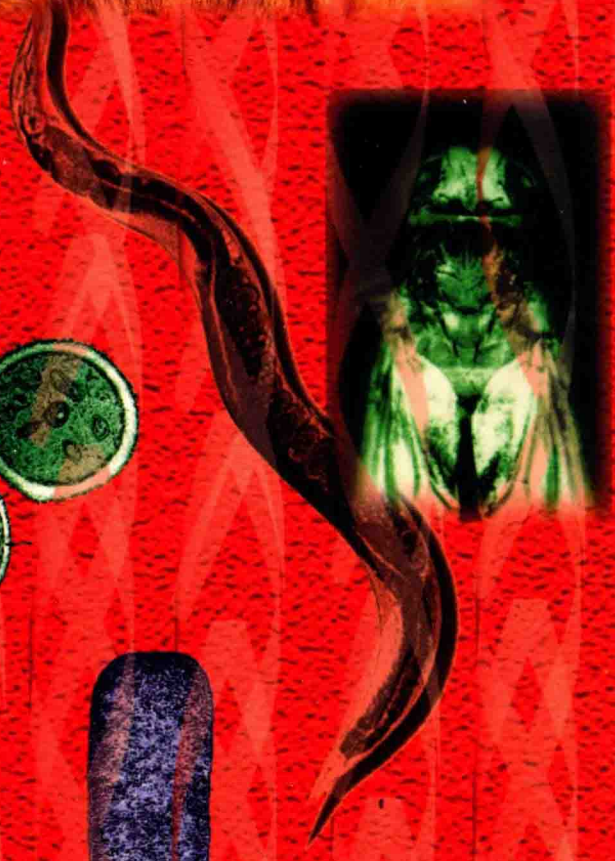
GENOME ANALYSIS

A LABORATORY MANUAL

VOLUME 4

MAPPING GENOMES



Volume Editors

Bruce Birren

Eric D. Green

Philip Hieter

Sue Klapholz

Richard M. Myers

Harold Riethman

Jane Roskams

COLD SPRING HARBOR LABORATORY PRESS

CSHL PRESS

GENOME ANALYSIS

A LABORATORY MANUAL

VOLUME 4

MAPPING GENOMES

Volume Editors

Bruce Birren
Whitehead Institute/MIT Center for Genome Research

Eric D. Green
National Human Genome Research Institute

Philip Hieter
University of British Columbia

Sue Klapholz
Stanford, California

Richard M. Myers
Stanford University School of Medicine

Harold Riethman
Wistar Institute

Jane Roskams
University of British Columbia

COLD SPRING HARBOR LABORATORY PRESS

GENOME ANALYSIS

A LABORATORY MANUAL
VOLUME 4 / MAPPING GENOMES

Printed in the United States of America
Book design by Emily Harste
Cover design by Tony Urgo

Library of Congress Cataloging-in-Publication Data

Genome analysis: a laboratory manual / editors Eric D. Green ... [et al.]
v. <1-2 >. cm.
Includes bibliographical references and index
Contents: v. 1. Analyzing DNA — v. 2. Detecting genes
ISBN 0-87969-496-3 (comb). — ISBN 0-87969-511-0 (comb).
ISBN 0-87969-495-5 (cloth). — ISBN 0-87969-510-2 (cloth).
v. 3. Cloning Systems — v. 4. Mapping Genomes
ISBN 0-87969-513-7 (comb). — ISBN 0-87969-515-3 (comb).
ISBN 0-87969-512-9 (cloth). — ISBN 0-87969-514-5 (cloth).
1. Gene mapping -- Laboratory manuals. 2. Nucleotide sequence -- Laboratory manuals. I. Green, Eric D.
QH445.2.G4645 1997
576.5'3—dc21
97-17117
CIP

Cover caption for paperback
Top: Photographs of human infant and mouse were reproduced with permission from Comstock, Inc.
Middle: (*Left*) *S. cerevisiae*; (*center*) *C. elegans*; (*right*) *Drosophila*
Bottom: E. coli

All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory Press, 10 Skyline Drive, Plainview, New York 11803-2500. Phone: 1-800-843-4388 in Continental U.S. and Canada. All other locations: (516) 349-1946. E-mail: cshpress@cshl.org. For a complete catalog of all Cold Spring Harbor Laboratory Press publications, visit our World Wide Web site: http://www.cshl.org.

