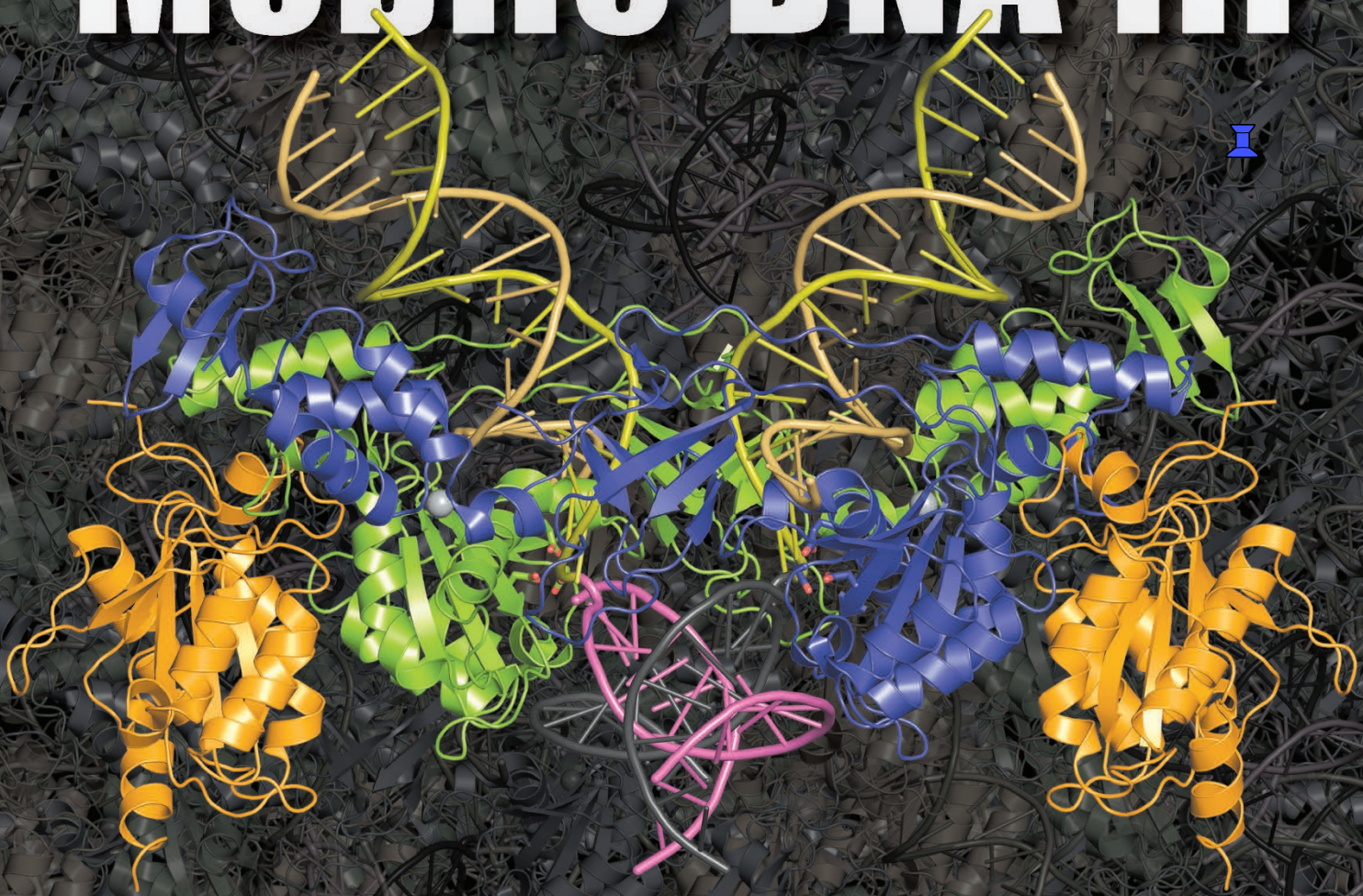
Mobile DNA III



Editor in Chief

Nancy L. Craig

Editors

Michael Chandler

Martin Gellert

Alan M. Lambowitz

Phoebe A. Rice

Suzanne Sandmeyer

Mobile DNA III

Mobile DNA III

Editor in Chief

Nancy L. Craig

Howard Hughes Medical Institute,
Department of Molecular Biology & Genetics,
Johns Hopkins University School of Medicine, Baltimore, MD

Editors

Michael Chandler

Laboratoire de Microbiologie et Génétique Moléculaire,
C.N.R.S., F-31062 Toulouse Cedex, France

Martin Gellert

National Institutes of Health,
Molecular Genetics Section, NIDDK, Bethesda, MD

Alan M. Lambowitz

Institute for Cellular Molecular Biology and
Department of Molecular Biosciences,
University of Texas at Austin, Austin, TX

Phoebe A. Rice

Department of Biochemistry and Molecular Biology,
University of Chicago, Chicago, IL

Suzanne B. Sandmeyer

Departments of Biological Chemistry and
Microbiology and Molecular Genetics,
University of California, Irvine, Irvine, CA

ASM Press, Washington, DC

Library of Congress Cataloging-in-Publication Data

Mobile DNA III / editor in chief, Nancy L. Craig, Johns Hopkins University School of Medicine, Baltimore, MD ; editors, Michael Chandler, Université Paul Sabatier, Toulouse, France, Martin Gellert, National Institutes of Health, Bethesda, MD, Alan M. Lambowitz, University of Texas, Austin, TX, Phoebe A. Rice, University of Chicago, Chicago, IL, Suzanne Sandmeyer, University of California, Irvine, CA.
pages cm
Includes bibliographical references and index.
ISBN 978-1-55581-920-0 (alk. paper)
1. Mobile genetic elements. I. Craig, Nancy Lynn, 1952- editor. II. Chandler, Michael (Molecular microbiologist), editor. III. Gellert, Martin, editor. IV. Lambowitz, Alan, editor. V. Rice, Phoebe A., editor. VI. Sandmeyer, Suzanne, editor. VII. Title: Mobile DNA 3. VIII. Title: Mobile DNA three.
QH452.3.M633 2015
572.8'69--dc23
2015011126
doi:10.1128/9781555819217

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Address editorial correspondence to: ASM Press, 1752 N St., N.W., Washington, DC 20036-2904, USA.

