Methods in Molecular Biology™

VOLUME 175

Genomics Protocols

Edited by

Michael P. Starkey
Ramnath Elaswarapu

HUMANA PRESS

METHODS IN MOLECULAR BIOLOGY™

Genomics Protocols

Edited by

Michael P. Starkey

and

Ramnath Elaswarapu

UK Human Genome Mapping Project Resource Centre, Cambridge, UK

Humana Press  **Totowa, New Jersey**

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

www.humanapress.com

This publication is printed on acid-free paper. ∞

ANSI Z39.48-1984 (American Standards Institute)

Permanence of Paper for Printed Library Materials.

Cover design by Patricia F. Cleary.

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: humana@humanapr.com; or visit our Website: www.humanapress.com

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging in Publication Data

Genomics protocols / edited by Michael P. Starkey and Ramnath Elaswarapu.
p. ; cm.—(Methods in molecular biology ; 175)
Includes bibliographical references and index.
ISBN 0-89603-774-6 (hardcover ; alk. paper)--ISBN 0-89603-708-8 (comb ; alk. paper)
1. Molecular genetics—Laboratory manuals. 2. Genomics–Laboratory manuals. I. Starkey, Michael P. II. Elaswarapu, Ramnath. III. Series QH440.5 G46 2001]
572.8—dc21

00-046106

Preface

We must unashamedly admit that a large part of the motivation for editing *Genomics Protocols* was selfish. The possibility of assembling in a single volume a unique and comprehensive collection of complete protocols, relevant to our work and the work of our colleagues, was too good an opportunity to miss. We are pleased to report, however, that the outcome is something of use not only to those who are experienced practitioners in the genomics field, but is also valuable to the larger community of researchers who have recognized the potential of genomics research and may themselves be beginning to explore the technologies involved.

Some of the techniques described in *Genomics Protocols* are clearly not restricted to the genomics field; indeed, a prerequisite for many procedures in this discipline is that they require an extremely high throughput, beyond the scope of the average investigator. However, what we have endeavored here to achieve is both to compile a collection of procedures concerned with genome-scale investigations and to incorporate the key components of "bottom-up" and "top-down" approaches to gene finding. The technologies described extend from those traditionally recognized as coming under the genomics umbrella, touch on proteomics (the study of the expressed protein complement of the genome), through to early therapeutic approaches utilizing the potential of genome programs via gene therapy (Chapters 27–30).

Although a number of the procedures described represent the tried and trusted, we have striven to include new variants on existing technologies in addition to exciting new approaches. Where there are alternative approaches to achieving a particular goal, we have sought assistance from an expert in the field to identify the most reliable technique, one suitable for a beginner in the field. Unique to the Methods in Molecular Biology series is the "Notes" section at the end of each chapter. This is a veritable Aladdin's cave of information in which an investigator describes the quirks in a procedure and the little tricks that make all the difference to a successful outcome.

The first section of the volume deals with the traditional positional cloning approach to gene identification and isolation. The construction of a high-resolution genetic map (Chapter 1) to facilitate the mapping of monogenic traits

and approaches to the analysis of polygenic traits (Chapter 2) are described. Identification of large numbers of single-nucleotide polymorphisms (Chapter 3) will pave the way for the construction of the next generation of genetic maps. Also described are such comparatively new technologies as genomic mismatch scanning (Chapter 4), for the mapping of genetic traits, and comparative genomic hybridization (Chapter 5), for the identification of gross differences between genomes.

Such studies are a prelude to the screening of large genomic clones, or clone contigs (Chapter 7). These transitions are made possible by the localization of genomic clones (Chapter 8) and the integration of the genetic and physical maps (Chapter 9) achieved by STS mapping. Identification of cDNAs mapping to the genomic clones implicated (Chapters 12–14) is the next step toward candidate gene identification. With the desire to acquire cDNAs capable of expressing authentic proteins, the emphasis in cDNA library construction is placed on a technology capable of delivering full-length cDNAs (Chapter 10).

One of the consequences of genome-scale sequencing programs has been the need to annotate large stretches of anonymous sequence data, and this has been the impetus for an explosion of bioinformatics programs targeted at gene prediction (Chapter 16). The use of model organisms (Chapter 17) to expedite gene discovery, on the basis of coding sequence similarites between genes with similiar functions, is another tool accessible to the gene hunter.

