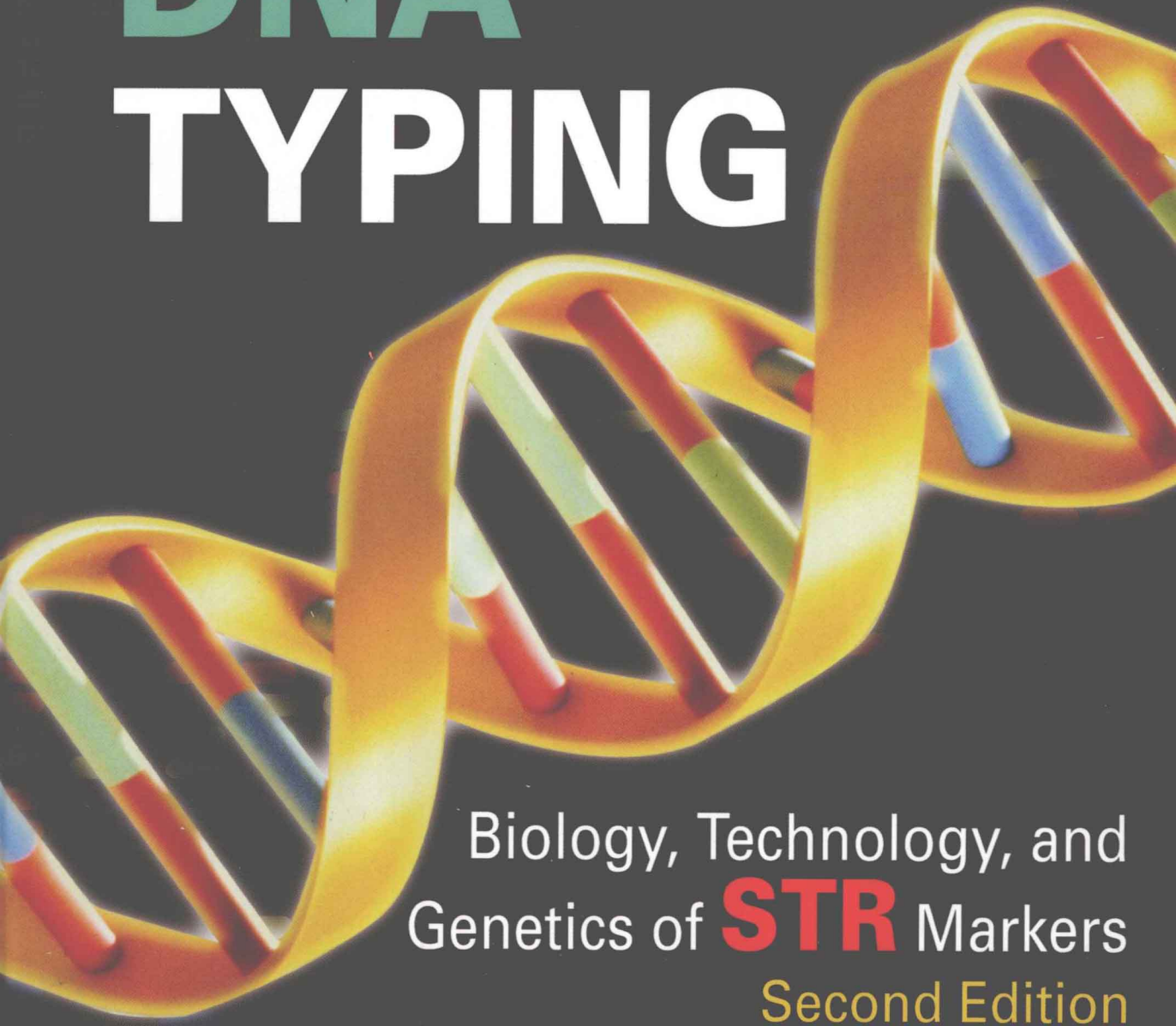
ELSEVIER
ACADEMIC
PRESS

JOHN M. BUTLER

FORENSIC DNA TYPING



Biology, Technology, and Genetics of STR Markers

Second Edition

FORENSIC DNA TYPING

BIOLOGY, TECHNOLOGY, AND GENETICS OF STR MARKERS

Second edition

John M. Butler

ELSEVIER
ACADEMIC
PRESS

Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

This book is printed on acid-free paper

Elsevier Academic Press
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
http://www.elsevier.com