Send orders to: ASM Press, P.O. Box 605, Herndon, VA 20172, USA.
Phone: 800-546-2416; 703-661-1593. Fax: 703-661-1501.
E-mail: books@asmusa.org
Online: http://estore.asm.org

Cover: An artistic representation of the retroviral intasome engaged with target DNA in the nucleus of a host cell. The illustration is based on the crystal structure of the prototype foamy virus strand transfer complex (Protein Databank ID 3OS0; for details see Maertens *et al*., Nature, 2010, 468, 326-9). Image provided by Dr. Peter Cherepanov, Cancer Research UK, London Research Institute, London, EN6 3LD United Kingdom.

Contents

Contributors *xi*
Preface *xxi*

I. INTRODUCTION

1. A Moveable Feast: An Introduction to Mobile DNA 3
   Nancy L. Craig

II. CONSERVATIVE SITE-SPECIFIC RECOMBINATION

2. An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective 43
   Makkumi Jayaram
3. The Serine Recombinases 73
   W. Marshall Stark
4. The λ Integrase Site-specific Recombination Pathway 91
   Arthur Landy
5. Cre Recombinase 119
   Gregory D. Van Duyne
6. The Integron: Adaptation On Demand 139
   José Antonio Escudero, Céline Loot, Akeksandra Nivina, and Didier Mazel

7. Xer Site-Specific Recombination: Promoting Vertical and Horizontal Transmission of Genetic Information 163
CAROLINE MIDONET AND FRANSOIS-XAVIER BARRE

8. The Integration and Excision of CTnDOT 183
MARGARET M. WOOD AND JEFFREY F. GARDNER

9. Site-specific DNA Inversion by Serine Recombinases 199
REID C. JOHNSON

10. Serine Resolvases 237
PHOEBE A. RICE

11. Phage-encoded Serine Integrases and Other Large Serine Recombinases 253
MARGARET C.M. SMITH

12. Hairpin Telomere Resolvases 273
KERRI KOBRYN AND GEORGE CHACONAS

13. Biology of Three ICE Families: SXT/R391, ICE*Bs1*, and ICE*St1*/ICE*St3* 289
NICHOLAS CARRARO AND VINCENT BURRUS

III. PROGRAMMED REARRANGEMENTS

14. V(D)J Recombination: Mechanism, Errors, and Fidelity 313
DAVID B. ROTH

15. Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination 325
JOYCE K. HWANG, FREDERICK W. ALT, AND LENG-SIEW YEAP

16. Programmed Genome Rearrangements in *Tetrahymena* 349
MENG CHAO YAO, JU-IAN CHAO, AND CHAO-YIN CHENG

17. Programmed Rearrangement in Ciliates: *Paramecium* 369
MIREILLE BETERMIER AND SANDRA DUHARCOURT

18. Programmed Genome Rearrangements in the Ciliate *Oxytricha* 389
V. TALYA YERLICI AND LAURA F. LANDWEBER

19. DNA Recombination Strategies During Antigenic Variation in the African Trypanosome 409
RICHARD MCCULLOCH, LIAM J. MORRISON, AND JAMES P.J. HALL

20. Recombination and Diversification of the Variant Antigen Encoding Genes in the Malaria Parasite *Plasmodium falciparum* 437
LAURA A. KIRKMAN AND KIRK W. DEITSCH

21. Mobile DNA in the Pathogenic *Neisseria* 451
Kyle P. Obergfell and H. Steven Seifert

22. *vls* Antigenic Variation Systems of Lyme Disease *Borrelia*: Eluding Host Immunity through both Random, Segmental Gene Conversion and Framework Heterogeneity 471
Steven J. Norris

23. Mating-type Gene Switching in *Saccharomyces cerevisiae* 491
Cheng-Sheng Lee and James Haber

24. A Unique DNA Recombination Mechanism of the Mating/Cell-type Switching of Fission Yeasts: a Review 515
Amar Klar

IV. DNA-ONLY TRANSPOSONS

25. Mechanisms of DNA Transposition 531
Allison Burgess Hickman and Fred Dyda

26. Everyman's Guide to Bacterial Insertion Sequences 555
Patricia Siguier, Edith Gourbeyre, Alessandro Varani, Bao Ton-Hoang, and Michael Chandler

27. Copy-out-Paste-in Transposition of IS*911*: A Major Transposition Pathway 591
Michael Chandler, Olivier Fayet, Philippe Rousseau, Bao Ton Hoang, and Guy Duval-Valentin

28. The IS*200*/IS*605* Family and "Peel and Paste" Single-strand Transposition Mechanism 609
Susu He, Alix Corneloup, Catherine Guynet, Laure Lavatine, Anne Caumont-Sarcos, Patricia Siguier, Brigitte Marty, Fred Dyda, Michael Chandler, and Boa Ton Hoang

29. Transposons Tn*10* and Tn*5* 631
David B. Haniford and Michael J. Ellis

30. Tn7 647
Joseph E. Peters

31. Transposable Phage Mu 669
Rasika M. Harshey

32. The Tn*3*-family of Replicative Transposons 693
Emilien Nicolas, Michael Lambin, Damien Dandoy, Christine Galloy, Nathan Nguyen, Cedric A. Oger, and Bernard Hallet

33. P Transposable Elements in *Drosophila* and other Eukaryotic Organisms 727
Sharmistha Majumdar and Donald C. Rio

34. Mariner and the ITm Superfamily of Transposons 753
MICHAEL TELLIER, CORENTIN CLAEYS BOUUAERT, AND RONALD CHALMERS

35. *hAT* Transposable Elements 773
PETER W. ATKINSON

36. *Mutator* and *MULE* Transposons 801
DAMON LISCH

37. Adeno-associated Virus as a Mammalian DNA Vector 827
MAX SALGANIK, MATTHEW HIRSCH, AND R. JUDE SAMULSKI

38. *Sleeping Beauty* Transposition 851
ZOLTÁN IVICS AND ZSUZSANNA IZSVÁK

39. *piggyBac* Transposon 873
KOSUKE YUSA

40. *Helitrons*, the Eukaryotic Rolling-circle Transposable Elements 891
JAINY THOMAS AND ELLEN J. PRITHAM