As an alternative to genetic studies, expression profiling seeks to identify candidate genes on the basis of their differential patterns of expression, either at the level of transcription or translation. A number of technologies, based on subtractive hybridization, differential display, and high throughput *in situ* hybridization are thus described (Chapters 18–22).

Functional characterization of isolated cDNAs is the next stage in establishing the likely candidature and thus potential utility of genes isolated as targets for therapeutic intervention. Predictions of protein structure and function (Chapter 23), mutagenesis (Chapter 24), or knockout studies (Chapter 25) can enable predictions of gene function. The yeast two-hybrid system (Chapter 26) is described at the level of monitoring interaction between individual proteins, but also on a potential genome scale.

In compiling *Genomics Protocols,* the aim—as with all other volumes in the Methods in Molecular Biology series—has been to produce a self-contained laboratory manual useful to both experienced practitioners and beginners in the field. We trust that we have been at least moderately successful. We must conclude by giving a vote of thanks to all the contributing authors, and to John Walker and the staff at Humana Press for seeing this project through.

Michael P. Starkey
Ramnath Elaswarapu

Contributors

TERRY J. AMISS • *Gene Therapy Center, University of North Carolina at Chapel Hill, Chapel Hill, NC*

DONALD S. ANSON • *Women's and Children's Hospital, North Adelaide, South Australia, Australia*

JAN P. A. BAAK • *Department of Pathology, Free University Hospital Amsterdam, Amsterdam, The Netherlands*

SIMON M. BROCKLEHURST • *Cambridge Antibody Technology, Melbourn, UK*

STEPHEN P. BRYANT • *Gemini Research Ltd., Cambridge, UK*

MATHIAS N. CHIANO • *Gemini Research Ltd., Cambridge, UK*

SANGDUN CHOI • *Division of Biology, California Institute of Technology, Pasadena, CA*

AMANDA COTTAGE • *Department of Pathology, Cambridge University, Cambridge, UK*

RONALD W. DAVIS • *Department of Biochemistry, Beckman Center, Stanford University School of Medicine, Stanford, CA*

PANAGIOTIS DELOUKAS • *The Sanger Centre, Cambridge, UK*

JOHAN T. DEN DUNNEN • *MGC-Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands*

YVONNE J. K. EDWARDS • *UK Human Genome Mapping Project Resource Centre, Cambridge, UK*

RAMNATH ELASWARAPU • *UK Human Genome Mapping Project Resource Centre, Cambridge, UK*

GREG ELGAR • *UK Human Genome Mapping Project Resource Centre, Cambridge, UK*

RUSSELL L. FINLEY, JR. • *Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, MI*

RUTH GJERSET • *Sidney Kimmel Cancer Center, San Diego, CA*

ERICA A. GOLEMIS • *Division of Basic Science, Fox Chase Cancer Center, Philadelphia, PA*

THOMAS M. GRESS • *Department of Internal Medicine I, University of Ulm, Ulm, Germany*
ALI HAGHIGHI • *Sidney Kimmel Cancer Center, San Diego, CA*
DAVID E. HARRIS • *The Sanger Centre, Cambridge, UK*
MARIO A. J. A. HERMSEN • *Department of Pathology, Free University Hospital Amsterdam, Amsterdam, The Netherlands*
FRANK T. HORRIGAN • *Department of Physiology, University of Pennsylvania School of Medicine, Philadelphia, PA*
SEAN J. HUMPHRAY • *The Sanger Centre, Cambridge, UK*
UNG-JIN KIM • *Division of Biology, California Institute of Technology, Pasadena, CA*
SUSAN J. KNAGGS • *Genomics Laboratory, Division of Medical and Molecular Genetics, UMDS, Guy's Hospital, London, UK*
NATALI KOLKER • *Department of Molecular Biotechnology, University of Washington, Seattle, WA*
RICHARD M. LAWN • *Falk Cardiovascular Research Center, Stanford University School of Medicine, Stanford, CA*
SVETLANA LEBEDEVA • *Sidney Kimmel Cancer Center, San Diego, CA*
MARGARET A. LEVERSHA • *Roy Castle International Centre for Lung Cancer Research, Liverpool, UK*
XING JIAN LOU • *Moleular Biology Systems Analysis, LumiCyte, Inc., Fremont, CA*
PAUL A. LYONS • *Department of Medical Genetics, Wellcome Trust Centre for Molecular Mechanisms in Disease, University of Cambridge, Cambridge, UK*
AHMED MANSOURI • *Department of Molecular and Cell Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany*
GERRIT A. MEIJER • *Department of Pathology, Free University Hospital Amsterdam, Amsterdam, The Netherlands*
DAN MERCOLA • *Sidney Kimmel Cancer Center, San Diego, CA; Cancer Center, University of California at San Diego, La Jolla, CA*
FARIDEH MIRZAYANS • *Department of Ophthalmology, Ocular Genetics Laboratory, University of Alberta, Edmonton, Alberta, Canada*
LEE MURPHY • *The Sanger Centre, Cambridge, UK*
DEBORAH A. NICKERSON • *Department of Molecular Biology, University of Washington, Seattle, WA*
CHRISTOF NIEHRS • *Division of Molecular Embryology, Deutsches Krebsforschungszentrum, Heidelberg, Germany*