GENOME ANALYSIS

A LABORATORY MANUAL

VOLUME 4

MAPPING GENOMES

GENOME ANALYSIS
A LABORATORY MANUAL

VOLUME 1
ANALYZING DNA

VOLUME 2
DETECTING GENES

VOLUME 3
CLONING SYSTEMS

VOLUME 4
MAPPING GENOMES

Series Editors

Eric D. Green
National Human Genome Research Institute

Bruce Birren
Whitehead Institute/MIT Center for Genome Research

Sue Klapholz
Stanford, California

Richard M. Myers
Stanford University School of Medicine

Philip Hieter
University of British Columbia

Preface

Although biologists have studied the genomes of numerous organisms for decades, the last ten years have brought an enormous increase in the pace of genome research. Problems of a previously unthinkable scope are now routinely solved, and the results of these studies are having a profound impact on biomedical research. Many of these advances are the direct result of the coordinated effort of the Human Genome Project, which was established to produce genomic maps and sequences for a set of well-studied organisms. Recent strategic and technological advances have made it feasible to clone, genetically manipulate, and analyze very large segments of DNA; to identify expressed sequences within large genomic regions; to identify DNA sequence variation associated with phenotypic variation; and to determine the nucleotide sequences of DNA cheaper and more efficiently than before. Indeed, to date, several bacterial genomes and that of baker's yeast have been sequenced in their entirety, and the sequencing of other genomes is well under way.

The purpose of this four-volume manual is to provide newcomers and experienced practioners alike with theoretical background, laboratory protocols, and resource materials for applying these powerful new techniques of genome analysis to the study of the very large number of genes and genomes yet to be characterized. We feel that there are several compelling reasons for producing such a manual. First, there are few sources that provide detailed information on the application of methods for genome analysis, due to the recent development of these techniques and the rapid rate at which they have evolved. Second, genome research has led to a new way of thinking that allows a completely different scale of question to be addressed. Many of these techniques, while originally developed for the systematic analysis of mammalian genomes, are now applied to specific biological questions in a variety of organisms. We hope to hasten further the increased understanding of basic biological phenomena by making this technology more widely applied. We have therefore provided step-by-step protocols with detailed explanations regarding why key manipulations are performed as described and which of these steps are most critical to success or prone to failure. We also describe how to recognize and avoid common problems and provide guidelines for troubleshooting various aspects of the protocols. Finally, we believe that a manual that covers the vast array of approaches needed at the different stages of a typical genome analysis project will find application in many laboratories.

This manual consists of chapters authored and edited by genome scientists who are experts in, and in many cases, the developers of, the described experimental techniques. The editors wrote some of the chapters themselves and worked closely with the other authors and the staff at Cold Spring Harbor Laboratory Press to develop cohesiveness, a consistent style,

and substantial cross-referencing. The methods are presented with a high level of detail and completeness to allow readers without experience with these methods to evaluate the strategies and successfully implement the protocols. Most of the editors have led laboratory courses at Cold Spring Harbor Laboratory, and thus have an appreciation for the importance of technology transfer and the amount of background explanation and detail that must accompany a protocol to make its use a success, especially in the hands of an inexperienced user. While we have not assumed that the user is experienced in the techniques of genome research, a basic knowledge of molecular biology techniques (i.e., recombinant DNA cloning) is a prerequisite for the successful use of most of the described methods. Appendices containing instructions for preparing reagents, protocols for basic methods used throughout the manual, safety information, and useful reference information are included at the end of each volume.

Volume 1 of this manual contains seven chapters describing basic techniques in genome analysis that are applicable to most of the experimental methods appearing in subsequent volumes. Chapter 1 provides basic protocols for isolating genomic DNA and performing standard manipulations, such as gel transfer for hybridization analysis. Chapter 2 describes protocols for isolating, manipulating, and analyzing high-molecular-weight DNA, many of which are based on technologies developed specifically in response to the need to clone, map, and sequence large genomes. The polymerase chain reaction is now a standard tool in almost all biomedical research laboratories, and the method has many key roles in genome research. Chapter 3 provides an overview of these PCR-based applications in genome analysis. Chapters 4, 5, and 6 describe large-scale DNA sequencing and include general protocols for dideoxy-mediated sequencing, as well as more specific methods for shotgun and directed sequencing strategies. Chapter 7 reviews perhaps the most critical set of all genome methods—those concerned with analyzing and accessing genomic information, particularly sequence data. This chapter is associated with an electronic version available at http:/www.cshl.org/books/g_a that contains supplemental information that should be useful to all researchers as well as to those interested in using the vast quantities of mapping and sequencing data generated by the genome project.

