



T. A. Brown

# GENE CLONING & DNA ANALYSIS

An Introduction

Seventh Edition

WILEY Blackwell



GENE  
CLONING  
**AND DNA**  
**ANALYSIS**



# GENE CLONING AND DNA ANALYSIS

An Introduction

**T.A. BROWN**

University of Manchester  
Manchester

Seventh Edition

**WILEY** Blackwell

This edition first published 2016 © 2016 by John Wiley & Sons, Ltd  
Sixth edition published 2010 by John Wiley & Sons, Ltd  
Fourth and fifth editions published 2001, 2006 by Blackwell Publishing Ltd  
First, second and third editions published 1986, 1990, 1995 by Chapman & Hall

*Registered office:* John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex,  
PO19 8SQ, UK

*Editorial offices:* 9600 Garsington Road, Oxford, OX4 2DQ, UK  
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK  
111 River Street, Hoboken, NJ 07030-5774, 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](http://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 UK 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.

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.

*Limit of Liability/Disclaimer of Warranty:* While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. 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*

Brown, T.A. (Terence A.), author.

Gene cloning and DNA analysis : an introduction / T.A. Brown. – Seventh edition.

p. ; cm.

Includes bibliographical references and index.

ISBN 978-1-119-07257-7 (cloth) – ISBN 978-1-119-07256-0 (pbk.)

I. Title.

[DNLNLM: 1. Cloning, Molecular. 2. DNA, Recombinant–analysis. 3. Sequence Analysis,  
DNA. QU 450]

QH442.2

572.8'633–dc23

2015015462

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

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

Set in 10/12pt Sabon by Aptara Inc., New Delhi, India

# Contents in Brief

## CONTENTS IN BRIEF

### Part I The Basic Principles of Gene Cloning and DNA Analysis 1

---

- 1 Why Gene Cloning and DNA Analysis are Important 3
- 2 Vectors for Gene Cloning: Plasmids and Bacteriophages 13
- 3 Purification of DNA from Living Cells 25
- 4 Manipulation of Purified DNA 47
- 5 Introduction of DNA into Living Cells 75
- 6 Cloning Vectors for *Escherichia coli* 93
- 7 Cloning Vectors for Eukaryotes 111
- 8 How to Obtain a Clone of a Specific Gene 135
- 9 The Polymerase Chain Reaction 157

### Part II The Applications of Gene Cloning and DNA Analysis in Research 173

---

- 10 Sequencing Genes and Genomes 175
- 11 Studying Gene Expression and Function 201
- 12 Studying Genomes 225

### Part III The Applications of Gene Cloning and DNA Analysis in Biotechnology 245

---

- 13 Production of Protein from Cloned Genes 247
- 14 Gene Cloning and DNA Analysis in Medicine 269
- 15 Gene Cloning and DNA Analysis in Agriculture 291
- 16 Gene Cloning and DNA Analysis in Forensic Science and Archaeology 311

Glossary 329  
Index 345





# Contents

## CONTENTS

*Preface to the Seventh Edition* xvii

*About the Companion Website* xix

## Part I The Basic Principles of Gene Cloning and DNA Analysis 1

---

### 1 Why Gene Cloning and DNA Analysis are Important 3

- 1.1 The early development of genetics 3
- 1.2 The advent of gene cloning and the polymerase chain reaction 4
- 1.3 What is gene cloning? 5
- 1.4 What is PCR? 6
- 1.5 Why gene cloning and PCR are so important 7
  - 1.5.1 Obtaining a pure sample of a gene by cloning 7
  - 1.5.2 PCR can also be used to purify a gene 8
- 1.6 How to find your way through this book 11
  - Further reading 12

### 2 Vectors for Gene Cloning: Plasmids and Bacteriophages 13

- 2.1 Plasmids 13
  - 2.1.1 Size and copy number 14
  - 2.1.2 Conjugation and compatibility 16
  - 2.1.3 Plasmid classification 16
  - 2.1.4 Plasmids in organisms other than bacteria 17
- 2.2 Bacteriophages 17
  - 2.2.1 The phage infection cycle 18
  - 2.2.2 Lysogenic phages 19
    - Gene organization in the  $\lambda$  DNA molecule 19
    - The linear and circular forms of  $\lambda$  DNA 19
    - M13 – a filamentous phage 22
  - 2.2.3 Viruses as cloning vectors for other organisms 24
  - Further reading 24

### 3 Purification of DNA from Living Cells 25

- 3.1 Preparation of total cell DNA 25
    - 3.1.1 Growing and harvesting a bacterial culture 26
    - 3.1.2 Preparation of a cell extract 28
    - 3.1.3 Purification of DNA from a cell extract 29
      - Removing contaminants by organic extraction and enzyme digestion 29
      - Using ion-exchange chromatography to purify DNA from a cell extract 30
      - Using silica to purify DNA from a cell extract 30
    - 3.1.4 Concentration of DNA samples 32
    - 3.1.5 Measurement of DNA concentration 33
    - 3.1.6 Other methods for the preparation of total cell DNA 34
  - 3.2 Preparation of plasmid DNA 35
    - 3.2.1 Separation on the basis of size 35
    - 3.2.2 Separation on the basis of conformation 37
      - Alkaline denaturation 37
      - Ethidium bromide–caesium chloride density gradient centrifugation 38
    - 3.2.3 Plasmid amplification 39
  - 3.3 Preparation of bacteriophage DNA 40
    - 3.3.1 Growth of cultures to obtain a high  $\lambda$  titre 41
    - 3.3.2 Preparation of non-lysogenic  $\lambda$  phages 41
    - 3.3.3 Collection of phages from an infected culture 43
    - 3.3.4 Purification of DNA from  $\lambda$  phage particles 43
    - 3.3.5 Purification of M13 DNA causes few problems 43
- Further reading 45