MARIA PACK • *RZPD Deutsches Resourcenzentrum für Genomforschung GmbH, Berlin, Germany*

LEONIDAS A. PHYLACTOU • *Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus*

NICOLAS POLLET • *Division of Molecular Embryology, Deutsches Krebsforschungszentrum, Heidelberg, Germany*

IOANNIS RAGOUSSIS • *Genomics Laboratory, Division of Medical and Molecular Genetics, UMDS, Guy's Hospital, London, UK*

MARK J. RIEDER • *Department of Molecular Biology, University of Washington, Seattle, WA*

RICHARD JUDE SAMULSKI • *Gene Therapy Center, University of North Carolina at Chapel Hill, Chapel Hill, NC*

KATJA SCHÄFER • *RZPD Deutsches Resourcenzentrum für Genomforschung GmbH, Berlin, Germany*

MARK SCHENA • *Department of Biochemistry, Beckman Center, Stanford University School of Medicine, Stanford, CA*

ILYA G. SEREBRIISKII • *Division of Basic Science, Fox Chase Cancer Center, Philadelphia, PA*

MICHAEL P. STARKEY • *UK Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, UK*

SUMIO SUGANO • *Department of Virology, The Institute of Medical Sciences, University of Tokyo, Tokyo, Japan*

YUTAKA SUZUKI • *Department of Virology, The Institute of Medical Sciences, University of Tokyo, Tokyo, Japan*

SCOTT L. TAYLOR • *Division of Development and Neurobiology, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

TIM THOMAS • *Department of Molecular and Cell Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany*

GARABET G. TOBY • *Division of Basic Science, Fox Chase Cancer Center, Philadelphia, PA, and Cell and Molecular Biology Graduate Group, University of Pennsylvania, Philadelphia, PA*

DANIELA TONIOLO • *Institute of Genetics, Biochemistry, and Evolution, CNR, Pavia, Italy*

OLIVER DORIAN VON STEIN • *InDex Pharmaceuticals AB, Stockholm, Sweden*

ANNE K. VOSS • *Division of Development and Neurobiology, Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

CHRISTINE WALLRAPP • *Department of Internal Medicine I, University of Ulm, Ulm, Germany*

MICHAEL A. WALTER • *Department of Ophthalmology, Ocular Genetics Laboratory, University of Alberta, Edmonton, Alberta, Canada*
MARTIN C. WAPENAAR • *MGC-Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands*
MARJAN M. WEISS • *Department of Gastroenterology, Free University Hospital Amsterdam, Amsterdam, The Netherlands*
GÜNTHER ZEHETNER • *Max-Planck-Institut für Molekulare Genetik, Berlin, Germany*

Contents

1

Construction of Microsatellite-Based, High-Resolution Genetic Maps in the Mouse

Paul A. Lyons

1. Introduction

The mapping of genes underlying either simple mendelian or complex traits can be broken down into a number of distinct stages—initial detection of the locus in a genome scan, determination of the most likely map location for the gene, and finally fine mapping of the locus. A number of experimental strategies for mapping genes in experimental organisms are available and their relative merits have been reviewed recently *(1)*. Whatever strategy is chosen, an essential prerequisite for any gene identification project is the ability to construct a high resolution genetic map around the locus of interest.