In Volume 2, comprehensive approaches are presented for identifying, isolating, and analyzing genes. Chapter 1 reviews strategies for gene discovery in mammalian systems, including approaches for gene identification, mapping, isolation of transcribed sequences, and assessment of candidate genes. Chapter 2 provides detailed protocols for the construction and screening of normalized cDNA libraries. Methods for gene isolation are presented in Chapters 3 (Direct cDNA Selection), 4 (Exon Trapping), and 5 (Gene Detection by the Identification of CpG Islands). Chapter 6 describes a variety of methods for the detection of DNA sequence variation, including protocols for identifying alterations in electrophoretic mobility and recognizing mismatches.

Volume 3 contains methods for using a variety of genomic cloning systems. Chapter 1 provides general methods for working with bacterial cloning systems, including procedures for construction and storage of bacterial genomic libraries, library screening, contig assembly, and chromosome walking. Chapters 2, 3, and 4 describe three bacterial cloning systems: cosmids, bacteriophage P1, and BACs, respectively. Each of these chapters provides a detailed description of cloning vectors and host strains, as well as methods for library construction, characterization, and handling. Chapter 2 includes protocols for subcloning DNA from YACs into cosmids and for preparing low- and high-density filter arrays of cosmid clones. Chapter 3 includes a general description of the bacteriophage P1 life cycle and methods for preparation of packaging extracts. Chapter 4 includes protocols for recovery and analysis of the large BAC DNA inserts. Finally, Chapter 5 provides a comprehensive treatment of the YAC cloning system, including a primer on yeast genetics and detailed protocols for library construction, screening, and characterization. Methods for working with individual YAC clones include protocols for long-range restriction mapping, isolation of insert ends, and YAC manipulation, such as recombination-based modification, amplification, and transfer of YACs between yeast strains and into mammalian cells by spheroplast fusion.

Volume 4 focuses on methods central to the generation and use of genomic maps. Chapter 1 is a comprehensive review of meiotic mapping in humans. Chapter 2 reviews genetic and comparative mapping in mice and includes protocols for mapping genes that are defined by a visible phenotype. Chapter 3 describes the use of DNA polymorphisms and provides protocols for the identification and analysis of single nucleotide polymorphisms (SNPs) and short tandem repeat polymorphisms (STRPs). Chapter 4 describes the use of DNA markers in physical mapping, including steps for the development of sequence-tagged sites (STS). In Chapter 5, the principles and applications of representational difference analysis (RDA) are presented along with detailed protocols for performing RDA. Chapter 6 provides an overview of somatic cell genetic and radiation hybrid mapping, with protocols for constructing and analyzing interspecific cell hybrids and radiation hybrids, as well as a description of database and software resources for radiation hybrid mapping. Chapter 7 is a comprehensive treatment of mapping by fluorescence in situ hybridization (FISH).

We strongly encourage users of this manual to heed all safety cautions noted in the protocols, in the Appendices, and in the instructions provided by manufacturers. We urge all investigators to be familiar with the safe use of reagents and laboratory equipment, as well as with national, state, local, and institutional regulations regarding the use and disposal of materials described in this manual.

The 1970s and 1980s brought the revolution of molecular biology—initially a field of research but more recently a fundamental set of techniques that have come to have key roles in virtually all studies of biological systems. The "genome revolution" of the 1990s will likely evolve in a similar fashion. Our hope is that the experimental techniques described in this four-volume manual will ultimately be useful for investigators focusing on the study of genomes, as well as for those wishing to manipulate and analyze genomes as a means of gaining insight into basic biological processes.

Eric D. Green
Bruce Birren
Philip Hieter
Sue Klapholz
Richard M. Myers

JOHN J. WASMUTH (1946–1995)

Dedication

We dedicate this volume of Genome Analysis to the memory of Dr. John J. Wasmuth, whose untimely passing left his many friends and colleagues much the poorer, both for the loss of his wonderful companionship and for the interruption of his seminal contributions to human and somatic cell genetics. Much of the material in this series was influenced by his pioneering work, particularly in the areas of somatic cell hybrids, mapping, and positional cloning. His reach extends from those who were lucky enough to work with him directly as trainees or colleagues, through those whose careers were enriched at a distance by his manifold discoveries and intellectual leadership, to the many people whose lives have been affected by one of the diseases for which he identified contributing genes. The entire human genetics community together cannot make up for the loss of his wisdom, but our wish to honor his memory ensures that we will never cease to try.

