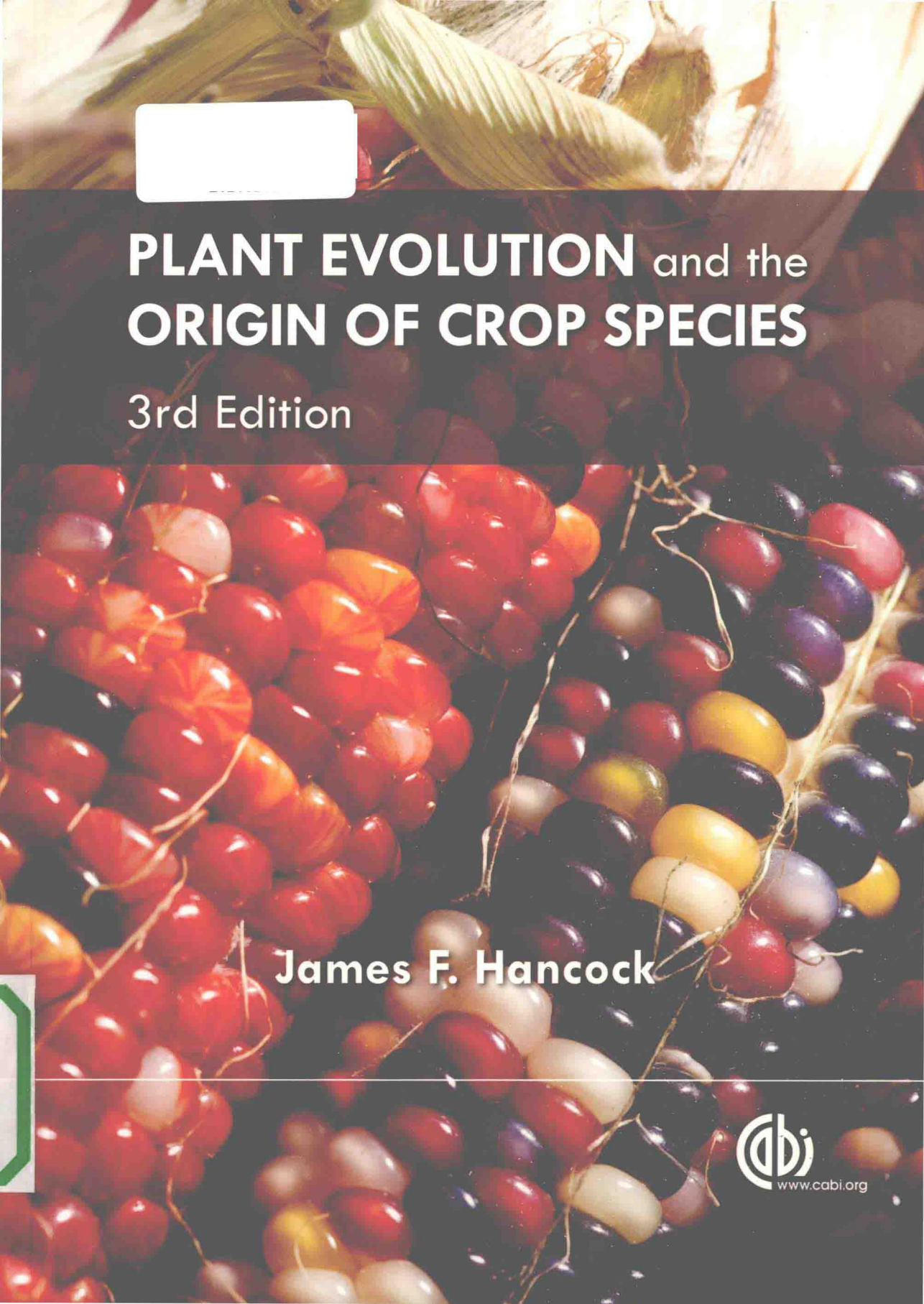
PLANT EVOLUTION and the
ORIGIN OF CROP SPECIES
3rd Edition
James F. Hancock
www.cabi.org