The focus of this chapter is the construction of such genetic maps using microsatellite markers in the mouse, however, the methodology described here is applicable to most experimental organisms for which microsatellite markers are available. The mapping process can be broken down into a number of discrete steps. The first step is selecting the experimental strategy and determining the numbers of mice required to give the desired resolution. For the purpose of this chapter it is assumed that a suitable experimental strategy has been chosen and the requisite number of mice have been bred. The next step is selection and polymerase chain reaction (PCR) optimization of a panel of microsatellite markers from the region of interest that are variant between the mouse strains being used. **Subheading 3.1.** of this chapter discusses criteria for selecting markers and provides sources of microsatellite markers available in the public databases. In **Subheading 3.2.** protocols are provided for the PCR optimization of selected microsatellite markers. The next step in the procedure is the preparation of DNA from samples for genotyping. **Subheading 3.3.**

From: *Methods in Molecular Biology, vol. 175: Genomics Protocols*
Edited by: M. P. Starkey and R. Elaswarapu

describes a protocol for the rapid extraction of DNA from mouse tails that is of a suitable quality for PCR analysis. **Sections 3.4.–3.6.** describe protocols for genotyping these DNA samples using either fluorescent or nonfluorescent-based approaches. The final step in the procedure, as outlined in **Subheading 3.7.**, is the construction of a genetic map from the genotyping data that has been obtained.

2. Materials

1. Tail buffer: 50 m*M* Tris-HCl, pH 8.0, 100 m*M* ethylenediaminetetraacetic acid (EDTA), 100 m*M* NaC1 and 1% (w/v) sodium dodecyl sulfate (SDS).
2. Proteinase K solution (10 mg/mL). Store in aliquots at –20°C.
3. Saturated NaC1 solution.
4. 1X $TE_{0.1}$: 10 m*M* Tris-HCl, pH 8.0, 0.1 m*M* EDTA.
5. 4 m*M* deoxyribonucleoside triphosphate (dNTPs).
6. Thermocycler (MJ Research, Watertown, MA).
7. PCR mix: Make 2000 reaction batches of PCR mix as follows. To 9.5 mL of dH_2O add 3 mL of 1OX TaqGold buffer (PE Biosystems, Warrington, UK) and 1.5 mL of 4 m*M* dNTPs. Mix and store at 4°C.
8. TaqGold DNA polymerase (PE Biosystems).
9. Nusieve agarose (Flowgen, Lichfield, UK).
10. Agarose loading buffer: 10 m*M* Tris-HCl, pH 8.0, 1 m*M* EDTA, 1% (w/v) SDS, 40% (w/v) sucrose, xylene cyanol, and bromophenol blue.
11. GS500 tamra-size standards (PE Biosystems).
12. Long-Ranger acrylamide/urea sequencing gel mix (Flowgen).
13. Acrylamide loading buffer: 90% (v/v) deionized formamide, 50 m*M* EDTA, pH 8.0, and dextran blue.
14. Deep-well titer plates (Beckman, High Wycombe, UK, cat. no. 267004).
15. Genescan and genotyper software (PE Biosystems).
16. ABI 377 Automated Sequencer (PE Biosystems).

3. Methods

Microsatellites are regions of DNA made up of repeating blocks of nucleotides where the length of the repeated unit is either 2 bp (dinucleotide repeats), 3 bp (trinucleotide repeats), or 4 bp (tetranucleotide repeats). Microsatellites are widely distributed throughout the mouse genome with the (CA)n dinucleotide repeat estimated as occurring 100,000 times *(2)*. In addition to being widely distributed, the number of repeat units, and hence the size of the microsatellite, varies between mouse strains, even among closely related inbred mouse strains. This variation in size can be readily followed by PCR amplification and gel electrophoresis, which makes microsatellites an ideal source of markers for genetic map construction *(3)*.

