

GENETIC

ANALYSIS

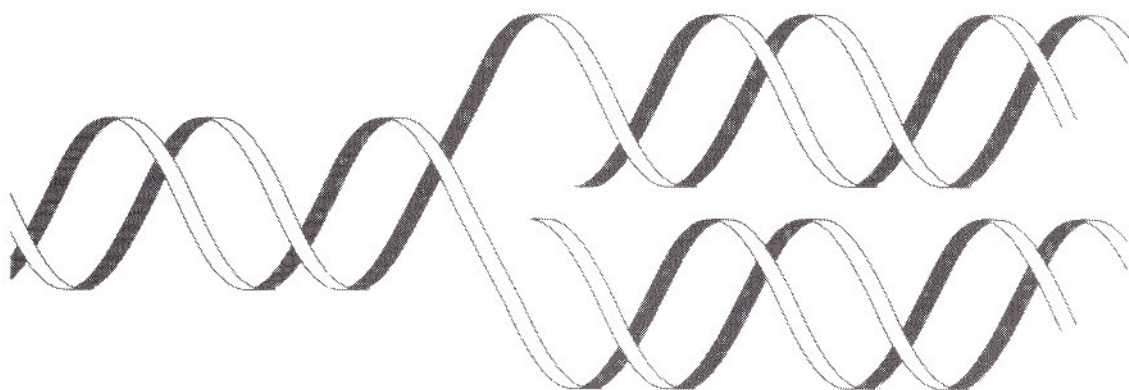
Principles, Scope and Objectives

JOHN R. S. FINCHAM

PhD, ScD, FRS, FRSE

Division of Biological Sciences

University of Edinburgh



b

Blackwell
Science

GENETIC ANALYSIS

GENETIC

ANALYSIS

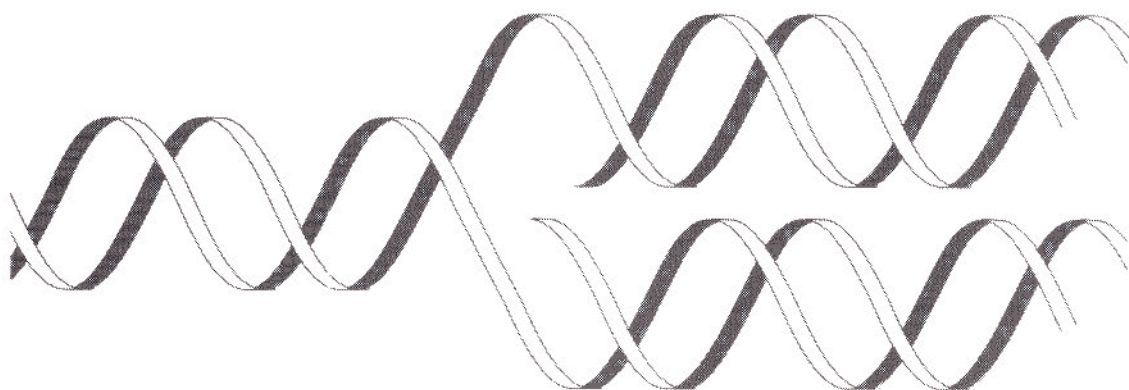
Principles, Scope and Objectives

JOHN R. S. FINCHAM

PhD, ScD, FRS, FRSE

Division of Biological Sciences

University of Edinburgh



b

Blackwell
Science

© 1994 by
Blackwell Science Ltd
Editorial Offices:
Osney Mead, Oxford OX2 0EL
25 John Street, London WC1N 2BL
23 Ainslie Place, Edinburgh EH3 6AJ
238 Main Street, Cambridge
Massachusetts 02142, USA
54 University Street, Carlton
Victoria 3053, Australia

Other Editorial Offices:

Arnette Blackwell SA
1, rue de Lille
75007 Paris
France

Blackwell Wissenschafts-Verlag GmbH
Kurfürstendamm 57
10707 Berlin
Germany

Blackwell MZV
Feldgasse 13
A-1238 Wien
Austria

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 copyright owner.

First published 1994

Set by Excel Typesetters Co., Hong Kong
Printed and bound in Great Britain
at the Alden Press, Oxford and Northampton

DISTRIBUTORS

Marston Book Services Ltd
PO Box 87
Oxford OX2 0DT
(Orders: Tel: 0865 791155
Fax: 0865 791927
Telex: 837515)

USA

Blackwell Science, Inc.
238 Main Street
Cambridge, MA 02142
(Orders: Tel: 800 759-6102
617 876-7000)

Canada

Oxford University Press
70 Wynford Drive
Don Mills
Ontario M3C 1J9
(Orders: Tel: 416 441-2941)

Australia

Blackwell Science Pty Ltd
54 University Street
Carlton, Victoria 3053
(Orders: Tel: 03 347-5552)

A catalogue record for this title
is available from the British Library

ISBN 0-632-03659-1

Library of Congress

Cataloging-in-Publication Data

Fincham, J.R.S.

Genetic analysis/John R.S. Fincham.
p. cm.

Includes bibliographical references
and index.

ISBN 0-632-03659-1

1. Genetics—Technique.

2. Nucleotide sequence.

3. Genome mapping.

I. Title.

[DNLM: 1. Genetics. 2. Genome.

3. Sequence Analysis, DNA.

QH 430 F492ga 1994]

QH441.F56 1993

574.87'322—dc20

Preface, viii

Introduction: the expanding scope of genetics, ix

1 Dissecting the genome using natural genetic systems, 1

The eukaryotic system, 1

Eukaryotes distinguished from prokaryotes, 1

Mitosis and chromosome structure, 4

Alternation of haploid and diploid phases in the sexual cycle, 7

Meiosis and the rules of classical genetics, 9

Meiosis, 9

Single-factor genetic ratios and their explanation, 12

Independent assortment of different allele-pairs, 14

Testing data for fit to theoretical ratios, 17

Linkage groups and the nature of crossovers, 18

Map units and map distance, 24

Placing linked genes in sequence, 24

Distinguishing between independent assortment and distant linkage, 25

Mapping centromeres using tetrads, 25

Sex linkage, 27

Segregation and linkage in human genetics, 28

Assigning linkage groups to chromosomes, 32

Genetic analysis of bacteria and bacteriophage, 34

Three modes of gene transfer in bacteria, 34

Time-of-entry mapping in *E. coli*, 37

Analogy with sexual recombination and segregation, 38

Stable partial diploids – F' plasmids, 40

Bacteriophage lambda – a virus in the chromosome, 41

Bacteriophage T4 – a model recombinational system, 42

Eukaryotic organelle genetics, 42

Exceptions to the rule of equal results from reciprocal crosses, 42

Chloroplast variants, 42

Mitochondrial variants, 43

Summary and perspective, 44

Further reading, 44

References, 44

CONTENTS

2 From mutations to genes, 47

Defining the gene by mutation and complementation, 47

What is it that the markers mark? 47

The collection of mutants, 47

Sorting of mutants into complementation groups, 48

An interim definition of the gene, 55

Mapping within the gene, 55

The detection of recombination within genes, 55

Mapping by reference to flanking markers, 58

Recombination frequency, 58

Deletion mapping, 58

The determination by genes of protein structure, 60

Colinearity of the gene and its encoded protein product, 60

Biochemistry of polypeptide synthesis, 61

Not all genes encode proteins, 64

Conclusions, 64

References, 66

3 The gene as DNA sequence, 67

Characterization of DNA fragments, 67

Restriction endonucleases, 67

Separating and sizing DNA fragments, 67

Probing for specific sequences, 68

Making restriction-site maps, 72

Cloning and cloning vectors, 72

Making recombinant DNA molecules, 72

Escherichia coli plasmid vectors, 72

Yeast 2- μ m plasmid and shuttle vectors, 76

Lambda (λ) bacteriophage and cosmid vectors, 76

Yeast artificial chromosomes (YACs), 79

Screening DNA libraries for functional genes, 82

Genomic libraries and cDNA libraries, 82

- Gene identification by complementation of mutants, 84
- Screening gene libraries with DNA probes, 85
- Designing gene-specific probes, 86
- Use of cDNA probes, 87
- Screening for gene expression, 87
- Probing across species or group boundaries, 88
- Positional cloning, 88
- Confirming the identity of cloned sequences, 88
 - DNA sequence and open reading frames, 88
 - Using the clone to disrupt the corresponding gene, 88
 - Hybridization back to chromosomes, 90
- Simultaneous mutagenesis and gene tagging, 91
 - The principle, 91
 - Transposon tags, 91
- Conclusion, 92
- References, 92