Plant Evolution and the Origin of Crop Species, 3rd Edition

James F. Hancock

Michigan State University, USA

www.cabi.org

CABI is a trading name of CAB International

CABI Head Office
Nosworthy Way
Wallingford
Oxfordshire OX10 8DE
UK

Tel: +44 (0)1491 832111
Fax: +44 (0)1491 833508
Email: cabi@cabi.org
Web site: www.cabi.org

CABI North American Office
38 Chauncey St
Suite 1002
Boston, MA 02111
USA

Tel: +1 617 395 4056
Fax: +1 617 354 6875
Email: cabi-nao@cabi.org

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

Library of Congress Cataloging-in-Publication Data
Hancock, James F.
Plant evolution and the origin of crop species / James F. Hancock.—3rd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84593-801-7 (alk.paper)
1. Crops--Evolution. 2. Crops--Origin. 3. Plants--Evolution. I. Title.

SB106.O74H36 2012
633--dc23

2012001017

ISBN-13: 978-1-84593-801-7 (HB)
ISBN-13: 978-1-78064-477-6 (PB)

First published (HB) 2012
First paperback edition 2014

Commissioning editor: Meredith Carroll
Editorial assistants: Gwenan Spearing and Alexandra Lainsbury
Production editor: Tracy Head

Artwork provided by Marlene Cameron.
Typeset by SPi, Pondicherry, India.
Printed and bound in the UK by CPI Group (UK) Ltd, Croydon, CR0 4YY

Plant Evolution and the Origin of Crop Species, 3rd Edition

Preface

It has been almost 20 years since the first edition of this book was published, and the amount of information available on plant evolution and crop origins has skyrocketed, particularly in the molecular arena. Where the molecular technologies have had the greatest impact is in tracing crop origins – determining parental species, finding out where crops were domesticated and identifying the genes that made the domestications possible.

Molecular studies have also played a key role in testing long-standing evolutionary theories. Most of the hypotheses of Anderson, Grant, Heiser and Stebbins have proven robust, even as experimental technologies have dramatically shifted from morphological and cytogenetic comparisons to gene chips and genome sequencing. Hypotheses concerning hybrid speciation, polyploidy and the role of chromosome rearrangements have been tested, retested and expanded as each new technology has emerged. The molecular studies have generated some surprises, such as the amount of chromosome repatterning and gene expression alterations associated with interspecific hybridization, and the high level of gene duplication found in all sequenced genomes, but in general each new molecular technology has supported the evolutionary hypothesis formulated decades ago.

In this book I have tried to use, whenever possible, the historical evidence of evolutionary phenomena and crop origins. This may give the book an "old fashioned" feel, but I think that the information is still relevant and should not be lost in a cloud of new technologies. I have chosen to work from the past up to the present, rather than the reverse. It is likely that the growing wealth of molecular information will necessitate an alternative approach in any future editions, but so far I think that I can hold my ground.

Acknowledgments

I would like to give special thanks to my wife Ann, who has so richly broadened my horizons with her passion and commitment to the living world around us. I would also like to thank the current crop of evolutionary biologists that I most admire: Norm Ellstrand, Loren Rieseberg, Tao Sang, Doug Schemske, Doug Soltis and Jonathan Wendel.

Contents

1

Chromosome Structure and Genetic Variability

Introduction

Evolution is the force that shapes our living world. Countless different kinds of plants and animals pack the earth and each species is itself composed of a wide range of morphologies and adaptations. These species are continually being modified as they face the realities of their particular environments.

In its simplest sense, evolution can be defined as a change in gene frequency over time. Genetic variability is produced by mutation and then that variability is shuffled and sorted by the various evolutionary forces. The way organisms evolve is dependent on their genetic characteristics and the type of environment they must face.

A broad spectrum of evolutionary forces, including migration, selection and random chance, alter natural species. These same forces operated during the domestication of crops, as will be discussed in Chapter 6. In the next four chapters, we will describe these parameters and how they interact; but before we do this, we will begin by discussing the different types of genetic variability that are found in plants. The primary requirement for evolutionary change is genetic variability, and mutation generates these building blocks. A wide range of mutations can occur at all levels of genetic organization from nucleotide sequence to chromosome structure. In this chapter, we will discuss how plant genes are organized in chromosomes, and then we will discuss the kinds of genetic variability present and its measurement.

