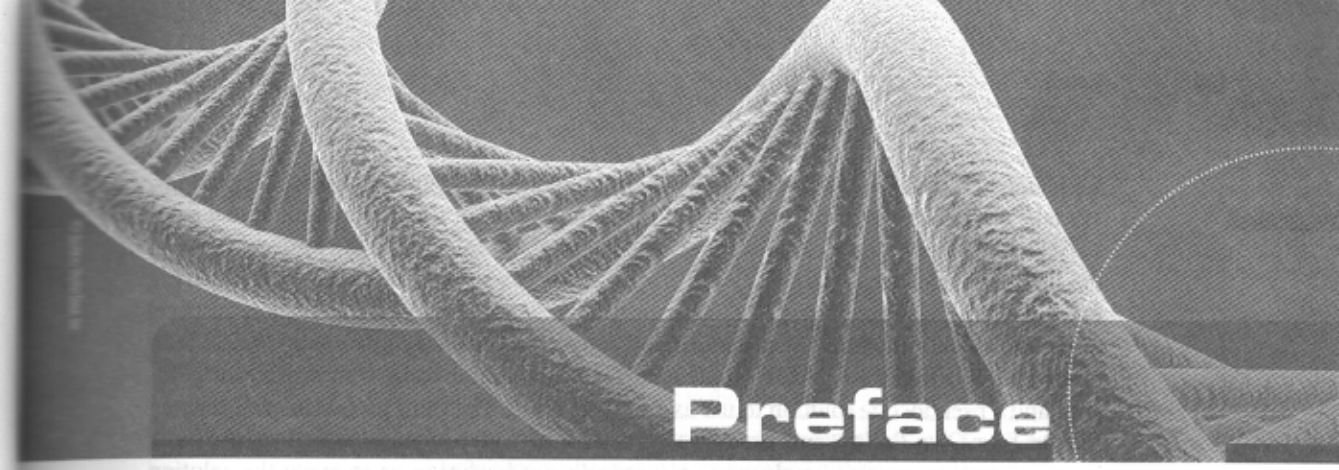

Preface

A Project-Based Approach to Bioinformatics

Why We Wrote this Book

Exploring Bioinformatics: A Project-Based Approach arose from the bioinformatics course that we team-teach at North Central College in the western suburbs of Chicago. Our course, in turn, came from our realization that (1) every working biologist uses bioinformatics in some way to make meaningful use of the mountains of genetic and genomic data available today, and (2) a lot of computer scientists are finding jobs in bioinformatics, doing research on bioinformatic algorithms or collaborating with biologists to solve bioinformatic problems. We began meeting our students' need for bioinformatics in both our fields by incorporating hands-on computational exercises in biology courses at all levels and using biological programming problems in computer-science courses. Eventually, we decided to bring the biologists and the computer scientists together to explore bioinformatics in more depth in a bioinformatics course and later an interdisciplinary bioinformatics minor.

From the beginning, we wanted our course to be inquiry-based, giving students an opportunity to work hands-on with real data. And, we believed that both computer science and biology students would benefit from understanding how bioinformatics tools work: the algorithms that make them tick. Anyone can run a BLAST search (see Chapter 4) or do a pairwise alignment (see Chapter 3), but to understand how changing parameters might produce a more biologically relevant and thus more informative BLAST search or alignment requires knowing something about how the algorithms work. Additionally, we wanted to teach our course on a level appropriate for students who've had one or two programming courses and one or two courses introducing molecular biology, genetics, and/or biochemistry. We had difficulty finding a text fully compatible with these goals, and the "labs" that we wrote for our students first grew to replace the books we were using and then evolved into this book. We think that students elsewhere can also benefit from our approach, and we tried to make this book adaptable to a variety of audiences and to the needs of a variety of instructors.

What's Different About *Exploring Bioinformatics*?