4 The evolving concept of the gene, 95

- The gene as a *cis*-acting unit, 95
- The gene as a unit of translation, 95
 - Single genes – single enzymic functions, 95
 - Single genes – multiple enzymic functions, 98
 - Single genes – polyproteins, 100
 - Single transcripts – multiple proteins – polar effects on translation, 101
- The gene as a unit of transcription, 102
 - Genomic DNA compared to cDNA – the detection of introns, 102
 - Comparison of genomic and cDNA restriction-site maps, 103
 - Introns defined by sequence analysis, 103
 - Defining the end-points of transcription, 106
- Processing of the primary transcript, 106
 - RNA polymerases – promoters and processing of transcripts, 106
 - Different products from the same primary transcript, 110
 - Genes nested within other genes' introns, 113
 - Retroviral genome expression – units of all kinds! 115
- The complex functions of promoters, 116
 - The *E. coli* paradigm, 116
 - Analysis of eukaryotic promoter regions by mutation, 117
- Enhancers – the expanding frontiers of the gene, 122
 - The ADH genes of *Drosophila*, 122
 - How do genes interact with their enhancers? The case of the β -globin locus control region, 123

- Locus-specific *trans*-acting control – transvection, 125
- Restructured genes and edited messages, 126
- Conclusion – what is the gene? 127
- References, 127

5 Analysis of the whole genome, 129

- Introduction, 129
- Total DNA – quantities and patterns of complexity, 129
 - C-values, 129
 - Degrees of repetition among genomic sequences, 129
- The nature of repetitive DNA, 132
 - Satellite DNA, 132
 - Functional genes in repetitive arrays, 132
 - Dispersed repeats, 133
- Chromosome characterization by microscopy, 138
 - Features visible in unspecifically stained chromosomes, 138
 - Chromosome banding, 140
- Chromosome sorting – chromosome-specific probes and libraries, 142
 - Pulsed-field gel electrophoresis (PFGE), 142
 - Fluorescence-activated chromosome sorting (FACS), 142
 - Use of human–rodent hybrid cells, 143
- Physical dissection of the chromosomes, 144
 - Getting chromosome-specific DNA markers – the use of PCR, 144
- Comprehensive genetic mapping with molecular markers, 147
 - Molecular markers present in populations – RFLPs, 147
 - Standard genotypes for genetic mapping, 149
 - Physical mapping, 151
- From physical map to complete DNA sequence, 153
 - Different gene densities in different organisms, 153
 - DNA methylation – undermethylated 'islands' as markers for active genes, 156
 - How many genes? 156
- DNA of the organelles, 157
 - Mitochondrial genomes – total analysis, 157
 - The mitochondrion as an alien system, 159
 - Chloroplast DNA, 159
- Conclusion: total analysis of genomes – what benefits to understanding? 160
- References, 160

6 Accounting for heritable variation, 163

- Introduction, 163
 - Background, 163
 - Kinds of genetic variation, 163
- Molecular identification of mutations of strong effect, 164
 - Using the mutation to find the gene, 164
 - Ways of screening known genes for sequence variation, 166
 - The diversity of mutant alleles in populations, 168
 - Screening for known mutations, 169
- DNA sequence variation without phenotypic consequences, 169
 - The scope for neutral mutation in protein-encoding genes, 169
 - Sequence variation outside coding sequences, 172
- Quantitative variation and its heritability, 172
 - Means and variances, 172
 - Heritability of quantitative variation, 173
 - Heritability assessed by correlations between relatives, 173
 - Estimates of heritability in humans, 175
 - How many genes? 176
 - What kinds of genes are involved in continuous variation? 177
- Attributing heritable quantitative variation to chromosomal loci, 178
 - Quantitative trait loci distinguishing inbred strains, 178
 - Looking for QTLs in outbred populations, 180
- Conclusions and perspectives, 181
- References, 182

7 Gene interactions and the genetic programme, 183

- Introduction, 183
 - Clues to gene function, 183
 - Molecular mechanisms of gene interaction, 184

- Epistasis and its interpretation, 184
- Examples from the yeasts, 185
 - Control of the cell cycle – an example of a protein cascade, 185
 - Galactose metabolism in budding yeast – transcriptional regulation and protein–protein interaction, 186
 - General control of amino acid biosynthesis – a protein kinase cascade that regulates translation, 189
- Another level of control – RNA splicing in *Drosophila* sexual development, 192
- The emergence of morphological pattern, 194
 - The maternal contribution – analysis in *Drosophila*, 194
 - The establishment of compartments in *Drosophila* development, 196
- Selector genes in *Drosophila* and other organisms, 201
 - Drosophila*, 201
 - Mouse, 203
 - Flowering plants, 204
- Timing of gene expression – a function for introns? 205
- Stabilizing gene activity – epigenetic effects, 207
 - Transcriptional activity/inactivity – mechanisms of maintenance, 207
 - X-chromosome inactivation in mammals, 209
 - Dosage compensation in *Drosophila*, 209
 - Drosophila* position–effect variegation and the maintenance of states of gene activity during development, 210
 - Imprinting of mammalian chromosomes, 213
 - The nature and significance of chromatin structure, 214
- Towards the complete description of the organism, 215
- Further reading, 216
- References, 216

Index, 219

Genetics increasingly dominates biological science. What began as a rather esoteric field, concerned with the mode of inheritance of minor variations and quirks, has become the main road to understanding how living organisms function. The great expansion of the scope of genetics is due to the reinforcement of classical genetic methodology with the newer molecular analysis.

The purpose of this book is to explain how modern genetics is actually done, to review what it tells us about the structure and functions of the genetic material, and to consider the extent to which all this new knowledge has solved, or looks like solving, the classical biological problems of variation and development. I have tried to show both the great power of genetic analysis and its present incompleteness – the enormous complexity of living systems will not yield to total analysis in the foreseeable future.

In order to keep the book to a reasonable length it has been necessary to be very selective in the choice of material. I have thought it best to consider a limited number of examples in some detail so as to show the kinds of analysis that are possible, even at the expense of marginalizing other important topics. A good deal of the more detailed information is presented in the figures and their legends. The bibliography is mainly aimed at sup-

.....

PREFACE

porting these specific examples, though I have also listed some more general sources of information.

This is not intended as a comprehensive genetics text-book. The emphasis is on 'mainline' eukaryotic organisms, with bacteria and viruses treated rather summarily, mainly in connection with their relevance to genetic manipulation. I have also concentrated on the problems that can be attacked through controlled investigations in laboratories, and have left aside questions of populational genetic change and evolution. These latter areas require different kinds of analysis and even different styles of thought.

J.R.S.F.
March 1994

Genetics can be said to have started with the rediscovery in 1900 of Mendel's rules, published in 1866, for the transmission of clearly defined distinguishing traits from one generation to the next. Originally formulated for peas, they were soon found to hold true for plants and animals generally. Clear-cut inherited differences could be attributed to hereditary units that soon came to be called genes. Initially, the genes were just symbols in a set of algebraic formulae, set up to describe the patterns of transmission of the differences. But, mainly as the result of the work of T.H. Morgan's school of fruit fly (*Drosophila melanogaster*) geneticists, the genes acquired a physical location. Each gene could be shown to reside at a specific position (locus) on one of the chromosomes of the cell nucleus.