V. LTR RETROTRANSPOSONS

41. Ty1 LTR-retrotransposon of Budding Yeast, *Saccharomyces cerevisiae* 927
M. JOAN CURCIO, SHEIL LUTZ, AND PASCALE LESAGE

42. Ty3, a Position-specific Retrotransposon in Budding Yeast 965
SUZANNE SANDMEYER, KURT PATTERSON AND VIRGINIA BILANCHONE

43. The Long Terminal Repeat Retrotransposons Tf1 and Tf2 of *Schizosaccharomyces pombe* 997
CAROLINE ESNAULT AND HENRY L. LEVIN

44. Retroviral Integrase Structure and DNA Recombination Mechanism 1011
ALAN ENGELMAN AND PETER CHEREPANOV

45. Host Factors in Retroviral Integration and Selection of Integration Target Sites 1035
ROBERT CRAIGIE AND FREDERIC D. BUSHMAN

46. Reverse Transcription of Retroviruses and LTR Retrotransposons
STEVEN H. HUGHES 1051

47. Mammalian Endogenous Retroviruses 1079
DIXIE L. MAGER AND JONATHAN P. STOYE

48. Retroviral DNA Transposition: Themes and Variations 1101
ANNA MARIE SKALKA

VI. NON-LTR RETROTRANSPOSONS

49. Integration, Regulation, and Long-Term Stability of R2 Retrotransposons 1127
Thomas H. Eickbush and Danna G. Eickbush

50. Site-Specific non-LTR retrotransponsons 1147
Haruhiko Fujiwara

51. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes 1165
Sandra R. Richardson, Aurélien J. Doucet, Huira C. Kopera, John B. Moldovan, José Luis Garcia-Perez, and John V. Moran

52. Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution 1209
Alan Lambowitz and Marlene Belfort

53. Diversity-generating Retroelements in Phage and Bacterial Genomes 1237
Huatao Guo, Diego Arambula, Partho Ghosh, and Jeff F. Miller

54. An Unexplored Diversity of Reverse Transcriptases in Bacteria 1253
Steven Zimmerly and Li Wu

55. Tyrosine Recombinase Retrotransposons and Transposons 1271
Russell T.M. Poulter and Margi I. Butler

Index 1293

Contributors

Frederick W. Alt
Department of Genetics, Harvard Medical School, Boston, MA 02115

Diego Arambula
Department of Microbiology, Immunology and Molecular Genetics, David Geffen School of Medicine, University. of California at Los Angeles, Los Angeles, CA 90095

Peter W. Atkinson
Department of Entomology and Institute for Integrative Genome Biology, University of California, Riverside, CA 92521

Francois-Xavier Barre
Centre de Genetique Moleculaire, C.N.R.S., Gif-sur Yvette, France 91198

Marlene Belfort
Department of Biological Sciences and RNA Institute, University of Albany, State University of New York, Albany, NY 12222

Mireille Betermier
Centre de Génétique Moléculaire, C.N.R.S., Gif-sur Yvette, France 91198

Virginia W. Bilanchone
Departments of Biological Chemistry and Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697

Vincent Burrus
Departement de biologie, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada

FREDERIC D. BUSHMAN
Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

MARGI I. BUTLER
Department of Biochemistry, University of Otago, Dunedin, New Zealand 09054

NICOLAS CARRARO
Departement de biologie, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada

GEORGE CHACONAS
Departments of Biochemistry and Molecular Biology; Microbiology, Immunology & Infectious Diseases, The University of Calgary, Calgary, ALB, Canada AB T2N 4N1

RONALD CHALMERS
School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH United Kingdom

MICHAEL CHANDLER
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

JU-LAN CHAO
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan 11529

CHAO-YIN CHENG
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan 11529

PETER CHEREPANOV
Cancer Research UK, London Research Institute, London, EN6 3LD United Kingdom

CORENTIN CLAEYS BOUUAERT
School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH United Kingdom

ALIX COUNELOUP
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

ROBERT CRAIGIE
Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20892

M. JOAN CURCIO
Laboratory of Molecular Genetics, University of Albany, State University of New York, Albany, NY 12208

DAMIEN DANDOY
Institut des Sciences de la Vie, Université Catholique de Louvain, Louvain-la-Neuve, B-1348, Belgium

KIRK W. DEITSCH
Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10065

Aurelien J Doucet
Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109

Sandra Duharcourt
Institut Jacques Monod, C.N.R.S., Paris, France 75205

Guy Duval-Valentin
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

Fred Dyda
Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20814

Danna G. Eickbush
Department of Biology, University of Rochester, Rochester, NY 14627

Thomas H. Eickbush
Department of Biology, University of Rochester, Rochester, NY 14627

Michael Joseph Ellis
Department of Biochemistry, University of Western Ontario, London, Ontario, N6A 5C1, Canada

Alan Engelman
Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, MA 02215

Jose Antonio Escudero
Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

Caroline Esnault
Program in Cellular Regulation and Metabolism, NICHD, National Institutes of Health, Bethesda, MD 20892

Hsiu-Fang Fan
Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan 00112

Olivier Fayet
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse Cedex, France 31062

Haruhiko Fujiwara
Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Japan 277-8562

Christine Galloy
Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

Jose L. Garcia-Perez
Department of Human DNA Variability, GENYO (Pfizer-University of Granada & Andalusian Regional Government Genomics & Oncology Center), 18016 Granada, Spain

JEFFREY F. GARDNER
Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

PARTHO GHOSH
Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA 92093

EDITH GOURBEYRE
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse Cedex, France 31062

PIOTR GUGA
Department of Bio-organic Chemistry, Polish Academy of Sciences, Lodz, POLAND

HUATAO GUO
Department of Molecular Microbiology and Immunology, University of Missouri School of Medicine, Columbia, MO 65212

CATHERINE GUYNET
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse Cedex, France 31062

