangeal and metatarsophalangeal joints	20
Pig	
Cryptorchidism	55
Hernia	50
Turkey	
Tibial dyschondroplasia	15

the presence or absence of hip dysplasia does not show single-locus inheritance, and that hip dysplasia occurs in a continuous range of severity (from normal to severely affected), the concepts of **incomplete penetrance** and **variable expressivity** have to be invoked. A genotype is said to be incompletely penetrant if less than 100 per cent of individuals with that genotype exhibit the phenotype with which it is normally associated. A genotype exhibits variable expressivity if individuals with that genotype differ in the extent to which they express the phenotype associated with that genotype. These two concepts can be useful for describing disorders for which there is clear evidence of segregation of a single gene of large effect on liability, but in which other factors (both genetic and non-genetic) occasionally exert an effect. For familial disorders, however, such as those discussed in this chapter, claiming single-locus inheritance is not helpful. The disadvantage of thinking in terms of a single gene if the inheritance data clearly do not fit single-gene inheritance, is that we continue to think in terms of carriers of a single gene, and we are therefore inclined to believe that if we can eradicate that gene, then the disorder will disappear. This approach will lead to great disappointment for breeders. On the other

Copyrighted material

hand, by regarding such disorders as multifactorial, we can forget about detecting carriers or removing a gene from a population, and concentrate instead on fruitful tasks. For such disorders, our willingness to use an affected animal in any mating will depend solely on the severity of its own disorder and on its relationship to other defective animals. We will not waste our time searching for Mendelian ratios. Instead, we will simply proceed with various matings, using as our guide the incidence of affected animals achieved in previous matings of that type.

In addition, by thinking in multifactorial terms for such disorders, we will avoid the common misconception that if the inheritance of a disorder is not compatible with simple dominant or simple recessive inheritance, its mode of inheritance is not known. In fact, the mode of inheritance is known, and appropriate breeding plans for the reduction in incidence of a particular disorder can easily be drawn up and put into practice on the basis of that understanding (as explained in Chapter 21).

Future resolution of multiple factors

A major aim of research into multifactorial disorders is to identify the important factors, both environmental and genetic, that contribute to variation in liability. Considerable progress has been made in this area, especially in regard to environmental factors. Much more progress can be expected in the future. On the genetic side, for example, much effort is being devoted to searching for single genes that make a major contribution to variation in liability. This is done by analysing the occurrence of the disorder in families (called **complex segregation analysis**; see Chapter 7). Unfortunately, this is an inexact science, and many claims of single genes being identified have not stood the test of time. However, as described in Chapter 7, the methods of analysis are being improved, and the search will be much more effective in the future. Increasingly, members of families in which a disorder occurs are being genotyped for tens of thousands of DNA markers via a SNP chip, enabling a genome-wide association (GWA) study to be conducted between the markers and the disorder phenotype. This leads to the identification of chromosomal regions that contribute to genetic variation for liability (called quantitative trait loci, QTL – see Chapter 13), and ultimately to the identification of genes that play important roles in determining variation in liability to the disorder.

It is important to note that the discovery of a single gene or even several genes that contribute to variation in liability for a disorder does not invalidate the multifactorial concept, and does not remove the need for control programmes based on the multifactorial concept. However, knowledge of DNA markers associated with liability, and of genes that contribute to liability, can be used to increase the effectiveness of control programmes, as described in the final section of Chapter 15.

Recurrence risks

On many occasions in this chapter, we have spoken about the percentage of affected offspring resulting from a certain type of mating. These percentages are called **recurrence risks**, because they give an indication of the risk of a particular disorder

Copyrighted material

recurring (i.e. occurring again), should that type of mating be repeated. The recurrence risks presented in this chapter are called **empirical recurrence risks** because they have been obtained by observing results of actual matings. Empirical recurrence risks can be very useful in planning and conducting breeding programmes aimed at reducing the frequency of certain disorders. For example, the figures for incidence of CSD shown in Fig. 6.5, which are really empirical recurrence risks, clearly show that the frequency of all grades of CSD will be decreased most rapidly by breeding only from parents that have no signs of CSD. If this is not possible, for example because the incidence of CSD is very high in a certain population of dogs, the empirical recurrence risks in Fig. 6.5 indicate that breeders should aim for matings in which the average severity of CSD is as low as possible.

There are, however, limitations to empirical recurrence risks, e.g. they cannot indicate the risk for a type of mating from which no offspring records are currently available.

These disadvantages can be overcome by using **theoretical recurrence risks**, which are predictions arising directly from fitting to the data either a single-gene model or a multifactorial model, whichever is appropriate. Because they are predicted from the model, theoretical recurrence risks can be calculated for any imaginable type of mating, whether or not such matings have ever occurred before. In the simplest cases, theoretical recurrence risks equal segregation ratios for single-gene models, and, for the multifactorial model, they can be calculated from the population incidence and heritability. In most cases, however, the calculation of theoretical recurrence risks involves complications that are beyond the scope of this book, and should be left to people with experience in this area.

Summary

- For the many familial defects and diseases (disorders) that do not follow simple Mendelian patterns of inheritance, it is useful to think in terms of multifactorial inheritance, in which liability to the disorder is determined by the combined effect of at least several factors, some of which are genetic and some of which are non-genetic (environmental)
- Liability varies on a continuous scale from low to high, reflecting increasing severity of the disorder
- The implications of multifactorial inheritance are:
  - the more severely an individual is affected, the more frequent and severe will be the disorder in the relatives, including offspring, of that affected individual
  - among normal individuals, the lower their genetic relationship with affected individuals and the greater the proportion of their relatives that are normal, the less frequent and severe will be the disorder in future relatives, including offspring
- Among the many familial disorders that can be considered as multifactorial are hip dysplasia and various forms of congenital heart disease

Copyrighted material

Further reading

Bruneau, B.G. (2008) The developmental genetics of congenital heart disease. *Nature*, **451**, 943–8.

Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to quantitative genetics*, 4th edn. Chapter 18. Longman, London.

Ginja, M.M., Silvestre, A.M., Gonzalo-Orden, J.M. and Ferreira, A.J. (2009) Diagnosis, genetic control and preventive management of canine hip dysplasia: A review. *Veterinary Journal* (epub ahead of print: doi: 10.1016/j.tvjl.2009.04.009).

Janutta, V. and Distl, O. (2008) Review on canine elbow dysplasia: pathogenesis, diagnosis, prevalence and genetic aspects. *Deutsche Tierarztliche Wochenschrift*, **115**, 172–81.

Zhang, Z., Zhu, L., Sandler, J., Friedenberg, S.S., Egelhoff, J., Williams, A.J., Dykes, N.L., Hornbuckle, W., Krotscheck, U., Moise, N.S., Lust, G. and Todhunter, R.J. (2009) Estimation of heritabilities, genetic correlations, and breeding values of four traits that collectively define hip dysplasia in dogs. *American Journal of Veterinary Research*, **70**, 483–92.

Zhu, L., Zhang, Z., Feng, F., Schweitzer, P., Phavaphutanon, J., Vernier-Singer, M., Corey, E., Friedenberg, S., Mateescu, R., Williams, A., Lust, G., Acland, G. and Todhunter, R. (2008) Single nucleotide polymorphisms refine QTL intervals for hip joint laxity in dogs. *Animal Genetics*, **39**, 141–6.

Zhu, L., Zhang, Z., Friedenberg, S., Jung, S.W., Phavaphutanon, J., Vernier-Singer, M., Corey, E., Mateescu, R., Dykes, N., Sandler, J., Acland, G., Lust, G. and Todhunter, R. (2009) The long (and winding) road to gene discovery for canine hip dysplasia. *Veterinary Journal*, **181**, 97–110.

Copyrighted material

7 Is it inherited?

'Is it inherited?' is a common question asked in relation to a wide range of disorders. Unfortunately, the answer for the majority of disorders is 'we don't know', because insufficient data have been collected and analysed to enable a decision to be made.

The situation is gradually improving as more studies are conducted. In order to conduct such studies, an understanding is required of Mendelian inheritance (Chapter 1), population genetics (Chapter 5), multifactorial traits (Chapter 6), and elementary statistics. The aim of this chapter is to show how knowledge in these areas can be put to practical use in investigating the mode of inheritance of a disorder.

It is important that such studies be conducted, not only from the immediate veterinary point of view, but also because medical researchers are increasingly on the lookout for inherited diseases of animals that could serve as models of human disease.

General evidence for a genetic aetiology

If genes make any contribution to the aetiology of a disorder, then, as discussed in Chapter 6, it follows that there will be a positive relationship between the chance of an individual being affected and the extent to which that individual has genes in common with affected individuals.

The most important practical implication of this relationship is seen in families and breeds. Members of the same family have more genes in common than members of different families, and members of the same breed have more genes in common than members of different breeds. If the incidence of a disorder is higher in some families than in others within a breed, or if the incidence is higher in some breeds than in others, one possible explanation is that there is a genetic contribution to the aetiology of the disorder.

Of course, evidence of variation in incidence between families and/or between breeds does not constitute proof of a genetic contribution to aetiology, because non-genetic factors common to members of a family and/or common to members of a breed may be sufficient to account for the observed variation in incidence. The first step, then, in attempting to disentangle non-genetic causes from genetic causes, is to remove or allow for the effects of identifiable non-genetic factors. This can be done only after a detailed investigation of all non-genetic factors that are thought

Copyrighted material

likely to affect the occurrence of the disorder. The trouble with this approach is that the investigator often has no prior indication as to which factors should be examined: each one that comes to mind as being possibly important must be investigated, in order to determine which ones (if any) are important. And, of course, the most important non-genetic factors may not be evident, or, if they are evident, may not be readily measured or even classified.

In many cases, variation between breeds and variation within breeds are directly related. Malignant hyperthermia syndrome (MHS; see Chapter 9), for example, has a completely genetic aetiology within pig breeds, in each of which it is the result of homozygosity for a recessive allele. There is also considerable variation in incidence between breeds, from well over 50 per cent in the Pietrain breed down to less than 5 per cent in Large Whites, and this between-breed variation can be explained in terms of different frequencies of the recessive allele in the different breeds.

However, such a relationship need not necessarily exist. In fact, the existence of a genetic contribution to aetiology within each of several breeds does not necessarily indicate that there is a genetic difference between breeds.

To continue with MHS as an example, it is evident that if the recessive allele occurs with the same frequency in two breeds, there will be no variation between those two breeds but there will still be considerable variation within each breed.

The final piece of general evidence suggesting a genetic contribution to aetiology is when the same, or a very similar, disorder is definitely inherited in another species of animal, or in humans. Single-gene disorders of the type discussed in Chapter 3 illustrate this point particularly well, and so too do inherited coat colours and patterns, which are described in Chapter 11. Of course there are exceptions to the rule, and we must be on the lookout continually for them; but the general principle holds very well. We can go one step further with X-linked loci, and predict quite confidently that if a particular disorder is X-linked in one species of mammal, it will be X-linked in most species of mammals.

The four types of simple, Mendelian inheritance

If a disorder is still familial, after removing or allowing for all conceivable non-genetic factors, the next step is to determine if the available data correspond in general terms to any of the four simple, Mendelian types of inheritance: autosomal dominant, autosomal recessive, X-linked dominant, and X-linked recessive.

Drawing on knowledge of Mendelian inheritance (Chapter 1) and of population genetics (Chapter 5), it is possible to assemble a list of criteria that, taken together, suggest a particular form of inheritance. The relevant criteria for each type of inheritance are summarized below.

Autosomal dominant

(1) The disorder is transmitted from generation to generation without skipping any generations.

Copyrighted material

(2) Every affected offspring has at least one affected parent, except in the case of a new mutant.
(3) Normal offspring from affected parents produce only normal offspring when mated to normals, and the same is true for all their descendants.
(4) Approximately equal numbers of males and females are affected.
(5) If the disorder is rare but not lethal, most matings producing affected offspring will be normal × affected ($aa \times Aa$), in which case the expectation is that one-half of each sex among the offspring will be affected. Thus the segregation frequency is $\frac{1}{2}$.
(6) If the disorder is lethal, it will be very rare, occurring sporadically with an incidence equal to twice the mutation rate.

Autosomal recessive

(1) The disorder may skip generations.
(2) All offspring of two affected parents are affected.
(3) Approximately equal numbers of males and females are affected.
(4) If the disorder is rare:
    (i) most affected individuals will have both parents normal;
    (ii) most matings producing affected offspring will be $Bb \times Bb$, for which the segregation frequency is $\frac{1}{4}$ (see later in this chapter for a discussion of a bias that arises in data collected from such matings);
    (iii) carriers (Bb) will usually mate with homozygous normals (BB), producing one-half carriers among their offspring; if, then, a carrier sire is mated to his own daughters, or to the daughters of another carrier sire, it follows that one-half of such matings will be $Bb \times Bb$, in which case the segregation frequency is $\frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$;
    (iv) matings between an affected animal and an unrelated normal animal usually produce only normal offspring;
    (v) affected × normal matings that do produce affected offspring must be $bb \times Bb$, in which case the segregation frequency is $\frac{1}{2}$;
    (vi) the average genetic relationship between normal parents of affected individuals is greater than between normal parents that have not produced affected individuals. (The reason for this is that the greater the genetic relationship between two individuals, the more likely they are to be carrying the same mutant gene; see Chapter 12.)

X-linked dominant

(1) Affected males when mated to normal females transmit the disorder to all their daughters but to none of their sons.
(2) Unless the disorder is very common, affected females when mated to normal males transmit the disorder to an average of one-half of their sons and one-half of their daughters.
(3) If the disorder is rare, its incidence in females is approximately twice that in males, in the general population.

Copyrighted material

(4) Every affected offspring has at least one affected parent, except in the case of a new mutant.

X-linked recessive

(1) The disorder may skip generations.
(2) All offspring of two affected parents are affected.
(3) Incidence is lower in females than in males, with the incidence of the disorder in females being approximately the square of the incidence in males, in the general population.
(4) If the disorder is rare:
   (i) most affected individuals are males, and result from matings among normal parents;
   (ii) most matings producing affected offspring will be $X^DX^d \times X^DY$, for which the segregation frequency is zero in females and $\frac{1}{2}$ in males (see later in this chapter for a discussion of a bias that arises in data collected from such matings);
   (iii) affected males when mated with normal unrelated females transmit the disorder to none of their offspring, but all of their daughters are carriers;
   (iv) affected females when mated with normal males transmit the disorder to all their male offspring but to none of their female offspring. All female offspring, however, are carriers.

Studying and analysing the data

Pedigrees

In many cases, it is convenient to start by drawing pedigrees or family trees. Pedigrees are most informative if drawn in a standard format, using the symbols illustrated in Fig. 7.1a. An example of a pedigree is given in Fig. 7.1b, in this case for hereditary multiple exostosis in a family of horses. Constructing pedigrees in this manner can be helpful in providing an initial impression of how a particular disorder is transmitted from one generation to the next. From Fig. 7.1b, for example, it is evident that males and females are affected in approximately equal proportions (six males and five females in generation IV), that matings of affected × affected can produce normal females, and that matings of affected × unrelated normals give rise to quite high proportions of affected offspring. The second observation above is not compatible with either form of X-linked inheritance or with autosomal recessive inheritance, but all three observations are compatible with autosomal dominance. Having examined a pedigree such as the one shown in Fig. 7.1b, a tentative conclusion could be drawn that the disorder in question, namely multiple exostosis in horses, is an autosomal dominant condition, assuming that III-1 is a new mutant. This conclusion could then be tested on further data.

The problems associated with drawing a pedigree are first that it is a time-consuming task, and second that in the case of animals, the occurrence of matings

Copyrighted material

Hidden page

between close relatives often results in a large number of intercrossing lines that are difficult to interpret. Despite these disadvantages, pedigrees can be quite useful, and are often drawn during the initial stages of investigation into the inheritance of a disorder. Fortunately, several software packages are now freely available for this purpose, e.g. Brian Kinghorn's PEDIGREE VIEWER (available from http://www-personal.une.edu.au/~bkinghor/pedigree.htm). Typically, the input for such packages is a three-column file: individual animal, sire, dam.

Segregation analysis

If the data on a particular disorder appear to correspond in general terms to one of the four sets of criteria outlined above, the next step is to determine more specifically whether the data are compatible with the respective simple Mendelian mode of inheritance. This is done by means of a segregation analysis. If specifically planned matings can be arranged, it is a simple matter to test agreement with a particular mode of inheritance, by comparing the expected segregation frequencies, which can be obtained from Tables 1.3 and 1.4 in Chapter 1, with those observed. For example, if the disorder is thought to be autosomal recessive, all normal individuals that have produced one or more affected offspring must be carriers. If matings among such known carriers are specifically arranged, we expect one-quarter of all offspring to be affected with the disorder, i.e. the expected segregation frequency is 0.25.

In many situations, however, it is not possible to arrange matings between known carriers and to observe all resultant offspring. Instead, it is much more usual for matings to come to the notice of the investigator after the event, and then only if an affected offspring is produced. This immediately introduces a bias into the data, because those carrier × carrier matings that by chance do not produce any affected offspring are automatically excluded from the data. The result is that even if the defect is really the result of an autosomal recessive allele, a segregation frequency greater than 0.25 is expected. This is because all the affected offspring have been included in the data, but not all of the normal offspring.

The existence of this bias was recognized as long ago as 1912, and since then many different methods of segregation analysis have been developed in order to cater for it. A worked example of the best method is given in Appendix 7.1 of *Veterinary Genetics*.

Although the simplest forms of segregation analysis are themselves quite straightforward, in practice there are often considerable difficulties encountered by investigators attempting to determine whether a disorder is Mendelian. For example, it is quite commonly found that while some of the available observations agree very closely with the general criteria for a certain type of inheritance, others are incompatible. Misclassification of phenotypes, as discussed in Chapter 6, is one possible reason. So too is the occurrence of new mutations, and variation in the age of animals when a disorder phenotype becomes visible (i.e. variable age of onset). In addition, complications can arise from sex-limited inheritance, genetic heterogeneity, X-inactivation, phenocopies, genetic anticipation associated with unstable trinucleotide repeats (all discussed in Chapter 3), and genomic imprinting (Chapter 1). In the

Copyrighted material

Hidden page

Hidden page

Hidden page

Hidden page

Hidden page

antigens, unless they have been specifically challenged with the appropriate foreign red blood cells.

Since animals do not normally carry antibodies to red-blood-cell antigens, it was often assumed that blood transfusions in animals can be conducted safely with any available blood, and that there is normally no need for blood typing before transfusion. However, transfusion with randomly-chosen, untyped blood may lead to an immediate transfusion reaction if, unbeknown to the practitioner, the recipient has been transfused previously with blood containing the same antigens. Even if this does not occur, it is likely that a random, untyped transfusion will sensitize the recipient to subsequent transfusions, or to the blood cells of the recipient's future offspring, if the recipient is a female (see below). Thus, whenever possible, it is advisable to obtain blood for transfusion from donors that have been typed as being compatible or negative for red-blood-cell antigens that are known to evoke strong antibody responses. Increasingly, animal blood banks are being established to satisfy this need.

The most clinically important antigens are A in dogs, B in cats, A_a and Q_a in horses, and A, F, and some B antigens in cattle. If untyped donors must be used, a simple cross-match should be conducted, in which a drop of plasma from the recipient is mixed on a slide with a drop of red-cell suspension from the donor. If agglutination occurs, it would be wise to find another donor. However, because this cross-match test is not always effective, a lack of agglutination does not guarantee that a transfusion reaction will not occur. Care should therefore be taken during transfusions, even if the cross-match test is negative.

Neonatal isoerythrolysis

Occasionally newborn foals that appear perfectly normal at birth become weak and dull within 24 hours of birth, and develop acute anaemia, jaundice, and haemoglobinuria. Their heart and respiratory rates become elevated, and they usually die within a few days. This disease is known as neonatal isoerythrolysis, NI, or haemolytic disease of the newborn. Judging from the above clinical signs, NI is associated with destruction of red blood cells. Why should this happen?

In the case of horses, the answer lies in feto–maternal haemorrhage, which occurs sometimes during pregnancy or birth, releasing red blood cells from the fetus into its dam's bloodstream. Consider the A blood-group system, which happens to be the most important in relation to NI, and consider the A_a antigen within that system. Suppose that the fetus has inherited the A_a antigen from its sire, i.e. both sire and fetus are positive for A_a (written as A_a+). Suppose also that the dam lacks the A_a antigen (written as A_a–). When the cells from the fetus enter the dam, she recognizes antigen A_a as non-self, because she does not have that antigen. She therefore produces anti-A_a antibodies in her serum. These anti-A_a antibodies are transferred along with all other antibodies into the dam's colostrum, which the foal drinks. The reason for the clinical signs of NI should now be evident: the anti-A_a antibodies are absorbed through the foal's gut and pass into its bloodstream, where they rapidly destroy all cells with A_a antigen on their surface.

Copyrighted material

Not all blood-group systems give rise to this problem. Indeed, only the A and Q systems are regularly implicated in horses, with the former being the most important. However, it seems that all donkeys exhibit a red-cell antigen that is absent in horses, which means that all mule fetuses carried in mares are at risk of NI. Fortunately, NI is rare in the first-born of any mare because the initial immune response is usually too slow to cause any trouble; the mare has not yet been sensitized. If challenged a second time, however, the mare quickly mounts an immune response which gives rise to NI.

There are certain actions that can be taken to alleviate NI. Treatment of affected foals involves either transfusion with whole blood from a suitable donor, or transfusion of washed red blood cells from the dam. The main requirement is that the cells given to the affected foal must not carry any antigens that are carried by the sire but not the dam of the foal, as it is these antigens against which the dam has produced antibodies. Thus the sire is *not* a suitable donor. In contrast, the dam could be a suitable donor, because none of the antibodies that the foal obtained from her colostrum will be directed against her own cells. But her serum contains the offending antibodies, produced by her against the foal's antigens inherited from its sire. Consequently, if the dam's cells are to be used for transfusion, they must first be washed with sterile saline, with the aim of removing all plasma and hence all offending antibody.

NI in foals can be prevented very simply by not allowing the foal access to its dam's colostrum for the first 24–36 hours, until protein molecules are no longer absorbed by the foal's small intestine. Such action should be taken only if the foal is known to be at risk. This can be determined by screening for the presence of antibody to the foal's cells, in the pregnant mare's serum at 4 weeks, 2 weeks, and 1 week before the expected parturition date. The cells used in this screening test can be from the sire of the foal or from a panel of horses known to be positive for antigens involved in NI. However, because NI occurs with a frequency of less than 1 per cent, this would be a very wasteful procedure if performed on all mares. A more efficient procedure would be to blood-type all mares, and then to screen during the last 4 weeks of pregnancy only those that are A_a negative, i.e. that lack antigen A_a. The reason for concentrating solely on A_a is that antibodies to this antigen are thought to be the cause of more than 80 per cent of all cases of NI in horses. At the very minimum, all mares that are thought to have previously produced an NI foal should be screened. If this is not possible, all subsequent foals of such mares should be given colostrum from a source other than their dam, as a precaution against NI.

NI is best known in horses but has also been reported in cats, dogs, cattle, and pigs. In the case of cattle and pigs, its cause was traced to the use of blood-based vaccines against babesiosis (tick fever) and anaplasmosis in cattle, and against swine fever (hog cholera) in pigs. These cases of NI have now been largely overcome by phasing out blood-based vaccines.

In the case of naturally occurring NI, it is worth noting that the genetic effect in all species is selection against heterozygotes. This is because the only situation in which the fetus has an antigen not carried by the dam is if the fetus inherited a different antigen from its sire, and must therefore be heterozygous.

Copyrighted material

Hidden page

Hidden page

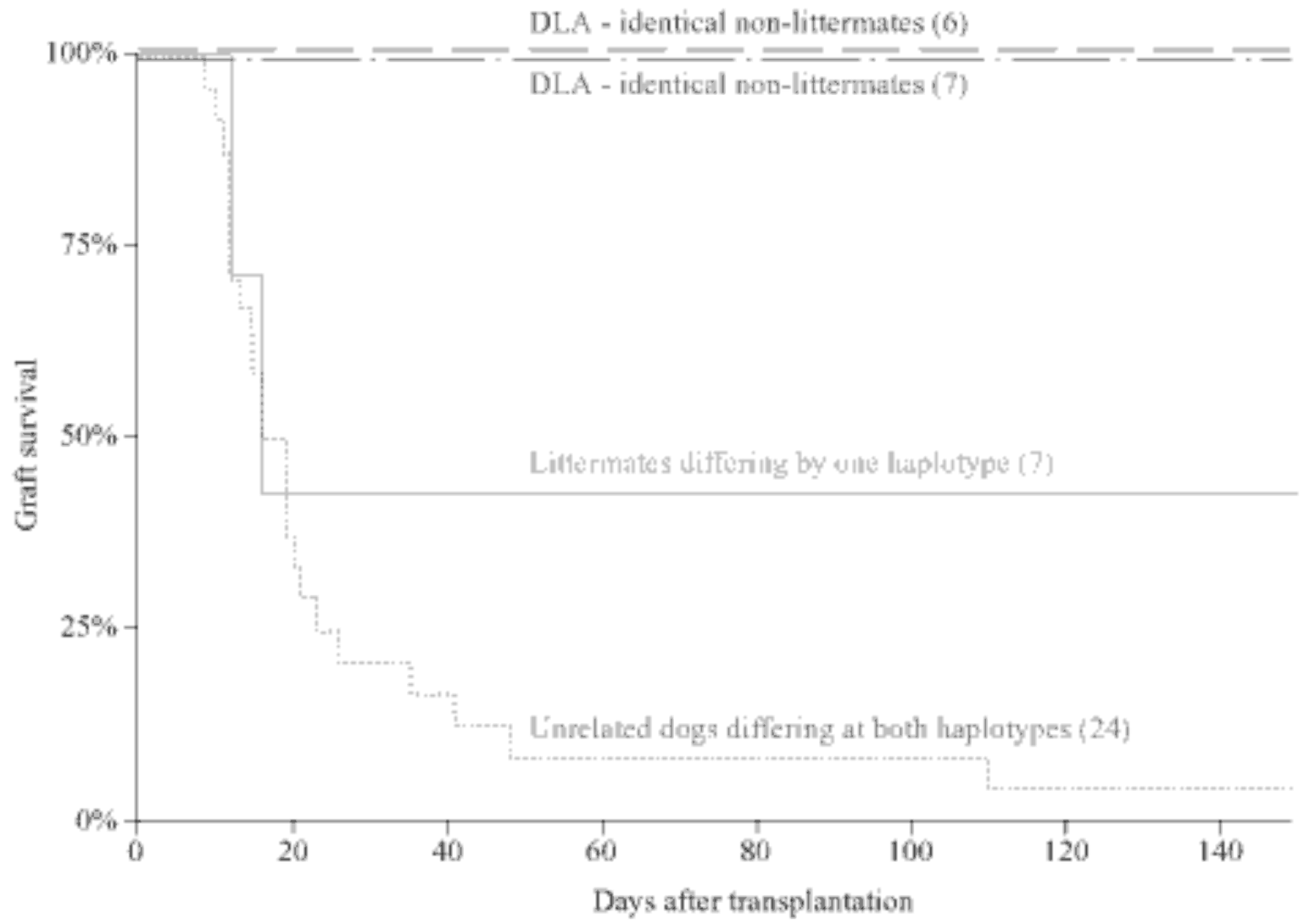


Fig. 8.3 Survival after kidney transplants in dogs. Numbers in brackets are the numbers of recipients in each group, excluding four recipients that died from causes other than rejection. (Drawn from data in Bijnen *et al.* (1980), *Transplantation*, **30**, 191–5.)

Although initial interest in the MHC arose from its role in tissue rejection, it has since become clear that the MHC plays a vital role in the immune response to pathogens and parasites. In fact, the MHC is the engine room of the immune response to disease. How does it perform this important role? Basically, it does so by enabling T cells to identify anything that is non-self (foreign). How does this occur? By a histoglobulin 'presenting' to T cells a small peptide fragment of the pathogen or parasite. The fragment is held within a groove (called the **peptide-binding region**, PBR, or **antigen-binding site**, ABS) of the histoglobulin. The T cells recognize the foreign peptide as non-self only if it is presented in the context of self, i.e. only if it is presented by a histoglobulin which the T cells recognize as self. This phenomenon is called MHC restriction. Once they have recognized the peptide as being foreign, the T cells activate either a cell-mediated immune response (if the foreign peptide has come from inside a host cell) or an antibody-mediated immune response (if it came from elsewhere).

Determining phenotype and genotype at the MHC

In order to investigate the role of the MHC in immunity to disease, it is necessary to be able to identify which histoglobulins are present in any individual. Also, in

Copyrighted material

Hidden page

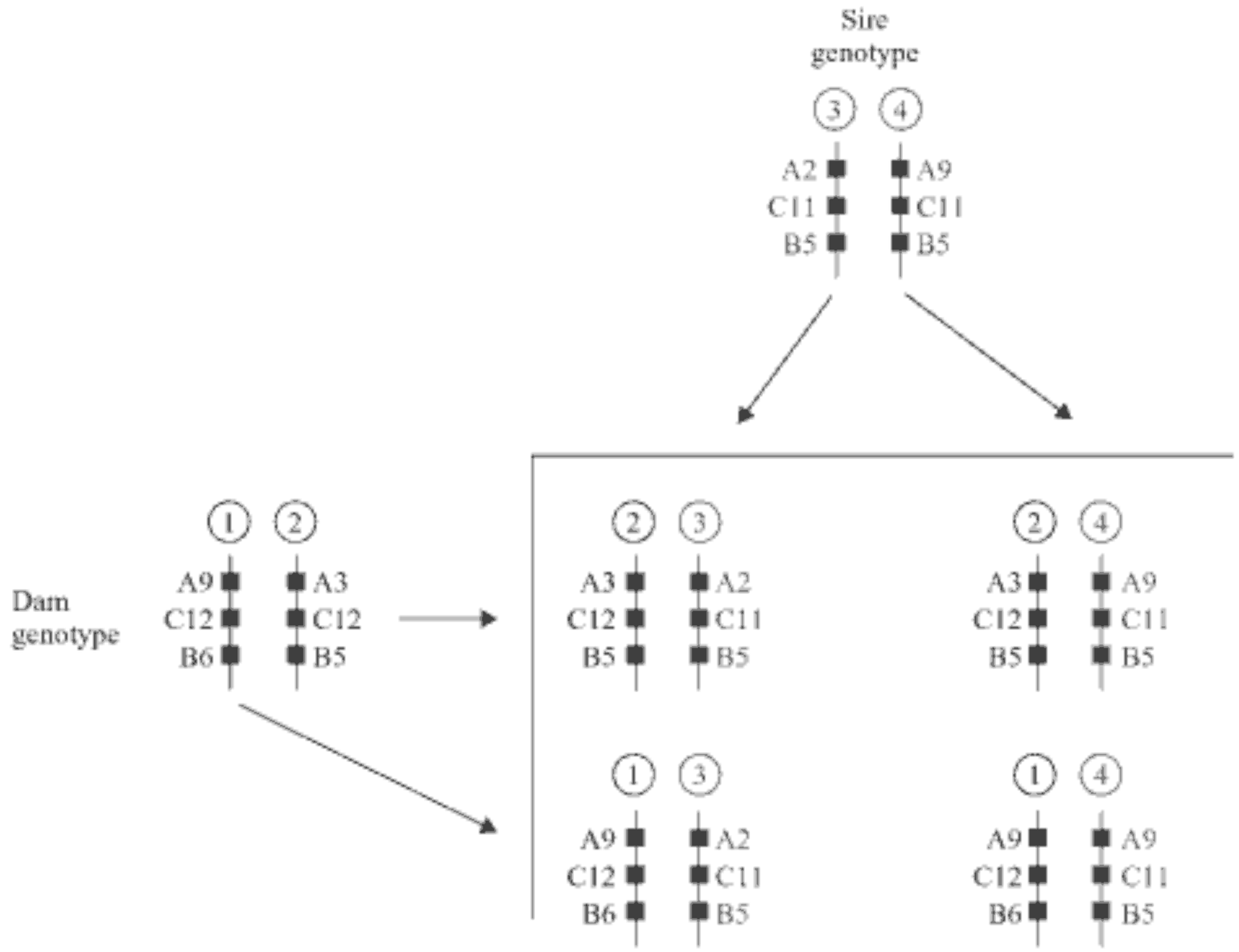


Fig. 8.4 Inheritance of major histocompatibility complex (MHC) haplotypes. For simplicity, only three class-I genes are shown. Haplotype numbers are shown in circles.

this results in only four different combinations of haplotypes in the offspring from a particular mating (called full-sibs), with all combinations occurring with equal frequency, namely 1/4. It follows that despite the immense polymorphism at the MHC, the chance that any two full-sibs have the same set of MHC genes is $1/4 \times 1/4 = 1/16$ (if we neglect the low chance of crossing-over). Also, for any individual, there is a chance of 1/4 that any one of its full-sibs has the same set of MHC genes as it does. Thus, full-sibs are a very useful potential source of tissue for transplantation.

The adaptive significance of MHC polymorphism

We saw earlier that most of the allelic variation within MHC genes occurs within just one or two exons. In fact, most of the variation is confined to 57 triplets within exons 2 and 3 in class I genes, 19 or 20 triplets in exon 2 of class II α-chain genes, and 15 or 16 triplets in exon 2 of class II β-chain genes. Interestingly, these are the triplets that encode the peptide-binding region (PBR) of histoglobulins. Furthermore, comparisons of the differences in sequence between alleles have shown that the rate of silent mutations (often called **synonymous base substitutions**) is much slower than the rate of mis-sense mutations (**non-synonymous base substitutions**) in the

Copyrighted material

triplets of the PBR, whereas exactly the opposite is true in non-PBR triplets. These results have been interpreted as indicating that natural selection favours the widest possible range of different amino-acid sequences in the PBR, but at the same time, positively discriminates against amino-acid substitutions in other regions of histoglobulins. It should be evident that this fits in perfectly with the different role of the PBR compared with the rest of the histoglobulin molecule: the PBR needs to be as variable as possible, so as to maximize the number of different foreign peptide fragments that can be presented to T cells. In contrast, the remainder of the histoglobulin has a constant role, irrespective of which peptide is being presented; its job is to anchor itself to the cell membrane, and to provide various forms of physical attachment by which T cells can interrogate the PBR region.

Associations between the MHC and disease

Given the picture presented above, is there evidence that certain histoglobulins are able to present certain pathogens more effectively than others, and so bring about a more effective immune response? In other words, is there any evidence for association between certain histoglobulins and certain diseases?

At the molecular level, there is convincing evidence of differences between histoglobulins in their ability to present foreign peptides. In addition, at the organismal level, there are definitely some well-documented examples of MHC–disease associations. However, most of these are in humans, most are involved with susceptibility rather than with resistance, and many involve diseases of an autoimmune nature. In domestic animals, despite a huge number of studies and many apparent associations, there are only a few that can withstand close scrutiny. The most convincing evidence to date has been obtained from poultry, in which the association between the B^{21} histoglobulin and susceptibility to Marek's disease is very solidly established.

In the future, it may become evident that the important thing is not to have a particular set of histoglobulins, but to be as heterozygous as possible at as many loci as possible, with the actual identity of histoglobulins being of secondary importance. Some tantalizing evidence along these lines comes from the Tasmanian devil, a species threatened with extinction because of the rapid spread of a fatal transmissible tumour. The tumour cells arose from just one animal and are spread via the devil's common habit of mouth-to-mouth combat. The inability of infected animals to mount an immune response against the foreign cells appears to be associated with the very low levels of MHC diversity in devils, which is presumably the result of a population bottleneck. If this turns out to be a widespread phenomenon, the aim that some people have of identifying the 'best' histoglobulins and then making populations homozygous for them, may be the wrong approach.

Other MHC associations

In view of the MHC's important role in immunity, it seems likely that it also holds the key to the design of effective vaccines. In particular, the effectiveness of a

Copyrighted material

vaccine depends on the extent to which histoglobulins can present the protective components of the vaccine to T cells. With PCR genotyping enabling MHC alleles to be distinguished down to the level of base sequence, there is great scope for vaccine manufacturers to collaborate with immuno-geneticists so as to ensure that commercial vaccines have the greatest chance of success.

In addition to transplantation and immunity, the MHC appears to play some role in choice of mate and in reproductive success. For example, some intriguing data from experiments with mice show that when given a choice between mating partners, both males and females tend to choose partners from whom they differ the most in terms of histoglobulins, i.e. there is a tendency for mates to be chosen on the basis of histo*in*compatibility. Furthermore, there is evidence suggesting that the discrimination between different MHC phenotypes is based on odours.

In relation to reproductive success, there is tantalizing evidence from humans that couples with a history of recurrent spontaneous abortion (RSA) are more similar in their histoglobulins, i.e. are more histocompatible, than fertile control couples. In addition, among embryos that survive to full term, those which are most histocompatible with their mothers tend to have the lightest birth weights. Some people argue that this is essentially a genetic phenomenon, reflecting the effects of increased homozygosity for defective alleles at loci located in the MHC (but not yet identified) which are involved in early embryonic development. Others favour the suggestion that successful implantation and pregnancy actually require a controlled immune response, which in turn requires the mother to recognize, and respond in a controlled manner to, the non-self embryonic histoglobulins – those that are the products of the MHC alleles inherited from the father. The RSA association with histoincompatibility could be explained on the basis that the greater the MHC similarity between father and mother, the smaller the chance that the mother will be able to mount the controlled immune response that is necessary for successful pregnancy.

The importance of the MHC in mate choice and reproductive success remains an open question at present, especially in relation to domestic animals. However, given the supreme importance of reproductive ability in domestic animals, the possible role of the MHC should be fully investigated, especially now that PCR technology enables such detailed and accurate genotyping.

Finally, mention must be made of the many studies that have investigated associations between the MHC and production traits of domestic animals. The main rationale for these studies has been the argument that if the MHC does have an influence on disease resistance, there may be a consequential effect on production – healthier animals are expected to be more productive. As with the disease-association studies, there are many claims of association between MHC and production traits, but on close examination, many turn out to be more apparent than real; and, except in poultry, none of the real associations has been sufficiently strong to justify their use in practical breeding programmes. It seems likely that the most fruitful areas of future MHC research in domestic animals will involve the immune response directly, rather than the possible indirect effects of MHC associations with production traits.

Copyrighted material

Hidden page

Further reading

Day, M.J. (2007) Immune system development in the dog and cat. *Journal of Comparative Pathology*, **137 Suppl 1**, S10–15.

Hansen, T.H. and Bouvier, M. (2009) MHC class I antigen presentation: learning from viral evasion strategies. *Nature Reviews Immunology*, **9**, 503–13.

Klein, J., Sato, A. and Nikolaidis, N. (2007) MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. *Annual Review of Genetics*, **41**, 281–304.

Munguia-Vega, A., Esquer-Garrigos, Y., Rojas-Bracho, L., Vazquez-Juarez, R., Castro-Prieto, A. and Flores-Ramirez, S. (2007) Genetic drift vs. natural selection in a long-term small isolated population: major histocompatibility complex class II variation in the Gulf of California endemic porpoise (Phocoena sinus). *Molecular Ecology*, **16**, 4051–65.

Owen, J.P., Delany, M.E. and Mullens, B.A. (2008) MHC haplotype involvement in avian resistance to an ectoparasite. *Immunogenetics*, **60**, 621–31.

Siddle, H.V., Kreiss, A., Eldridge, M.D., Noonan, E., Clarke, C.J., Pyecroft, S., Woods, G.M. and Belov, K. (2007) Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. *Proceedings of the National Academy of Sciences USA*, **104**, 16221–6.

Teng, G. and Papavasiliou, F.N. (2007) Immunoglobulin somatic hypermutation. *Annual Review of Genetics*, **41**, 107–20.

Yamazaki, K. and Beauchamp, G.K. (2007) Genetic basis for MHC-dependent mate choice. *Advances in Genetics*, **59**, 129–45.

Copyrighted material

9 Pharmacogenetics

Given the vast array of different biochemical pathways that are involved in reactions to drugs, there are obviously many genes that affect drug metabolism. The aim of this chapter is to provide an overview of the genetic control of drug metabolism.