### 4 Manipulation of Purified DNA 47

- 4.1 The range of DNA manipulative enzymes 48
  - 4.1.1 Nucleases 48
  - 4.1.2 Ligases 50
  - 4.1.3 Polymerases 51
  - 4.1.4 DNA-modifying enzymes 52
- 4.2 Enzymes for cutting DNA: Restriction endonucleases 53
  - 4.2.1 The discovery and function of restriction endonucleases 54
  - 4.2.2 Type II restriction endonucleases cut DNA at specific nucleotide sequences 55
  - 4.2.3 Blunt ends and sticky ends 55
  - 4.2.4 The frequency of recognition sequences in a DNA molecule 57
  - 4.2.5 Performing a restriction digest in the laboratory 58
  - 4.2.6 Analysing the result of restriction endonuclease cleavage 59
    - Separation of molecules by gel electrophoresis 59
    - Visualizing DNA molecules in an agarose gel 60
  - 4.2.7 Estimation of the sizes of DNA molecules 61
  - 4.2.8 Mapping the positions of different restriction sites in a DNA molecule 62

4.2.9 Special gel electrophoresis methods for separating larger molecules 63

#### 4.3 Ligation: Joining DNA molecules together 66

4.3.1 The mode of action of DNA ligase 66

4.3.2 Sticky ends increase the efficiency of ligation 67

4.3.3 Putting sticky ends on to a blunt-ended molecule 67

Linkers 68

Adaptors 68

Homopolymer tailing 70

4.3.4 Blunt end ligation with a DNA topoisomerase 71

Further reading 74

## 5 Introduction of DNA into Living Cells 75

### 5.1 Transformation: The uptake of DNA by bacterial cells 76

5.1.1 Not all species of bacteria are equally efficient at DNA uptake 77

5.1.2 Preparation of competent *E. coli* cells 78

5.1.3 Selection for transformed cells 78

### 5.2 Identification of recombinants 79

5.2.1 Recombinant selection with pBR322: Insertional inactivation of an antibiotic resistance gene 80

5.2.2 Insertional inactivation does not always involve antibiotic resistance 81

### 5.3 Introduction of phage DNA into bacterial cells 83

5.3.1 Transfection 83

5.3.2 *In vitro* packaging of  $\lambda$  cloning vectors 83

5.3.3 Phage infection is visualized as plaques on an agar medium 86

### 5.4 Identification of recombinant phages 86

5.4.1 Insertional inactivation of a *lacZ'* gene carried by the phage vector 87

5.4.2 Insertional inactivation of the  $\lambda$  *cl* gene 87

5.4.3 Selection using the Spi phenotype 88

5.4.4 Selection on the basis of  $\lambda$  genome size 88

### 5.5 Introduction of DNA into non-bacterial cells 88

5.5.1 Transformation of individual cells 88

5.5.2 Transformation of whole organisms 90

Further reading 90

## 6 Cloning Vectors for *Escherichia coli* 93

### 6.1 Cloning vectors based on *E. coli* plasmids 94

6.1.1 The nomenclature of plasmid cloning vectors 94

6.1.2 The useful properties of pBR322 94

6.1.3 The pedigree of pBR322 95

6.1.4 More sophisticated *E. coli* plasmid cloning vectors 95

pUC8: A Lac selection plasmid 97

pGEM3Z: *In vitro* transcription of cloned DNA 98

- 6.2 Cloning vectors based on  $\lambda$  bacteriophage** 99
  - 6.2.1 Segments of the  $\lambda$  genome can be deleted without impairing viability 99
  - 6.2.2 Natural selection was used to isolate modified  $\lambda$  that lack certain restriction sites 100
  - 6.2.3 Insertion and replacement vectors 102
    - Insertion vectors 102
    - Replacement vectors 102
  - 6.2.4 Cloning experiments with  $\lambda$  insertion or replacement vectors 103
  - 6.2.5 Long DNA fragments can be cloned using a cosmid 103
  - 6.2.6  $\lambda$  and other high-capacity vectors enable genomic libraries to be constructed 104
- 6.3 Cloning vectors for the synthesis of single-stranded DNA** 106
  - 6.3.1 Vectors based on M13 bacteriophage 107
  - 6.3.2 Hybrid plasmid–M13 vectors 108
- 6.4 Vectors for other bacteria** 109
  - Further reading 110

