
Preface

Bioinformatics is a relatively recent discipline. The term first appeared in scientific papers at the beginning of the 1990s, but this fact can be misleading. Already in the 1960s, when research laboratories were able to afford computers with good graphical performance, the scientific community had started to use them to analyze biological data. From the point of view of informatics and biology, substantial progress has been made since then.

Hardware has become more efficient; the speed and graphical performance of personal computers now are astonishing compared with those of just 10 years ago. The process of developing software is made easier and more straightforward every day. Another important aspect has been the development and spread of the World Wide Web, an instrument that has changed our conception of communications and made a deep impact on the scientific community. Furthermore, biological data have been exponentially accumulating, thanks to new, powerful techniques available in every laboratory.

Bioinformatics has synergistically exploited new technologies, giving rise to a new scientific discipline, with its own history and even some revolutions. We can define bioinformatics as the science that uses the instruments of informatics to analyze biological data in order to formulate hypotheses about life. However, a scientific definition needs to be operational rather than semantic, and the aim of this book is to describe bioinformatics methods and tools in terms of their ability to help us solve biological problems.

Nevertheless, before starting we must address frequently asked questions: "Who is a bioinformatician?" "What is his or her cultural background?" "Should a bioinformatician be proficient in programming or should he or she know biology in detail?" The confusion arises from the fact that we use the word bioinformatician to indicate expert users of the available tools as well as developers of new and more powerful methods. Of course the background in the two cases could and should not be the same.

In the first case, it is important to have a good biological knowledge to be able to understand the results of bioinformatics analyses and, at the same time, to write simple programs. In the second, expertise in statistical methods, algorithms, and programming and a basic biological knowledge are required. In both cases, though, it is essential to understand biological problems and methods and the rational basis of available bioinformatics tools. This is a must if we wish to use them correctly or improve them.

Therefore, the aim of this book is to describe the rationale and the limitations of the methods and tools available to the biological community at large. It is directed to students who want to have an idea about what bioinformatics is before deciding whether it is worth getting deeper into the subject and to those who, having decided

to pursue a career in experimental biology, want to have a grasp of the methods they will undoubtedly need during their research.

The questions that we will address concern ways of storing and (more importantly) retrieving the enormous amount of biological data produced every day (Chapter 1) and the methods to decrypt the information encoded by a genome (Chapter 2), to detect and exploit the evolutionary and functional relationships among biological elements (Chapters 3, 4, and 5), and to predict the three-dimensional structure of a protein (Chapters 6, 7, 8, and 9).

Chapter 10 offers a window to what the future holds, although in such a young and quickly evolving field as bioinformatics, we have learned that it is hard to predict what will come next, even in the very near future. This is proving to be even more difficult than predicting the structure and the function of a biological macromolecule!

The future will challenge us with new methodologies for tackling new and old problems, but some fundamental aspects will not change. We will always need to apply new methods to the same types of biological data, to implement them efficiently, and, most of all, to be aware of the power and limitations of these methods, in order to evaluate the meaningfulness of their results and extract information useful to solving biological problems.

Note: At the end of most chapters is a list of problems. Some of the input data for the problems can be downloaded in electronic format from the publisher's Web site, www.crcpress.com.