Genetic polymorphisms affecting drug metabolism

One of the most important sets of genes concerned with drug metabolism is the cytochrome-P450 gene family. The number of different functional P450 genes in a species is at least 60 and could be greater than 200. The products of these genes are haem-containing enzymes called mono-oxygenases, which are in the first line of 'defence' against 'foreign' chemicals, performing many different biochemical functions including hydroxylation, dehalogenation, dealkylation, deamination, and reduction, as a necessary first step in detoxification.

While there is still much to be learnt about these enzymes, especially in domestic animals, there is already substantial evidence of considerable polymorphism at many of the P450 genes, giving rise to marked differences in reaction to a range of drugs including debrisoquin, sparteine, dextrometorphan, tricyclic antidepressants, opioids, beta-adrenergic receptor antagonists, mephenytoin, mephobarbital, hexobarbital, and diazepam. Mutant forms of P450 enzymes inhibit the metabolism of these drugs, resulting in adverse drug reactions, including presentation with extreme side effects.

Another well documented polymorphism in drug metabolism concerns the locus for the enzyme *N*-acetyltransferase (NAT), which is involved in numerous acetylation reactions. Faulty alleles at this locus result in substantially slower deactivation of a wide range of drugs including isoniazid, sulphamethazine and several other sulphonamides, hydralazine, procainamide, dapsone, para-aminobenzoic acid, phenelzine, and aminoglutethimide; as well as substances such as clonazepam, nitrazepam, benzidine, 2-amino-fluorene, beta-naphthylamine, and even caffeine! Once again, the clinical importance of the mutant alleles arises from the side effects induced by failure to deactivate the chemical.

Copyrighted material

Genetics and anaesthesia

One of the best understood cases of the genetic basis of drug response in animals concerns reaction to the anaesthetic halothane in pigs, which has been used as a test for malignant hyperthermia syndrome (MHS). Reactors usually show signs of stiffening in the hindquarters after 2 minutes exposure to halothane, associated with rapid hyperthermia which soon leads to death if halothane is not removed. Reaction to halothane is the result of homozygosity for a recessive allele, which is usually given the symbol *hal* or *n*.

Apart from its obvious importance in relation to MHS, the *hal* allele is also important because there is a very strong association between adverse reaction to halothane and the economically important set of traits known as porcine stress syndrome (PSS), which is characterized by sudden death after only a minor stress, and by pale, soft, exudative (PSE) meat. In contrast to these undesirable features, the *hal* allele also has a substantial advantage: it increases the yield of lean meat. (As mentioned in Chapter 5, this appears to be another example of selection favouring heterozygotes, which explains why the *hal* allele occurred at relatively high frequencies in many populations throughout the world.) Because of its obvious importance, tens of thousands of pigs were halothane-tested throughout the world each year during the 1980s, and the halothane locus was the subject of a huge international research effort. In 1991, a major breakthrough occurred when a Canadian research team led by David MacLennan reported the isolation and cloning of the halothane gene in pigs. Furthermore, by comparing the nucleotide sequence of the two alleles (normal and *hal*), MacLennan and colleagues showed that the *hal* allele differed from the normal allele by 18 nucleotide substitutions. However, only one of these nucleotide substitutions creates an amino acid substitution; the remainder are silent mutations. This means that all of the adverse effects of the *hal* allele (MHS, PSS, and PSE), plus its advantageous effects, are the result of the simplest of all possible mutations, namely a single base substitution. This occurs at the 1843rd nucleotide (where C is replaced by T), which causes an amino-acid substitution (arginine is replaced by cysteine) at the 615th position in the polypeptide chain.

The polypeptide product of the halothane gene has the unlikely name of ryanodine receptor, because it was first identified as a protein that happens to bind very strongly to a plant alkaloid called ryanodine. This protein is actually a calcium release channel (CRC) in the sarcoplasmic reticulum of skeletal muscle, allowing calcium ions to flow into the surrounding muscle tissue, resulting in muscle contraction.

The CRC molecule is extremely large, consisting of a single chain of approximately 5000 amino acids, which obviously means that the total coding sequence of the gene is approximately 15 000 bases long. This makes the single base-substitution difference between the normal and *hal* alleles all the more remarkable; the two alternative forms of the polypeptide differ by only one amino acid in a chain of 5000 amino acids! This simple difference occurs at a position in the polypeptide which has a critical influence on calcium flow, which in turn gives rise to all of the effects (both advantageous and disadvantageous) of the *hal* allele.

The cloning of the CRC gene enabled extensive DNA testing to be conducted in populations around the world. Essentially, the same mutation is responsible for all

Copyrighted material

cases of MHS in pigs. In humans, the situation is not so straightforward; in some families the same mutation as in pigs is associated with MHS, but in other families, MHS segregates independently from the CRC gene, indicating that there are other genes involved in MHS. Given this genetic heterogeneity in humans, we should not be surprised if sometime in the future we find cases of MHS in pigs which are not associated with the CRC gene.

Warfarin resistance

Warfarin is an anticoagulant that is used extensively in medicine; it is also used as a rodent poison.

Only five years after its introduction as a rodenticide in Britain, resistant rats were detected in the field. Nine years later (in 1967, when estimates first became available), the frequency of resistant rats was nearly 50 per cent in areas in which warfarin was used regularly, and it seemed that this frequency would continue to increase. However, the frequency of resistant rats has remained at an intermediate level ever since. What has caused this apparently stable intermediate equilibrium?

The answer lies in an understanding of the blood coagulation process. In simple terms, coagulation is a complex cascade of biochemical reactions involving various clotting factors. Some of these factors, namely the proteins II, VII, IX, and X, require vitamin K for their activation. Specifically, vitamin K is a co-factor for the carboxylation phase of activation, in which certain glutamic-acid amino acids in the clotting-factor polypeptides are converted to γ-carboxyglutamic acid (called Gla) by the enzyme γ-glutamylcarboxylase. This carboxylation enables the clotting factors to bind calcium ions, which in turn enables the factors to attach themselves to phospholipids on the surface of platelets. This is how blood clotting is initiated. By the time each vitamin K molecule has completed its carboxylation task, it has become vitamin K epoxide, which is then reduced by the enzyme epoxide reductase back to its original form, in readiness for the next round of carboxylation. The process of carboxylation followed by reduction is called the vitamin-K cycle.

Warfarin has a structure very similar to vitamin K, and therefore binds to epoxide reductase, preventing it from recycling the vitamin K epoxide. The result is a build-up of vitamin K epoxide, a deficiency of vitamin K, and hence a deficiency of clotting factors.

There are several different forms of inherited resistance to warfarin in rodents. One of the best understood involves an autosomal locus with an allele, *R*, which encodes a slightly different form of epoxide reductase that has a lower affinity for warfarin. However, because vitamin K and warfarin have such a similar molecular structure, the mutant form of epoxide reductase also has a lower affinity for vitamin K. The practical consequence of this is that animals homozygous for the resistant allele have a 20-fold higher requirement for vitamin K in their diet, to compensate for the lower affinity of the enzyme produced by this allele. If this increased demand cannot be met, homozygotes for the resistant allele suffer from the same bleeding disorders as those induced by warfarin in susceptible animals, namely disorders caused by deficiencies in the vitamin-K-dependent blood clotting factors.

Copyrighted material

Several studies of the relative fitness of each genotype resulting from the action of these factors have been conducted. Using the results from two such studies in Britain, the following somewhat simplified account illustrates the main aspects of the warfarin story.

Because *RR* homozygotes have a higher requirement for vitamin K than heterozygotes (*RS*, where *S* is the susceptible, wild-type allele), the genotype *RR* has the lowest relative fitness in the absence of warfarin, with the *RS* genotype having a fitness somewhere between the two homozygotes, e.g. 0.46, 0.77, and 1.00 for *RR*, *RS*, and *SS* respectively. From Chapter 5, it should be evident that the result of this set of relative fitnesses is that the *R* allele is either removed from the population, or is maintained at a low frequency by a selection–mutation balance.

When warfarin is introduced, both *RS* heterozygotes and *RR* homozygotes are equally resistant. But *RR* homozygotes still have a lower fitness because of their greater requirement for vitamin K. And *SS* homozygotes have a lower fitness than *RS* heterozygotes because they are susceptible to warfarin. In the presence of warfarin, therefore, there is heterozygote superiority for fitness. In one study, the relative fitnesses in the presence of warfarin were 0.37, 1.00, and 0.68 for *RR*, *RS*, and *SS*, respectively. It is left to the reader to verify, using the concepts discussed in Chapter 5, that the expected consequence of this change in relative fitness is an increase in the frequency of the *R* allele to a new equilibrium of 0.34, which corresponds to a frequency of resistant rats of approximately 0.56 at birth in any generation.

This is a classic example of a stable polymorphism. Allowing for the likely deaths of at least some of the *RR* rats as a result of vitamin K deficiency, and for their apparently smaller body size and consequent lack of social dominance, the above account is an adequate explanation for the observed frequency of resistant rats. It is also an adequate explanation for the gradual decrease in the frequency of resistant rats observed in areas where warfarin use has been discouraged after previous extensive use.

The warfarin story is perhaps the most dramatic example of the practical implications of genetic variation in response to drugs. Although the use of warfarin as a rodenticide is only indirectly relevant to the use of it or any other drug in domestic animals or in humans, the warfarin story does emphasize that differences between individuals do exist in relation to drug response, and that these differences are sometimes under genetic control.

Not surprisingly, the effect of warfarin can be mimicked by mutations at any of the loci encoding the enzymes involved in the vitamin-K cycle. For example, a familial deficiency of all of the vitamin-K-dependent clotting factors (II, VII, IX, and X), reported in Devon Rex cats, appears to be caused by a mutation in the locus for γ-glutamylcarboxylase.

Multifactorial pharmacogenetics

In the examples described above, alleles at a single locus have a large effect on the metabolism of a particular drug. This is illustrated in Fig. 9.1. However, Fig. 9.1 also shows that there is substantial variation *within* each of the two groups. This

Copyrighted material

Hidden page

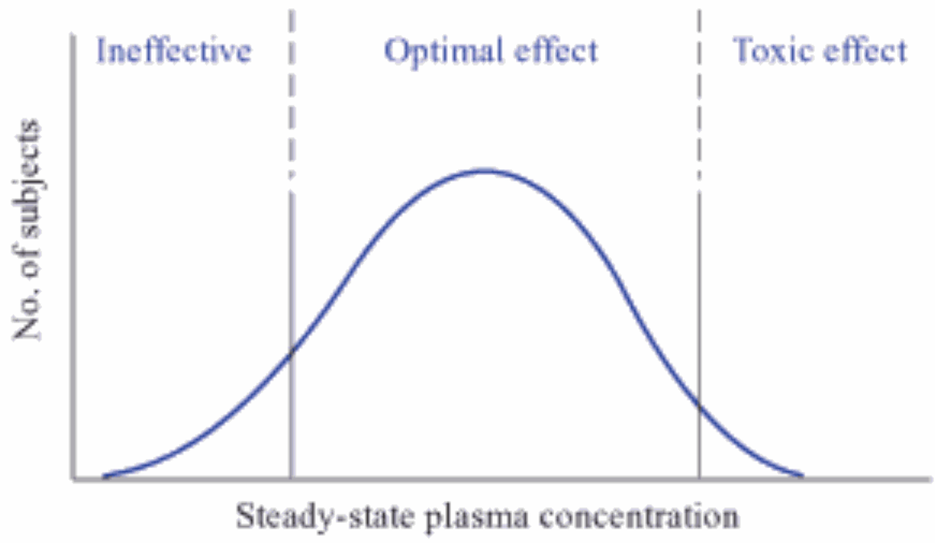


Fig. 9.2 The continuous distribution of drug concentration resulting from the action of many genes and many non-genetic factors. (Reprinted by permission from Springer Science+Business Media: Vogel and Motulsky (1979), *Human genetics: problems and approaches*, p. 261, fig. 4.60.)

which live on the dead litter at the base of many pastures in the North Island of New Zealand. The exact biochemistry of this disease is not yet known, but there is substantial evidence for multifactorial variation in response to the toxin: some animals are not affected at all, others are affected at a subclinical level, which can cause substantial loss of production, and others show severe clinical signs including photosensitivity and tissue damage, sometimes resulting in death. The heritability of response to the toxin is around 40 per cent, which tells us that there are substantial genetic differences between individuals in their ability to cope with the toxin. Consistent with this heritability, selection for increased resistance to a standard dose of the toxin has resulted in a dramatic decrease in the incidence of the disease, from around 65 per cent down to around 20 per cent in only 6 years.

Summary

- Given the vast array of different biochemical pathways that are involved in reactions to drugs, there are obviously many genes that affect drug metabolism
- One of the most important sets of genes concerned with drug metabolism is the cytochrome-P450 gene family, for which there is considerable polymorphism, giving rise to marked differences in reaction to a range of drugs
- Malignant hyperthermia in pigs is a single-locus defect arising from an adverse reaction to halothane, with reaction being recessive to non-reaction; this polymorphism is caused by a single base substitution in an enormous gene, whose coding region comprises around 15 kb
- Warfarin resistance in rodents is a single-locus condition with resistance being dominant, but also with homozygotes for the resistant allele having greatly reduced fitness because of a requirement for up to 20 times more vitamin K, which normal diets cannot meet – giving rise to selection favouring heterozygotes, which accounts for the continued polymorphism observed at this locus
- Response to many drugs is most likely to be multifactorial

Copyrighted material

Further reading

Fleischer, S., Sharkey, M., Mealey, K., Ostrander, E.A. and Martinez, M. (2008) Pharmacogenetic and metabolic differences between dog breeds: their impact on canine medicine and the use of the dog as a preclinical animal model. *AAPS Journal*, **10**, 110–19.

Gomase, V.S., Tagore, S., Changbhale, S.S. and Kale, K.V. (2008) Pharmacogenomics. *Current Drug Metabolism*, **9**, 207–12.

Ishizuka, M., Okajima, F., Tanikawa, T., Min, H., Tanaka, K.D., Sakamoto, K.Q. and Fujita, S. (2007) Elevated warfarin metabolism in warfarin-resistant roof rats (Rattus rattus) in Tokyo. *Drug Metabolism and Disposition*, **35**, 62–6.

Ishizuka, M., Tanikawa, T., Tanaka, K.D., Heewon, M., Okajima, F., Sakamoto, K.Q. and Fujita, S. (2008) Pesticide resistance in wild mammals–mechanisms of anticoagulant resistance in wild rodents. *Journal of Toxicological Science*, **33**, 283–91.

Martinez, M., Modric, S., Sharkey, M., Troutman, L., Walker, L. and Mealey, K. (2008) The pharmacogenomics of P-glycoprotein and its role in veterinary medicine. *Journal of Veterinary Pharmacology and Therapeutics*, **31**, 285–300.

Phua, S.H., Dodds, K.G., Morris, C.A., Henry, H.M., Beattie, A.E., Garmonsway, H.G., Towers, N.R. and Crawford, A.M. (2009) A genome-screen experiment to detect quantitative trait loci affecting resistance to facial eczema disease in sheep. *Animal Genetics*, **40**, 73–9.

Rost, S., Pelz, H.J., Menzel, S., MacNicoll, A.D., León, V., Song, K.J., Jäkel, T., Oldenburg, J. and Müller, C.R. (2009) Novel mutations in the VKORC1 gene of wild rats and mice – a response to 50 years of selection pressure by warfarin? *BMC Genetics*, **10**, 4.

Wardrop, D. and Keeling, D. (2008) The story of the discovery of heparin and warfarin. *British Journal of Haematology*, **141**, 757–63.

Copyrighted material

10 Hosts, parasites, and pathogens

'… can we doubt that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind?'
(Darwin, 1859)

Parasites and pathogens exert a profound influence on animal health and production throughout the world. It is therefore very important that we understand as much as possible about them. Even more importantly, we must understand the implications of attempts by humans to bring them under control.

The aim of this chapter is to provide such an understanding. We shall start by considering host–pathogen interactions, and shall then discuss the genetic basis of resistance in hosts and in parasites and pathogens. We shall finish with some illustrations of ways in which attempts are now being made to combat parasites and pathogens.

Host–pathogen interactions

In this section, we shall consider three very different examples of interaction between hosts and parasites or pathogens, with the aim of providing an indication of the variety of ways in which such interactions occur.

Myxomatosis in rabbits

In 1859 a homesick Englishman, who had grown a little tired of seeing nothing but kangaroos on his property near Geelong in Australia, imported some rabbits from England. He thought it would improve the hunting. Little did he know that this innocent importation would give rise to a plague of rabbits so serious that it would threaten the very existence of the pastoral industries in Australia.

Nothing done by humans had much effect on the rabbit population until the CSIRO released a virulent strain of the myxoma virus into the rabbit population in 1950. Having never before been exposed to the virus, the rabbits were very susceptible; their liability to myxomatosis was very high – almost 100 per cent of infected

Copyrighted material

rabbits died. The combination of susceptible host and virulent pathogen assured the scheme of initial success; myxomatosis spread very rapidly, and rabbits died by the thousands. But the very success of the scheme in its early days ensured that it could not be a success in the longer term. The CSIRO appreciated this because they realized that the death of large numbers of rabbits imposed very strong natural selection for resistance in the host, and also very strong natural selection for avirulence in the pathogen.

If there had been no genetic variation for liability to myxomatosis in the rabbit population, this very strong natural selection would have had no effect. But there was genetic variation (the heritability of liability to myxomatosis is around 35 per cent), and those animals with genes for low liability had a greater chance of surviving to reproduce and hence to pass on their 'resistance' genes to their offspring. Similarly, had there been no genetic variation in the virus and had there been no mechanisms for new variation to be generated in the virus, virulence would not have changed. But either there was genetic variation in the virus, or (as seems more likely) new variation was created by mutation. As soon as a mutant occurred that decreased the virulence, the rabbits infected with the less virulent virus stayed alive longer and enabled more 'offspring' of the less-virulent strain to be produced.

It is not surprising, therefore, that strains of rabbits gradually emerged with varying degrees of increased resistance to the virus, and strains of virus emerged with varying degrees of decreased virulence.

New, virulent strains of virus were developed in the laboratory and were released, with considerable effectiveness in the period immediately following their release. Once again, however, the very success of the new strain of virus imposed very strong natural selection for decreased virulence in the virus and increased resistance in the host, with the result that fewer and fewer rabbits died.

In a situation like this, there is obviously a very dynamic interaction between host and pathogen. Humans can and do tamper with this interaction from time to time, but their ability to exert a lasting effect in the direction they desire is limited by biological realities of which we all should be very much aware.

That is not to say that the myxomatosis programme should never have been launched by the CSIRO; on the contrary, the programme was very successful. It is simply to say that in order to avoid disappointment, the limitations imposed by biological realities should be appreciated before such a programme is commenced.

The host–pathogen interaction exemplified by the rabbit and the myxoma virus is a typical one. We shall now discuss an example of host–pathogen interaction that is just as important but which is somewhat less typical. Indeed, in this particular example, there is debate about the nature of the pathogen.

Prion diseases

There is a class of diseases called spongiform encephalopathies, which are fatal neurological diseases that have been known for many years in sheep and goats (where the disease is called scrapie), and in humans (where the diseases are kuru disease, Creutzfeldt–Jakob disease, and Gerstmann–Sträussler–Scheinker syndrome). There

Copyrighted material

is still controversy about the nature of the infectious agent that causes these diseases. However, it seems likely that it is a modified form of a protein encoded by a gene in the host. The name given to this infectious protein particle is **prion**.

The host gene is called the prion protein (*PrP*) gene, which is a normal part of the genome of mammals and birds. Its polypeptide product, called PrP^{C}, is a naturally occurring protein attached to the outer surface of neurones, and, to a lesser extent, of lymphocytes and other cells; its natural role (including the maintenance of long-term memory) appears to be as a copper-dependent anti-oxidant. The actual infectious agent, called PrP^{Sc}, appears to be a modified form of PrP^{C}, but it still has the same amino acid sequence. One of the more favoured theories for the spread of infection is that when PrP^{Sc} molecules enter a previously uninfected host, they convert the naturally occurring PrP^{C} molecules, produced by the host *PrP* gene, into infectious particles, which ultimately cause clinical signs in that animal, and which also can spread to other animals, both horizontally (by infection) and vertically (via maternal transmission).

This intriguing and unusual example of host–pathogen interaction has puzzled scientists for decades, but in 1986 prion diseases became a major international public-health issue when cattle in Britain developed a completely new disease called bovine spongiform encephalopathy (BSE or mad cow disease). The cause of this new disease is now generally thought to have been the transfer of the scrapie agent from sheep to cattle through the feeding of sheep offal to cattle, although the possibility of it having resulted from a mutation in the bovine *PrP* gene (creating an allele that encodes an infectious form of the peptide) cannot be ruled out. The scrapie agent certainly causes fatality in cattle, but it does not appear to be infectious in cattle, nor is it capable of moving from cattle to any other hitherto uninfected species. As a consequence, initial fears that it could be spread to humans simply by eating sheep or cattle meat seem to be unfounded.

In sheep, there is strong evidence that different alleles at the *PrP* locus confer different levels of susceptibility to scrapie. Most of the important allelic variation occurs in just three coding triplets (136, 154, and 171), and alleles are named according to the amino acids encoded by these coding triplets, using the single-letter code presented in Table 1.2. For example, the most resistant allele encodes a form of the prion protein with alanine (A) at position 136 and arginine (R) at positions 154 and 171; this allele is therefore named *ARR*. In contrast, the most susceptible allele is *VRQ* (i.e. valine at 136, arginine at 154, and glutamine at 171). In January 2002, a Ram Genotyping Scheme was launched as part of the National Scrapie Plan for Great Britain. This scheme involved the use of a DNA test to genotype breeding sheep at the *PrP* locus, followed by the issuing of breeding certificates only to those sheep with favourable genotypes. By 2008, the incidence of scrapie had decreased to such an extent that the scheme was judged to have achieved its aims, and was closed.

The occurrence of BSE has given a new impetus to research into prion diseases, as has the recent discovery of a similar disease in cats and in various captive non-domestic animals such as the Greater Kudu, the cheetah, and the puma. Much will be learnt about this most unusual type of host–pathogen interaction in the years ahead.

Copyrighted material

Hidden page

expression site, in place of the previous duplicated gene, which by this time has been removed and lost.

There are three other aspects of antigenic variation that should be mentioned. The first is that it occurs even in the absence of antibody, which indicates that trypanosomes are somehow 'pre-programmed' to produce a sequence of antigens even in the absence of external stimuli. The second point is that if trypanosomes are removed from a host and are recycled through a tsetse fly, they revert to producing one of the basic antigens when they enter a new host, irrespective of which antigen was last produced in the previous host. This also indicates a degree of 'pre-programming'. Finally, it is obvious that the practical implication of antigenic variation is that it is very difficult to produce a vaccine that will be effective against the large number of different antigens that each host is likely to encounter.

The phenomenon of antigenic variation is an important example of host–pathogen interaction. It is all the more important because it occurs not only in trypanosomes but in other protozoa including members of the genus *Plasmodium* (which causes malaria) and the genus *Babesia* (which causes babesiosis or tick fever, an important disease of cattle that was mentioned in Chapter 8).

Resistance in hosts

There is no shortage of examples of genetic variation in hosts for resistance to pathogens or parasites. Indeed, this type of variation is present in most, if not all, populations of hosts, in relation to most, if not all, parasites and pathogens that have been investigated. This genetic variation exists between and within breeds, and is usually multifactorial. Among the many documented examples are genetic variation for resistance to various bacteria, viruses, and protozoa in chickens; for resistance to bacteria in turkeys; for resistance to bacteria, viruses, and protozoa in fish; for resistance to insects (fly strike), bacteria (fleece rot, footrot), nematodes, and various viruses in sheep; for resistance to bacteria (atrophic rhinitis and scours) in pigs; and for resistance to arthropods (cattle tick, horn fly, stable fly), nematodes, trypanosomes, and bacteria (mastitis) in cattle. We shall consider just two examples below, in which genes for resistance have been identified.

Resistance to neonatal scours in pigs

As we saw in Chapter 5, a major cause of neonatal scours in pigs are the strains of *Escherichia coli* that have a cell-surface antigen called F4. However, not all piglets are susceptible to F4 *E. coli*. In particular, only those piglets with an F4 receptor on the walls of their intestines are susceptible; those that lack the receptor are resistant. As reported in Chapter 5, the presence or absence of the F4 receptor is determined by two alleles at an autosomal locus, with the allele for presence of the receptor being completely dominant to that for lack of the receptor. Thus resistance to neonatal *E. coli* scours in pigs is under the control of two alleles at a single locus, with resistance being recessive to susceptibility.

Copyrighted material

Resistance to Marek's disease in chickens

Marek's disease in chickens is a neoplastic disease in which the growth of tumour cells is caused by a DNA virus. When a population of the Cornell random-bred control strain of chickens was selected for resistance to Marek's disease, mortality from the disease decreased dramatically from around 50 per cent to less than 10 per cent in only four generations. In another line taken from the same control population, selection for susceptibility increased the mortality to over 90 per cent in the same period. In terms of the concepts discussed in Chapter 6, selection had produced large changes in liability to Marek's disease. The fact that selection was able to alter liability so quickly indicates that there was substantial genetic variation for liability in the base population. In other words, there was substantial genetic variation in the host for resistance to the pathogen.

As briefly mentioned in Chapter 8, it has since been shown that there is a strong association between the MHC histoglobulin B^{21} and resistance to Marek's disease. In fact, much of the change in liability in the selection lines described above was the result of changes in the frequency of the B^{21} allele. This is one of the first cases where a gene that contributes to multifactorial variation in host resistance has been identified. As described at the end of this chapter, many more such discoveries will be made in the future.

Resistance in parasites and pathogens

We shall now turn to parasites and pathogens, first to review the extent of genetic variation in resistance in them, and then to examine the extent to which selection has exploited this variation. We shall consider just one example each from insects, helminths, and microorganisms. Similar examples exist for most, if not all, parasites and pathogens.

Resistance to insecticides in sheep blowflies

The Australian sheep blowfly, *Lucilia cuprina*, is a major cause of loss of income in the Australian wool industry, through damage done by larvae to live sheep, and the consequent loss of wool production and death of sheep. Attempts by humans to control the blowfly have until recently concentrated mainly on spraying (jetting) with insecticides.

The use of insecticides soon gave rise to a predictable pattern of response in the blowfly. In 1955, for example, an organochlorine called dieldrin was released for use, and it was extremely effective at first, giving protection for around 8 weeks or longer after jetting. Within 3 years, however, its effectiveness had dramatically fallen, and sheep were being struck by blowflies only a week or two after jetting. It was said that the blowfly had become resistant to the chemical. In 1958, a new insecticide was released, this time an organophosphate called diazinon. It, too, was very effective at first, but, as with dieldrin, it soon lost its effectiveness.

Copyrighted material

Other chemicals were released in an attempt to keep one step ahead of the blowfly but in each case, it was only a matter of time before the blowfly became resistant.

It is now known that the blowflies became resistant to each chemical in turn because of very strong natural selection favouring alleles for resistance at one or more loci in the blowfly. These resistance alleles usually operate by encoding an enzyme that is able to detoxify the insecticide, or by encoding a variant of the enzyme against which the insecticide acts, such that the variant is still able to function in the presence of the insecticide. Not surprisingly, some of these resistance alleles are located at loci encoding cytochrome-P450 enzymes, which, as we saw in Chapter 9, play a key role in detoxifying drugs. The other two types of resistance loci are those encoding carboxylesterases and glutathione-S-transferases. Prior to the introduction of a particular chemical, the relevant resistance alleles are usually maintained at a low frequency in the overall population of flies by a balance between mutation and selection, which arises because resistance alleles, in the absence of the chemical, are usually disadvantageous. The introduction of a new insecticide produces an immediate change in the relative fitnesses of the three genotypes *RR*, *RS*, and *SS* (where *R* is the resistance allele and *S* is the susceptibility allele), with the result that *RR* now has the highest relative fitness and *SS* the lowest. The inevitable consequence of this change in relative fitness is an increase in the frequency of the *R* allele: the population of flies becomes more resistant.

Switching from one insecticide to another within the same class of chemical compound produces very little change if the mode of action of the chemical, and hence the nature of resistance, is the same in each case. On the other hand, replacing one type of compound with another may cause equally strong selection in favour of a resistance allele at a different locus, in which case the above story is repeated. If the first resistance allele also still exerts some effect on the new insecticide (if there is **cross-resistance**), it remains at a high frequency even after the original insecticide is withdrawn from use. This is what sometimes happens in practice. If there is no cross-resistance, the withdrawal of the original insecticide causes the relative fitnesses of the three genotypes to revert to their original values, and the resistance allele decreases in frequency to its original very low value (provided, of course, that it had not already become fixed in the population).

By using standard gene-mapping techniques, researchers have determined the location of resistance loci in the blowfly genome. There are, for example, two different loci with alleles that confer resistance to organophosphates; one on chromosome 4 and another on chromosome 6. The resistance alleles are incompletely dominant to the susceptibility alleles, and have been recorded at a very high frequency (0.98–1.00) in several natural populations of blowflies. Dieldrin resistance is determined by an allele at a third locus, located on chromosome 5. When dieldrin was withdrawn in 1958, the frequency of the resistance allele was around 0.4. Ten years later, its frequency had fallen to around 0.01, where it has remained ever since. The fact that resistance to organophosphates was developing during the period in which dieldrin resistance was declining indicates that there is no cross-resistance between dieldrin and organophosphates.

Copyrighted material

The above account indicates that the reason for blowflies becoming resistant to insecticides is basically very simple: it represents an excellent illustration of the principles of population genetics.

Not surprisingly, the real situation is a little more complicated than outlined above. There are, for example, modifying genes of relatively small effect that play a role in insecticide resistance. In other words, there is multifactorial resistance in addition to single-gene resistance. The longer a particular insecticide is used, the more are those modifying genes subjected to selection, with the result that the major resistance gene becomes less and less likely to decrease in frequency when the insecticide is eventually withdrawn. Despite complications such as this, the above picture highlights the important general principles involved in the development of insecticide resistance in many species of insect, and of resistance to chemicals in general.

We can conclude that the inevitable consequence of the widespread use of an insecticide is that insects will become resistant to it.

Resistance to anthelmintics

Once again, the same principles apply. For example, just three years after the release in the USA of drenches based on the compound thiabendazole (TBZ), resistant strains of *Haemonchus contortus* were detected. With the continual widespread use of this and other benzimidazoles (BZ) in several major sheep-producing regions of the world, resistance to various forms of BZ has now become a major problem. Worse still, worms that are resistant to BZ sometimes show resistance to other compounds used as drenches, including organophosphates, salicylanilides, and substituted nitrophenols.

Researchers have shown that in some cases, resistance is due to a particular allele with a relatively large effect at a single locus. For example, most of the variation in resistance to levamisole in *Trichostrongylus colubriformis* is due to a sex-linked recessive allele. In other cases, however, resistance appears to be multifactorial, being primarily determined by alleles of relatively small effect at an unknown number of loci.

Resistance to antibiotics

In the years following the introduction and widespread use of penicillin, strain after strain of resistant bacteria was isolated. Other antibiotics were introduced, but resistant strains soon emerged. Worse still, many such strains were resistant to more than one antibiotic. The emergence of resistance was too rapid and widespread to be explained solely by the conventional process described for other species of pathogens and parasites in earlier sections, namely the increase in frequency of resistance genes, resulting from selection during the vertical transmission of those genes from generation to generation in the pathogen or parasite. In fact, the rapid and widespread emergence of antibiotic resistance in bacteria has been mainly the result of the ability

Copyrighted material

of bacteria to transfer genes horizontally (between contemporary individuals, within generations) as well as vertically (between generations).

There are three methods by which bacteria transfer genes horizontally. They are **transformation** (the release of naked DNA from one cell, and its uptake by another cell), **transduction** (the transfer of DNA from cell to cell by means of a bacteriophage), and **conjugation** (the transfer of DNA from one cell to another, following the joining together – mating – of the two cells). Antibiotic resistance genes can be transferred horizontally by all three methods, with conjugation probably being the most important. This process involves the transfer from one cell to another of a duplicate copy of a plasmid, which, as we saw in Chapter 2, is a circular segment of DNA that functions independently of the bacterial chromosome. The importance of conjugation lies in the fact that many important genes for antibiotic resistance are located on plasmids. Plasmids that carry one or more resistance genes are called **R factors**.

The importance of horizontal transfer of resistance arises because of the large number of bacteria present in the external and internal environment of animals and of humans. The use of antibiotics creates an environment favouring the survival of strains that possess R factors. These strains then act as a reservoir for the transfer of R factors to any other strains (including pathogenic strains) that happen to be present from time to time.

Many different types of R factors have been identified, including those that have a single gene for resistance to almost any of the antibiotics currently in use. But there are also plasmids known to have genes for resistance to more than one antibiotic. At first it was hard to see how such plasmids could arise, because the chance of more than one relevant mutation occurring in the same plasmid is very small. The answer to this puzzle came with the discovery in the mid-1970s of **transposons**, which are a type of transposable genetic element (TGE). As described in Chapter 1, TGEs are segments of DNA that can move from one site to another within and between chromosomes. In bacteria, transposons can move within and between plasmids and the main bacterial chromosome. They can also insert themselves into and remove themselves from bacteriophage DNA. The most important aspect of transposons is that they contain one or more genes in addition to those required for transposition. Unfortunately, these additional genes often confer resistance to antibiotics.

Although most transposons carry only one gene for antibiotic resistance, it is easy to see how plasmids with multiple resistance genes have arisen so rapidly; they simply accumulate the appropriate transposons. In fact, a plasmid carrying genes for resistance to several different antibiotics is really just a plasmid containing the appropriate transposons.

There is abundant evidence of the role played by transposition in the creation of plasmids with multiple resistance, and of the role played by conjugation in enabling the rapid spread of such plasmids, and hence of multiple resistance, both within and between species of bacteria. There is also abundant evidence to indicate that the extensive use of antibiotics throughout the world confers a strong selective advantage for bacteria carrying multiple resistance.

Copyrighted material

Control of parasites and pathogens

Having illustrated the fact that parasites and pathogens show widespread genetic variation for resistance, we shall now consider some of the ways in which humans are attempting to control them, or at least to combat them.

Screw-worm flies

There are two types of screw-worm fly: the Old World fly (*Chrysomya bezziana*) and the New World fly (*Cochliomyia hominivorax*). They are both parasites of warm-blooded animals. The damage they cause arises from the predilection of the larval stage for open wounds. The Old World larvae, for example, cause up to 30 per cent loss of calves in New Guinea as the result of navel strike, and, at the height of their activity (in the 1950s), the New World larvae resulted in an annual loss of around $100 million in cattle production in the southern states of the USA. Because of its economic importance, considerable effort has been directed at eradicating the screw-worm fly from the USA, with a large degree of success.

The biggest contribution to this success has been a method of biological control known as the **Sterile Insect Release Method**, SIRM. This method involves rearing large numbers of larvae of both sexes in the laboratory, and exposing the late-stage pupae to doses of radiation sufficiently high to render the resultant flies sterile. The sterile flies are then dropped from aeroplanes at rates of between 1600 and 4000 per square mile per week. The principle of the method is simply that if sufficient sterile flies can be released, most (and preferably all) wild-type flies mate with a sterile fly rather than with a wild-type (fertile) fly. Obviously this occurs only if very large numbers of sterile flies are released.

A judicious combination of SIRM and dipping of cattle resulted in the complete eradication of the screw-worm fly from south-eastern USA by 1959, only 2 years after the programme had begun. The south-western eradication programme, which was a much more ambitious task, was commenced in 1962. By 1971 a fly-free barrier 200–300 miles wide had been established along the USA–Mexico border, and all areas north of this barrier were free of the fly. The Mexican–American Commission for the Eradication of Screw-worm (MACES) then gradually moved the barrier southwards. By 1991, the fly had been eradicated from Mexico. Obviously, the narrower the stretch of land over which the barrier has to extend, the easier and cheaper it is to maintain. Because of this, the barrier has since been moved down to Panama.

Despite the success of the eradication programme, the screw-worm fly was not yet prepared to admit complete defeat. In 1988, an outbreak was reported in Libya, most probably having arisen from the importation of infested meat from South America. The establishment of the screw-worm fly on the African continent caused very serious concern in many countries, because it seemed likely that it could spread throughout Africa, and thence to southern Europe, the Middle East, and Asia. In northern Africa alone, it was estimated that the annual cost of treatment and prevention would be $360 million and the threat to wildlife and humans (which it also attacks), especially in central Africa, was extremely serious.

Copyrighted material

In the face of this potential catastrophe, the Food and Agriculture Organization of the United Nations (FAO) set up a Screw-worm Task Force which established the Screw-worm Emergency Centre for North Africa (SECNA) in mid-1990. After intensive investigations of possible courses of action, it was decided that an SIRM programme would be most appropriate, and that MACES would be asked to supply the sterile pupae. Although this would mean chartering aircraft to fly the pupae across the Atlantic Ocean from Mexico to Libya, it would be cheaper and quicker than establishing a new facility in Libya.

Shipments commenced in December 1990 at the rate of 3.5 million flies per week, and were gradually increased until by May 1991, 40 million flies were being flown across the Atlantic Ocean each week. This programme continued until October 1991, by which time it was evident that the screw-worm fly had been eradicated from Libya. The total cost of this operation was around $75 million, of which $25 million was provided by Libya, with the remainder coming from other countries. Given the devastation that would have resulted if the fly had become established, it is not surprising to learn that an independent economic analysis of the programme indicated a benefit : cost ratio of 50 : 1.

The use of SIRM in helping to eradicate the screw-worm fly, both in North America and in Libya, is commonly regarded as one of the most successful of all human attempts at biological control.

Although it has come to be regarded as a form of genetic control, the SIRM itself involves no genetics. It does, however, have genetic implications in relation to problems caused by adaptation of laboratory-reared flies, and also in relation to issues such as the possibility of selection for mating preference. (The SIRM obviously imposes very strong natural selection for the ability of wild-type flies to identify and mate with wild-type flies rather than laboratory-reared flies.) In addition, the sterilization procedure can be regarded as giving rise to dominant lethal alleles.

Other insects

The conventional SIRM involves the maintenance of laboratory-reared insects right up to the adult stage (including irradiation at the late pupal stage) and the release of adults from aeroplanes or other vehicles. If a genetic source of sterility could be used as an alternative to radiation, the provision of labour and facilities for irradiation would be unnecessary and the insects could be released at the larval stage, which is much simpler and far cheaper. Another substantial advantage of larval release is that the resultant adults are likely to be much better acclimatized, having pupated naturally in the soil, and having then emerged naturally into the area where they are expected to perform.

What are potential sources of genetic sterility? A possible answer could be some of the chromosomal rearrangements which were described in Chapter 4. Of course, in that chapter, we were discussing the effect of chromosomal rearrangements on fecundity of animals. But exactly the same principles apply to insects. With this in mind, there was considerable research into the creation (by once-only irradiation) of strains of laboratory-raised insects that carry one or more reciprocal translocations

Copyrighted material

between autosomes and the Y chromosome. All males in such stocks are heterozygous for the translocation, and consequently show the reduced fecundity that is characteristic of such chromosomal rearrangements. If the autosome involved in the translocation contains normal (wild-type) alleles at all loci, and if the translocation is introduced into a laboratory population that is homozygous for a deleterious recessive allele (e.g. for eye colour or wing deformities) at a locus on the same autosome, males survive (because they have one wild-type allele on the translocated autosome, in addition to being homozygous for the deleterious allele on their normal autosomes) but females do not, because, not having a Y chromosome, they lack the translocated autosome and hence lack the wild-type allele.

The combination of reduced fecundity and deleterious recessives has the potential for powerful biological control of insects. At the time of writing, however, none of this has yet come to fruition.

Worms

In principle, the techniques described above for the biological control of insects are also applicable to worms. In practice, relatively little research has been conducted in this area.