## **7 Cloning Vectors for Eukaryotes** 111

- 7.1 Vectors for yeast and other fungi** 111
  - 7.1.1 Selectable markers for the 2  $\mu$ m plasmid 112
  - 7.1.2 Vectors based on the 2  $\mu$ m plasmid: Yeast episomal plasmids 112
  - 7.1.3 A YEp may insert into yeast chromosomal DNA 113
  - 7.1.4 Other types of yeast cloning vector 115
  - 7.1.5 Artificial chromosomes can be used to clone long pieces of DNA in yeast 116
    - The structure and use of a YAC vector 116
    - Applications for YAC vectors 118
  - 7.1.6 Vectors for other yeasts and fungi 118
- 7.2 Cloning vectors for higher plants** 119
  - 7.2.1 *Agrobacterium tumefaciens*: nature's smallest genetic engineer 119
    - Using the Ti plasmid to introduce new genes into a plant cell 120
    - Production of transformed plants with the Ti plasmid 122
    - The Ri plasmid 123
    - Limitations of cloning with *Agrobacterium* plasmids 123
  - 7.2.2 Cloning genes in plants by direct gene transfer 124
    - Direct gene transfer into the nucleus 125
    - Transfer of genes into the chloroplast genome 125
  - 7.2.3 Attempts to use plant viruses as cloning vectors 126
    - Caulimovirus vectors 127
    - Geminivirus vectors 127
- 7.3 Cloning vectors for animals** 127
  - 7.3.1 Cloning vectors for insects 128
    - P elements as cloning vectors for *Drosophila* 128
    - Cloning vectors based on insect viruses 129

- 7.3.2 Cloning in mammals 130
  - Viruses as cloning vectors for mammals 130
  - Gene cloning without a vector 131
- Further reading** 132

## 8 How to Obtain a Clone of a Specific Gene 135

- 8.1 The problem of selection** 135
  - 8.1.1 There are two basic strategies for obtaining the clone you want 136
- 8.2 Direct selection** 137
  - 8.2.1 Marker rescue extends the scope of direct selection 138
  - 8.2.2 The scope and limitations of marker rescue 139
- 8.3 Identification of a clone from a gene library** 140
  - 8.3.1 Gene libraries 140
    - Not all genes are expressed at the same time 140
    - mRNA can be cloned as complementary DNA 142
- 8.4 Methods for clone identification** 143
  - 8.4.1 Complementary nucleic acid strands hybridize to each other 143
  - 8.4.2 Colony and plaque hybridization probing 144
    - Labelling with a radioactive marker 145
    - Non-radioactive labelling 146
  - 8.4.3 Examples of the practical use of hybridization probing 146
    - Abundance probing to analyse a cDNA library 147
    - Oligonucleotide probes for genes whose translation products have been characterized 148
    - Heterologous probing allows related genes to be identified 150
    - Southern hybridization enables a specific restriction fragment containing a gene to be identified 151
  - 8.4.4 Identification methods based on detection of the translation product of the cloned gene 153
    - Antibodies are required for immunological detection methods 153
    - Using a purified antibody to detect protein in recombinant colonies 154
    - The problem of gene expression 155
- Further reading** 155

## 9 The Polymerase Chain Reaction 157

- 9.1 PCR in outline** 157
- 9.2 PCR in more detail** 159
  - 9.2.1 Designing the oligonucleotide primers for a PCR 159
  - 9.2.2 Working out the correct temperatures to use 162
- 9.3 After the PCR: Studying PCR products** 164
  - 9.3.1 Gel electrophoresis of PCR products 164
  - 9.3.2 Cloning PCR products 166
  - 9.3.3 Problems with the error rate of *Taq* polymerase 167

- 9.4 Real-time PCR enables the amount of starting material to be quantified 169
  - 9.4.1 Carrying out a quantitative PCR experiment 169
  - 9.4.2 Real-time PCR can also quantify RNA 171
- Further reading 171

## Part II The Applications of Gene Cloning and DNA Analysis in Research 173

### 10 Sequencing Genes and Genomes 175

- 10.1 Chain-termination DNA sequencing 176
  - 10.1.1 Chain-termination sequencing in outline 176
  - 10.1.2 Not all DNA polymerases can be used for sequencing 178
  - 10.1.3 Chain-termination sequencing with *Taq* polymerase 179
  - 10.1.4 Limitations of chain-termination sequencing 180
- 10.2 Next-generation sequencing 182
  - 10.2.1 Preparation of a next-generation sequencing library 182
    - DNA fragmentation 183
    - Immobilization of the library 184
    - Amplification of the library 184
  - 10.2.2 Next-generation sequencing methods 185
    - Reversible terminator sequencing 186
    - Pyrosequencing 187
  - 10.2.3 Third-generation sequencing 188
  - 10.2.4 Directing next-generation sequencing at specific sets of genes 188
- 10.3 How to sequence a genome 189
  - 10.3.1 Shotgun sequencing of prokaryotic genomes 190
    - Shotgun sequencing of the *Haemophilus influenzae* genome 190
    - Shotgun sequencing of other prokaryotic genomes 193
  - 10.3.2 Sequencing of eukaryotic genomes 194
    - The hierarchical shotgun approach 194
    - Shotgun sequencing of eukaryotic genomes 196
    - What do we mean by 'genome sequence'? 198
- Further reading 198

### 11 Studying Gene Expression and Function 201

- 11.1 Studying the RNA transcript of a gene 202
  - 11.1.1 Detecting the presence of a transcript and determining its nucleotide sequence 203
  - 11.1.2 Transcript mapping by hybridization between gene and RNA 204
  - 11.1.3 Transcript analysis by primer extension 205
  - 11.1.4 Transcript analysis by PCR 206