JAMES HABER
Department of Biology and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454

JAMES HALL
Department of Biology, University of York, York, YO10 5DD United Kingdom

BERNARD HALLET
Institut des Sciences de la Vie, Universté Catholique de Louvain, Louvain-la-Neuve, BELGIUM B-1348

DAVID B. HANIFORD
Department of Biochemistry, The University of Western Ontario, London, Ontario, N6A 5C1 Canada

RASIKA HARSHEY
Department of Molecular Biosciences, Institute of Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712

SUSU HE
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., Toulouse Cedex, France 31062

ALISON BURGESS HICKMAN
Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, MD 20814

MATTHEW HIRSCH
Gene Therapy Center, Department of Ophthalmology, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599

STEPHEN H. HUGHES
HIV Drug Resistance Program, National Cancer Institute at Frederick, NIH, Frederick, MD 21702

Joyce K. Hwang
Program in Cellular and Molecular Medicine; Department of Genetics, Harvard Medical School, Boston, MA 02115

Ken Ishikawa
Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute at Frederick, NIH, Frederick, MD 21702

Zoltan Ivics
Division of Medical Biotechnology, Paul Ehrlich Institute, Langen, Germany 63225

Zsuzsanna Izsvak
Max Delbruck Center for Molecular Medicine, Max Delbruck Center for Molecular Medicine, Berlin, Germany 13092

Makkuni Jayaram
Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

Reid C Johnson
Department of Biological Chemistry, University of California, Los Angeles School of Medicine, Los Angeles, CA 90095

Aashiq H. Kachroo
Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

Laura Kirkman
Department of Internal Medicine, Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10065

Amar J. S. Klar
Gene Regulation and Chromosome Biology Laboratory,, National Cancer Institute at Frederick, NIH, Frederick, MD 21702

Kerri Kobryn
Department of Microbiology and Immunology, College of Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5E5 Canada

Huira C. Kopera
Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109

Michael Lambin
Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

Alan M. Lambowitz
Institute for Cellular Molecular Biology and Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

Laura F. Landweber
Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544

ARTHUR LANDY
Division of Biology and Medicine, Brown University, Providence, RI 02912

LAURE LAVATINE
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

CHENG-SHENG LEE
Department of Biology and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454

PASCALE LESAGE
Institut Universitaire d'hématologie, C.N.R.S., F-75475 Paris Cedex 10, France

HENRY L. LEVIN
Program in Cellular Regulation and Metabolism, NICHD, National Institutes of Health, Bethesda, MD 20892

DAMON LISCH
Department of Botany and Plant Pathology, University of California, Berkeley, Berkeley, CA 94720

CELINE LOOT
Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

SHEILA LUTZ
Laboratory of Molecular Genetics, University of Albany, State University of New York, Albany, NY 12208

CHIEN-HUI MA
Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

DIXIE L. MAGER
Terry Fox Laboratory, British Columbia Cancer Agency and Dept. of Medical Genetics, University of British Columbia, Vancouver, British Columbia, V5Z 1L3, Canada

SHARMISTHA MAJUMDAR
Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720

BRIGITTE MARTY
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

DIDIER MAZEL
Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

RICHARD MCCULLOCH
Wellcome Trust Centre for Molecular Parasitology, University of Glasgow, Glasgow, G12 8TA, United Kingdom

CAROLINE MIDONET
Centre de Génétique Moléculaire, C.N.R.S., Gif-sur Yvette, 91198 France

Jeff F. Miller
Department of Microbiology, Immunology and Molecular Genetics; The California Nano Systems Institute, University of California at Los Angeles, Los Angeles, CA 90095

John B. Moldovan
Cellular and Molecular Biology Graduate Program, University of Michigan Medical School, Ann Arbor, MI 48109

Sharon Moore
Gene Regulation and Chromosome Biology Laboratory,, National Cancer Institute at Frederick, NIH, Frederick, MD 21702

John V. Moran
Departments of Human Genetics; Cellular and Molecular Biology, University of Michigan Medical School, Ann Arbor, Michigan 48109

Liam Morrison
Roslin Institute, University of Edinburgh, Midlothian, EH25 9RG United Kingdom

Nathan Nguyen
Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

Emilien Nicolas
Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

Aleksandra Nivina
Unité Plasticité du Génome Bactérien, Institut Pasteur, Paris, 75724 France

Steven J Norris
Department of Pathology and Laboratory Medicine, University of Texas Health Medical School, Houston, TX 77225

Kyle P. Obergfell
Department of Microbiology and Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611

Cedric A. Oger
Institut des Sciences de la Vie, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

Kurt Patterson
Departments of Biological Chemistry and Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697

Joseph E. Peters
Department of Microbiology, Cornell University, Ithaca, NY 14853

Russell Poulter
Department of Biochemistry, University of Otago, Dunedin, 09054 New Zealand

Ellen J. Pritham
Department of Human Genetics, University of Utah, Salt Lake City, UT 84112

PHOEBE A. RICE
Department of Biochemistry and Molecular Biophysics, University of Chicago, Chicago, IL 60637

SANDRA R RICHARDSON
Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109

DONALD RIO
Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720

DAVID B. ROTH
Department Pathology and Laboratory Medicine and the Center for Personalized Diagnostics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

PHILIPPE ROUSSEAU
Laboratoire de Microbiologie et Génétique Moléculaires, C.N.R.S., F-31062 Toulouse Cedex, France

PAUL A. ROWLEY
Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712

MAXIM SALGANIK
Gene Therapy Center/Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599

RICHARD JUDE SAMULSKI
Gene Therapy Center, University of North Carolina, Chapel Hill, Chapel Hill, NC 27599

SUZANNE B. SANDMEYER
Departments of Biological Chemistry and Microbiology and Molecular Genetics, University of California, Irvine, Irvine, CA 92697

ANNE SARCOS
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

H. STEVEN SEIFERT
Department of Microbiology and Immunology, Northwestern University Feinberg School of Medicine, Chicago, IL 60611

PATRICIA SIGUIER
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

ANNA MARIE SKALKA
Department of Microbiology, Fox Chase Cancer Center, Philadelphia, PA 19111

