

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
<http://www.springer.com/series/7651>

DNA Barcodes

Methods and Protocols

Edited by

W. John Kress and David L. Erickson

*Department of Botany, National Museum of Natural History,
Smithsonian Institution, Washington, DC, USA*

 **Humana Press**

Editors

W. John Kress, Ph.D.
Department of Botany
National Museum of Natural History
Smithsonian Institution
Washington, DC, USA

David L. Erickson, Ph.D.
Department of Botany
National Museum of Natural History
Smithsonian Institution
Washington, DC, USA

ISSN 1064-3745 ISSN 1940-6029 (electronic)
ISBN 978-1-61779-590-9 ISBN 978-1-61779-591-6 (eBook)
DOI 10.1007/978-1-61779-591-6
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2012931933

© Springer Science+Business Media, LLC 2012

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Humana Press, c/o Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Humana Press is part of Springer Science+Business Media (www.springer.com)

Foreword

The diversity of life in a hectare of reef, a county of grassland, or a shipload of imports challenges biologists called to identify the species comprising biodiversity, functioning as ecosystems, or invading ports. The sequences of black-and-white barcodes that empower a newly hired clerk to wave a wand over a cart full of goods swiftly, print an itemized receipt infallibly, and order replacements invisibly call forth a vision of an analog for identifying species. The resemblance of barcodes on commercial products to sequences of DNA shown as black-and-white bars on electrophoretic gels reinforced the vision back in 2003 in the founding meetings of the barcode of life movement. This book edited by early adopters of DNA barcodes, John Kress and David Erickson, proves the barcode of life has arrived in environmental science. In less than a decade, they and the other authors in this volume have realized the vision of a short DNA sequence on a uniform locality of the genome to identify species rapidly and accurately.

Because the currency in biology is species, their identification is no academic diversion. Biologists count the rise and fall of biodiversity in species. Regulators designate endangered species by their identified populations and reserve land where they identify the endangered. Governments appraise the success of preservation in the currency of species. Inspectors define quarantines in identified species. Biologists carry the weight of these consequences as they select an exact name from almost two million known species names or conjecture that a specimen may belong to one of millions more unknown species.

Written as a sequence of four discrete nucleotides—CATG—along a uniform locality on a genome, a barcode of life provides a “digital” identifying feature, supplementing the more analog gradations of shapes, colors, and behaviors. A library of digital barcodes will provide an unambiguous reference that will facilitate identifying species invading and retreating across the globe and through centuries.

Making a difficult task harder, many species metamorphose into different forms as they cycle through stages in their lives. Eggs may become caterpillars and caterpillars become butterflies but, of course, all remain the same species carrying the same genes. Different species may resemble one another or be too small to distinguish easily but each carries different alleles and thus barcodes, which can unmask their identity. Furthermore, an inspector of unloaded cargo on a dock or an analyst of the remains of diets in a stomach may be called to identify species from only a snippet, a hair, or a fin. The fragment may be unrecognizable, but it will faithfully carry the identifying barcode of the source.

Since Carl Linnaeus (1707–1778) developed systematic naming, ranking, and classifying of organisms, biologists have produced master keys to all knowledge about a species in the form of binomial names. Biologists use distinguishing features, such as shape, color, or number of legs in taxonomic keys first to assign binomials, like *Homo sapiens*, and then to associate the names of the organisms with biological knowledge about the species and its relatives. Of course, the bank of names suffers some problems, such as when several names are applied to one species [1]. And, biologists continuously debate criteria for species. The diversity of life from bacteria to whales renders any single rule inadequate for defining all species. Nevertheless a few basic criteria, such as that distinct species do not interbreed and meld their genetic sequences, serve for many groups.

Since Charles Darwin (1809–1882) proposed a branching pattern of evolution in *On the Origin of Species*, biologists have sought to arrange a phylogenetic system of species on an evolutionary tree of life. A tree of life illustrates every introductory biology text. Barcoding will reveal whether a newly collected specimen belongs to a species already on the tree. Or if a specimen is a truly new species, barcoding will help place it as a new leaf among known species on the proper branch of the tree of life.

Whatever the criteria for defining and recognizing species, their inheritance and their genes must differ to maintain species distinctions generation after generation. Since the molecular discoveries of the mid-twentieth century, genes intimate a code comprising sequences of the four nucleotides that constitute DNA. Even before the barcoding movement now embodied in the 200 member organizations from 50 countries of the Consortium for the Barcode of Life (CBOL), scientific revisions of species boundaries included DNA analysis, and the ability to distinguish new species included DNA divergences.

