_____

Getting started in Bioinformatics

Martin Gollery[*]

Center for Bioinformatics, University of Nevada at Reno, Department of Biochemistry / MS330;
E-mail: marty.gollery@gmail.com; Phone: +775 784 7042; * Corresponding author
published online March 24, 2006

Editorial Message:

Several times a month, people that want to get started in the field of Bioinformatics contact me, looking for a job, a student position or a connection to someone else who has a job or postdoc opening. This is to be expected, perhaps. After the dot-com crash a few years ago, people were leaving the Internet and telecom fields in droves to get retrained. The 'Next Big Thing' that was being hyped at the time was Bioinformatics, and there were many new programs starting up that were available to train them. Now they are finished and want to take the next step.

The problem is, that the reality of the Bioinformatics job market does not match the hype of a few years back. There are positions to be had, but also a good deal of competition. How do you separate your resume beyond the usual grade point averages and class rankings? Bioinformatics is a broad field, and beyond the basics there is much less standardization in programs compared to a much more mature field such as Chemistry or physics. The result is that employers frequently don't know what classes you have taken based simply on your degree, and generally do not have the time to investigate more thoroughly. You need a way to convince people that you can really do the job, and solve real-world problems, not merely pass classes.

The solution is to do a useful project (or multiple projects) that people will actually use. If you become the person that is known for developing a certain program, then everyone that uses your program will know your name. Even people that have not yet seen your program will be able to download it and see what you can do. Even a clever Perl script or an improvement to an existing program can be beneficial to your career- after all, people will figure that if you can solve some of their computational problems, the probability is pretty high that you will be able to do it again.

Now, where will you find the perfect project to work on? Professors, Biotech or Pharma scientists, or anyone you meet at conferences, who is working in an area that interests you. Most of these people have much more work than they can actually accomplish. Ask them what you can do to help them out. They will be much more inclined to talk to you when you are offering to volunteer some help than if you are pestering them for work!

Another source for inspiration is the OpenBioinformatics Foundation. They have several ongoing community projects (bioperl.org, biojava.org, etc) that can always use new developers. Pick your favorite language, or perhaps the one that you would most like to learn. The developers on these projects are willing to help the newcomers, and can offer plenty of tips to help get you started.

Another great open source project to get involved with is the European Molecular Biology Open Software Suite, or EMBOSS. This program has recently lost funding, so they can really use the help. EMBOSS consists of some 200 programs, many of which could be optimized, parallelized, extended or improved in some way. A list with dozens of areas requiring software development may be found at http://emboss.sourceforge.net/apps/proposed.html.

If you do find a mentor that has a good project for you, get the information that you need to be sure you are working on the right things, get all the advice that you can, then leave him alone. Don't pester someone that has already helped you- he or she does not owe you a job. You might get a recommendation out of the deal, and that can be very valuable, but do not try to attach yourself to that one person. Looking desperate is never attractive.

In the longer term, you need to build a network of associates that know your work. This is another advantage to working with Bioperl or one of the other OpenBioinformatics projects. Many of the developers on these projects have developed working relationships in the field that may prove to be very useful.

Once you have successfully completed one project, later efforts will be easier. As people begin to use what you have developed, they will begin to contact you, rather than you having to contact them. This is a much better situation to be in! After all, if they are asking you for information, you will feel much more confident in asking for advice or help in return.

For the established Bioinformatics researchers who are reading this, I recommend that you begin keeping a list of those projects that you don't have time for- both small and large. Look for things that will be useful to as many people

as possible. If it only works with your system, then you haven't done anybody a favor.

Secondly, look for projects that can be done by one person. We want to encourage newcomers to the field, not overwhelm them. Adding a component to an existing package, or building connecting code from one package to another, or perhaps developing an interface are all useful skills to be learned and can frequently be accomplished by a single worker.

Finally, don't rely on the final product. Acknowledge that you are handing this task to a beginner that you do not control. Volunteers have to do something else to put food on the table, and this can interfere with completing your project. Set a reasonable deadline, and ask for progress reports. If it is not going to be finished in time, reassign it to someone else on your team. Even if it is completed quickly, the product may not work as you want or expected. While the developer should always strive to deliver the best possible product, your attitude should be that this is basically free work, and any benefit that you may derive from it is a bonus.

The field of Bioinformatics is a rapidly evolving area that is highly useful as a tool for modern biology. It is also a growing field, but it has had some bumps in the past few years, and has not lived up to the hyperbole of the past. Those that are interested in this field need to understand that it is a challenging science with a steeper learning curve than other biosciences due to its' highly multidisciplinary nature. Encouraging top quality scientists to learn Bioinformatics is a worthy goal, and we should all join together to assist in their training.