4273?: a novel platform for bioinformatics education

Gaining an understanding of bioinformatics is now par for the course when it comes to training undergraduate students in biology. However, facilitating a comprehensive course in bioinformatics is often hampered by the constraints of a typical classroom environment, specifically the software and hardware requirements. To grasp more in-depth concepts, beyond what is offered from basic tools available online, Daniel Barker from the University of St Andrews, UK and colleagues propose the re-creation of a bioinformatics research environment. To this end they developed the 4273*?* platform for teaching bioinformatics, on the inexpensive Raspberry Pi computer, modified for the classroom setting and including their Open Access bioinformatics course - as explained in their recent study published in *BMC Bioinformatics*. **Barker and leading bioinformatician Ian Korf from University of California, Davis, USA discuss the importance of bioinformatics training for biologists, and the potential for 4273*?* to improve the way this is implemented.**

Why is teaching bioinformatics so important now?

IK The main reason is that biology is becoming much more quantitative. Sequencing and other high-throughput technologies are revolutionizing many biology-related industries. In the past, people would focus on a few genes, for example, but today it's possible to look at all genes. That's a very powerful paradigm, but you need a different 'microscope' when looking at genomic data, and it takes some training to understand how to handle and analyze large data sets.

Biologists are adaptable people in general, and I find that the typical biologist can become a good bioinformatician. Many would-be bioinformaticians get scared away because of the arcane syntax of the command line, their lack of computer programming experience, or a feeling that their mathematics skills are insufficient. These are just fears. If you hold their hand for a little while they can get through the scary bits, they will emerge on the other side self-empowered and with a new perspective on problem solving. It will impact everything from grocery shopping to genome analysis.

What are the challenges in organizing an undergraduate bioinformatics course at a university?

IK Most universities have plenty of computers, but they aren't typically configured to analyze genomic data. Most bioinformatics takes place in a Linux/Unix/Mac environment yet most labs have Windows boxes. It's easy enough to install Linux on a PC either as a second OS or in a virtual machine, but the University might not want to do that and may have policies against it. So one solution is to have a separate set of Linux machines for bioinformatics. That can get expensive. Once you have the computers, you have to set them up with an appropriate bioinformatics environment too, and that can be complicated by the ever changing nature of the field.

DB At my university, the computer classrooms run Windows, which is one of the least useful operating systems for bioinformatics. A few years ago we used the Windows computers only to log in to a remote Linux server, but this still has limitations and isn't much fun. Students are denied administrator access – so they are more dependent on staff, to install software and so on, than would be likely in a subsequent career.

I have not found costs to be a big problem. Laboratory practical classes in other areas of the curriculum can be expensive, requiring consumables and technical preparation every time they run. For bioinformatics, equipment has to be purchased every few years but the intervening years are much cheaper.

IK One of the greatest obstacles to teaching bioinformatics is the teachers themselves. Bioinformatics is an eclectic field drawing from molecular biology, statistics, computer science, mathematics and other disciplines. Not many teachers have such a diverse education.

DB Another problem has been textbooks. There are plenty, but it has been hard to find anything with the right balance. I was very happy when 'UNIX and Perl to the Rescue' came out (Keith Bradnam and Ian Korf, 2012). It covers the underlying skills well. Still, it is not intended to be a complete bioinformatics textbook.

What is a Raspberry Pi computer and how could it be useful in teaching bioinformatics?

DB The Raspberry Pi is a small, low-cost, general-purpose computer. It was developed by the Raspberry Pi foundation to get schoolchildren interested in programming again.

The Raspberry Pi Model B is like the guts of a mobile phone, plus network and USB sockets. Instead of a hard disk the Raspberry Pi uses an SD card. You use another computer to copy an operating system onto this. After connecting everything to the Pi and powering up, you have something that looks identical to Linux running on any other computer. The hardware also makes it easy to interface with electrical devices like robots, though we have not used this in 4273?.

The Pi has a 700 MHz CPU and (now) 512 GB of RAM. Overall performance seems about the same as a 10-year-old laptop. For most things you use a computer for, this is actually fine. Overall you have a cheap, tiny, portable computer that runs Linux.