Gene and Chromosome Structure

Both gene and chromosome structure are complex. Genes are composed of coding regions called exons and non-coding regions called introns. Both the introns and the exons are transcribed, but the introns are removed from the final RNA product before translation (Fig. 1.1). There are also short DNA sequences, both near and far from the coding region, that regulate transcription but are not transcribed themselves. "Promotor" sequences are found immediately before the protein coding region and play a role in the initiation of transcription, while "enhancer" sequences are often located distant from the coding region and regulate levels of transcription.

Each chromosome contains not only genes and regulatory sequences, but also a large number of short, repetitive sequences. Some of these are concentrated near centromeres in the densely stained portions of chromosomes called heterochromatic regions, and may play a role in the homologous pairing of chromosomes and their separation. However, there are numerous other repeating units that are more freely dispersed over chromosomes and do not appear

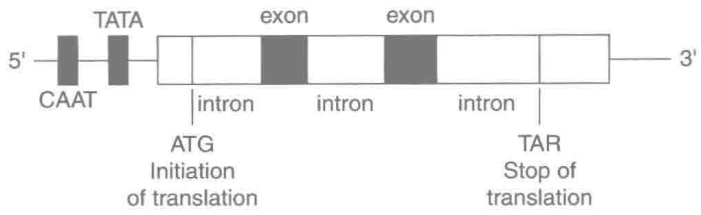

Fig. 1.1. Organization of a typical eukaryotic gene. A precursor RNA molecule is produced from which the introns are excised and the exons are spliced together before translation. The CAAT and TATA boxes play a role in transcription initiation and enhancement.

to have a functional role. These have been described by some as "selfish" or "parasitic", as their presence may stimulate further accumulation of similar sequences through transposition, a topic we will discuss more fully later.

The overall amount of DNA in nuclei can vary dramatically between taxonomic groups and even within species (Wendel *et al.*, 2002; Bennetzen *et al.*, 2005). The total DNA content of nuclei is commonly referred to as the genome. There is a 100-fold variation in genome size among all diploid angiosperms, and congeneric species vary commonly by threefold (Price *et al.*, 1986; Price, 1988). In some cases, genomic amplification occurs in a breeding population in response to environmental or developmental perturbations (Walbot and Cullis, 1985; Cullis, 1987). Most of these differences occur in the quantity of repetitive DNA and not in unique sequences.

Except in the very small genome of *Arabidopsis thaliana* (Barakat *et al.*, 1998), it appears that genes are generally found near the ends of chromosomes in clusters between various kinds of repeated sequences (Schmidt and Heslop-Harrison, 1998; Heslop-Harrison, 2000). The amount of interspersed repetitive DNA can be considerable, making the physical distances between similar loci highly variable across species. However, the gene clusters may be "hotspots" for recombination, making recombination-based genetic lengths much closer than physical distances (Mézard *et al.*, 2006).

Types of Mutation

There are four major types of mutation: (i) point mutations; (ii) chromosomal sequence alterations; (iii) chromosomal additions and deletions; and (iv) chromosomal number changes. Point mutations arise when nucleotides are altered or substituted, for example, when the base sequence CTT becomes GTT. Chromosomal sequence alterations occur when the order of nucleotides is changed within a chromosome. Chromosomal duplications and deletions are produced when portions of chromosomes are added or subtracted. Chromosomal numerical changes arise when the number of chromosomes changes.

Point mutations

Nucleotide changes occur spontaneously due to errors in replication and repair at an average rate of 1×10^{-6} to 10^{-7}. These estimates have come largely from unicellular organisms such as bacteria and yeast, which are easy to manipulate and have tremendous population sizes, but good estimates have also been obtained in higher plants using enzymes (Kahler *et al.*, 1984) and a variety of seed traits (Table 1.1). Mutation rates can be increased by numerous environmental agents such as ionizing radiation, chemical mutagens and thermal shock.

Sequence alterations

Three types of DNA sequence alterations occur: translocations, inversions and transpositions. Small numbers of redundant nucleotide blocks may be involved or whole groups of genes. Translocations occur when nucleotide sequences are transferred from one chromosome to another. In homozygous individuals, nuclear translocations have no effect on fertility, but in heterozygous individuals only a portion of

Table 1.1. Spontaneous mutation rates of several endosperm genes in *Zea mays* (Stadler, 1942).