The product barcode analogy lent impetus to the continuing matching of species and genetic differences. Commercial barcodes must be uniform across shelves and warehouses. For animals, concentration on the single segment of the mitochondrial COI gene across the far wider shelves of life imparted the necessary uniformity to avoid a Tower of Babel. Conceiving the series of nucleotides CATG as bars and their presence and order as digital bars opened the door to rapid and unambiguous connection of specimens. Instead of connecting biological specimens to shelves and suppliers, the DNA barcode of life connects them to curated collections in museums and herbaria, lifting their utility. It would also connect specimens to the biological literature of binomial names. DNA barcodes offer a globally consistent way to propose provisional or candidate species that experts have not yet honored with a full description and binomial name.

Worries at first evoked by DNA barcoding have not been realized. It has heightened the nuance of the species concept, not diminished it. It has widened humanity's view of diversity, not reduced diversity to ciphers. It has excited wonder at the knowledge hard-won through earlier techniques and accessible through the master key of binomial names. It has enhanced the need for systematists to match the flood of barcodes with a sound array of binomial names. Barcoding is not a mere slogan and an inadequate analogy. It is a now proven tool for understanding biodiversity.

Recurring to the need for uniformity to avoid a Tower of Babel, the choice of a segment of the mitochondrial COI gene has excelled for almost all animal taxa. This barcode region meets four basic specifications: the locality must be present in all barcoded species; it must be shaved as short as possible; the locality must have sequences stable within a species through many generations; and it must nevertheless have sequences variable enough to distinguish species. As this book reports, botanists have now also found barcode regions that are proving successful from carrots and chamomile to oats and pines. Fungal barcodes are not far behind.

Some observers do ask a single, searching question about the barcode of life arriving in environmental science: When will it be small, cheap, and convenient enough for nonexperts, even children? In particular, when will the needed equipment shrink to the size of a laptop or a handheld barcoder? In fact, even today the key machines have shrunk until they fit comfortably on a desk or tabletop.

The analogy of the newly hired clerk faultlessly pricing the cart of goods suggests the ability to make taxonomic expertise go further, and very far if a handheld barcoder were present. Clues that such a goal will be achieved lie in reports of students detecting endangered marine species on sale in supermarkets, identifying insect traces in their homes, and

analyzing tea leaves with inexpensive equipment. As well as enabling specialized scientists to do more and lift the value of specimen collections, barcoding promises to enable laymen to appreciate the diversity of life.

The array of opportunities offered by DNA barcodes must rest on a sound foundation of binomial names with associated, vouchered, and identified specimens—housed in readily accessible museum collections. A sound foundation of binomials based on new and existing natural history collections stands as the first priority for the success of DNA barcodes. Fortunately, the Global Names Architecture project associated with the Encyclopedia of Life has already amassed 19 million common and scientific names and is reconciling them for the two million or so species estimated to be known already. Within 5 years, we could celebrate the achievement of the international Barcode of Life (iBOL) project: access to the barcodes of an array of five million specimens sequenced from 500,000 species. Voucher specimens, which are prepared, curated, databased, often digitally imaged, and stored in natural history collections, will support this effort.

Already, in just a handful of years, the DNA barcode of life database (www.boldsystems.org) has soared above 1.2 million specimens from about 150,000 species. Already, as the chapters in this book show, the library of barcodes linked to names and curated specimens is multiplying the knowledge of a marine ecologist about a reef, the quality of surveillance for invasive species, and the accuracy of labeling of food products. Such successes will motivate and sustain the further building of the reference library of barcodes and the removal of obstacles for its quick, frugal realization and use.

Our vision, first inspired by a barcode wand in the hand of a supermarket clerk, is comparable magic for an ichthyologist on a research vessel with featureless fish larvae, a child on a woodland trail, or an inspector at a port infallibly identifying a species. Reading this book, we learn that science can make magic.

New York, NY, USA

*Jesse H. Ausubel
Alfred P. Sloan Foundation*

Reference

1. Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. *Trends Ecol Evol* 25:686–691.

