## Bayesian Evolutionary Analysis with BEAST

What are the models used in phylogenetic analysis and what exactly is involved in Bayesian evolutionary analysis using Markov chain Monte Carlo (MCMC) methods? How can you choose and apply these models, which parameterisations and priors make sense, and how can you diagnose Bayesian MCMC when things go wrong?

These are just a few of the questions answered in this comprehensive overview of Bayesian approaches to phylogenetics. From addressing theoretical aspects of the field to providing pragmatic advice on how to prepare and perform phylogenetic analysis, this practical guide also includes coverage of the interpretation of analyses and visualisation of phylogenies. The software architecture is described and a guide to developing BEAST 2.2 extensions is provided to allow these models to be extended further.

With an accompanying website (http://beast2.org/) providing example files and tutorials, this one-stop reference to applying the latest phylogenetic models in BEAST 2 will provide essential guidance for all users – from those using phylogenetic tools, to computational biologists and Bayesian statisticians.

**Alexei J. Drummond** is Professor of Computational Biology and Principal Investigator at the Allan Wilson Centre of Molecular Ecology and Evolution. He is the lead author of the BEAST software package and has gained a reputation in the field as one of the most knowledgeable experts for Bayesian evolutionary analyses.

**Remco R. Bouckaert** is a computer scientist with a strong background in Bayesian methods. He is the main architect of version 2 of BEAST and has been working on extensions to the BEAST software and other phylogenetics projects in Alexei Drummond's group at the University of Auckland.

# Bayesian Evolutionary Analysis with BEAST

ALEXEI J. DRUMMOND
University of Auckland, New Zealand

REMCO R. BOUCKAERT
University of Auckland, New Zealand

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

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/9781107019652

First published 2015

Printed in the United Kingdom by Bell and Bain Ltd

*A catalogue record for this publication is available from the British Library*

*Library of Congress Cataloguing in Publication data*
Drummond, Alexei J.
Bayesian evolutionary analysis ; with BEAST 2 / Alexei J. Drummond, University of Auckland,
New Zealand, Remco R. Bouckaert, University of Auckland, New Zealand.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-01965-2 (Hardback)
1. Cladistic analysis–Data processing. 2. Bayesian statistical decision theory. I. Bouckaert, Remco R.
II. Title.
QH83.D78 2015
578.01′2–dc23 2014044867

ISBN 978-1-107-01965-2 Hardback

Additional resources for this publication at http://beast2.org/book.html

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.

Cambridge University Press
978-1-107-01965-2 - Bayesian Evolutionary Analysis with BEAST
Alexei J. Drummond and Remco R. Bouckaert
Frontmatter
More information

# Contents

Cambridge University Press
978-1-107-01965-2 - Bayesian Evolutionary Analysis with BEAST
Alexei J. Drummond and Remco R. Bouckaert
Frontmatter
More information

# Preface

This book consists of three parts: theory, practice and programming. The *theory part* covers theoretical background, which you need to get some insight in the various components of a phylogenetic analysis. This includes trees, substitution models, clock models and, of course, the machinery used in Bayesian analysis such as Markov chain Monte Carlo (MCMC) and Bayes factors.

In the *practice part* we start with a hands-on phylogenetic analysis and explain how to set up, run and interpret such an analysis. We examine various choices of prior, where each is appropriate, and how to use software such as BEAUti, FigTree and DensiTree to assist in a BEAST analysis. Special attention is paid to advanced analysis such as sampling from the prior, demographic reconstruction, phylogeography and inferring species trees from multilocus data. Interpreting the results of an analysis requires some care, as explained in the post-processing chapter, which has a section on troubleshooting with tips on detecting and preventing failures in MCMC analysis. A separate chapter is dedicated to visualising phylogenies.

BEAST 2.2 uses XML as a file format to specify various kinds of analysis. In the third part, the XML format and its design philosophy are described. BEAST 2.2 was developed as a platform for creating new Bayesian phylogenetic analysis methods, by a modular mechanism for extending the software. In the *programming part* we describe the software architecture and guide you through developing BEAST 2.2 extensions.

We recommend that everyone reads Part I for background information, especially introductory Chapter 1. Part II and Part III can be read independently. Users of BEAST should find much practical information in Part II, and may want to read about the XML format in Part III. Developers of new methods should read Part III, but will also find useful information about various methods in Part II.