This phase of genetic analysis is described in Chapter 1. It was based entirely on the natural breeding systems of the organisms concerned – not only the sexual cycle of plant and animals but also the very different modes of gene transfer found in bacteria. These natural systems have tended to be overshadowed in recent years by the 'genetic engineering' made possible by molecular technology, but they still provide the geneticist with an essential set of analytical tools. Moreover, they are what goes on all the time in the real world outside the laboratory, which in itself is a more than sufficient reason for knowing about them.

Until the advent of molecular biology, the genes remained intangible. The chromosome locus was merely the site of the determinant of a difference – usually a difference between the normal form of the organism and an aberrant variant, or *mutant*. For several decades after the formal establishment of the chromosome theory of Mendelian inheritance, there remained considerable scepticism about its general importance. The apparently trivial or freakish character of many of the inherited differences used in classical genetics encouraged some to argue that, even if genes existed, they were responsible only for superficial quirks superimposed on an essentially invariant species-specific substructure. This view became increasingly implausible as the number and range of known Mendelian variants was increased, especially through the use of radiation and other mutagenic treat-

INTRODUCTION: THE EXPANDING SCOPE OF GENETICS

ments. It became apparent that no feature of the organism was immune to the effects of mutation. All parts of the living system appeared to be dependent on the integrity of whatever it was that resided at the chromosome loci.

Nevertheless, it remained true for a long time that genes were detectable only in so far as they mutated. As late as the 1950s, at least one distinguished geneticist, Richard Goldschmidt, argued that the 'gene' was created by the mutation – that the 'mutant gene' was just a scar on the chromosome which, in its unscarred state, was an integrated whole, not divided into functionally distinct components.

What, more than anything else, gave solidity to the gene concept, was the detailed study, especially in *Drosophila*, of recurrent mutations at the same chromosome locus. The effects of such *allelic* mutations tended to be variations on a common theme. By the criterion of non-complementation (discussed in Chapter 2) they seemed to represent different degrees of defect in the same function of the organism. Whatever it was that resided at the chromosome locus, it clearly had a high degree of functional specificity.

As genetic mapping was pursued to a higher level of resolution, first in *Drosophila* and then in even more detail in micro-organisms, it was discovered that the functional units, or genes, were not indivisible, as had previously been assumed, but consisted of linear arrays of individually mutable sites. And as the effects of gene mutations were analysed to the biochemical level, it became clear in certain cases that mutations within a gene caused changes in the sequence of amino acids in a protein, and that the amino acid sequence of the

protein was encoded in the linear structure of the gene.

This phase of genetic analysis, culminating in the concept of the gene as a repository of linearly encoded information for the synthesis of a specific macromolecule, will be described in Chapter 2. It proceeded against the background of the proof (by transformation experiments) of the genetic function of DNA and concurrently with the elucidation of the biochemical mechanisms whereby DNA is transcribed into RNA, which is then translated into protein structure.

Until the 1970s, however, it still remained true that genes were recognized only through the effects of their mutations. The gene as a molecular structure in itself remained elusive. The great change came with the development of new molecular techniques, first for detection (probing) of specific DNA sequences and then for amplifying and purifying genes or gene fragments as molecular clones. These revolutionary developments are described in Chapter 3. They gave physico-chemical reality to the genes and showed them to be tracts of DNA. Genes at the DNA level conform to the expectations of classical genetics in that they reside within specific and relatively short chromosome segments but, as we shall see in Chapter 4, they turn out to display a complexity and variety of structure and function that could not previously have been imagined.

As a consequence of the molecular revolution, the agenda of genetic analysis has become radically changed. Formerly it started with genetic variation and attempted to define and map the genes responsible. The challenge was to account for inherited variation in terms of gene differences. Now genes of all kinds can be detected and mapped, whatever their functions and whether they mutate or not. Moreover, once genes have been isolated as DNA, alterations of any desired kind can be made in them to order outside the cell, and the altered genes reintroduced into the living system for observation of the consequences. Analysis of gene function is no longer necessarily dependent on random mutation.

With the powerful combination of classical and molecular methods of analysis, the scope and ambition of genetics is greatly increased. It becomes possible to pose questions about maximum

objectives, and the extents to which they are likely to be realized. The final three chapters of this book address what may be considered as the three grand objectives of genetics.

The first, dealt with in Chapter 5, is the *genome project*, at present attracting a great deal of attention as applied to humans, but being pursued in several other species as well. The project is to make a complete molecular map and ultimately obtain complete DNA sequences of all the chromosomes, at least for one representative individual. For more complex organisms it is a formidable task, but feasible given sufficient resources. Once a complete DNA sequence has been obtained, it should be possible to recognize the genes, or potential genes, by computer-based scanning. Finding out what all the genes do will be much more difficult. It will be possible to deduce much, though not everything, about the structures of the proteins that they encode, but the only way of finding out what a given protein does for the organism is to see what difference it makes when it is lost or modified, either by random mutation or by DNA engineering. This brings us back to the analysis of genetic variants with which genetics started, but working from the gene to the effect rather than the other way round. This approach is feasible with yeast, far more difficult with the mouse, and possible with humans only at one remove – by using the mouse as a proxy or model.

A second grand objective of genetics, which is considered in Chapter 6, is to account for natural inherited variation in terms of defined gene differences at the molecular level. A comprehensive account is clearly out of the question for any species. It would involve the completion of a whole genome project for every individual. There is, however, some prospect of being able to define some of the gene differences that make relatively large contributions to populational variation, even if their individual effects are not completely clear-cut. Gene differences of small effect – and the effects grade all the way down to zero – are generally not worth pursuing. Nevertheless, if a particular variant form of a gene, whatever the magnitude of its effect, has already been defined in one individual, it can be relatively straightforward to screen for it in other individuals. It is all a question of knowing what to look for.

The final ambition of genetics, dealt with in the final chapter, is to account for the entire development of the organism in terms of the information encoded in the genome. The problem here lies in the sheer complexity of the operation. It would be quite unrealistic to think in terms of a number, however large, of parallel and separate connections between genes and traits of the developed organism. In reality, the system is a network, with a very large number of primary elements (the 10 000 or 100 000 genes and the proteins that they directly encode) forming innumerable cross-connections and loops. We cannot be sure that a

complete description will ever be possible for any but the simplest organisms. But, as Chapter 7 attempts to show, we do begin to see some of the general principles of interaction that, with innumerable subtle variations, may account in principle for the development of living systems. The most fruitful approach is again through the study of heritable variation, but now tracing cause-effect relationships in both directions – from observable inherited differences back to the genes, and (‘genetics in reverse’) from defined changes in the genes out to the phenotype.