MARGARET C.M. SMITH
Department of Biology, University of York, York, YO10 5DD United Kingdom

W. MARSHALL STARK
Institute of Molecular Cell and Systems Biology, University of Glasgow, Glasgow, G12 8WW, United Kingdom

Jonathan P. Stoye
Division of Virology, MRC National Institute for Medical Research, London, NW7 1AA, United Kingdom

Michael Tellier
School of Life Sciences, University of Nottingham, Nottingham, NG7 2UH, United Kingdom

Jainy Thomas
Department of Human Genetic, University of Utah, Salt Lake City, UT 84112

Bao Ton Hoang
Laboratoire de Microbiologie et Génétique Moléculaire, C.N.R.S., F-31062 Toulouse Cedex, France

Gregory D. Van Duyne
Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104

Alessandro Varani
Departamento do Tecnología, UNESP - Univ. Estadual Paulista, Jaboticaba. SP, BRAZIL 14870

Yuri Voziyanov
School of Biosciences, Louisiana Tech University, Ruston, LA 71272

Margaret M. Wood
Academic Advising Center, Simmons College, Boston, MA 02115

Li Wu
Department of Biological Sciences, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

Meng-Chao Yao
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan 11529

Leng Siew Yeap
Program in Cellular and Molecular Medicine; Department of Genetics, Harvard Medical School, Boston, MA 02115

V. Talya Yerlici
Department of Molecular Biology, Princeton University, Princeton, NJ 8544

Kosuke Yusa
Stem Cell Genetics, Wellcome Trust Sanger Institute, Cambridge, CB10 1SA, United Kingdom

Steven Zimmerly
Department of Biological Sciences, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

Preface

Nowhere do the cooperative powers of DNA sequencing, high-resolution protein structure, biochemistry and molecular genetics shine more intensely than on Mobile DNAs. In Mobile DNA II, we knew that almost half of the human genome is comprised of retroelements. What discoveries since Mobile DNA II could surpass that claim? Very simply: everywhere DNA is dynamic, and we now meet the elegant protein machines, co-evolved DNA partners, and diverse RNA choreographers. These pages hold something for every reader, beginning with the introductory overview of mechanisms. Novices will find some of the most lucid reviews of these complex topics available anywhere. Specialists will be able to pick and choose advanced reviews of specific elements, but will be drawn in by unexpected parallels and contrasts among the elements in diverse organisms. Biomedical researchers will find documentation of recent advances in understanding immune-antigen conflict between host and pathogen. Biotechnicians will be introduced to amazing tools for *in vivo* control of designer DNAs. And long-time aficionados will simply fall in love all over again.

Questions still abound about the Transposable Elements (TE) described in this volume. Perhaps none is more profound than the basis and consequences of TE diversity even among related genomes. Active DNA TE show perhaps the most disparate distribution among organisms, being dominant in prokaryotes, and in some animals, including some insects and fish, but with the exception of certain bats, virtually absent in mammals. Plants illustrate expansion of genomes, mediated not only by increasing ploidy, but also by expansion of DNA-based TE and Long Terminal Repeat (LTR) retrotransposons. Although reverse transcriptases are found throughout all kingdoms, autonomous retroelements simply explode together with their non-autonomous partners in mammals with remarkably species-specific types. These differences in mobile DNAs define self and mate, sister species, host and pathogen.

The most striking impression from these pages must be the raw power of genetic material to refashion itself to any purpose. DNA exchange between bacteria and their environment blurs the boundaries between host, transposon, and phage, as organisms secrete and take up DNA, stash genetic material in integrons for future use, conjugate, are attacked by phage and fight back. Delving into mechanisms, we see single-stranded hairpin structures and G quartets that anchor rearrangements in multiple ways; chemically diverse nucleophilic centers–hydroxyls couched in pentose, tyrosine or serine moieties–that covalently bond or attack directly in strand-transfer reactions. Proteins act as clamps to topologically constrain DNA or act as mechanical swivels, linking and unlinking mobilizing strands. Resolution of transposition intermediates might also involve host replication or recombination machinery. More recently discovered helitrons offer unexpected opportunities for expansion of DNA-based elements by rolling-circle replication.

RNA, the primal, catalytic nucleic acid, is in evidence everywhere. In retroelements, RNA partners with reverse transcriptase to deliver on transcriptional expansion of autonomous and non-autonomous TE sequences. Group II introns in bacteria likely gave rise to eukaryotic organelle group II self-splicing, retro-homing introns, Long INterspersed Elements (LINEs), telomerase reverse transcriptases and in addition, spliceosomal introns. Phylogenetic analysis of bacterial genomes previously revealed group II introns, diversity-generating retroelements Diversity-Generating Retroelements (DGR), and retrons, but next generation sequencing now identifies a multitude of novel reverse transcriptases of unknown function.

In ciliates, *Paramecium, Tetrahymena* and *Oxytrichia,* RNA directs massive genome reduction between germ-line and somatic nuclei, mediated by ancient transposase-like enzymes. LINEs containing restriction-enzyme like- or AP-endonucleases dominate in some eukaryotic cells. Others are dominated by LTR retrotransposons and their offspring, the retroviruses; stripped down Penelope-like elements with GIY-YIG endonucleases; DIRS elements with tyrosine recombinases: and attendant non-autonomous elements.

Exceptional elements provide evidence for the interaction of domains over evolutionary time, including LTR retrotransposons encoding envelope proteins, retroviruses replicating intracellularly, and DIRS elements in which retroelement RT/RNaseH is associated with a Crypton-type DNA element tyrosine recombinase.

Nowhere is the sharp focus of structural biology and biochemistry more apparent than in studies of key retroelement enzymes reverse transcriptase and integrase motivated by the quest for inhibitors of human immunodeficiency virus (HIV) replication. Reverse transcriptase structures for multiple retroviruses, as well as now one retrotransposon, demonstrate the robustness of the palm, thumb, fingers model. However, as a caution against too much generalization, subdomains are re-arranged in monomeric, homodimeric, and heterodimeric forms in different enzymes, and catalytic activities operate in *cis* or *trans* within the complex, depending on the enzyme. The structure of full length retrovirus integrase notoriously resisted high resolution structural analysis, but now has rewarded efforts of many labs with key insights (cover of this volume). These include a surprising dimer-dimer interface where active sites are juxtaposed to a trapped, and dramatically bent and widened, major groove target. Whereas LTR retrotransposons target integration to transcriptionally-repressed regions through interactions with heterochromatic protein domains or Pol III-transcribed genes thought to repress Pol II transcription, next generation sequencing has surfaced less dramatic, but significant, retrovirus integration bias, favoring transcriptionally-active regions.