The BEAST software can be downloaded from http://beast2.org and for developers, source code is available from https://github.com/CompEvol/beast2/. There is a lot of practical information available at the BEAST 2 wiki (http://beast2.org), including links to useful software such as Tracer and FigTree, a list of the latest packages and links to tutorials. The wiki is updated frequently. A BEAST users' group is accessible at http://groups.google.com/group/beast-users.

# Acknowledgements

Many people made BEAST what it is today. Andrew Rambaut brought the first version of BEAST to fruition with AJD in the ‘Oxford years’ and has been one of the leaders of development ever since. Marc Suchard arrived on the scene a few years later, precipitating great advances in the software and methods, and continues to have a tremendous impact. All of the members of the core BEAST development team have been critical to the software’s success.

Draft chapters of this book were greatly improved already by feedback from a large number of colleagues, including in alphabetical order, Richard Brown, David Bryant, Rampal Etienne, Alex Gavryushkin, Sasha Gavryushkina, Russell Gray, Simon Greenhill, Denise Kühnert, Tim Vaughan, David Welch, Walter Xie.

Paul O. Lewis created the idea for Figure 1.6. Section 2.3 is derived from work by Joseph Heled. Tanja Stadler co-wrote Chapter 2. Some material for Chapters 7 and 10 is derived from messages on the BEAST mailing list and the FAQ of the BEAST wiki. Section 8.4 is partly derived from ‘A rough guide to SNAPP’ (Bouckaert and Bryant 2012). Walter Xie was helpful in quality assurance of the software, in particular regression testing of BEAST ensuring that the analyses are valid. Parts of Chapter 4 derive from previous published work by AJD and co-authors.

The development of BEAST 2 was supported by four meetings funded by the National Evolutionary Synthesis Center (www.nescent.org). AJD was funded to write this book by a Rutherford Discovery Fellowship from the Royal Society of New Zealand.

CAMBRIDGE

Cambridge University Press
978-1-107-01965-2 - Bayesian Evolutionary Analysis with BEAST
Alexei J. Drummond and Remco R. Bouckaert
Frontmatter
More information

# Summary of most significant capabilities of BEAST 2

| | | |
|---|---|---|
| Analysis | Estimate phylogenies from alignments<br>Estimate dates of most recent common ancestors<br>Estimate gene and species trees<br>Infer population histories<br>Epidemic reconstruction<br>Estimate substitution rates<br>Phylogeography<br>Path sampling<br>Simulation studies | |
| Models | Trees | Gene trees, species trees, structured coalescent, serially sampled trees |
| | Tree-likelihood | Felsenstein, Threaded, BEAGLE<br>Continuous, ancestral reconstruction<br>SNAPP<br>Auto partition |
| | Substitution models | JC96, HKY, TN93, GTR<br>Covarion, stochastic Dollo<br>RB, subst-BMA<br>BLOSUM62, CPREV, Dayhoff, JTT, MTREV, WAG |
| | Frequency models | Fixed, estimated, empirical |
| | Site models | Gamma site model, mixture site model |
| | Tree priors | Coalescent constant, exponential, skyline<br>Birth–death Yule, birth–death sampling skyline<br>Yule with calibration correction<br>Multispecies coalescent |
| | Clock models | Strict, relaxed, random local clock |
| | Prior distributions | Uniform, 1/X, normal, gamma, beta, etc. |
| Tools | BEAUti | GUI for specifying models<br>Support for hierarchical models<br>Flexible partition and parameter linking<br>Read and write models<br>Extensible through templates<br>Manage BEAST packages |
| | BEAST | Run analysis specified by BEAUti |
| | ModelBuilder | GUI for visualising models |
| | LogCombiner | Tool for manipulating log files |

| | | |
|---|---|---|
| | EBSPAnalyser<br>DensiTree<br>TreeAnnotator<br>TreeSetAnalyser<br>SequenceSimulator | Reconstruct population history from EBSP analysis<br>Tool for analysing tree distributions<br>Tool for creating summary trees from tree sets<br>Tool for calculating statistics on tree sets<br>Generate alignments for simulation studies |
| Check pointing | Resuming runs when ESS is not satisfactory<br>Exchange partial states to reduce burn-in | |
| Documentation | Tutorials, Wiki, User forum<br>This book | |
| Package support | Facilitate fast bug fixes and release cycles independent of core release cycle<br>Package development independent of core releases | |