



PERSPECTIVES IN BIOANALYSIS

VOLUME 2

NEW HIGH THROUGHPUT TECHNOLOGIES FOR DNA SEQUENCING AND GENOMICS



KEITH R. MITCHELSON

(EDITOR)

PERSPECTIVES IN BIOANALYSIS

NEW HIGH THROUGHPUT TECHNOLOGIES FOR DNA SEQUENCING AND GENOMICS

EDITOR

KEITH R. MITCHELSON

CAPITALBIO CORPORATION:
NATIONAL ENGINEERING RESEARCH CENTRE
FOR BEIJING BIOCHIP TECHNOLOGY
18 LIFE SCIENCE PARKWAY
CHANGPING DISTRICT
BEIJING 102206
CHINA

AND

MEDICAL SYSTEMS BIOLOGY RESEARCH CENTER
TSINGHUA UNIVERSITY SCHOOL OF MEDICINE
BEIJING 100084
CHINA

VOLUME 2

AMSTERDAM – BOSTON – HEIDELBERG – LONDON – NEW YORK – OXFORD
PARIS – SAN DIEGO – SAN FRANCISCO – SINGAPORE – SYDNEY – TOKYO

Elsevier

Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2007

Copyright © 2007 Elsevier B.V. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions @elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at <http://elsevier.com/locate/permissions>, and selecting *Obtaining permission to use Elsevier material*

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data

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

ISBN-13: 978-0-444-52223-8

ISBN-10: 0-444-52223-9

ISSN: 1871-0069

For information on all Elsevier publications
visit our website at books.elsevier.com

Printed and bound in Italy

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

PERSPECTIVES IN
BIOANALYSIS

VOLUME 2

Cover image: Ian Tattersall, American Museum of Natural History

此为试读, 需要完整PDF请访问: www.ertongbook.com

Contributors

Numbers in parentheses indicate the pages where the authors' contributions can be found.

P. Adams (303), Department of Mathematics, University of Queensland, St. Lucia, Queensland 4072, Australia

J. J. Austin (357), Department of Environmental Biology, University of Adelaide, North Terrace, Adelaide 5005, South Australia, Australia

Dirk van den Boom (97), Sequenom Corporation, 3595 John Hopkins Court, San Diego, CA 92121, USA

I. Braslavsky (209), Department of Physics and Astronomy, Clippinger 251B, Ohio University, Athens, OH 45701, USA

D. E. Bryant (303), Department of Mathematics, University of Queensland, St. Lucia, Queensland 4072, Australia

J. C. Carter (303), Leukaemia Foundation Queensland Laboratories, Queensland Institute of Medical Research, Herston, Queensland 4006, Australia

J. Cheng (45), Capitalbio Corporation: National Engineering Research Center for Beijing Biochip Technology, Beijing, 18 Life Science Parkway, Changping District, Beijing 102206, China; and the Medical Systems Biology Research Center, Tsinghua University School of Medicine, Beijing 100084, China

D. A. E. Cochran (303), Agen Biomedical Limited, Durbell Street, Acacia Ridge, Queensland 4110, Australia

J. R. Edwards (187), Columbia Genome Center, Columbia University College of Physicians and Surgeons, Russ Berrie Medical Science Pavilion, 1150 St. Nicholas Avenue, New York, NY 10032, USA

M. Ehrich (97), Sequenom Corporation, 3595 John Hopkins Court, San Diego, CA 92121, USA

C. W. Fuller (119), GE Healthcare, 800 Centennial Avenue, Piscataway, NJ 08855, USA

P. S. Gooding (357), Agricultural Division – AGRF, Plant Genomics Centre, University of Adelaide, Hartley Grove, Waite Campus PMB1, Glen Osmond, SA 5064, Australia

D. B. Hawkes (3,303), AGRF, Institute of Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia

B. Hebert (209), Department of Physics, McGill University, Rutherford Physics Building 228, 3600 University Street, Montreal, Quebec H3A 2T8, Canada

J. Herschleb (265), Laboratory for Molecular and Computational Genomics, UW Biotechnology Center, Laboratory of Genetics and Department of Chemistry, University of Wisconsin, 425 Henry Mall, Madison 53706, USA

F. Hillenkamp (97), Institute for Medical Physics and Biophysics, University of Münster, Robert-Koch-Str. 31, Münster D-48149, Germany