- 11.2 Studying the regulation of gene expression** 207
  - 11.2.1 Identifying protein binding sites on a DNA molecule 209
    - Gel retardation of DNA–protein complexes 209
    - Footprinting with DNase I 210
    - Modification interference assays 212
  - 11.2.2 Identifying control sequences by deletion analysis 212
    - Reporter genes 213
    - Carrying out a deletion analysis 215
- 11.3 Identifying and studying the translation product of a cloned gene** 216
  - 11.3.1 HRT and HART can identify the translation product of a cloned gene 216
  - 11.3.2 Analysis of proteins by *in vitro* mutagenesis 216
    - Different types of *in vitro* mutagenesis techniques 218
    - Using an oligonucleotide to create a point mutation in a cloned gene 220
    - Other methods of creating a point mutation in a cloned gene 220
    - The potential of *in vitro* mutagenesis 223
- Further reading** 223

## **12 Studying Genomes** 225

- 12.1 Genome annotation** 225
  - 12.1.1 Identifying the genes in a genome sequence 226
    - Searching for open reading frames 226
    - Simple ORF scans are less effective at locating genes in eukaryotic genomes 227
    - Gene location is aided by homology searching 228
    - Comparing the sequences of related genomes 229
    - Identifying the binding sites for regulatory proteins in a genome sequence 230
  - 12.1.2 Determining the function of an unknown gene 231
    - Assigning gene function by experimental analysis requires a reverse approach to genetics 231
    - Specific genes can be inactivated by homologous recombination 232
- 12.2 Studies of the transcriptome and proteome** 233
  - 12.2.1 Studying the transcriptome 234
    - Studying transcriptomes by microarray or chip analysis 234
    - Studying a transcriptome by SAGE 235
    - Sequencing a transcriptome by RNA-seq 236
    - Advantages of the different methods for transcriptome analysis 237
  - 12.2.2 Studying the proteome 237
    - Separating the proteins in a proteome 238
    - Identifying the individual proteins after separation 239
  - 12.2.3 Studying protein–protein interactions 240
    - Phage display 241
    - The yeast two-hybrid system 242
- Further reading** 243

## Part III The Applications of Gene Cloning and DNA Analysis in Biotechnology 245

- ### 13 Production of Protein from Cloned Genes 247
- 13.1 Special vectors for the expression of foreign genes in *E. coli* 249
    - 13.1.1 The promoter is the critical component of an expression vector 251
      - The promoter must be chosen with care 251
      - Examples of promoters used in expression vectors 253
    - 13.1.2 Cassettes and gene fusions 254
  - 13.2 General problems with the production of recombinant protein in *E. coli* 257
    - 13.2.1 Problems resulting from the sequence of the foreign gene 257
    - 13.2.2 Problems caused by *E. coli* 258
  - 13.3 Production of recombinant protein by eukaryotic cells 259
    - 13.3.1 Recombinant protein from yeasts and filamentous fungi 260
      - Saccharomyces cerevisiae* as the host for recombinant protein synthesis 260
      - Other yeasts and fungi 261
    - 13.3.2 Using animal cells for recombinant protein production 262
      - Protein production in mammalian cells 262
      - Protein production in insect cells 262
    - 13.3.3 Pharming: Recombinant protein from live animals and plants 263
      - Pharming in animals 263
      - Recombinant proteins from plants 265
      - Ethical concerns raised by pharming 265
- Further reading 266
- ### 14 Gene Cloning and DNA Analysis in Medicine 269
- 14.1 Production of recombinant pharmaceuticals 269
    - 14.1.1 Recombinant insulin 270
      - Synthesis and expression of artificial insulin genes 270
    - 14.1.2 Synthesis of human growth hormones in *E. coli* 271
    - 14.1.3 Recombinant factor VIII 274
    - 14.1.4 Synthesis of other recombinant human proteins 275
    - 14.1.5 Recombinant vaccines 275
      - Producing vaccines as recombinant proteins 276
      - Recombinant vaccines in transgenic plants 277
      - Live recombinant virus vaccines 279
  - 14.2 Identification of genes responsible for human diseases 280
    - 14.2.1 How to identify a gene for a genetic disease 282
      - Locating the approximate position of the gene in the human genome 282
      - Linkage analysis of the human *BRCA1* gene 283
      - Identification of candidates for the disease gene 284



- 14.3 Gene therapy** 286
  - 14.3.1 Gene therapy for inherited diseases 286
  - 14.3.2 Gene therapy and cancer 288
  - 14.3.3 The ethical issues raised by gene therapy 288
- Further reading** 290

## **15 Gene Cloning and DNA Analysis in Agriculture** 291

- 15.1 The gene addition approach to plant genetic engineering** 292
  - 15.1.1 Plants that make their own insecticides 292
    - The  $\delta$ -endotoxins of *Bacillus thuringiensis* 292
    - Cloning a  $\delta$ -endotoxin gene in maize 293
    - Cloning  $\delta$ -endotoxin genes in chloroplasts 295
    - Countering insect resistance to  $\delta$ -endotoxin crops 296
  - 15.1.2 Herbicide-resistant crops 298
    - 'Roundup Ready' crops 298
    - A new generation of glyphosate-resistant crops 299
  - 15.1.3 Other gene addition projects 300
- 15.2 Gene subtraction** 302
  - 15.2.1 Antisense RNA and the engineering of fruit ripening in tomato 302
    - Using antisense RNA to inactivate the polygalacturonase gene 302
    - Using antisense RNA to inactivate ethylene synthesis 304
  - 15.2.2 Other examples of the use of antisense RNA in plant genetic engineering 304
- 15.3 Problems with genetically modified plants** 305
  - 15.3.1 Safety concerns with selectable markers 305
  - 15.3.2 The terminator technology 306
  - 15.3.3 The possibility of harmful effects on the environment 307
- Further reading** 308