In the foreseeable future, animal producers will have to combat worms as best they can, by drenching. Fortunately, there is now a wealth of information available on the extent of cross-resistance between different classes of anthelmintics, which forms the basis of recommendations for the timing and rotation of drenches, with the aim of slowing the spread of resistance.

Bacteria

The long-term outlook for the control of bacteria is not good. As described earlier, bacteria have very effective means of transferring genes for antibiotic resistance horizontally as well as vertically, and the continued widespread use of antibiotics has led to an alarming increase in the occurrence of resistant bacteria. One of the few steps that could be taken to ease the selection pressure for resistant bacteria would be to limit the use of antibiotics, for example, by banning the use of antibiotics as additives in animal food. But this use of antibiotics is such a lucrative business that there is very strong opposition to suggestions that it should be banned, with the result that antibiotics are still used far more widely than is necessary for purely therapeutic needs. It seems that we will have to learn to live with ever-increasing resistance to antibiotics in bacteria.

Increasing the level of resistance in hosts

In addition to the direct means of controlling pathogens and parasites described above, it is possible to exert some indirect influence over them, by increasing the level of resistance in hosts.

Copyrighted material

Hidden page

Practical implications of host–pathogen interactions

In using any of the above approaches to alter resistance in the host, we must never forget the practical implications of the dynamic equilibria discussed at the beginning of this chapter. If you change one side of the ledger (e.g. the level of resistance in the host), the other side (pathogen or parasite) is automatically exposed to natural selection for overcoming whatever change has occurred. Ideally, the change in the host should be sufficient to provide an insurmountable hurdle to the pathogen or parasite. The challenge is in determining which change in the host is most likely to create such a hurdle. Whichever path is taken, we should not be complacent about the powers of natural selection to respond to the challenges that will be provided by the development of increased resistance in hosts.

Summary

- In most situations in which hosts and parasites/pests/pathogens interact with each other, a dynamic equilibrium results as each side attempts to get the upper hand
- A good example of this is the dynamic equilibrium between rabbits and the myxoma virus that causes myxomatosis
- A most challenging example of host–pathogen interaction concerns the prion diseases, in which it appears that the infective agent is actually a particular configuration of the peptide product of a host gene
- Some pathogens keep one step ahead of their host's immune response by the process of antigenic variation, which is the occurrence of a sequence of different antigenic variants all arising from the single population of pathogens that originally entered the host
- There is considerable genetic variation for resistance in hosts to parasites and pathogens
- There is also, unfortunately, considerable genetic variation in pests, parasites and pathogens for resistance to chemicals, which explains why resistance develops
- In fact, the most likely consequence of the use of almost any chemical is that the organism against which it is directed will become resistant to it; and the more effective the chemical is initially, the faster will resistance develop
- For some insect pests, biological control in the form of the sterile insect release method has been successful
- Given the ubiquity of genetic variation for resistance in hosts, selection for resistance is almost certain to increase resistance, but only via the continual exposure of the hosts to the pest, parasite, or pathogen, which is not desirable
- DNA markers for resistance are a possible solution to this quandary, enabling the selection of naturally-resistant animals without the need for all animals to be exposed

Copyrighted material

Further reading

Aguzzi, A., Baumann, F. and Bremer, J. (2008) The prion's elusive reason for being. *Annual Review of Neuroscience*, **31**, 439–77.

Alexander, J.L. (2006) Screwworms. *Journal of the American Veterinary Medical Association*, **228**, 357–67.

Anon. (2008) Special issue: prion disease in animals. *Veterinary Research*, **39** (4).

Beaumont, C., Chapuis, H., Protais, J., Sellier, N., Menanteau, P., Fravalo, P. and Velge, P. (2009) Resistance to Salmonella carrier state: selection may be efficient but response depends on animal's age. *Genetics Research*, **91**, 161–9.

Beraldi, D., McRae, A.F., Gratten, J., Pilkington, J.G., Slate, J., Visscher, P.M. and Pemberton, J.M. (2007) Quantitative trait loci (QTL) mapping of resistance to strongyles and coccidia in the free-living Soay sheep (*Ovis aries*). *Internationa Journal for Parasitology*, **37**, 121–9.

Bishop, S.C. and Morris, C.A. (2007) Genetics of disease resistance in sheep and goats. *Small Ruminant Research*, **70**, 48–59.

Bishop, S.C., Axford, R.F.E., Nicholas, F.W. and Owen, J.B. (eds.) (2009) *Breeding for disease resistance in farm animals*, 3rd edn. CABI Publishing, Wallingford, UK.

Boerlin, P. and Reid-Smith, R.J. (2008) Antimicrobial resistance: its emergence and transmission. *Animal Health Research Reviews*, **9**, 115–26.

Conington, J., Cao, G., Stott, A. and Bünger, L. (2008) Breeding for resistance to mastitis in United Kingdom sheep, a review and economic appraisal. *Veterinary Record*, **162**, 369–76.

Courtin, D., Berthier, D., Thevenon, S., Dayo, G.K., Garcia, A. and Bucheton, B. (2008) Host genetics in African trypanosomiasis. *Infection Genetics and Evolution*, **8**, 229–38.

Darwin, C. (1859) *On the origin of species by means of natural selection*, 1st edn. John Murray, London.

Dawson, M., Moore, R.C. and Bishop, S.C. (2008) Progress and limits of PrP gene selection policy. *Veterinary Research*, **39**, 25.

Doeschl-Wilson, A.B., Vagenas, D., Kyriazakis, I. and Bishop, S.C. (2008) Challenging the assumptions underlying genetic variation in host nematode resistance. *Genetics Selection Evolution*, **40**, 241–64.

Duncan, E.J., Dodds, K.G., Henry, H.M., Thompson, M.P. and Phua, S.H. (2007) Cloning, mapping and association studies of the ovine ABCG2 gene with facial eczema disease in sheep. *Animal Genetics*, **38**, 126–31.

Fisher, P., Hedeler, C., Wolstencroft, K., Hulme, H., Noyes, H., Kemp, S., Stevens, R. and Brass, A. (2007) A systematic strategy for large-scale analysis of genotype phenotype correlations: identification of candidate genes involved in African trypanosomiasis. *Nucleic Acids Research*, **35**, 5625–33.

Ffrench-Constant, R.H. (2007) Which came first: insecticides or resistance? *Trends in Genetics*, **23**, 1–4.

Gilleard, J.S. and Beech, R.N. (2007) Population genetics of anthelmintic resistance in parasitic nematodes. *Parasitology*, **134**, 1133–47.

Heifetz, E.M., Fulton, J.E., O'Sullivan, N.P., Arthur, J.A., Wang, J., Dekkers, J.C. and Soller, M. (2007) Mapping quantitative trait loci affecting susceptibility to Marek's disease virus in a backcross population of layer chickens. *Genetics*, **177**, 2417–31.

Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J. and Bishop, S.C. (2008) Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics*, **178**, 1109–15.

Copyrighted material

Ibeagha-Awemu, E.M., Kgwatalala, P., Ibeagha, A.E. and Xin Zhao, X. (2008) A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mammalian Genome*, **19**, 226–45.

James, C.E., Hudson, A.L. and Davey, M.W. (2009) Drug resistance mechanisms in helminths: is it survival of the fittest? *Trends in Parasitology*, **25**, 328–35.

Kaminsky, R., Ducray, P., Jung, M., Clover, R., Rufener, L., Bouvier, J., Weber, S.S., Wenger, A., Wieland-Berghausen, S., Goebel, T., Gauvry, N., Pautrat, F., Skripsky, T., Froelich, O., Komoin-Oka, C., Westlund, B., Sluder, A. and Maser, P. (2008) A new class of anthelmintics effective against drug-resistant nematodes. *Nature*, **452**, 176–80.

Li, A.Y., Davey, R.B., Miller, R.J., Guerrero, F.D. and George, J.E. (2008) Genetics and mechanisms of permethrin resistance in the Santa Luiza strain of *Boophilus microplus* (Acari: Ixodidae). *Journal of Medical Entomology*, **45**, 427–38.

Martinez, J.L. (2008) Antibiotics and antibiotic resistance genes in natural environments. *Science*, **321**, 365–7.

Moen, T., Sonesson, A.K., Hayes, B., Lien, S., Munck, H. and Meuwissen, T.H. (2007) Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genetics*, **8**, 53.

Nieuwhof, G.J., Conington, J. and Bishop, S.C. (2009) A genetic epidemiological model to describe resistance to an endemic bacterial disease in livestock: application to footrot in sheep. *Genetics Selection Evolution*, **41**, 19.

Prusiner, S.B. and McCarty, M. (2006) Discovering DNA encodes heredity and prions are infectious proteins. *Annual Review of Genetics*, **40**, 25–45.

Reiner, G., Kliemt, D., Willems, H., Berge, T., Fischer, R., Kohler, F., Hepp, S., Hertrampf, B., Daugschies, A., Geldermann, H., Mackenstedt, U. and Zahner, H. (2007) Mapping of quantitative trait loci affecting resistance/susceptibility to *Sarcocystis miescheriana* in swine. *Genomics*, **89**, 638–46.

Rosario-Cruz, R., Almazan, C., Miller, R.J., Dominguez-Garcia, D.I., Hernandez-Ortiz, R. and de la Fuente, J. (2009) Genetic basis and impact of tick acaricide resistance. *Frontiers in Bioscience*, **14**, 2657–65.

Sahana, G., Lund, M.S., Andersson-Eklund, L., Hastings, N., Fernandez, A., Iso-Touru, T., Thomsen, B., Viitala, S., Sorensen, P., Williams, J.L. and Vilkki, J. (2008) Fine-mapping QTL for mastitis resistance on BTA9 in three Nordic red cattle breeds. *Animal Genetics*, **39**, 354–62.

Stanford, M.M., Werden, S.J. and McFadden, G. (2007) Myxoma virus in the European rabbit: interactions between the virus and its susceptible host. *Veterinary Research*, **38**, 299–318.

Stear, M.J., Boag, B., Cattadori, I. and Murphy, L. (2009) Genetic variation in resistance to mixed, predominantly *Teladorsagia circumcincta* nematode infections of sheep: from heritabilities to gene identification. *Parasite Immunology*, **31**, 274–82.

Vitezica, Z.G., Moreno, C.R., Lantier, F., Lantier, I., Schibler, L., Roig, A., Francois, D., Bouix, J., Allain, D., Brunel, J.C., Barillet, F. and Elsen, J.M. (2007) Quantitative trait loci linked to PRNP gene controlling health and production traits in INRA 401 sheep. *Genetics Selection Evolution*, **39**, 421–30.

Copyrighted material

11 Single genes in animal breeding

In previous chapters, we discussed many traits that are determined by a single gene, including blood-cell antigens and a large variety of disorders. There are many other single-gene traits, some of which are of particular importance to animal breeding. The aim of this chapter is to discuss such traits, and to consider some of the practical uses to which they can be put.

Coat colour

Coat colour in mammals is caused by the presence in hair and wool of pigment granules consisting of melanins in a protein framework. The melanins are formed by a series of metabolic pathways that convert the amino acid tyrosine into either **eumelanins** (dark colour) or **phaeomelanins** (light colour). Eumelanins are often described as black, but they include brown and derivatives of black and brown. Phaeomelanins contain sulphur. They are often described as yellow, but they can range from a rather bright yellow through to red. Melanin production occurs in organelles called **melanosomes** within cells called **melanocytes**. Once a melanosome has become filled with pigment, it is called a pigment granule, and is secreted from the cell. Many genes are involved in the production of pigments. We shall now discuss some interesting examples.

Examples of coat-colour genes

A coat-colour gene of particular interest is dominant white spotting, which was illustrated in the photograph of the tortoiseshell cat in Chapter 1 (Fig. 1.15). There are two interesting effects of this gene. First, in animals with some white fur, the white is usually on the belly and other ventral surfaces rather than on the back or dorsal surfaces. Second, the larger the total area of white, the larger are the individual patches of orange and non-orange in tortoiseshell cats. The explanation for these two observations comes from research in mice, where the white-spotting gene is known to be one of a number of genes that regulate normal growth and proliferation

Copyrighted material

of cells. In fact, it encodes a protein called mast/stem cell growth factor receptor (MGFR) that protrudes through the cell membrane, relaying 'messages' from outside to inside the cell. As its name implies, the transmembrane domain of the protein is a receptor for a **growth factor** (a protein produced by one type of cell, that acts on another type of cell). The domain inside the cell has tyrosine kinase activity. When a growth factor binds to the receptor on the outside of the cell, this stimulates tyrosine kinase activity inside the cell, which sets off a cascade of phosphorylations, resulting in activation of transcription factors (Chapter 1), which in turn activate genes, resulting in multiplication of stem cells, including melanocyte precursor cells, in the developing embryo. This whole process is known as a **signal transduction pathway**. During embryonic development, the melanocyte precursor cells migrate from the neural crest down either side of the body, eventually meeting at the centre of the belly. The cells then proliferate sideways until they meet neighbouring cells, thereby filling up all available areas, resulting in a solid mass of melanocytes over the entire body.

The white-spotting allele produces a defective transmembrane protein that is less efficient at relaying messages. This results in fewer melanocyte precursor cells, possibly with poorer mobility. The migration consequently stops part-way down the sides and, because there are fewer cells, each cell proliferates to a greater extent before meeting a neighbouring cell, giving rise to larger pigmented areas derived from each precursor cell. This provides an explanation for the two observed effects of the white-spotting allele; the white areas are those that were never reached by the melanocyte precursor cells; and the larger the area not reached by these cells, the larger is each patch of cells derived from a single cell.

Another interesting aspect of the white-spotting gene is that if it is activated at the wrong time, the result can be excess and uncontrolled proliferation of stem cells; in other words, cancer. In fact, at some time in the past, a retrovirus 'picked up' a copy of the white-spotting gene and incorporated it into its own genome. When this retrovirus infects mice, it activates its own copy of this gene at inappropriate times, causing sarcoma. Retroviral genes that cause cancer are called **oncogenes**. The original host version of an oncogene is called a **proto-oncogene**. Thus, the white-spotting gene is actually a proto-oncogene. In this particular case, the proto-oncogene was discovered and named as *c-kit* (where '*c*' indicates that it is a cellular gene) long before its identity with the white-spotting gene was established. The corresponding oncogene is called *v-kit*, where '*v*' indicates the viral form of the gene. Mutations at the *c-kit* gene in humans cause piebaldism, which is the human equivalent of white spotting.

There is one more interesting point. In mice, there is another coat-colour gene called *steel*, which is on a different chromosome to white spotting, and whose mutants give rise to the same phenotype as white spotting. The protein encoded by *steel* is in fact the growth factor (mast/stem cell growth factor, MGF) that binds to the receptor portion of the white-spotting gene product – it is the so-called **ligand** for the white-spotting receptor. Thus, the white-spotting phenotype can result either from a faulty receptor or a faulty growth factor. If either gene is faulty, the appropriate message does not get through, and cell proliferation is incomplete. Similarly,

Copyrighted material

Hidden page

functional *ASIP* gene (the black allele) and a duplication (the white allele) will have white fleece, because one continually functional version of the *ASIP* gene is sufficient to give ubiquitous expression of agouti-signalling protein, and hence white fleece. Consequently, we say that the black allele is recessive; not because of any inherent property of that allele, but simply because one white allele is sufficient to produce entirely white wool.

The really intriguing aspect of this locus in sheep is that there is substantial variation in the number of copies of the duplication, i.e. this is a good example of copy number variation (CNV), first mentioned in Chapter 2. In two groups of white Merino sheep, the number of copies of the *ASIP* gene ranged from three (i.e. two non-functional *ASIP* genes and one duplication) to six (two non-functional *ASIP* genes and four duplications). The existence of this variation imposes a serious challenge in devising a simple molecular genotyping test for carriers of the black allele, because an animal with, for example, four copies of the *ASIP* gene, could be homozygous for an allele comprising the non-functional *ASIP* gene and one duplication, in which case it would never produce a black lamb; or it could be heterozygous for the black allele (one non-functional *ASIP* gene) and an allele with three copies (one non-functional *ASIP* gene and two duplications). It remains to be seen whether there is any association between multifactorial traits, such as the extent of whiteness of the fleece, and the number of copies of the duplication.

An equally fascinating gene is the one responsible for greying in horses. The phenotype is characterized by normal pigmentation at birth, followed by gradual loss of pigmentation as the horse ages, until, at around 6–8 years old, it becomes white, while retaining dark skin pigmentation. A disadvantage of this gene is a relatively high incidence of dermal melanomas, and an associated decrease in life expectancy. This classic phenotype is caused by a 4.6-kb intronic duplication in the gene for syntaxin-17 (*STX17*). This duplication appears to increase the expression of syntaxin-17 and/or a neighbouring gene that encodes nuclear receptor subfamily 4, group A, member 3 (*NR4A3*). Although the details are still to be worked out, it appears that over-expression of one or both of these two genes leads to a hyperproliferation of hair-follicle melanocytes, leading to a gradual depletion of progenitor stem cells, and hence to a gradual loss of hair pigmentation as the animal ages. At the same time, the over-expression of one or both of these genes leads to a proliferation of dermal melanocytes in glabrous skin (skin that naturally lacks hair), which predisposes to melanoma development.

A summary of coat-colour genes that have been characterized at the molecular level in at least one species of domesticated animal, is presented in Table 11.1.

Complications

Although the inheritance of coat colours is straightforward in principle, there are several complications that arise in practice.

One of these is **epistasis**, or interaction between loci. An example of epistasis is the complete masking of the various tabby alleles at the tabby locus by the

Copyrighted material

Hidden page

Hidden page

Hidden page

another transforming growth factor. The discovery that mutations in this gene have a profound effect on meat yield and quality opens the way for elucidating the role of myostatin in meat production, which in turn will suggest novel (possibly non-genetic) ways of enhancing meat yield and quality. There are also genes of large effect on muscular hypertrophy in sheep, namely *callipyge* and *rib eye muscling* (Carwell). Particularly interesting is that the *callipyge* form of hypertrophy represents an epigenetic phenomenon: the *callipyge* phenotype is seen only in heterozygotes that have inherited the allele from their sire – an unusual form of inheritance called **polar over-dominance**. The mutation is a base substitution in a non-coding section of a 90-kb region of DNA that contains six genes: *DLK1*, *DAT*, and *PEG11* have paternal allele-specific expression whereas *GTL2*, *PEG11AS* and *MEG8* are expressed from the maternal allele. When an animal inherits the *callipyge* mutation from its sire but not its dam, the mutation increases the expression of the *DLK1* and *PEG11* genes on the same (paternal) chromosome; the mutation is said to be ***cis*-acting**. The increased expression of these two genes is associated with the muscle hypertrophy of the *callipyge* trait.

Dwarf poultry

There are alleles at several different loci in the domestic chicken that give rise to birds with a markedly lower mature body size. Most of these so-called dwarfing genes seriously affect viability or hatchability and are therefore of no commercial interest. There is, however, a Z-linked dwarfing gene that appears to have some commercial value. It encodes the chicken growth hormone receptor (GHR), and the *dw* allele at this locus produces a defective transcript that is only three-quarters the length of the normal transcript. In at least one strain of broilers, this is caused by a deletion in the GHR gene (Appendix 3.1).

For practical purposes, the *dw* allele can be regarded as recessive. Thus dwarf males are $Z^{dw}Z^{dw}$ and dwarf females are $Z^{dw}W$. Several broiler-breeder lines homozygous for *dw* have been developed. These *dw* lines are used as a source of female parents that are mated to males from a normal broiler strain. Mainly because of reduced feed consumption and a smaller space requirement, *dw* broiler-breeder lines are more economical than normal broiler lines.

Genes for sexing chickens

Poultry breeders are continually faced with the need to separate day-old chicks into males and females. The conventional method by which sexes are distinguished is vent sexing, which requires skilled operators and which is therefore relatively expensive.

A much cheaper alternative, now widely used commercially, is to exploit genetically-determined differences between the sexes which can be distinguished easily in day-old chicks by unskilled workers. The most commonly used differences arise from segregation of alleles at two Z-linked loci: the feathering locus, with the slow-feathering allele (K^+) being dominant to rapid-feathering (k^+); and a feather-colour

Copyrighted material

Hidden page

Hidden page

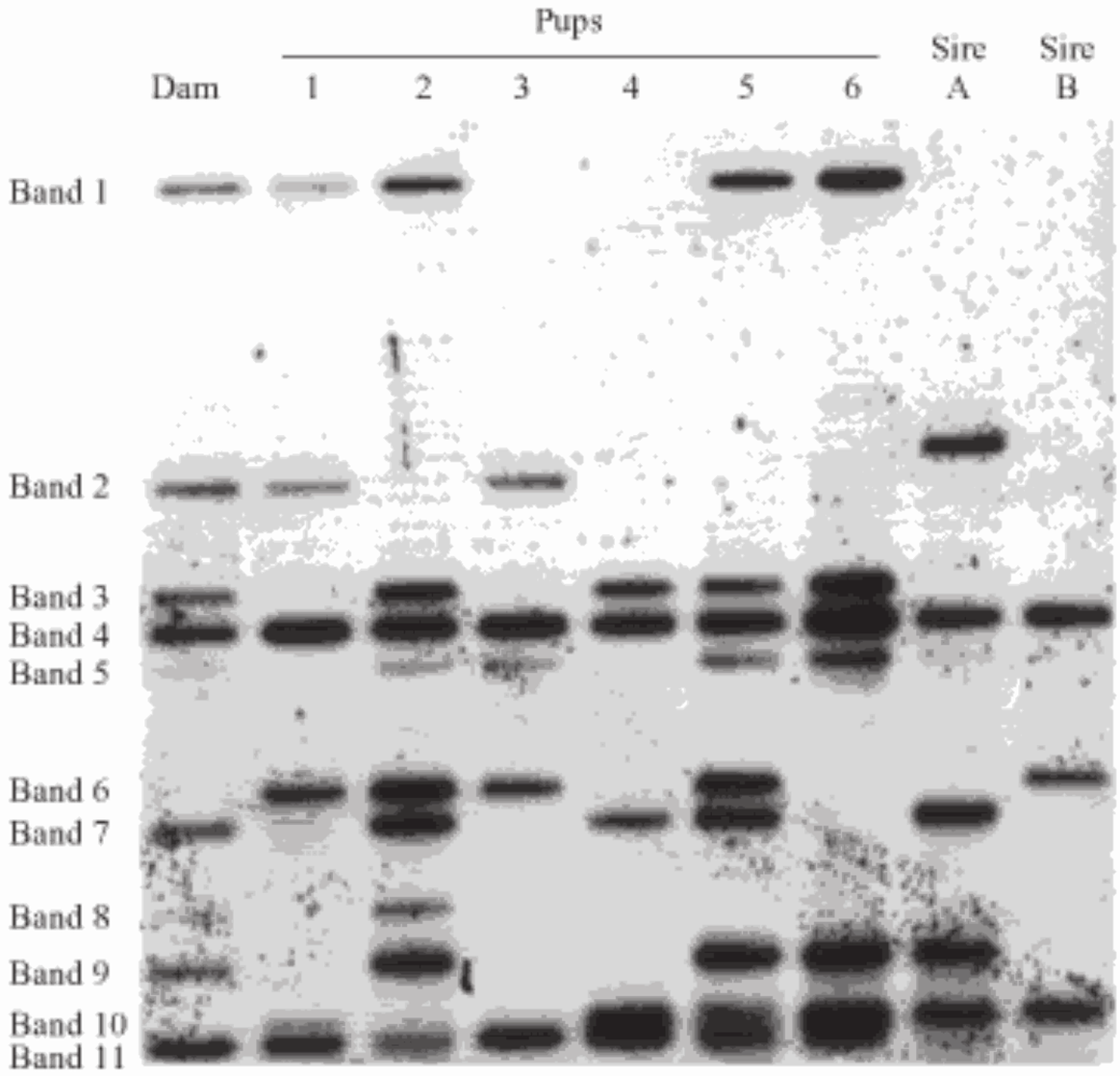


Fig. 11.2 A paternity test in dogs using Southern analysis with the enzyme *Hae*III and a multi-locus human minisatellite probe comprising 29 tandem repeats of a 16-bp core sequence. (Reproduced by permission of I. Hermans from Hermans *et al.* (1991) *New Zealand Veterinary Journal*, 39, 61–4.)

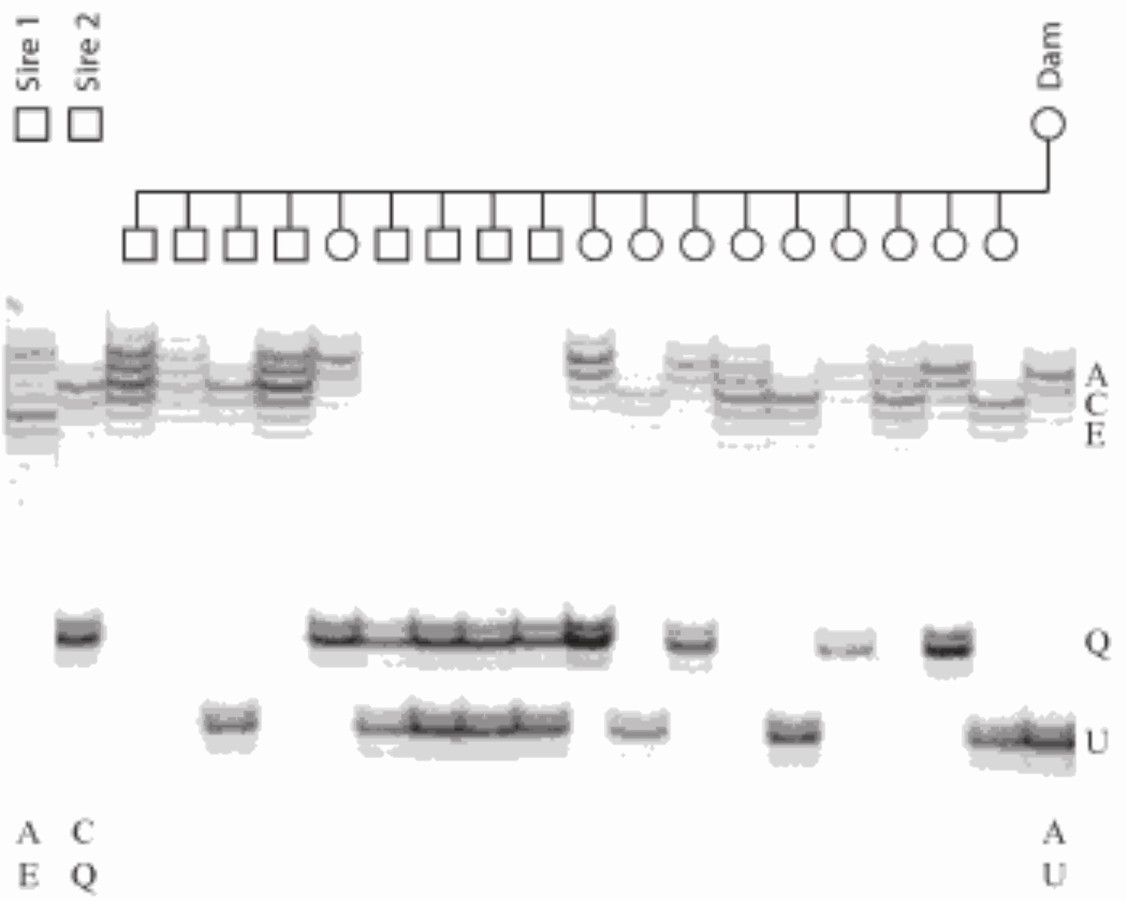


Fig. 11.3 A microsatellite paternity test in sheep, involving one locus with five alleles (A, C, E, Q, U; labelled far right). The presence of a set of bands (stutter bands) for each allele is not unusual with microsatellites. The genotypes of the dam and the two possible sires are given at the bottom of the gel. Sire 1 is excluded by all the offspring with allele Q; sire 2 is not excluded by any offspring. (Reproduced courtesy of J. Maddox.)

Copyrighted material

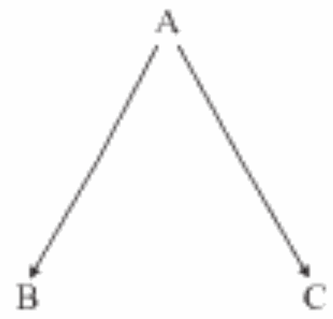
Hidden page

12 Relationship and inbreeding

The aim of this chapter is to provide an understanding of the concepts of relationship and inbreeding, which have important practical applications in the breeding of companion animals, the control of inherited diseases, the management of zoo populations and endangered species, the estimation of heritability, and in artificial selection programmes.

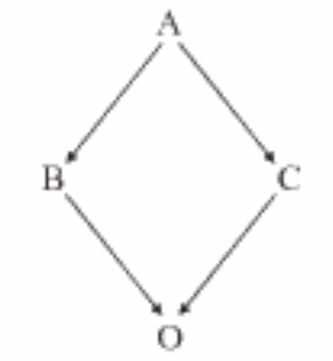
The inbreeding coefficient

The simplest case involves considering a single locus in two offspring, B and C, that have one parent, A, in common. Using arrows to indicate the direction of inheritance, this situation can be represented as follows:



A is called the **common ancestor** of B and C. At any locus, there is a chance that B and C each received a copy of the same gene (the same segment of DNA) that existed in A. When this happens, we say that B and C have genes that are **identical by descent** from A. Genes are identical by descent if they are copies of a segment of DNA that occurs in a common ancestor.

Suppose that B and C mate and produce an offspring, O:



At any locus, if B and C have genes that are identical by descent, there is a chance that O inherits two genes that are identical by descent, one from A through B, and the other from A through C. We are now ready for a definition: the **inbreeding coefficient** of an individual is the probability that the two genes present at a locus

Copyrighted material

in that individual are identical by descent. The inbreeding coefficient is usually given the symbol F.

If an individual has two genes that are identical by descent at a certain locus, the individual must be homozygous at that locus. Thus the inbreeding coefficient is a reflection of homozygosity. However, it does not measure homozygosity in an absolute sense. The reason for this will be explained later in this chapter.

We shall now consider relationship in some detail, before returning to inbreeding.

Relationship

The most common situation, and the one most easily understood, is where neither of the related individuals is inbred. In this section, therefore, unless otherwise stated we shall assume that relatives are not inbred, i.e. we shall assume that $F = 0$ for each relative.

Direct relationship

The most basic relationship is that between a parent (P) and its offspring (O). At any locus an offspring has two genes, one of which (the paternal gene) is a copy of one of its father's genes, and the other of which (the maternal gene) is a copy of one of its mother's genes. Because the same is true for all loci, we can say that an offspring has exactly $\frac{1}{2}$ of its genes in common with each of its parents. We can now state a simple definition: the **relationship** between two non-inbred individuals is the expected proportion of genes that the two have in common. The relationship between offspring and parent, therefore, is $\frac{1}{2}$.

What is the relationship between offspring (O) and grandparent (GP)? In order to answer this question, we need to follow the transmission of genes from a particular grandparent (GP_1) to a parent (P) to an offspring (O). It is obvious that GP_1 transmits exactly $\frac{1}{2}$ of its genes to P who in turn transmits exactly $\frac{1}{2}$ of its genes to O. On average therefore, O receives $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ of its genes from GP_1. Thus, the relationship between an offspring and any one of its grandparents is $\frac{1}{4}$.

By an argument similar to that used above, the relationship between an offspring and any one of its great-grandparents is $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$. It should now be evident that the relationship between an individual and an ancestor decreases by $\frac{1}{2}$ for each generation that separates the individual from that ancestor. Thus, the relationship between an individual and an ancestor is $(\frac{1}{2})^n$, where n is the number of generations between the individual and the ancestor in the pathway of direct descent from the ancestor to the individual.

It follows that irrespective of how superior a particular ancestor may have been, if it occurs only once in an individual's pedigree and several generations back in the pedigree, there is only a small chance that the individual in question inherited any of that ancestor's genes. It is therefore pointless to select an individual solely because its pedigree contains a distant ancestor that happened to be very famous.

Copyrighted material

The situation is a little different if a particular ancestor occurs more than once in a pedigree. In this case, each occurrence of the ancestor in the pedigree provides an independent opportunity for the individual to have inherited the ancestor's genes. The total relationship is the sum of the independent contributions from each pathway of direct descent. Thus, if an ancestor occurs several times in a pedigree, and not too far back in the pedigree, the relationship between the individual and the ancestor could be quite high. This is called **line breeding**.

The relationships considered above are called **direct relationships** because they refer to relationships that can be traced in a direct line or pathway of descent from an ancestor to a descendant. The other type of relationship is where two individuals have an ancestor in common, and may therefore each have inherited the same gene from that ancestor. This is called **collateral relationship**.

Collateral relationship

We shall return to the simplest situation described earlier, which involves two offspring, B and C, and their parent, A. Having one parent in common, B and C are called **half-sibs**. As we saw earlier, if B and C each receive a copy of the same gene at any locus in A, we can say that B and C have genes that are identical by descent from A. We can now consider a more general definition of relationship: the **relationship** between any two individuals is the expected number of genes at a locus in one individual that are identical by descent with a randomly chosen gene at the same locus in the other individual. This definition is correct in all situations, whether the related individuals are inbred or not. If the related individuals are not inbred, this definition is equivalent to the simpler one given earlier.

As with direct relatives, the relationship between two collateral relatives decreases by $\frac{1}{2}$ for each generation that separates the two relatives along a certain pathway. The only difference is that in the case of collateral relatives, the pathway between them consists of two lines of descent; from the common ancestor to one relative and from the common ancestor to the other relative. To emphasize this, the total number of generations separating two collateral relatives is written as $n + n'$, where n is the number of generations between the common ancestor and the first relative, and n' is the number of generations between the common ancestor and the second relative. Thus for any pathway, the relationship between two collateral relatives is $(\frac{1}{2})^{n+n'}$.

Sometimes there may be more than one pathway through a certain common ancestor, and sometimes there may be more than one common ancestor. In all these cases, the overall relationship is the sum of the relationships contributed by each pathway.

The inbreeding coefficient revisited

We are now in a position to derive an equation for calculating the inbreeding coefficient, by giving further consideration to the case where individual A has two offspring B and C, who mate to produce offspring O.

Copyrighted material

Consider a randomly chosen gene that B passes on to O. In order to calculate the inbreeding coefficient of O, we need to calculate the probability that the other gene at the same locus in O (the gene from C) is identical by descent with the gene from B. We can do this by noting that the relationship between B and C is the expected number of genes in C that are identical by descent with a randomly chosen gene at the same locus in B. However, the randomly chosen gene in B is the gene that was passed from B to O. Since there is a chance of $\frac{1}{2}$ that any gene in C is passed on to O, it follows that the probability that O receives a gene from C that is identical by descent with the gene that O received from B, is equal to the relationship between B and C, multiplied by $\frac{1}{2}$.

In general, the **inbreeding coefficient** of any individual is $\frac{1}{2}$ the relationship between the parents of that individual. To calculate the inbreeding coefficient of any individual, therefore, all we have to do is to calculate the relationship between its parents, and then multiply by $\frac{1}{2}$. If the relationship of the parents is not known, the inbreeding coefficient of their offspring is assumed to be zero. The justification for this will be given later in this chapter.

This is the time to mention another measure of relationship between two individuals, namely **coancestry**, which is simply half the relationship as defined above. In other words, the coancestry between two individuals is equal to the inbreeding coefficient of an offspring that would result if they were to mate.

A general expression for relationship and inbreeding

We can now derive a general expression for calculating the relationship between any two individuals, and hence determining the inbreeding coefficient of their offspring.

First we need to consider the effect on relationship if the common ancestor in a certain pathway is itself inbred, i.e. if the common ancestor has an inbreeding coefficient greater than zero.

If the common ancestor is inbred, there is a greater chance that it transmits the same gene to each of the collateral relatives descended from it, which increases the chance that the two relatives each have genes that are identical by descent. We can cater for this by multiplying the contribution of each pathway by $(1 + F_A)$, where F_A is the inbreeding coefficient of the common ancestor in that pathway. As before, if we know nothing of the parents of a certain common ancestor, the inbreeding coefficient of that ancestor is assumed to be zero.

Combining all of the above discussions, we end up with the relationship between any two individuals being given by

$$a = \sum_{i=1}^{p} \left(\frac{1}{2}\right)^{n_i + n_i'} (1 + F_{A_i})$$

where a is the symbol for relationship, Σ indicates summation, p is the number of pathways, n_i is the number of generations between the common ancestor and the

Copyrighted material

first relative (along the ith pathway), n_i' is the number of generations between the common ancestor and the second relative (along the ith pathway), and F_{A_i} is the inbreeding coefficient of the common ancestor in the ith pathway. In the case of direct relationships, there is only one line of descent for each pathway, and so $n_i' = 0$ for each pathway.

The relationship calculated by the above formula is called the **additive relationship**, which is why it is given the symbol a. It is also sometimes called the **numerator relationship**. It corresponds to the exact definition of relationship, namely the expected number of genes at a locus in one individual that are identical by descent with a randomly chosen gene at the same locus in the other individual. If neither of the relatives is inbred, it also equals the expected proportion of genes in common between the two relatives.

Once the additive relationship has been calculated for any pair of individuals, it is a simple matter to calculate the inbreeding coefficient, F, of the offspring of these two individuals as

$$F = \frac{1}{2}a$$

Examples of the use of these two formulae are given in Appendix 13.1 of *Veterinary Genetics* (Nicholas, 1987).

In principle, the use of these two equations for calculating relationships and inbreeding coefficients is relatively straightforward. In practice, however, it can be very complex and tedious, especially for large pedigrees. Fortunately, these are exactly the type of calculations that can be performed readily by computers, and there are now several public-domain packages that calculate relationships and inbreeding coefficients from three-column lists of individual, sire, and dam. A particularly good one is Brian Kinghorn's PEDIGREE VIEWER (already mentioned in Chapter 7 and available at http://www-personal.une.edu.au/~bkinghor/pedigree.htm).

The base population

It was stated earlier that the inbreeding coefficient does not measure homozygosity in an absolute sense. In fact, it measures the decrease in heterozygosity relative to a base population in which all individuals are assumed to be unrelated and to have zero inbreeding. In other words, it measures the extent to which an individual is less heterozygous than individuals that are assumed to have zero inbreeding.

How can measures of relationship and inbreeding be of any use when their actual magnitudes are dependent on the arbitrary choice of a base population, the members of which are conveniently assumed to have no ancestors in common? At first sight, it is difficult to reconcile this with the fact that if we go back a sufficient number of generations, any two individuals must have ancestors in common and hence must be related. In practice, this does not present a problem, so long as we remember that coefficients of relationship and inbreeding are relative, rather than absolute,

Copyrighted material

quantities. Thus, if an individual has a zero inbreeding coefficient relative to a particular base population, we know that this individual has the same level of heterozygosity as members of that base population. If the actual level of heterozygosity in that base population is quite low, due, say, to substantial mating of close relatives among the ancestors of that base population, then the individual with an inbreeding coefficient of zero actually has quite a low level of heterozygosity. The important point is, however, that if a second individual has an inbreeding coefficient of 0.3 relative to the same base population, we immediately know that this second individual is 30 per cent less heterozygous than the individual whose inbreeding coefficient is zero. This highlights the fact that inbreeding coefficients enable the relative levels of heterozygosity to be compared among individuals whose pedigrees can be traced back to the same base population.

Inbreeding in populations

Inbreeding is the mating of related individuals. In previous sections, we have seen that the greater the relationship between two individuals, the greater is the inbreeding coefficient of their offspring. We have also seen that the greater the inbreeding coefficient of an individual, the lower is the chance of that individual being heterozygous. It follows that the greater the extent of inbreeding within a population, the lower the frequency of heterozygotes within that population. In the extreme case where all members of a population are completely inbred ($F = 1$), all individuals are homozygous at all loci, and the frequency of heterozygotes is zero.

From this, it is apparent that inbreeding changes genotype frequencies: it decreases the frequency of heterozygotes and increases the frequency of homozygotes. However, it is important to realize that inbreeding does not, on average, change gene frequencies: the effect of the decrease in heterozygosity is simply to relocate genes from heterozygotes to homozygotes.