The eukaryotic system, 1

Eukaryotes distinguished from prokaryotes, 1
Mitosis and chromosome structure, 4
Alternation of haploid and diploid phases in the sexual cycle, 7

Meiosis and the rules of classical genetics, 9

Meiosis, 9
Single-factor genetic ratios and their explanation, 12
Independent assortment of different allele-pairs, 14
Testing data for fit to theoretical ratios, 17
Linkage groups and the nature of crossovers, 18
Map units and map distance, 24
Placing linked genes in sequence, 24
Distinguishing between independent assortment and distant linkage, 25

Mapping centromeres using tetrads, 25

Sex linkage, 27

Segregation and linkage in human genetics, 28

Assigning linkage groups to chromosomes, 32

Genetic analysis of bacteria and bacteriophage, 34

Three modes of gene transfer in bacteria, 34

Time-of-entry mapping in *E. coli*, 37

Analogy with sexual recombination and segregation, 38

Stable partial diploids – F' plasmids, 40

Bacteriophage lambda – a virus in the chromosome, 41

Bacteriophage T4 – a model recombinational system, 42

Eukaryotic organelle genetics, 42

Exceptions to the rule of equal results from reciprocal crosses, 42

Chloroplast variants, 42

Mitochondrial variants, 43

Summary and perspective, 44

Further reading, 44

References, 44

The eukaryotic system

Eukaryotes distinguished from prokaryotes

Gregor Mendel established his simple laws of heredity in peas, and T.H. Morgan and his colleagues used the fruit fly *Drosophila melanogaster* to show that Mendelian factors were located on physical structures, the chromosomes of the cell nucleus. The classical principles of genetics established by these pioneers were soon shown to apply to other higher plants and animals, to several kinds of fungi and algae, and to human beings. All of these very diverse organisms qualify as *eukaryotes*, a class defined by common features of cell organization which are quite different from those of the *prokaryotes* which, broadly speaking, are the bacteria and allied forms. Another characteristic of eukaryotes is their system of sexual

1

DISSECTING THE GENOME USING NATURAL GENETIC SYSTEMS

reproduction (see pp. 7–9) which is, as we shall see later in this chapter, very different from the modes of genetic mixing found in prokaryotes.

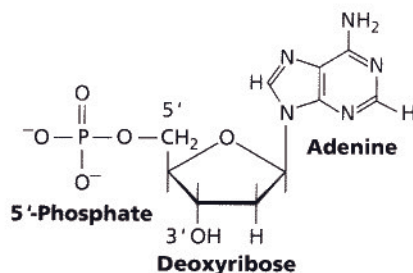
The formal rules governing hereditary transmission in eukaryotes were, as explained in this chapter, all worked out in the absence of any knowledge of deoxyribonucleic acid (DNA). But DNA is now known to be the material in which genetic information is encoded, and is central to all our thinking about genetic mechanisms. Although this chapter will not be so much concerned with molecules as later ones, it is well to put DNA in the centre of the stage from the start. Box 1.1 contains basic information about DNA structure and the way that specific sequences of DNA bases are faithfully transmitted through cycles of replication.

In eukaryotic, as opposed to prokaryotic cells, the bulk of the DNA is present in a cell *nucleus*, bounded by a nuclear membrane, often called the nuclear envelope. The nuclear DNA is divided between a number of microscopically visible bodies called chromosomes, literally 'coloured bodies' in reference to their staining by a range of microscopists' dyes. Another feature of eukaryotic cells is the presence of minor fractions of the DNA in internal membrane-bound organelles – namely *mitochondria*, the centres of energy-generating oxidative metabolism, and, in green plants, *chloroplasts*, the centres of photosynthesis. These organelles, both in the non-chromosomal organization of their DNA and in their sensitivity to antibiotics, seem to have bacterial affinities, and it has been conjectured that they may be the specialized and reduced descendants of prokaryotes that established symbiotic relationships with the

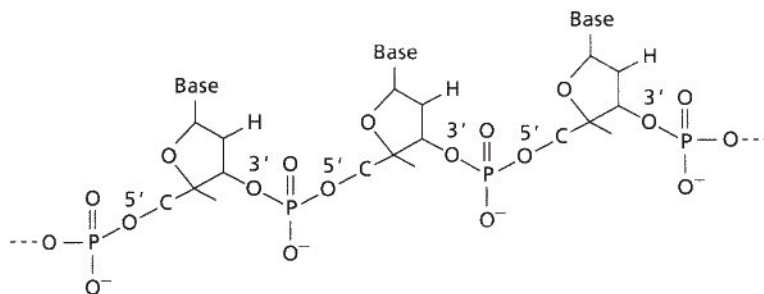
Box 1.1 DNA structure and replication

DNA consists of chains of deoxyribonucleotide units, each containing one of four alternative bases – adenine or guanine (purines) or cytosine or thymine (pyrimidines).

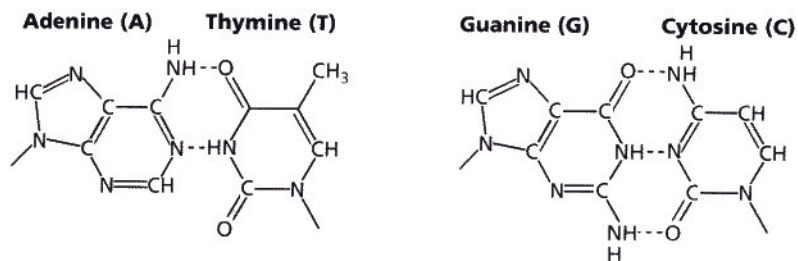
Deoxyadenylic acid (deoxyadenosine 5'-monophosphate) as an example of deoxynucleotide structure:



The DNA chain has the 3'-hydroxyl of the deoxyribose of one unit condensed with the 5'-phosphate of the next:



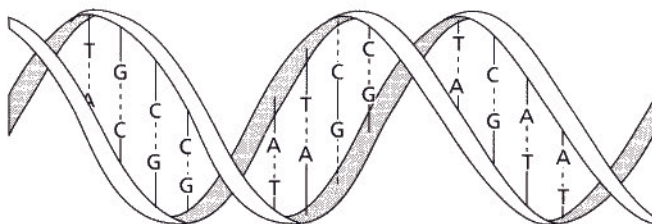
Double-stranded DNA consists of two chains of opposite orientation (one 3'-to-5' and the other 5'-to-3') held together in a double-helix by specific hydrogen bonding between adenine–thymine and guanine–cytosine base-pairs:



continued

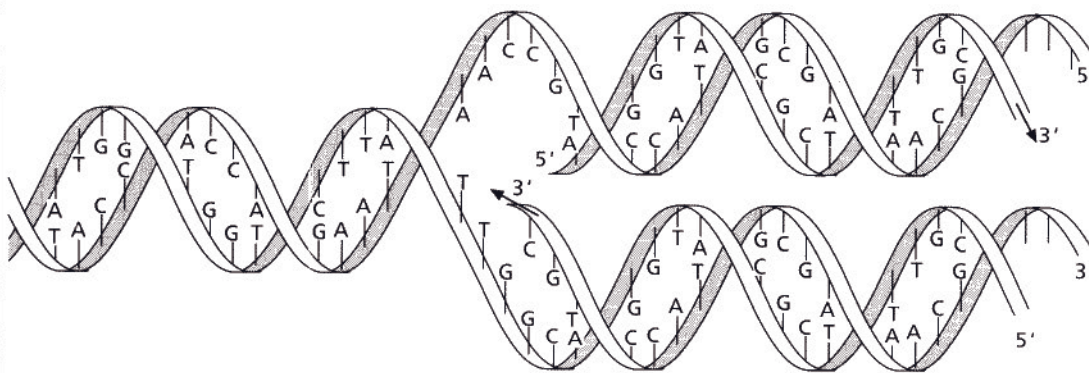
Box 1.1 *Continued*

Skeletal model of the double-helix (duplex):



Note. The base sequence along one strand is the exact complement of, and therefore fixed by, the sequence along the other strand.

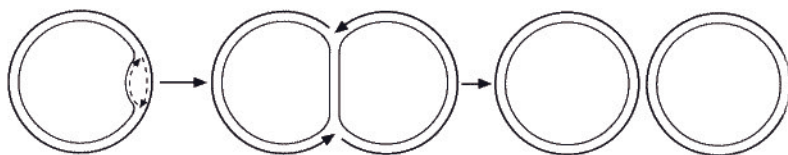
Replication of double-stranded DNA is catalysed by DNA polymerase and a number of accessory proteins. It proceeds by the progressive unwinding of the duplex at a replication fork and the synthesis of a new strand alongside each of the old ones. New synthesis is by successive addition of deoxynucleotide units to the 3' ends of the growing chains. The deoxynucleotide to be added at each step is selected by the enzyme for correct base fit with the pre-existing strand, i.e., A opposite T, G opposite C. Because of the opposite 5'-3' polarities of the two strands in the duplex, the new strands have to be synthesized in opposite directions. One can grow continuously with the moving replication fork and the other has to be synthesized 'backwards' in initially discontinuous fragments that are subsequently joined up.