## **16 Gene Cloning and DNA Analysis in Forensic Science and Archaeology** 311

- 16.1 DNA analysis in the identification of crime suspects** 312
  - 16.1.1 Genetic fingerprinting by hybridization probing 312
  - 16.1.2 DNA profiling by PCR of short tandem repeats 312
- 16.2 Studying kinship by DNA profiling** 315
  - 16.2.1 Related individuals have similar DNA profiles 315
  - 16.2.2 DNA profiling and the remains of the Romanovs 315
    - STR analysis of the Romanov bones 315
    - Mitochondrial DNA was used to link the Romanov skeletons with living relatives 317
    - The missing children 318
- 16.3 Sex identification by DNA analysis** 318
  - 16.3.1 PCRs directed at Y chromosome-specific sequences 318
  - 16.3.2 PCR of the amelogenin gene 319

- 16.4 Archaeogenetics: Using DNA to study human prehistory** 320
- 16.4.1 The origins of modern humans 320
    - DNA analysis has challenged the multiregional hypothesis 321
    - DNA analysis shows that Neanderthals are not the direct ancestors of modern Europeans 322
    - The Neanderthal genome sequence suggests there was interbreeding with *H. sapiens* 323
  - 16.4.2 DNA can also be used to study prehistoric human migrations 324
    - Modern humans may have migrated from Ethiopia to Arabia 324
    - Colonization of the New World 325
- Further reading** 328

*Glossary* 329

*Index* 345

# *Preface to the Seventh Edition*

## PREFACE TO THE SEVENTH EDITION

Anyone who works with DNA is well aware of the dramatic changes that have taken place during the past few years in DNA sequencing methodology. To reflect these advances, in this new edition of *Gene Cloning and DNA Analysis: An Introduction* I have completely remodelled the chapter on DNA sequencing to give the new 'next-generation' methods equal prominence alongside the traditional approaches to DNA sequencing, and also to modernize the description of the ways in which genome sequences are generated. Elsewhere, I have stressed the importance of RNA-seq as a means of studying transcriptomes, and ChIP-seq for locating protein-binding sites. These changes correct the major weakness of the Sixth Edition, which was written just before these methods came into mainstream use.

Elsewhere, I have made the usual updates, especially in Part III where I have tried to keep pace with the increasingly rapid developments in the applications of gene cloning and DNA analysis in industry, medicine and agriculture. I have also rewritten the last part of the final chapter, on archaeogenetics, in order to present some of the new information on the human past that has been revealed by the Neanderthal and Denisovan genome sequences. As always, my primary aim is to ensure that *Gene Cloning* remains an introductory text that begins at the beginning and does not assume that the reader has any prior knowledge of the techniques used to study genes and genomes.

For the n-th time I must thank my wife Keri for the unending support that she has given to me in my decision to use up evenings and weekends writing this and other books.

**T.A. Brown**  
University of Manchester



# *About the companion website*

## ABOUT THE COMPANION WEBSITE

This book is accompanied by a companion website:

[www.wiley.com/go/brown/cloning](http://www.wiley.com/go/brown/cloning)

The website includes:

- Powerpoints of all figures from the book for downloading
- PDFs of tables from the book



# PART I

## The Basic Principles of Gene Cloning and DNA Analysis

- 1** | Why Gene Cloning and DNA Analysis are Important 3
- 2** | Vectors for Gene Cloning: Plasmids and Bacteriophages 13
- 3** | Purification of DNA from Living Cells 25
- 4** | Manipulation of Purified DNA 47
- 5** | Introduction of DNA into Living Cells 75
- 6** | Cloning Vectors for *Escherichia coli* 93
- 7** | Cloning Vectors for Eukaryotes 111
- 8** | How to Obtain a Clone of a Specific Gene 135
- 9** | The Polymerase Chain Reaction 157





# Chapter 1

## Why Gene Cloning and DNA Analysis are Important

### *Chapter contents*

#### CHAPTER CONTENTS

- 1.1 The early development of genetics
- 1.2 The advent of gene cloning and the polymerase chain reaction
- 1.3 What is gene cloning?
- 1.4 What is PCR?
- 1.5 Why gene cloning and PCR are so important
- 1.6 How to find your way through this book

In the middle of the 19th century, Gregor Mendel formulated a set of rules to explain the inheritance of biological characteristics. The basic assumption of these rules is that each heritable property of an organism is controlled by a factor, called a **gene**, that is a physical particle present somewhere in the cell. The rediscovery of Mendel's laws in 1900 marks the birth of **genetics**, the science aimed at understanding what these genes are and exactly how they work.