Inbreeding depression

In Chapter 5, we saw that harmful recessive genes are maintained in populations by a balance between selection and mutation, and that the majority of these genes are hidden in heterozygotes. Since inbreeding decreases the frequency of heterozygotes, it tends to bring these harmful recessive genes out into the open, in the form of homozygotes. Consider, for example, a recessive gene with a frequency of 0.05. In a non-inbred, random-mating population, the frequency of homozygotes for this gene, and hence the frequency of the recessive phenotype, is $(0.05)^2 = 2.5$ per thousand. The frequency of heterozygotes is $2 \times 0.05 \times 0.95 = 0.095 = 95$ per thousand. An individual that has an inbreeding coefficient of $F = 0.4$ relative to this base population is 40 per cent less heterozygous; its actual level of heterozygosity is $(1 - 0.4)\ (0.095) = 0.057$. The loss in heterozygosity is $0.4 \times 0.095 = 0.038 = 38$ per thousand. Because this loss in heterozygosity is apportioned equally between the two types of homozygotes (19 per thousand to each homozygote), the

Copyrighted material

Hidden page

A particularly interesting example of an inbred line is the herd of White cattle that has existed in isolation on the Chillingham estate in north-eastern England, at a total population size in the tens rather than hundreds, for more than 300 years; possibly 700 years. With just one sire being involved in nearly all matings until he is overthrown by a son, most of the matings in this herd are sire–daughter or full-sib, resulting in the most intense form of inbreeding possible in cattle. Not surprisingly, when this herd was genotyped for a set of 25 DNA markers, there was a very high level of homozygosity (24 out of the 25 loci). Why has this inbred line survived? The obvious answer is that all of the lethal alleles originally present in the population were, by chance, lost during inbreeding in this particular herd. In other words, the Chillingham White cattle are just like an inbred line that, by chance, survives. For every such surviving line, there are many lines that will have gone to extinction because of the chance fixation of a lethal allele.

Summary

- The inbreeding coefficient of an individual is the probability that the two genes (segments of DNA) at a locus in that individual are identical by descent; it is a measure of the extent to which an individual is less heterozygous and hence more homozygous than individuals in a reference (base) population assumed to be unrelated with zero inbreeding
- The most basic relationship between two individuals is that between parent and offspring: because each offspring receives copies of exactly $\frac{1}{2}$ of its parent's genes, parent and offspring have exactly $\frac{1}{2}$ of their genes in common, and are said to have a relationship of $\frac{1}{2}$
- The relationship between any two individuals that are in a direct line or pathway of descent is called a direct relationship, and equals $(\frac{1}{2})^n$ where n is the number of generations separating the two individuals
- Individuals that share a common ancestor are called collateral relatives
- The relationship between any two individuals is the expected number of genes at a locus in one individual that are identical by descent with a randomly chosen gene at the same locus in the other individual; if any two individuals are not inbred (i.e. have zero inbreeding coefficients), the relationship between them equals the expected proportion of genes that they share
- The inbreeding coefficient of an offspring is $\frac{1}{2}$ the relationship between its parents
- The mating of related individuals is called inbreeding, the result of which is a decrease in the frequency of heterozygous genotypes and a corresponding increase in the frequency of homozygous genotypes
- Inbreeding changes genotype frequencies, but does not, on average, change gene frequencies
- Inbreeding is often associated with decreases in average performance (known as inbreeding depression), especially for viability and reproductive ability

Copyrighted material

Further reading

Ala-Honkola, O., Uddström, A., Pauli, B.D. and Lindström, K. (2009) Strong inbreeding depression in male mating behaviour in a poeciliid fish. *Journal of Evolutionary Biology*, **22**, 1396–406.

Alvarez, G., Ceballos, F.C. and Quinteiro, C. (2009) The role of inbreeding in the extinction of a European royal dynasty. *PLoS One*, **4**, e5174.

Archie, E.A., Hollister-Smith, J.A., Poole, J.H., Lee, P.C., Moss, C.J., Maldonado, J.E., Fleischer, R.C. and Alberts, S.C. (2007) Behavioural inbreeding avoidance in wild African elephants. *Molecular Ecology*, **16**, 4138–48.

Bonte, D. (2009) Inbreeding depresses short and long distance dispersal in three congeneric spiders. *Journal of Evolutionary Biology*, **22**, 1429–34.

Carolino, N. and Gama, L.T. (2008) Inbreeding depression on beef cattle traits: Estimates, linearity of effects and heterogeneity among sire-families. *Genetics Selection Evolution*, **40**, 511–27.

Charpentier, M.J., Widdig, A. and Alberts, S.C. (2007) Inbreeding depression in non-human primates: a historical review of methods used and empirical data. *American Journal of Primatology*, **69**, 1370–86.

Gulisija, D., Gianola, D. and Weigel, K.A. (2007) Nonparametric analysis of the impact of inbreeding on production in Jersey cows. *Journal of Dairy Science*, **90**, 493–500.

Jensen, H., Bremset, E.M., Ringsby, T.H. and Saether, B.E. (2007) Multilocus heterozygosity and inbreeding depression in an insular house sparrow metapopulation. *Molecular Ecology*, **16**, 4066–78.

Kruuk, L.E. and Hill, W.G. (eds) (2008) Evolutionary dynamics of wild populations. *Proceedings of the Royal Society B*, **275**, 593–750.

Luong, L.T., Heath, B.D. and Polak, M. (2007) Host inbreeding increases susceptibility to ectoparasitism. *Journal of Evolutionary Biology*, **20**, 79–86.

McParland, S., Kearney, F. and Berry, D.P. (2009) Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genetics Selection Evolution*, **41**, 16.

Ross-Gillespie, A., O'Riain, M.J. and Keller, L.F. (2007) Viral epizootic reveals inbreeding depression in a habitually inbreeding mammal. *Evolution*, **61**, 2268–73.

Szulkin, M., Garant, D., McCleery, R.H. and Sheldon, B.C. (2007) Inbreeding depression along a life-history continuum in the great tit. *Journal of Evolutionary Biology*, **20**, 1531–43.

Townsend, A.K., Clark, A.B., McGowan, K.J., Buckles, E.L., Miller, A.D. and Lovette, I.J. (2009) Disease-mediated inbreeding depression in a large, open population of cooperative crows. *Proceedings of the Royal Society B*, **276**, 2057–64.

Visscher, P.M., Smith, D., Hall, S.J.G. and Williams, J.A. (2001) A viable herd of genetically uniform cattle – deleterious alleles seem to have been purged in a feral strain of inbred cows. *Nature*, **409**, 303.

Copyrighted material

Hidden page

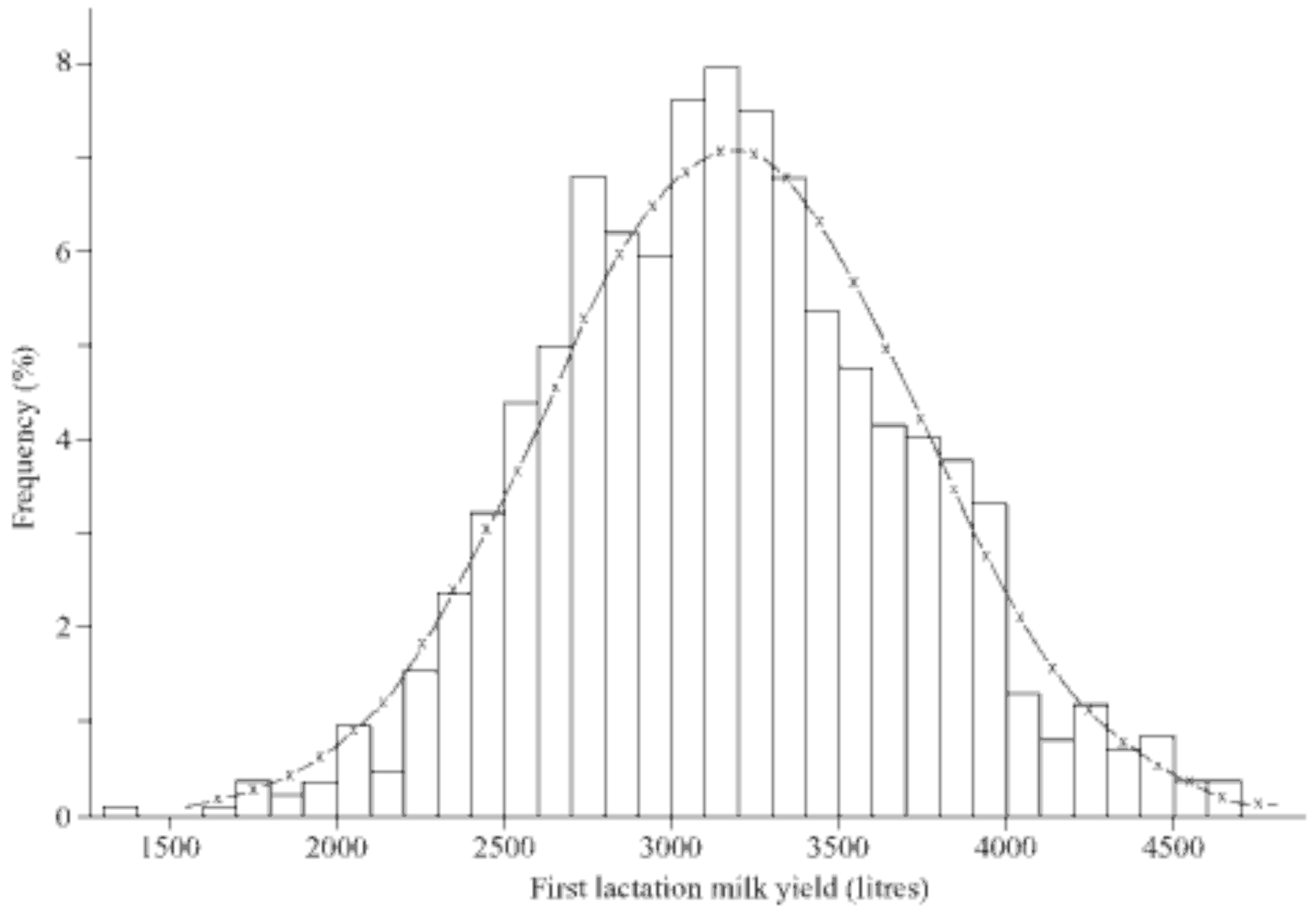


Fig. 13.1 Frequency distribution of first-lactation milk yield in a population of Holstein-Friesian cows. If P_i is the performance (milk yield) of the ith heifer, if n is the total number of heifers, and if Σ indicates summation from $i = 1$ to $i = n$, the mean is estimated as $\Sigma P_i / n$ and the variance is estimated as $\left[\Sigma P_i^2 - (\Sigma P_i)^2/n\right]/(n-1)$. In this population, the mean is 3180 litres and the variance is 315352 (litres)^2. A Normal distribution with this mean and variance has been superimposed on the frequency distribution.

distributed, e.g. faecal worm-egg counts, it is usually possible to find a simple transformation, such as square-root or logarithmic, that enables the useful properties of the Normal distribution to be exploited.

How does continuous variation arise from the action of discrete genes? In order to answer this question, we shall start by considering just one locus. Imagine that there are two alleles at this locus, with allele *A* increasing milk yield by 1 litre and allele *a* giving no increase in milk yield. Imagine further that there is no dominance at this locus, so that the three genotypes *aa*, *Aa*, and *AA* contribute 0, 1, and 2 litres to milk yield. If the frequency of each allele in a herd of cows is 0.5, the frequency distribution of milk yield resulting from variation at this one locus consists of the three classes and is symmetrical, as shown in Fig. 13.2a. Suppose now that exactly the same situation exists at a second locus, with the three genotypes *bb*, *Bb*, and *BB* also contributing 0, 1, and 2 litres to milk yield. Considering these two loci together, the resultant frequency distribution now has five classes and is shown in Fig. 13.2b. As we increase the number of loci contributing to the quantitative trait, we increase the number of classes. Not all alleles have the same effect on milk yield. If there are several loci with a very small effect, the difference between some of the classes is as small as the error of measurement of milk yield, and it is no longer possible

Copyrighted material

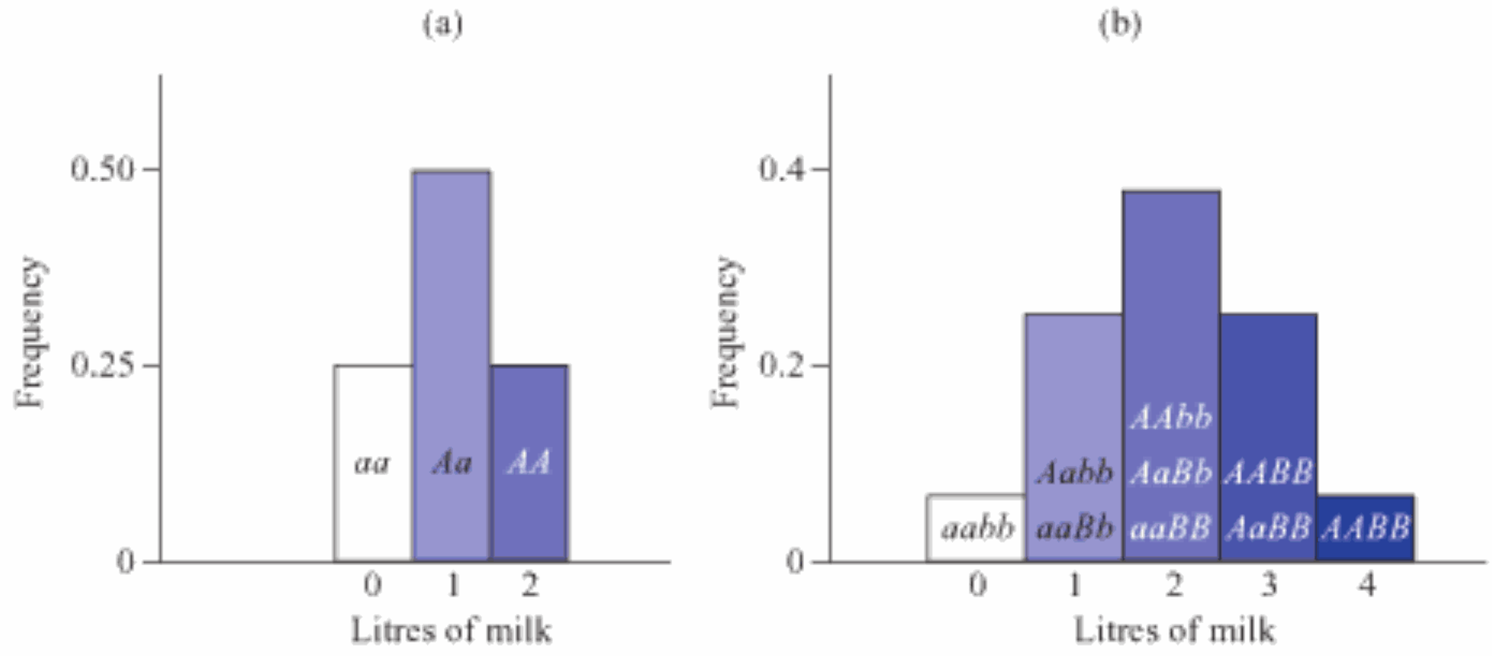


Fig. 13.2 Frequency distribution of milk yield resulting from variation at (a) a single locus and (b) two loci.

always to place heifers in their correct class: the edges of the different classes are blurred. Besides the effect of alleles at a number of loci, there are many non-genetic (often called 'environmental') factors that affect milk yield, and these also tend to blur the edges of the discontinuous classes. If the effect of a particular allele is very large relative to the total variation, discrete classes are still evident even in the presence of non-genetic factors. However, the majority of alleles affecting quantitative traits have an effect that is sufficiently small so as to be not readily detectable.

We can conclude that quantitative traits are determined by the combined action of alleles at many loci (most of which have a relatively small effect on the trait), plus non-genetic (environmental) factors.

The performance of an individual animal

The performance of an animal for a particular trait is called its **phenotypic value**, P, for that trait. The two factors that together determine phenotypic value are **genotypic value**, G, and non-genetic (environmental) deviation, E. The genotypic value of an animal refers to the combined effect of all the animal's genes at all loci that affect the trait, taking account of the way in which the genes are paired together into a genotype. The environmental deviation represents the combined effect of all non-genetic factors that have influenced the phenotypic value. For any trait in any animal, G is determined at conception, while E represents the combined effect of all factors that have influenced the trait in that animal between conception and the time when P is measured.

The above paragraph can be summarized by writing the phenotypic value of an animal for a particular trait as

$$P = G + E \tag{13.1}$$

Copyrighted material

Hidden page

Hidden page

in all non-genetic factors that have influenced the trait from the moment of conception to the time when phenotypic value was measured.

There is one important difference to note between equations (13.1)–(13.4) and equation (13.5). In the first four equations, each component can have a positive or negative sign. This means that we cannot think in terms of a phenotypic value being divided into proportions attributable to A or D or I or E. On the other hand, the components of V_P are always positive and so they do represent proportions of V_P. We can see this most easily by dividing through equation (13.5) by V_P to give

$$1 = \frac{V_A}{V_P} + \frac{V_D}{V_P} + \frac{V_I}{V_P} + \frac{V_E}{V_P} \tag{13.6}$$

The relative contribution of any component to V_P is simply the proportion of V_P attributable to that component.

Heritability

Breeding values are of prime importance in selection programmes, because the main aim when deciding which animals to use as breeders is to choose those that are going to have the best offspring. We are likely, therefore, to be interested in the proportion of phenotypic variance attributable to variance in breeding values. This is represented by the fraction V_A/V_P, which is called **heritability** and which is given the symbol h^2. Strictly speaking, it is called heritability in the **narrow sense**, in contrast to the ratio V_G/V_P, which is heritability in the **broad sense**, and which describes the relative contribution of genotypic variance to V_P. Heritability was mentioned briefly in Chapter 6. Since V_A is never greater than V_G, it is evident that narrow-sense heritability is never greater than broad-sense heritability. Because the concept of heritability is most commonly used in relation to selection programmes, and because it is breeding values that are of prime importance in selection programmes, the term heritability is normally used in the narrow sense unless otherwise stated.

Response to selection

The aim of selection programmes is to select those animals with the highest breeding values. We have also seen that heritability indicates the relative contribution of variation in breeding values to phenotypic variance. We can link these two observations together and explain why heritability is such an important concept, by considering two extreme situations in relation to response to selection.

Consider a population of animals at breeding age. We shall call these animals **candidates** for selection. Suppose that a measurement of phenotypic value for each candidate has been made, so that candidates can be ranked and V_P can be calculated. The top 20 per cent of candidates are to be selected as parents of the next generation, and the remaining 80 per cent are to be culled.

Copyrighted material

Hidden page

Hidden page

it is not surprising that heritability estimates for the same trait vary from one estimate to the next. However, for any particular trait there is a certain degree of similarity between estimates. Table 13.1 lists representative heritability values for various traits in domestic animals. In general, those traits most closely associated with viability and reproductive ability have lower heritability than other traits. This has important practical implications, as we shall see in Chapter 17.

Correlations between traits

As we saw in Chapter 11, it is common for a single gene to affect more than one trait. When the traits are quantitative (i.e. determined by many genes and many non-genetic factors), the extent to which two traits are determined by the same set of genes is called the **genetic correlation**, r_G. Another measure of association between two traits is the **phenotypic correlation**, r_P, which is the correlation between the performance in each trait; it reflects the extent to which performance in one trait is associated with performance in the other trait.

Correlations between traits are estimated using statistical analyses very similar to those used for estimating heritabilities. Increasingly, multi-trait REML analyses are being used to obtain estimates of the heritability of each of a number of traits, together with estimates of correlations between all pairwise combinations of the traits. Values for genetic and phenotypic correlations among some traits commonly measured in domestic animals are given in Table 13.2. These correlations have important practical implications in breeding programmes, as we shall see in Chapter 15.

Quantitative trait loci (QTL)

We have seen that genetic variation in quantitative traits is the result of segregation at many loci. These loci used to be called **polygenes**, but are now called **quantitative trait loci** (QTL). Typically, there is a range in the size of effects of the QTL for any trait, from a few with relatively large effects, down to a large number having very small effects. The technical developments in molecular biology (described in Chapter 2) now make it feasible to identify QTL.

If a large number of animals are genotyped for a set of mapped markers on all chromosomes, and are also measured for a quantitative trait, then, by considering the interval between each pairwise combination of adjacent markers, it is possible to conduct analyses which identify the most likely map locations for QTL, and their effect on the trait. This is called **interval mapping**. A summary of many QTL-mapping results for livestock is provided on the AnimalQTLdb website: http://www.animalgenome.org/QTLdb/. Some results for behavioural traits are included in the Further Reading at the end of this chapter. Just as we would expect, there are tens

Copyrighted material

Table 13.1 Representative values of heritability. For other traits see Table 6.1

	Per cent
Horses	
Log of earnings – jumping	20
Best time	25
Pulling ability	25
Dairy cattle	
Protein yield	25
Milk yield	25
Butterfat yield	25
Beef cattle	
Survival to weaning	5
Weaning weight	25
Postweaning gain	35
Efficiency of gain	35
Final weight	40
Marbling	40
Percentage retail product	65
Sheep	
Probability of conceiving (fertility)	5
Number of lambs per pregnancy (fecundity)	15
Clean fleece weight	40
Fibre diameter	45
Staple length	50
Greasy fleece weight	60
Pigs	
Litter size	10
Killing-out percentage	20
Daily gain	25
Food conversion ratio	35
Eye muscle area	50
Carcase lean proportion	50
Fish (salmon/trout)	
Survival	5
Body length	10
Body weight	20
Chickens	
Fertility	5
Hatchability	5
Liveability	5
Hen-housed egg production	10
Body weight	50
Egg weight	55
Dogs	
Greeting during social contact	5
Aggression on sudden appearance	10
Startle reaction to metallic noise	15
Interest during play	20

Copyrighted material

Table 13.2 Examples of genetic (r_G) and phenotypic (r_P) correlations

	r_G	r_P
Dairy		
Milk yield: fat yield	+0.80	+0.90
Milk yield: protein yield	+0.90	+0.95
Fat yield: protein yield	+0.90	+0.95
Pigs		
Growth rate: food conversion ratio	−0.80	−0.75
Growth rate: food intake	+0.40	+0.45
Growth rate: backfat thickness	−0.25	−0.10
Sheep		
Clean fleece weight: yield	+0.50	+0.40
Clean fleece weight: staple length	+0.40	+0.30
Clean fleece weight: greasy fleece weight	+0.60	+0.80
Chickens		
Egg number: age at first egg	−0.35	−0.25
Egg number: body weight	−0.20	0.00
Egg number: egg weight	−0.30	−0.05

of QTL affecting many traits, spread widely over the genome. Of course, not all QTL are segregating in all breeds.

With DNA-marker chips containing tens of thousands of markers now becoming available, the accuracy and precision of estimates of QTL location and effect can be quite high. As soon as markers associated with QTL have been identified, they can be used in selection programmes. This use of markers in selection is described in Chapter 15.

As QTL are estimated with ever-increasing accuracy and precision, and as genomes become better annotated, increasingly the actual nucleotide polymorphism responsible for a QTL will be identified (so-called **quantitative trait nucleotide**, QTN). An early example of the success of this approach was reported in 2002, when a QTL for fat content in dairy cattle was shown to be the result of a mis-sense mutation in the gene for an enzyme involved in lipid metabolism, namely diacylglycerol acyltransferase (DGAT1). Interestingly, two of the early QTN identified involve mutations in non-coding regions that affect meatiness – one in pigs and one in sheep. The pig mutation is a base substitution (G → A) in an intron of the gene for insulin-like growth factor 2 (IGF2). This mutation removes the recognition site for a transcription factor that represses transcription, resulting in three times the normal quantity of IGF2 mRNA in muscle, leading to increased muscling (and decreased fatness). The sheep QTN is also a G → A base substitution, this time in the 3′ UTR of the gene for myostatin in Texel sheep. The mutation creates a target site for two microRNAs (miRNA) which (via the antisense phenomenon described in Chapter 2) lead to the destruction of myostatin mRNA and hence to a lack of myostatin, which in turn results in increased muscle mass.

Copyrighted material

Hidden page

Hidden page

Barton, N.H. and Keightley, P.D. (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**, 11–21.

Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to quantitative genetics*, 4th edn Longman Cheshire, London.

Georges, M. (2007) Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. *Annual Review of Genomics and Human Genetics*, **8**, 131–62.

Jones, P., Chase, K., Martin, A., Davern, P., Ostrander, E.A. and Lark, K.G. (2008) Single-nucleotide-polymorphism-based association mapping of dog stereotypes. *Genetics*, **179**, 1033–44.

Kukekova, A.V., Trut, L.N., Chase, K., Shepeleva, D.V., Vladimirova, A.V., Kharlamova, A.V., Oskina, I.N., Stepika, A., Klebanov, S., Erb, H.N. and Acland, G.M. (2008) Measurement of segregating behaviors in experimental silver fox pedigrees. *Behavior Genetics*, **38**, 185–94.

Olsen, H.G., Nilsen, H., Hayes, B., Berg, P.R., Svendsen, M., Lien, S. and Meuwissen, T. (2007) Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. *BMC Genetics*, **8**, 32.

Rothschild, M.F., Hu, Z.L. and Jiang, Z. (2007) Advances in QTL mapping in pigs. *International Journal of Biological Science*, **3**, 192–7.

Spady, T.C. and Ostrander, E.A. (2008) Canine behavioral genetics: pointing out the phenotypes and herding up the genes. *American Journal of Human Genetics*, **82**, 10–18.

Van Buskirk, J. and Willi, Y. (2006) The change in quantitative genetic variation with inbreeding. *Evolution*, **60**, 2428–34.

van der Waaij, E.H., Wilsson, E. and Strandberg, E. (2008) Genetic analysis of results of a Swedish behavior test on German Shepherd Dogs and Labrador Retrievers. *Journal of Animal Science*, **86**, 2853–61.

Visscher, P.M., Hill, W.G. and Wray, N.R. (2008) Heritability in the genomics era - concepts and misconceptions. *Nature Reviews Genetics*, **9**, 255–66.

Weller, J. (2009) *Quantitative trait loci analysis in animals*, 2nd edn. CAB International, Wallingford.

Wirén, A., Gunnarsson, U., Andersson, L. and Jensen, P. (2009) Domestication-related genetic effects on social behavior in chickens – effects of genotype at a major growth quantitative trait locus. *Poultry Science*, **88**, 1162–6.

Copyrighted material

Hidden page

The factors that must be taken into account include randomization, equality of treatment, the type of comparison, the number of sires, and the number of repeat measurements.

The question of randomization arises at several levels. In many situations it is best if each population is represented by a random sample of animals. If this is not possible, great care should be taken to see that the method and extent of selection of animals is as similar as possible for each population. Also, if the animals are to be evaluated according to the performance of their progeny when mated to a so-called **tester** population, the mates must be chosen at random from the tester population. Furthermore, if the comparison is to be conducted at more than one site, there must be a random allocation of stock from each population to each site.

The requirement that all stock be treated equally is so obvious that it should not need to be mentioned. However, because there are sometimes quite large profits or losses to be made depending on the outcome of a comparison, all possible precautions should be taken to ensure that the comparison is conducted in a manner that minimizes the chances of any population receiving favoured treatment.

Another factor that affects the efficiency of distinguishing between populations is the type of comparison that is conducted. For example, a **direct comparison** (all populations under test being compared at the same location) is more efficient than an **indirect comparison** (each population being measured in a different location, with comparisons being made indirectly through a tester population measured in all locations). Also, a comparison between straight-bred offspring (where each offspring in the comparison is the progeny of a male and a female belonging to one of the populations under test) is more efficient than a comparison between cross-bred offspring (where the offspring being compared are obtained from mating just one sex, usually male, from the populations under test, to a random sample of members of a tester population).

Another factor that plays a major role in determining the efficiency of a comparison is the number of animals measured in each population under test. There are two important generalizations that can be made in relation to this factor. First, the efficiency of a comparison is determined primarily by the number of sires represented in each population under test, rather than by the number of offspring per sire. Indeed, comparisons involving only, say, two or three sires per population are virtually worthless, even if hundreds of offspring from each sire have been measured, because the average performance of the offspring will be much more a reflection of the average breeding value of those particular sires than of the population from which the sires came. For this reason, when planning a comparison, it is vitally important to ensure that each population under test is represented by sufficient sires to ensure a reasonable representation of the population. Another point to note is that unless the traits being measured have a very low repeatability (see Chapter 15), it is usually better to measure a different set of offspring each season or year than to take repeated measurements on the same set of offspring.

Copyrighted material

Genotype–environment interaction

Comparisons generally involve testing the performance of groups of animals under one or a limited number of environments. The Food and Agricultural Organization (FAO), for example, conducted a large-scale comparison of various Friesian strains from around the world, by using appropriate semen on a large number of cows on collective farms in Poland. Do the results obtained in Poland apply to other countries as well? In other words, would the bulls within countries of origin, and the countries of origin themselves, rank in the same order if the trial had been conducted in, say, Canada or New Zealand?

If the results of comparisons vary depending on the environments in which they are conducted, we say there is **genotype–environment interaction**, G × E. In the examples of G × E illustrated in Fig. 14.1, three different cow genotypes were

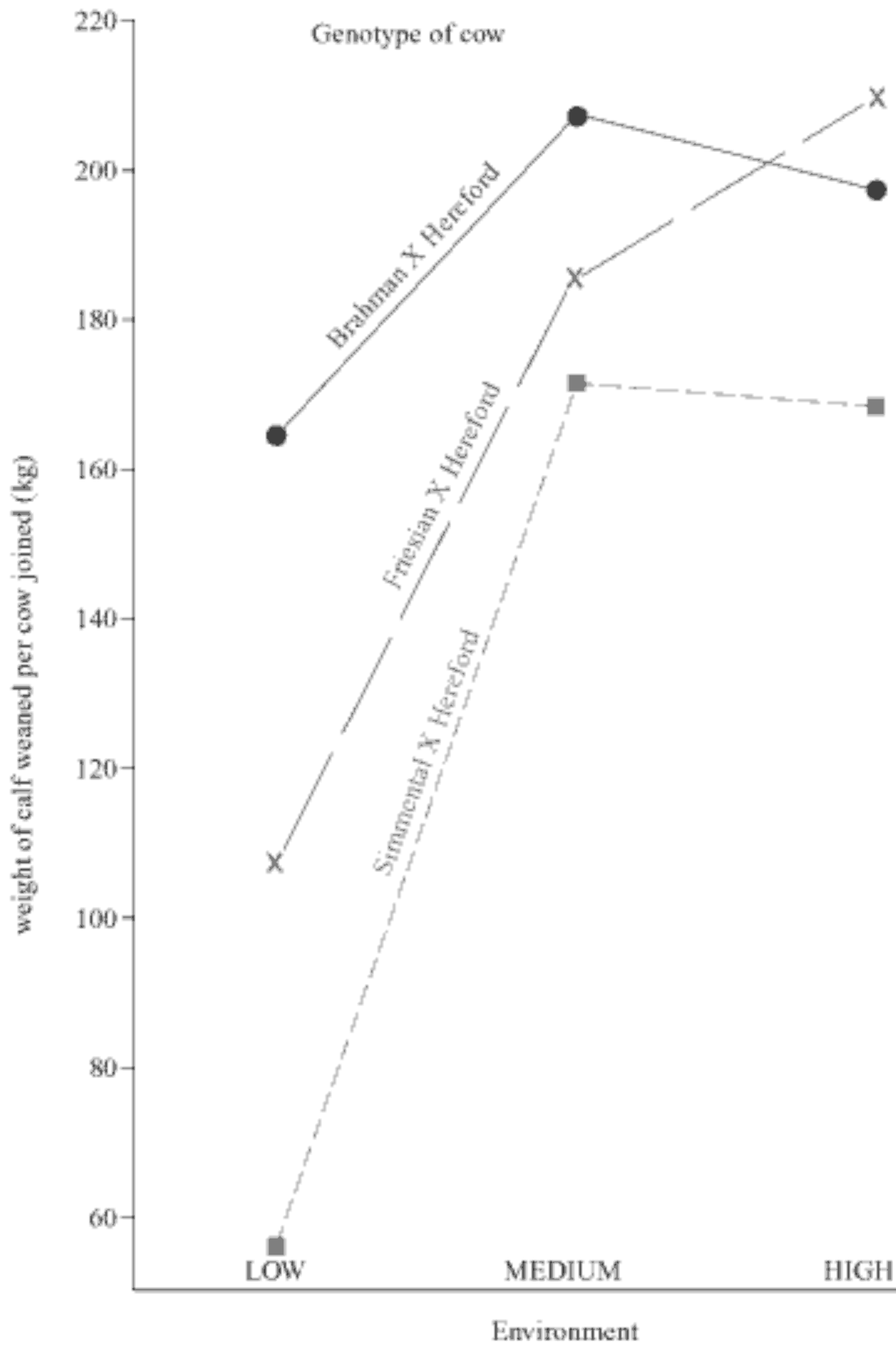


Fig. 14.1 Examples of genotype–environment interaction (G × E) in beef production. (Drawn from data in Barlow *et al.* (1978), *Wool Technology and Sheep Breeding*, **26**, 5–12.)

Copyrighted material

evaluated under low, medium, and high levels of nutrition on a research station in NSW, Australia. All cows were mated to Hereford bulls. The results of this comparison illustrate the two types of G × E that occur. The first involves just a change in relative performance (low versus medium environments), while the second involves a change in ranking (medium versus high environments).

An interesting approach to estimating the extent of G × E is to regard performance in each environment as a different trait, and then to estimate the genetic correlation (Chapter 13) between performance in the two environments. This is done as a matter of course by Interbull, an organization that conducts international genetic evaluations using dairy data from around the world. For protein yield in Holstein–Friesians, as an example, the genetic correlation between performance in the USA and performance in Australia is around 0.8, whereas between New Zealand and Australia it is around 0.9. These estimates reflect the extent to which the dairy production system in New Zealand is more similar to that in Australia than is the production system in the USA. The lower the genetic correlation between two countries, the less emphasis should be placed on evaluation results from one country as a guide to progeny performance in the other country. Fortunately, the standard methods of genetic evaluation (Chapter 15) can readily take account of these genetic correlations, enabling international evaluations to place the appropriate emphasis on data from each country.

The problem with G × E is that in many cases it is difficult to predict in advance whether or not such an interaction is likely to be important. All that can be done in practice is to conduct the comparison in all relevant environments, which is usually not possible, or to interpret the results from a different environment cautiously, in the light of what is known in general about G × E in relation to the traits concerned.

Summary

- Effective selection between breeds requires well-planned breed comparisons
- The efficiency of a comparison is determined primarily by the number of sires represented in each population under test, rather than by the number of offspring per sire
- It is usually better to measure a different set of offspring each season or year than to take repeated measurements on the same set of offspring
- It is not unusual for the relative performance of breeds or strains to alter if they are evaluated in more than one environment, and sometimes rankings may even change; such alterations indicate genotype–environment interaction (G × E), the potential importance of which must be taken into account when evaluating the results of comparative trials
- the extent of G × E can be quantified by regarding performance in each environment as a different trait, and then estimating the genetic correlation (Chapter 13) between performance in the two environments

Copyrighted material

Further reading

Fanatico, A.C., Pillai, P.B., Emmert, J.L. and Owens, C.M. (2007) Meat quality of slow- and fast-growing chicken genotypes fed low-nutrient or standard diets and raised indoors or with outdoor access. *Poultry Science*, **86**, 2245–55.

Fulkerson, W.J., Davison, T.M., Garcia, S.C., Hough, G., Goddard, M.E., Dobos, R. and Blockey, M. (2008) Holstein–Friesian dairy cows under a predominantly grazing system: interaction between genotype and environment. *Journal of Dairy Science*, **91**, 826–39.

Haile-Mariam, M., Carrick, M.J. and Goddard, M.E. (2008) Genotype by environment interaction for fertility, survival, and milk production traits in Australian dairy cattle. *Journal of Dairy Science*, **91**, 4840–53.

Lillehammer, M., Goddard, M.E., Nilsen, H., Sehested, E., Olsen, H.G., Lien, S. and Meuwissen, T.H. (2008) Quantitative Trait Locus-by-Environment Interaction for Milk Yield Traits on *Bos taurus* Autosome 6. *Genetics*, **179**, 1539–46.

Lillehammer, M., Arnyasi, M., Lien, S., Olsen, H.G., Sehested, E., Odegard, J. and Meuwissen, T.H. (2007) A genome scan for quantitative trait locus by environment interactions for production traits. *Journal of Dairy Science*, **90**, 3482–9.

Lillehammer, M., Odegard, J. and Meuwissen, T.H. (2007) Random regression models for detection of gene by environment interaction. *Genetics Selection Evolution*, **39**, 105–21.

Macdonald, K.A., Verkerk, G.A., Thorrold, B.S., Pryce, J.E., Penno, J.W., McNaughton, L.R., Burton, L.J., Lancaster, J.A., Williamson, J.H. and Holmes, C.W. (2008) A comparison of three strains of Holstein–Friesian grazed on pasture and managed under different feed allowances. *Journal of Dairy Science*, **91**, 1693–707.

Shariati, M.M., Su, G., Madsen, P. and Sorensen, D. (2007) Analysis of milk production traits in early lactation using a reaction norm model with unknown covariates. *Journal of Dairy Science*, **90**, 5759–66.

Singh, R., Cheng, K.M. and Silversides, F.G. (2009) Production performance and egg quality of four strains of laying hens kept in conventional cages and floor pens. *Poultry Science*, **88**, 256–64.

Steinheim, G., Odegard, J., Adnoy, T. and Klemetsdal, G. (2008) Genotype by environment interaction for lamb weaning weight in two Norwegian sheep breeds. *Journal of Animal Science*, **86**, 33–9.

Strandberg, E., Brotherstone, S., Wall, E. and Coffey, M.P. (2009) Genotype by environment interaction for first-lactation female fertility traits in UK dairy cattle. *Journal of Dairy Science*, **92**, 3437–46.

Copyrighted material

Hidden page

It is convenient to think of accuracy on a scale from 0 per cent, where the clue provides no information on breeding value, to 100 per cent, where the clue is the true breeding value itself. If accuracy is 0 per cent, animals are, in effect, ranked at random; selecting those animals at the top of the list produces no response to selection, on average. At the other extreme, if accuracy is 100 per cent, animals are ranked according to their true breeding value, and response to selection is maximized.

Strictly speaking, **accuracy of selection** for any trait is the correlation between the available clues and the true breeding value for that trait. The symbol for accuracy is r_{AC}, where r is the usual symbol for correlation, A represents true breeding value, and C stands for clues or criteria.

Clues to a candidate's breeding value for a trait

At regular time intervals during a selection programme, groups of candidates become available for selection and therefore have to be ranked in terms of estimated breeding value, EBV.

There are many potential clues to a candidate's true breeding value.

The simplest and most commonly available clue is a single measurement of its own performance. Measurement of the performance of candidates is called **performance testing**, and selection on the results of performance testing is called **individual selection** or **mass selection**. Other potential clues to a candidate's true breeding value include single measurements on a relative of a candidate (e.g. parent, sib, or offspring), or on groups of relatives, such as full-sibs (called a **dam family**) or half-sibs (called a **sire family**). Selection on ancestor performance is called **pedigree selection**, and on sib performance is called **sib selection**. Evaluating candidates on the performance of their offspring is called **progeny testing**. For some traits, it is possible to measure performance more than once on the same animal. Fleece weight, for example, can be measured each year on each sheep, and maternal ability of beef cows (as judged by the weaning weight of their calves) can be measured on the same cow each time she weans a calf. Such traits are said to be repeatable, and the correlation between repeated measurements on the same animal is called **repeatability**, which ranges from 0 per cent to 100 per cent.

For each of these sources of clues, it is possible to derive an expression for accuracy of selection, as shown in Appendix 16.2 of *Veterinary Genetics*. In all cases, accuracy is a function of heritability; the higher the heritability of a trait, the greater the accuracy of selection.