3.1. Microsatellite Marker Selection

1. Sources of microsatellite markers: Over the past decade a large effort has gone into generating, characterizing, and mapping microsatellite markers. The largest effort has come from Eric Lander and colleagues at the Whitehead Institute in Cambridge, MA, who have generated a map of over 6000 markers, with an average spacing of one every 0.2 c*M*, throughout the mouse genome *(4)*. Information regarding microsatellite markers developed at the Whitehead Institute, including primer sequences, chromosomal location, and allele sizes in a panel of inbred strains is readily accessible via the Internet *(5)*. Another major source of marker information is the Mouse Genome Database, which is maintained by the Jackson Laboratories *(6)*. This database acts as a central repository for mouse genetic mapping data, including marker information, and is updated on a regular basis.
2. Marker selection: An important consideration when selecting microsatellite markers for use is how the genotyping will be performed, that is, whether markers will be analyzed using fluorescence-based gel systems or nonfluorescence-based systems. For nonfluorescence-based genotyping analyzed on agarose gels, the allele sizes need to vary by at least 10% to be resolvable. For fluorescence based gel systems, this is not a consideration, as differences as small as 2 bp can be resolved. Another consideration is whether or not markers will be pooled for gel electrophoresis in which case markers with nonoverlapping allele size ranges should be chosen.

3.2. PCR Optimization of Microsatellite Markers

1. Prepare 10X working dilutions of each microsatellite primer pair as follows: Dilute the forward and reverse stock primers together in a single tube to a final concentration of 25 µg/mL of each primer.
2. For each microsatellite primer pair being titrated, prepare a master mix as follows:
   a. Aliquot 105 µL of PCR mix into a microfuge tube.
   b. Add 22.5 µL of 10X primer dilution and 1.5 µL of TaqGold polymerase.
   c. Mix by vortexing briefly and place on ice.
3. Set up PCR reactions in three microtiter plates as follows at room temperature: For each primer pair being titrated, aliquot 5 µL of mouse genomic DNA (8 µg/mL) into four wells of the microtiter plate. To each well add 1.5 µL of either 10 m*M*, 20 m*M*, 30 m*M*, or 40 m*M* $MgCl_2$ solution and 8.5 µL of master mix (final reaction volume 15 µL). If appropriate, overlay with one drop of mineral oil.
4. Centrifuge the microtiter plates briefly and place on a thermocycler.
5. PCR the first microtiter plate as follows: 94°C 10 min followed by 36 cycles of 94°C for 10 s, 55°C for 20 s, and 72°C for 20 s. For the two subsequent PCR plates, adjust the 55°C annealing temperature to 53°C and 50°C, respectively (*see* **Note 1**).
6. Prepare a 2% (w/v) agarose gel in 1X TBE buffer.
7. Add 1.5 µL of agarose loading buffer to the samples in each microtiter plate and centrifuge briefly to mix. Load 10 µL of sample onto a 2% agarose gel and electrophorese until the xylene cyanol dye has migrated approx 2 cm.

8. Determine the optimal PCR conditions by visualizing the PCR products on a UV-transilluminator. Select the Mg^{2+} concentration and annealing temperature that gives a strong, discrete band of the expected size PCR product (*see* **Note 2** and **Fig. 1A**).

3.3. DNA Extraction from Mouse Tails

1. Cut 1 cm of tail and place in a 1.5-mL microfuge tube on ice (*see* **Note 3**).
2. To each tail sample add 400 µL of tail buffer and 10 µL of proteinase K solution.
3. Incubate at 42°C overnight in a shaking incubator.
4. To each sample add 200 µL of saturated NaC1 solution. Mix well by shaking for 30 s, do not vortex.
5. Centrifuge at 18,000g for 20 min at room temperature in a benchtop centrifuge.
6. Transfer the DNA containing supernatant to a fresh 1.5-mL microfuge tube being careful not to disturb the pellet.
7. Add 800 µL of 100% ethanol to each sample and mix by gentle inversion (*see* **Note 4**).
8. Pellet the DNA precipitate by centrifuging at 18,000g for 3 min at room temperature.
9. Remove the supernatant and wash the pellet with 500 µL of 70% ethanol to remove excess salt.
10. Centrifuge at 18,000*g* for 1 min, carefully remove the supernatant and allow the DNA pellet to air dry briefly.
11. Gently resuspend the DNA pellet in 200 µL of 1X $TE_{0.1}$ (*see* **Note 5**).
12. Measure the DNA concentration of each stock solution at OD_{260} with a spectrophotometer.
13. Prepare a working dilution (8 µg/mL) of each sample by diluting in 1X $TE_{0.1}$. To facilitate downstream sample processing, prepare the dilutions in 96-well format deep-well titer plates.
14. Store the working dilutions at 4°C and the stock DNAs at –20°C.