IK Raspberry Pi simplifies several problems. You can outfit each student with a computer for almost nothing. The peripherals (mouse, keyboard, monitor) cost more than the computer if you had to buy them. The 4273*?* package simplifies the software problem too, since it has enough bioinformatics software to get you started. It's not going to solve anyone's big data needs, but it's a great practice environment.

What are the benefits and limitations of the Raspberry Pi platform?

DB A major benefit is that it's cheap. From a software point of view, the availability of Raspbian – a standard, reliable GNU/Linux operating system – is a huge advantage.

Some students complain about the generally low speed of the Raspberry Pi, but I don't see this as a problem. Now most things we do on computers are almost instant (email, word processing, even statistical analyses), students are surprised to have to wait an hour or so for results, but this is a useful experience to go through. It's good preparation for research, where analyses can take days or weeks.

Currently, graphics are the main limitation on the Raspberry Pi. This makes the Raspberry Pi a little sluggish and irritating for things like browsing the Web. People are working hard to address this in software.

How is the Raspberry Pi approach to teaching bioinformatics different from other existing approaches?

IK Many bioinformatics courses use a GUI to teach bioinformatics. This could be Galaxy, iPlant, or some other alternative to working with the command line. And you could do this with the Raspberry Pi too, but the philosophy of 4273*?* is to teach students the nuts and bolts of bioinformatics, not just the surface. To do that, you have to be at the command line.

Ultimately, the Raspberry Pi is just a personal computer. By today's standards, it's not very powerful, but it's more powerful than the computers we had in the glory days of the Human Genome Project. The ridiculously low price means that the barriers to bioinformatics education (or any IT education) are very low.

DB We lend a Raspberry Pi to each student on the module, giving them full administrator access. This means they get experience installing software. Each student is personally responsible for their own files, system maintenance, back-ups and so on, as is the case for most post-docs in the life sciences.

If a student were to somehow ‘wreck’ their Raspberry Pi, the SD card can just be wiped and reinstalled, which takes around 20 minutes. If the hardware were damaged somehow, it is cheap to replace. The Pi can easily be taken home, where it can be plugged into most monitors or TVs.

A fairly similar educational experience could be achieved using laptops, but this would be more expensive and less fun.

What unique skills will students gain from taking the Open Access 4273*?* Bioinformatics for Biologists course?

DB They will learn ‘real’ bioinformatics. Where we can, we have based case-studies on research by ourselves and others. We aim to leave students in a good position to discover which bioinformatics tools they require, with the skills to be able to use them – after the usual poking around in online manuals and so on that we all have to do.

This should give the students far more power for any future life sciences career. It is not hard to come up with interesting questions that massive amounts of DNA sequence can help us solve. It is not hard, and increasingly not expensive, to obtain these data – or they may be available to download already. The tricky step, and one at which many life sciences PhD projects are stalled, is how to make effective use of the data.

Do you have plans to expand or modify the course in the future?

DB We have ambitious plans for 4273*?*. At the moment, it is closely based on a final-year undergraduate module at the University of St Andrews. It should be expanded to include material on different bioinformatics subjects and at different levels. We are loyal to the Raspberry Pi but it would be useful to add other platforms as well. Also, 4273*?* should be frequently updated. I would like to see it expanding into new areas and dropping obsolete material as time goes by. In this way it could have a long life, compared to more static forms of teaching material such as textbooks which – when it comes to bioinformatics software – can go out of date too fast.

What advice or tips would you give to people planning to organize a bioinformatics course?

DB Respond to questions from students, but – particularly for assessed coursework – not too quickly. If a student can’t work out that they’ve omitted a semicolon or something, and their analysis is totally failing, they may get frustrated and drop you an email. But in the next hour or two they may well solve the problem themselves, which is a better educational experience than pasting text from your reply. Students should learn that we all go through these kinds of problems on a daily basis. Also, when giving advice, be as fair as possible. If one student asks a question about an assessment, CC your reply to all students on the module.

IK Bioinformatics can be divided into users and developers, but this dichotomy shouldn't be encouraged. Users should know what is going on inside programs, and developers should understand the practicality of their algorithms.

Biology lab classes are often robotic exercises in following recipes. Resist doing this in bioinformatics. People will learn more if you tell them the goal and let them figure out how to get there. There isn't much they can break in a computer, so give them the freedom to explore.

Local computing is more reliable than cloud computing.

The field changes quickly, so course materials must be updated regularly.