### 1.1 The early development of genetics

For the first 30 years of its life this new science grew at an astonishing rate. The idea that genes reside on **chromosomes** was proposed by W. Sutton in 1903, and received experimental backing from T.H. Morgan in 1910. Morgan and his colleagues then developed the techniques for **gene mapping**, and by 1922 had produced a comprehensive analysis of the relative positions of over 2000 genes on the four chromosomes of the fruit fly, *Drosophila melanogaster*.

Despite the brilliance of these classical genetic studies, there was no real understanding of the molecular nature of the gene until the 1940s. Indeed, it was not until the

experiments of Avery, MacLeod, and McCarty in 1944, and of Hershey and Chase in 1952, that anyone believed that deoxyribonucleic acid (DNA) is the genetic material. Up until then it was widely thought that genes were made of protein. The discovery of the role of DNA was a tremendous stimulus to genetic research, and many famous biologists (Delbrück, Chargaff, Crick, and Monod were among the most influential) contributed to the second great age of genetics. During the 14 years between 1952 and 1966, the structure of DNA was elucidated, the genetic code cracked, and the processes of transcription and translation described.

## 1.2 The advent of gene cloning and the polymerase chain reaction

These years of activity and discovery were followed by a lull, a period of anticlimax when it seemed to some molecular biologists (as the new generation of geneticists styled themselves) that there was little of fundamental importance that was not understood. In truth, there was a frustration that the experimental techniques of the late 1960s were not sophisticated enough to allow the gene to be studied in any greater detail.

Then, in the years 1971–1973 genetic research was thrown back into gear by what at the time was described as a revolution in experimental biology. A whole new methodology was developed, enabling previously impossible experiments to be planned and carried out, if not with ease, then at least with success. These methods, referred to as **recombinant DNA technology** or **genetic engineering**, and having at their core the process of **gene cloning**, sparked another great age of genetics. They led to rapid and efficient **DNA sequencing** techniques that enabled the structures of individual genes to be determined, reaching a culmination at the turn of the century with the massive genome sequencing projects, including the human project which was completed in 2000. They led to procedures for studying the regulation of individual genes, which have allowed molecular biologists to understand how aberrations in gene activity can result in human diseases such as cancer. The techniques spawned modern **biotechnology**, which puts genes to work in production of proteins and other compounds needed in medicine and industrial processes.

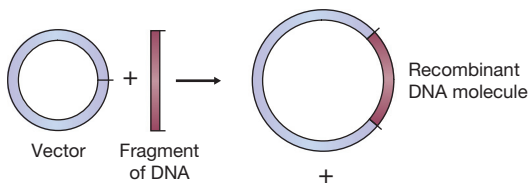
During the 1980s, when the excitement engendered by the gene cloning revolution was at its height, it hardly seemed possible that another, equally novel and equally revolutionary process was just around the corner. According to DNA folklore, Kary Mullis invented the **polymerase chain reaction (PCR)** during a drive along the coast of California one evening in 1985. His brainwave was an exquisitely simple technique that acts as a perfect complement to gene cloning. PCR has made easier many of the techniques that were possible but difficult to carry out when gene cloning was used on its own. It has extended the range of DNA analysis and enabled molecular biology to find new applications in areas of endeavour outside of its traditional range of medicine, agriculture, and biotechnology. Archaeogenetics, molecular ecology, and DNA forensics are just three of the new disciplines that have become possible as a direct consequence of the invention of PCR, enabling molecular biologists to ask questions about human evolution and the impact of environmental change on the biosphere, and to bring their powerful tools to bear in the fight against crime. Forty years have passed since the dawning of the age of gene cloning, but we are still riding the rollercoaster and there is no end to the excitement in sight.

## 1.3 What is gene cloning?

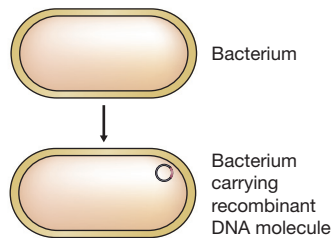
What exactly is gene cloning? The easiest way to answer this question is to follow through the steps in a gene cloning experiment (Figure 1.1):

- 1 A fragment of DNA, containing the gene to be cloned, is inserted into a circular DNA molecule called a **vector**, to produce a **recombinant DNA molecule**.
- 2 The vector transports the gene into a host cell, which is usually a bacterium, although other types of living cell can be used.
- 3 Within the host cell the vector multiplies, producing numerous identical copies, not only of itself but also of the gene that it carries.
- 4 When the host cell divides, copies of the recombinant DNA molecule are passed to the progeny and further vector replication takes place.
- 5 After a large number of cell divisions, a colony, or **clone**, of identical host cells is produced. Each cell in the clone contains one or more copies of the recombinant DNA molecule. The gene carried by the recombinant molecule is now said to be cloned.

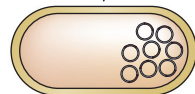
### 1 Construction of a recombinant DNA molecule



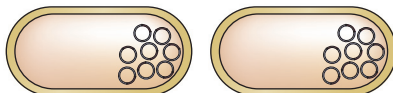
### 2 Transport into the host cell



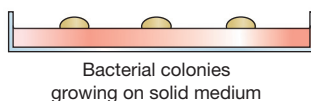
### 3 Multiplication of recombinant DNA molecule



### 4 Division of host cell



### 5 Numerous cell divisions resulting in a clone



**Figure 1.1**

The basic steps in gene cloning.