Elsevier Academic Press
84 Theobald's Road, London WC1X 8RR, UK
http://www.elsevier.com

Library of Congress Catalog Number: 2004114214

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0-12-147952-8

Acquisitions Editor: *Mark Listewnik*
Associate Acquisitions Editor: *Jennifer Soucy*
Developmental Editor: *Pamela Chester*
Marketing Manager: *Christian Nolin*
Cover Designer: *Eric DeCicco*

05 06 07 08 09 8 7 6 5 4 3 2 1

FORENSIC DNA TYPING

FOREWORD

Forensic DNA Typing charts the progress and development of DNA applied to criminal forensics, providing vivid demonstrations of the amazing potential of the method, not only to convict the guilty but also to exonerate the innocent. John Butler has created a text that caters to all audiences, covering the basics of DNA structure and function and describing in detail how the techniques are used. In addition, the extensive use of D.N.A. (Data, Notes, and Application) Boxes in the text enables the reader to dip in and out as he or she pleases.

Probably the most important development of recent years is the universal use of polymerase chain reaction (PCR) to replicate DNA molecules *in vitro*. This has led to the rapid development of new platforms and biochemistry that have revolutionized the methods used to carry out DNA analysis. These new technologies are clearly explained in great detail in this book, with lavish illustrations. The culmination of recent advances has led to the instigation of massive National DNA offender databases using short tandem repeat (STR) loci. For example, since its inception in 1995, the England and Wales National DNA database (NDNADB) now has more than 2.75 million reference DNA profiles from suspects and offenders alike, against which all crime stains are routinely compared. Many more countries throughout the world have since followed suit. The social benefits of such databases are considerable – individuals who commit major crimes such as murder usually already have a criminal record. UK policy enables the collection of DNA profiles from all offenders regardless of the seriousness of the crime. Consequently, those who re-offend can be quickly identified and apprehended. In the US thirteen different STRs are combined together into one or two-tube reactions known as multiplexes to provide data for the Combined DNA Index System (CODIS). When a complete DNA profile is obtained the probability of a chance match with a randomly chosen individual is usually less than one in one trillion using these 13 CODIS loci.

Other areas are explored in detail including mitochondrial, Y chromosomal DNA and use of forensic science in wildlife crime such as poaching. Recently, as a result of terrorist attacks, new areas of forensic DNA profiling have arisen in response. Foremost amongst these is the field of microbial forensics, which is used to identify pathogens such as anthrax.

Although it is very difficult to anticipate all future developments, STRs are probably the system of choice for the foreseeable future although other systems, especially single nucleotide polymorphisms (SNPs), have been suggested. SNPs may find a special niche to analyze very highly degraded material and may play a valuable role in future mass disasters. However, there is no doubt that the utility of both STRs and SNPs will benefit from new biochemistry and new platforms such as microchips. Automation, miniaturization and expert systems will all play an increasingly important role over the coming years. The main aims of the new technology can be summarized: to enable faster processing; to reduce costs; to improve sensitivity; to produce portable instruments; to de-skill and to automate the interpretation process; to improve success rates; to improve quality of the result and to standardize processes. The next few years will probably see a new revolution as this new technology comes of age and becomes widely available.

John Butler reviews these new innovations in great detail – he is to be congratulated for preparing such a readable book that will appeal to everyone from the layperson, the lawyer and the scientist alike.

PETER GILL, Ph.D.
Birmingham, UK
December 2004

INTRODUCTION

An expert is one who knows more and more about less and less until they know absolutely everything about nothing...

(First part by Nicholas Butler, Bartlett's 585:10)

The work is its own reward.