3.4. PCR Amplification

1. For each microsatellite to be genotyped, prepare a master mix as follows: For each DNA sample add 7 µL of PCR mix, 1.5 µL of 10X $MgCl_2$ (as previously determined in **Subheading 3.2.**), 1.5 µL of 10X primer dilution and 0.1 µL of TaqGold polymerase. Mix by vortexing.
2. Aliquot 5 µL of genomic DNA (8 µg/mL) into a microtiter plate, add 10 µL of master mix and overlay with a drop of mineral oil, if necessary.
3. Centrifuge briefly and place on a thermocycler.
4. Perform PCR as follows: 94°C for 10 min followed by 36 cycles of 94°C for 10 s, *X*°C for 20 s and 72°C for 20 s, where *X* equals the optimal annealing temperature determined in **Subheading 3.2.** (*see* **Note 1**).
5. Store PCR products at –20°C prior to analysis.

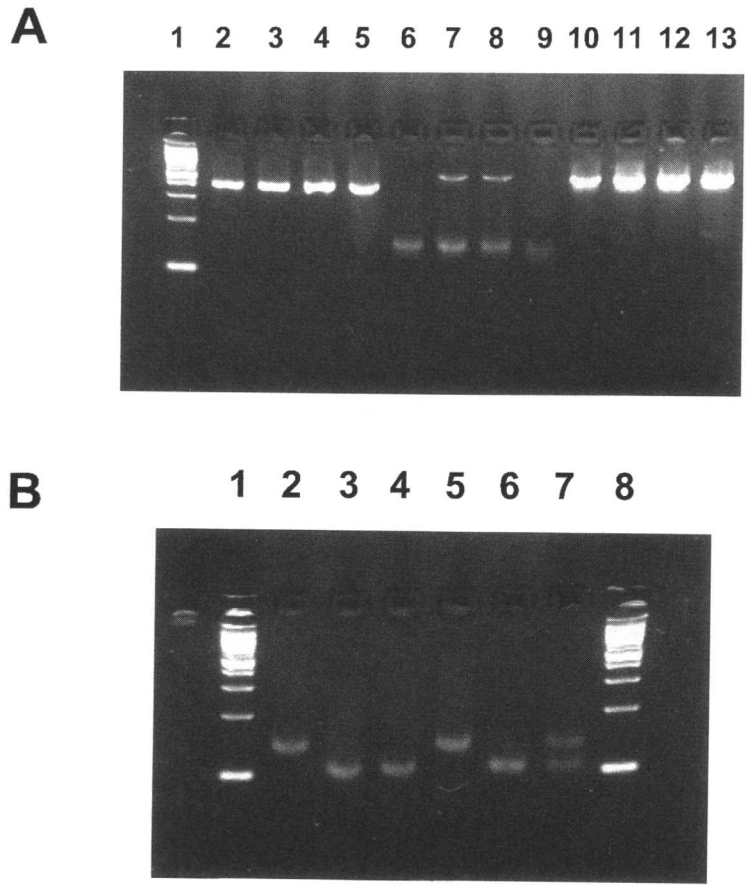

Fig. 1. PCR optimization of microsatellite markers. **(A)** Magnesium titrations of *DlNds31* (lanes 2–5), *DlNds32* (lanes 6–9), and *D4Nds26* (lanes 10–13) at 1 m*M* Mg^{2+} (lanes 2, 6, and 10), 2 m*M* Mg^{2+} (lanes 3, 7, and 11), 3 m*M* Mg^{2+} (lanes 4, 8, and 12), and 4 m*M* Mg^{2+} (lanes 5, 9, and 13). Lane 1 molecular-weight markers. **(B)** Amplification of C57BL/10 (lanes 2 and 5), NOD (lanes 3 and 6), and (NODxC57BL/10)F1 (lanes 4 and 7) DNA with *D3Nds6* using TaqGold (lanes 2-4) or Amplitaq (lanes 5–7) DNA polymerase. Lanes 1 and 8 are molecular-weight markers.

3.5. Analysis of PCR Products by Gel Electrophoresis

3.5.1. Agarose-Resolvable PCR Products

1. Prepare a 3% (w/v) Nusieve agarose/1% (w/v) agarose gel in 1X TBE.
2. Using a multichannel pipet add 1.5 µL of agarose loading buffer to each sample and centrifuge briefly to mix.