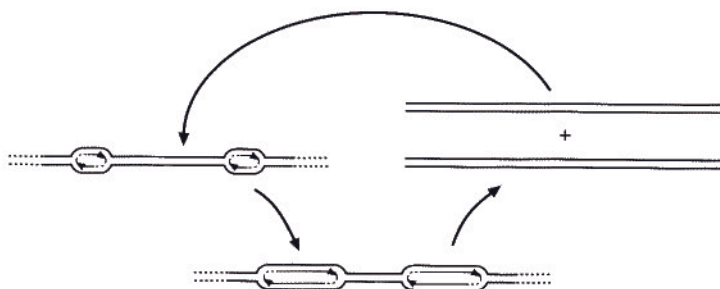
continued on p. 4

Box 1.1 Continued

The DNA of prokaryotes, plasmids and eukaryotic mitochondria and chloroplasts organelles is generally in the form of a closed loop and replicated in both directions from a single origin, with divergent replication forks.



Chromosomal DNA replication proceeds from multiple origins, of the order of 100–1000 per chromosome, with separate replication ‘bubbles’ eventually merging.



primitive eukaryotic cell. We return to their genetic significance at the end of this chapter.

In contrast to eukaryotes, prokaryotic cells have no internal organelles, and their DNA, usually in the form of a single large folded molecule, is attached to the cell membrane and not enclosed in a nuclear envelope.

Mitosis and chromosome structure

During ordinary cell division (though not in meiosis, see p. 9), each division of the cell is preceded by the division of each of the chromosomes that constitute the cell nucleus. Each daughter cell receives a nucleus with the same set of chromosomes as was present in the mother cell.

The process is called mitosis and is illustrated in Fig. 1.1.

The chromosome, in its undivided state, consists of a single very long molecule of double-stranded DNA, packed together with special proteins to form a still deeply problematic complex called *chromatin*. In preparation for mitosis, the chromosomal DNA is replicated (see Box 1.1, p. 3). The two daughter duplexes become separated into two parallel *chromatids* which remain held together at one point, the *centromere*. During the *prophase* stage the chromosomes become progressively condensed until, at the *metaphase* stage, they are sufficiently compact to be visible as discrete, countable bodies under the microscope.

Metaphase is characterized by the appearance

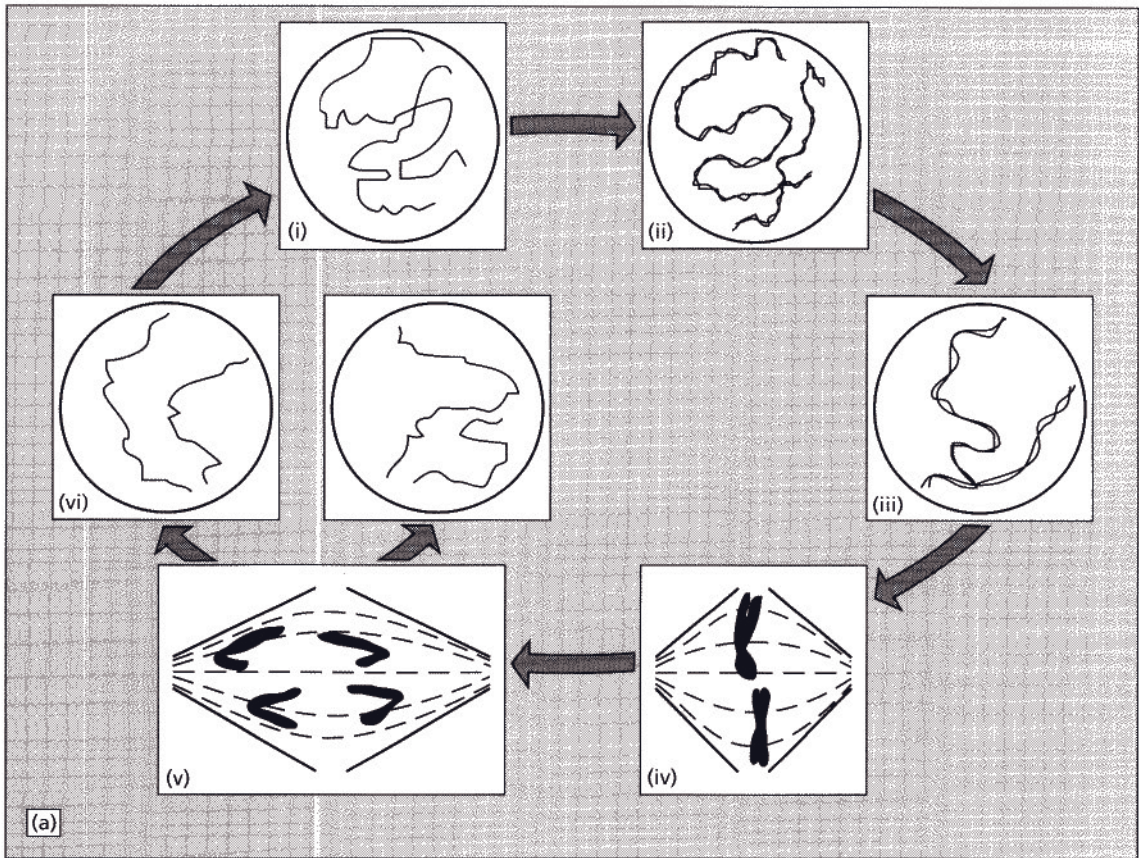


Fig. 1.1 (a) The mitotic cycle in a typical eukaryote. (i) Interphase before replication (G1); (ii) interphase after replication (G2); (iii) prophase, with chromosomes becoming contracted; (iv) metaphase, with chromosomes maximally contracted and aligned on the equator of the division spindle, the nuclear envelope having dissolved; (v) anaphase, with daughter chromosomes pulled apart by their centromeres to the spindle poles; (vi) telophase – nuclear envelopes form round daughter chromosome groups and chromosomes lose their compaction. For simplicity only two chromosomes are shown. The centrosomes (spindle pole bodies), which are seen in animals and fungi but not in flowering plants, are not shown here.

(Continued overleaf.)

of the *division spindle*, an array of parallel protein fibres with contractile properties, and the disappearance (except in fungi) of the nuclear envelope. The centromeres of the metaphase chromosomes become attached to spindle fibres and come to lie in a plane, the spindle equator, midway between the spindle poles. Metaphase is succeeded by *anaphase*, when each centromere splits into two daughter centromeres and the chromatids become separate daughter chromosomes. The daughter centromeres are pulled apart by their attached spindle fibres towards the spindle poles, with the flanking chromosome arms trailing

behind them. At *telophase*, the chromosomes lose their compact structure, and the two groups of daughter chromosomes become enclosed in nuclear envelopes to provide working nuclei for the two products of cell division.

