

Preface

The motivation for this book is to teach Master and PhD students how to develop a software like *MrBayes*. During my time as Postdoc, several very bright students asked me how I learned to develop a *RevBayes*. Everyone was looking for a class that teaches the basic skills, but I could never point them to a good resource. Now as a PI/group leader/advisor, I'm writing this book/teaching resource for all the students and researchers who want to get going with developing a software for Bayesian phylogenetics.

When you want to learn how to program, the most important part is to practice. You can read a lot of books or take classes about the theory and concepts of programming. This will not make you a good programmer; writing a lot of code will. Imagine you want to swim a long distance race. Of course you need to learn how to swim first, but then it is a lot of practice to build up your strength and endurance muscles. To improve you time for the race, you might go back to improve your technique, but more often you will have to simply practice more. The same can be said about programming. Having this in mind, I have designed this book to only explain the necessary theory while we come across it. Of course, at the beginning I will explain some necessary basic concepts in the first few chapters. From then on, I focus more on learning by doing.

Although *MrBayes* was originally written as a c program, we will learn here how to use c++. This book is by no means intended to properly teach all the features of c++. There are many other books dedicated solely to the topic of teaching c++ and I advise interested readers to search for one of them. I would also never dare to say that I could write a book about programming in c++ as well as the experts do. This book instead is written specifically with computational and evolutionary biologists in mind as students.

My goal for this book is that readers who will follow the book until the end will have written their own version of *MrBayes* from scratch. The full functionality of *MrBayes* is of course too much to cover within a single book. Nevertheless, I am convinced that the basic principles used within *MrBayes* can be covered and a fully working software for Bayesian phylogenetic inference can be developed within a semester long class. Thus, I call this project and software *BabyBayes* to honor the inspiration taken from *MrBayes* and *RevBayes* but emphasize that it's not a full grown version.

In this book you will (1) learn how to program an academic software using c++; (2) learn about statistical phylogenetics; and (3) learn about and how to develop and perform a Bayesian statistical analysis. Each of these three learning goals will be tightly connected. We will develop a Markov chain Monte Carlo algorithm to perform a Bayesian statistical analysis, although I could have chosen to use Maximum Likelihood instead. To perform the Bayesian statistical analysis, you will need to develop a statistical model, which will be a standard phylogenetic model. Again, other models could have been chosen and hopefully by the end you have your own ideas how to extend or replace the models taught in this book. And finally, to put the two together and actually write a software, I chose to teach c++ here, again, any other programming language could be chosen too. I hope you will see the motivation and big picture behind my choices while we progress through the book.

The whole book is intended to accompany a programming class. Each chapter should be covered within one week and the class should take a total of 12 to 15 weeks. The main concepts and code is explained in each chapter, and at the end of the chapter you will find exercises to work through at home.

Besides the present book, I'm also aware of a new resource developed by Paul Lewis: Paul Lewis' Programming Tutorial. Although my own book might have a similar intention as Paul's, his tutorial reads more advanced than the present book. Thus, I would recommend to the interested readers to work through this book first and then through Paul's to deepen your understanding of programming and statistical methods in phylogenetics.

I'm also aware of efforts by John Huelsenbeck to write a programming book for computational biologists. His book has inspired my own writing. Unfortunately, his book is not available yet. Please stay tuned!