Gene	Character	No. of gametes tested	Mutation rate
R	Aleurone color	554,786	0.00049
I	Color inhibitor	265,391	0.00011
Pr	Purple color	647,102	0.000011
Su	Sugary endosperm	1,678,736	0.000002
C	Aleurone color	426,923	0.000002
Y	Yellow seeds	1,745,280	0.000002
Sh	Shrunken seeds	2,469,285	0.000001
Wx	Waxy starch	1,503,744	<0.000001

the gametes are viable due to duplications and deficiencies (Fig. 1.2). Translocations are widespread in a number of plant genera, including *Arachis, Brassica, Clarkia, Campanula, Capsicum, Crepis, Datura, Elymus, Galeopsis, Gossypium, Hordeum, Paeonia, Gilia, Layia, Madia, Nicotiana, Secale, Trillium* and *Triticum*.

Populations are generally fixed for one chromosomal type, but not all. Translocation heterozygotes are common in *Paeonia brownii* (Grant, 1975), *Chrysanthemum carinatum* (Rana and Jain, 1965), *Isotoma petraea* (James, 1965), and numerous species of *Clarkia* (Snow, 1960). Probably the most extreme example of translocation heterozygosity is in *Oenothera biennis* where all of its nuclear chromosomes contain translocations and a complete ring of chromosomes is formed at meiosis in heterozygous individuals (Cleland, 1972). In some cases, translocations have resulted in the fusion and fission of non-homologous chromosomes with short arms ("Robertsonian translocations").

Inversions result when blocks of nucleotides rotate 180°. Nuclear inversions are called *pericentric* when the rotation includes the centromere and *paracentric* when the centromeric region remains unaffected (Figs 1.3 and 1.4). Like translocations, individuals that are heterozygous produce numerous inviable gametes, but only if there is a crossover between chromatids; all the gametes of homozygotes are fertile regardless of crossovers.

Inversion polymorphisms have been described in a number of plant genera. One of the best documented cases of an inversion heterozygosity within a species is in *Paeonia californica*, where heterozygous plants are common throughout the northern range of the species (Walters, 1952). As we will describe more fully in the chapter on speciation, inversions on six chromosomes distinguish *Helianthus annuus, Helianthus petiolaris* and their hybrid derivative *Helianthus anomalus* (Rieseberg *et al.*, 1995). Tomato and potato vary by five inversions (Tanksley *et al.*, 1992), and pepper and tomato by 12 inversions (Livingstone *et al.*, 1999). The chloroplast genome of most angiosperm species has a large inverted repeat (Fig. 1.5), but its structure is highly conserved across families. Only a few species do not have the repeat and no intra-populational variation has been described (Palmer, 1985).

Transposition occurs when nucleotide blocks move from place to place in the genome (Bennetzen, 2000a; Fedoroff, 2000). Fragments from multiple chromosomal loci can even be fused to form new open reading frames (Jiang *et al.*, 2004; Hanada *et al.*, 2009). There are two major classes of transposons: DNA and RNA transposable elements. The RNA transposable elements (retroelements) amplify via RNA intermediates, while the DNA transposons rely on actual excision and reinsertment. Both classes of transpositions have been found in all plant species where detailed genetic analysis has been performed, and in many plant species, mobile elements actually make up the majority of the nuclear genome (SanMiguel and Bennetzen, 1998; Bennetzen, 2000a). Most of the transposons are inserted into non-coding regions, but sometimes they enter exons and when they do, they can have extreme effects on phenotype. The wrinkled-seed character described by Mendel is caused by a transpose-like insertion into the gene encoding a starch branching enzyme (Bhattacharyya *et al.*, 1990). Much of the flower color variation observed in the morning glory is due to the insertion and deletion of transposable elements (Clegg and Durbin, 2000; Durbin *et al.*, 2001).