It is at first sight difficult to believe that the relatively bulky metaphase chromatid could represent a single double-stranded DNA molecule. The quantity of DNA in the largest human chromosome, to take one example, would be about 8 cm in length if free of protein and fully extended. The chromosome at metaphase is of the order of 10 000 times shorter than that. Current

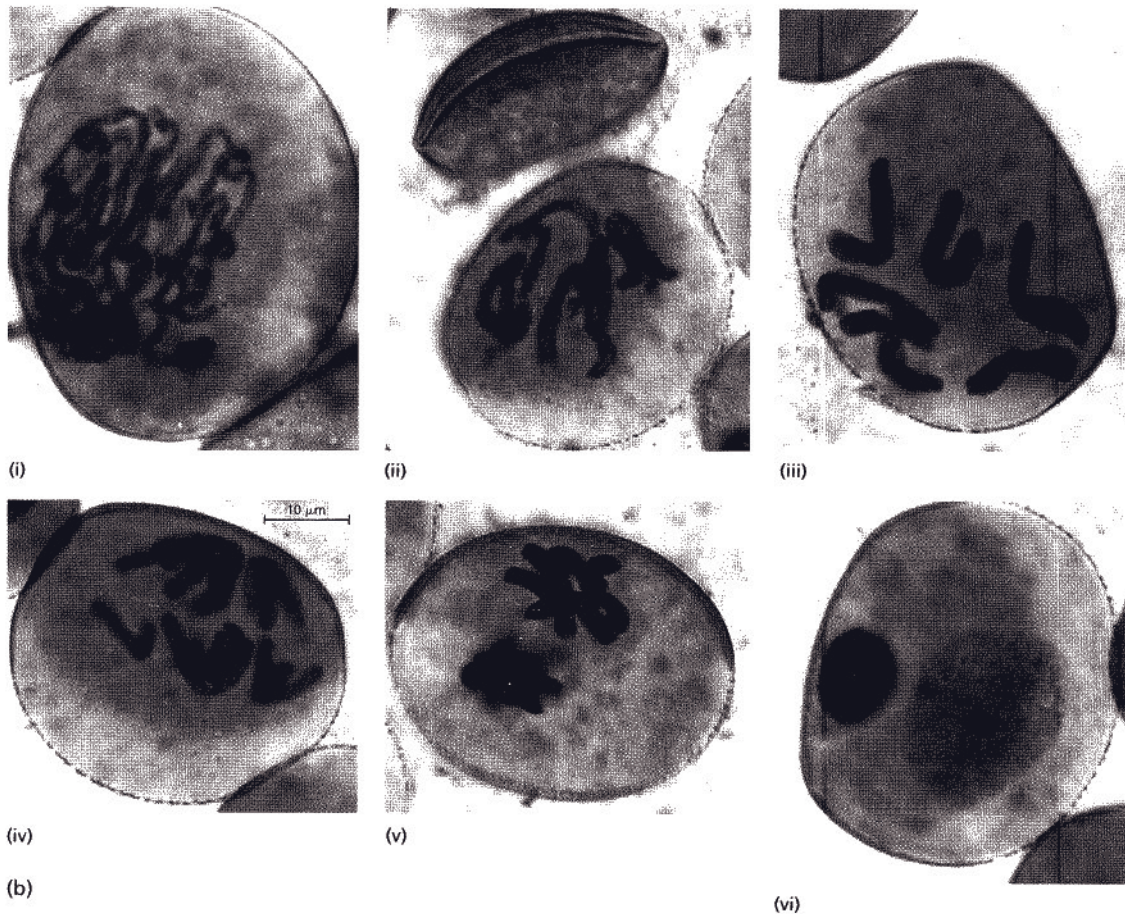


Fig. 1.1 (continued) (b) Stages of mitosis in pollen grains of *Tradescantia paludosa*, a plant species with very large chromosomes: (i) prophase; (ii) pro-metaphase; (iii) metaphase; (iv) anaphase; (v) early telophase; (vi) post-mitosis with two nuclei in the pollen grain. The more diffuse nucleus does not divide again, and the more condensed one divides in the pollen tube to form the two gamete nuclei (cf. Fig. 1.4).

understanding of how the necessary contraction and thickening is achieved is summarized in Fig. 1.2. The best understood part of chromatin structure is the first level of contraction – the formation of *nucleosomes* by the wrapping of the DNA around bead-like complexes of proteins of the *histone* class, and the helical packing of the nucleosome string to form a chromatin fibre. Further contraction is achieved by the attachment of the fibre in loops to a chromosome matrix, which itself may be helically folded as a final stage of compaction.

Metaphase is the stage at which chromosomes are most easily characterized and counted. Every

species has its own characteristic chromosome set, constant in number and form. The most important criteria used for distinguishing between chromosomes microscopically are: (i) length; (ii) position of the centromere, which may be central, off-centre or terminal; and (iii) the pattern of dark and light regions seen after certain special staining procedures (banding techniques, see p. 140).

Depending on the stage of the sexual life cycle, the cell nucleus usually contains either a single (*haploid*, n) set of chromosomes, with every one different, or a double (*diploid*, $2n$) set, with two chromosomes of each kind. Polyploid species, with

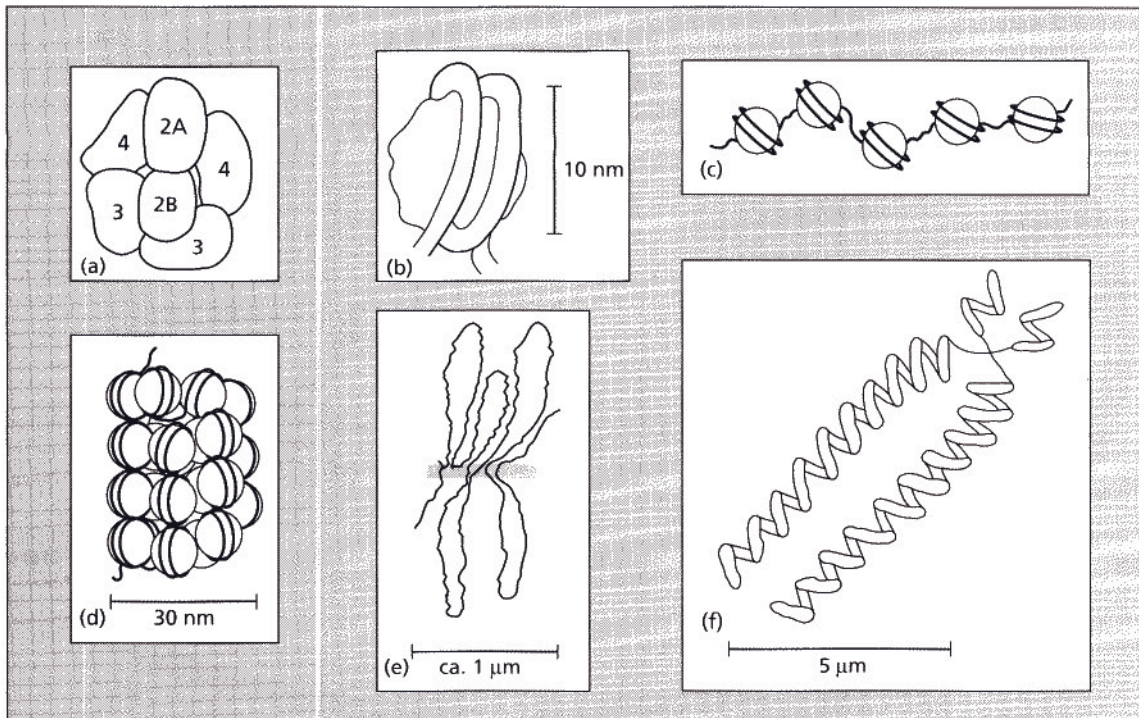


Fig. 1.2 Chromosome structure. (a) The structure of the protein core of the nucleosome, with two molecules each of histones 2A, 2B, 3 and 4 packed together to form a spherical protein core (one 2A and one 2B are out of sight behind the particle). (b) The winding of the DNA duplex around each nucleosome core. (c) The extended nucleosome string. (d) The packing of the nucleosome string to form the approximately $0.03\ \mu\text{m}$ (30 nm, 300 Å) chromatin fibre. The compacted fibre is thought to be stabilized by the binding of a histone H1 molecule between each nucleosome and the next. After Widom & Klug (1985). (e) Looping of the chromatin fibre by its attachment at intervals to the nuclear matrix. At metaphase, shown here, the elements of the matrix coalesce to form a continuous linear chromosome scaffold. Drawn from the photograph of Marsden & Laemmli (1979). (f) Helical folding of the chromosome scaffold at metaphase (human chromosomes). Drawn from photographs of Boy de la Tour & Laemmli (1988).