Combining clues from more than one source

In practical selection programmes, clues to a candidate's true breeding value are often available from more than one source. In selecting hens for egg production, for example, data are usually available on the hen's own performance, on the average

Copyrighted material

performance of the full-sib family of which she is a member (her dam-family average), and on the average performance of the half-sib family of which she is a member (her sire-family average). Corresponding male candidates can be ranked according to the same sire-family and dam-family averages, but obviously without any measurement of their own performance being included. With pigs, both male and female candidates are often ranked in terms of their own performance and that of their sibs.

The accuracy of selection increases as more clues are used. How can the clues be combined together into one overall estimate of a candidate's breeding value, and how can the resultant accuracy be determined?

Imagine that we have k different clues,

$$C_1, C_2, \ldots, C_k$$

where a clue may be:

(1) a single measurement or the average of repeated measurements on a candidate or on any relative, or
(2) the average performance of a group of the candidate's relatives, where the record available on each relative is either a single measurement or the average of repeated measurements.

The single overall estimate of a candidate's true breeding value that is obtained from these different clues is called a **selection index**. It is the sum of all the clues, with each clue being weighted in such a way as to maximize the accuracy of the index as an estimate of the candidate's true breeding value. The index is written as

$$I = \text{EBV} = b_1C_1 + \ldots + b_kC_k \qquad (15.1)$$

where I is the estimated breeding value, EBV, and where the b values are weighting factors calculated so as to maximize the correlation between the index, I, and the candidate's true breeding value, A, i.e. to maximize accuracy of selection.

The general principles for calculating appropriate weighting factors so as to maximize accuracy are very straightforward – they involve a simple application of the standard method of maximizing a function, as taught in high-school calculus courses:

(1) write down an expression for the parameter to be maximized (in this case, accuracy);
(2) differentiate the expression with respect to each b value;
(3) set the differentials equal to zero, thereby creating a set of so-called **normal equations**; and
(4) solve these equations simultaneously for the b values.

The actual information required to calculate the b values is:

Copyrighted material

Hidden page

take correct account of the available clues from all the known relatives of a candidate. Another important advantage of BLUP is that because it can provide EBVs for all animals in a population (using the so-called **animal model**), a good indication of response to selection (**genetic trend**) can be obtained simply by calculating the average EBV of each age group within the population, and then plotting average EBV against year of birth.

Because of these substantial advantages, BLUP has become the method of choice for calculating EBVs. Further discussion of the BLUP method and a simple practical example are given in Appendix 16.3 of *Veterinary Genetics*.

Correlated traits

Selection for one trait almost always produces changes in other traits. A change in an unselected trait resulting from selection of another trait is called a **correlated response**. The magnitude of the correlated response depends, among other things, on the correlations between the two traits, which were discussed in Chapter 13. These correlations have important practical implications for breeding programmes.

First, the direction and magnitude of correlated response in an unselected trait depends on the sign and the magnitude, respectively, of the genetic correlation between the unselected trait and the selected trait. If, for example, the genetic correlation between two traits is negative, an increase in the trait being selected produces a decrease in the unselected trait.

Second, the existence of genetic correlations raises the possibility of improving a trait that is difficult or expensive or even impossible to measure in practical conditions, by selecting on a correlated trait that is easy and cheap to measure. The practice of improving one trait by selecting on another is called **indirect selection**. It is used, for example, in broiler chickens, where growth rate has a high positive genetic correlation with food conversion efficiency. This means that food conversion efficiency can be improved by selecting solely on growth rate, without going to the considerable extra expense of actually measuring food consumption.

Third, the existence of genetic correlations provides an important new source of clues to a candidate's breeding value for any trait; the performance of a candidate and/or of any relative in any trait that is correlated with the trait to be improved, provides additional clues to the candidate's breeding value for the trait to be improved. Thus, even if the trait to be improved can be measured easily and cheaply on candidates themselves, it may be useful to measure the performance of the candidate and/or of relatives for correlated traits, and to make use of all these measurements when estimating the breeding value of candidates.

This can be achieved by using a selection index or BLUP, in exactly the same manner as described earlier, except that now the list of potential clues includes correlated traits as well as, or instead of, the trait to be improved. For example, a sheep selection programme might involve the measurement of fleece weight and fibre diameter on all candidates. Even if no pedigree records are available, it would still

Copyrighted material

be possible to use several different selection indexes, each using the same two clues, namely fleece weight and fibre diameter. For example, a selection index could be used for calculating each candidate's EBV for fleece weight, and another index could be used for calculating each candidate's EBV for fibre diameter. Other indexes could be derived for calculating each candidate's EBV for any trait that is correlated with the two measured traits, but which is not actually measured in the flock, e.g. body weight.

In each case, the clues for each index are exactly the same, namely the candidate's own performance in the two traits; the only differences between the indexes are the b values. For the fleece-weight index, the b values are calculated so as to maximize the correlation between the index and true breeding value for fleece weight; for the fibre-diameter index, the b values maximize the correlation between the index and true breeding value for fibre diameter; and for the body-weight index, the b values maximize the correlation between the index and true breeding value for body weight.

For all indexes, the b values are calculated using exactly the same approach as described for single-trait selection. However, because the clues now consist of more than one trait, estimates of genetic and phenotypic correlations among the traits are required in addition to the heritabilities needed to calculate b values when the clues consist of measurements on only one trait.

If pedigree records are kept, the performance of relatives (for each trait) can be used as additional clues in each index. Alternatively, EBVs for each trait can be calculated using so-called multi-trait BLUP, which combines all available clues from all traits measured on the candidate and on all known relatives, automatically providing the optimum weight (b value) for each clue.

Irrespective of which method is used, the result is a list of EBVs for each trait for each candidate, so that candidates can be ranked for each trait. While this is very useful information, it is not the end of the matter, because when replacements are being selected, we need to rank candidates according to overall merit, taking account of their EBV for each trait.

Selection for more than one trait

There are five steps involved in selecting for more than one trait.

Step 1: decide on the breeding objectives

This involves drawing up a list of the traits that you would like to improve, irrespective of whether they are cheap or expensive or even impossible to measure on the candidate (called **breeding objectives**).

For example, in a pig selection programme, you may wish to improve growth rate (G), food conversion ratio (F), and percentage lean in the carcase (L). The first trait is easy to measure on candidates; the second is expensive to measure, and hence may be recorded on only a limited number of candidates; while the third trait is

Copyrighted material

impossible to measure on candidates, because it can be recorded only from carcases.

Step 2: decide on the selection criteria

As we have already seen, selection criteria are the traits that are actually measured on candidates and/or their relatives. These may be the same as the breeding objectives, or they may be different; there are no limitations on which traits can be used as selection criteria, providing that they are measurable, and providing that estimates of the relevant genetic and phenotypic parameters are available.

For the pig example just mentioned, suppose that the only two traits that are actually measured are growth rate (G) and backfat thickness (B). The first of these traits is one of the objectives, and is correlated with the other objectives; the second measured trait is not an objective, but is correlated with the objectives (see Table 13.2).

Step 3: calculate the EBV for each breeding objective for each candidate

This involves using one of the approaches described earlier.

Each candidate's EBV for growth rate could be calculated just from the candidate's own performance for that trait. But the EBV would be estimated more accurately if a selection index were used to combine information from both clues (growth rate and backfat thickness) from each candidate, and preferably from relatives as well. For the other two objectives (the ones that are not actually measured), a selection index has to be used to combine clues from the measured traits on the candidates and/or on relatives. Alternatively, maximum accuracy of selection can be achieved if multi-trait BLUP is used to calculate the EBVs for each objective for each candidate.

Step 4: determine the net economic value of improvement in each breeding objective

For each objective, the **net economic value**, v, is the additional profit expected from a unit improvement in that objective alone. This step is often not as simple as it might at first appear. However, software packages for calculating net economic values are gradually becoming available to individual breeders, who, with the help of suitably trained advisers, can calculate a set of values appropriate to their particular circumstances.

In the past, and commonly still today, breeders have guessed the net economic value for each of their objectives. In many cases, this has been done subconsciously. The important thing to realize is that whenever there is more than one objective in a breeding programme, net economic values must be determined in one way or another before selection can be conducted. Furthermore, this statement is as true for the prehistoric humans who domesticated animals, as it is for present-day breeders.

Copyrighted material

Step 5: carry out selection

Once the net economic value has been either guessed or calculated for each breeding objective, there are two main ways in which selection is conducted.

(1) Independent culling levels

This method involves selecting animals which exceed a particular level (the culling level) for each objective. In other words, an animal is not selected, i.e. is culled, if its EBV is worse than the culling level for any one objective, irrespective of how much its EBV may be better than the culling level for each of the other objectives. The actual culling levels are chosen so as to give, overall, a certain proportion of candidates selected. The method of calculating the correct culling levels is rather complicated, but in many cases the necessary calculations can now be performed on a computer. Once calculated, the culling levels can be used in any selection programme for which they are relevant, without further use of a computer.

(2) Index selection

This method involves ranking and selecting candidates according to their EBV for overall merit which is the weighted sum of the EBV for each objective, with the weights being the net economic values. This is yet another form of selection index. If, as in our pig example, there were two breeding objectives, the index would be:

$$I = \text{EBV for overall merit} = v_1\text{EBV}_1 + v_2\text{EBV}_2$$

where v_1 is the net economic value and EBV_1 is the estimated breeding value for the first breeding objective; and similarly for the second objective.

In most circumstances, index selection is the best option.

The importance of inbreeding and genetic drift

In Chapter 12 we saw that the mating of relatives (inbreeding) results in a decrease in performance (inbreeding depression), especially for traits associated with viability and/or reproductive ability; as the level of inbreeding increases, so too does the incidence of embryonic mortality (spontaneous abortion), stillbirths, and single-gene recessive disorders. In addition, because inbreeding increases homozygosity, it results in a decrease in genetic variance, which, in turn, reduces the potential for improving a population by selection.

In Chapter 5, we saw that the smaller the population, the more likely it is that gene frequencies change due to chance (genetic drift) rather than due to selection.

In addition to their relevance described in the earlier chapters, inbreeding and genetic drift are very relevant to improvement programmes, where there is a great temptation to concentrate on just a few animals with very high EBVs. The fewer

Copyrighted material

the parents in a selection programme, the greater is the rate of inbreeding, and the greater is the chance that genetic drift results in the loss of favourable alleles (which reduces the improvement that can be achieved). In addition, the greater is the chance that genetic drift results in the whole improvement programme making no progress, or even going backwards!

The importance of inbreeding and genetic drift in a closed population can be quantified in terms of a parameter called **effective population size**, N_e, which (as shown in Appendix 16.1 of *Veterinary Genetics*) can be calculated as $N_e = (4sdL)/(s + d)$, where s is the number of males first used as sires in any year, d is the number of females first used as dams in any year, and L is the average age of all male and female parents when their offspring are born (i.e. the average of male and female **generation intervals**), in years. As also shown in *Veterinary Genetics*, the increase in inbreeding per year, ΔF, is approximately $\Delta F = 1/(2N_eL)$.

How, then, do we try to ensure that ΔF is at an acceptably low level? In practice, we generally follow two rather arbitrary rules, which are:

(1) avoid mating very close relatives such as full-sibs, or parents and offspring, and
(2) ensure that effective population size is sufficient for ΔF to be less than an arbitrarily chosen value, such as 1 per cent per year.

From the above equation, this amounts to ensuring that N_e is larger than $50/L$. In very general terms, this is also a satisfactory criterion for avoiding the worst excesses of genetic drift and inbreeding depression.

In populations in which pedigrees are recorded, it is possible to keep track of the actual change in level of inbreeding over time, by calculating the inbreeding coefficient for each animal from all available pedigree information (as described in Chapter 12), then calculating the average F of all animals born in each year, and then plotting average F against year of birth. An estimate of ΔF can then be calculated as the regression of average F on year of birth.

In recent years, algorithms have been developed for optimizing the result of a selection programme by accounting simultaneously for the desire to maximize response and to limit future inbreeding. One such algorithm is Total Genetic Resource Management (TGRM™), developed by Professor Brian Kinghorn, which explores many possible mating options amongst the selected animals available at any given time, in search of the best compromise between maximizing the average breeding value of the offspring, and limiting the extent of relationship (quantified as coancestry) among those offspring. This and other approaches to achieving the best compromise are described in the list of references at the end of the chapter.

Sire-reference schemes

In the dairy industry, the widespread use of artificial insemination (AI) means that most dairy herds are linked together by the use of common bulls. Consequently,

Copyrighted material

Hidden page

Hidden page

using GWS for the selection of their next crop of young bulls (several years before they could be selected on daughter performance); and there is likely to be rapid application of the technology in many different situations in the livestock-breeding sector. Just as relevant is the potential of GWS in selection against multifactorial disorders such as hip dysplasia in dogs (Chapter 6); and for behavioural traits that are expensive to measure. Exactly the same principles apply for such disorders and behavioural traits as apply for production traits in livestock: in all cases, the best applications of the technology will see it being used for increasing the accuracy of EBVs, especially for animals on which measurements of their own phenotype are not available.

Summary

- Candidates for selection on a particular trait are ranked according to estimated breeding value, EBV, for that trait, which is determined from available clues to their true breeding value for that trait
- The potential clues to a candidate's true breeding value for a particular trait are:
  - A single measurement of the trait or the average of repeated measurements of the trait on the candidate or on any relative of the candidate, and
  - The average performance of a group of the candidate's relatives, e.g. full-sibs, half-sibs, or progeny, where the record available on each relative is either a single measurement of the trait or the average of repeated measurements of the trait
- If more than one of these clues are to be used to estimate the breeding value of each candidate for a particular trait, the clues are combined together into a selection index, which is a weighted sum of clues, with each clue weighted according to the heritability and (if there are repeat measurements) repeatability of the trait, the relationship between the source of the clue and the candidate, and the number of measurements that have contributed to that clue
- The most efficient way to calculate an EBV from a range of clues is to use a statistical technology called Best Linear Unbiased Prediction (BLUP)
- This technology is applied in many practical improvement programs around the world: for each candidate, all available clues from all measured traits on all relatives of a candidate are used to calculate an EBV for each trait
- When more than one trait is involved (a so-called multi-trait BLUP), estimates of genetic and phenotypic correlations between all pairwise combinations of traits are required, in addition to estimates of heritability of each trait
- A correlated response to selection is a change in an unselected trait resulting from selection of another trait; the magnitude of correlated response depends on the genetic correlation (r_G) between the two traits

Continued

Copyrighted material

- The existence of correlated traits extends the potential clues to a candidate's true breeding value for any trait: the EBV of a candidate for any trait can be calculated from an index incorporating the performance of the candidate and/or relatives for the trait in question and/or for any correlated trait
- The final step is to combine the EBVs for each objective into a single EBV for overall genetic merit, which is a weighted sum of the EBV for each breeding objective, where the weight for each breeding objective is the additional profitability of a unit increase in that breeding objective, assuming that all other breeding objectives do not change
- The importance of inbreeding and genetic drift in a closed population can be quantified in terms of a parameter called effective population size, N_e
- To avoid the worst excesses of inbreeding and genetic drift in a selection programme, N_e should be larger than $50/L$, where L is the average age of all male and the female parents when their offspring are born (i.e. the average of male and female generation intervals), in years
- Now that tens of thousands or even hundreds of thousands of DNA markers are available, it is possible to use genome-wide selection (genomic selection), which involves selecting on EBVs calculated from genotypes at the several hundred DNA markers that account for the vast majority of genetic variation in the trait
- In effect, these genotypes are simply another class of clues to each candidate's true breeding value

Further reading

Berglund, B. (2008) Genetic improvement of dairy cow reproductive performance. *Reproduction in Domestic Animals*, **43 Suppl 2**, 89–95.

Daetwyler, H.D., Villanueva, B. and Woolliams, J.A. (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*, **3**, e3395.

Enns, R.M. and Nicoll, G.B. (2008) Genetic change results from selection on an economic breeding objective in beef cattle. *Journal of Animal Science*, **86**, 3348–57.

Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to quantitative genetics*, 4th edn. Longman Cheshire, London.

Goddard, M.E. and Hayes, B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, **10**, 381–391.

Haile-Mariam, M., Bowman, P.J. and Goddard, M.E. (2007) A practical approach for minimising inbreeding and maximising genetic gain in dairy cattle. *Genetics Selection Evolution*, **39**, 369–89.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. and Goddard, M.E. (2009) Invited review: genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*, **92**, 433–43.

Hayes, B.J., Visscher, P.M. and Goddard, M.E. (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genetical Research*, **91**, 47–60.

Copyrighted material

Hinrichs, D., Wetten, M. and Meuwissen, T.H. (2006) An algorithm to compute optimal genetic contributions in selection programs with large numbers of candidates. *Journal of Animal Science*, **84**, 3212–18.

Janutta, V., Hamann, H. and Distl, O. (2008) Genetic and phenotypic trends in canine hip dysplasia in the German population of German shepherd dogs. *Berliner Munchen Tierarztliche Wochenschr*, **121**, 102–9.

Kinghorn, B.P., Meszaros, S.A. and Vagg, R.D. (2002) Dynamic tactical decision systems for animal breeding. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19–23 August 2002, Montpellier, France*, CD-ROM, paper 23–07 (See also http://www.xprime.com.au/)

Lassen, J., Sorensen, M.K., Madsen, P. and Ducrocq, V. (2007) An approximate multitrait model for genetic evaluation in dairy cattle with a robust estimation of genetic trends. *Genetics Selection Evolution*, **39**, 353–67.

Peixoto, M.G., Verneque, R.S., Teodoro, R.L., Penna, V.M. and Martinez, M.L. (2006) Genetic trend for milk yield in Guzerat herds participating in progeny testing and MOET nucleus schemes. *Genetics and Molecular Research*, **5**, 454–65.

Pong-Wong, R. and Woolliams, J.A. (2007) Optimisation of contribution of candidate parents to maximise genetic gain and restricting inbreeding using semidefinite programming. *Genetics Selection Evolution*, **39**, 3–25.

Taubert, H., Agena, D. and Simianer, H. (2007) Genetic analysis of racing performance in Irish greyhounds. *Journal of Animal Breeding and Genetics*, **124**, 117–23.

Copyrighted material

Hidden page

Hidden page

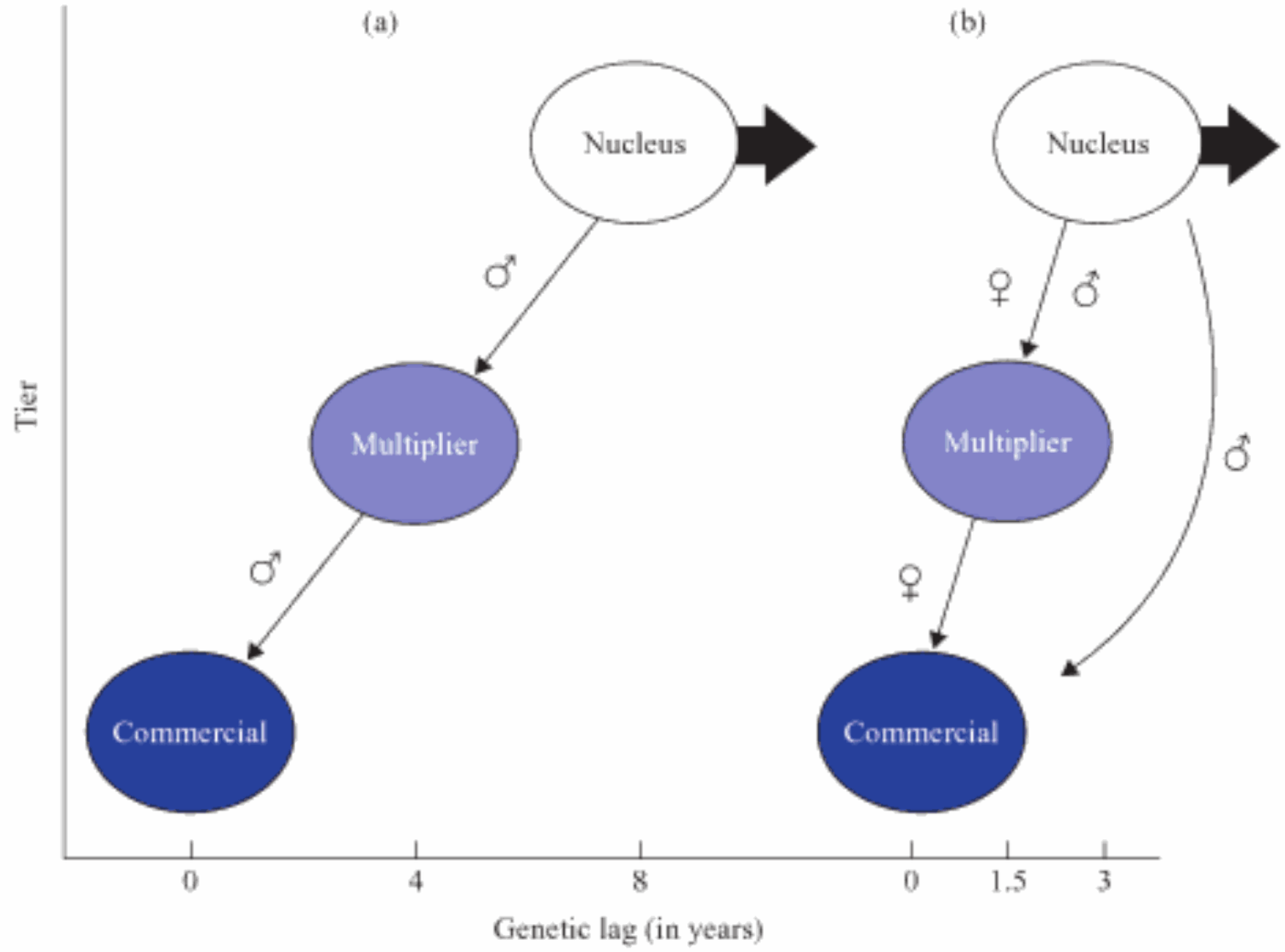


Fig. 16.2 The reduction in improvement lag brought about by altering the source of sires and dams used in the lower tiers of a pig pyramid. Thick arrows indicate the direction of genetic improvement in the nucleus tier, and thin arrows show the direction of gene flow between tiers. In this example, the lag is that expected when unselected animals are passed down. If selected animals are passed down, the lag is smaller. (After Bichard (1978) New Developments in Scientific Pig Breeding [supplement to *Pig Farming*, **27**(11)], 2–6.)

a higher performance than some animals in a higher tier. To the extent that higher performance is evidence of a superior breeding value, it would make good sense to transfer superior animals from a lower tier into a higher tier, and even (if they are sufficiently superior) into the nucleus tier. This amounts to opening the nucleus. Since traditionally the nucleus tier has consisted of studs of 'pedigree' animals registered with a breed society, the above proposal raises the possibility of introducing commercial (non-'pedigree') animals into 'pedigree' herds or flocks, which of course is not exactly in keeping with traditional stud-breeding philosophy. And yet there are definite advantages to be gained by opening the nucleus: the annual response to selection is increased, and the rate of inbreeding in the nucleus is substantially reduced.

For example, in a well-designed sheep or beef-cattle open-nucleus scheme, response to selection could be increased by 10–15 per cent, and the rate of inbreeding could be halved, in comparison with a closed-nucleus scheme of the same overall size.

One popular form of an open-nucleus breeding scheme involves a group of breeders agreeing to co-operate in the formation and subsequent running of an open

Copyrighted material

nucleus, in return for a regular supply of breeding stock (mostly males) from the nucleus, for use in their own herd or flock. There are several such schemes in practice now in sheep and to a lesser extent in cattle. They are often called **group breeding schemes** or **co-operative breeding schemes**.

One of the biggest practical difficulties with group breeding schemes is that very close and continuing co-operation is required between the participating breeders. The successful establishment of a group breeding scheme is therefore as much a matter of sociology as of genetics. However, a sufficient number of schemes have been running successfully for a sufficient time now to indicate that they can, given the right circumstances, be very beneficial to the co-operating breeders and to the industry as a whole.

If the advantages of open-nucleus schemes are as great as described, why do pig and poultry breeders still persist with closed-nucleus schemes? The main reason is that in pig and poultry breeding, nucleus herds and flocks are usually maintained under strict quarantine, with the aim of excluding as many diseases as possible. Opening these nucleus populations to regular importations from other herds or flocks involves an unacceptably high risk of introducing disease. This is an important point. It illustrates clearly that practical breeding programmes cannot be designed solely on the basis of genetic requirements. Quite often there are non-genetic constraints, such as requirements in relation to disease, which have a major influence on the final design of a breeding programme.

Information nucleus

In all the cases discussed above, nucleus herds or flocks are the primary source of genetic improvement. In recent years, a new type of nucleus has been proposed – a nucleus that is the primary source of phenotypic and genotypic information for the calculation of EBVs for traits that are difficult and/or expensive to measure ('novel' traits); information that is not likely to be collected in a commercial situation. A typical information nucleus comprises several flocks or herds representative of a range of environmental and/or production systems within an animal industry. The nucleus is co-funded and operated in a coordinated manner by the industry. Each year, semen from a large number of young sires from the leading seed-stock providers of the industry is distributed across females from all the flocks/herds, aiming for an equal number of offspring from each male in each flock/herd, with a total of at least, say, 50 measurable offspring per sire, which are phenotyped for the 'novel' traits. In the manner of a typical progeny test, this phenotypic information is then used in the calculation of EBVs of the young sires for the 'novel' traits. Importantly, since many offspring have contributed to each EBV, these EBVs have relatively high accuracy. These same sires (and possibly their offspring) are then subjected to a genome-wide scan with a SNP chip (the GWA phase; Chapter 15), providing a powerful data set for identifying a panel of a few hundred SNPs that can be used in genome-wide selection (the prediction phase; Chapter 15) throughout the entire

Copyrighted material

Hidden page

17 Crossing

Crossing is the mating of individuals from different populations (strains, breeds, or species). It is the second major method of exploiting genetic variation, selection being the first.

Animals that result from crossing are called **cross-breds**, as distinct from animals that result from matings within a population, which are called **straight-breds** or **pure-breds**.

Many people think of crossing in terms of crosses between inbred lines, partly because of the considerable success achieved by this breeding method in plants, especially in maize. But as described in Chapter 12, the development of inbred lines in animals is very wasteful and costly, and hence is hardly ever used now, except possibly by one or two poultry-breeding organizations. Instead, crossing in animals is generally carried out between populations that have not been deliberately inbred but which have been isolated from each other for varying lengths of time.

The aim of this chapter is to describe the main methods of crossing that are used in practice, and to comment on the implications of each method.

Regular crossing

Regular or systematic crossing occurs when the same cross is made on a regular basis, with the aim of producing a particular type of progeny. Regular crossing may be advantageous from two points of view.

First, the cross-bred progeny usually show **heterosis** or **hybrid vigour** for certain traits. Heterosis occurs when the average performance of cross-bred progeny is superior to the average performance of the two parents. Various types of heterosis are recognized, including **parental heterosis** (either **maternal** or **paternal**), referring to the performance of animals as parents, and **individual heterosis**, referring to non-parental performance. The extent of heterosis for a particular trait can vary considerably, depending on the environment and on the populations being crossed. However, in general, (1) heterosis is greatest in traits most closely associated with reproduction and viability, and (2) the greater the genetic diversity between two populations of domesticated animals, the greater the heterosis in crosses between them.

Copyrighted material

The reasons for these conclusions can be briefly summarized as follows. Heterosis occurs only if there is non-additive gene action (dominance and/or epistasis) at loci affecting the trait; and the greater the difference in gene frequencies between the two populations being crossed, the greater the heterosis. Recalling from Chapter 12 that inbreeding depression occurs only if there is non-additive gene action, it is not surprising to find in general that the same traits that show heterosis also show inbreeding depression.

The second advantage of regular crossing is that **complementarity** may occur. This refers to the additional profitability obtained from crossing two populations, resulting not from heterosis but from the manner in which two or more traits complement each other.

For example, consider two populations of pigs, one (A) with a very high food conversion ratio but low litter size, and the other (B) with poor food conversion ratio but high litter size. Suppose that boars from A are regularly mated with sows from B to produce cross-bred pigs that are raised solely for meat production. Even if there is no heterosis for food conversion ratio, i.e. even if the average performance of the cross-bred pigs is exactly intermediate between that of the two parental populations, and if the litter size of B sows is the same when mated to either A or B boars, the overall profitability of this particular operation is likely to be much greater than the profitability of running either A or B alone, because the same total number of pigs is produced as from the most prolific population, B, but the food conversion ratio of each offspring is higher than that of offspring from B × B matings.

Certain crosses show more complementarity than others, depending on the extent to which the populations differ in reproductive performance and in production traits, and also depending on the 'direction' of the cross. The latter point can be illustrated by noting that there is far greater complementarity when the most prolific population is used as a source of dams rather than of sires.

There are two basic methods of regular crossing: specific crossing and rotational crossing.

Specific crossing

The simplest form of specific crossing is the **two-way cross**, in which animals from one population, A, are regularly mated to animals of a second population, B:

$$\begin{array}{c} A \times B \\ \downarrow \\ (AB) \end{array}$$

In this diagram A and B are straight-bred parents, and (AB) represents the cross-bred progeny, which are known as **first-cross** or **F1** (first filial) progeny.

Sometimes the male parent in the cross always comes from one population and the female parent always comes from the other population, so as to obtain maximum benefit from complementarity. When this happens, the two parental populations are called the **sire line** and the **dam line**, respectively. Obviously, to maintain the lines

Copyrighted material

as separate entities, both males and females must be mated within a sire line and within a dam line; but only one of the sexes from each line (males in the sire line, and females in the dam line) makes a contribution to the cross.

The two-way specific cross produces offspring that are 100 per cent heterozygous, in the sense that they have one gene from each parental population at every locus. As such, they show 100 per cent of individual heterosis. However, two-way crossing does not provide any opportunity for the breeder to benefit from maternal or paternal heterosis, because the parents in a two-way cross are never themselves cross-bred; they are always straight-bred. In order to exploit maternal and/or paternal heterosis, other types of crossing must be used.

One of these is the **backcross**, which involves mating cross-bred animals from a two-way cross back to one of the parental breeds:

$$\begin{array}{ccc} (AB)\times A & \text{or} & (AB)\times B \\ \downarrow & & \downarrow \\ (AB)A & & (AB)B \end{array}$$

With backcrossing, the cross-bred parent is 100 per cent heterozygous, and hence shows 100 per cent parental heterosis. However, the backcross progeny are the result of the fusion of a gamete containing all, say, A genes (from the parental breed) and a gamete containing an average of one-half A genes and one-half B genes (from the cross-bred parent). This means that at half of their loci, on average, backcross progeny have two A genes, i.e. are homozygous for genes from the A population, and at the other half of their loci, they have one A gene and one B gene, i.e. they are heterozygous. In other words, backcross progeny on average are 50 per cent less heterozygous than first-cross, AB, progeny. Consequently, they show on average only 50 per cent of individual heterosis.

A better way to exploit heterosis is to use a **three-way cross**, in which a first-cross animal, AB, is mated to an animal from a third population, C. Once again, because it is generally more profitable to improve female reproductive ability, a three-way cross is usually of the form

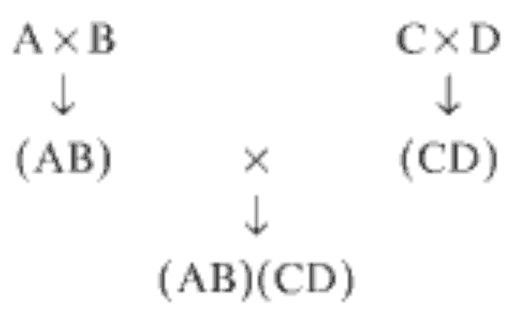
$$\begin{array}{c} (AB)\,♀\times C\,♂ \\ \downarrow \\ (AB)C \end{array}$$

This cross enables full utilization of maternal heterosis because the dam is 100 per cent heterozygous. It also enables full utilization of individual heterosis because the progeny are also 100 per cent heterozygous, being AC at one half of their loci, and BC at the other half, on average. It should be noted that even though first-cross and three-way-cross progeny both show full utilization of individual heterosis, in the sense that they are both 100 per cent heterozygous, they may not show the same amount of heterosis. The reason for this is that different crosses are involved in each case and, as noted earlier, the amount of heterosis is greater with some crosses than with others.

Copyrighted material

Because it is so successful at exploiting heterosis in the female parent and in the growing, commercial animal, and because it also enables exploitation of complementarity, the three-way cross is used widely throughout the world.

The final form of specific crossing is the **four-way cross**, in which the cross-bred progeny from two separate two-way crosses are mated to produce commercial progeny:



Since both the male and female parents of the commercial progeny are cross-bred, this cross enables full exploitation of both paternal and maternal heterosis, as well as individual heterosis. Once again, it should be noted that the actual amount of individual heterosis in the commercial progeny may not be the same as that observed in, for example, either of the two types of first-cross progeny.

Rotational crossing

In all forms of specific crossing, both parents of the commercial progeny are either straight-bred or are the cross-bred progeny of straight-bred parents. It follows that a commercial breeder who wishes to produce only the final commercial progeny has to obtain all breeding replacements, both male and female, from another source. Being obliged to introduce all breeding replacements from other populations is not always desirable, from a general management point of view and, especially in the case of pigs and poultry, from a disease point of view. It also deprives interested breeders of any possibility of breeding at least some of their own replacements.

Rotational crossing is an alternative form of regular crossing that overcomes some of these difficulties. It generally involves the use of males from two or three different populations, in regular sequence. If rotational crossing is continued in a commercial herd or flock for several years, and if all replacement females are bred in the herd or flock, all females will soon be cross-bred, containing varying proportions of genes from the two or three populations from which males were obtained. Symbolically, we have the situation as illustrated in Table 17.1.

A major disadvantage of rotational crossing is that it does not allow any exploitation of complementarity, because the populations involved in the crossing cannot be restricted in use to a single purpose, as they are in specific crossing. Because of this, rotational crossing is a much more attractive proposition where the populations available for crossing show little or no complementarity, but do show some economically useful heterosis.

Copyrighted material

Hidden page

Table 17.2 Fraction of heterosis, as indicated by average heterozygosity, expected in the most common types of regular crossing

Type of crossing	Fraction of heterosis		
	Individual	Maternal	Paternal
Straight-bred	0	0	0
Two-breed cross			
A♀ × B♂	1	0	0
Backcross			
AB♀ × (A♂ or B♂)	1/2	1	0
(A♀ or B♀) × AB♂	1/2	0	1
Three-breed cross			
AB♀ × C♂	1	1	0
C♀ × AB♂	1	0	1
Four-breed cross			
AB♀ × CD♂	1	1	1
Rotational cross			
two sire breeds	2/3	2/3	0
three sire breeds	6/7	6/7	0

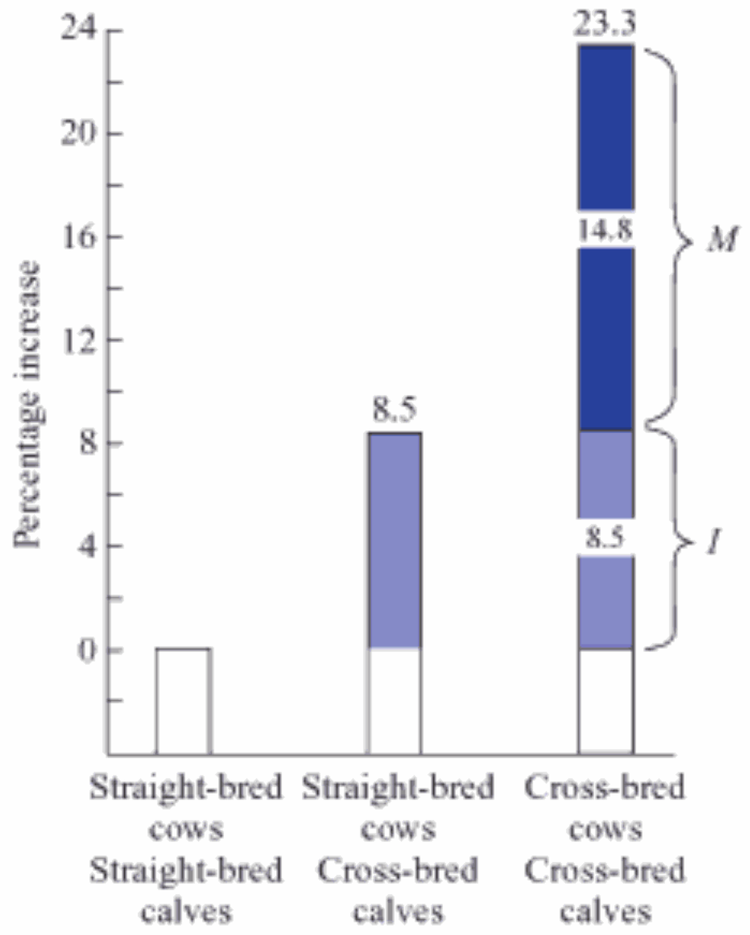


Fig. 17.1 Heterosis in weight of calf weaned per cow joined, as a result of mating straight-bred cows to bulls of a different breed (*centre*), or mating first-cross cows to bulls of a third breed (*right*). Results were obtained from all relevant crosses among Hereford, Angus, and Shorthorn cattle. *I* = individual heterosis; *M* = maternal heterosis. (After Cundiff (1982), in Beef Research Program Progress Report No. 1 (ARM-NC-21), pp. 3–5, Roman L. Hruska US Meat Animal Research Center, Nebraska.)

Copyrighted material

Table 17.3 Performance observed or expected from various types of regular crossing in cattle, sheep, and pigs

Type of regular crossing	Heterosis utilized[1]	Weight of calf weaned per cow joined[2]	Weight of lamb weaned per ewe joined[2]	Pigs reared per sow per year[2]
Straight-bred	nil	100	100	100
Two-breed cross	*I*	108	107	106
Backcross	(1/2) *I* + *M*	119	–	–
Three-breed cross				
C♀ × AB♂	*I* + *P*	–	137	–
AB♀ × C♂	*I* + *M*	123	152	118
Two-breed rotation	(2/3) *I* + (2/3) *M*	116	–	112
Three-breed rotation	(6/7) *I* + (6/7) *M*	120	–	–
Other				
AB♀ × AB♂	(1/2) *I* + *M* + *P*	–	166	–

[1] *I* = individual heterosis; *M* = maternal heterosis; *P* = paternal heterosis.
[2] As a percentage of straight-bred performance.

only as a general guide. Second, the traits considered in Table 17.3 are closely associated with reproduction and/or viability and hence show substantial heterosis. Other traits less closely associated with reproduction and/or viability show less heterosis, and some traits (such as carcase traits) show essentially zero heterosis.

Crossing to produce a synthetic

An alternative to regular crossing is to perform one or a few crosses between two or more populations to produce a single population of animals containing a mixture of genes from each population. A single population that is a mixture of various populations is called a **synthetic** or **composite**. Once a synthetic has been formed, the main aim is to improve it as rapidly as possible by selecting within it, following the procedures in Chapter 15. In many situations, the result of selection within a synthetic is a new breed. For example, the Santa Gertrudis, the Jamaican Hope, the Norwegian Red and White, the Luing, the Australian Milking Zebu, the Australian Friesian Sahiwal, and the Belmont Red are all cattle breeds formed comparatively recently by selection within a synthetic, and similar examples can be found in other species of domestic animals. Some synthetics are based almost entirely on just one traditional breed, and were founded with selected animals from a number of different strains within the breed. Others, such as many of the present-day broiler chicken populations maintained by breeding companies, are mixtures of several different breeds, each of which was thought to have at least some desirable traits in relation to the aim of the breeders concerned.

A particularly popular combination in the formation of synthetics in cattle has been that of *Bos taurus* with *Bos indicus*, in order to produce new breeds capable of improved production under tropical and semi-tropical conditions. The aim has usually been to combine the improved production ability of *Bos taurus* breeds with the heat tolerance and tick resistance of *Bos indicus* breeds.

Copyrighted material

Of course, if we go back far enough in time, it becomes evident that many of the long-established breeds that we know today started as synthetics of one form or another. In some cases, the crossing phase lasted a long time. But sooner or later, in most cases, the resultant synthetic was closed and a new breed began to emerge. Increasingly, it is being recognized that closing a synthetic for ever is not the best option. On the contrary, in most cases it is advantageous to introduce some new genetic variation from time to time, and this continues to be done in a number of cases.