(Sherlock Holmes, *The Adventure of the Norwood Builder*)

Several significant things have happened since the first edition of *Forensic DNA Typing* was published in January 2001. The Human Genome Project published a draft sequence of the human genome in February 2001 and completed the 'finished' reference sequence in April 2003. In addition, human mitochondrial DNA population genomics is underway and more than a thousand full mitochondrial genomes have been published. Technology for DNA sequencing and typing continues to advance as does our understanding of genetic variation in various population groups around the world. These milestones are a tribute to the progress of science and will benefit the field of forensic DNA typing.

The literature on the short tandem repeat (STR) markers used in forensic DNA testing has more than doubled in the four years since writing the first edition. More than 2000 publications now detail the technology and report the allele frequencies for forensically-informative STR loci. Hundreds of different population groups have been studied; new technologies for rapidly typing DNA samples have been developed, and standard protocols have been validated in laboratories worldwide. Yet DNA results are still sometimes challenged in court – not usually because of the technology, which is sound – but rather the ability of practitioners to perform the tests carefully and correctly. A major purpose of this book is to help in the training of professionals in the field of forensic DNA testing. The knowledge of forensic scientists, lawyers, and students coming into the field will be enhanced by careful review of the materials found herein.

The advent of modern DNA technology has resulted in the increased ability to perform human identity testing. Individual identification is desirable in a number of situations including the determination of perpetrators of violent

crime such as murder and rape, resolving unestablished paternity, and identifying remains of missing persons or victims of mass disasters.

In the past few years, the general public has become more familiar with the power of DNA typing as the media has covered efforts in identifying remains from victims of the World Trade Center twin towers collapse following the terrorist attacks of 11 September 2001, the O.J. Simpson murder trial, the President Clinton–Monica Lewinsky scandal, and the identification of the remains in the Tomb of the Unknown Soldier. In addition, our perceptions of history have been changed with DNA evidence that revealed Thomas Jefferson may have fathered a child by one of his slaves.

These cases have certainly attracted widespread media attention in recent years, however, they are only a small fraction of the thousands of forensic DNA and paternity cases that are conducted each year by public and private laboratories around the world. The technology for performing DNA typing has evolved rapidly since the 1990s to the point where it is now possible to obtain results in a few hours on samples with only the smallest amount of biological material.

This book will examine the science of current forensic DNA typing methods by focusing on the biology, technology, and genetic interpretation of short tandem repeat (STR) markers, which encompass the most common forensic DNA analysis methods used today. The materials in this book are intended primarily for two audiences: forensic scientists who want to gain a better understanding of STRs and professionals in the law enforcement and legal communities who find it hard to comprehend the complexities of DNA profiling. This text should also directly benefit college students learning more about forensic DNA analysis in an academic environment. The references cited at the end of each chapter provide a fairly comprehensive view of this dynamic field.

This book is also intended to aid forensic DNA laboratories in meeting the training requirements stated in the DNA Advisory Board Quality Assurance Standards. These standards are striving to improve the quality of work performed in forensic laboratories by requiring technical managers and DNA examiners to have training in biochemistry, genetics and molecular biology in order to gain a basic understanding of the foundation of forensic DNA analysis. See Standard 5.2.1 and 5.3.1 in Appendix IV of this book.

NEW MATERIAL IN THIS SECOND EDITION

Since the first edition was written in the winter months of 2000, the published literature has grown dramatically on the topic of STR typing and its use in forensic DNA testing. With more than 2000 papers now available describing STR markers, technology for typing these STRs, and allele frequencies in various populations around the world, the scientific basis for forensic DNA typing

is sound. The basic foundational material in the first edition is still relevant and thus has remained essentially unchanged. However, ten new chapters have been added to accommodate the explosion of new information since the turn of the century.