$3n$, $4n$, $6n$, etc., chromosomes, are widespread and particularly important in plants, but are neglected in this book.

Alternation of haploid and diploid phases in the sexual cycle

The essence of sexual reproduction does not lie in the distinction between male and female – that is a common but not universal feature of sexually reproducing organisms – but rather in the alternation of haploid and diploid phases in the life cycle. The germ cells (*gametes*) – eggs and sperm in animals, ova and pollen tubes in flowering plants, morphologically undifferentiated cells in budding yeast – are haploid. The union of the

germ cells leads to fusion of their nuclei (*karyogamy*) and initiates the diploid phase in which the parental haploid chromosome sets are present together. The transition back to haploid, which is essential for the next turn of the sexual cycle, occurs through the process of *meiosis*, described in the next section. Karyogamy and meiosis are the two key events that punctuate the sexual cycle.

Different groups of eukaryotes differ greatly in the relative duration and prominence of their haploid and diploid phases. In animals the only haploid cells are the gametes. Diploidy is maintained through many rounds of mitosis during the entire growth and development of the animal except in female oocytes and male spermatocytes, which undergo meiosis to form eggs and sperm,

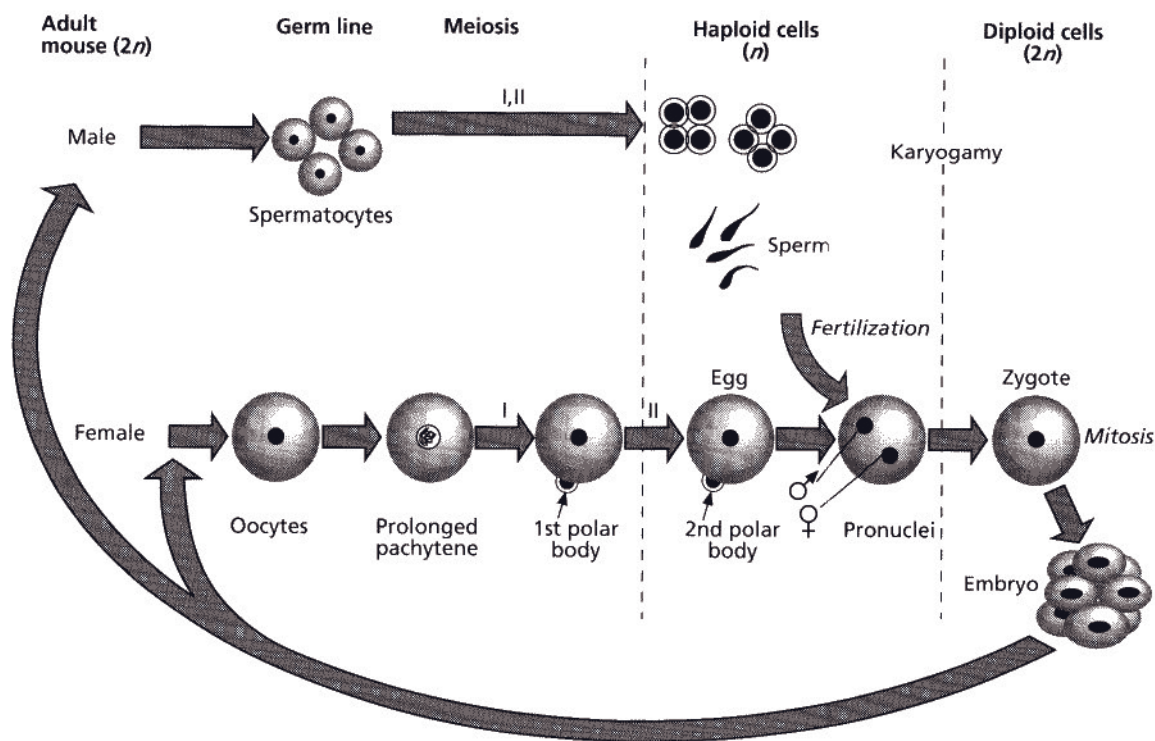


Fig. 1.3 The sexual life cycle of a mammal such as the mouse. The only haploid (n) cells are the immediate products of meiosis. Meiosis in the spermatocyte results in tetrads of spermatids that can all develop into sperm. Meiosis in the oocyte starts soon after the establishment of the female germ line, but is not completed until after fertilization by a sperm cell, which provides the haploid male pronucleus. The haploid female pronucleus is the only surviving product of female meiosis; one product of the first division and one product of the second division are discarded in the first and second polar bodies. Male and female pronuclei fuse to form the diploid ($2n$) zygote, which then undergoes cleavage to initiate embryonic development. In mammals the oocytes are stored with meiosis arrested at pachytene. *Drosophila* differs only in that female meiosis is not initiated at all until fertilization of the oocyte.

respectively. Figure 1.3 summarizes the sexual cycle in the mouse.

Flowering plants come close to the animal pattern in that (leaving aside polyploid species) their vegetative structures are entirely diploid. They differ from animals in a brief propagation of the haploid phase, following meiosis but preceding karyogamy, within the pollen tube and the embryo sac of the ovule (Fig. 1.4). Readers with a broader interest in plants may note that ferns, mosses and liverworts have more extensive haploid phases, amounting in ferns to the small and short-lived but free-living green prothallus and in mosses and liverworts to the whole green plant with the exception of the spore-bearing capsule, which is diploid.

Most fungi, including the experimentally im-

portant filamentous Ascomycetes, are haploid, with meiosis following immediately after karyogamy. But budding yeast, *Saccharomyces cerevisiae*, the fungus most prominently featured in this book, is unusual in being able to propagate itself vegetatively in either the haploid or the diploid phase (Fig. 1.5). Haploid cells are not distinguishably male and female, but exist in two self-incompatible but mutually compatible mating types. So long as the mating types are kept separate the haploid condition can be maintained, with the reservation that some strains, including most wild strains, are able to switch mating type. Diploid cultures are stable so long as they are well nourished, but are induced to undergo meiosis by starvation.