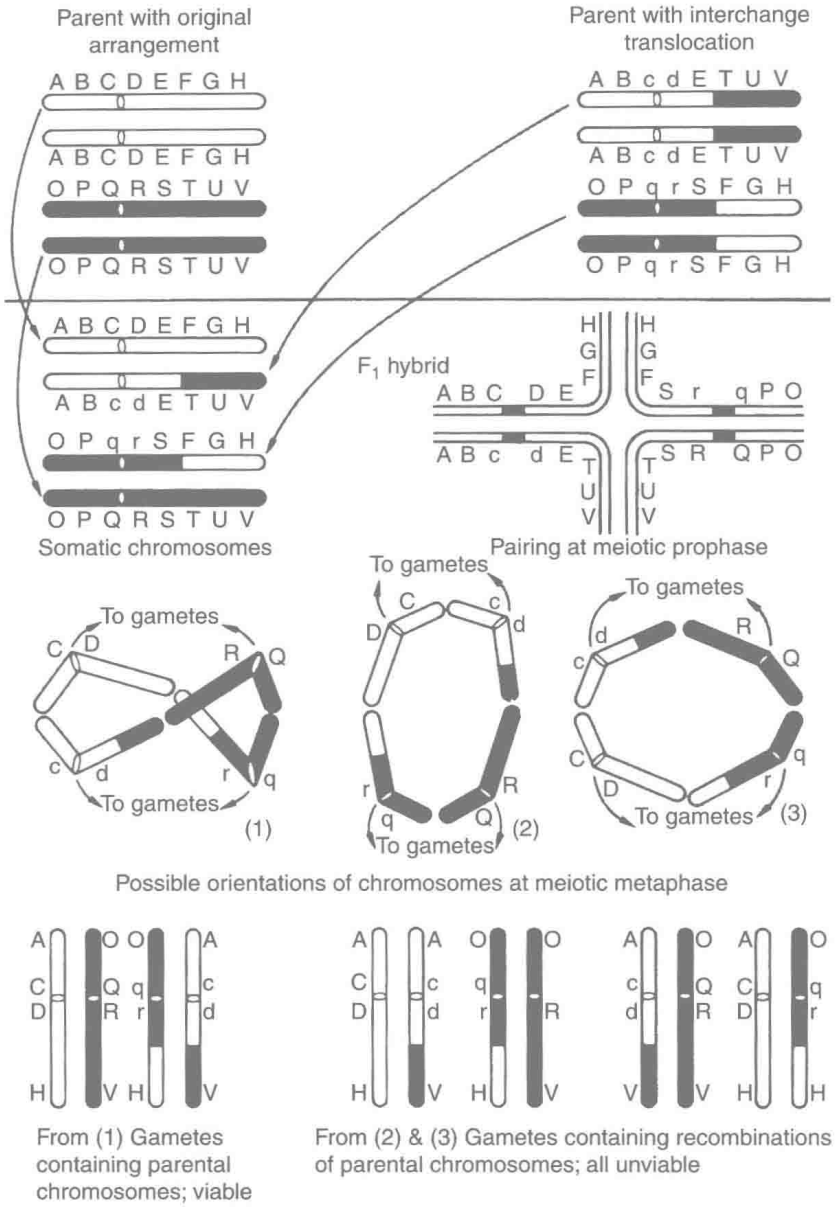

Fig. 1.2. Types of gametes produced by a plant heterozygous for a translocation. A ring of chromosomes is formed at meiosis and depending on how the chromosomes orient at metaphase and separate during anaphase, viable or inviable combinations of genes are produced. (Used with permission from T. Dobzhansky, © 1970, *Genetics of the Evolutionary Process*, Columbia University Press, New York.)

The DNA transposons range in size from a few hundred bases to 10 kb, and the most complex members are capable of autonomous excision, reattachment and alteration of gene expression. They all have short terminal inverted repeats (TIRs); the most complex ones encode an enzyme called transposase that recognizes the families' TIR and performs the excision and reattachment.