This distribution has been shown now in two cases to be mediated by interactions between integrase and epigenetic mark-associated proteins.

While it has been argued that mobile elements are "selfish DNAs", these pages are replete with examples of the positive contributions of mobile elements to host genome function. Bacterial transposons encode and mobilize selectable markers including antibiotic resistance, detoxifying enzymes, and conjugation and virulence functions. In eukaryotic cells, mobile elements contribute to chromosome structure: constituting centromeres or telomeres in some organisms and seeding heterochromatin in others. TE constitute a large fraction of transcription factor binding sites and provide an ongoing source of novel combinations which are responsive to stress signaling, MAP kinase activation and other developmental signals. Insertions of LINEs and Alu elements affect RNA processing because they encode cryptic splice sites, termination signals, and can target RNA editing.

Exapted mobile DNA coding sequences appear in novel contexts: transposases have evolved into the RAG endonuclease for V(D)J immunoglobulin gene diversification and into heterochromatic factor CENP-B; a reverse transcriptase evolved into telomerase; retrovirus envelope proteins became the trophoblast fusion protein syncytin. There are other examples of TE Open Reading Frames (ORFs) under selection, but with, as yet, unknown functions. Endogenous retroviruses forego prior allegiances and join strategies to resist new infections. For example, Fv–1, a retroviral Gag relic, thwarts incoming retroviruses of similar type. Repeated TE sequences are susceptible to DNA rearrangements via non-allelic recombination, aborted transposition, and generation of pseudo-genes–all of which might ultimately contribute to the resiliency of host genomes.

TE exploit their hosts as well. The bacterial XerCD tyrosine recombinases which function in bacteria to unlink multimeric chromosomes are exploited to integrate phage genomes or mediate invasion of the host chromosome on behalf of certain plasmid-borne mobile elements. Transposases, resolvases and integrases *in vivo* likely associate with host factors as they are joined with host genomes. TE are generally tightly controlled by host regulation so that some display opportunistic bursts of activity during specific windows of development. This is exemplified by yeast Ty transcription in response to MAP kinase signaling and activation of reverse transcription by DNA checkpoints. A common theme more generally is TE activation during stress. Diverse retroelements are derepressed during periods of germ cell development ensuring their vertical spreading in populations.

Despite these examples of cooperation, mobile DNAs are also in conflict with their hosts. RNA interference likely arose in part to combat mobilization of retroelements. Invaders of one sort or another engage in a dizzying unscored dance with their hosts. One result of this conflict is rapid evolution of genes encoding host innate immunity restriction factors, which for retroviruses include ones that prematurely uncoat incoming viruses, starve reverse transcriptases for nucleotides, and deaminate cytidines in replicating cDNA. Some of these same factors also suppress movement of endogenous retroelements.

Programmed variation is used by invaders and hosts alike for purposes of immune evasion or resistance, respectively. Examples include *Salmonella* Hin invertase flipping a promoter sequence to switch between expression of different antigenic flagellar proteins and DGR directing mutagenic reverse transcription of a template transcript coupled with directed conversion of a target expression site. *Neisseria gonorrhea, Borrelia burgdorferi*, *Trypanosoma brucei*; and *Plasmodium falciparum*, agents of gonorrhea, Lyme disease, sleeping sickness, and malaria, respectively, use amazingly diverse mechanisms to program variation of their antigenic surfaces for immune evasion. To counter this assault, there

is programmed variation of host immune proteins. In human immunoglobulin production, a DDE TE-derived RAG site-specific endonuclease initiates V(D)J switching, followed by transcription-activated somatic hyper-mutation (activation-induced cytidine deaminase), nuclease introduction of DSB, and final joint formation by redundant NHEJ pathways.

Next generation sequencing and development of methods for rapid TE mapping have greatly improved understanding of the distribution of TE as well as the utility of transposons for functional genomics. The bacterial Tn5 system has been exploited in particular for *in vitro* mutagenesis and next generation sequencing libraries by collapsing fragmentation and adapter ligation into a single step. P, Hermes, piggyBac, and Sleeping Beauty transposons have wide activity in eukaryotic systems and have been harnessed for genome-wide profiling, gene disruption and tagging, and genome modification. Retroviruses are additionally used in lineage tracing. The controlled, high-frequency mobilization of Mutator has made it indispensable for gene discovery in maize.

In medical research, understanding the impact of DNA mobilization is critical. In addition to individual TE, other mobile DNAs such as plasmids, Integrative Conjugative Elements (ICE) and both transposon-borne and chromosomal integrons are bacterial reservoirs of mobilizable antibiotic resistance. HIV, malaria, and sleeping sickness, and other pathogens, too numerous to mention here, remain threats to global health. Mobile element vectors transposons piggyBac, Sleeping Beauty, lenti-retroviruses and adenoviruses are being used as vectors to introduce exogenous DNAs in research, and in clinical trials. They differ with respect to targeting, excision footprints, payload size, and host activity profiles. Their mechanisms of DNA breakage and joining were among the systems first analyzed, now enabling them to be harnessed and used extensively for genome engineering including with developmentally-regulated expression, inactivation, and self-deletion strategies to enable probing essential or tissue- specific functions.

What challenges remain? One goal is to connect key findings from basic research, to clinical developments in drug resistance and genetic engineering. This volume is based on the considerable increase in understanding of molecular mechanisms of mobilization in the last decade. However, we have likely seen only the tip of the iceberg of how mobile DNAs affect the day to day biology of their hosts.