Preface

The use of a universally accepted short DNA sequence for identification of species has been proposed for application across all forms of life. Such a “DNA barcode,” a term first coined less than a decade before the publication of the present book, in its simplest definition is one or more short gene sequences (<700 base pairs) taken from a standardized portion of the genome that are used to identify species through reference to DNA sequence libraries or databases. *DNA Barcodes: Methods and Protocols* is a compendium of the latest information on generating, applying, and analyzing DNA barcodes across the Tree of Life, including animals, fungi, protists, algae, and plants. The volume is divided into five major sections: I. Introduction; II. DNA Barcodes for the Tree of Life; III. Generating DNA Barcode Data; IV. Applications of DNA Barcode Data; and V. Case Studies on DNA Barcodes.

In preparing the volume, we recognized that DNA barcoding is much more than the sequencing of one or two genes from an organism. The endeavor has come to encompass many elements, from campaigns that provide a deterministic framework for how to build specimen libraries, to the bioinformatic systems needed to track the many samples, sequences, and meta-data that are linked to each individual. Our ability to apply DNA barcode data in diverse contexts is also critical, for just as a library of books is not useful without those who would apply the knowledge contained therein, so it is true that in applying DNA barcode data we fulfill its promise. To that end, this volume is intended as a roadmap, linking methods ranging from standard wet-lab methods, to methods of bioinformatics, statistical and ecological analysis and methods to guide future, large-scale collections campaigns.

In the Introductory Section, background material is provided on the rationale for the use of DNA barcodes as well as a short history of the development of the concept of DNA barcoding to the different domains of life. In Section II, detailed protocols and methodologies for barcoding various types of organisms are presented. Although the field of barcoding is still in its infancy, specific methodologies have now been developed for organisms across the Tree of Life, and these chapters describe the most successful methods employed thus far. Although some of these protocols are still evolving, some have become more or less standard for particular taxonomic groups.

Section III covers more broadly applicable topics that apply to barcoding any type of organism, such as sample acquisition and archiving, laboratory tracking of tissues and sequences, DNA extraction and amplification, using “mini-barcodes” for samples with degraded DNA, and generating barcodes with next-generation sequencing technology. In addition to these chapters on specific laboratory methodologies, Section IV is devoted to the applications of DNA barcodes in the fields of systematics, phylogeny, and community ecology. These chapters focus on analytic methods that in many cases are still in their infancy of development, but will be critical for those biologists who want to go beyond generating sequences for particular taxa and actually apply DNA barcodes to answer specific

questions in ecology and evolution. Topics include evaluating the efficacy of DNA barcodes, species discovery using DNA barcodes, constructing phylogenetic trees using DNA barcode sequence data, and applying such phylogenies to understanding test hypotheses concerning the assembly of species into communities.

In order to better understand how barcodes can be applied across specific taxonomic groups or to specific ecological situations, Section V provides two cases studies of ongoing, large-scale campaigns. The first case is a worldwide initiative to barcode the fishes of the world and the second example applies DNA barcoding to all tree species is represented in a worldwide network of forest dynamics plots (CTFS) as a tool for understanding community evolution an ecological forensics. DNA barcoding is a new and powerful basic research tool with exceptional potential for the incorporation of new technologies and for future applications. The volume closes with a vision by the Editors on the future of DNA barcoding.

This book should be of benefit and interest to all biologists and technicians interested in the relevance and application of molecular biology and DNA sequencing to identification, taxonomy, evolution, and ecology. We would like to thank all of the authors of the chapters included in this book for opening up their laboratories for all readers to see how their protocols were developed and how they work to generate and analyze DNA barcodes. All of the contributors to this volume recognize that DNA barcoding is a rapidly changing field and that new methods are being proposed almost on a monthly basis and we thank them for sharing their most up-to-date information. We all hope that the methods and protocols contained herein will be helpful in advancing these efforts.