Retroelements (RNA transposons) are the most abundant class of transposons and they make up the majority of most large plant genomes. They transpose through reverse-transcription of RNA intermediates, and as a result they do not excise when they transpose, resulting in amplification. The most abundant class of retroelements in plants are the long

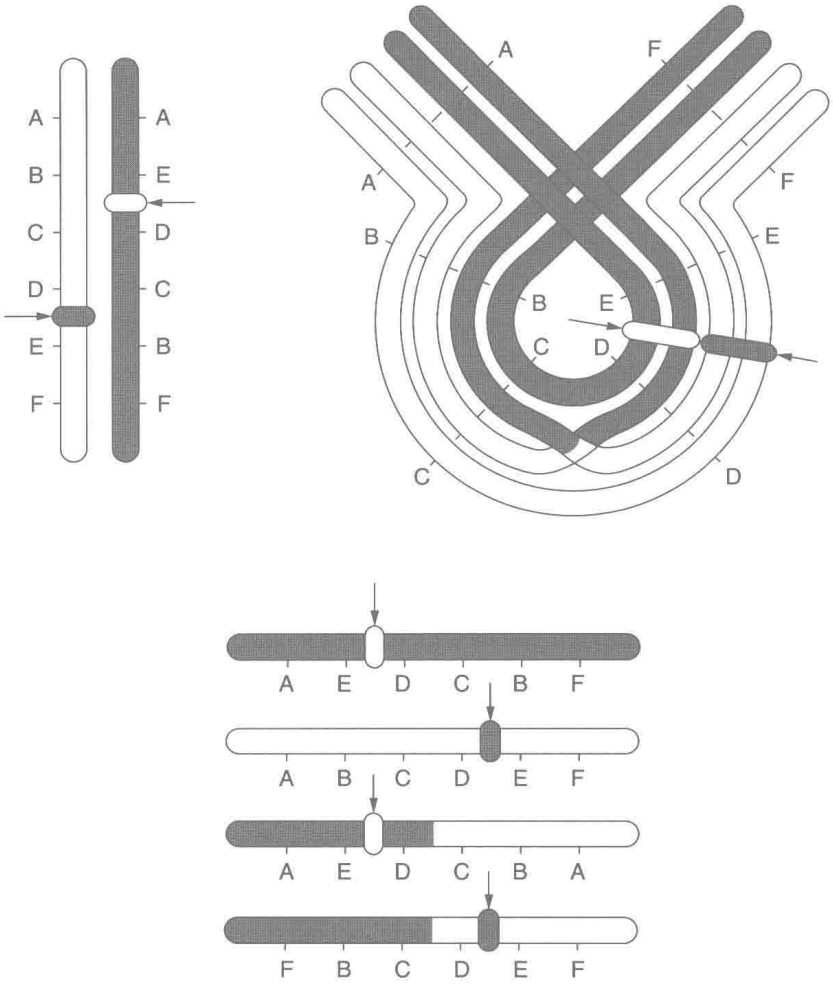

Fig. 1.3. Chromosome types produced after crossing over in an individual heterozygous for a pericentric inversion. Note the two abnormal chromatids, one with a duplication and the other with a deficiency. (Used with permission from T. Dobzhansky, © 1970, *Genetics of the Evolutionary Process*, Columbia University Press, New York.)

terminal repeat (LTR)-retrotransposons, varying in size from a few hundred to over 10,000 nucleotides.

Duplications and deficiencies

Chromosomal deficiencies occur when nucleotide blocks are lost from within a chromosome, while duplications arise when nucleotide sequences are multiplied. These are caused by unequal crossing-over at meiosis or translocation (Burnham, 1962). They also can occur when DNA strands mispair during replication of previously duplicated sequences (Levinson and Gutman, 1987). As previously mentioned, the genome is filled with high numbers of short, repeated sequences that vary greatly in length. So great, in fact, that some of them such as SSRs (simple sequence repeats) have proven valuable as molecular markers to distinguish species, populations and even individuals. Gene amplifications are so common that Wendel (2000) has suggested that "one generalization that has been confirmed and extended by the data emerging from the global thrust in genome sequencing and mapping is that most 'single-copy' genes belong to larger gene families, even in putatively diploid organisms". As more and more crop genomes are sequenced, it is becoming very clear that gene duplications are widespread. The fraction