The main guidelines to be followed in crossing to produce a synthetic are:

(1) ensure that the animals used in the original crossings have the highest possible average breeding value for the relevant traits (it is no use starting a synthetic with inferior animals);
(2) maximize variance in breeding values amongst the foundation animals in the synthetic, by using as many unrelated animals as possible from each of the contributing populations, bearing in mind the previous requirement that they should have the highest possible average breeding value.

With respect to heterosis, the simplest expectation is that if there are n populations contributing equally to the synthetic, then $(1 - 1/n)$ of the heterosis present in the F1 (the first-cross progeny) is retained in the **F2** (the progeny resulting from matings among F1s), and in subsequent generations. However, if the synthetic is not maintained at an adequate size, or if selection within the synthetic is very intense, inbreeding depression may whittle away the remaining heterosis. The main advantage of synthetics is that only one population has to be maintained, rather than the two or more parental populations required for a regular crossing programme.

Grading-up

Grading-up involves a succession of backcrosses from one population into another population, with the aim of introducing a new gene or genes into one of the populations, or with the aim of substituting one population for the other.

Introducing a new gene or genes

Suppose, for example, that some breeders wish to create a polled strain of a horned breed, and have one mutant animal that is heterozygous for the dominant polled gene, *P*, at their disposal. The polled animal does not necessarily have to belong to the same breed in which the polled strain is to be produced; its only requirement is to be able to produce fertile offspring when mated to members of that breed. The breeding programme to be followed in introducing the polled gene to a horned breed is called **introgression**, and is illustrated in Table 17.4. The procedure is very straightforward. All that is required is for all horned offspring to be culled each year so that only polled offspring remain to be mated back to the horned breed. A very critical requirement is that a large and representative sample of animals from the horned breed should be used in the backcrossing programme, to ensure that the polled strain has as much genetic variation as the horned breed from which it arose.

Copyrighted material

Table 17.4 The use of grading-up to create a polled strain of a horned breed

Generation	Mating programme	Average proportion of genes from horned breed in polled offspring
0	Polled × Horned	
1	Selected polled offspring × Horned	$1/2 = (1/2)^1$
2	Selected polled offspring × Horned	$3/4 = 1 - (1/2)^2$
3	Selected polled offspring × Horned	$7/8 = 1 - (1/2)^3$
t		$1 - (1/2)^t$

In most situations, and especially if the above guideline has been followed, three or at the most four crosses, i.e. the original cross and two or three backcrosses, are sufficient to achieve the original aim. By this stage, the average proportion of genes from the horned breed in the new strain is either $\frac{7}{8}$ (two backcrosses) or $\frac{15}{16}$ (three backcrosses) and the polled offspring are indistinguishable from members of the horned breed, especially if there has been selection among polled offspring for traits associated with the horned breed, during grading-up.

The final phase of a grading-up programme for the introduction of a new gene depends on whether the gene is dominant or recessive. If it is recessive, the new strain can be formed immediately from all offspring exhibiting the trait; they will all breed true for the recessive gene. If, however, the gene is dominant, e.g. the polled gene, all offspring showing the trait are heterozygous and will not therefore breed true. The task is then to arrange matings among these animals, and in subsequent generations to distinguish carriers from non-carriers, with the aim of identifying and culling all carriers. The general requirements for this are outlined in Chapter 21.

DNA markers linked to the gene of interest can be of immense help during the final phase of introgression if the introgressed gene is dominant. If the gene is recessive, linked markers will be of substantial use during the backcrossing phase in enabling identification of carriers of the introgressed gene.

In practice, many introgression programmes involve the introduction of desirable quantitative traits rather than simply a single desirable gene. For example, there has been much interest in introducing the high fecundity of Chinese pigs into European pigs. Exactly the same principles apply for introgressing a quantitative trait as for a single gene. And if sufficient DNA markers are available, they will be of substantial help in enabling the identification of the particular segments of chromosomes that are to be introgressed, i.e. the segments of chromosomes that carry fecundity genes from the Chinese pigs.

Copyrighted material

In fact, the use of **marker-assisted introgression** (MAI) is likely to be one of the most immediate practical applications of DNA markers in animal breeding.

Breed substitution

The late 1960s and early 1970s witnessed an awakening of worldwide interest in long-established continental European breeds of cattle such as Charolais and Simmental. A few animals and many doses of semen from a number of such breeds were exported to many countries in which these breeds were previously unknown, and in each country a race began to establish 'purebred' herds of these 'migrant' breeds as quickly as possible, by means of a grading-up procedure like that illustrated in Table 17.5.

In the early stages of establishing a breed using this type of grading-up, much controversy often arises as to the number of backcrosses that should be required before progeny are eligible for official registration as 'purebreds'. Such controversies illustrate a lack of understanding of the genetic basis of grading-up.

The most important point to realize about grading-up is that, except for half-breds, there is variation among the progeny of all other generations, with respect to the proportion of 'local' and 'migrant' genes.

For example, it is possible that a so-called three-quarter-bred could have more than $\frac{31}{32}$ migrant genes. Equally likely, it could have just over one-half migrant genes. The reason for this is that segregation during the formation of gametes in half-breds and in all subsequent crosses gives rise to a spectrum of different types of gametes, ranging theoretically from a gamete containing only local genes, to a gamete consisting solely of migrant genes (neglecting the effect of crossing-over). Thus, as shown in Table 17.5, the proportion of migrant genes in three-quarter-breds and all subsequent crosses can range from $\frac{1}{2}$ to 1, if we neglect the effects of crossing-over. The

Table 17.5 Grading-up to a migrant breed, M, from any local animals, L

Generation	Mating programme	Designation (grade) of animals in square brackets	Proportion of migrant genes in animals in square brackets		
			Min.[1]	Av.	Max.[1]
0	L × M				
1	[LM] × M	1/2 bred	1/2	1/2	1/2
2	[(LM)M] × M	3/4 bred	1/2	3/4	1
3	[{(LM)M}M] × M	7/8 bred	1/2	7/8	1
4	[({(LM)M}M)M] × M	15/16 bred	1/2	15/16	1
5	etc.	31/32 bred	1/2	31/32	1

[1] Assuming no crossing-over.

Copyrighted material

most likely proportion of migrant genes in any cross is the average, which is the proportion that is used to describe that cross; and extreme deviations from these proportions are less likely than small deviations.

But the fact remains that by the time breeders have reached the second or third backcross, they may have offspring with proportions of migrant genes ranging from nearly $\frac{1}{2}$ to nearly 1. Although breeders cannot tell exactly what proportion of migrant genes exists in any particular animal simply by looking at it, they can be certain that, on average, animals whose appearance corresponds more closely to the migrant breed are likely to have a larger than average proportion of migrant genes, within any particular grade.

If breeders do any selection in favour of animals showing migrant-like appearance during their grading-up programme, by the time they have reached the second backcross, i.e. $\frac{7}{8}$, their selected progeny will on average have a proportion of migrant genes far higher than $\frac{7}{8}$, and probably higher than $\frac{31}{32}$.

It is now obvious why the controversies about the required number of backcrosses are futile, even if the only aim of a grading-up programme was to obtain the highest possible proportion of migrant genes within a particular herd or flock.

But there is a more important reason why the debates are pointless. It is that breeders would improve economically important traits more rapidly if they stopped grading-up after the first or at most the second backcross, and concentrated solely on continual selection. In this way, the migrant genes are allowed to find their own optimum proportion for the particular production system.

Continual backcrossing simply removes all the local genes, and achieves nothing more than replicas of migrant animals that were alive 15 or 20 years ago. Such a programme implies that the local animals have absolutely nothing to offer by way of adaptation to local environments (both natural and managerial) that are often very different from those that exist in the country of origin of the migrant breed, In fact, of course, local animals are usually well adapted to their local environment, and therefore they do have some genes that should make a useful contribution to the new breed being developed.

We can conclude that continual backcrossing in a grading-up programme is both unnecessary and undesirable.

Summary

- Regular crossing involves the same cross being made on a regular basis, with the aim of producing a particular type of progeny
- The main benefits of regular crossing arise from complementarity (the additional productivity resulting from the manner in which two or more traits complement each other) and heterosis (superiority in performance of crossbred offspring over the average of parents)
- Heterosis is greatest in traits most closely associated with ability to survive and reproduce

Continued

Copyrighted material

- The greater the genetic differences between two populations of domesticated animals, the greater the heterosis in crosses between them
- Regular crossing can be either specific (e.g. two-way cross, three-way cross, backcross) or rotational (in which males from two or more populations are mated on a rotational basis with females obtained from earlier crosses in the series)
- Crossing can be used to create a synthetic or composite, which is a single population comprising a mixture of genes from two or more other populations; the aim in producing a synthetic is to maximize genetic variation and then to select within that population as rapidly as possible
- Most breeds, including most old established breeds, originated as synthetics
- Grading-up involves a succession of backcrosses from one population into another population, with the aim of introducing a new gene or specific genes into one of the populations, or with the aim of substituting one population (migrant) for the other (local)
- In the second case, there is often controversy as to how many backcrosses to the migrant population should be required before the resultant animals can be registered as 'pure'
- Such controversies are pointless for two reasons: first, if any selection in favour of migrant appearance has occurred during the first one or two backcrosses, the proportion of migrant genes actually present in offspring is likely to be far higher than the average figures of $\frac{3}{4}$ and $\frac{7}{8}$ expected for all offspring; second, it is very likely that the local population has some useful genes, especially in terms of local adaptation
- Consequently, in most cases it is best to forget about grading-up and instead to combine together the local and migrant population into a synthetic, and then to select in this synthetic for economic performance in the local environment, allowing the optimum proportion of local and migrant genes to be achieved automatically

Further reading

Abasht, B. and Lamont, S.J. (2007) Genome-wide association analysis reveals cryptic alleles as an important factor in heterosis for fatness in chicken F2 population. *Animal Genetics*, **38**, 491–8.

Ask, B., van der Waaij, E.H., Stegeman, J.A. and van Arendonk, J.A. (2006) Genetic variation among broiler genotypes in susceptibility to colibacillosis. *Poultry Science*, **85**, 415–21.

Dechow, C.D., Rogers, G.W., Cooper, J.B., Phelps, M.I. and Mosholder, A.L. (2007) Milk, fat, protein, somatic cell score, and days open among Holstein, Brown Swiss, and their crosses. *Journal of Dairy Science*, **90**, 3542–9.

Hielscher, A., Brandt, H., Erhardt, G. and Gauly, M. (2006) Heterosis analysis of *Haemonchus contortus* resistance and production traits in Rhoen sheep, Merino Land sheep and crossbred lambs. *Veterinary Parasitology*, **141**, 279–84.

Copyrighted material

Hidden page

18 Selection and regular crossing

Traditionally, many stud breeders have regarded crossing as an insult to them and to their animals, and as a threat to their livelihoods. The aim of this chapter is to explain why these opinions are incorrect, and to illustrate that the most profitable future for both stud breeders and commercial breeders lies in breeding programmes that involve both selection and regular crossing.

Selection

A selection programme within a population can lead to a steady improvement in average breeding value and hence in profitability. However, the average level of inbreeding also shows a steady increase, even if mating is entirely at random among the selected candidates. In addition, the greater the selection applied, the greater is the increase in average level of inbreeding. Thus, successful selection programmes in closed populations incur inbreeding depression which, as we saw in Chapter 12, has the greatest effect on traits associated with viability and reproductive ability. If effective population size (N_e) is not too small, the increase in inbreeding depression may be unimportant from a practical point of view. However, if N_e is relatively small, the steady increase in inbreeding depression may begin to adversely affect profitability.

Of course, this problem can be offset in practice by the occasional introduction of unrelated animals into the otherwise closed population. However, as soon as breeders start buying animals from other sources, the response in their own population is determined in part by the breeding values of the purchased animals; and if response to selection within their own population has been quite rapid, they may have difficulty finding other sources from which to purchase stock that match the average breeding value of their own stock. This is becoming an increasingly common dilemma.

Selection and regular crossing

We saw in Chapter 17 that the traits that show the greatest inbreeding depression are those that also show the greatest heterosis when different populations are crossed.

Copyrighted material

This observation immediately suggests a solution to the above dilemma: we can get the best of both worlds by continually selecting within each of several populations and by making regular crosses between them in order to produce the final commercial product.

In this way we are able to get the most out of all genetic variation; continual selection exploits variation in breeding values, V_A, and regular crossing exploits non-additive genetic variation, V_D and V_I.

The major difficulty with breeding programmes involving both selection and regular crossing is the organizational and management problems associated with maintaining at least two different straight-bred populations together with various cross-bred populations. In some situations, this difficulty is overcome by co-operation among several breeders, with each breeder maintaining just one population. The financial benefits from combining continual selection with regular crossing are sufficient to make this extra effort well worthwhile in most circumstances.

An important practical point is that regular crossing can occur only if someone is maintaining and improving the straight-bred lines. Thus, rather than presenting a threat to stud breeders, regular crossing actually provides them with a substantial incentive to maintain and improve their straight-bred populations.

In discussing regular crossing in this section, we have concentrated solely on the benefits to be gained from heterosis. However, it was emphasized in Chapter 17 that there is more to crossing than just heterosis. In fact, it was pointed out that even in the absence of heterosis, certain crosses show complementarity. Thus, in certain circumstances, regular crossing may still be an economically viable proposition, even if there is negligible inbreeding depression in the relevant closed populations.

Furthermore, the benefits from complementarity can be enhanced by selection for different traits in the lines that are used in the crossing programme. If, for example, one line is always used as the sire of the final commercial generation in a three-way cross, it can be selected specifically for traits that are important in the commercial generation, e.g. growth rate and carcase quality. In contrast, a line that is used solely as the female parent of the F1 hybrid in a three-way cross, can be selected less for growth and carcase traits, and more for reproductive traits, probably utilizing information from relatives to help overcome the relatively low heritability of such traits. This leads to the development of specialized **sire lines** and **dam lines**.

Recalling some points made in Chapters 15 and 17, we can conclude that by carrying out selection within populations and regular crossing between populations, we are able to exploit additive gene effects within populations (through selection), additive gene effects between populations (through complementarity), and non-additive gene effects (through heterosis).

A final point should be made in relation to synthetics or composites. Sometimes, there are good reasons for crossing various populations on a once-only basis to produce a synthetic. Sooner or later, however, a synthetic has to be considered in the same category as any other population. Thus, even if inbreeding depression is not currently a problem in a particular synthetic, it may still be economically advantageous to use that synthetic in a regular crossing programme, if the synthetic exhibits useful complementarity with one or more other populations.

Copyrighted material

Hidden page

19 Biotechnology and the future

Much of this book has been concerned with the application of various forms of biotechnology to the breeding of animals. The aim of this chapter is to fill in the remaining gaps – to comment upon those aspects of biotechnology that have not been sufficiently covered in earlier chapters. In so doing, we shall be looking into the future, which is always a risky business. The predictions that follow, therefore, are made in the full knowledge that not all of them are likely to come to fruition.

Artificial insemination (AI)

AI was the first form of reproductive technology to have large-scale practical implications for genetic improvement. As it is basically a means of increasing the reproductive potential of males, AI enables fewer males to be selected as parents of the next generation. Its genetic effect, therefore, is to increase the intensity of selection of males. As soon as it became a practical reality in the dairy industry in the 1950s, breeding programmes were devised to enable its full genetic potential to be realized. Basically, these programmes involve progeny testing of dairy bulls, by distributing the semen from a team of young bulls across a number of participating herds, in each of which the daughters of the various young bulls are directly compared. Although this involves a long generation interval (because the bulls are not selected until their daughters have completed their first lactation), it results in a relatively high accuracy of selection, and the potential genetic gain is substantially greater than that which is possible under natural mating. In dairy cattle, for example, AI can give rise to genetic improvement of around 1 per cent of the mean production per year. (While this may seem to be a very small figure, remember that the gain is cumulative, i.e. an additional gain of the same magnitude is achieved each year. Also, once each year's gain has been achieved, it remains for ever after, without the need for additional input to maintain it.)

Critics of AI have pointed to its potential for increasing the rate of inbreeding, with all the attendant disadvantages, especially in relation to autosomal recessive defects, as discussed in Chapter 12. AI certainly does have this potential, and, in closed populations, care must be taken to ensure that the effective population size is sufficiently large for the rate of inbreeding to be kept at an acceptably low level.

Copyrighted material

In practice, most AI improvement programmes have a sufficiently large effective population size, and, in addition, intermittent immigration ensures that the level of inbreeding remains at satisfactorily low levels. What the critics of AI have generally failed to realize is that, at the level of commercial production, AI has the potential to *decrease* the rate of inbreeding, because it provides a far wider choice of unrelated males than is ever available under natural mating. This is very important, because it means that AI can be a very powerful tool for preventing the occurrence of inherited defects. Thus, rather than trying to condemn AI as a dangerous technology, as some well-meaning veterinary authorities try to do from time to time, all interested parties should be continually trying to encourage its correct use.

Given its many advantages, it seems likely that AI will continue to be used extensively in dairy improvement, and in any other form of animal production where the female reproductive cycle can be readily monitored or controlled.

Multiple ovulation and embryo transfer (MOET)

Just as AI increases male reproductive potential, so does MOET increase female reproductive potential. Thus, the genetic consequence of MOET is an increase in female selection intensity. Alternatively, MOET can be used to decrease the female generation interval, by enabling replacements to be obtained from only the younger females in a population.

For industries in which the traditional form of improvement programme involves performance testing (e.g. beef and sheep), MOET offers considerable advantages, resulting in approximately 50 per cent additional response per year.

When superimposed on an AI improvement programme involving progeny-testing of males, MOET has relatively little effect, because the intensity of selection of females to breed the next team of males is already quite high. However, if MOET is used in circumstances where selection of males on the basis of progeny-testing is replaced by selection of males at a far younger age on the basis of their female sibs' performance, then, although the accuracy of selection is reduced, the generation interval is reduced to an even greater extent, with the result that the rate of genetic improvement per year is actually as great as, if not greater than, a progeny-testing programme.

Progeny-testing programmes are large-scale affairs, involving the recording of performance of thousands or tens of thousands of animals, and requiring the co-operation of many commercial farmers. In contrast, a MOET improvement programme can be conducted quite successfully in a nucleus population of just a few hundred animals, which can be run as a single herd or flock under the control of a just a few people. Since relatively few animals have to be recorded, it is feasible to measure traits additional to those traditionally measured. For example, in a dairy MOET nucleus, it is possible to measure individual food intake and various blood metabolites, and to genotype animals for relevant DNA markers. In contrast, in a state-wide or national progeny-testing programme, such additional activities are out

Copyrighted material

of the question both economically and practically. In addition, because a MOET nucleus inevitably has substantial biotechnological expertise on hand, such herds or flocks are ideally placed to adopt new biotechnologies such as those discussed below. Finally, in developing countries, where large-scale recording schemes and widespread use of AI are almost out of the question, a MOET nucleus may be the only realistic way to conduct an improvement programme, because performance recording and the facilities and skills necessary for conducting MOET are required only in the nucleus. In all circumstances, the benefits of improvement in a MOET nucleus can be spread to the commercial level by the provision of natural-service bulls from the nucleus, or, if this source is insufficient to meet demand, from one or more multipliers. Alternatively, if the appropriate technology is available, AI of nucleus-bred males can be used.

Just as with AI, MOET has the potential to increase the rate of inbreeding. If it is used in a closed population, the effective population size must be sufficiently large to avoid an unacceptably high rate of inbreeding. In practice, again as with AI, occasional immigration is a very useful means of avoiding inbreeding problems. The challenge, of course, is to find unrelated animals of at least equal, and preferably superior, genetic merit. In practice, most MOET nucleus populations remain open for many years; each year, a proportion of the matings involve semen from outside bulls. If the resultant offspring have sufficiently high EBVs, they are retained, which means that the genes from the imported semen are incorporated into the nucleus. If the offspring are not good enough, they are culled.

In evaluating the potential of MOET, it helps to realize that this technology in effect changes a monotocous species into a polytocous species; in other words, MOET turns a cow into a sow. It follows that MOET is unlikely to have much effect on genetic improvement in naturally polytocous species. Consequently, cattle (both dairy and beef), sheep, and goats are the species that have most to gain from the use of MOET nucleus populations.

In vitro maturation (IVM) and *in vitro* fertilization (IVF) of ova

These two technologies have considerable potential. From the point of view of commercial animal production, they offer the potential for artificial insemination of semen to be replaced by artificial inembryonation of mass-produced specific-purpose embryos. Depending on the industry's needs, these embryos could be straight-breds, or F1 hybrids, or even three-way or four-way hybrids.

In relation to genetic improvement, IVM and IVF have the potential to reduce the variability in yield of embryos, and thereby increase the effectiveness of conventional MOET programmes. In addition, by enabling the collection of eggs from very young females, they hold the promise of reducing the generation interval and hence increasing the rate of improvement in a MOET nucleus.

IVM and IVF also raise new possibilities. For example, all selected males could be mated to all selected females, thereby creating a family of paternal half-sibs, a

Copyrighted material

Hidden page

One of the major uses of this technology has been the production of a new generation of vaccines directed against a wide range of pathogenic and parasitic diseases, following the cloning and appropriate expression of the gene(s) for the most antigenically protective protein portion of the pathogen or parasite. Some of these vaccines are already on the market, and are very successful; others have not been successful.

An interesting example of a vaccine based on a recombinant protein is the vaccine against the cattle tick, *Boophilus microplus*. After the discovery that extracts from the gut of the tick provide protective immunity, the most protective antigen, called BM86, was identified by successive rounds of protein fractionation and vaccination trials. Peptide sequences derived from the purified protein were used to design oligonucleotide probes which were used to screen a tick cDNA library. This process identified the full-length mRNA sequence encoding the protective protein. Recombinant forms of the protein were then produced in either bacteria or yeast, purified and incorporated into a vaccine. Ticks feeding on immunized cows were heavily damaged and often died or reproduced very poorly. This led to decreased numbers of ticks on pasture and decreased tick burdens on the cattle.

Although vaccines are a very important example of the use of recombinant proteins, there are many other uses as well. In fact, the possibilities for recombinant proteins are endless. Unfortunately, so are the technical and financial challenges involved in translating an idea into a commercially viable product. There are also public concerns about the consumption of food from animals that have received recombinant proteins such as somatotropin (growth hormone). In addition, there is concern about the effect of such proteins on animals (will they be pushed too far?), and on the livelihood of farmers (will the dramatic increases in production per animal lead to a reduction in the number of viable farms?). In the case of bovine somatotropin, there is also concern about the potential for its undisclosed use in dairy herds to produce biased EBVs. The recent licensing of bovine and porcine somatotropin in several countries will provide test cases for the acceptance of this form of biotechnology.

Transgenesis

The basic technology of transgenesis was described in Chapter 2. One of the initial major aims of transgenesis in farm animals was to obtain extra improvement in economically important traits, e.g. by inserting additional copies of the growth-hormone gene. However, the effects of the transgene need to be quite large before this use of transgenesis is more effective than conventional selection. The reason for this is that once the transgenic animal has been created, it takes several generations to introgress the transgene into a population, and to evaluate its effect. During this time, substantial additional gains will be made by conventional selection. Consequently, by the time the transgenic line is ready for commercial release, its non-transgenic competitors will have made substantial additional progress by conventional selection.

Copyrighted material

In view of these substantial limitations, it seems likely that the greatest benefits of transgenesis in relation to animal production will come from introducing novel effects, such as utilizing novel foodstuffs, or overcoming problems associated with welfare (e.g. natural disease resistance) or pollution (e.g. pigs transgenic for phytase are able to utilize phosphorus which otherwise is a major pollutant), or producing new forms of conventional animal products that have a more desirable balance of amino acids. The potential for transgenesis in this area is limited only by our ability to devise novel approaches, and our knowledge of the existence and modes of function of genes.

In the shorter term, the most likely use of transgenesis will be in the production of human polypeptides required for pharmaceutical use. However, because just one or two transgenic animals will be sufficient to produce the world's requirements for various human proteins, this use of transgenesis is not likely to have an impact on practical animal production.

Important issues that must be faced in relation to transgenesis are the legal questions (who owns the transgene, and should transgenic animals be patented?) and the ethical and animal-welfare questions. As with animal cloning (see below), many people feel that transgenesis is one step too far. Some scientists working with transgenesis are already devoting considerable time to informing other scientists and members of the public about what is happening, and in return are receiving useful feedback as to what is deemed to be acceptable. This is a laudable approach, and should be followed by scientists working in any area of biotechnology.

Animal cloning

The production of genetically identical animals (animal cloning) is not new. In fact, this has been occurring naturally for millions of years – we are all familiar with identical twins that result from the natural splitting of an embryo soon after fertilization. In the last few decades of the 20th century, it became possible for embryo splitting to be performed in the laboratory, with each portion of the split embryo being inserted into a different foster mother and being brought to term in the usual way. This form of embryo cloning has been widespread in livestock for some time. The main limitation of this technology is the relatively small number of identicals that can be obtained from a single embryo. To overcome this limitation, cloning by nuclear transfer was introduced in the 1980s. This involves obtaining unfertilized ova from a recipient female, removing the nucleus from each ovum, and replacing it with a nucleus (or more commonly a whole cell) from a tissue or cultured cells. At first (in the mid-1980s), the cell was from undifferentiated embryonic tissue; by 1995 it was a cell taken from a line of undifferentiated embryonic cells that had been cultured for several weeks; and the next year, it was a cell taken from a culture of adult (differentiated) cells. The birth of Dolly the sheep, announced in February 1997, created enormous interest because it showed that, contrary to general expectation, it is possible to culture adult, differentiated cells (in this case, mammary cells)

Copyrighted material

Hidden page

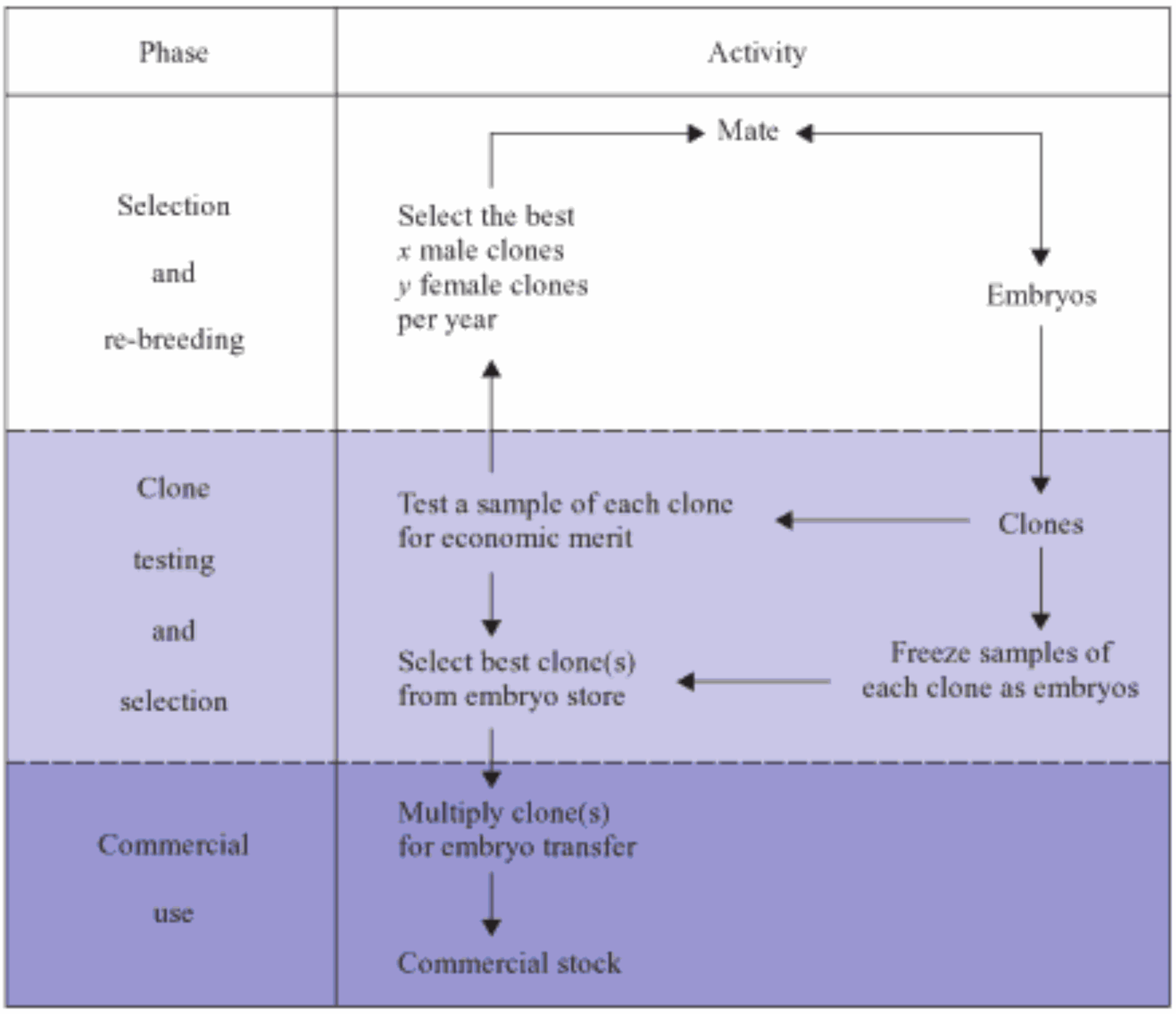


Fig. 19.1 An improvement programme involving the use of embryo cloning. (Drawn from Smith, C. (1989) *Animal Production*, 49, 49–62.)

(embryos) from the best nucleus matings, by clone-testing each embryo. Once the best clones have been identified, it is simply a matter of going to the freezer and extracting a few more embryos from those best clones, and growing them into adult animals, which are then mated to produce the next round of clones. Figure 19.1 illustrates how embryo cloning could be incorporated into an improvement programme.

At the level of commercial production, embryo cloning could have profound effects. For example, in the case of dairy production, sufficient cows could be inembryonated with a female dairy hybrid (two-way, three-way, or four-way cross) to produce the next generation of commercial cows. This involves only a portion of all dairy cows. The remainder could be given cloned embryos for use in the beef industry, e.g. F1 hybrid males for use as natural-service sires in beef herds, F1 hybrid females for use as dams in beef herds, or three-way or four-way cross slaughter-generation beef embryos.

An interesting consequence of the widespread practical application of embryo cloning is that the average genetic merit of the clones that are used in commercial production is substantially higher than the average genetic merit of the nucleus herds from which they are chosen. This is the exact opposite of the lag which occurs in a traditional breeding pyramid, which was described in Chapter 16.

Copyrighted material

With the potential for large-scale benefits, it seems that embryo cloning is certain to become a widespread reality in the future. However, there are some problems that must be considered. For example, there is concern that the widespread use of cloned animals will create problems associated with monoculture. If circumstances change, e.g. if a new and virulent form of a pathogen arises, then whole populations could be eliminated. However, this argument overlooks the fact that a wide range of different clones will be available to commercial farmers. In addition, the clones will be used only at the commercial level; it will be absolutely necessary to maintain as much genetic variation as possible at the nucleus level, so that continual genetic improvement can be achieved, and to provide the range of genotypes that will be essential to cater for the changing circumstances certain to be encountered by commercial producers. A potentially serious problem with embryo cloning is the concern felt by some people at the very thought of animal clones. The fact that many of our most widely-used agricultural plant species are clones, e.g. grapes, bananas, apples, may not provide any solace to people who believe that cloning animals is one step too far along the path of human intervention. These concerns should not be dismissed lightly.

Summary

- Artificial insemination (AI) enables fewer males to be selected as parents of the next generation; its genetic effect is to increase the intensity of selection of males
- Critics of AI point to its potential for increasing the rate of inbreeding, with all the attendant disadvantages, especially in relation to autosomal recessive defects
- What the critics of AI have generally failed to realize is that, at the level of commercial production, AI has the potential to *decrease* the rate of inbreeding compared with natural mating, because it provides a far wider choice of unrelated males than is ever available under natural mating
- Multiple ovulation and embryo transfer (MOET) increases female reproductive potential, and hence can lead to an increase in female selection intensity or a decrease in the female generation interval, by enabling replacements to be obtained from only the younger females in a population
- In effect, MOET changes a monotocous species into a polytocous species, and hence cattle (both dairy and beef), sheep and goats are the species that have most to gain from the use of MOET nucleus populations
- *In vitro* maturation and *in vitro* fertilization of ova offer the potential for artificial insemination to be replaced by artificial inembryonation with mass-produced specific-purpose embryos, which could be straight-breds, or F1 hybrids, or even three-way or four-way hybrids

Continued

Copyrighted material

- The ability to manipulate sex ratio will not have a substantial impact in a selection program, but it can have substantial benefit at the commercial level of some industries, where one sex is preferred over the other
- Transgenesis will continue to be a useful means for producing certain foreign polypeptides, but it is unlikely to have an effect on animal agriculture, unless it can create animals with a very desirable attribute
- Animal cloning has been happening naturally for thousands of years (identical twins)
- Animal cloning by nuclear transfer from differentiated adult cells has considerable potential for conservation of threatened breeds or species, and for genetic improvement programmes

Further reading

Allan, M.F. and Smith, T.P.L. (2008) Present and future applications of DNA technologies to improve beef production. *Meat Science*, **80**, 79–85.

Bertschinger, H.J., Meltzer, D.G. and van Dyk, A. (2008) Captive breeding of cheetahs in South Africa – 30 years of data from the de Wildt Cheetah and Wildlife Centre. *Reproduction in Domestic Animals*, **43 Suppl 2**, 66–73.

Dinnyes, A., Tian, X.C. and Yang, X. (2008) Epigenetic regulation of foetal development in nuclear transfer animal models. *Reproduction in Domestic Animals*, **43 Suppl 2**, 302–9.

Galli, C., Lagutina, I., Duchi, R., Colleoni, S. and Lazzari, G. (2008) Somatic cell nuclear transfer in horses. *Reproduction in Domestic Animals*, **43 Suppl 2**, 331–7.

Galli, C. and Lazzari, G. (2008) The manipulation of gametes and embryos in farm animals. *Reproduction in Domestic Animals*, **43 Suppl 2**, 1–7.

Hill, W.G., Salisbury, B.A. and Webb, A.J. (2008) Parentage identification using SNP genotypes: application to product tracing. *Journal of Animal Science*, **86**, 2508–17.

Laible, G. (2009) Enhancing livestock through genetic engineering – recent advances and future prospects. *Comparative Immunology, Microbiology and Infectious Diseases*, **32**, 123–37.

Leboeuf, B., Delgadillo, J.A., Manfredi, E., Piacere, A., Clement, V., Martin, P., Pellicer, M., Boue, P. and de Cremoux, R. (2008) Management of goat reproduction and insemination for genetic improvement in France. *Reproduction in Domestic Animals*, **43 Suppl 2**, 379–85.

Long, J.A. (2008) Reproductive biotechnology and gene mapping: tools for conserving rare breeds of livestock. *Reproduction in Domestic Animals*, **43 Suppl 2**, 83–8.

Oback, B. (2008) Climbing mount efficiency – small steps, not giant leaps towards higher cloning success in farm animals. *Reproduction in Domestic Animals*, **43 Suppl 2**, 407–16.

Rath, D. and Johnson, L.A. (2008) Application and commercialization of flow cytometrically sex-sorted semen. *Reproduction in Domestic Animals*, **43 Suppl 2**, 338–46.

Smith, G.C., Pendell, D.L., Tatum, J.D., Belk, K.E. and Sofos, J.N. (2008) Post-slaughter traceability. *Meat Science*, **80**, 66–74.

Whitelaw, C.B., Lillico, S.G. and King, T. (2008) Production of transgenic farm animals by viral vector-mediated gene transfer. *Reproduction in Domestic Animals*, **43 Suppl 2**, 355–8.

Copyrighted material

Hidden page

Hidden page

inferred. But measuring sufficient phenotypes (and then relating those phenotypes to identifiable animals) is often a major challenge. Finally, it must be noted that V_A for any trait is a function of heterozygosity at all loci that contribute to the genetic variation for that trait. This highlights the importance of heterozygosity as a measure of genetic diversity.

Measurement of genetic diversity among populations

Genetic diversity among populations can be measured in terms of variation in physical appearance or measured performance, variation in genotypes at a group of loci, and variation in DNA sequence at one or more loci.

Differences in physical appearance or measured performance

Animals can be classified for a range of conformational traits, and these subjective classifications can be quantified by the use of an agreed scoring system. There will be variation among assessors in the subjective scores they allocate, but this type of scoring system is remarkably robust to differences between assessors.

One might imagine that measured performance of traits would be more straightforward than subjective assessment of conformation. However, often the only data available are measurements on each of a number of populations, each measured in their own environment. With such data, the confounding of genetic and environmental effects invalidates most or all of the comparisons. Only if there has been at least one population represented across all environments can useful comparisons of measured performance be made.

Once collected, the subjective data can be subjected to numerical taxonomy analyses, or parameters such as the mean character difference between all pairwise combinations of breeds can be estimated. Descriptions of these methods of analysis are beyond the scope of this book – details are available in Jordana *et al.* (1995) listed under Further reading.

Differences in allele frequencies

The extent to which allele frequencies differ between two populations is called the **genetic distance** between them. As with heterozygosity and allelic diversity, this is best estimated from a representative set of loci.

Differences in DNA sequence

The genetic diversity between two populations can also be estimated by counting the number of bases at which the sequence of a particular locus differs between the two populations. For comparisons across wide ranges of species, the same type of estimate can be obtained by comparing amino-acid sequences of a particular peptide.

Copyrighted material

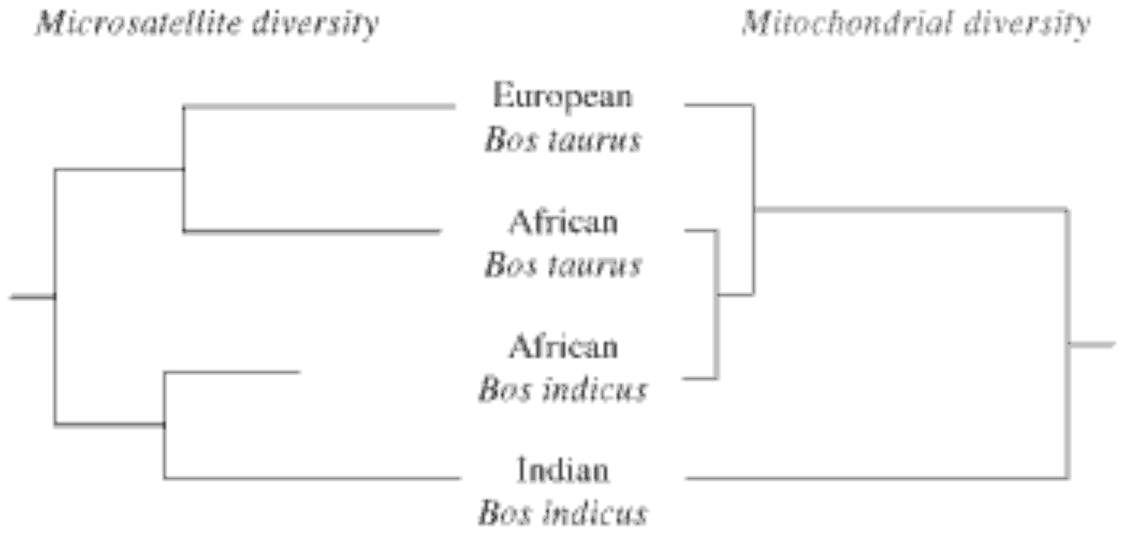


Fig. 20.1 Phylogenetic trees of four of the main groups of cattle. (*Left*) Derived from allele frequencies at microsatellite loci; (*right*) derived from base sequence in mitochondrial DNA. (Reproduced with permission from Bradley *et al.* (1998) *Evolutionary Anthropology*, **6**, 79–86.)