In the human genome alone, retroelements provide promoters for long non-coding and other RNAs of completely unknown function; Alu elements redirect RNA processing and delivery, and mobilization is occurring during neuronal development and in cancer with unknown consequences, just to mention a few. Finally, endogenization of a gamma retrovirus in Australian koalas is ongoing and those studies should provide insights into retroelement-host interaction. How have transposition events after separation from other great apes contributed to traits that make us human? What transposition events will provide key substrates for future evolution? And of course, perhaps the ultimate question, could we survive as a species were there no transposition?

We give our heartfelt thanks to all the authors who contributed diverse and fascinating chapters to Mobile DNA III. We express special thanks to Patti Kodeck, Administrative Assistant to Editor in Chief N. Craig, who mediated recruitment of and communications with authors and interactions between them and the publisher. Finally, our most sincere thanks to all of our supporters at ASM Press for their dedication in producing this volume, but especially to: Gregory Payne, Senior Editor; Larry Klein, Production Editor; Christine Charlip, Director of Administration; and Cathy Balogh, Administrative Assistant for Production.

Introduction

I

Mobile DNA, 3rd Edition
Nancy L. Craig, Michael Chandler, Martin Gellert, Alan M. Lambowitz, Phoebe A. Rice and Suzanne Sandmeyer

doi:10.1128/microbiolspec.MDNA3-0062-2014

Nancy L. Craig[1]

1

A Moveable Feast: An Introduction to Mobile DNA

INTRODUCTION

DNA has two critical functions: to provide the cell with the information necessary for macromolecular synthesis and to transmit that information to progeny cells. Genome sequence stability is important for both these functions. Indeed, cells devote significant resources to various DNA repair processes that maintain genome structure and repair alterations that can arise from DNA synthesis errors and assaults from both endogenous and exogenous sources. DNA sequence variation, however, provides the substrate for adaptation, selection, and evolution.

Genomes are, however, highly dynamic. Notably, they vary not only at the single or several base pair level (although such changes can be transformative and even deadly), but they also change by DNA rearrangements, that is, the movement of DNA segments that may be many kb (or even longer) in length. Such rearrangements can have enormous impacts on genome structure, function, and evolution.

The DNA rearrangements discussed here generally involve specific DNA sequences, or in some cases RNA sequences, that are recognized and acted on by specialized recombination proteins or recombinases that promote DNA breakage followed by joining of the broken DNAs to new sites. The involvement of a sequence-specific recombinase is what distinguishes site-specific recombination from homologous recombination, which can occur between any two DNA segments as long as they are homologous to each other, as in RecA- and Rad51-dependent recombination. In some cases, the specialized recombinase is a sequence-specific nuclease that targets homologous recombination to a specific DNA site.

In some rearrangements, the recombinase alone breaks, exchanges, and joins DNA by using covalent protein-DNA intermediates. In other cases, DNA synthesis is also essential in these rearrangements. Notably, this DNA synthesis can involve not only conventional DNA synthesis in which a DNA polymerase uses DNA as a template, but also reverse transcription in which a novel DNA polymerase, reverse transcriptase, uses an RNA template to generate new DNA. A very wide number of other cellular processes can influence or be required for DNA rearrangements, including transcriptional activation of particular sites, DNA bending by bending proteins, DNA supercoiling, and many variations in chromatin structure, as well as DNA repair reactions including DNA end joining. Although a purified recombinase may execute DNA breakage and joining *in vitro*, it is critical to remember that this reaction and its consequences will be enormously influenced by its cellular environment.

[1]Department of Molecular Biology and Genetics, Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

Although DNA rearrangements can provide very useful rapid and focused changes in genetic information, they are also very hazardous. Unrepaired DNA breaks can result in DNA mis-segregation and are often lethal. Not surprisingly then, DNA rearrangements often occur in elaborate nucleoprotein complexes that organize and juxtapose the participating DNAs and promote breakage and joining in carefully choreographed steps. The frequency of DNA rearrangements is usually highly regulated, often by restricting to low levels the recombinase that initiates or mediates the rearrangements.

Mobile DNAs also include a diverse variety of discrete mobile genetic elements, such as transposable elements, that move themselves or copies of themselves from place to place within and between genomes. Thus, in some cases, a copy of the element remains at its original site and there is a new copy at the new insertion site. This type of replicative mechanism leads to an enormously high element copy number, especially in some eukaryotic genomes. The majority of the maize genome, for example, is derived from a particular kind of transposon. High copy numbers of repetitive sequences result in increased susceptibility to nonallelic homologous recombination events that can lead to deletions, inversions, translocations, and other chromosomal rearrangements.

The movement of a transposable element within a single genome can have substantial genotypic and phenotypic consequences. The insertion of a transposable element into a gene can lead to gene disruption but even nearby insertions can effect gene expression as many elements carry regulatory signals, such as enhancers and promoters, as well as splice sites and transcription termination signals. Excision of an element also changes the donor site. Thus, the intracellular translocation of a mobile element results in genetic variation. The range of target sites used by the elements ranges from insertion into specific sites or regions that provide a "safe harbor" for the element with reduced negative consequence on the host, to preferences for actively transcribed regions to facilitate element expression to virtually random insertion, which can thus result in genetic variation anywhere within the host genome.

DNA rearrangements also play a crucial role in the interactions between viral chromosomes and their hosts, as well as the proper replication and segregation of host chromosomes. Many viruses integrate into and excise from host genomes, although in some cases integration is irreversible, such as with HIV-1. All of these reactions involve at least specific sites on the viral genome that are acted upon by site-specific recombinases and which sometimes involve specific sites on the host genome. Recombination between specific sites to promote chromosome monomerization plays a key role in chromosome transmission in bacteria.

The translocation of mobile elements encoding a wide variety of determinants including genes encoding antibiotic resistances, virulence determinants, and diverse metabolic pathways from plasmids to chromosomes and from viruses and DNA fragments that are transduced or transformed into a cell, can also result in permanent chromosomal acquisition of these determinants. This sort of horizontal gene transfer involving mobile elements is rampant in bacteria and contributes greatly to genetic variation. There are also an increasing number of examples of horizontal gene transfer involving mobile elements in eukaryotes.

