

How I Began My Training in Bioinformatics

Manuel J. Torres

University of California at Davis, Davis, California 95616

I have always had a strong and abiding interest in the life sciences and earned an undergraduate degree in Biochemistry from San Francisco State University in 1979. My earliest memories of computers go back to 1980 or so, about the time Apple was starting the personal computer revolution. My passion for bioinformatics developed slowly and as a direct result of these two events. In 1994, I was working at Bayer Biotechnology in the Protein Purification facility in the Berkeley, California, production site. Literally overnight, the company installed a computer network in the laboratories to streamline operations and to monitor production tasks. It was no longer possible to read a pH meter or UV-light spectrophotometer without interfacing with a computer. I asked lots of questions, What is Windows? "It's DOS with a GUI (graphic user interface)," came the answer. What is a GUI? "It's the interface to the operating system," came the answer. What is an operating system? I pleaded. "Manny," came the exasperated response, "go back to school!" So I did. In 1995 I matriculated into the MA Program in Cell and Molecular Biology at San Francisco State University.

One day in Maureen Whalen's Recombinant DNA course, we were using the UC San Diego supercomputer center by telnet in order to do some basic sequence analysis exercises on the software program GCG. I was confronted with a blinking light once I had logged on with my assigned user name and password. I asked more questions. What is the blinking light? "The cursor, you type in UNIX commands at the cursor," Dr. Whalen explained. What is UNIX? I pleaded. Dr. Whalen became very serious, "Manuel, you must read about these things and find out for yourself if that's what interests you!" So that's what I did. I was a full-time graduate student in biology and began studying computer science in my spare time. I found that the computer courses at the local community college were more to my liking

and speed. I read a book on UNIX and took an introductory programming course in Pascal. I learned C++ and began writing and compiling programs first on a Borland PC compiler and then on a UNIX server using GNU software. I wrote a rudimentary string-search program in C++ for a class project. I looked for evidence of isochores and matrix attachment regions distributed along the recently sequenced yeast genome. The human genome would be sequenced in about 5–7 years according to the most recent predictions of that time. I became extremely excited at the thought of writing programs to parse the human genetic code. I learned from writing my first string-search program that I had a very minimal understanding of what a gene was and how to look for intergenic regions on a chromosome. So I decided I needed to learn more genetics. Also, I was having difficulty finding a principal investigator to sponsor a research project in bioinformatics. I applied to the Genetics Graduate Group at UC Davis and matriculated into the PhD program in 1997.

My experience at UC Davis has been excellent. I work with John Yoder in the Department of Vegetable Crops. We are investigating plant-plant interactions and use a functional genomics approach to predict gene functions in a parasitic plant model. John is a very flexible and progressive thinker. We are very much alike in that way and are very compatible. I have continued to learn computer science on my own by taking night classes, attending seminars, conferences, and workshops, in addition to being a full-time graduate student. I have learned everything I can about operating systems and I am comfortable working with almost any computer. I have learned UNIX, including how to write PERL scripts on UNIX machines and most recently on a Wintel PC. The field of bioinformatics is an excellent career choice for young people with an undergraduate background in the life sciences, math,

or computer science. However, the field is multidisciplinary so you will have to work hard to learn widely disparate scientific fields of inquiry. You will need to put on your thinking caps for this new and exciting major!

For young people who think they might be interested in a career in bioinformatics or life-scientists looking for a career change, I recommend learning several key areas of computer science that are essential in bioinformatics. Learn about networks and the Internet. UNIX is the de facto Internet and network standard, learn as much as possible about shell scripts, file transfer protocols, and all Internet protocols. Bioinformatics is something people do across networks. Learn computer programming and data structures. Any language will do and PERL might be a good place to start. PERL is the Swiss army knife of scripting languages. PERL is a powerful

tool for the kind of text file manipulation tasks that every bioinformatician needs in his toolkit. Learn lots of math and gain at least a rudimentary understanding of statistics and probability. Bioinformatics involves modeling of biological systems and a good understanding of mathematics and the theory of probability will be very helpful. Finally, learn something about database design. All of your work will generate data, lots of it. You will need a strong understanding of how to store and retrieve data. Of course it almost goes without saying that you will need a strong life sciences curriculum to round out your skill set. Learning bioinformatics is difficult because there are few formal educational programs and fewer role models. But then, being a pioneer is never easy. Please do not hesitate to email me (mjtorres@ucdavis.edu) if you have any questions.