Phylogenetic trees

Each of the above measures of genetic diversity among populations is typically estimated for all pairwise combinations of a set of populations (breeds or isolated populations or species). From these, it is possible to construct a phylogenetic tree, showing the relationship of each population to every other population. In the simplest type of tree, the length of the branches separating any two populations is directly proportional to the extent to which those populations differ in terms of appearance, performance, allele frequencies or sequence. Two simple phylogenetic trees are shown in Fig. 20.1, one (left) derived from allele frequency data and the other (right) from sequence data. In this particular case, the two phylogenies tell partly different stories, because the allele frequency data come from autosomes, while the sequence data come from mitochondrial DNA, which is inherited maternally. Both trees confirm that the largest differences are between *taurus* and *indicus* cattle. They also show that for nuclear genes, African *indicus* cattle are closest to Indian *indicus*, but for mitochondrial sequences, they are closest to African *taurus*. The most likely explanation is that African *indicus* cattle arose from continual crossing of (mostly) male Indian *indicus* cattle with female native African *taurus* cattle.

Importance of genetic diversity

Why is genetic diversity important? Essentially, it is the only means by which populations can immediately respond to selection (artificial and/or natural) either now or when circumstances change in the future. In the case of domesticated species, genetic diversity includes a potentially valuable 'bank' of alleles conferring adaptation to circumstances of animal agriculture that, even though they may not be significant now, could become of major importance in feeding and clothing humanity in the future.

Copyrighted material

We saw in Chapter 8 that the MHC is a set of closely linked loci whose gene products are intimately involved in host immunity – these self peptides present non-self molecules to the passing-parade of T cells, thereby enabling the immune system to respond to pathogens and parasites. We also saw the extreme levels of polymorphism that exist at the MHC, maintained at least in part by selection favouring heterozygotes, because the greater the range of alleles across MHC loci, the greater the range of non-self peptides that can be presented as non-self to the T cells. Given the continuing importance of disease as a threat to all animals, maintaining genetic diversity at the MHC is obviously important. Or, to look at it from the opposite angle, a population that is homozygous at the MHC is at the mercy of any pathogen that can mutate to a form that cannot be presented by the limited array of histoglobulins available, and hence can evade the host's immune response.

As we saw in Chapter 5, there are many recessive alleles that are harmful in the present context and which, consequently, are maintained at a low frequency by a selection/mutation balance. If circumstances change, it is possible that the previously harmful allele now confers an advantage, in which case the population can automatically respond to the new selective forces, resulting in an increase in frequency of the now-favourable allele. However, if that allele has been lost from the population, there can be no response to selection until a favourable allele is created by mutation, which could be too late to save the population.

The same principle applies for traits determined by the action of alleles at many loci. As we saw in equation 13.7 in Chapter 13, the extent to which a population can respond (R) to directional selection for a quantitative trait is determined by the product of heritability (h^2) and selection differential (S), namely $R = h^2S$. This prediction applies equally to artificial selection for sustainable productivity of livestock, and to natural selection for viability and/or reproductive ability following, for example, a change of environment for a non-domesticated species. Since heritability is a simple function of V_A, as shown above and in Chapter 13, it follows that the ability of a population to respond to selection for a quantitative trait is dependent on the magnitude of V_A.

Loss of genetic diversity

Genetic diversity can be lost as a result of artificial selection and/or as a result of decline in population size because of natural or human-directed causes.

Domesticated species

The biggest threat to genetic diversity in livestock species in recent times has been selection between breeds – the favouring of a small number of breeds for livestock production throughout the world, almost irrespective of whether or not the favoured breeds are well adapted to local circumstances. In some tropical countries, for example, imported *Bos taurus* cattle have replaced indigenous breeds. Only later

Copyrighted material

does it become evident that the imports are not as productive in the new environment as they are in the countries in which they are well adapted. In the case of dairy cattle, the global trend has been towards a particular strain of Holstein–Friesian, namely from the USA. The short-term gains from replacing the local Holstein–Friesian strain with the USA strain have been considerable, and there is still much genetic variation within the global population of largely USA Holstein–Friesians. However, it is very likely that some locally adapted Holstein–Friesian genetic variation has been lost in some countries. And in the meantime, other breeds that may in the future be seen to be more efficient at converting grass into milk solids have been reduced to relatively small populations.

Artificial selection within a breed also has the potential to reduce genetic diversity. Indeed, the theoretical result of continual directional selection is fixation of the most favourable allele at every locus contributing to variation in the trait being selected. Once this point is reached, there is homozygosity at every locus relevant to that trait – all additive genetic variation for that trait has been exhausted. In practice, even long-term directional selection experiments in laboratory species have rarely ended up with complete homozygosity – indeed, in most experimental populations, there does not seem to have been any appreciable reduction in additive genetic variation despite many generations of successful directional selection. The reasons for this most likely include greater-than-expected creation of new genetic variation by mutation, and natural selection opposing artificial selection.

Another potential cause for concern is the inbreeding effect of selection. The more intense the selection and the greater the use of information from relatives, the greater is the tendency for animals selected as parents to be related to each other, and hence the greater the increase in level of inbreeding within the population. As we saw in Chapter 12, the inevitable result of inbreeding is a decrease in heterozygosity and a consequent increase in homozygosity. This loss of heterozygosity has two effects: first, there is less additive genetic variation, and hence less response to selection; and second, there is inbreeding depression (decrease in performance as the level of inbreeding increases).

There is another effect of directional selection. The greater the intensity of selection, the smaller the effective population size, and hence the greater the extent of genetic drift. As we saw in Chapter 5, genetic drift refers to changes in gene frequency caused entirely by chance, resulting from the sampling of finite numbers of genes that is inevitable in all finite populations. As we also saw in Chapter 5, the end result of genetic drift is random fixation or loss of alleles from the population; and the smaller the effective population size, the greater is the chance of fixation or loss. Because fixation or loss of alleles is loss of genetic diversity, it follows that artificial directional selection has the potential to lead to an increased loss of genetic diversity through genetic drift.

Non-domesticated species

For non-domesticated species, the main threat to genetic diversity is human encroachment. As populations become smaller, heterozygosity declines because of inbreeding

Copyrighted material

Hidden page

gather relevant data on breeds that seem threatened with extinction, and to act, where possible, to save them. In Ireland, for example, a conservation programme for the Kerry breed of cattle has been underway since 1982. It consists of a modest government subsidy for each calf registered, government support for the maintenance in a national park of the largest remaining herd, and access to software for the calculation of relationships between potential mates, with the aim of minimizing the rate of inbreeding.

Interestingly, the two areas that are probably of greatest concern are at either end of the spectrum of animal improvement. At one end we have a large variety of locally adapted native populations (often in developing countries) that are under threat from the influx of 'improved' breeds and strains from developed countries. At the other end we have an increasing number of poultry selection lines that are discarded when yet another independent poultry breeder is taken over by a larger and often multinational breeding company.

There are no easy solutions to these problems. It is comforting, however, to see increasing activity in conservation around the world, led by the FAO's Global Animal Genetic Resources Programme, with its strong emphasis on local ownership of indigenous genetic diversity, leading to the development of sustainable local breeding programmes embracing the concept of conservation by utilization. In 2007, this programme culminated in the publication of a detailed account of the state of the world's animal genetics resources, and agreement amongst delegates from 109 countries at a conference in Interlaken, Switzerland, on a declaration of responsibility and a global plan of action for animal genetic resources. Listed under FAO in Further Reading at the end of this chapter, the two key documents are freely available from http://www.fao.org/.

If it is not possible to develop such an *in-situ* conservation programme, the next best option is some form of *ex-situ* conservation, which could involve the storage of ova or semen or zygotes. As mentioned in Chapter 19, the creation of whole animals from differentiated somatic cells, as first exemplified by Dolly, offers the prospect of conserving a breed for which there is no immediate need, by simply storing a small sample of tissue from each of a representative group of animals.

Non-domesticated species

Here the challenges are daunting but not impossible. When a species is threatened with extinction in the wild, one possibility is to capture the survivors, and to attempt to breed them in captivity, with the ultimate aim of increasing the population size sufficiently that descendants can be released back into the wild. These so-called **captive breeding programmes** are becoming increasingly popular, especially in zoos, and some have already achieved their aim. However, it is important to realize the limitations of such programmes. For example, even if the population size were eventually to grow to hundreds or even thousands, the level of heterozygosity is the same as in the few original survivors, because increasing the population size does not in itself create new alleles. In fact, of course, the only source of new alleles is mutation, and, unfortunately, mutation is a painfully slow means of regenerating lost

Copyrighted material

alleles. Another limitation is that the longer a population is bred in captivity, the more strongly has that population been selected for adaptation to captivity, which could compromise its chances of surviving when released into the wild.

Another limitation is inbreeding depression of fitness traits. Suppose, for example, that among the survivors of a particular species, only one pair produces offspring. When these parents die, and the only surviving offspring are ready to reproduce, there is no alternative but to mate full-sibs, which results in offspring with 25 per cent inbreeding, which in turn could drastically reduce the chance of this next generation of offspring surviving. It is possible, therefore, that a captive breeding programme could actually end in the extinction of the species, despite the best intentions of the people involved. However, if the survivors were all closely related in the first place, there may be relatively little inbreeding depression in subsequent generations, because the survivors already share many genes. Overall, the important conclusion to be drawn is that captive breeding programmes are not without their risks, and that people should not be too optimistic about the results of any particular programme. In many cases, however, there is no other option, and some programmes have been very successful.

At a less extreme level, it is becoming increasingly common for zoos to keep pedigree records, and to calculate the relationship between all possible mating pairs that could occur in a particular species or population. These calculations can be made jointly for several different zoos, with a view to exchanging animals (or semen) so as to maximize the avoidance of inbreeding, and hence to minimize inbreeding depression. A somewhat different approach, which is best for retaining genetic diversity, is to arrange matings so as to minimize the average mean kinship of the population, where the mean kinship of an animal is one-half the average relationship of that animal to all other animals in the population.

As with domesticated species, there is always the last resort of *ex-situ* conservation by storage of gametes or zygotes. However, this form of technology is available for only a very limited range of species. As also with domesticated species, cloning from differentiated cells offers a longer-term prospect.

Summary

- If a population can be genotyped for some DNA markers, its level of genetic diversity can be estimated as observed heterozygosity, expected heterozygosity, allelic diversity, and effective number of alleles
- The genetic distance between two populations is the extent to which their allelic frequencies differ
- The genetic diversity between two populations can also be estimated by counting the number of bases at which the sequence of a particular locus differs between the two populations

Continued

Copyrighted material

- Each of the above measures of genetic diversity among populations is typically estimated for all pairwise combinations of a set of populations (breeds or isolated populations or species), from which it is possible to construct a phylogenetic tree, showing the relationship of each population to every other population
- Genetic diversity is important because it is the only means by which populations can immediately respond to selection (artificial and/or natural) either now or when circumstances change in the future
- The biggest threat to genetic diversity in livestock species in recent times has been selection between breeds
- For non-domesticated species, the main threat to genetic diversity is human encroachment
- To maintain around 90 per cent of a population's genetic diversity for 100 years, the effective population size must be around $500/L$; and to retain evolutionary potential in the longer term (where generation intervals are not relevant), an effective population size of at least 500 is required
- Captive breeding programmes conducted by zoos and other organizations are good examples of the practical application of population and quantitative genetics

Further reading

Boakes, E.H., Wang, J. and Amos, W. (2007) An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity*, **98**, 172–82.

Colleau, J.-J. and Avon, L. (2008) Sustainable long-term conservation of rare cattle breeds using rotational AI sires. *Genetics Selection Evolution*, **40**, 415–32.

Edmands, S. (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, **16**, 463–75.

FAO (2007) *The state of the world's animal genetic resources for food and agriculture*, Rischkowsky B, and Pilling D (eds). FAO, Rome. (freely available from http://www.fao.org/)

FAO (2007) *Global plan of action for animal genetic resources and the Interlaken declaration*. FAO, Rome. (freely available from http://www.fao.org/)

Frankham, R. (2008) Genetic adaptation to captivity in species conservation programs. *Molecular Ecology*, **17**, 325–33.

Frankham, R., Ballou, J.D. and Briscoe, D.A. (2004) *A Primer of Conservation Genetics*. Cambridge University Press, Cambridge.

Frankham, R., Ballou, J.D. and Briscoe, D.A. (2009) *Introduction to Conservation Genetics*, 2nd edn. Cambridge University Press, Cambridge.

Fredrickson, R.J., Siminski, P., Woolf, M. and Hedrick, P.W. (2007) Genetic rescue and inbreeding depression in Mexican wolves. *Proceedings of the Royal Society B*, **274**, 2365–71.

Copyrighted material

Glowatzki-Mullis, M.L., Muntwyler, J., Bäumle, E. and Gaillard, C. (2009) Genetic diversity of Swiss sheep breeds in the focus of conservation research. *Journal of Animal Breeding and Genetics*, **126**, 164–75.

Gulisija, D. and Crow, J.F. (2007) Inferring purging from pedigree data. *Evolution*, **61**, 1043–51.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.-L., Harshman, J., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Steadman, D.W., Witt, C.C. and Yuri, T. (2008) A phylogenomic study of birds reveals their evolutionary history. *Science*, **320**, 1763–8.

Hall, S.J.G. (2004) *Livestock Biodiversity: Genetic Resources for the Farming of the Future*. Wiley-Blackwell, Oxford.

Hogg, J.T., Forbes, S.H., Steele, B.M. and Luikart, G. (2006) Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B*, **273**, 1491–9.

IUCN *Red List of Threatened Species*. IUCN, Gland, Switzerland (website http://www.iucn.org/ or http://www.iucnredlist.org).

Jamieson, I.G., Wallis, G.P. and Briskie, J.V. (2006) Inbreeding and endangered species management: is New Zealand out of step with the rest of the world? *Conservation Biology*, **20**, 38–47.

Jordana, J., Pares, P.M. and Sanchez, A. (1995) Analysis of genetic relationships in horse breeds. *Journal of Equine Veterinary Science*, **15**, 320–28.

Leberg, P.L. and Firmin, B.D. (2008) Role of inbreeding depression and purging in captive breeding and restoration programmes. *Molecular Ecology*, **17**, 334–43.

Lipinski, M.J., Lutz Froenickea, L., Baysaca, K.C., Billingsa, N.C., Leuteneggerb, C.M., Levyc, A.M., Longerid, M., Niinie, T., Ozpinarf, H., Slaterg, M.R., Pedersenb, N.C. and Lyons, L.A. (2008) The ascent of cat breeds: Genetic evaluations of breeds and worldwide random-bred populations *Genomics*, **91**, 12–21.

Mills, J.S. (2006) *Conservation of Wildlife Populations: Demography, Genetics and Management*. Wiley-Blackwell, Oxford.

Roldan, E.R., Gomendio, M., Garde, J.J., Espeso, G., Ledda, S., Berlinguer, F., del Olmo, A., Soler, A. J., Arregui, L., Crespo, C. and Gonzalez, R. (2006) Inbreeding and reproduction in endangered ungulates: preservation of genetic variation through the Organization of Genetic Resource Banks. *Reproduction in Domestic Animals*, **41 Suppl 2**, 82–92.

Scherf, B.D. (ed.) (2000) *World Watch list for domestic animal diversity*, 3rd edn. Food and Agriculture Organization of the United Nations, Rome (available from http://dad.fao.org/).

Searle, J.B. (2008) The genetics of mammalian invasions: a review. *Wildlife Research*, **35**, 185–92.

Vonholdt, B.M., Stahler, D.R., Smith, D.W., Earl, D.A., Pollinger, J.P. and Wayne, R.K. (2008) The genealogy and genetic viability of reintroduced Yellowstone grey wolves. *Molecular Ecology*, **17**, 252–74.

Windig, J.J., Meuleman, H. and Kaal, L. (2007) Selection for scrapie resistance and simultaneous restriction of inbreeding in the rare sheep breed 'Mergellander'. *Preventive Veterinary Medicine*, **78**, 161–71.

Copyrighted material

Hidden page

Hidden page

that is learnt about the genetic basis of an inherited disor der, the more it is likely that a non-genetic treatment will be developed.

Transplantation and corrective surgery

While somewhat different from the non-genetic factors discussed above, transplantation and corrective surgery can both be regarded as non-genetic means of alleviating inherited disorders.

Transplantation of bone marrow, liver, or kidney, from normal animals to affected animals, has the potential to provide the necessary enzyme or blood-clotting factor in individuals suffering inherited deficiencies of these polypeptides. This type of treatment has produced positive results for disorders such as citrullinaemia, haemolytic anaemia due to pyruvate kinase deficiency, Gaucher's disease, and von Willebrand's disease.

Corrective surgery is also sometimes used to alleviate an inherited disorder. For example, some veterinarians have claimed that cutting the pectineus muscle alleviates hip dysplasia in dogs. The evidence for this is somewhat controversial but the operation is quite popular in some countries. There are numerous other examples where surgery has been used to 'hide' structural defects.

Genetic effects of environmental control

Non-genetic control certainly alleviates inherited disorders. In the case of PKU, the simple change in diet has essentially eradicated the disorder. It has not, of course, removed the mutant alleles that cause PKU, and this raises an important question about non-genetic control of inherited disorders. If individuals carrying defective alleles are enabled to survive and reproduce, and hence pass on harmful or even lethal alleles, will there be an increase in the frequency of these alleles in the future, giving rise to an ever-increasing number of individuals requiring treatment?

The basic principles of population genetics tell us that if we remove selection pressure from a previously harmful allele so that all genotypes have equal fitness, on average the frequency of each allele will remain constant, except for the effect of mutation, which will tend to increase the frequency of the previously undesirable allele, but at an imperceptibly low rate.

Thus, in principle, there will be no extra problems created in the future by the use of non-genetic control of inherited disorders, provided that all genotypes have an equal chance of contributing alleles to the next generation. In practice, however, this proviso is often not met. For example, if an animal commands a large stud fee but has an inherited disorder, it is tempting to correct or treat the disorder and then to use that animal as extensively as possible. In effect, this increases the relative fitness of the undesirable genotype and consequently increases the frequency of the undesirable allele or alleles. In practice, therefore, non-genetic control of genetic disorders should be used with caution, and animals so treated should not be allowed to make large contributions to the next generation. They can, however, be used repeatedly in cross-bred matings to produce commercial (non-pedigreed) animals, such as those

Copyrighted material

Hidden page

Pedigree analysis

Pedigree analysis involves studying available pedigrees, with the aim of estimating the probability that a normal prospective parent is homozygous. While straightforward in principle, the mathematics are quite complex. However, algorithms have been written to perform the necessary calculations, and software incorporating such algorithms is becoming increasingly available to breeders (see, for example, the WINCANIS pedigree software at http://www.geocities.com/willowind_dals/pedigree). Such software is very useful as a means of decreasing the frequency of undesirable alleles without involving breeders in any major expense or trouble: the computer programme tells them which animals are most likely to be homozygous, and it is these animals that should be used in future matings.

Test matings

Test matings involve considerable extra effort, expense and time but they can be very useful. The most common options involve mating the prospective parent with:

(1) homozygotes for the recessive allele, or
(2) known heterozygotes, or
(3) progeny of the prospective parent, or
(4) a random sample of the population.

There are two possible outcomes from a test mating. If an offspring from a test mating is affected, and hence is homozygous for the undesirable allele (*aa*), that offspring must have received one of its *a* alleles from the prospective parent. Obviously, in this case the prospective parent must be a carrier. The other possible outcome is that the offspring is normal, in which case there are two alternative explanations: either the prospective parent is homozygous for the normal allele (*AA*), or it is really a carrier (*Aa*) but by chance remained undetected because it passed on its *A* allele rather than its *a* allele. The aim of a test-mating programme is to distinguish as efficiently as possible between these two alternatives.

Consider matings with homozygous recessives. If the prospective parent is a carrier, a test mating of this type can be written as $Aa \times aa$, with the expected results being one-half *Aa* (normal) offspring and one-half *aa* (affected) offspring. Recalling that the carrier parent remains undetected if an *Aa* offspring is produced, the chance of a carrier parent remaining undetected after producing one offspring is 0.5. As the genotype of each offspring is independent of the genotypes of all other offspring, the chance of a carrier parent remaining undetected after two offspring is $0.5 \times 0.5 = (0.5)^2 = 0.25$; after three offspring it is $(0.5)^3 = 0.125$; and after n offspring it is $(0.5)^n$. Thus, for each additional offspring, the chance of a carrier parent remaining undetected decreases by one-half.

The important question that has to be asked in practice is: how many offspring do we need to observe in order for there to be only a very small chance of a carrier parent remaining undetected? Suppose, for example, that we would like the chance

Copyrighted material

of a carrier remaining undetected to be no greater than 5 per cent. In this case, we require sufficient offspring per prospective parent to ensure that $(0.5)^n$ is less than or equal to 5 per cent. By solving the equation $(0.5)^n \leq 0.05$, we conclude that $n = 5$ is the required number of offspring. If we want the chance to be 1 per cent, then the required number is $n = 7$.

It should be obvious that we can never prove that a parent is homozygous normal in a test mating. All we can do is to reduce the chance of the parent being undetected as a carrier, to some acceptably low level. On the other hand, as soon as the first affected offspring is produced, we have proved that the parent of that offspring is a carrier.

The principle is the same for the other types of test matings listed above; the main difference is in the chance of a carrier remaining undetected. Full details of how this is derived for each type of test mating are provided in *Veterinary Genetics* (Nicholas, 1987, pp. 319–22). The results are shown in Table 21.1. It can be seen that matings to homozygous recessives are most efficient. If such individuals are not available, the next best option is to use known carriers. Although the last two types of test matings appear to be much less efficient, they do have certain advantages. For example, they do not involve the expense of establishing and maintaining a special group of tester animals. More importantly, they provide a test for undesirable recessive alleles at all loci rather than just at one locus, as is the case for the first two types of matings. In comparing the last two types of matings, Table 21.1 suggests that it is more efficient to mate a prospective parent to its offspring than to a random sample of the population, unless the recessive allele is quite common in the population. Against this, however, is the fact that the use of offspring as testers greatly increases the time required to test a prospective parent. The final type of mating is the one that occurs naturally in progeny-testing schemes such as those occurring throughout the world in the dairy industry. Notice that the efficiency of the test

Table 21.1 The number of offspring required in test matings for a recessive gene. In the fourth type of test mating, q is the frequency of the recessive gene in the general population

Prospective parent mated to:	Chance that a carrier remains undetected with n offspring	Number of offspring required to reduce the chance of non-detection to:		
		0.05	0.01	0.001
1. Homozygote recessives	$(0.5)^n$	5	7	10
2. Known carrier	$(0.75)^n$	11	16	24
3. Offspring of prospective parent	$(0.875)^n$	23	35	52
4. Individuals chosen at random	$(1 - 0.5q)^n$, where q equals			
	0.2	29	44	66
	0.1	59	90	135
	0.01	598	919	1379

Copyrighted material

Hidden page

Finding useful DNA markers

One way to search for a useful marker is the so-called **candidate-gene** approach. This involves investigating what is known about the biochemistry and/or the genetics of the disorder in all species in which it has been reported, with a view to identifying genes that could possibly correspond to the disorder locus. An example of this approach is provided by the search for a marker for the halothane locus. A wealth of basic physiological and biochemical research over many years led to the suggestion that the gene for the calcium-release channel (CRC, also known as the ryanodine receptor – see Chapter 9) might be the halothane gene. As soon as the *CRC* gene was cloned in one species of mammal (human), it was a relatively straightforward task to clone it in other species, and to test whether it was linked to the halothane gene in each species. In pig families, and in some (but not all) human families, the recombination fraction turned out to be zero, i.e. the *CRC* gene is completely linked to the halothane gene. As described in Chapter 9, further research showed that the *CRC* gene is actually the halothane gene, and a C → T substitution at position 1843 is the actual cause of the defective (*hal*) allele in populations of pigs throughout the world. Since this base substitution creates or destroys the recognition sequence for several restriction enzymes, it was a simple matter to develop a PCR–RFLP genotyping test for this mutation. Pigs can therefore be genotyped at the halothane locus on the basis of the polymorphism that actually causes the disorder (Fig. 21.1).

Not all candidate-gene searches turn out to be as successful as this one: many candidate genes turn out not to be linked to the disorder locus.

A set of mapped markers provides a powerful alternative to the candidate-gene approach. If the markers are equally spaced along all chromosomes, at least two of the markers must be linked to each locus in the genome (one on either side). As described in Chapter 2, tens of thousands of mapped markers are becoming available in domestic species. Encouragingly, as described in Chapter 3, there is now strong evidence that by genotyping as few as 20 animals (split into roughly equal numbers of affecteds and normals) with tens of thousands of DNA markers that cover all regions of the nuclear genome, it is possible to identify a region as small as 1–10 Mb

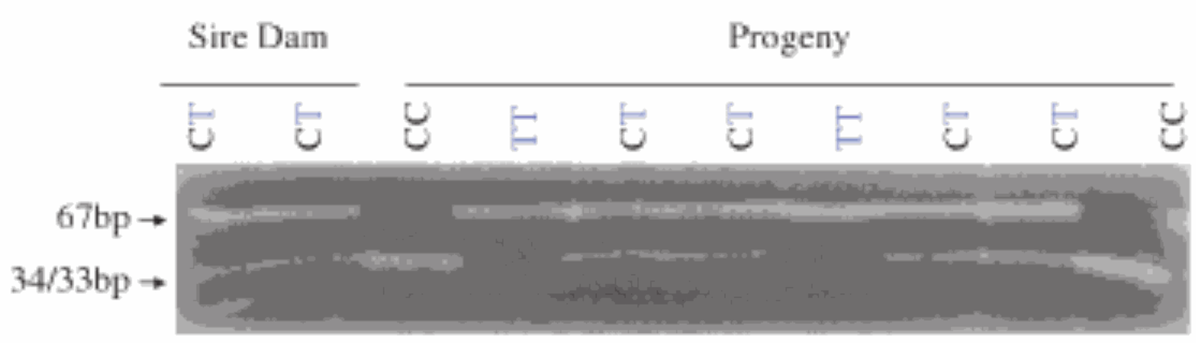


Fig. 21.1 A PCR–RFLP test (using the enzyme *Hha*I) for malignant hyperthermia syndrome (MHS) in pigs, which determines the genotype at the site of the causal mutation in the calcium-release-channel gene. CC = homozygous normal, CT = heterozygous, TT = homozygous for *hal*. (Reproduced courtesy of I.P. Hughes.)

Copyrighted material

in which the defective gene must be located. With 1 Mb of genomic DNA containing on average around 10 genes, even with a region of 10 Mb it is feasible to look through the list of genes in that region and to narrow the field down to a handful of likely candidates (based on the likely biochemical and physiological effect of a deficiency in the gene product or on knowledge that a particular gene, when mutated in another species, gives rise to a similar disorder). In several cases of the application of this strategy, as described in Chapter 3, sequencing of the most likely candidate gene in affecteds and controls has identified the molecular basis of the disorder.

Using DNA markers in the control of single-gene disorders

Markers other than the causal mutation

At one extreme, the marker might be completely unlinked to the disorder locus, e.g. it could be on a different chromosome, but a recent importation (even of just one animal) could have generated substantial linkage disequilibrium. In this case, the marker is of only very limited use, because its non-random association with the defective allele (the very thing that makes it useful) decreases by one-half every generation. At the other extreme, the marker might be located very close to the disorder locus (it could even be located within the locus, but still be many kilobases away from the causal mutation), in which case the linkage disequilibrium could have existed for many generations, and will be decaying at an imperceptibly slow rate. In this case, the marker will be a very useful predictor of disorder-locus genotype at the population level for many, many years.

When the marker is the causal mutation

When the DNA-marker polymorphism actually corresponds to the causal mutation, as with MHS, it is a straightforward matter to determine whether any animal is a homozygote or a heterozygote, simply by genotyping that animal for the causal mutation (Fig. 21.1). For an up-to-date list of disorders for which there is a known molecular basis, and hence a direct DNA test, see OMIA at www.omia.angis.org.au.

A potential limitation

New mutations are occurring all the time, albeit at a slow rate. This means that even if a stage is reached where the actual causal mutation of a particular inherited disorder has been identified and is used as the basis of a genotyping test, it is always possible that a new mutation will occur at a different site in the same gene, or even in another gene, giving rise to the same clinical signs. We must always be prepared, therefore, for a well-established DNA genotyping test to 'miss' some carriers of defective alleles. In haemophilia A and other X-linked disorders in humans, for example, about one-third of affected individuals have novel mutations.

This means that we can never hope to keep up with the occurrence of harmful alleles, and we should therefore be prepared for any DNA test to fail sooner or later. Depending on the size of the gene, it may be possible without too much trouble to identify the new mutation by sequencing.

Copyrighted material

Gene therapy

Gene therapy is a logical extension of transplantation. One of the major problems of transplantation, namely the problem of rejection, can be overcome if the patient's own cells are removed, 'repaired' by the addition of normal genes, and then replaced. Alternatively, normal genes can be inserted into the appropriate tissue *in vivo* by indirect means, such as via a harmless virus or a liposome carrying the normal host gene. Considerable research effort is being devoted to gene therapy, and substantial progress has been made. There are still, however, major practical problems to be solved.

Even if it can be made to work regularly and effectively, gene therapy is applicable only to those disorders in which the relevant gene is expressed in a tissue that is accessible, either directly or indirectly. In any case, it is doubtful whether such therapy will be of much practical use in domestic animals. The reason for this is that the molecular technology for heterozygote detection is likely to progress at least as fast as that of gene therapy, so that by the time gene therapy becomes a practical reality, heterozygote detection will most likely be available for the majority of single-gene disorders. Once this stage is reached, it will be far easier to prevent the birth of affected animals by genotyping prospective parents, and arranging appropriate matings. Only in species like humans, where this type of control is unacceptable, is the addition of normal genes to previously defective cells likely to have much impact. However, most of the development work for human gene therapy uses animal models of human disorders, in which veterinarians have a major interest.

Genetic control of multifactorial disorders

The genetic control of multifactorial disorders is best illustrated by hip dysplasia in dogs.

Hip dysplasia is probably the best known of all animal disorders that are familial but not due to a single gene; it has been the subject of more research, and more controversy, than any other similar disorder. One of the causes of controversy is that hip dysplasia is traditionally diagnosed not on clinical signs but by subjective evaluation of a radiograph, as illustrated in Chapter 6. The problem with this approach is that in most populations, the incidence of abnormal hips as diagnosed by radiography is much higher than the incidence of clinical hip dysplasia (CHD). This immediately creates a credibility gap between veterinarians and breeders, because all too often a dog that can jump a fence six feet high is diagnosed as being dysplastic according to its radiograph. The solution to this problem is for breeders and veterinarians to distinguish between the breeding objective (the trait that we wish to improve) and the selection criterion (the trait that is actually measured), as described in Chapter 15. In the present context, the breeding objective is clearly CHD. Since it is very difficult to measure CHD, and it may not be expressed at a young age, the selection criterion in most hip dysplasia control programmes is 'radiographic' hip dysplasia (RHD), which is readily assessable at an early age. The fact that RHD is assessed

Copyrighted material

subjectively does not detract from its usefulness as a selection criterion. All that is required is that RHD be assessed on an arbitrary scale, and that this assessment has a positive genetic correlation (see Chapter 13) with CHD. Although no estimates of this correlation have been made, the available evidence indicates that it is positive and sufficiently high to justify the use of RHD as the selection criterion in control programmes.

The heritability of RHD has been estimated on many occasions, with estimates usually falling in the range 0.25–0.40. This is sufficiently high to justify a selection programme based on simple mass selection, which is selection of individuals according to their own phenotype. As mentioned above, phenotype for RHD is usually assessed on an arbitrary scale. For example, in the control programme run jointly by the British Veterinary Association (BVA) and the Kennel Club (KC) (which is now the standard scheme for almost all breeds in the UK, New Zealand, and Australia), nine different radiographic features of each hip are each assessed on a scale from 0 (ideal) to 6 (worst), giving a total potential range of subjective scores from 0 to 108. With such a large range of possible scores, selection on RHD is somewhat similar to selection on a continuously varying trait such as body weight; to an extent, the BVA/KC scoring system has changed RHD from a threshold trait to a conventional multifactorial (quantitative) trait. It is not exactly the same, however, because the scores are not actually measurements, but instead are ratings; a radiographic feature with a score of, say, 4 does not reflect hip dysplasia twice as bad as a score of 2. And the same applies to most hip-dysplasia schemes throughout the world, with the exception of the PennHip scheme, which is based on actual measurements of features in radiographs (taken under three different views: distraction, compression, and hip-extended).

It is straightforward to apply all the principles described in Chapter 15 to calculate EBVs for liability for any disorder for all candidates, whether based on scores or measurements. Armed with such EBVs, breeders can then calculate the average liability of offspring from any potential mating, which is simply the average of the two parental EBVs. Matings can then be chosen so as to minimize offspring liability. This is a very powerful tool for decreasing the liability of a population, and hence decreasing the incidence and severity of disorders. The only requirements are a scheme by which the disorder is scored or measured in an agreed manner, with the data being assembled in a database alongside pedigrees. The data can be in just two categories (affected/normal) or in many categories (as in the BVA/KC hip dysplasia scheme). The services of a quantitative geneticist will be required initially to provide advice on the method of calculating the EBVs. Once established, such a scheme can operate with little intervention. It has the virtue of imposing no arbitrary breeding bans on animals with phenotypes worse than a particular arbitrary standard. And it works!

Use of DNA markers

As mentioned in Chapters 6 and 7, as tens of thousands of mapped DNA markers become increasingly available in each domestic species, sets of markers will be

Copyrighted material

Hidden page

tens of thousands of defective alleles. Even though the rate of mutation is very low, the inevitable consequence of the above biological facts (as we saw in Chapter 5) is that almost all animals, including the purest of purebreds (plus the readers of this book, and its author!), are carrying several severely deleterious alleles. Given this biological reality, it is obviously pointless to try to achieve a situation where a population is free of all harmful alleles. In fact, the only way to achieve this state would be to cull almost all animals!

As soon as these biological realities are grasped, it is obvious that we have to learn to live with harmful alleles. For autosomal recessive disorders, which are by far the most common, it has already been stressed that so long as at least one parent is not carrying the harmful allele, *no* affected offspring will be produced. In other words, the incidence of the disorder can be reduced to zero, simply by ensuring that each mating involves at least one non-carrier. Obviously the harmful allele is still present in the population, but the disorder itself will have disappeared, which is all we can realistically hope for.

In order to achieve this, it is necessary to have a simple means of genotyping individuals for known harmful alleles. Wherever such tests exist, breed societies and artificial-breeding organizations should be encouraged to genotype breeding animals for known harmful alleles, and to widely publicize the results of the tests. An atmosphere needs to be developed in which breeders are rewarded for telling others about which harmful alleles their animals carry, rather than, as in most cases at present, being persecuted or, at the very least, being financially penalized through lost sales of breeding stock. Breeders must realize that having harmful alleles is a biological inevitability rather than something to be ashamed of. Moreover, if they are willing to go public about the occurrence of harmful alleles, the incidence of inherited disorders can be dramatically reduced, because this information enables breeders to avoid those matings that are likely to produce animals with inherited disorders.

The problems created by some breed standards

Most breeds have an official standard to which all breeders aspire, and which forms the basis for the decisions of judges at shows. Unfortunately, the official standards of some breeds actually encourage breeders to select for inherited disorders. The most obvious examples are those in which the breed standard (or a favoured phenotype) is based on heterozygosity for a defective gene, e.g. Scottish Fold cats, Manx cats, Merle dogs, and Overo horses (see Chapter 11). However, there are many other, more subtle, cases (especially in companion animals) where the breed standard encourages breeders to select for undesirable features. Overcoming these problems involves changing the culture of breed societies, which is obviously not a simple task. All that can be done in the short term is to be aware of the problems created by some of the current breed standards, and to try to encourage a gradual change in philosophy, with the eventual aim of having breed standards that are a reflection of healthy animals, and which are not conditional upon the presence of defective alleles.

Copyrighted material

Crossing: thinking outside the square

There are several important lessons that breeders of companion animals can take from the widespread use of crossing in production animals, as was discussed in Chapter 17.

Grading-up

Many breed societies maintain what are, in effect, closed studbooks; prohibiting the introduction of new genetic variation from other breeds, this confines breeders to continually 'puddling' in their own gene pools. When a breed is numerically small, this can be particularly troublesome, with breeders having few alternatives but to arrange matings amongst close relatives, and with particular disorders being so frequent that they are almost a hallmark of the breed. A clear solution is grading-up, which appears to be so radical as to be beyond consideration. But in reality, it is not nearly as radical as it appears to be.

We saw in Chapter 17 that it is possible to introduce a new gene or genes into a particular (target) breed by doing an outcross to whatever population has the desired gene or genes, and then repeatedly backcrossing to the target breed. We can now reinterpret the conclusions in Chapter 17 in the light of questions concerning the opening-up of closed breeds to doses of foreign genes. Importantly, the lessons of Chapter 17 tell us that by the second or third backcross, there will be offspring with proportions of genes from the target breed ranging from nearly $\frac{1}{2}$ to nearly 1. In other words, there is substantial variation in the proportion of genes from the target breed after the first backcross. What does this mean in practice? Consider, for example, crossing a Labrador (high incidence of hip dysplasia) with a Greyhound (zero incidence of hip dysplasia), and then doing several backcrosses to Labradors, selecting for (a) Labrador appearance and (b) low EBV for hip dysplasia. By the third backcross, there will be dogs that are indistinguishable from purebred Labradors, but which have far lower liability to hip dysplasia.

A practical example of such a scheme is provided by Dr Bruce Cattanach, an experienced Boxer breeder and international judge, who outcrossed to Corgis to introduce the bob-tail trait (short tail; single-locus, autosomal dominant) into Boxers. By the second backcross, he had animals that were indistinguishable from Boxers. Third backcross animals were registered as Boxers by the Kennel Club, and had success in the show ring.

This is a wonderful example of how outcrossing followed by backcrossing can very quickly recreate a breed with new genetic variation.