New topics such as single nucleotide polymorphisms (SNPs) and Y chromosome testing have gained greater acceptance within the forensic community since 2000 and therefore have become areas of expansion in this edition. A very comprehensive look at mitochondrial DNA and its application to forensic DNA analysis is included in Chapter 10. There is updated information on new DNA extraction procedures, real-time PCR for DNA quantification, multi-capillary electrophoresis instruments, and 5-dye chemistries that are now used in many forensic DNA laboratories. Citations have expanded to include more than 500 new literature references enabling readers to find original source material or to conduct extensive background research on the various topics covered herein and more than 50 new figures and 45 new tables containing helpful information have been added in this second edition of *Forensic DNA Typing*.

Statistical issues with data analysis and interpretation that were missing in the first edition are covered in Chapters 19–23 in this new edition. Extensive examples are provided for each equation discussed and corresponding population data can be found in Appendix II to enable readers to review the source of conclusions reached. Another appendix includes the description of a hypothetical case from start to finish in an attempt to bring together the information discussed throughout the book and to aid in training students and professionals in the field.

In this edition, we utilize Data, Notes, and Applications (D.N.A.) Boxes to cover specific topics of general interest. Many of the high-profile cases included in the last chapter of the first edition, such as the O.J. Simpson trial, are now scattered throughout the book near the sections dealing with the science or issues behind these cases. It is hoped that these D.N.A. Boxes will help readers see the practical value of forensic DNA typing.

AN OVERVIEW OF THE BOOK CHAPTERS

The book has been divided into three primary sections covering the biology, technology, and statistical analysis (genetics) of STR markers. Within each section, the chapters progress from basic introductory information to on-going 'cutting-edge' research. The first few chapters in particular are meant as introductory material for those readers who might be less familiar with DNA or as a review of useful materials for more advanced readers. The biology section is contained in Chapters 2 through 11, the technology section involves Chapters 12 through 18, and the genetics section may be found in Chapters 19 to 23. The final chapter examines the use of DNA testing in mass disaster victim

identification efforts, which include the greatest national tragedy in U.S. history, the events of 11 September 2001.

BIOLOGY SECTION

The book begins with an overview and history of DNA and its use in human identification. An actual criminal investigation where DNA evidence proved crucial is used to illustrate the value of this technology to law enforcement. Chapter 2 provides some basic information on DNA structure and function while Chapter 3 covers the processes involved in preparing samples for DNA amplification via the polymerase chain reaction, which is discussed further in Chapter 4. Chapter 5 focuses on the 13 commonly used STR markers in the United States today with details about naming of alleles and unique characteristics of each marker. Chapter 6 goes into the biology of STR markers including stutter products, non-template addition, microvariants, and null alleles. These aspects can complicate data interpretation if they are not understood properly. Chapter 7 discusses issues that are unique to the forensic DNA community, namely mixtures, degraded DNA samples, PCR inhibition, and contamination, all of which impact forensic casework since many samples do not come from a pristine, controlled environment. What was previously Chapter 8 in the first edition that discussed additional markers used in conjunction with STRs to aid in human identification has now been expanded upon in four additional chapters. The new Chapter 8 covers single nucleotide polymorphisms (SNPs) and technologies for typing them. Chapter 9 reviews Y chromosome markers for specifically identifying the male contributor of a sample and Chapter 10 discusses maternally inherited mitochondrial DNA, the use of which often provides results in situations involving highly degraded DNA. Finally, we touch on the use of non-human DNA to aid forensic investigations in Chapter 11 through a discussion of animal, plant, and microbial DNA testing.

TECHNOLOGY SECTION

The technology portion of the book begins in Chapter 12 with a discussion of DNA separations using slab gel and capillary electrophoresis. Fluorescent detection methods are the primary topic of Chapter 13. This chapter has a number of colorful figures featuring the fluorescent dyes in use today. A description of the most widely used DNA analysis instruments in modern forensic laboratories is presented in Chapter 14. This chapter covers the ABI Prism 310 Genetic Analyzer, the 16-capillary array ABI Prism 3100, and reviews the Hitachi FMBIO II Fluorescence Imaging System used in conjunction with slab gel electrophoresis. Issues surrounding genotyping of STR results are the focus of Chapter 15 and Chapter 16 reviews laboratory validation and quality assurance of DNA analysis.