Exploring Bioinformatics doesn't try to be encyclopedic. Each chapter covers a major area of bioinformatics, one which we think is fundamental. We don't try to cover every aspect of that topic, however; we instead focus on a specific biological problem and how bioinformatics can be applied to that problem. We try to include "just enough" biological background to understand the context of the problem, with a brief biological introduction in each chapter

supplemented by a *BioBackground* box to help students with limited biology backgrounds. We take an "under the hood" look at an important algorithm used in computational solutions to the problem so that programmers and non-programmers alike can understand what the bioinformatics tools are doing. We then use web-based bioinformatics tools involving the algorithm (or related ones) in a hands-on *Web Exploration* project that explores aspects of the biological problem. For programming courses, a *Guided Programming Project* then offers an opportunity to actually implement the algorithm in code, with significant guidance and a pseudocode solution to use as a starting point. The third project in each chapter, the *On Your Own Project*, encourages students to work independently to understand (and, in programming courses, implement) a more advanced solution or to apply the solution to a more complex problem. The exercises and questions found throughout the book are intended not merely to test students' knowledge but also to develop it, as we feel that the best learning takes place when students are actively engaged in finding their own solutions.

Exploring Bioinformatics is intended to be flexible. Students can learn how computational algorithms are applied to a particular problem and find research-rich exercises to apply existing web-based bioinformatics tools to real data without having to do any programming. Students in programming courses will find pseudocode and discussions of algorithms that are not language-specific, allowing solutions to be implemented in any desired language. Although exercises and supporting data are provided, instructors could apply the same concepts to any desired dataset if they would like to incorporate their own research or interests into their courses. The text is closely tied to resources for both students and instructors available on a companion website, and instructors can, as desired, download working programs for students to use in lieu of writing their own solutions as well as test data, answers for exercises, and other features. *More to Explore* boxes suggest ways that students or instructors can readily take their inquiry beyond what is in the chapter.

What's New in the Second Edition

After we published our first edition, we started to hear reports of how instructors were using our text and what they wished they could do with it, and we realized that we needed to make the content more flexible in this second edition. To that end, we have removed its dependence on the PERL programming language (though PERL code along with solutions in other languages are still available on the website) in favor of flexible pseudocode solutions. We have also made our discussions of the bioinformatics algorithms more central to the chapters and more accessible to students in courses where no programming is required. We have added new material reflecting changes in how bioinformatics is used, such as next-generation sequencing, metagenomic analysis, and statistical methods like hidden Markov models. In the process, we re-wrote nearly all of the text with better and more applicable biological problems, more authentic data sets, more real-world problem solving with the web-based bioinformatics tools, clearer explanations of algorithms, better figures, more references, and more useful questions and exercises. We would be happy to hear your reactions and your suggestions for further improvement down the road.

How to Use *Exploring Bioinformatics*

The following table lists the main elements of each chapter, their intended use, and the exercises or other assessments that can be completed to reinforce concepts and skills. Each chapter focuses on a major area of bioinformatics and presents a biological problem to which that bioinformatic concept can be applied. Although later chapters do draw on material from

Using *Exploring Bioinformatics: a Project-Based Approach, Second Edition*

Chapter Element	Goal	Assessment
Introduction and Conceptual Development		
Understanding the Problem	Overview the biological problem to be addressed.	BioConcept Questions
BioBackground	As necessary (depending on biology background), learn key biological concepts relevant to the problem.	
Bioinformatics Solutions	Gain a broad understanding of how computational techniques can be applied to the biological problem.	
Understanding the Algorithm	Explore in detail how a computational algorithm relevant to the biological problem works "under the hood."	Test Your Understanding
Chapter Project		
Web Exploration	Apply existing, freely available web-based tools to an aspect of the biological problem, using real-world data.	Web Exploration Questions
Guided Programming Project	Implement an algorithmic solution to an aspect of the biological program in a desired programming language (programming courses only).	Putting Your Skills Into Practice
On Your Own Project	Work independently, using skills developed in the chapter, to further explore the biological problem either through more in-depth programming or more in-depth analysis.	Solving the Problem and/or Programming the Solution
Optional Elements		
Learning Tools	Download an exercise, spreadsheet, or other aid to assist in understanding a biological or computational concept.	
More to Explore	Continue investigating a topic beyond the exercises given in the chapter with additional tools or ideas for exploration.	
Connections	Applications of the chapter's ideas to additional real-world problems or current questions.	
References and Supplemental Reading	Sources for algorithms discussed in the text, background reading on the biological problems, and additional reading to better understand biological ideas or bioinformatics solutions.	

earlier chapters, where time is limited, an instructor could select a group of chapters to cover in a course or omit chapters s/he does not wish to cover.