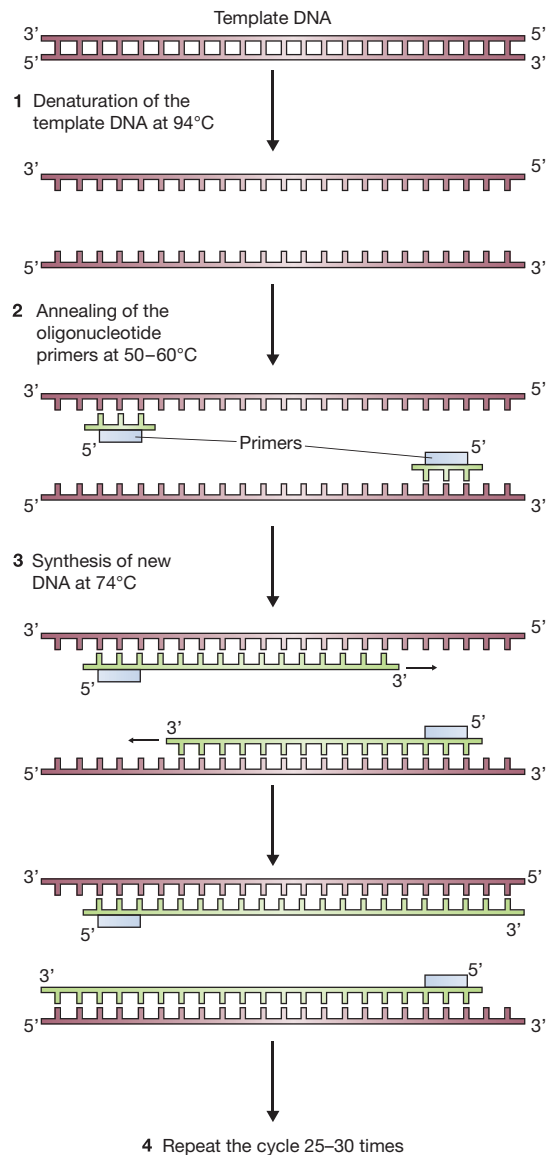
## 1.4 What is PCR?

The polymerase chain reaction is very different from gene cloning. Rather than a series of manipulations involving living cells, PCR is carried out in a single test tube simply by mixing DNA with a set of reagents and placing the tube in a thermal cycler, a piece of equipment that enables the mixture to be incubated at a series of temperatures that are varied in a preprogrammed manner. The basic steps in a PCR experiment are as follows (Figure 1.2):

- 1 The mixture is heated to 94 °C, at which temperature the hydrogen bonds that hold together the two strands of the double-stranded DNA molecule are broken, causing the molecule to **denature**.

**Figure 1.2**

The basic steps in the polymerase chain reaction.



- 2 The mixture is cooled down to 50–60 °C. The two strands of each molecule could join back together at this temperature, but most do not because the mixture contains a large excess of short DNA molecules, called **oligonucleotides** or **primers**, which **anneal** to the DNA molecules at specific positions.
- 3 The temperature is raised to 74 °C. This is a good working temperature for the **Taq DNA polymerase** that is present in the mixture. We will learn more about **DNA polymerases** on p. 51. All we need to understand at this stage is that the **Taq** DNA polymerase attaches to one end of each primer and synthesizes new strands of DNA, complementary to the **template** DNA molecules, during this step of the PCR. Now we have four stands of DNA instead of the two that there were to start with.
- 4 The temperature is increased back to 94 °C. The double-stranded DNA molecules, each of which consists of one strand of the original molecule and one new strand of DNA, denature into single strands. This begins a second cycle of denaturation–annealing–synthesis, at the end of which there are eight DNA strands. By repeating the cycle 30 times the double-stranded molecule that we began with is converted into over 130 million new double-stranded molecules, each one a copy of the region of the starting molecule delineated by the annealing sites of the two primers.

## 1.5 Why gene cloning and PCR are so important

As can be seen from Figures 1.1 and 1.2, gene cloning and PCR are relatively straightforward procedures. Why, then, have they assumed such importance in biology? The answer is largely because both techniques can provide a pure sample of an individual gene, separated from all the other genes in the cell.

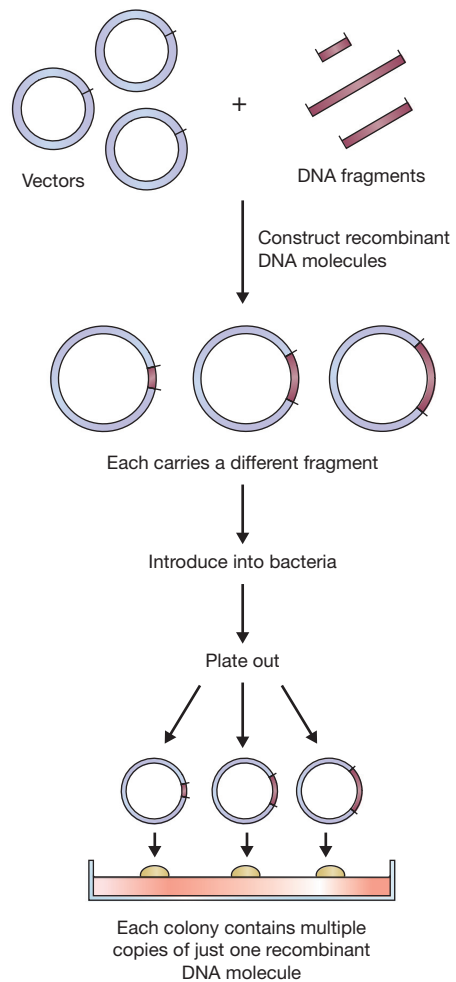
### 1.5.1 Obtaining a pure sample of a gene by cloning