Acknowledgments

The editors and authors would like to thank the following:

Susan Allen
Chip Asbury
David Basiji
Anne Bergmann
Pat Brown
Nigel Carter
Mari Christensen
Thomas Cremer
Pieter de Jong
Christine Disteche
Anne Fertitta
Cynthia Friedman
Dominique Giorgi
Andi Gnirke
Connie Kelly
Judy Kirk
Leonid Kruglyak
Hillary Massa
John McPherson
Heather Mefford
Myron Peto
Pamela Rabbitts
Sylvie Rouquier
Mike Singer
David Spector
Ger van den Engh
Hiroki Yokota

We would also like to acknowledge the staff at Cold Spring Harbor Laboratory Press:

Executive Director: John Inglis
Assistant to the Director: Elizabeth Powers

Developmental Editors: Judy Cuddihy, Michele Ferguson, Marilu Hoeppner
Production Editor: Dorothy Brown
Project Coordinator: Mary Cozza
Desktop Editor: Susan Schaefer
Book Design: Nancy Ford, Emily Harste

Contributors

Aravinda Chakravarti, *Case Western University, Cleveland*
Peter D'Eustachio, *New York University Medical Center, New York*
William Dietrich, *Harvard Medical School, Cambridge*
Pui-Yan Kwok, *Washington University School of Medicine*
Nikolai Lisitsyn, *University of Pennsylvania School of Medicine, Philadelphia*
Audrey Lynn, *Case Western University, Cleveland*
Tara Matise, *The Rockefeller University, New York*
John McPherson, *Washington University School of Medicine, St. Louis*
Richard M. Myers, *Stanford University, California*
Debbie Nickerson, *University of Washington, Seattle*
Roger Reeves, *Johns Hopkins University School of Medicine, Baltimore*
Barbara J. Trask, *University of Washington, Seattle*
Douglas Vollrath, *Stanford University School of Medicine, California*
John J. Wasmuth,* *University of California, Irvine*
James L. Weber, *Marshfield Medical Research Foundation, Wisconsin*

*Deceased

Abbreviations and Acronyms

In addition to standard abbreviations for metric measurements (e.g., ml) and chemical symbols (e.g., HCl), the abbreviations and acronyms below are used throughout this manual.

5FoA	5-fluoroorotic acid
A	adenosine (RNA) or deoxyadenosine (DNA) residue
ACEDB	a *C. elegans* database
AHC medium	acid-hydrolyzed casein medium
AMCA	7-amino-4-methylcoumarin-3-acetic acid
amp^r	β-lactamase gene conferring resistance to ampicillin
Amp^r	ampicillin-resistance phenotype
AMV	avian myeloblastosis virus
AP PCR	arbitrarily primed PCR
APM	affected pedigree member
ARS	autonomously replicating sequence
AT-2	Artificial Transposon-2
ATCC	American Type Culture Collection
ATP	adenosine triphosphate
ATPγS	adenosine-5′-*O*-(3-thiotriphosphate)
AV-FIGE	asymmetric voltage field-inversion gel electrophoresis
B pulse	treatment of cells in culture with thymidine analog, bromodeoxyuridine (BrdU), following cell cycle arrest in early S phase
BAC	bacterial artificial chromosome
BAP	bacterial alkaline phosphatase
BP filter	band-pass filter
BLAST	basic local alignment search tool
bp	base pair

