## TOPOLOGICAL DATA ANALYSIS FOR GENOMICS AND EVOLUTION Topology in Biology

Biology has entered the age of Big Data. A technical revolution has transformed the field, and extracting meaningful information from large biological data sets is now a central methodological challenge. Algebraic topology is a well-established branch of pure mathematics that studies qualitative descriptors of the shape of geometric objects. It aims to reduce comparisons of shape to a comparison of algebraic invariants, such as numbers, which are typically easier to work with. Topological data analysis is a rapidly developing subfield that leverages the tools of algebraic topology to provide robust multiscale analysis of data sets. This book introduces the central ideas and techniques of topological data analysis and its specific applications to biology, including the evolution of viruses, bacteria and humans, genomics of cancer, and single cell characterization of developmental processes. Bridging two disciplines, the book is for researchers and graduate students in genomics and evolutionary biology as well as mathematicians interested in applied topology.

RAÚL RABADÁN is a Professor at Columbia University, New York. He is Director of the Program for Mathematical Genomics at Columbia University, New York, and the NCI Physics and Oncology Center for Topology of Cancer Evolution and Heterogeneity. Dr Rabadán received his Ph.D. in Theoretical Physics in 2001 and went on to conduct research at the European Laboratory for Particle Physics (CERN) in Switzerland, and at the Institute for Advanced Study (IAS) in Princeton, New Jersey. At Columbia University, he leads a highly interdisciplinary laboratory with researchers from the fields of mathematics, physics, computer science, engineering, and medicine, with the common goal of solving biomedical problems through quantitative computational models.

ANDREW J. BLUMBERG is a Professor in the Department of Mathematics at the University of Texas, Austin. He completed his Ph.D. at the University of Chicago under the supervision of Peter May and Michael Mandell, and was later a National Science Foundation postdoctoral fellow at Stanford. He also spent a year as a member at the Institute for Advanced Study (IAS) in Princeton, New Jersey. His pure mathematics research focuses primarily on homotopy theory and algebraic topology and his applied research focuses on the development of topological and geometric techniques for studying genomic data.

## TOPOLOGICAL DATA ANALYSIS FOR GENOMICS AND EVOLUTION

Topology in Biology

RAÚL RABADÁN

Columbia University, New York

ANDREW J. BLUMBERG University of Texas, Austin



## **CAMBRIDGE** UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India

## 79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org Information on this title: www.cambridge.org/9781107159549 DOI: 10.1017/9781316671665

© Raúl Rabadán and Andrew J. Blumberg 2020

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2020

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data Names: Rabadán, Raúl, author. | Blumberg, Andrew J., author. Title: Topological data analysis for genomics and evolution : topology in biology / Raúl Rabadán, Columbia University, New York, Andrew J. Blumberg, University of Texas, Austin. Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2019. | Includes bibliographical references and index. Identifiers: LCCN 2019002342 | ISBN 9781107159549 (hardback : alk. paper) Subjects: LCSH: Bioinformatics – Mathematical models. | Computational biology. | Mathematical analysis. Classification: LCC QH324.2 .R33 2019 | DDC 570.285–dc23 LC record available at https://lccn.loc.gov/2019002342

ISBN 978-1-107-15954-9 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.