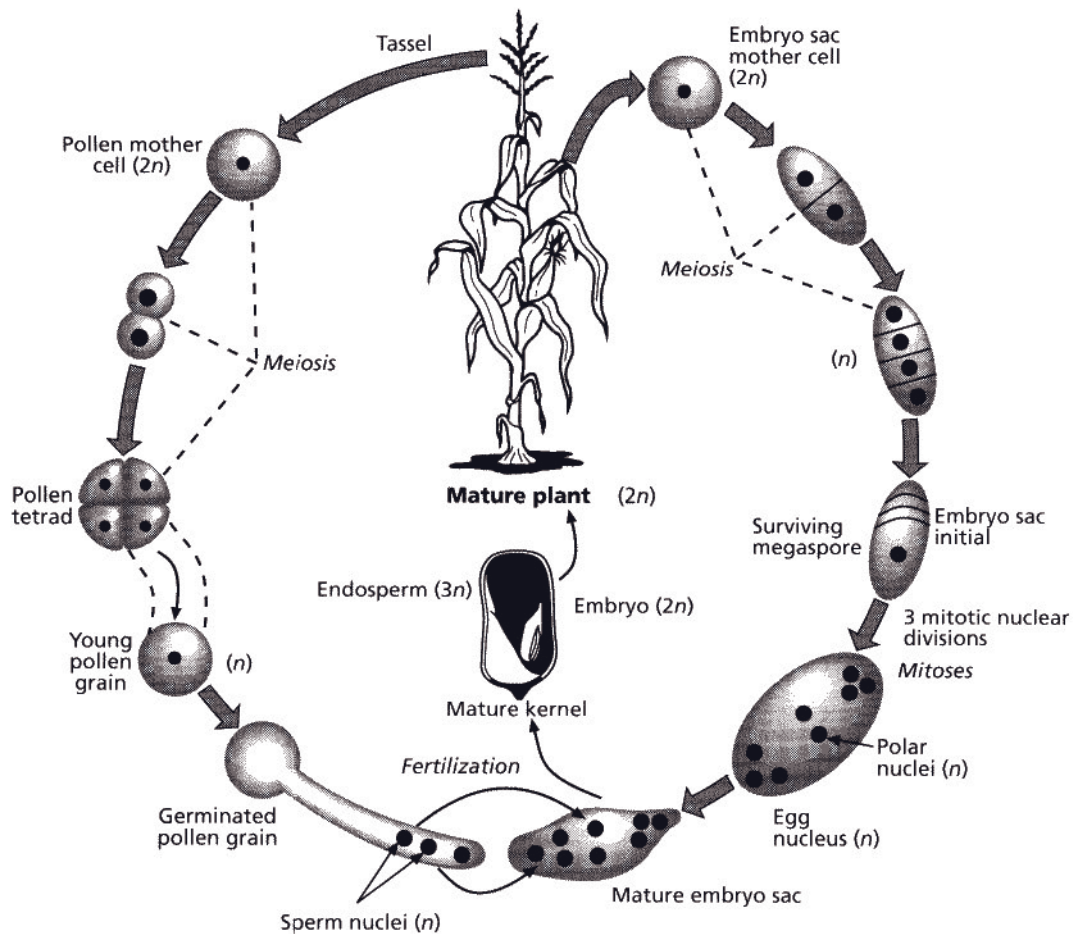


Fig. 1.4 The alternation of haploid and diploid phases in the life cycle of maize (*Zea mays*). Following meiosis, there are three rounds of haploid mitosis in the embryo sac before egg formation and, in the pollen grain and pollen tube, two haploid mitotic divisions. The endosperm, which provides a food store in the seed, is often triploid, the product of fusion of two embryo sac nuclei and one pollen tube nucleus. Based on Srb *et al.* (1965).

Meiosis and the rules of classical genetics

Meiosis

The process of meiosis, with some local variations some of which we shall deal with later, conforms to a standard pattern throughout the eukaryotic world. Its principal stages are illustrated in Fig. 1.6, and may be summarized as follows.

1 DNA replication occurs before the start of meiosis, and there is no further DNA replication (apart from some basic repair—replication associated with crossing over; see below) until after it is

finished. The $2n$ chromosomes, notwithstanding the fact that initially they do not look double, start meiosis already replicated. They actually comprise four copies of the genome, which, through the two divisions of meiosis, are segregated from one another to provide one copy for each of the four meiotic products.

2 Pairs of matching (*homologous*) chromosomes become closely aligned point-for-point along their entire lengths. At the *pachytene* stage, when pairing (*synapsis*) is complete, alignment is so close that, even with the best resolution of the light microscope, there may be difficulty in seeing any space between the homologues. Electron micro-

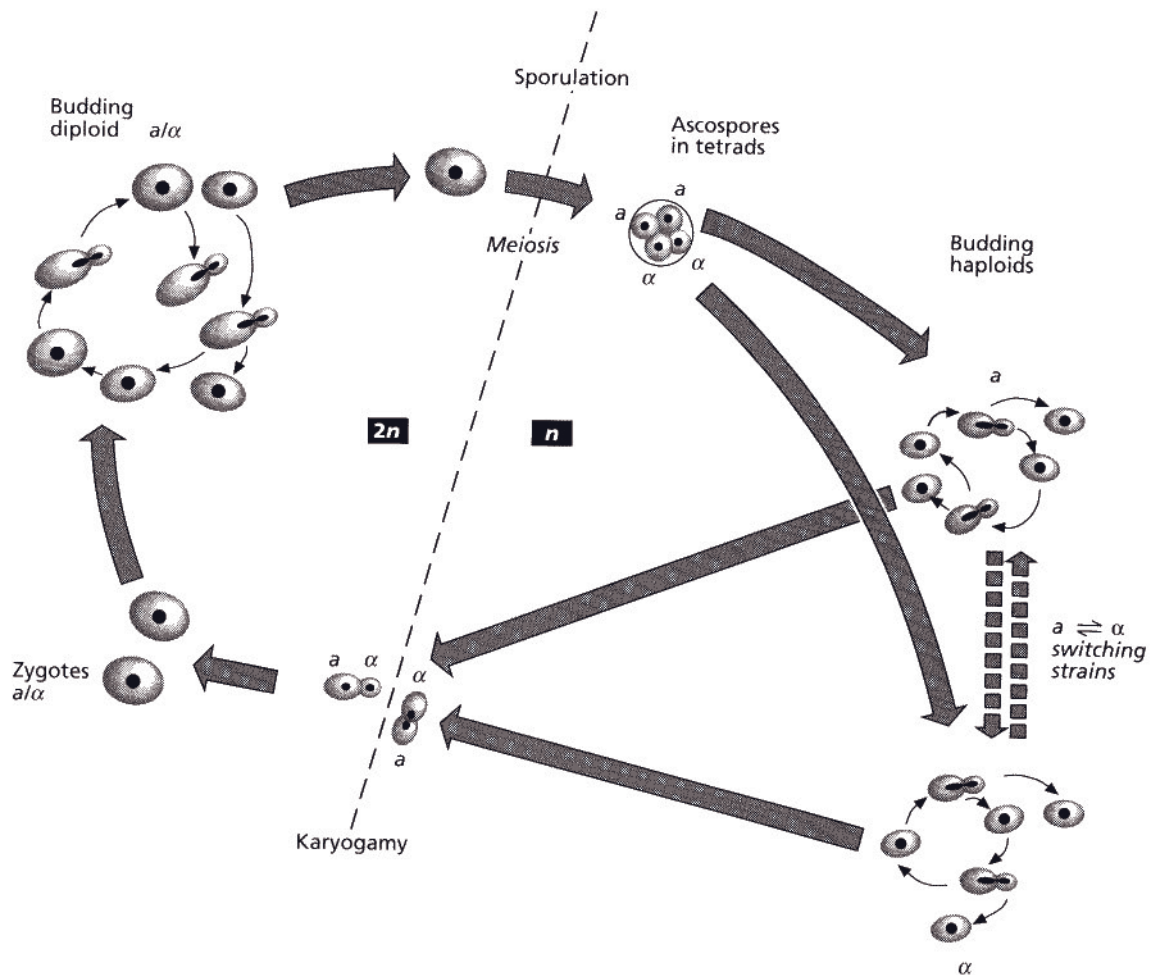


Fig. 1.5 The alternation of haploid and diploid phases in budding yeast, *Saccharomyces cerevisiae*. Haploid cells are of two alternative mating types a and α ; all diploids are of genotype a/α and meiotic tetrads each contain 2 a and 2 α ascospores. In the presence of a switching gene *HO*, which is prevalent in the wild yeast, a and α mating types interconvert frequently by controlled DNA rearrangement, so any haploid quickly reverts to the diploid condition. The standard laboratory strains are homozygous for the non-switching allele *ho*, and so a and α strains can be propagated stably as haploids until brought together.

scopy reveals a protein ribbon, called the *synaptonemal complex*, running between each pair of synapsed chromosomes and consisting of a central element and two lateral elements, the latter apparently in direct contact with the chromosomes. The lateral elements bind to silver, permitting brilliantly clear electron microscopic visualization (Fig. 1.7).

3 The transition from pachytene to diplotene is marked by two striking visible changes. Firstly, the chromosomes can now be seen to be divided

lengthwise into chromatids except at their centromeres, where the chromatids are still attached together. Secondly, the formerly synapsed chromosomes have largely separated, and are prevented from falling apart altogether by the presence between each pair of one or more cross-connections (*chiasmata* – chiasma in the singular). Each chiasma involves just one chromatid from each chromosome; the choice of chromatids for participation in any one chiasma appears to be a matter of chance. There is convincing micro-