Washington, DC, USA

*W. John Kress
David L. Erickson*

Contents

<i>Foreword</i>	<i>v</i>
<i>Preface</i>	<i>ix</i>
<i>Contributors</i>	<i>xiii</i>
PART I INTRODUCTION	
1 DNA Barcodes: Methods and Protocols	3
<i>W. John Kress and David L. Erickson</i>	
PART II DNA BARCODES FOR THE TREE OF LIFE	
2 Introduction to Animal DNA Barcoding Protocols	11
<i>Lee A. Weigt, Amy C. Driskell, Andrea Ormos, Christopher Meyer, and Allen Collins</i>	
3 DNA Barcodes for Insects	17
<i>John James Wilson</i>	
4 DNA Barcoding Methods for Invertebrates	47
<i>Nathaniel Evans and Gustav Paulay</i>	
5 DNA Barcoding Amphibians and Reptiles	79
<i>Miguel Vences, Zoltán T. Nagy, Gontran Sonet, and Erik Verbeyen</i>	
6 DNA Barcoding Fishes	109
<i>Lee A. Weigt, Amy C. Driskell, Carole C. Baldwin, and Andrea Ormos</i>	
7 DNA Barcoding Birds: From Field Collection to Data Analysis	127
<i>Dario A. Lijtmaer, Kevin C.R. Kerr, Mark Y. Stoeckle, and Pablo L. Tubaro</i>	
8 DNA Barcoding in Mammals	153
<i>Natalia V. Ivanova, Elizabeth L. Clare, and Alex V. Borisenko</i>	
9 Methods for DNA Barcoding of Fungi	183
<i>Ursula Eberhardt</i>	
10 Methods for DNA Barcoding Photosynthetic Protists Emphasizing the Macroalgae and Diatoms	207
<i>Gary W. Saunders and Daniel C. McDevit</i>	
11 DNA Barcoding Methods for Land Plants	223
<i>Aron J. Fazekas, Maria L. Kuzmina, Steven G. Newmaster, and Peter M. Hollingsworth</i>	
PART III GENERATING DNA BARCODE DATA	
12 Field Information Management Systems for DNA Barcoding	255
<i>John Deck, Joyce Gross, Steven Stones-Havas, Neil Davies, Rebecca Shapley, and Christopher Meyer</i>	

13 Laboratory Information Management Systems for DNA Barcoding. 269
*Meaghan Parker, Steven Stones-Havas, Craig Starger,
and Christopher Meyer*

14 DNA Extraction, Preservation, and Amplification 311
Thomas Knebelberger and Isabella Stöger

15 DNA Mini-barcodes 339
Mebrdad Hajibabaei and Charly McKenna

16 Ways to Mix Multiple PCR Amplicons into Single 454 Run
for DNA Barcoding 355
Ryuji J. Machida and Nancy Knowlton

PART IV APPLICATIONS OF DNA BARCODE DATA

17 The Practical Evaluation of DNA Barcode Efficacy 365
John L. Spouge and Leonardo Mariño-Ramírez

18 Plant DNA Barcodes, Taxonomic Management,
and Species Discovery in Tropical Forests 379
Christopher W. Dick and Campbell O. Webb

19 Construction and Analysis of Phylogenetic Trees Using DNA Barcode Data 395
David L. Erickson and Amy C. Driskell

20 Phylogenetic Analyses of Ecological Communities Using DNA Barcode Data 409
Nathan G. Swenson

PART V CASE STUDIES USING DNA BARCODES

21 FISH-BOL, A Case Study for DNA Barcodes 423
Robert D. Ward

22 Generating Plant DNA Barcodes for Trees in Long-Term
Forest Dynamics Plots 441
W. John Kress, Ida C. Lopez, and David L. Erickson

23 Future Directions 459
David L. Erickson and W. John Kress

Index 467

Contributors

- JESSE H. AUSUBEL • *Program for the Human Environment, The Rockefeller University, New York, NY, USA*
- CAROLE C. BALDWIN • *Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA*
- ALEX V. BORISENKO • *Biodiversity Institute of Ontario & Integrative Biology, University of Guelph, Guelph, ON, Canada*
- ALLEN COLLINS • *Department of Invertebrate Zoology, Smithsonian Institution, NMNH, Washington, DC, USA*
- ELIZABETH L. CLARE • *Biodiversity Institute of Ontario & Integrative Biology, University of Guelph, Guelph, ON, Canada*
- NEIL DAVIES • *Berkeley Natural History Museums, University of California at Berkeley, Berkeley, CA, USA*
- JOHN DECK • *Berkeley Natural History Museums, University of California at Berkeley, Berkeley, CA, USA*
- CHRISTOPHER W. DICK • *Department of Ecology and Evolutionary Biology and Herbarium, University of Michigan, Ann Arbor, MI, USA*
- AMY C. DRISKELL • *Laboratories of Analytical Biology, Smithsonian Institution, NMNH, Suitland, MD, USA*
- URSULA EBERHARDT • *CBS-KNAW Fungal Biodiversity Centre, Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands*
- DAVID L. ERICKSON • *Department of Botany, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA*
- NATHANIEL EVANS • *Florida Museum of Natural History, University of Florida, Gainesville, FL, USA*
- ARON J. FAZEKAS • *Department of Integrative Biology, University of Guelph, Guelph, ON, Canada*
- JOYCE GROSS • *Berkeley Natural History Museums, University of California at Berkeley, Berkeley, CA, USA*
- MEHRDAD HAJIBABAEI • *Biodiversity Institute of Ontario & Integrative Biology, University of Guelph, Guelph, ON, Canada*
- PETER M. HOLLINGSWORTH • *Royal Botanic Garden Edinburgh, Edinburgh, UK*
- NATALIA V. IVANOVA • *Biodiversity Institute of Ontario & Integrative Biology, University of Guelph, Guelph, ON, Canada*
- KEVIN C.R. KERR • *Department of Natural History, Royal Ontario Museum, Toronto, ON, Canada*
- THOMAS KNEBELSBERGER • *Senckenberg Research Institute, German Centre for Marine Biodiversity Research (DZMB), Wilhelmshaven, Germany*
- NANCY KNOWLTON • *Department of Invertebrate Zoology, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA*