Perhaps the most profound example of the effect of mobile elements on eukaryotic genome evolution is the nuclear invasion of mobile group II introns into the nuclear genome from bacterial symbionts to form spliceosomal introns.

Cell type can also have substantial impact on DNA rearrangements. The elaborate DNA breakage and joining reactions that underlie immunoglobulin diversification are actually terminal differentiation events restricted to particular somatic cells. There is increasing interest in the somatic movement of transposable elements, which can also have profound organismal impact. The movement of human transposable elements in somatic tissue is associated with a variety of cancers, although it remains to be determined if such events can cause oncogenic transformation or are rather a consequence of transformation. The movement of transposable elements in neuronal tissue in several organisms raises the interesting possibility that such rearrangements are a deliberate strategy for neural plasticity.

Such terminal differentiation events involving DNA rearrangements are incompatible with the bacterial lifestyle, except in a few known cases such as spore formation by a subset of cells. By contrast, reversible DNA inversions that vary promoter or ORF orientation are well known in bacteria.

Thus, DNA rearrangements can contribute substantially to genetic variation. The frequency and potential advantage of the resulting variation must be carefully balanced with genome stability to avoid its potential for population-wide genomic catastrophe.

Although not exclusively so, the focus of this work is on the mechanism and regulation of DNA rearrangements. How do specific DNA (and sometimes RNA) sequences recognize each other and how do they assemble to form the machines in which DNA rearrangements occur? What are the mechanisms for DNA

strand breakage and joining? What processes determine when and where these reactions occur? How are actions at multiple DNA sites, for example, the two ends of a transposable element and its target DNA, coordinated? Importantly, how are nonproductive breakage and joining events avoided and how is intact duplex DNA regenerated?

Mobile DNAs are "natural" genome engineers. Although not a focus of this work, many of the mobile elements discussed here have been harnessed to facilitate researcher-directed rearrangements both *in vitro* and *in vivo*. Mobile elements are used for "random" insertional genome mutagenesis both *in vivo* and *in vitro*, as well as for "targeted mutagenesis." Many mobile elements are used as vectors in both homologous and heterologous systems.

TARGETED DNA BREAKS LEAD TO GENE REPLACEMENT

DNA Double Strand Breaks Stimulate Homology-Dependent DNA Repair

Homologous recombination occurs without requirement for any particular sequence, depending only on base pairing between the participating DNA strands. However, the frequency of homologous recombination is stimulated by the presence of broken DNA, in particular double strand breaks. These breaks stimulate recombination because the action of nucleases and helicases at these breaks leads to the generation of DNA with single stranded 3′ trails that are the preferred substrate for DNA pairing mediated by RecA- and Rad51-like proteins. By interacting with a donor site, this pairing of 3′-OH ends can initiate homology-dependent DNA repair, which copies DNA sequence information from the donor site into the broken DNA target site. This repair leads to the replacement, or modification, of an existing gene or insertion of a new gene. The insertion of many mobile DNAs into a new site is targeted by double strand breaks by highly site-specific endonucleases.

There's No Place Like Home: Homing Endonucleases

Homing endonucleases (HENs) are highly site-specific endonucleases (1). Although often associated with other genetic elements (see below), freestanding HEN genes can themselves be mobile DNA elements. If a HEN cleavage site lies within an "empty allele" of DNA that flanks the HEN ORF, cleavage of that target site can initiate homology-dependent DNA synthesis that will transfer a copy of the HEN gene to that double strand break at the nuclease target site (Fig. 1).

HEN genes are also often found in other genetic elements such as self-splicing RNA introns, that is, group I introns, and self-splicing proteins, that is, inteins. Thus, if the HEN introduces a double strand break into the "empty" allele of a site occupied by the intron or intein, targeted DNA repair introduces a copy of the DNA encoding the intron or intein into that target site. Because the RNA intron can splice out of the RNA containing it and the protein intein can splice out of the protein containing it, the insertion of these elements is generally phenotypically silent.

Alternative Life Styles: Switching Mating Type in *Saccharomyces cerevisiae*

These yeasts have two different haploid cell types, mating type **a** and mating type α, which can mate to form diploids. During sporulation, meiotic recombination shuffles the two parental genomes, generating diverse haploid progeny. To facilitate diploid formation, haploids can switch mating type from mating type **a**

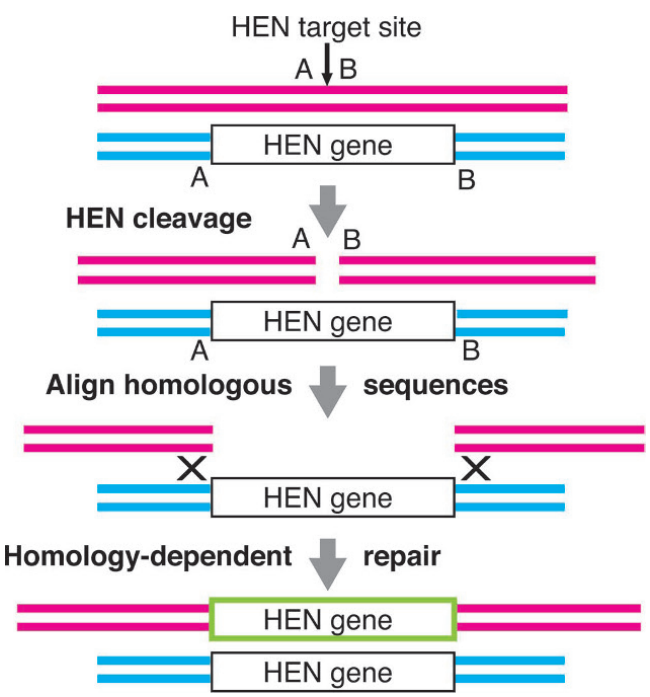


Figure 1 A targeted DNA double strand break can lead to gene insertion. Introduction of a site-specific double strand break by a homing endonuclease (HEN) in a homologous DNA duplex lacking the HEN gene targets homology-dependent DNA synthesis (green) that introduces a copy of the HEN gene to the broken DNA.
doi:10.1128/microbiolspec.MDNA3-0062-2014.f1