Bq	Becquerel
BrdU	bromodeoxyuridine
BSA	bovine serum albumin
C	cytidine (RNA) or deoxycytidine (DNA) residue
C. albicans	*Candida albicans*
cam^r	gene conferring resistance to chloramphenicol
Can	canavanine
C-banding	centromere banding
CCD	charge-coupled device
CCM	chemical cleavage at mismatches
CDGE	constant denaturant gel electrophoresis
CDI	carbodiimide modification
cDNA	complementary DNA
C. elegans	*Caenorhabditis elegans*
CEN	centromere
CEPH	Centre D'Etudes du Polymorphisme Humain
CERN	European Nuclear Research Council
CFLP	cleavage fragment length polymorphism
cfu	colony-forming units
CGH	comparative genome hybridization
CHAPS	3-([3-cholamidopropyl]-dimethylammonio)-1-propanesulfonate
CHEF	contour-clamped homogeneous electric field
CHL	Chinese hamster lung (cells)
CHO	Chinese hamster ovary (cells)
Ci	Curie
CIP (also known as CIAP)	calf intestinal alkaline phosphatase
cM	centiMorgans
COD-FISH	chromosome orientation and direct FISH
cpm	counts per minute
cR	centiRay
CTAB	cetyltrimethylammonium bromide
CTP	cytosine triphosphate
CY3, CY5	sulfoindocyanine derivatives
Cyh	cycloheximide
DAPI	4′,6-diamidino-2-phenylindole
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
DDBJ	DNA Database of Japan
ddF	dideoxy fingerprinting
ddGTP	dideoxyguanosine triphosphate
ddNTP	dideoxynucleoside triphosphate
DD PCR	differential display PCR
ddTTP	dideoxythymidine triphosphate
DEPC	diethyl pyrocarbonate
DGGE	denaturing gradient gel electrophoresis
dGTP	deoxyguanosine triphosphate
DHLPC	denaturing high-performance liquid chromatography

dITP	deoxyinosine triphosphate
DMD	Duchenne muscular dystrophy
D. melanogaster	*Drosophila melanogaster*
DMEM	Dulbecco's modified Eagle's medium
DMF	*N,N*-dimethylformamide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNMP	deoxynucleoside monophosphate
dNTP	deoxynucleoside triphosphate
DOL assay	dye-labeled oligonucleotide ligation assay
DOP PCR	degenerate-oligomer-primed PCR
D-PBS	Dulbecco's phosphate-buffered saline
dpm	disintegrations per minute
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
dUTP	deoxyuridine triphosphate
EBI	European Bioinformatics Institute
EBV	Epstein-Barr virus
ECM	enzyme mismatch cleavage method
EC number	Enzyme Commission number
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EEO	electroendosmosis
EGTA	ethylene glycol-bis(β-amino-ethyl ether) *N,N,N′,N′*-tetraacetic acid
ELISA	enzyme-linked immunosorbent assay
E-mail	electronic mail
EMBL	European Molecular Biology Laboratory
EPPS	*N*-(2-hydroxyethyl)piperazine-*N′*-(3-propanesulfonic acid)
ES	embryonic stem
EST	expressed sequence tag
EUCIB	European Backcross Collaborative Group
FACS	fluorescence activated cell sorting
FAM	5-carboxy-fluorescein
FBS	fetal bovine serum
FCS	fetal calf serum
FIGE	field-inversion gel electrophoresis
FISH	fluorescence in situ hybridization
FITC	fluorescein isothiocyanate
Flp-ter	fractional length measurement, a measure of the distance from p-ter to the probe site expressed as a fraction of the total length of the chromosome
FPLC	fast-performance liquid chromatography
FRET	fluorescence resonance energy transfer
FRT	FLP (Flp recombinase) recombination target
ftp	file transfer protocol
F.W.	formula weight

G	guanosine (RNA) or deoxyguanosine (DNA) residue
G_0, G_1, G_2	gap phases of the cell cycle
GBA	genetic bit analysis
G-banding	Giemsa banding
GDB	Genome Database
gDGGE	genomic denaturing gradient gel electrophoresis
GSS division	GenBank division for genome survey sequences
GTP	guanosine triphosphate
HA	heteroduplex analysis
HAP	hydroxyapatite
HAT	hypoxanthine aminopterin thymidine (medium)
HBSS	Hanks' balanced salt solution
HEPES	*N*-(2-hydroxyethyl)piperazine-*N*´-(2-ethanesulfonic acid)
HGMP	Human Genome Mapping Program
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HMBD	histidine-tagged methyl-CpG-binding domain
HMW DNA	high-molecular-weight DNA
HPLC	high-performance liquid chromatography
HPRT	hypoxanthine phosphoribosyl transferase
HSCR	Hirschsprung disease
HSV-TK	herpes simplex virus thymidine kinase
HT	hypoxanthine and thymidine
HTG division	GenBank division for data from high-throughput genome sequencing centers
HTML	hypertext markup language
http	hypertext transfer protocol
H. wingei	*Hansenula wingei*
IBD	identity-by-descent
IBS	identity-by-state
IgG	immunoglobulin G
IMEM	Iscove's modified Eagle's medium
IPTG	isopropylthio-β-D-galactoside
IRS PCR	interspersed-repetitive-sequence-based PCR
kan^r	gene conferring resistance to kanamycin
Kan^r	kanamycin-resistance phenotype
kb	kilobase pair
kD	kilodalton
KGB	potassium glutamate buffer
LARS	leucyl-tRNA synthetase
lb	pound(s)
LB medium/plate	Luria-Bertani medium/plate
LDB	Genetic Location Database
LIDS	lithium dodecyl sulfate
LINE	long interspersed nuclear element
lod	logarithm of the odds
LP filter	long-pass filter