T. P. Jarvie (153), 454 Life Sciences Corporation, 20 Commercial Street, Branford, CT 06405, USA

J. Ju (187), Columbia Genome Center, Columbia University College of Physicians and Surgeons, Russ Berrie Medical Science Pavilion, 1150 St. Nicholas Avenue, New York, NY 10032, USA

J. M. Keith (303), Institute of Molecular Bioscience and Department of Mathematics, University of Queensland, St. Lucia, Queensland 4072, Australia

D. H. Kim (187), Columbia Genome Center, Columbia University College of Physicians and Surgeons, Russ Berrie Medical Science Pavilion, 1150 St. Nicholas Avenue, New York, NY 10032, USA

J. R. Knight (153), 454 Life Sciences Corporation, 20 Commercial Street, Branford, CT 06405, USA

G. A. Kowalchuk (357), Netherlands Institute of Ecology, PO Box 40, 6666 ZG Heteren, The Netherlands

S. Kumar (119), GE Healthcare, 800 Centennial Avenue, Piscataway, NJ 08855, USA. Present Address: 21 Muirhead Court, Belle Mead, NJ 08502, USA

J. W. Lee (245), Oak Ridge National Laboratory, Oak Ridge, TN 37831-6194, USA

M. Margulies (153), 454 Life Sciences Corporation, 20 Commercial Street, Branford, CT 06405, USA

A. McGrath (327), Australian Genome Research Facility, University of Queensland, St. Lucia, Queensland 4072, Australia

A. Meller (245), The Department of Physics and Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215, USA

A. E. Men (3), AGRF, Institute of Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia

K. R. Mitchelson (3,303), Capitalbio Corporation: National Engineering Research Centre for Beijing Biochip Technology, 18 Life Science Parkway, Changping District, Beijing 102206, China; and Medical Systems Biology Research Center, Tsinghua University School of Medicine, Beijing 100084, China

D. C. Schwartz (265), Laboratory for Molecular and Computational Genomics, UW Biotechnology Center, Laboratory of Genetics and Department of Chemistry, University of Wisconsin, 425 Henry Mall, Madison 53706, WI, USA

J. F. Simons (153), 454 Life Sciences Corporation, 20 Commercial Street, Branford, CT 06405, USA

J. R. Stephen (357), Agricultural Division – AGRF, Plant Genomics Centre, University of Adelaide, Hartley Grove, Waite Campus PMB1, Glen Osmond, SA 5064, Australia

R. Turakulov (3), AGRF, Walter and Eliza Hall Institute for Medical Research, 1G Royal Parade, Parkville, Victoria 3050, Australia

Q. Xiong (45), Capitalbio Corporation: National Engineering Research Center for Beijing Biochip Technology, 18 Life Science Parkway, Beijing 102206, China

S. Zhou (265), Laboratory for Molecular and Computational Genomics, UW Biotechnology Center, Laboratory of Genetics and Department of Chemistry, University of Wisconsin, 425 Henry Mall, Madison 53706, WI, USA

Preface

Since the independent invention of DNA sequencing by Sanger and by Gilbert 30 years ago, it has grown from a small-scale technique capable of reading several kilobase-pair of sequence per day into today's multibillion dollar 'industry', with large Sequencing Centers for large-scale delineation of entire genomes, and supporting DNA sequencing activity at some level at virtually all Universities and larger hospitals throughout the world. We are now in a "post-genomic era" with possibly more than 150 billion base-pair of sequence information held at international Bioinformatics Centers, yet DNA sequencing continues as a major diagnostic and research activity in many areas of life science and medicine. This growth has spurred the development of new sequencing technologies that do not involve either electrophoresis or Sanger sequencing chemistries. Sequencing by synthesis (SBS) involves multiple parallel micro-sequencing addition events occurring on a surface, where data from each round is detected by imaging. The recent plan to sequence a complete Neandertal genome (*Homo sapiens neanderthalensis*) by the 454 Life Sciences and the Max Planck Institute using SBS underlines the relevance of these new sequencing technologies to post-genomic science. This will be the second entire human genome project, and it is expected to radically advance knowledge of the human genomic biology and provide profound new insights into genetic diseases in man.