To understand exactly how cloning can provide a pure sample of a gene, consider the basic experiment from Figure 1.1, but drawn in a slightly different way (Figure 1.3). In this example, the DNA fragment to be cloned is one member of a mixture of many different fragments, each carrying a different gene or part of a gene. This mixture could indeed be the entire genetic complement of an organism – a human, for instance. Each of these fragments becomes inserted into a different vector molecule to produce a family of recombinant DNA molecules, one of which carries the gene of interest. Usually, only one recombinant DNA molecule is transported into any single host cell, so that although the final set of clones may contain many different recombinant DNA molecules, each individual clone contains multiple copies of just one molecule. The gene is now separated away from all the other genes in the original mixture, and its specific features can be studied in detail.

In practice, the key to the success or failure of a gene cloning experiment is the ability to identify the particular clone of interest from the many different ones that are obtained. If we consider the **genome** of the bacterium *Escherichia coli*, which contains just over 4000 different genes, we might at first despair of being able to find just one gene among all the possible clones (Figure 1.4). The problem becomes even more overwhelming when we remember that bacteria are relatively simple organisms and that the human genome contains about five times as many genes. As explained in Chapter 8, a variety

**Figure 1.3**

Cloning allows individual fragments of DNA to be purified.



of different strategies can be used to ensure that the correct gene can be obtained at the end of the cloning experiment. Some of these strategies involve modifications to the basic cloning procedure, so that only cells containing the desired recombinant DNA molecule can divide and the clone of interest is automatically **selected**. Other methods involve techniques that enable the desired clone to be identified from a mixture of lots of different clones.

Once a gene has been cloned there is almost no limit to the information that can be obtained about its structure and expression. The availability of cloned material has stimulated the development of many different analytical methods for studying genes, with new techniques being introduced all the time. Methods for studying the structure and expression of a cloned gene are described in Chapters 10 and 11, respectively.

### 1.5.2 PCR can also be used to purify a gene

The polymerase chain reaction can also be used to obtain a pure sample of a gene. This is because the region of the starting DNA molecule that is copied during PCR

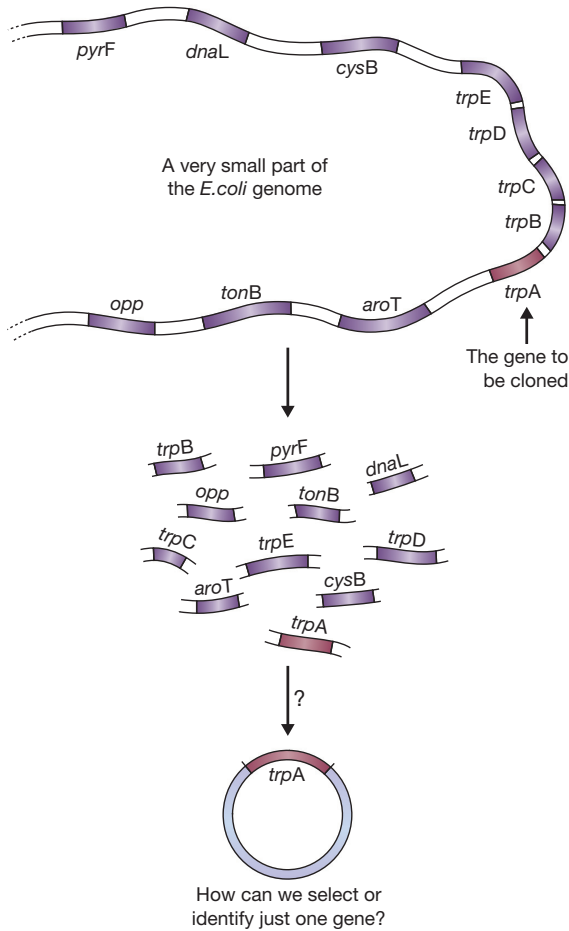


Figure 1.4

The problem of selection.

is the segment whose boundaries are marked by the annealing positions of the two oligonucleotide primers. If the primers anneal either side of the gene of interest, many copies of that gene will be synthesized (Figure 1.5). The outcome is the same as with a gene cloning experiment, although the problem of selection does not arise because the desired gene is automatically 'selected' as a result of the positions at which the primers anneal.

A PCR experiment can be completed in a few hours, whereas it takes weeks – if not months – to obtain a gene by cloning. Why then is gene cloning still used? This is because PCR has two limitations:

- In order for the primers to anneal to the correct positions, on either side of the gene of interest, the sequences of these annealing sites must be known. It is easy to synthesize a primer with a predetermined sequence (see p. 149), but if the sequences of the annealing sites are unknown then the appropriate primers cannot be made. This means that PCR cannot be used to isolate genes that have not been studied before – that has to be done by cloning.