Alternative DNA analysis technologies such as mass spectrometry and microchips are reviewed in Chapter 17 along with robotics and expert systems for automated data analysis. The final chapter in the technology section, Chapter 18, discusses the use of computer DNA databases to solve crimes. Large national DNA databases will continue to benefit law enforcement for many years to come by connecting violent crimes and serial criminal activity with otherwise unknown perpetrators.

GENETICS AND STATISTICAL ANALYSIS SECTION

The genetics and statistical analysis section begins with a review of genetic principles and statistics in Chapter 19. Chapter 20 discusses Hardy–Weinberg and linkage equilibrium and the role of these calculations in checking performance of genetic markers in population databases. Calculations for DNA profile frequency estimates including random match probabilities and likelihood ratios are covered in Chapter 21. Chapter 22 discusses approaches to interpreting mixtures or partial profiles resulting from degraded DNA. Finally, Chapter 23 deals with kinship and paternity testing situations. Throughout all of these sections clear examples are used to demonstrate how equations are applied to the calculations required.

MASS DISASTER VICTIM IDENTIFICATION: UTILIZING BIOLOGY, TECHNOLOGY, AND GENETICS

Forensic DNA laboratories may be called upon to assist in victim identification following mass disasters such as airplane crashes or terrorist attacks. Situations where remains are highly fragmented prevent the use of fingerprints or dental records to rapidly determine the identity of each victim and DNA analysis often becomes the only method to bring closure to the chaos of such an event. To recognize the increasing role that DNA information is playing in mass disaster victim identification, we discuss here the application of STRs, mitochondrial DNA, and single nucleotide polymorphisms in the identification of individuals who died in the terrorist attacks of 11 September 2001 at the World Trade Center twin towers, the Pentagon, and Shanksville, Pennsylvania. The original information from the first edition on the Waco Branch Davidian fire (April 1993) and the airline crash of Swissair Flight 111 (September 1998) are retained to provide historical perspective.

APPENDICES

There are seven appendices at the back of the book that provide valuable supplemental material. Appendix I describes all reported alleles for the 13 CODIS

STR loci as of January 2004. Sequence information, where available, has been included along with the reference that first described the noted allele. As most laboratories now use either a Promega GenePrint® STR kit or an Applied Biosystems AmpFℓSTR® kit for PCR amplification, we have listed the expected size for each allele based on the sequence information. Appendix II lists some STR allele frequency information from U.S. populations of African-Americans, Caucasians, and Hispanics. This information is used for all statistical calculations performed in the book. Appendix III is a compilation of companies and organizations that are suppliers of DNA analysis equipment, products, and services. Approximately 100 companies are listed along with their addresses, phone numbers, internet web pages, and a brief description of their products and/or services. Appendix IV contains the DNA Advisory Board (DAB) Quality Assurance Standards that pertain to forensic DNA testing laboratories and convicted offender DNA databasing laboratories in the United States. These standards are important for laboratory validation and maintaining high quality results as DNA testing becomes more prevalent. Appendix V includes the DAB recommendations on statistics. Appendix VI reviews the National Research Council's *The Evaluation of Forensic DNA Evidence*, better known as NRC II, and the application of its recommendations to STR typing. Finally, Appendix VII provides two example forensic cases in an attempt to put the information contained in this book within a proper context.

ACKNOWLEDGMENTS

I express a special thanks to colleagues and fellow researchers who kindly provided important information and supplied some of the figures for this book. These individuals include Martin Bill, George Carmody, Mike Coble, David Duewer, Dan Ehrlich, Nicky Fildes, Lisa Forman, Ron Fourney, Lee Fraser, Chip Harding, Debbie Hobson, Bill Hudlow, Alice Isenberg, Margaret Kline, Carll Ladd, Demris Lee, Steve Lee, Bruce McCord, Steve Niezgoda, Richard Schoske, Jim Schumm, Bob Shaler, Melissa Smrz, Amanda Sozer, Kevin Sullivan, and Lois Tully. I am indebted to the dedicated project team members, past and present, who work with me at the U.S. National Institute of Standards and Technology: Peter Vallone, Margaret Kline, Janette Redman, Mike Coble, David Duewer, Jill Appleby, Amy Decker, Christian Ruitberg, and Richard Schoske. It is a pleasure to work with such supportive and hard-working scientists.