- W. JOHN KRESS • *Department of Botany, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA*
- MARIA L. KUZMINA • *Biodiversity Institute of Ontario & Integrative Biology, University of Guelph, Guelph, ON, Canada*
- DARÍO A. LIJTMAER • *Ornithology, “Bernardino Rivadavia”, Buenos Aires, Argentina*
- IDA C. LOPEZ • *Department of Botany, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA*
- RYUJI J. MACHIDA • *Department of Invertebrate Zoology, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA*
- LEONARDO MARIÑO-RAMÍREZ • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*
- DANIEL C. MCDEVIT • *Department of Biology, Centre for Environmental & Molecular Algal Research, University of New Brunswick, Fredericton, NB, Canada*
- CHARLY MCKENNA • *Department of Integrative Biology and Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada*
- CHRISTOPHER MEYER • *Department of Invertebrate Zoology, MRC-163, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA*
- ZOLTÁN T. NAGY • *Royal Belgian Institute of Natural Sciences, Joint Experimental Molecular Unit, Brussels, Belgium*
- STEVEN G. NEWMASER • *Department of Integrative Biology, University of Guelph, Guelph, ON, Canada*
- ANDREA ORMOS • *Laboratories of Analytical Biology, Smithsonian Institution, NMNH, Suitland, MD, USA*
- MEAGHAN PARKER • *Department of Invertebrate Zoology, MRC-163, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA*
- GUSTAV PAULAY • *Florida Museum of Natural History, University of Florida, Gainesville, FL, USA*
- GARY W. SAUNDERS • *Department of Biology, Centre for Environmental & Molecular Algal Research, University of New Brunswick, Fredericton, NB, Canada*
- REBECCA SHAPLEY • *Structured Data Group, Google Research, Google Inc., Mountain View, CA, USA*
- GONTRAN SONET • *Royal Belgian Institute of Natural Sciences, Joint Experimental Molecular Unit, Brussels, Belgium*
- JOHN L. SPOUGE • *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA*
- CRAIG STARGER • *Department of Invertebrate Zoology, MRC-163, Smithsonian Institution, National Museum of Natural History, Washington, DC, USA*
- MARK Y. STOECKLE • *The Rockefeller University, New York, NY, USA*
- ISABELLA STÖGER • *Bavarian State Collection of Zoology (ZSM), Munich, Germany*
- STEVEN STONES-HAVAS • *BioMatters Ltd., Auckland, New Zealand*
- NATHAN G. SWENSON • *Department of Plant Biology, Michigan State University, East Lansing, MI, USA*
- PABLO L. TUBARO • *Ornithology, “Bernardino Rivadavia”, Buenos Aires, Argentina*
- MIGUEL VENCES • *Division of Evolutionary Biology Zoological Institute, Technical University of Braunschweig, Braunschweig, Germany*

ERIK VERHEYEN • *Royal Belgian Institute of Natural Sciences, Vertebrate department, Brussels, Belgium*

ROBERT D. WARD • *Wealth from Oceans Flagship, CSIRO Marine and Atmospheric Research, Tasmania, Australia*

CAMPBELL O. WEBB • *Arnold Arboretum of Harvard University, Cambridge, MA, USA*

LEE A. WEIGT • *Laboratories of Analytical Biology, Smithsonian Institution, NMNH, Suitland, MD, USA*

JOHN JAMES WILSON • *Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada*