GENES

BENJAMIN LEWIN

GENES

BENJAMIN LEWIN
Editor, *Cell*

John Wiley & Sons
New York Chichester Brisbane
Toronto Singapore

Production supervised by Linda Indig
Manuscript editor was Mondy Dana,
supervised by Vivian Kahane Cooper
Cover and text designed by Laura Ierardi
Illustrations by John Balbalis with
the assistance of the Wiley
illustration department.

Library of Congress Cataloging in Publication Data:

Lewin, Benjamin
Genes.

Includes bibliographies and indexes.
1. Genetics. I. Title.

QH430.L487 1983 575.1 82-20022
ISBN 0-471-09316-5

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

PREFACE

The title of this book almost makes a preface superfluous. *Genes* is simply about genes, recognizing what amounts to a new field whose extraordinary progress has all but overwhelmed the traditional discipline of genetics. My aim is to cut through the enormous mass of information that has accumulated recently, to discern general principles and describe the state of the art in this exciting area. This text asks: what is a gene, how is it reproduced, how is it expressed, what controls its expression?

The underlying theme of this book is that the gene has at its disposal a vast repertoire of strategies for survival, different examples of which are displayed in various systems. Reflecting the perspective of current research, procaryotic and eucaryotic molecular biology are given equal weight. Both are now part of the same story. The starting point is the issue of how a gene is represented in protein; and from the protein, we work backward at the molecular level, as it were, to the DNA itself.

As a comprehensive introduction to the molecular biology of the gene, this book assumes no prior knowledge and is up to date with current research. I hope that this overview will make this rapidly advancing subject more accessible and readily allow readers to proceed to more advanced works. In view of the size of the relevant literature, it would only be confusing to rely on the citation of individual research articles. Each chapter therefore concludes with a bibliography to suggest useful reviews and some research articles that lead more deeply into the subject.

I have tried to illustrate all important points diagramatically, and where appropriate the illustrations attempt to give some feeling for the scale and relationships of the elements involved. This preface would certainly be incomplete without acknowledgment of the considerable artistic endeavors of John Balbalis to realise this aim.

One of the pleasures of writing this book has been the ensuing discussions with my friends and colleagues who have commented on it. Many improvements have resulted from the generous efforts of Sankar Adhya, Sidney Altman, French Anderson, David Clayton, Nicholas Cozzarelli, Bernard Davis, Igor Dawid, Arg Efstratiadis, Nina Federoff, Alice Fulton, Joe Gall, Nicholas Gillham, Philip Hanawalt, Ira Herskowitz, Lee Hood, Joel Huberman, George Khoury, Nancy Kleckner, Marilyn Kozak, Charles Kurland, Art Landy, Jeffrey Miller, Masayasu Nomura, Charles Radding, Jeff Roberts, Rich Roberts, Gerry Rubin, Robert Schimke, David Schlessinger, David Shafritz, Phil Sharp, Allen Smith, Phang-C. Tai, Susumu Tonegawa, Harold Varmus, Alex Varshavsky, and Hal Weintraub. Finally, it hardly needs saying that the book would have been much less fun to write without the enthusiastic participation of my family.

Benjamin Lewin

CONTENTS

PART 1
THE NATURE OF GENETIC INFORMATION

CHAPTER 1
WHAT IS A GENE? A GENETIC VIEW 3

A PARTICULATE FACTOR OF INHERITANCE 4
THE INDEPENDENCE OF DIFFERENT GENES 7
THE ROLE OF CHROMOSOMES IN HEREDITY 7
GENES LIE ON CHROMOSOMES 9
CHROMOSOMES CONTAIN A LINEAR ARRAY OF GENES 12
THE GENETIC MAP IS CONTINUOUS 15
ONE GENE—ONE PROTEIN 16
A NEW DEFINITION: THE CISTRON 17
A NOTE ABOUT TERMINOLOGY 20

CHAPTER 2
WHAT IS A GENE? A BIOCHEMICAL VIEW 21

THE GENETIC MATERIAL IS DNA 22
MORE EVIDENCE FOR THE ROLE OF DNA 23
THE COMPONENTS OF DNA 26
DNA IS A DOUBLE HELIX 28
ABOUT ALTERNATIVE DOUBLE-HELICAL STRUCTURES 31
SUPERCOILING MAY BE IMPOSED ON THE DOUBLE HELIX 36
RNA ALSO HAS SECONDARY STRUCTURE 37
DNA CAN BE DENATURED AND RENATURED 39
NUCLEIC ACIDS HYBRIDIZE BY BASE PAIRING 41
THE MOLECULAR BASIS OF MUTATION 42
MUTATIONS ARE CONCENTRATED AT HOTSPOTS 45
THE RATE OF MUTATION 46

CHAPTER 3
WHAT IS A GENE? MOLECULAR STRUCTURE 49

A DIRECT APPROACH TO GENE STRUCTURE 50
RESTRICTION ENZYMES CLEAVE DNA INTO SPECIFIC FRAGMENTS 50
CONSTRUCTING A RESTRICTION MAP 52
SOME NICETIES OF RESTRICTION MAPPING 53
SEQUENCING DNA 54
ARE GENES AND PROTEINS COLINEAR? 56
EUCARYOTIC GENES CAN BE INTERRUPTED 58
OVERLAPPING AND ALTERNATIVE GENES 60
WHAT'S IN A GENE? 61

CHAPTER 4
BREAKING THE GENETIC CODE 62

DNA NEEDS ONLY TO CODE FOR A SEQUENCE OF AMINO ACIDS 62
THE GENETIC CODE IS READ IN TRIPLETS 65
THE APPARATUS FOR SEQUENTIAL PROTEIN SYNTHESIS 67
CODONS REPRESENTING AMINO ACIDS 68
THE NATURE OF SIGNALS FOR TERMINATION 70
IS THE CODE UNIVERSAL? 72
TRANSLATION IN OVERLAPPING READING FRAMES 73

CHAPTER 5
FROM GENE TO PROTEIN 75
RIBOSOMES ARE THE SITE OF PROTEIN SYNTHESIS 76
SEARCHING FOR THE MESSENGER 76
TRANSFER RNA IS THE ADAPTOR 78
RIBOSOMES TRAVEL IN CONVOY 82

PART 2
HOW PROTEINS ARE SYNTHESIZED

CHAPTER 6
THE ASSEMBLY LINE FOR POLYPEPTIDE SYNTHESIS 87
THE RIBOSOMAL SITES OF ACTION 87
INITIATION: A SPECIAL INITIATOR tRNA 89
INITIATION NEEDS 30S SUBUNITS AND ACCESSORY FACTORS 91
THE BRIEF FREEDOM OF 30S SUBUNITS 92
PICKING OUT THE INITIATOR tRNA 93
EUCARYOTIC INITIATION INVOLVES MANY FACTORS 94
THE IMPORTANCE OF BEING eIF2 95
THE ORDER OF EVENTS IN PROCARYOTES AND EUCARYOTES 95
ELONGATION: BRINGING AMINOACYL-tRNA INTO THE A SITE 96
GTP CLEAVAGE OCCURS AFTER RIBOSOME BINDING 97
PEPTIDE BOND FORMATION IS A RIBOSOMAL FUNCTION 98
THE TRANSLOCATION STEP 99
GETTING THE ENERGY FOR RIBOSOMAL ACTION 101
TERMINATION: COMPLETING PROTEIN SYNTHESIS 102

CHAPTER 7
TRANSFER RNA: THE TRANSLATIONAL ADAPTOR 103
THE UNIVERSAL CLOVERLEAF 104
tRNA CONTAINS MANY MODIFIED BASES 104
AN L-SHAPED TERTIARY STRUCTURE 107
SYNTHETASES SORT THROUGH tRNAS AND AMINO ACIDS 110
THE CHARGING STEP 111
CODON-ANTICODON RECOGNITION INVOLVES WOBBLING 114
BASE MODIFICATION MAY CONTROL CODON RECOGNITION 115
MITOCHONDRIA HAVE MINIMAL tRNA SETS 117
MUTANT tRNAS MAY READ DIFFERENT CODONS 118
SUPPRESSOR tRNAS COMPETE FOR THEIR CODONS 121
TRANSFER RNA MAY INFLUENCE THE READING FRAME 122

CHAPTER 8
THE RIBOSOME TRANSLATION FACTORY 125
RIBOSOMES ARE COMPACT RIBONUCLEOPROTEIN PARTICLES 126
THE STRUCTURE OF RIBOSOMAL RNA 129
RIBOSOMAL PROTEINS OCCUPY SPECIFIC LOCATIONS 131
INTERACTIONS BETWEEN RIBOSOMAL PROTEINS AND rRNA 133
RIBOSOMAL SUBUNITS CAN BE DISSOCIATED AND RECONSTITUTED 133
MUTANTS IN RIBOSOME ASSEMBLY 134
SUBUNIT ASSEMBLY IS LINKED TO TOPOLOGY 135
ALL RIBOSOMAL COMPONENTS CAN BE MUTATED 136
RIBOSOMES HAVE SEVERAL ACTIVE CENTERS 138
BINDING 30S SUBUNITS TO mRNA 140
THE ACCURACY OF TRANSLATION 140

CHAPTER 9
THE MESSENGER RNA TEMPLATE 143
THE TRANSIENCE OF BACTERIAL MESSENGERS 143
THE CONSTITUTION OF BACTERIAL MESSENGERS 144
TRANSLATION OF POLYCISTRONIC MESSENGERS 146
A FUNCTIONAL DEFINITION FOR EUCARYOTIC mRNA 148
EUCARYOTIC mRNA MAY BE POLYADENYLATED AT THE 3′ END 149
ISOLATION OF mRNA VIA POLY(A) 150
EUCARYOTIC mRNA HAS A METHYLATED CAP AT THE 5′ END 151
THE POWER OF *IN VITRO* TRANSLATION SYSTEMS 153
INITIATION MAY INVOLVE PAIRING BETWEEN mRNA AND rRNA 154
SMALL SUBUNITS MAY MIGRATE TO INITIATION SITES ON EUCARYOTIC mRNA 157
LINKAGES BETWEEN PROTEIN SYNTHESIS AND LOCATION 158

PART 3
PRODUCTION OF THE TEMPLATE

CHAPTER 10
RNA POLYMERASES: THE BASIC TRANSCRIPTION APPARATUS 165
WHAT IS RNA POLYMERASE? 166
SUBUNIT STRUCTURE OF BACTERIAL RNA POLYMERASE 167

SIGMA FACTOR CONTROLS BINDING TO DNA 167
THE SIGMA FACTOR CYCLE 169
THE CORE ENZYME SYNTHESIZES RNA 170
FUNCTIONS OF THE CORE SUBUNITS 171
PHAGE RNA POLYMERASES MAY BE "MINIMUM" ENZYMES 172
COMPLEX EUCARYOTIC RNA POLYMERASES 172

CHAPTER 11
PROMOTERS: THE SITES FOR INITIATING TRANSCRIPTION 174
DEFINING THE STARTPOINT *IN VIVO* AND *IN VITRO* 175
BINDING SITES FOR *E. COLI* RNA POLYMERASE 176
SEQUENCE HOMOLOGIES IN *E. COLI* PROMOTERS 179
UP AND DOWN PROMOTER MUTATIONS 181
POINTS OF CONTACT IN THE PROMOTER 183
RECOGNITION AND UNWINDING OF DNA 184
POSITIVE REGULATION AT THE PROMOTER 185
POSSIBLE CONSENSUS SEQUENCES FOR RNA POLYMERASE II 186
IN VITRO AND *"IN VIVO"* SYSTEMS 187
RNA POLYMERASE II FUNCTIONS ACCURATELY *IN VITRO* 189
A DISCREPANCY BETWEEN THE *IN VIVO* AND *IN VITRO* BOUNDARIES 189
A DOWNSTREAM PROMOTER FOR RNA POLYMERASE III 192

CHAPTER 12
GLOBAL SWITCHES IN INITIATION 195
THE SPORULATION LIFE-STYLE 195
ALTERNATIVE PHAGE SIGMA FACTORS 198
EACH SIGMA MAY HAVE ITS OWN -35 AND -10 CONSENSUS SEQUENCES 198
NEW PHAGE RNA POLYMERASES 200

CHAPTER 13
TERMINATION AND ANTITERMINATION 202
TERMINATORS MUST BE IDENTIFIED *IN VITRO* 202
TERMINATION MAY OR MAY NOT NEED THE RHO FACTOR 203
A DIGRESSION ABOUT INVERTED REPEATS 203
CORE ENZYME PAUSES AT PALINDROMES 205
HOW DOES RHO FACTOR WORK? 206
MUTATIONS IN THE RHO FACTOR 209
ANTITERMINATION IS A PHAGE-CONTROL MECHANISM 209
ANTITERMINATION DEPENDS ON SPECIFIC DNA SITES 212
MORE SUBUNITS FOR RNA POLYMERASE? 214
DIFFICULTIES IN EUCARYOTES 216

PART 4
CONTROL OF PROCARYOTIC GENE EXPRESSION

CHAPTER 14
THE OPERON: THE LACTOSE PARADIGM 219
INDUCTION AND REPRESSION ARE CONTROLLED BY SMALL MOLECULES 219
GENE CLUSTERS ARE COORDINATELY REGULATED 220
A REGULATOR GENE CONTROLS THE STRUCTURAL GENES 221
THE CONTROL CIRCUIT OF THE OPERON 222
CONSTITUTIVE MUTATIONS DEFINE REPRESSOR ACTION 223
THE OPERATOR IS *CIS*-DOMINANT 225
UNINDUCIBLE MUTATIONS CAN OCCUR IN THE PROMOTER OR REPRESSOR 226
HOW DOES REPRESSOR BLOCK TRANSCRIPTION? 226
CONTACTS IN THE OPERATOR 228
THE INTERACTION OF REPRESSOR SUBUNITS 230
REPRESSOR AS A DNA-BINDING PROTEIN 231
GETTING OFF DNA 232
STORING SURPLUS REPRESSOR 233
A PARADOX OF INDUCTION 235

CHAPTER 15
CONTROL CIRCUITS: A PANOPLY OF OPERONS 236
DISTINGUISHING POSITIVE AND NEGATIVE CONTROL 236
THE TRYPTOPHAN OPERON IS REPRESSIBLE 238
A MODIFICATION OF COORDINATE REGULATION 240
THE TRYPTOPHAN OPERON IS CONTROLLED BY ATTENUATION 240
ALTERNATIVE SECONDARY STRUCTURES CONTROL ATTENUATION 241
THE GENERALITY OF ATTENUATION 245
REPRESSION CAN OCCUR AT MULTIPLE LOCI 246
THE ARABINOSE OPERON HAS DUAL CONTROL 247
THE CROWDED *ARA* REGULATORY REGION 249
A DOUBLE PROMOTER FOR THE GALACTOSE OPERON 250
CATABOLITE REPRESSION IMPOSES PREFERENCE FOR GLUCOSE 251
AUTOGENOUS CONTROL OF RIBOSOMAL PROTEIN TRANSLATION 252
AUTOGENOUS CONTROL AND MACROMOLECULAR ASSEMBLIES 255
HARD TIMES PROVOKE THE STRINGENT RESPONSE 255

CHAPTER 16
LYTIC CASCADES AND LYSOGENIC REPRESSION 258
THE LYTIC CYCLE HAS DISTINCT STAGES 259
LYTIC DEVELOPMENT IS CONTROLLED BY A CASCADE 259
FUNCTIONAL CLUSTERING IN PHAGES T7 AND T4 261
HOW LAMBDA EXERCISES ITS LYTIC CASCADE 264
LYSOGENY IS MAINTAINED BY AN AUTOGENOUS CIRCUIT 265
REPRESSOR IS A DIMER WITH DISTINCT DOMAINS 268
REPRESSOR BINDS COOPERATIVELY AT EACH OPERATOR 269
HOW IS REPRESSOR SYNTHESIS ESTABLISHED? 272
ANTIREPRESSOR IS NEEDED FOR LYTIC INFECTION 275
A DELICATE BALANCE: LYSOGENY VERSUS LYSIS 275

PART 5
CONSTITUTION OF THE EUCARYOTIC GENOME

CHAPTER 17
EUCARYOTIC GENOMES: A CONTINUUM OF SEQUENCES 281
THE C-VALUE PARADOX DESCRIBES VARIATIONS IN GENOME SIZE 281
REASSOCIATION KINETICS DEPEND ON SEQUENCE COMPLEXITY 283
EUCARYOTIC GENOMES CONTAIN SEVERAL SEQUENCE COMPONENTS 285
NONREPETITIVE DNA COMPLEXITY CAN ESTIMATE GENOME SIZE 286
EUCARYOTIC GENOMES CONTAIN REPETITIVE SEQUENCES 287
MODERATELY REPETITIVE DNA CONSISTS OF MANY DIFFERENT SEQUENCES 288
MEMBERS OF REPETITIVE SEQUENCE FAMILIES ARE RELATED BUT NOT IDENTICAL 289
MODERATELY REPETITIVE DNA IS INTERSPERSED WITH NONREPETITIVE DNA 290

CHAPTER 18
STRUCTURAL GENES: AS REPRESENTED IN mRNA 292
ARE STRUCTURAL GENES UNIQUE OR REPETITIVE? 292
MOST STRUCTURAL GENES LIE IN NONREPETITIVE DNA 293
HOW MANY NONREPETITIVE GENES ARE EXPRESSED? 294
ESTIMATING GENE NUMBERS BY THE KINETICS OF RNA-DRIVEN REACTIONS 295
GENES ARE EXPRESSED AT WIDELY VARYING LEVELS 297
OVERLAPS BETWEEN mRNA POPULATIONS 298

CHAPTER 19
DEALING WITH DNA 300
ANY DNA SEQUENCE CAN BE CLONED IN BACTERIA 300
CONSTRUCTING THE CHIMERIC DNA 303
COPYING mRNA INTO DNA 306
ISOLATING SPECIFIC GENES FROM THE GENOME 307
SHOTGUN CLONING TO FORM LIBRARIES 309
EUCARYOTIC GENES CAN BE TRANSLATED IN BACTERIA 310

CHAPTER 20
STRUCTURAL GENES: INTERNAL ORGANIZATION 312
THE DISCOVERY OF INTERRUPTED GENES 313
VISUALIZING INTERRUPTED GENES BY ELECTRON MICROSCOPY 313
RESTRICTION MAPPING OF INTERRUPTED GENES 316
CHARACTERIZING GENOMIC DNA FRAGMENTS 318
GENES COME IN ALL SHAPES AND SIZES 320
INTRONS IN GENES CODING FOR rRNA AND tRNA 323
INTRONS ARE NONREPETITIVE AND EVOLVE RAPIDLY 323
EXON-INTRON JUNCTIONS HAVE A CONSENSUS SEQUENCE 324
ONE GENE'S INTRON CAN BE ANOTHER GENE'S EXON 325
AN INTRON THAT MAY CODE FOR A REGULATOR PROTEIN 328
HOW HAVE INTERRUPTED GENES EVOLVED? 332

PART 6
CLUSTERS OF RELATED SEQUENCES

CHAPTER 21
STRUCTURAL GENES: EXTERNAL RELATIONSHIPS 337
THE MULTIPLE TYPES OF GLOBIN PROTEINS 338
GLOBIN GENES ARE ORGANIZED IN CLUSTERS 338
UNEQUAL CROSSING-OVER REARRANGES GENE CLUSTERS 340
MANY α THALASSEMIAS RESULT FROM UNEQUAL CROSSING OVER 342
NEW GENES GENERATED IN β THALASSEMIAS 343
GENE CLUSTERS SUFFER CONTINUAL REORGANIZATION 344
AN EVOLUTIONARY TREE FOR GLOBIN GENES 345
SEQUENCE DIVERGENCE INDICATES EVOLUTIONARY SEPARATION 346
TWO TYPES OF DIVERGENCE IN DNA 346
USING THE CLOCK TO FOLLOW THE GLOBIN GENES 347
MECHANISMS FOR MAINTAINING ACTIVE SEQUENCES 349
PSEUDOGENES ARE DEAD ENDS OF EVOLUTION 350

GENE FAMILIES ARE COMMON FOR ABUNDANT PROTEINS 351

CHAPTER 22 GENOMES SEQUESTERED IN ORGANELLES 353
ORGANELLE GENES SHOW NONMENDELIAN INHERITANCE 353
ORGANELLE GENOMES ARE CIRCULAR DNA MOLECULES 355
ORGANELLES EXPRESS THEIR OWN GENES 356
THE LARGE MITOCHONDRIAL GENOME OF YEAST 359
THE COMPACT MITOCHONDRIAL GENOME OF MAMMALS 360
RECOMBINATION OCCURS IN (SOME) ORGANELLE DNAS 362
REARRANGEMENTS OF YEAST MITOCHONDRIAL DNA 363

CHAPTER 23 IDENTITY AND VARIATION IN TANDEM GENE CLUSTERS 365
THE REPETITIVE NATURE OF HISTONE GENES 366
A VARIETY OF TANDEM GENE CLUSTERS FOR HISTONES 367
GENES FOR rRNA AND tRNA ARE REPEATED 368
A TANDEM REPEATING UNIT CONTAINS BOTH rRNA GENES 369
SOME rRNA GENES ARE EXTRACHROMOSOMAL 371
ABOUT NONTRANSCRIBED SPACERS AND PROMOTERS 372
5S GENES AND PSEUDOGENES ARE INTERSPERSED 373
AN EVOLUTIONARY DILEMMA 374
BACTERIAL rRNA GENES FORM MIXED OPERONS WITH tRNA GENES 375
tRNA GENES MAY LIE IN CLUSTERS 376

CHAPTER 24 ORGANIZATION OF SIMPLE SEQUENCE DNA 378
THE ALU FAMILY 378
INVERTED REPEATS RENATURE INSTANTLY 379
HIGHLY REPETITIVE DNA FORMS SATELLITES 381
SATELLITE DNA OFTEN LIES IN HETEROCHROMATIN 382
ARTHROPOD SATELLITES HAVE VERY SHORT IDENTICAL REPEATS 383
MAMMALIAN SATELLITES CONSIST OF HIERARCHICAL REPEATS 384
RECONSTRUCTING THE STAGES OF MOUSE SATELLITE DNA EVOLUTION 387
VARIATIONS IN THE PRESENT REPEATING UNIT 388
THE CONSEQUENCES OF UNEQUAL CROSSING-OVER 389
CROSSOVER FIXATION COULD MAINTAIN IDENTICAL REPEATS 390

PART 7 REACHING MATURITY: RNA PROCESSING

CHAPTER 25 CUTTING AND TRIMMING STABLE RNA 395
PHOSPHODIESTER BONDS CAN BE CLEAVED ON EITHER SIDE 396
RNAase III RELEASES THE PHAGE T7 EARLY mRNAS 397
RNAase III SEPARATES rRNAS FROM THEIR PRECURSOR 399
CLEAVAGE SITES IN THE PATHWAY FOR EUCARYOTIC rRNA RELEASE 401
tRNAS ARE CUT AND TRIMMED BY SEVERAL ENZYMES 402

CHAPTER 26 MECHANISMS OF RNA SPLICING 405
YEAST tRNA SPLICING INVOLVES CUTTING AND REJOINING 405
THE EXTRAORDINARY SPLICING OF TETRAHYMENA rRNA 408
NUCLEAR RNA SPLICING FOLLOWS PREFERRED PATHWAYS 410
SPLICING JUNCTIONS MAY BE INTERCHANGEABLE 412
MUTATIONS IN CONSENSUS SEQUENCES CAN AFFECT SPLICING 415
IS snRNA INVOLVED IN SPLICING? 416

CHAPTER 27 CONTROL OF RNA PROCESSING 420
hnRNA IS LARGE AND UNSTABLE 420
mRNA IS DERIVED FROM hnRNA 422
THE IMPORTANCE OF POLYADENYLATION 424
hnRNA IS MORE COMPLEX THAN mRNA 425
IS THERE CONTROL AFTER TRANSCRIPTION? 426
MODELS FOR CONTROLLING GENE EXPRESSION 428
THE POTENTIAL OF CELLULAR POLYPROTEINS 431

PART 8 THE PACKAGING OF DNA

CHAPTER 28 ABOUT GENOMES AND CHROMOSOMES 437
CONDENSING VIRAL GENOMES INTO THEIR COATS 438
THE BACTERIAL GENOME IS A FOLDED NUCLEOID 441
THE NUCLEOID CONTAINS MANY SUPERCOILED LOOPS 443

THE CONTRAST BETWEEN INTERPHASE CHROMATIN AND MITOTIC CHROMOSOMES 444
THE EUCARYOTIC CHROMOSOME AS A SEGREGATION DEVICE 447
THE EXTENDED STATE OF LAMPBRUSH CHROMOSOMES 450
POLYTENY FORMS GIANT CHROMOSOMES 452
TRANSCRIPTION DISRUPTS THE CHROMOSOME STRUCTURE 454