This volume brings together some of the new developments in DNA sequencing technology. Reviews of complementary developments in Sanger and SBS sequencing chemistries (Kumar and Fuller), capillary electrophoresis and micro-device integration (Xiong and Cheng), MS sequencing (Ehrich *et al.*) and applications (Mitchelson *et al.*) set the framework. Reviews of new developments in SBS technology (Margulies *et al.*), the chemistry of nucleotide-dye SBS sequencing (Edwards *et al.*), and steps toward realizing the sequencing of single DNA molecules by cyclic synthesis (Hebert and Braslavsky), nano-pore sequencing (Lee and Meller) and optical mapping of arrayed DNA (Schwartz *et al.*) indicate the latest advances. Finally, bioinformatics tools for genome assembly (McGrath), sequencing ancient and environmental DNA samples (Kowalchuk *et al.*) and support for SBS sequence assembly (Keith *et al.*) discuss many issues relevant to new applications using SBS sequencing. The authors hope this volume will provide stimulus to both students and researchers interested in this vital field of chemistry and technological innovation.

July 2006

Keith Mitchelson

Contents

<i>Contributors</i>	xi
<i>Preface</i>	xv

ENABLING TECHNOLOGIES

Chapter 1. Overview: Developments in DNA Sequencing

Keith R. Mitchelson, David B. Hawkes, Rustam Turakulov and Artem E. Men	3
--	---

1. Introduction	4
2. Advanced sequencing technologies	9
3. Solid-phase array sequencing devices	15
4. Future technologies	22
5. Applied short-read genomic sequencing	25
6. Summary	35
References	36

Chapter 2. Chip Capillary Electrophoresis and Total Genetic Analysis Systems

Qiang Xiong and Jing Cheng	45
----------------------------	----

Abstract	46
1. Introduction	46
2. Chip design and fluid manipulation	48
3. Materials and fabrication	51
4. Detection	57
5. Surface modification	65
6. Applications	68
7. DNA sequencing	74
References	87

Chapter 3. Comparative Sequence Analysis by MALDI-TOF Mass Spectrometry – Utilizing the Known to Discover the New

Mathias Ehrich, Franz Hillenkamp and Dirk van den Boom 97

Abstract	97
1. The concept of comparative sequencing	98
2. MALDI-TOF MS-based nucleic acid analysis	99
3. The base-specific cleavage assay	100
4. Applications for comparative sequencing	103
5. Summary	112
6. Outlook	112
Acknowledgements	115
References	115

Chapter 4. Advances in Dye-Nucleotide Conjugate Chemistry for DNA Sequencing

Shiv Kumar and Carl W. Fuller 119

Abstract	119
1. Introduction	119
2. Fluorescent DNA sequencing	121
3. Energy transfer dye terminators	125
4. Terminal phosphate-labeled nucleotides	144
5. Conclusions	146
References	146

SEQUENCING BY SYNTHESIS PLATFORMS

Chapter 5. The 454 Life Sciences Picoliter Sequencing System

Marcel Margulies, Thomas P. Jarvie, James R. Knight and Jan Fredrik Simons 153

Abstract	153
1. Introduction	154
2. The 454 life sciences picoliter sequencing system	155
3. Applications	170
4. Discussion	182
Acknowledgments	184
References	184

Chapter 6. An Integrated System for DNA Sequencing by Synthesis

John R. Edwards, Dae Hyun Kim and Jingyue Ju 187

Abstract 187

1. Introduction 187

2. DNA sequencing by synthesis methodology 189

3. Conclusion 203

Acknowledgments 203

References 203

SINGLE-MOLECULE SEQUENCING**Chapter 7. Single-Molecule Fluorescence Microscopy and its Applications to Single-Molecule Sequencing by Cyclic Synthesis**

Benedict Hebert and Ido Braslavsky 209

Abstract 210

1. Introduction 210

2. Background 212

3. DNA sequencing by cyclic synthesis 219

4. Data analysis 230

5. Error sources in base calling 234

6. Performance 237

7. Applications 238

8. Conclusions 238

Acknowledgments 239

References 239

Chapter 8. Rapid DNA Sequencing by Direct Nanoscale Reading of Nucleotide Bases on Individual DNA chains

