

Computer buffs chase a common language

Feb 26 2002

Open-source software is ideal for managing human genome information, but research units are not working together

Science and Education Correspondent

ONE whiteboard, 20 hackers and the mysteries of the human genome are the base ingredients for the world's first ever biohackathon, now under way in Cape Town.

It brings together the world's top bioinformatics programmers one woman and a group of young men (their average age is 26) who make the powerful software that allows scientists to manage the massive amounts of data that accompany research into the human genome.

Unlike traditional licensed software, open-source software is usually free. But more importantly, it comes complete with its programming code, enabling the user to customise the programmes or remove bugs.

It can be developed much faster than licensed software and is particularly well suited to the fast pace of developments in the life sciences, including research into the human genome.

Around the world, different research groups have given rise to their own programming communities, operating in different "open source" programming languages such as Java, Perl and Python.

The big problem facing the bioinformatics community is that because they have worked out their own solutions for their own projects, their programmes do not necessarily "talk" to each other, says Tania Broveak Hide, MD of local firm Electric Genetics.

And that is why she has gone into partnership with US publisher O'Reilly & Associates, and brought the open-source bioinformatics community together for the first time to "crack code".

They call themselves hackers, but there is nothing underhand or illegal in what they do. They just have a few things in common with hackers their youth, a passion for computers, and since most of them are students, they tend to punch out their programmes long after midnight.

The group first met in Tucson, Arizona, two weeks ago, and are now in Cape Town for the second part of their ambitious project.

They sit classroom fashion in front of their laptops, with their tasks laid out on the whiteboard.

There is no leader and tasks are identified and allocated by consensus.

You might expect egos to collide in such an anarchic set up, but it is quite the opposite, director of the Open Bioinformatics Foundation in the US, Chris Dagdigan says. "It's quite simple. Whoever writes the code wins the argument."

Some hackers are working on the "libraries" of programmes that enable people to build their own applications, hoping to make the different pieces compatible, even though they are written in different programming languages.

The hackers also hope to make data exchange more efficient and improve the visualisation tools used to look at genome sequences.

"At the moment, genome data is posted at centres all over the world, rather like houses that are managed by different estate agents. We want to create a system that will enable you to make a single query and search for data from all over the world."

Since the mapping of the human genome two years ago, scientists have a pretty good idea of the sequence of genetic material, or DNA, it contains says Broveak Hide.

Scientists know that the sequence of DNA spells out the exact instructions for the creation and destruction of each and every cell in our bodies. In other words, the human genome is the road map for our genes the units that determine what we look like and, when they go wrong, which diseases we will succumb to.

Although scientist have identified genes associated with more than 30 disorders, their understanding is limited, says Ewan Birney of the European Bioinformatics Institute. Trying to understand which bits of the human genome do what is a bit like trying to translate Russian with an incomplete phrase book.

Since understanding the mechanisms of disease offers hope of prevention and cure, the human genome is the subject of intense research worldwide, particularly by the big pharmaceutical companies.

"The amount of data on the human genome in the public domain has grown exponentially in the past few years. Managing that data is probably the single biggest challenge facing computational biology today," says Dagdigian,

"As the body of knowledge expands, scientists can increasingly do experiments on computers instead of in laboratories, giving them tremendous efficiency gains," says Birney.

"Competitive advantage comes from what you do with these programming tools, not the complex algorithms within the tools themselves," explains Dagdigian.

As a result open source programmers traditionally share information, he says, a philosophy that marries well with the way science has worked for centuries.

"If you publish your research in peer review journals, you have to tell the scientific community the steps you took to get your results. That concept of sharing materials and methods fits with the open source bioinformatics approach," says Dagdigian.

With its hope of delivering new and freely available software to the scientific community by the end of this week, the biohackathon might sound like a socialist's dream.

But do not be misled, says Birney. Just as an athlete's stamina, not her shoes, will help her win a marathon, this software is merely the enabling tool for scientists and pharmaceutical companies in their high-risk, high-return race to find drugs and vaccines.

It is up to them what they do with it.