CHAPTER 29
NUCLEOSOME PARTICLES AND THE STRUCTURE OF CHROMATIN 456
THE PROTEIN COMPONENTS OF CHROMATIN 457
CHROMATIN CONTAINS DISCRETE PARTICLES 458
THE NUCLEOSOME IS THE BASIC SUBUNIT OF ALL CHROMATIN 460
THE CORE PARTICLE IS HIGHLY CONSERVED 461
DNA IS COILED AROUND THE HISTONE OCTAMER 463
DNA IS SYMMETRICALLY EXPOSED TO NUCLEASES 465
THE UNRESOLVED QUESTION OF THE PERIODICITY OF DNA 468
THE ARRANGEMENT OF HISTONES AND DNA 469
NUCLEOSOME ASSEMBLY VERSUS CHROMATIN REPRODUCTION 470
NUCLEOSOME ASSEMBLY REQUIRES NONHISTONE PROTEINS 473
THE PATH OF NUCLEOSOMES IN THE CHROMATIN FIBER 474
LOOPS, DOMAINS, AND SCAFFOLDS 478

CHAPTER 30
NUCLEOSOMES IN ACTIVE CHROMATIN 479
ARE NUCLEOSOMES ARRANGED IN PHASE? 480
THE SPECIFICITY OF MICROCOCCAL NUCLEASE 482
ARE TRANSCRIBED GENES ORGANIZED IN NUCLEOSOMES? 482
THE DNAase-SENSITIVE DOMAINS OF TRANSCRIBABLE CHROMATIN 485
NONHISTONE PROTEINS CONFER DNAase SENSITIVITY 487
HISTONES SUFFER TRANSIENT MODIFICATIONS 489
H2A IS CONJUGATED WITH UBIQUITIN ON A SUBSET OF NUCLEOSOMES 491
GENE EXPRESSION IS ASSOCIATED WITH DEMETHYLATION 492
SOME MODELS FOR THE CONTROL OF METHYLATION 493
DNAase HYPERSENSITIVE SITES LIE UPSTREAM FROM ACTIVE PROMOTERS 494
NUCLEASE-SENSITIVE AND PROTECTED REGIONS 497

PART 9
PERPETUATION OF DNA

CHAPTER 31
THE REPLICON: UNIT OF REPLICATION 503
DNA SYNTHESIS IS SEQUENTIAL AND SEMICONSERVATIVE 504
THE BACTERIAL GENOME IS A SINGLE REPLICON 507
CONNECTIONS BETWEEN DNA REPLICATION AND CELL DIVISION 509
EACH EUCARYOTIC CHROMOSOME CONTAINS MANY REPLICONS 511
ISOLATING THE ORIGINS OF YEAST REPLICONS 513
REPLICATION CAN PROCEED THROUGH EYES, ROLLING CIRCLES, OR D LOOPS 515
PLASMID INCOMPATIBILITY IS CONNECTED WITH COPY NUMBER 518

CHAPTER 32
THE TOPOLOGY OF DNA REPLICATION 521
DESCRIBING THE TOPOLOGY OF DNA 522
TOPOLOGICAL MANIPULATION OF DNA 523
GYRASE INTRODUCES NEGATIVE SUPERCOILS IN DNA 526
EUCARYOTIC DNA POLYMERASES 527
PROCARYOTIC DNA POLYMERASES HAVE SEVERAL ENZYMATIC ACTIVITIES 528
DNA SYNTHESIS IS SEMIDISCONTINUOUS 531
OKAZAKI FRAGMENTS ARE PRIMED BY RNA 533

CHAPTER 33
THE ENZYMATIC APPARATUS FOR DNA REPLICATION 537
THE COMPLEXITY OF THE BACTERIAL REPLICATION APPARATUS 537
INITIATING SYNTHESIS OF A SINGLE DNA STRAND 538
MOVEMENT OF THE PRIMOSOME 541
INITIATING REPLICATION AT DUPLEX ORIGINS 544
THE REPLICATION APPARATUS OF PHAGE T4 547
THE REPLICATION APPARATUS OF PHAGE T7 549
THE PROBLEM OF LINEAR REPLICONS 550

CHAPTER 34
SYSTEMS THAT SAFEGUARD DNA 553
THE OPERATION OF RESTRICTION AND MODIFICATION 554
THE ALTERNATE ACTIVITIES OF TYPE I ENZYMES 556
THE DUAL ACTIVITIES OF TYPE III ENZYMES 559
DEALING WITH INJURIES IN DNA 560
EXCISION-REPAIR SYSTEMS IN *E. COLI* 563

RECOMBINATION-REPAIR SYSTEMS IN *E. COLI* 565
AN SOS SYSTEM OF MANY GENES 566
MAMMALIAN REPAIR SYSTEMS 568

CHAPTER 35
RETRIEVAL AND RECOMBINATION OF DNA 570
RECOMBINATION REQUIRES SYNAPSIS OF HOMOLOGOUS DUPLEX DNAS 571
BREAKAGE AND REUNION OCCURS VIA HETERODUPLEX DNA 572
ISOLATION OF RECOMBINATION INTERMEDIATES 574
THE STRAND-EXCHANGE FACILITY OF RecA 577
RecA AND THE CONDITIONS OF RECOMBINATION 579
GENE CONVERSION ACCOUNTS FOR INTERALLELIC RECOMBINATION 581
SPECIALIZED RECOMBINATION RECOGNIZES SPECIFIC SITES 583
STAGGERED BREAKAGE AND REUNION IN THE CORE 584

PART 10
THE DYNAMIC GENOME: DNA IN FLUX

CHAPTER 36
TRANSPOSABLE ELEMENTS IN BACTERIA 589
THE DISCOVERY OF TRANSPOSITION IN BACTERIA 590
INSERTION SEQUENCES ARE BASIC TRANSPOSONS 591
COMPOSITE TRANSPOSONS HAVE *IS* MODULES 593
ONLY ONE MODULE OF TN10 IS FUNCTIONAL 595
THE MODULES OF TN5 ARE ALMOST IDENTICAL BUT VERY DIFFERENT 596
TRANSPOSITION INVOLVES REPLICATIVE RECOMBINATION 598
TRANSPOSITION OF TN3 PROCEEDS BY COINTEGRATE RESOLUTION 600
SOME UNUSUAL FEATURES OF TRANSPOSING PHAGE MU 604
SALMONELLA PHASE VARIATION OCCURS BY INVERSION 606