James Weifu Lee and Amit Meller 245

Abstract 245

1. Introduction 246

2. DNA sequencing by nanoelectrode-gated electron-tunneling conductance spectroscopic molecular detection 248

3. DNA sequencing by massively parallel optical readout of nanopore arrays and design polymer 256

4. Conclusion	260
Acknowledgments	261
References	261

Chapter 9. A Single Molecule System for Whole Genome Analysis

Shiguo Zhou, Jill Herschleb and David C. Schwartz	265
---	-----

Abstract	266
1. Introduction	266
2. The optical mapping system	273
3. The optical mapping system: image acquisition, processing, and analysis	280
4. Applications of optical mapping	287
5. Comparison of optical mapping and alternate methods for genome analysis	292
6. Optical sequencing	294
References	298

SEQUENCING VALIDATIONS AND ANALYSIS

Chapter 10. Sequencing Aided by Mutagenesis Facilitates the *De Novo* Sequencing of Megabase DNA Fragments by Short Read Lengths

Jonathan M. Keith, David B. Hawkes, Jacinta C. Carter, Duncan A. E. Cochran, Peter Adams, Darryn E. Bryant and Keith R. Mitchelson	303
--	-----

Abstract	304
1. Introduction	304
2. Principles of SAM sequencing	307
3. Simulated SAM sequencing	309
4. Analysis of SAM sequencing target assemblies	312
5. Discussion	319
References	325

Chapter 11. Genome Sequencing and Assembly

Annette McGrath	327
-----------------	-----

Abstract	327
1. Introduction	328
2. Approaches to genome sequencing	328
3. Problems inherent with genome assemblies	335

4. A mathematical model of shotgun sequencing	338
5. Genome assembly approaches and programs	339
6. New generation sequence assembly tools	343
7. Assembly of genomes by comparative means	347
8. Assembly of sequence data from emerging sequencing technologies	348
References	350

Chapter 12. Valid Recovery of Nucleic Acid Sequence Information from High Contamination Risk Samples – Ancient DNA and Environmental DNA

George A. Kowalchuk, Jeremy J. Austin, Paul S. Gooding and John R. Stephen	357
--	-----

Abstract	357
1. Introduction	358
2. Features of high contamination and artifact risk samples	359
3. Amplification and/or recovery of nucleic acids in the laboratory	363
4. Consideration in laboratory set-up	365
5. Looking to the future	367
References	368
Subject Index	373

Enabling Technologies

Chapter 1

Overview: Developments in DNA Sequencing

Keith R. Mitchelson,^{1,2} David B. Hawkes,³ Rustam Turakulov⁴ and Artem E. Men³

¹*Capitalbio Corporation: National Engineering Research Centre for Beijing Biochip Technology, 18 Life Science Parkway, Changping District, Beijing 102206, China*

²*Medical Systems Biology Research Center, Tsinghua University School of Medicine, Beijing 100084, China*

³*AGRF, Institute of Molecular Bioscience, University of Queensland, St. Lucia, Queensland 4072, Australia*

⁴*AGRF, Walter and Eliza Hall Institute for Medical Research, 1G Royal Parade, Parkville, Victoria 3050, Australia*

Contents

1. Introduction	4
1.1. Biotechnological implications of ultra-high-throughput sequencing capability	6
2. Advanced sequencing technologies	9
2.1. Capillary electrophoresis and Sanger sequencing	9
2.2. High-throughput capillary-array sequencing	9
2.3. Signal detection dyes and detectors	10
2.4. Microchip electrophoresis	11
2.5. Capillary electrophoretic sequencing on microcapillary chips	11
2.6. Sequencing by mass spectrometry	14
3. Solid-phase array sequencing devices	15
3.1. Ultra-sensitive detectors and sequencers	15
3.2. Sequencing by synthesis	15
3.3. Single DNA molecule sequencing	19
3.4. Hybridization re-sequencing	22
4. Future technologies	22
4.1. Nanopore membranes	23
4.2. Direct electrical detection of DNA synthesis	25
5. Applied short-read genomic sequencing	25
5.1. Genotyping by re-sequencing	25
5.1.1. Polony genotyping	26
5.1.2. Pyrosequencer genotyping	26
5.1.3. Polymorphism ratio sequencing	26
5.1.4. BEAMing	27
5.2. PaleoGenomics	27
5.3. Neanderthal genomics	28
5.4. MetaGenomics	29