The initial sections of each chapter lay out the biological and bioinformatic concepts needed to complete that chapter's projects. For students with less background in biology, the *BioBackground* boxes provide "just enough" introduction to the topic. *Understanding the Algorithm* is the heart of the chapter's conceptual foundation, walking students through a key computational algorithm to provide the basis either for using existing bioinformatic tools or implementing the algorithm in their own code. *BioConcept Questions* and *Test Your Understanding* exercises provide a convenient means of assessing student learning; instructors can assign a selection of these questions or devise exercises of their own.

Students will spend most of their time for each chapter working on the *Chapter Project*, which has three parts. In the *Web Exploration*, web-based bioinformatic tools are used to approach the biological problem, with *Web Exploration Questions* providing exercises ranging from simple to more sophisticated to test students' mastery of the programs. For programming courses, the *Guided Programming Project* helps students implement the algorithm in

whatever programming language is used in their course, and *Putting Your Skills Into Practice* requires them to use their programming ability to extend or improve the solution. The *On Your Own* project, which has options for both programming and non-programming courses, gives an opportunity for more independent investigation. Each chapter lists specific learning objectives and offers suggestions and options for effective use of the projects in programming and non-programming courses. Many chapters have additional downloadable *Learning Tools* to help students grasp the problem or algorithm under discussion.

We intend for this book to be usable in a variety of ways to suit the interests and goals of instructors and their students. Instructors can select chapters, project segments, and exercises that suit their courses best, add their own exercises, or use their own data. We hope this book will be an effective tool to help build undergraduates' understanding of this new and essential field at the interface of biology and computer science. We welcome your comments and ideas as you explore bioinformatics.

Resources

For the Student

Hosted on the Navigate Companion Website (go.jblearning.com/Bioinformatics2ecw), students will find the following resources available for download:

- **Web links**: Links to access the tools and websites used in the chapter projects or referenced elsewhere in the text.
- **Learning Tools**: Visual demonstrations, simulations, and hands-on exercises to help students understand important concepts.
- **Python and Perl syntax guides**: A breakdown of key syntax needed for each chapter's programming exercises for both PERL and Python.
- **Sequences and Test Data** used in the chapter projects.

Access to this companion site is included with each new, printed copy of the text.

For the Instructor

The following resources will be made available to all instructors who adopt this text:

- **Keys for chapter exercises**: Answer keys for the Web Exploration, Test Your Understanding, and BioConcept Questions should the instructor choose to assign these questions as homework.
- **Programming solutions**: Complete, executable code corresponding to the Putting Your Skills Into Practice exercises and the On Your Own Project for each chapter will be provided in both PERL and Python; these files can be used as keys or could be downloaded and given to students in non-programming courses for use in completing the exercises.
- **Additional Sequences and Data Files**: Where appropriate, sequences that students will download or generate as part of the chapter's exercises are provided for the instructor's convenience.
- **Updates**: If additional tools are added in the future (e.g., in response to instructor feedback) or if corrections need to be made, updated files will be added to the website.

The instructors' resources are available via secure download; please contact your sales representative for more details.

Acknowledgments

We are grateful for the support of our faculty colleagues and acknowledge the North Central College biology and computer-science students who "test-drove" the exercises and commented on the text. We particularly thank Michael Holler for his invaluable contributions. We have also benefited from feedback from the participants in a 2010 HHMI faculty bioinformatics workshop at Lewis and Clark College who worked through early versions of many exercises used in this edition, contributed to the development of the concepts, and provided valuable suggestions.

Caroline St. Clair
Jonathan E. Visick