CHAPTER 37
MOBILE ELEMENTS IN EUCARYOTES 609
YEAST TY ELEMENTS RESEMBLE BACTERIAL TRANSPOSONS 610
SEVERAL TYPES OF TRANSPOSABLE ELEMENTS RESIDE IN *D. MELANOGASTER* 611
ABOUT COMPLEX LOCI AND CHROMOSOME WALKING 614
INSERTIONS AT THE *W* LOCUS DEFINE A COMPLEX TARGET 616
THE ROLE OF TRANSPOSABLE ELEMENTS IN HYBRID DYSGENESIS 619
CONTROLLING ELEMENTS IN MAIZE ARE TRANSPOSABLE 620
DS MAY TRANSPOSE OR CAUSE CHROMOSOME BREAKAGE 622
DS TRANSPOSITION IS CONNECTED WITH REPLICATION 624
YEAST HAS SILENT AND ACTIVE LOCI FOR MATING TYPE 625
SILENT AND ACTIVE CASSETTES HAVE THE SAME SEQUENCES 628
UNIDIRECTIONAL TRANSPOSITION IS INITIATED BY THE RECIPIENT (*MAT*) LOCUS 630

CHAPTER 38
ELEMENTS THAT MOVE IN AND OUT OF THE GENOME 632
THE RETROVIRUS LIFE CYCLE INVOLVES TRANSPOSITION-LIKE EVENTS 633
RETROVIRUSES MAY TRANSDUCE CELLULAR SEQUENCES 635
RNA-DEPENDENT TRANSPOSITIONS MAY HAVE OCCURRED IN THE CELL 637
TISSUE-SPECIFIC VARIATIONS OCCUR IN THE *DROSOPHILA* GENOME 638
SELECTION OF AMPLIFIED GENOMIC SEQUENCES 641
EXOGENOUS SEQUENCES CAN BE INTRODUCED BY TRANSFECTION 645
TRANSFECTED DNA CAN ENTER THE GERM LINE 646

CHAPTER 39
GENERATION OF ANTIBODY DIVERSITY 649
IMMUNOGLOBULIN GENES ARE ASSEMBLED FROM THEIR PARTS 652
THE DIVERSITY OF GERM-LINE INFORMATION 654
JOINING REACTIONS GENERATE ADDITIONAL DIVERSITY 656
RECOMBINATION OF V AND C GENES GENERATES DELETIONS AND REARRANGEMENTS 657
SOME POSSIBLE CAUSES OF ALLELIC EXCLUSION 660
FURTHER DNA RECOMBINATION CAUSES CLASS SWITCHING 662
EARLY HEAVY-CHAIN EXPRESSION CAN BE CHANGED BY RNA PROCESSING 663
SOMATIC MUTATION GENERATES ADDITIONAL DIVERSITY 665
COMPLEXITY OF MAJOR HISTOCOMPATIBILITY LOCI 666

GLOSSARY 671

INDEX 687

GENES

BENJAMIN LEWIN
Editor, *Cell*

John Wiley & Sons
New York Chichester Brisbane
Toronto Singapore

PART 1
THE NATURE OF GENETIC INFORMATION

In calling the structure of the chromosome fibers a code-script we mean that the all-penetrating mind could tell from their structure whether the egg would develop, under suitable conditions, into a black cock or into a speckled hen, into a fly or a maize plant, a beetle, a mouse or a woman. . . . But the term code-script is, of course, too narrow. The chromosome structures are at the same time instrumental in bringing about the development they foreshadow. They are law-code and executive power—or, to use another simile, they are architect's plan and builder's craft—in one.

Erwin Schrödinger, 1945

CHAPTER 1
WHAT IS A GENE? A GENETIC VIEW

The concept of the gene has been the focus of some hundred years of work to establish the basis of heredity. It has been scrutinized from so many perspectives that there is no simple one-line answer to satisfy the question: what is a gene?

A gene is a sequence of DNA that carries the information representing a protein. Until very recently this would have been an adequate (if incomplete) biochemical description. The sequence of DNA could be identified as a continuous stretch of nucleotides, related to the protein sequence by a colinear readout. But now it is clear that the sequence representing protein is not always continuous; it may be interrupted by sequences not concerned with specifying the protein. So genes may be in pieces that are put together during the process of gene expression.

A gene can be identified as a cluster of mutations all of which prevent the production of the protein that it represents. Although this remains true, the properties of these clusters become more complex for interrupted genes. This issue has been moot, because most interrupted genes are found in situations in which detailed genetic analysis is not possible; but in principle, the genetic view of the gene now needs modification.

The large number of genes that make up the genome of any species are organized into a comparatively small number of chromosomes. The genetic material of each chromosome consists of an extremely long stretch of DNA, containing many genes in a linear order. How many genes are present *in toto* has been a puzzle for a long time. Recently the view that each gene may reside by itself as a unique entity has been superseded by the realization that, in many cases, there may be clusters of related genes that constitute small families.

The genome generally has been viewed as rather stable, subject to changes in overall constitution and organization only on an extraordinarily slow evolutionary time scale. This contrasts with recent evidence that in some instances there may be rearrangements that occur regularly; and there may be components of the genome that are relatively mobile.

The concept of the gene has therefore undergone an evolution in which, although many of its traditional properties remain, exceptions have been found to show that none constitutes an absolute rule. Starting from the discovery of the gene as a fixed unit of inheritance, its properties were defined in terms of its residence at a definite position on the chromosome, this in turn leading to the view that the genetic material of the chromosome is a continuous length of DNA representing many genes. By tracing the development of these ideas, we can arrive at an operational description of the gene; although it is impractical to provide a capsule definition.