Several other people deserve specific recognition for their support of this endeavor. The information reported in this book was in large measure made possible by a comprehensive collection of references on the STR markers used in forensic DNA typing. For this collection now numbering more than 2000 references, I am indebted to the initial work of Christian Ruitberg for tirelessly collecting and cataloging these papers and the steady efforts of Janette Redman to monthly update this STR reference database. A complete listing of these references may be found at http://www.cstl.nist.gov/biotech/strbase. George Carmody reviewed the statistical materials in Chapters 19–24 and provided many valuable comments.

My wife Terilynne, who carefully reviewed the manuscript and made helpful suggestions, was always a constant support in the many hours that this project took away from my family. As the initial editor of all my written materials, Terilynne helped make the book more coherent and readable.

Since I was first exposed to forensic DNA typing in 1990 when a friend gave me a copy of Joseph Wambaugh's *The Blooding* to read, I have watched with wonder as the forensic DNA community has rapidly evolved. DNA testing that once took weeks can now be performed in a matter of hours. I enjoy being a part of the developments in this field and hope that this book will help many others come to better understand the principles behind the biology, technology, and genetics of STR markers.

ABOUT THE AUTHOR

John Marshall Butler grew up in the Midwest and enjoying science and law decided to pursue a career in forensic science at an early age. After completing an undergraduate education at Brigham Young University in chemistry, he moved east to pursue his graduate studies at the University of Virginia. While a graduate student, he enjoyed the unique opportunity of serving as an FBI Honors Intern and guest researcher for more than two years in the FBI Laboratory's Forensic Science Research Unit. His Ph.D. dissertation research, which was conducted at the FBI Academy in Quantico, Virginia, involved pioneering work in applying capillary electrophoresis to STR typing. After receiving his Ph.D. in 1995, Dr. Butler obtained a prestigious National Research Council postdoctoral fellowship to the National Institute of Standards and Technology (NIST). While a postdoc at NIST, he designed and built STRBase, the widely used Short Tandem Repeat Internet Database (http://www.cstl.nist.gov/biotech/strbase) that contains a wealth of standardized information on STRs used in human identity applications. Dr. Butler then went to California for several years to work as a staff scientist and project leader at a startup company named GeneTrace System to develop rapid DNA analysis technologies involving time-of-flight mass spectrometry. In the fall of 1999, he returned to NIST to lead their efforts in human identity testing with funding from the National Institute of Justice.

Dr. Butler received the Presidential Early Career Award for Scientists and Engineers from President George W. Bush in a White House ceremony held in July 2002. In September 2003, he was awarded the Scientific Prize of the International Society of Forensic Genetics, the first American to be given this honor by his scientific peers. Following the terrorist attacks of 11 September, 2001, Dr. Butler's expertise was sought to aid the DNA identification efforts, and he served as part of the distinguished World Trade Center Kinship and Data Analysis Panel (WTC KADAP). He is also a regular invited guest and participant in the semi-annual meetings of the FBI's Scientific Working Group on DNA Analysis Methods (SWGDAM). In addition, he serves on the Department of Defense Quality Assurance Oversight Committee for DNA Analysis and as a guest editor for the *Journal of Forensic Sciences*. His more than 65 publications in the field make him one of the most prolific active authors in the field with

articles appearing regularly in every major forensic science journal. He has been an invited speaker to numerous national and international forensic DNA meetings and in the past few years has spoken in Germany, France, England, Portugal, Cyprus, and Australia. He is well-qualified to present the information found in this book, much of which has come from his own research efforts over the past decade. In addition to his busy scientific career, Dr. Butler and his wife serve in their community and church and are the parents of five children, all of which have been proven to be theirs through the power of STR typing.

To my wife Terilynne and our children
Amanda, Marshall, Katy, Emma, and Ethan

CONTENTS