LTR	long terminal repeat
M	molar
M phase	mitosis phase of the cell cycle
M13 RF	M13 replicative form
MACAW	multiple alignment construct and analysis workbench
Mb	megabase pair
MBD	methyl-CpG-binding domain
Mbytes	megabytes
α-MEM	α-minimum essential medium
MCD	multiple complete digest
MES	2-(*N*-morpholino)ethanesulfonic acid
MGD	Mouse Genome Database
MMLV	Moloney murine leukemia virus
MNB	minimum number of breaks
m.o.i.	multiplicity of infection
MOPAC	mixed oligonucleotide-primed amplification of cDNA
MOPS	3-(*N*-morpholino)propanesulfonic acid
MRD	mismatch repair detection
MREC	mismatch repair enzyme cleavage
mRNA	messenger RNA
m.w.	molecular weight
N	normal
β-NAD	β-nicotinamide adenine dinucleotide
NCBI	National Center for Biotechnology Information
NCHGR-DOE	National Center for Human Genome Research/Department of Energy
NGM	nematode growth medium
NIGMS	National Institute for General Medical Sciences
NIH	National Institutes of Health
NMR	nuclear magnetic resonance
NOR staining	nuclear organizing region staining
NP-40	Nonidet P-40
OFAGE	orthogonal field alternation gel electrophoresis
OLA	oligonucleotide ligation assay
OMIM	On-line Mendelian Inheritance of Man
ORF	open reading frame
ori	origin of replication
p arm	short arm of human chromosome
p-ter	p-arm telomere
PAC	P1-derived artificial chromosome
PACE	programmable autonomously controlled electrodes
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PERT	phenol emulsion reassociation technique

PFG	pulsed-field gel
PFGE	pulsed-field gel electrophoresis
pfu	plaque-forming unit
PIC	polymorphism information content
PIPES	piperazine-*N,N´*-bis(2-ethanesulfonic acid)
PL	prehead lysate
PMSF	phenylmethylsulfonyl fluoride
PN buffer	phosphate buffer containing NP-40
poly(A)$^+$	polyadenosine residues
PPL	packaging protein lysate
PRINS labeling	primed in situ labeling
PTT	protein truncation test
q arm	long arm of human chormosome
q-ter	q-arm telomere
Q-banding	quinacrine banding
QFD-banding	Q-banding by fluorescence using DAPI
r_{avg}	average radius
RACE	rapid amplification of cDNA ends
RAPD	random amplified DNA polymorphism
RARE cleavage	RecA-assisted restriction enzyme cleavage
R-banding	replication or reverse banding
RC	recombinant congenic strains
RCRE	rare-cutting restriction enzyme
RDA	representational difference analysis
REF	restriction enzyme fingerprinting
RF	replicative form
RFLP (also known as RFLV)	restriction fragment length polymorphism (variant)
RGE	rotating gel electrophoresis
RH	radiation hybrid
RHdb	Radiation Hybrid Database
RI strain	recombinant inbred strain
RNA	ribonucleic acid
RNase	ribonuclease
ROX	6-carboxy-"X"-rhodamine (X refers to the eXtra julolidine rings of the fluorophore)
rpm	revolutions per minute
rRNA	ribosomal RNA
RT-PCR	reverse transcription followed by PCR
RZPD	German Human Genome Project RZPD (Ressourcen Zentrum Primar Datenbank)
S phase	DNA synthesis phase of the cell cycle
S. cerevisiae	*Saccharomyces cerevisiae*
SC medium	synthetic complete medium
SD medium	synthetic dextrose medium
SDP	strain distribution pattern
SDS	sodium dodecyl sulfate
SINEs	short interspersed repeated DNA sequences
SNP	single nucleotide polymorphism