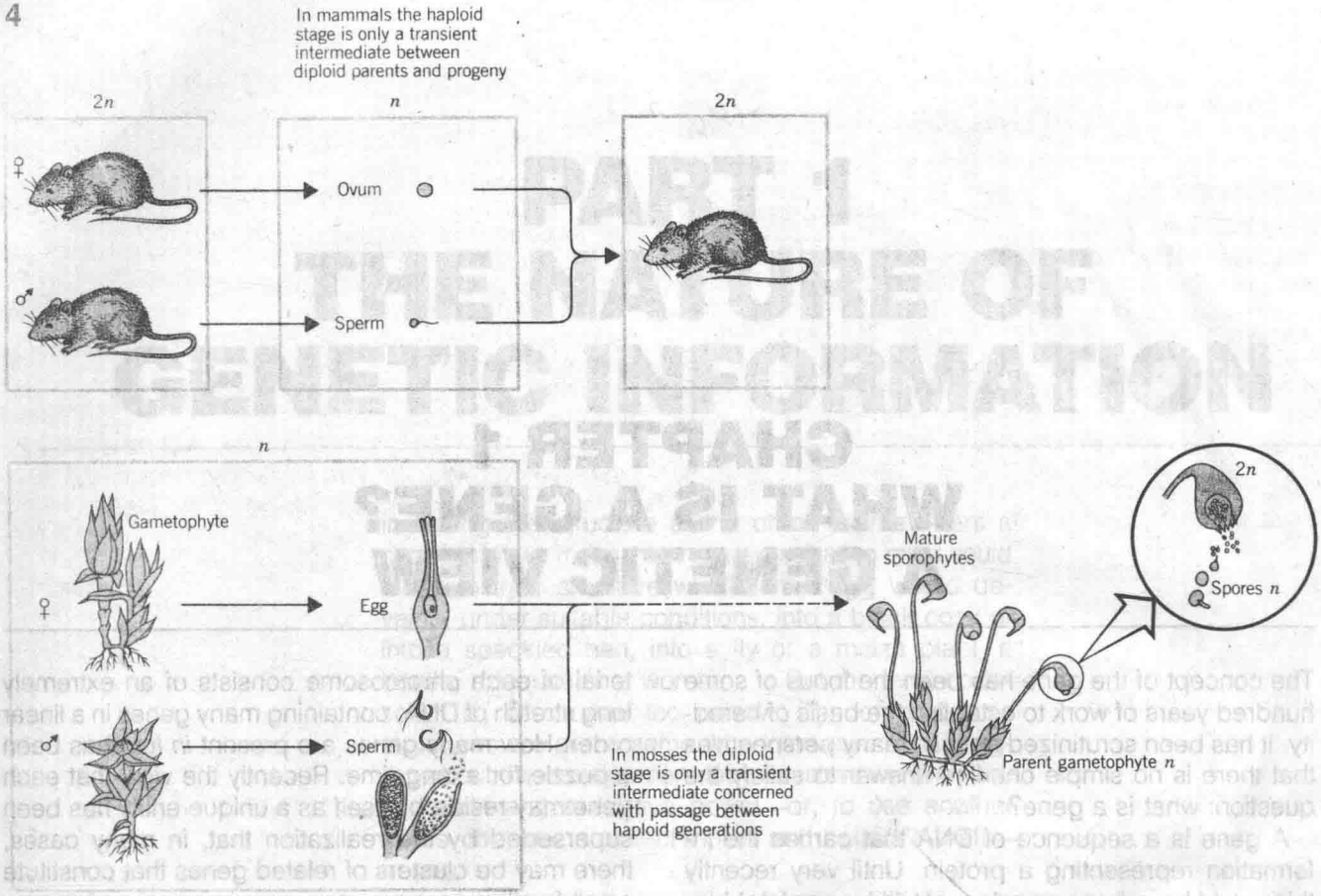

Figure 1.1
When eucaryotes perpetuate their genes through an alternation of diploid and haploid states, either type of state may provide the predominant "adult" type while the other is concerned solely with gamete and zygote formation. The mammals and the mosses are extreme examples in which the adult generations are diploid and haploid, respectively.
Red indicates diploid (2n) tissue; grey indicates haploid (n) tissue.

A PARTICULATE FACTOR OF INHERITANCE

"The constant characters that appear in the group of plants may be obtained in all the associations that are possible according to the mathematical laws of combination. Those characters which are transmitted visibly, and therefore constitute the characters of the hybrid, are termed the dominant, and those that become latent in the process are termed recessive. The expression 'recessive' has been chosen because such characters withdraw or disappear entirely in the hybrids, but nevertheless reappear unchanged in their progeny in predictable proportions, without any essential alterations. Transitional forms were not observed in any experiment."

This (abbreviated) quotation from Mendel's remarkable paper of 1865 introduces the fundamental concept of genetics: there is a unit of heredity consisting of some factor that is passed from parent to progeny. This particulate factor is, of course, what we now know as the gene. Mendel's work elucidated the general behavior of the gene in inheritance. Implicit in this is the concept that the organism is the gene's way of expressing and perpetuating itself, a point illustrated by the alternation of generations shown in **Figure 1.1.**

In striking contrast to the confusion of earlier thought

about heredity, Mendel's work established genetics as an experimental science through the introduction of several critical features. First, the initial set of plants was **true breeding** for the character under consideration: each gave rise to plants showing only that characteristic and none of the alternatives. Then the inheritance of only the one character was initially followed. With this established, the simultaneous behavior of two or more traits could be investigated. The experiments were placed on a quantitative basis by counting the numbers of each progeny type resulting from each mating.

The characteristics of the gene discernible from Mendel's work are summarized in what are known as his first and second laws. The first law deals with the properties of the individual gene. The organism has two copies of each gene: in modern terminology it is said to be **diploid.** Only one of the two copies is passed from parent to offspring through the **gametes** (sex cells). Thus when the gametes unite, the **zygote** (fertilized egg) gains one copy from each parent, restoring the situation in which every organism has one copy of paternal origin and one of maternal origin.

A gene may exist in alternative forms that result in the expression of a different characteristic (such as red versus white flower color). These forms are called **alleles.** The *law of independent segregation* states that these alleles do not affect each other when present in the same plant, but segregate unchanged by passing into different gametes when the next generation forms.

In a true-breeding organism, a **homozygote,** both alleles are the *same*. But a mating between two parents each of which is homozygous for a *different* allele generates a hybrid or **heterozygote.** If one allele is **dominant** and the other is **recessive,** the organism will have the appearance or **phenotype** only of the dominant type (so the heterozygote is indistinguishable from the true-breeding dominant parent).

But Mendel's first law recognizes that the genetic constitution or **genotype** of the hybrid comprises the presence of both alleles. This is revealed, as shown in **Figure 1.2,** when the hybrid is crossed with another hybrid to form the second generation. The critical point is that the alleles do not mix ("no transitional forms were observed"), but are physical entities whose interaction is at the level of expression.

