

Preface

Many times I would find myself wondering how people choose what to do in their professional lives. After giving years of work to experimental virology, and some more to botany, microbial genetics, and teaching biological sciences, I settled at a research career in computational biology. And all this time, I was interested as much in biology as in the different ways in which people think about their research. How do we decide which problem to study? Why do some scientific questions sound interesting and important, and others do not? And why do different people have different opinions on what is interesting and important?

I noticed that much of what turns out to be interesting and important—in science and elsewhere—happens when two seemingly unrelated things suddenly reveal some sort of similarity. The pleasure of such discovery, of course, is only comparable to the joy of finding a difference between two things that were previously thought to be the same. Thus, I realized that I am interested in similarities and differences, and in patterns and motifs. And if this is what you are after, then computational biology is a good line of work.

As a “local bioinformatics specialist” at my institute, I spend a lot of time talking to the “noncomputational” biologists. My colleagues often tell me that they are more interested in ways to think about science than in actual applications and protocols. Remarks such as “I have read about database search statistics, and I think I understand how this algorithm works—but tell me how you decide which of these weak sequence similarities are more important than the others!” are common. So, it seems that the myriads of bioinformatics texts that are published these days need a reader’s companion, which talks about prejudices, preferences, and priorities.

This is my attempt on such a companion. It is not intended as a comprehensive source on genome comparisons or other issues of computational biology. I wrote mostly about things that are of interest to me: For example, most of this book is concerned with the protein world, and there is almost no discussion of nucleotide sequence analysis. There is also very little mathematics, statistics, or computer science in the book, even though the practice of bioinformatics requires dealing with models, equations, and algorithms. Rather, this book is about scientific ideas that I believe to be the most important in computational biology and in its most accomplished branch, comparative genomics. I am also trying to show that the era of completely sequenced genomes is a truly novel age of biology, and that comparative genomics is the science for this age.

This book would not be possible without collaboration, friendship, and, throughout the years, many conversations with Eugene Koonin and Alexey Kondrashov. Luna Han, my editor, helped me to define where this book should be going and gently persuaded me to stay on track, and the members of the Bioinformatics Center at the Stowers Institute for Medical Research held the fort all the while.

Most important, my family put up with everything. I thank my wife Irina Sorokina—all the good things in my life for so many years are because of you, my love—and my children Alexandra, Nikolai, and Natalia.