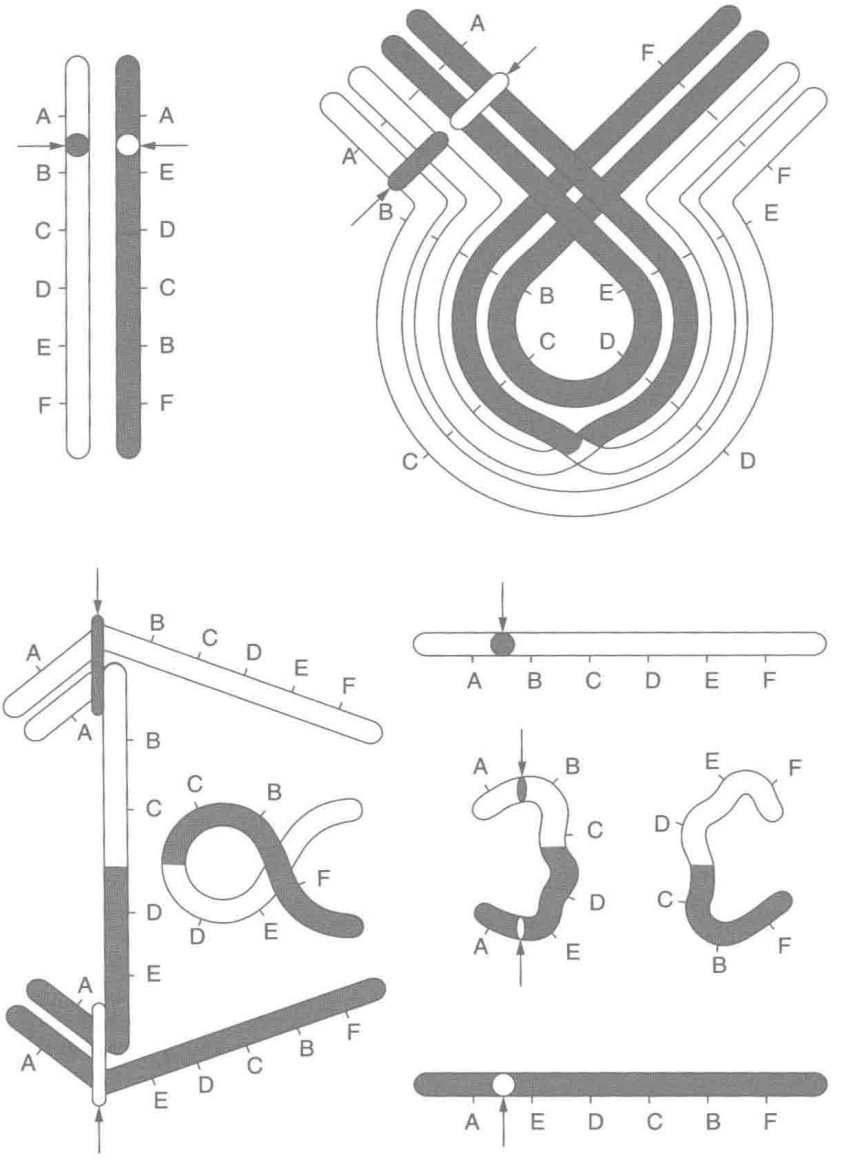

Fig. 1.4. Chromosome types produced after crossing over in an individual heterozygous for a paracentric inversion. Note the chromosomal bridge and the resulting chromatids with deletions. (Used with permission from T. Dobzhansky, © 1970, *Genetics of the Evolutionary Process*, Columbia University Press, New York.)

of the genome represented by duplications has been estimated to be 72% in maize (Ahn and Tanksley, 1993; Gaut and Doebley, 1997) and 60% in *Arabidopsis thaliana* (Blanc *et al.*, 2000).

Clusters of duplicated genes are often found scattered at multiple locations across the genome. When Blanc *et al.* (2000) compared the sequence of duplications on chromosomes 2 and 4 of *Arabidopsis thaliana*, they identified 151 pairs of genes, of which 59 (39%) showed highly similar nucleotide sequences. The order of these genes was generally maintained on the two chromosomes, except for a small duplication and an inversion. When they compared the sequence of these duplicated regions to the rest of the genome, they found 70% of the genes to be present elsewhere. The genes were duplicated in 18 large translocations and several smaller ones (Fig. 1.6).