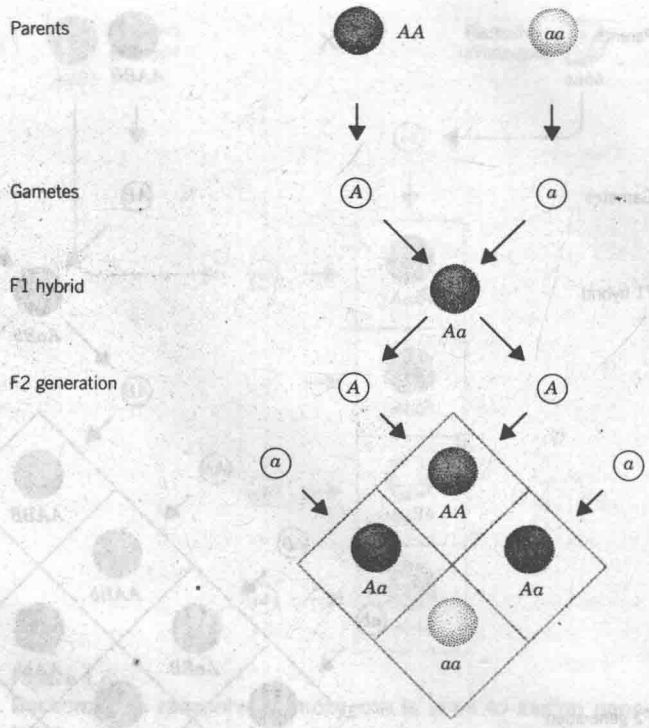

Figure 1.2
Alleles show independent segregation.

The two parents are homozygous: *AA* has two copies of the dominant allele; *aa* has two copies of the recessive allele. Each forms only one type of gamete, so that the **F1** (first hybrid generation) is uniformly hybrid as *Aa*. Because *A* is dominant over *a*, the *phenotype* of *Aa* is the same as that of *AA* (indicated by the color). The phenotype of the recessive homozygote *aa* is indicated by the lack of color.

Each F1 hybrid forms both *A* and *a* gametes in equal amounts. Upon mating these unite randomly to generate an **F2** (second hybrid generation) consisting of: 1 *AA* : 2 *Aa* : 1 *aa*. Since the *AA* and *Aa* have the same phenotype, this gives the classic 3 : 1 ratio of dominant : recessive types.

In cases in which the heterozygote *Aa* has a phenotype intermediate between the parental *AA* and *aa*, the F1 would be distinct, and in the F2 the ratio of phenotypes would be 1 dominant : 2 intermediate : 1 recessive.

Although the characteristics studied by Mendel (fortunately) showed complete dominance, this is not necessarily always the case. Alleles may exhibit **incomplete (partial) dominance** or no dominance (sometimes known as **codominance**). In the latter case, the heterozygote is distinguished from the homozy-

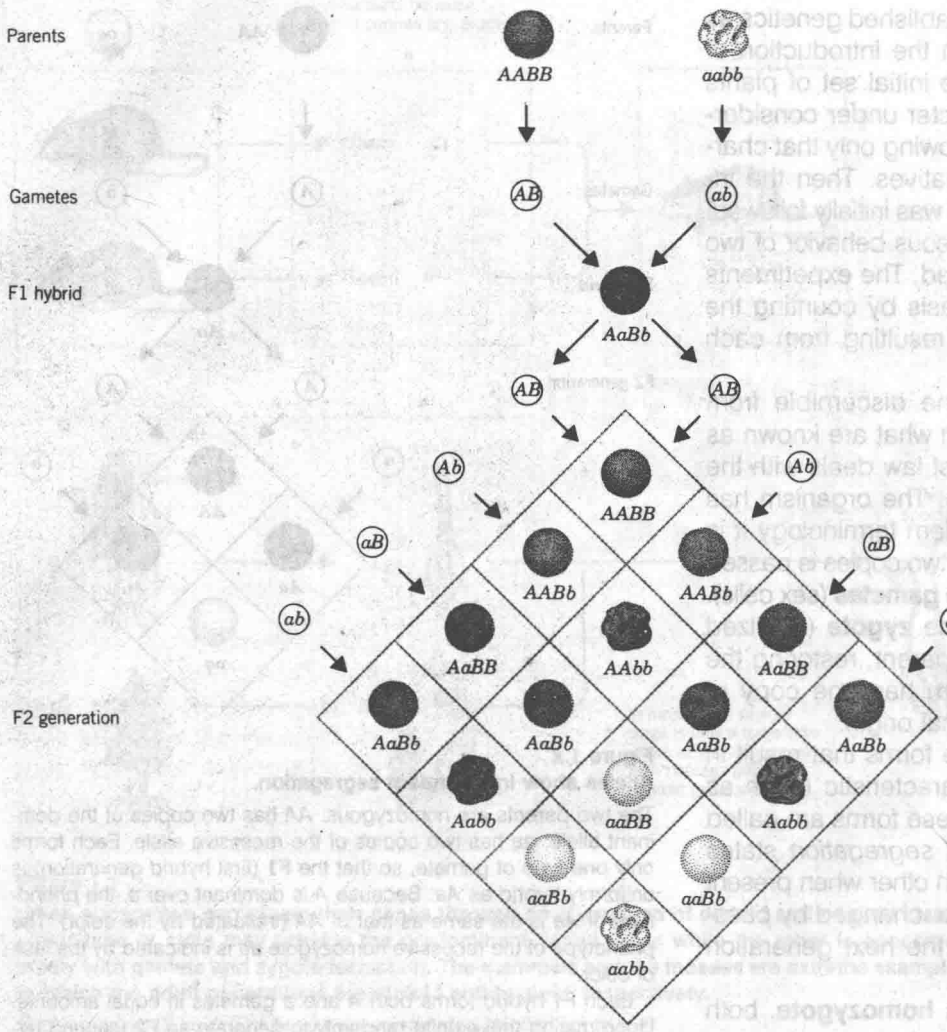

Figure 1.3
Different genes assort independently.

One parent is homozygous for two dominant genes, *A* determining color, and *B* determining shape (shown by round structure). The other parent is homozygous for the recessive alleles *a* and *b* (characteristics shown by no color and wrinkled shape). The F1 is uniform with the dominant characteristics.

The F1 parents produce gametes in which there is independent segregation of alleles and independent assortment of genes, so that equal amounts are produced of each of the four possible types of gamete. These unite randomly to form 9 genotypic classes, which because of the dominance relationships appear as the four phenotypic classes: 9 colored-smooth : 3 colored-wrinkled : 3 noncolored-smooth : 1 noncolored-wrinkled.

Note that the same numbers are present of each reciprocal genotype; for example, the two parents (one each of *AABB* and *aabb*) or the recombinant classes (one each of *AAbb* and *aaBB*). The 3 : 1 ratios are maintained for each individual segregating character.

The number of phenotypic classes will be greater if one or both of the characters is not dominant (so that heterozygotes appear different from either homozygote). It will be less if two genes affect a single characteristic. So if both *A* and *B* were needed for the production of color, a ratio of 9 colored : 7 noncolored would be seen in the F2.