Regular crossing

One of the striking features of F1 animals produced by crossing two breeds is their relatively high degree of uniformity of phenotype (including appearance, behaviour, etc.). This has been recorded for centuries in livestock breeding, and was also evident in Dr Cattanach's crosses. The reasons for this uniformity are obvious: each F1

Copyrighted material

Hidden page

Hidden page

Glossary

ABS	antigen-binding site
AI	artificial insemination
ASA	allele-specific amplification
ASIP	agouti-signalling protein
ASO	allele-specific oligo
ASS	argininosuccinate synthetase
BAC	bacterial artificial chromosomes
BC	backcross
BLAD	bovine leucocyte adhesion deficiency
BLUP	best linear unbiased prediction
bp	base pair
BrdU	bromodeoxyuridine
BSE	bovine spongiform encephalopathy
BZ	benzimidazoles
cDNA	complementary DNA/copy DNA
CDS	coding sequence
CHD	clinical hip dysplasia
cM	centiMorgan
CNP	copy number polymorphism
CNV	copy number variation
CRC	calcium release channel
CSD	conotruncal septum defects
DD	ductus diverticulum
DGAT1	diacylglycerol acyltransferase
DGGE	denaturation gradient gel electrophoresis
DNA	deoxyribonucleic acid
DUMPS	deficiency of uridine monophosphate synthase
EBV	estimated breeding value
EST	expressed sequence tag
FACS	fluorescence-activated cell sorting
FAO	Food and Agriculture Organization of the United Nations
FISH	fluorescence *in situ* hybridization
G6PD	glucose 6-phosphate dehydrogenase
GHR	growth hormone receptor

Copyrighted material

GP	grandparent
GWA	genome-wide association
GWS	genome-wide selection
HbA	haemoglobin A
HbB	haemoglobin B
HRM	high-resolution melt
IGF2	insulin-like growth factor 2
INSDC	International Nucleotide Sequence Database Collaboration
ISCNDB	International System for Cytogenetic Nomenclature of Domestic Bovids
IUCN	International Union for the Conservation of Nature
IVM	*in vitro* maturation
IVF	*in vitro* fertilization
kb	kilobase
LBTS	lethal brachygnathia trisomy syndrome
LD	linkage disequilibrium
LINE	long interspersed elements
LWFS	lethal white foal syndrome
MACES	Mexican–American Commission for the Eradication of Screw-worm
MAF	minor allele frequency
MAI	marker-assisted introgression
MAS	marker-assisted selection
Mb	megabase
MCMC	Markov chain Monte Carlo
MD	muscular dystrophy
MGF	mast/stem cell growth factor
MGFR	mast/stem cell growth factor receptor
MHC	major histocompatibility complex
MHS	malignant hyperthermia syndrome
MOET	multiple ovulation and embryo transfer
mRNA	messenger RNA
miRNA	microRNA
NAT	*N*-acetyltransferase
ncRNA	non-protein-coding RNA
NF	nombre fondamental
NI	neonatal isoerythrolysis
NOR	nucleolar organizer region
O	offspring
ORF	open reading frame
P	parent
PBR	peptide-binding region
PCP-C-I	procollagen I carboxy-proteinase
PCP-N-I	procollagen I amino-proteinase
PCR	polymerase chain reaction

Copyrighted material

Hidden page

Index

Bold indicates appearing in a chapter summary; ***bold italic*** indicates appearing in a figure.

acceptor site, 21
accuracy of selection (r_{AC}), 236
acrocentric chromosome, 3
activator, 22, **37**
additive, 219, 220
  genetic variance (V_A), 221–2, **228**, 268, 281, 284, 285
  relationship, 211, 223, 238
adenine, 12, ***13***, ***14***
allele, 24–5, **37**
  frequency; *see* gene frequency
alleles
  effective number of, 281, **288**
  multiple, 25
allele-specific
  amplification (ASA), 63
  oligo (ASO), 63
allelic diversity, 281, **288**
alpaca, karyotype, 4
amino acid, 16, 17
  motif, 22
anastomosis, vascular, 114, 115
ancestor, common, 207
aneuploidy, 104, 119
animal
  cloning, 275–6, **279**
  model, 239
annealing, in PCR, 56, ***56***, 57
anthelmintics, resistance to, 186
antibiotics, resistance to, 186–7
antibody, 158, 159–60, ***159***, 170
  diversity, 160, 170
anticipation, genetic, 95
anticoding strand, of DNA, 20
anticodon, 17
antigen, 158
  red-cell, 160–62
antigen-binding site (ABS), of MHC, 165
antigenic variation, **192**
  in trypanosomes, 183
antisense, 226
  RNA, 80, **83**
  strand, of DNA, 20
  technology, 80–81
artificial
  insemination (AI), 270–71, **278**
  selection, 127, 230
assembly
  genome, 35
  sequence, 74–6
autosomal
  dominant inheritance, 151–2
  recessive inheritance, 152
autosome, 3, 4

BAC, 51
  fingerprint map, 73
  library, 73
backcross, 70, 256, 259, 260, **265**
bacterial artificial chromosome (BAC), 51
bacteriophage lambda, 48, ***48***, 49
banding of chromosomes, 3–5
  C, 4
  DAPI/AMD, ***41***
  G, 4, ***39***
  GBG, 4
  GTG, 4, ***40***, ***42***
  Q, 4
  R, 4
  RBA, 4
  RBG, 4, ***43***, ***44***
Barr body, 32–3, ***33***, 34, 103
base,
  in DNA, 12
  pair (bp), 14
  population, 211–2
  substitution
    non-synonymous, 167
    synonymous, 167
best linear unbiased prediction; *see* BLUP
biochemical screening, 297, **306**
biotechnology, 270–79
bivalent, in meiosis, 7
blood-group system, 160–62, 170
blot
  dot, 63
  reverse dot, 63
blotting
  northern, 58
  Southern, 58, ***59***
  south-western, 58
  western, 58
blowfly, sheep, 184–6
blunt end, 46, **81**
BLUP, 238–9, 240, 241, **246**
bottleneck, population, 134
breed
  comparison, 230–31
  standards, 303
  structure, 249–53
  substitution, 263–4
breeding
  objective, 240, **247**
  programme, captive, 287, **289**
  scheme
    closed nucleus, 249–50, ***251***, **253**
    co-operative, 252, **253**
    group, 252, **253**
    open nucleus, 250–51, **253**
value (A), 219, **227**, **228**
  estimated, 235–42, **246**, **247**, 301
  variance (V_A), 221–2, **228**
  true, 235

Copyrighted material

broad-sense heritability, 221
build, of genome sequence assembly, 75

CAAT box, ***19***, 20
calibration phase, of GWS, 245
callipyge, 101, 202
cancer, and chromosomal aberrations, 112
candidate
  for artificial selection, 221–2, 236–42, **246**
  gene, 96, 298
cap site, 21
capping, of RNA, ***19***
captive breeding programme, 287, **289**
carpet wool, 200
carrier
  of translocation, 107
  of defective allele, 86, 87, 91, 93, 125, 130, 131, 132, **136**, 152, 153, 155, 156, 198, 262, 294–7, 299, 302, **306**
Carwell, 202
cat
  coat colour, 32
  genome sequenced, 75
  karyotype, 2, 4, **39**
  tortoiseshell, 105
cattle
  chromosome, **72**
  comparative map with human, **75**
  genome sequenced, 75
  idiogram, **6**
  karyotype, 4, **5**
  RH maps, 72
  transgenic, 78
cDNA, 51–3, 57, 58, 60, 64, **81**, 85, 97
  library, 51, **52**, 53, **81**, 274
centimorgan, *see* cM
central test, 244
centric
  fission, 111, 119
  fusion translocation, 110, ***110***, **111**, **118**, 120
centromere, 1, 7, **8**, ***10***, 11, ***12***, **37**
chain-terminating sequencing, 53, ***54***
checkerboard, 9, 25, 26, 27, ***30***, ***31***
chiasma(ta), 7, **8**
chicken
  sexing, 202–3, ***203***
  genome sequenced, 75
  karyotype, 4, ***41***
chimeric, 114
chimerism, XX/XY, 114
chip, 64
  expression, 53, 60, **82**, 227
  SNP, 69, 70, **82**, **83**, 96, **97**, 136, 147, 156, 191, 204, 226, 244, 245, 280, **306**
chromatid, 1, 7, **8**, ***10***, 11, ***12***, 28, **37**
  non-sister, 28, **29**
  sister, 28
chromosomal
  aberrations, 103–20
    aneuploidy, 104, 119
    deletion, 111–2, 116, **118**, 119
    insertion, 120
    list of, 119–20
    monosomy, 103
    trisomy, 103, 119
  intersex, 117, **118**
chromosome painting, 72
chromosome, 1–4, **37**
  acrocentric, 3, 4
  banding, 3–5
  metacentric, 3, 4
  p arm, 3
  painting, 72, **72**
  q arm, 3
  sub-metacentric, 3
cis-acting, 202
citrullinaemia, in cattle, 85
cleavage site, 46, 47, ***48***
clinical screening, 294
cloning
  animals, 275–6, **279**
  DNA, 49–51, **81**
  embryo, 277–8, **277**
closed-nucleus breeding scheme, 249–50, ***251***, **253**
clues, to a candidate's breeding value, 236–8, **246**
cM, 31, **32**, 74, 163
CNV, 69–70, 198–9, 227
coancestry, 210
coat colour, 195–200
  table of genes, 199
  tortoiseshell, 32, ***33***, 34
coding
  region, of DNA, 16
  sequence, of DNA, 16
  strand, of DNA, 17, ***18***, ***19***, 20
co-dominant, 87, 90
  disorder, **97**
codon, of mRNA, 17
coefficient of selection, 127
collateral relationship, 209, 210, **214**
commercial layer, in breed structure, 249, ***250***, ***251***, **253**
common ancestor, 207
comparative
  gene mapping, 73, **75**, 96
  positional candidate gene, 96
comparison
  between breeds, 230–31
  of crosses, 258–60
  direct, 231
  indirect, 231
complementarity, 255, **264**, 268
complementary DNA, *see* cDNA
complex segregation analysis, 147
composite population, 260–61, **265**, 268
concordant, 71
conjugation, and resistance to antiobiotics, 187
consensus sequence, 20, 23
conservation
  genetics, 280–90
  of genetic diversity, 286–8, **289**
  of synteny, 73, **83**
contig, 73, 74, 75
continuous variation, 216–29
control of inherited disorders, 291–306
co-operative breeding scheme, 252, **253**
copy DNA, *see* cDNA
copy number
  polymorphism (CNP), 69–70
  variation, *see* CNV
corrective surgery, impact of, 293
correlated
  response to selection, 239–40, **246**
  traits, 239–40
correlation
  genetic (r_G), 224, 226, **228**, **246**
  genetic, and G × E, 233, **233**
  phenotypic (r_P), 224, 226, **228**
cos ends, 49
cosmid, 49
CpG island, 35
criteria, selection, 235, 241
cross-bred, 254
crosses, comparison of, 258–60
crossing, 254–66, 304
  backcross, 256, 259, 260, **265**
  four-way, 257, 259, 260, **265**, 272, 277, **278**

Copyrighted material

crossing (*continued*)
to produce a synthetic, 260–61, **265**
regular, 254–60, **264**, **265**, 304
rotational, 257–8, 259, 260, **265**
specific, 255–7, **265**
three-way, 256, 259, 260, **265**, 272, 277, **278**
two-way, 255, 259, 260, **265**, 272, 277, **278**
crossing-over, ***29***, ***30***, ***31***, **37**
in meiosis, 7
cross-over, 28, 29, 71
cross-resistance, 185
culling, 294
levels, independent, 242
cycle sequencing, 58
cytogenetics, 1–4, 103–20
cytosine, 12, ***13***, ***14***

dam
family, 236
line, 255, 268
deletion
mutation, 23, 24, 62, 64, 89, 90, 95
chromosomal, 111–2, 116, **118**, 119
denaturation
gradient gel electrophoresis, *see* DGGE
in PCR, 56, ***56***, 57
deoxynucleotide, ***56***
deoxyribose, 11
depression, inbreeding, 212–4, **214**, 242, 267
deviation, 219
DGGE, 62
diagnosis, veterinary, 63–4
Dicer fragment, 81, **83**
dideoxy sequencing, 53, ***54***
digestion
of DNA, 47, ***48***, ***50***, ***59***, ***61***, **81**
partial, 51, 52
diploid, 3, **37**
direct
comparison, 231
relationship, 208–9, **214**
repeat, 36
disequilibrium, linkage, *see* LD
disjunction, 7, ***8***, ***10***, 11, ***12***, 108, 111
disorders, single-locus, genetic
control of, 294–99
diversity
allelic, 281, **288**
antibody, 160, 170
DNA cloning, 49–51, ***50***, **81**
DNA, 11–16, **37**
expression chip, 58, 60
fingerprint, 64–8, ***66***, 67, 204, ***205***
gene, 164
length in a diploid cell, 15
ligase, 15, 49, ***50***
markers, in control of single-locus disorders, 297–8, **306**
polymerase, 14, 51, 52, ***54***, ***56***, 82
repetitive, 36, **37**
replication, 14, ***15***
sequencing, 53–5
single-copy, 35
types of, 35–6
unique, 35
dog
genome sequenced, 75
karyotype, 4, ***40***
Dolly, 275, 287
dominance
deviation (D), 219, **227**
variance (V_D), 221–2, **228**
incomplete, 87
dominant, 86, 90
disorder, **97**
gene, selection against, 127–8
donkey, karyotype, 4
donor site, 21
dosage compensation, 32, 34, **37**
dot blot, 63
double
helix, 14, ***14***, **37**
muscling, 201–2
double-stranded RNA, *see* dsRNA
drift, genetic, 123, 124, 133–4, 135, **137**, 242–3, 244, **247**, 285, 286
drug metabolism, 172, **177**
dsRNA, 81, **83**
dwarf poultry, 202

EBV, 235–42, 244, 245, **246**, **247**, 272, 274, 276, 301–2, 304, **306**
economic value, net (v), 241
effective
number of alleles, 281, **288**
population size (N_e), 243, **247**, 267, 286
embryo cloning, 277–8, **277**
embryonic
loss due to chromosomal aberration, 106
stem cell, 78
enhancer, 19, ***19***, 22
environmental
control of inherited disorders, 291–4
deviation (E), 218, **227**
variance (V_E), 221–2, **228**
epigenetics, 35, **37**, 202
epistasis, 199, 219–20
epistatic
deviation (I), 220, **227**
variance (V_I), 221–2, **228**
equilibrium
gene frequency, selection/mutation balance, 128, 130
gametic, 135
linkage, 135
EST, 53
estimated breeding value, *see* EBV
eukaryote, 16
eumelanin, 195
euploidy, 104
evolution of karyotypes, 113
exon, ***19***, 21
expected heterozygosity, 281, **288**
expressed sequence tag, *see* EST
expression
chip (microarray), 53, 58, 60, **82**, 227
library, 76
QTL, 226
vector, 76
expressivity, variable, 145–7
ex-situ conservation programme, 287
extension, in PCR, 56, ***56***, 57

5′
cap, 21
end of a DNA segment, 12, ***13***, 14, ***15***, 16, ***18***, 18, ***19***, 21, 36
F1 (first filial), 255
generation, 70
F2 generation, 70
facial eczema, in sheep, 176–7
FAO, 189, 232, 280, 286, 287
FACS, 71
fingerprint, DNA, 64–8, ***66***, 67, 204, ***205***
FISH, 72, **72**
fish, transgenic, 78
fission, centric, 111, 119
fitness, 127
relative, 127
fixation of a gene in a population, 134

Copyrighted material

fluorescence in situ hybridization, *see* FISH
fluorescence-activated cell sorting, *see* FACS
Food and Agriculture Organization of the United Nations, *see* FAO
founder effect, 134
four-way cross, 257, 259, 260, **265**, 272, 277, **278**
fragile site, 112, 119
frameshift mutation, 24, ***94***, 95
freemartin, 114–5, **118**
frequency distribution, ***217***, ***218***
fundamental number, of chromosomes, 113
fusion, centric, 110

G × E, 232–3, ***232***, **233**
gametic
  disequilibrium, *see* LD
  equilibrium, 135
GC box, ***19***, 20
GenBank, 55
gene, 18–23, ***19***, 24–5
  cloning, 49
  construct, 77
  frequency, 121–2
  jumping, 36
  knockout, 79
  mapping, 70–74
  mutation, 23
  regulation, 22–3, 81
  replacement, targeted, 78
  split, 20–22
  structural, 35, **37**
  therapy, 78, 300
generation interval (L), 243, **247**, 286
genes
  number of, 27, 35, 70, 76, 92, 130, 302
  size of, 21
genetic
  anticipation, 95, 155
  code, 15–16, 17, 18
  control
    of multifactorial disorders, 300–302, **306**
    of single-locus disorders, 294–99
correlation (r_G), 224, 226, **228**, **246**
  and G × E, 233, **233**
distance among populations, 282, **288**
diversity
  conservation of, 286–8, **289**
  loss of, 284–6
  measurement of, 280–83
drift, 123, 124, 133–4, 135, **137**, 242–3, 244, **247**, 285, 286
effect of non-genetic control of inherited disorders, 293–4, **305**
  genomics, 226
  heterogeneity of disease, 87–90, **97**, 155
  trend, 239
  variance (V_G), 221
genome, 3
  assembly, 35, 74–6
genome-wide
  association (GWA) analysis, see GWA
  scan, **228**
  selection, *see* GWS
genomic
  library, 51, 52, **81**
  selection; see GWS
genotype, 87
  frequency, 121–2
genotype-environment interaction, *see* G × E
genotypic
  value (G), 218, **227**
  variance (V_G), 221
germ-line mutation, 24
Global Animal Genetic Resources Programme, of FAO, 287
goat
  karyotype, 4, ***43***
  transgenic, 78
gonadal intersex, 117, **118**
grading-up, 261–4, **265**, 304
group breeding scheme, 252, **253**
growth factor, 196
GT-AC rule, 21
guanine, 12, ***13***, ***14***
GWA, 147, 245
GWS, 46, **228**, 244–6, **247**
  GWA phase, 245
  prediction phase, 245

haemophilia in dogs, 93
halothane, adverse reaction to, 173, **177**
haploid, 7, **37**
haploidy, 106, **118**
haplotype, 135–6, 245
  MHC, inheritance of, 167, ***167***
Hardy-Weinberg
  frequencies, 123–6, **136**
  law, 123–6, **136**
    and multiple alleles, 125
    and recessive genes, 124–5
    and X-linked genes, 125–6
hemizygous, 125
heritability, 140–41, 143, 145, 177, 221–4, 238, 281, 284
  broad sense, 143, 221
  estimation of, 223–4, **228**
  of liability, 146, 180
  narrow sense, 143, 145, 221, **228**
  realized, 223
  table of estimates, 225
hermaphrodite, 116
  true, 116
heterogametic sex, 9
heterogeneity of disease, genetic, 87–90, **97**, 155
heterosis, 254–60, **264**, 267, 268
  individual, 254
  maternal, 254
  parental, 254
  paternal, 254
heterozygosity
  expected, 281, **288**
  observed, 280, **288**
heterozygote, ***61***
  double, 28, ***30***, ***31***
  selection against, 132–3, 162
  selection favouring, 131–2, **137**, 168, 175, **177**
  for translocation, 107, 109
heterozygous, 25
high-resolution melt (HRM) analysis, 63
hip dysplasia in dogs, 139–40, ***140***, 300–301
  control of, 291–2
histocompatibility, 163
histoglobulin, 163–4, ***163***, 170
histone, 11
homeo domain, 23
homeobox, 23
homeotic gene, 23
homogametic sex, 9
homologue, 3, ***8***, 28, ***29***, **37**
homozygosity mapping, 96
homozygote, ***61***
  double, 28, ***30***, ***31***
  for translocation, 109
homozygous, 25
horns, absence of, 201

Copyrighted material

horse
genome sequenced, 75
karyotype, 4, ***45***
host-pathogen interaction, 179–83
hosts, resistance in, 183–4
HRM analysis, 63
in MHC genotyping, 166
hybrid vigour; see heterosis
hybridization, interspecific, 113

identical by descent, 207
idiogram, 4
cattle, *6*
immunogenetics, 158–71
immunoglobulin, 159–60
imprinting, 34–5, **37**, 155–6, 202
improvement lag, 250, ***251***, **253**
in vitro
fertilization (IVF) of ova, 272–3, **278**
maturation (IVM) of ova, 272–3, **278**
inactivation, 32, 35, **37**
inborn error of metabolism, 85–7, **97**
inbred line, of cattle, 214
inbreeding, 242–3, 267, 270, 271, 272, **278**
coefficient, 207–8, 209–12, **214**
depression, 212–4, **214**, 242, 267
in populations, 212, **214**
incomplete
dominance, 87
penetrance, 145–7
indel, 24
independent culling levels, 242
index selection, 237, 239, 242, **246**
indirect
comparison, 231
selection, 239
individual
heterosis, 254
selection, 236
information nucleus, 252
inheritance, 25, 26, 151–3
autosomal
dominant, 86, 90, 151–2
recessive, 86, 90, 152
co-dominant, 87, **97**
dominant, **97**
incompletely dominant, 87
multifactorial, **148**, 175–6, ***176***, ***177***, **177**
polar over-dominance, 202
recessive, **97**
sex-limited, 87, 155
X-linked dominant, 152–3
X-linked recessive, 153
inherited disorders
control of, 291–306
paradox of, 292–3, **305**
INSDC, 55
insecticide, resistance to, 184–6
insert DNA, 49, **81**
insertion mutation, 24, 36, 62, 64, 95, 116, 203
in-situ conservation programme, 287
insurance scheme, 302
integrated map, 74, **83**
Interbull, 233, 244
interference, RNA, 81, **83**
International Nucleotide Sequence Database Collaboration, *see* INSDC
International Union for the Conservation of Nature, *see* IUCN
intersex, **118**
chromosomal, 117, **118**
classification of, 116–7
gonadal, 117, **118**
phenotypic, 117, **118**
interspecific hybridization, 113
interval mapping, 224
introgression, 261
marker-assisted, 263
intron, ***19***, 21
inversion, 75, 111–2, **118**
paracentric, 111, 120
pericentric, 111, 120
inverted repeat, 36
isochromosome, 112, **118**, 120
isoerythrolysis, neonatal, 161–2
IUCN, 280
IVF, 272–3, **278**
IVM, 272–3, **278**

jumping gene, 36

karyotype, 1, 2, 4, 5, 11, 37, ***39–45***
cat, ***39***
chicken, ***41***
dog, ***40***
evolution of, 113
goat, ***43***
horse, ***45***
pig, ***42***
sheep, ***44***
kilobase (kb), 14
knockout, gene, 79

lag, improvement, 250, ***251***, **253**
LD, 135–6, **137**, 245–6, 299, **306**
leader sequence, ***19***, 20
leucine zipper, 22
liability, 139–43, ***141***, **142**, **148**, 179–80, 301
heritability of, 146, 180
library
BAC, 73
cDNA, 51, 53, 52, **81**
genomic, 51, **52**, **81**
plasmid, **52**
ligand, 196
ligation, 49
LINE, 64
line breeding, 209, 213
linkage, 27–32, **37**
disequilibrium, *see* LD
equilibrium, 135
group, **83**
map, 32, 70, 74, **83**
live recombinant vaccine, 77
llama, karyotype, 4
location database (LDB), 73
loci, 24–5
locus, 24–5, **37**
long interspersed element, *see* LINE
loss
of genes from a population, 134
of genetic diversity, 284–6
Lyon hypothesis, 34
lysosomal storage disease, 91

macrochromosome, in birds, 3
MAF, 122
major histocompatibility complex, *see* MHC
malignant hyperthermia syndrome, see MHS
map
BAC fingerprint, 73
comparative, **75**
distance, 31, **32**, 32
integrated, 74, **83**
linkage, 70, 74, **83**
physical, 70, 74, **83**
mapping
comparative, 73
function, ***32***, 32
gene, 70–74
interval, 224
marker-assisted
introgression, 263
selection (MAS), 244

Copyrighted material

Markov chain Monte Carlo (MCMC), 156
marsupial, 34
MAS, 244
mass selection, 236
massively-parallel sequencing, 55, **82**
maternal heterosis, 254
mean, 216, ***217***
megabase (Mb), 14
meiosis, 7–11, 29, ***30–31***, **37**, 103–4, ***104***
  in female, ***8***
  in male, ***10***
melanocyte, 195
melanosome, 195
Mendelian inheritance, 25–7, **37**, 151–3
messenger RNA, *see* mRNA
metacentric chromosome, 3
metallothionein promoter, 79
metaphase, in mitosis, 11
methylation, 35
metric traits, *see* quantitative traits
MHC, 163–9, ***163***, 170, 284
  association with disease, 168, 184
  and mate choice, 169
  restriction, 165
MHS, 173–4, **177**, **298**
microarray, 69
  expression, **82**
microchromosome, in birds, 3
microcloning, of single chromosomes, 72
microdissection, of chromosomes, 72
microRNA (miRNA), 81, **83**
microsatellite, 65, 67, ***68***, **82**
  in pedigree checking, 204, ***205***
migration, 123
minisatellite, 65
  in pedigree checking, 204, ***205***
minor allele frequency, *see* MAF
miRNA, 81, **83**
mis-sense mutation, 24, 92, 167
mitosis, 11, ***12***, **37**
mixoploidy, 106
MOET, 271–2, **278**
monosomy, 103, 106, **117**, **118**
mosaic, 34, 105
mRNA, 16, 17, ***18***, **37**, 51, ***52***
multifactorial
  disorders, genetic control of, 300–302, **306**
  inheritance, 139–43, ***141***, ***142***, **148**, 175–6, ***176***, 177, ***177***
multigene family, 35
multiple
  allele, 25
  ovulation and embryo transfer, *see* MOET
multiplex PCR, 57
multiplier, in breed structure, 249, ***250***, ***251***, **253**
muscular hypertrophy, 201–2
mutation, 23–4, **37**, 123
  deletion, 23, 24, 62, 64, 89, 90, 95
  frameshift, 24, ***94***, 95
  gene, 23
  germ-line, 24
  insertion, 24, 36, 62, 64, 95, 116, 203
  mis-sense, 24, 92, 167
  nonsense, 23, 92
  point, 23
  silent, 24, 167
  somatic, 24, 160
myxomatosis in rabbits, 179–80

narrow-sense heritability, 221, **228**
National Scrapie Plan for Great Britain, 181
natural selection, 127, 180
ncRNA, 21
N_e, 243, **247**, 267
neonatal isoerythrolysis (NI), 161–2
net economic value (v), 241
next-generation sequencing, 55, 75, **82**
nombre fondamental (NF), 113
non-additive
  genetic deviation, 220
  genetic variance (V_{NA}), 221
non-disjunction, 103, 104, ***104***
non-genetic
  deviation (E), 218, **227**
  variance (V_E), 221, **228**
non-orange, coat colour, 32, ***33***
non-protein-coding RNA, 21
non-recombinant, ***29***
  chromosome, ***8***
  gamete, 28, 29
nonsense mutation, 23, 92
non-sister chromatid, 28, ***29***
non-synonymous base substitution, 167
NOR, 22
normal equations, 237
northern blotting, 58
nucleolar organizer region (NOR), 22
nucleolus, 22, ***33***
nucleotide, 11, 12, ***13***
nucleus
  in breed structure, 249, ***250***, ***251***, **253**
  information, 252
numerator relationship, 211

objectives, breeding, 240
observed heterozygosity, 280, **288**
Okazaki fragment, 15
oligonucleotide, 56
OMIA, 98
oncogene, 196
online Mendelian inheritance in animals, *see* OMIA
open reading frame, 16, 75
open-nucleus breeding scheme, 250–51, **253**
opossum, genome sequenced, 75
orange, coat colour, 32, ***33***
overall merit, selection for, 242
Oxford grid, 75

p arm, of chromosome, 3
painting, chromosome, 72, ***72***
palindrome, 36, 46
paracentric inversion, 111, 120
paradox of inherited disorders, 292–3, **305**
parasites
  control of, 188–90
  resistance in, 184–6
parental heterosis, 254
partial digestion, 51, ***52***
paternal heterosis, 254
paternity test, 204, ***205***
pathogens
  control of, 190
  resistance in, 186
PCR, 55–8, **56**, **82**
  amplification cycle, 57
  annealing phase, 56, ***56***, 57
  denaturaion phase, 56, ***56***, 57
  in diagnosis, 63–4
  extension phase, 56, ***56***, 57
  in MHC genotyping, 166
  multiplex, 57
  product, 57
  real-time quantitative (Q-PCR), 57
  reverse transcription (RT-PCR), 57

Copyrighted material

pedigree, 153–5, ***154***
  analysis, 295, **306**
  checking, 204–5, ***205***, **206**
  selection, 236
  software, 155
penetrance, incomplete, 145–7
peptide, 16
  synthesis, ***18***
peptide-binding region (PBR), of MHC, 165, 167–8
performance testing, 236
pericentric inversion, 111, 120
phaeomelanin, 195
phage lambda, 48, ***48***, 49
pharmacogenetics, 172–8
phenocopy, 91–2, **97**, 155
phenotype, 87
phenotypic
  correlation (r_P), 224, 226, **228**
  intersex, 117, **118**
  value (P), 218, **227**
  variance (V_P), 221–2, **228**
phylogenetic tree, 283, ***283***, **289**
physical map, 70, 74, **83**
pig
  genome sequenced, 75
  karyotype, 4, ***42***
  transgenic, 78
plasmid, 49, ***50***
  library, ***52***
  recombinant, 80
  and resistance to antibiotics, 187
platypus, genome sequenced, 75
pleiotropy, 132
point mutation, 23
polar body in female meiosis,
  first, 7, ***8***
  second, ***8***, 9
polar over-dominance, 202
polledness, 201
poly-A tail, 21
polyadenylation
  of RNA, ***19***
  signal, 21
polygene, 224
polymerase chain reaction, *see* PCR
polymorphic, 57
polymorphism, 62, 69, **137**
  extreme in MHC, 166–7, **170**
polypeptide
  from cloned DNA, 76–7
  synthesis, 16–17, ***18***
polyploidy, 106, **118**, 120
  induced in fish, 107
population size, effective, *see* N_e
prediction phase, of GWS, 245
primer, for PCR, 56
prion, 180–82
proband, 154, ***154***
probe
  chip, 60
  in Southern blotting, 58, **59**
processed pseudogene, 164
progeny
  difference, 219
  testing, 236
prokaryote, 16
prolificacy in sheep, 200–201
promoter, ***19***, 20, 22, 79
pronucleus, 77, **78**
proportion of loci that are polymorphic, 281
protein, 15–16
  recombinant, 273–4
  synthesis, 16–18
proto-oncogene, 196
proximal end, of chromosome, 74
pseudo-autosomal region, 9, ***10***
pseudogene, 164
  processed, 164
pseudohermaphrodite, 116
  female, 116
  male, 116
Punnett square, 9, 25, 27, ***30***, ***31***
pure-bred, 254
purine, 12, ***13***, ***14***
pyramid, breed, 249, ***250***, **253**
pyrimidine, 12, ***13***, ***14***
pyrosequencing, 53

q arm, of chromosome, 3
QTL, 147, 191, 224–7, **228**
  expression, 226
quadrivalent, in meiosis, 108–9, ***109***
quantitative
  RT-PCR, 57
  trait locus, *see* QTL
  trait nucleotide (QTN), 226
  traits, 216–29
  variation, 216–29

R factor, 187
rabbit
  genome sequenced, 75
  karyotype, 4
radiation hybrid (RH), 71, 72
random
  amplified polymorphic DNA, *see* RAPD
  mating, 122, **136**
RAPD, 57
Rare Breeds Survival Trust, 286
realized heritability, 223
real-time
  quantitative PCR (Q-PCR), 57
  RT-PCR, 57
recessive, 86, 90
  disorder, **97**
  gene, selection against, 129
reciprocal translocation, 107–9, **118**, 120
recognition sequence, 46, 47
recombinant, 28, **29**
  chromosome, ***8***
  DNA, 49–51
  gamete, 28, 29
  plasmid, ***50***, 80
  protein, 273–4
  vaccine, 274
recombination, 29, **37**, 71
  fraction, 28, 29, ***30***, 31, ***31***, 32, 32, **83**, 135
  hotspot, 65
  in meiosis, 7
recurrence risk, 147–8
  empirical, **148**
  theoretical, **148**
red-cell antigen, 170
redundancy, in genetic code, 16, 24
reference sire, 244
regular crossing, 254–60, **264**, **265**, 304
  and selection, 267–9
regulation, gene, 81
regulatory protein, **37**
relationship, 208–9, 210–11, **214**
  additive, 211, 238
  collateral, 209, **214**
  direct, 208–9, **214**
  numerator, 211
relative fitness, 127
relatives, resemblance between, 223
REML, 223, 224
repeat
  direct, 36
  inverted, 36
  unit, 36
  unstable, 155
repeatability, 236, 238
repetitive DNA, 36, **37**
replication, of DNA, 14, ***15***
repressor, 22, **37**
resemblance between relatives, 223
resistance
  to anthelmintics, 186
  to antibiotics, 186–7
  in hosts, 183–4

Copyrighted material

selection for, 190–91
to insecticides, 184–6
to Marek's disease in chickens, 183–4
in parasites and pathogens, 184–7
to scours in pigs, 183
resource family, for linkage mapping, 70
response
element, 23
to selection (R), 221–3, ***222***, **228**, 284
restricted maximum likelihood, *see* REML
restriction
enzyme, 46, 47, 49, **81**
fragment length polymorphism, *see* RFLP
map, ***48***, 49
site, ***61***
retrovirus, 51
reversal, of sex, 102, 106, 120
reverse
dot blot, 63
transcriptase, 51, **52**, **81**
transcription, 51
PCR (RT-PCR), 57
RFLP, ***61***, 62, **82**
in MHC genotyping, 166
RH map, 72
ribonucleic acid, 16
ribosomal RNA, 21–2
ribosome, 16, ***18***, 20
RISC, 81, **83**
RNA, 16
antisense, **83**
double-stranded, 81, **83**
interference (RNAi), 81, **83**
polymerase, 16, ***18***, 20
small interfering (siRNA), 81, **83**
splicing, ***19***, 21
RNA-induced silencing complex, *see* RISC
Robertsonian translocation, 110, ***110***, ***111***, **118**, 120
rotational crossing, 257–8, 259, 260, **265**
rRNA, 21–2
RT-PCR, 64

Sanger sequencing, 53, ***54***
satellite DNA, 65
scrapie, 180–82
screening
biochemical, 297, **306**
clinical, 294
screw-worm fly, 188–9
segregation analysis, 155–6
complex, 147, 156
segregation, 25, 27
independent, 26, **37**
ratio, 25, 26, 27
selection, 123, 127–33, **136**
accuracy of (r_{AC}), 236
against a recessive gene, 129
against heterozygotes, 132–3, 162
artificial, 127, 230
between populations, 230–34
coefficient, 127
criteria, 235, 241
differential (S), 222, **228**, 284
against a dominant gene, 127–8
favouring heterozygotes, 131–2, **137**, 168, 175, **177**
index, 237, 239, 242, **246**
indirect, 239
individual, 236
mass, 236
natural, 127, 180
pedigree, 236
and regular crossing, 267–9
for resistance in hosts, 190–91
response to (R), 221–3, ***222***, **228**
sib, 236
within populations, 235–48
selection/mutation balance, **136–7**
for a dominant gene, 128
for a recessive gene, 129–31
sense strand, of DNA, 20
sequence
databases, 55
read, 55
sequence-tagged site (STS), 67
sequencing
cycle, 58
DNA, 53–5
massively-parallel, 55, **82**
next-generation, 55, 75, **82**
pyrosequencing, 53
by synthesis, 55, **81**
sex
chromatin, 33
chromosome, 3, 7
non-disjunction of, 103–4, ***104***
determination, 3, 115–6
ratio, control of, 273
reversal, 102, 106, 120
sexing chickens, 202–3, ***203***
sex-limited inheritance, 87, 155
sex-linkage, 27
sex-linked, 27
sheep
blowfly, 184–6
karyotype, 4, ***44***
transgenic, 78, **78**
short interspersed element, *see* SINE
sib selection, 236
signal transduction pathway, 196
silent mutation, 24, 167
SINE, 64, 65
single nucleotide polymorphism, *see* SNP
single-copy DNA, 35
single-gene
disorders, investigating with a SNP chip, 96
traits, list of, 98–102
single-sperm typing, 71
single-stranded conformational polymorphism, *see* SSCP
sire
family, 236
line, 255, 268
sire-reference scheme, 243–4
SIRM, 188–9, **192**
siRNA, 81, **83**
sister chromatid, 28
small interfering RNA, *see* siRNA
SNP, 68–9, 70, 71, **82**, 122, 131, **206**
chip, 69, 70, **82**, **83**, 136, 147, 191, 204, 280, **306**
in GWS, 244–5
in investigating single-gene disorders, 96, **97**
somatic mutation, 24
somatic-cell hybrid, 71
Southern
analysis, 58, **59**, 64, ***66***, ***205***
blotting, 58, **59**
south-western blotting, 58
specific crossing, 255–7, **265**
splice site, 21
splicing
in recombinant DNA, 49
of RNA, ***19***, 21
split genes, 20–22
spongiform encephalopathy, 180–82
sporadic case, 156
SSCP, 62
start triplet, ***19***, 20
in genetic code, 16
stem cell, embryonic, 78
sterile insect release method, *see* SIRM

Copyrighted material

sticky end, 47, **81**
stop triplet, ***19***, 20
  in genetic code, 16
straight-bred, 254
structural gene, 18, ***19***, 20, 35, **37**
structure, of breeds, 249–53
sub-metacentric chromosome, 3
synapsis, 7, ***8***, 9, ***10***, 28
synaptonemal complex, 7
synonymous base substitution, 167
syntenic group, 71
synteny, conservation of, 73, **83**
synthetic population, 260–61, **265**, 268

3′ end of a DNA segment, 12, ***13***, 14, ***15***, 16, ***18***, 18, ***19***, 21, 36
2*n* number of chromosomes, 3, 4
tandemly-repeated DNA, 22, 64–8, 205
tandem translocation, 110, 120
Taq polymerase, ***56***, 57
TATA box, ***19***, 20
temperature gradient gel electrophoresis, *see* TGGE
template
  DNA (for PCR), 56, ***56***
  strand, 16, ***18***, ***19***
test matings, 295–7, **306**
tester population, 231
tetraploidy, 106, 107, 120
TGE, 36, **37**, 65, 187
TGGE, 62
TGRM™, 243, 286
thermal cycler, 57
three-way cross, 256, 259, 260, **265**, 272, 277, **278**
threshold, 139–43, ***141***, ***142***
thresholds, multiple, 141–3, ***142***
thymine, 12, ***13***, ***14***, 16
tissue rejection, 163, 164
tissue typing, 166
tortoisehell coat colour, 32, ***33***, 34, 105, 126, 195–6
Total Genetic Resource Management™, *see* TGRM™
trailer sequence, ***19***, 20
transcription, 16, **18**, 20, **37**
  factor, 20
  initiation site, ***19***, 20
  termination site, ***19***, 20
transcriptome, 60
transduction, and resistance to antibiotics, 187
transfer RNA, 16
transformation, ***50***
  and resistance to antibiotics, 187
transfusion, blood, 161
transgene, 77
transgenesis, 77–80, ***78***, **83**, 191, 274–5, **279**
transgenic
  cattle, 78
  fish, 78
  goat, 78
  pig, 78
  sheep, 78, ***78***
translation, 17, **37**
translocation, **118**
  centric fusion, 110, ***110***, ***111***, **118**, 120
  heterozygote, 107, 109
  homozygote, 109
  reciprocal, 107–9, **118**, 120
  Robertsonian, 110, ***110***, ***111***, **118**, 120
  tandem, 110, 120
transplantation, 293
  kidney, in dogs, 164, ***165***
transposable genetic element, *see* TGE
transposon, and resistance to antibiotics, 187
trend, genetic, 239
trinucleotide repeat, unstable, 95
triplet code, of DNA, 16, 17
triploidy, 106, 107, 120
trisomy, 103, **117**, **118**, 119
  autosomal, 105–6, 119
trivalent, in meiosis, 111, ***111***
tRNA, 16
true breeding value, 235
trypanosomiasis, African, 182–3
two-way cross, 255, 259, 260, **265**, 272, 277, **278**
type of gene action, 86–7, 90–91, **97**

unstable trinucleotide repeat, 95
untranslated region, *see* UTR
uracil, 16
urea cycle, ***86***
UTR, 20

vaccine
  live recombinant, 77
  recombinant, 274
variable
  expressivity, 146
  number of tandem repeats, *see* VNTR
variance, 216, ***217***, 220
  in breeding values (V_A), 221–2, **228**, 268, 281, 284
  in dominance deviations (V_D), 221–2, **228**
  in epistatic deviations (V_I), 221–2, **228**
  genetic (V_G), 220, 221
  in non-genetic (environmental) deviations (V_E), 221–2, **228**
  in phenotypic values (V_P), 221–2, **228**
variation, 220–21
  in base sequence, 60–63
  between breeds, 150–51
  within breeds, 150–51
  quantitative, 216–29
vascular anastomosis, 114, 115
vector, in DNA cloning, 49, **81**
VNTR, 64–8, ***66***, 204

W chromosome, in birds, 3
warfarin, resistance to, 174–5, **177**
western blotting, 58
wild-type, 32
World Watch List, of FAO, 280

X chromosome, 3, 33, 34
X-inactivation, 34, **37**, 155
X-linkage, *see* X-linked
X-linked, 27, 32, 34, 73, 105, 125–6, 151, 201
  disorders, 90–91, 93, 101, 102, 116, 299
  dominant inheritance, 152–3
  recessive inheritance, **136**, 153
XX male, 116, 120
XX/XY chimerism, 114
XY female, 102, 116, 120

Y chromosome, 3
YAC, 51
yeast artificial chromosome, *see* YAC
Y-linked, 27

Z chromosome, in birds, 3
Z-inactivation, 34
zinc finger, 22, 23, 76
zoo-FISH, ***72***, 73
zygote, 7

Copyrighted material