Bioinformatics in China

Clare Sansom (Birkbeck College, London)

The Cyberbiochemist theme for this special issue of *The Biochemist* devoted to China hardly needs stating: an overview, such as is possible in the space available, of bioinformatics – and genomics – research in the country that is the most populous on earth and that boasts the fastest growing economy of any major nation.

Had such a survey been carried out early in the bioinformatics era in, say, the early 1990s, it would have had to be a short one. In a perspective article on Chinese bioinformatics published in *PLoS Computational Biology* in 2008[1], Liping Wei of Peking University in Beijing and Jun Yu of the Beijing Institute of Genomics comment that the field had a "delayed and patchy start" in their country. The first bioinformatics research centre in the country was set up in Beijing as late as 1996. Wei and Yu note, however, that the field has recently enjoyed steady growth, buoyed by general increases in government funding for research and large numbers of well-trained professionals returning from overseas to take up positions in their home country. They publish citation statistics that, although fairly crude (and dating only to 2006), back up this point: in the first years of this century, the number of bioinformatics publications in PubMed from China has grown faster than the number of such publications worldwide. I estimated today, very roughly, that the proportion of all papers in PubMed with at least one author with a Chinese affiliation has risen from about 1.6% in 2000 to about 7.3% so far in 2011. This is still fairly low compared with its estimated 19.3% of world population, but China is classified by the World Bank as a middle-income country with its per capita GDP rated as 94th in the world, and vast swathes of its population, mainly in rural areas, still exist on less than US$2 per day.

As one of the most powerful emerging economies, and with a government that takes a very progressive line on science investment, China is well placed to set its own agenda for bioscience research. Several of the world's leading centres for high-throughput 'next-generation' genome sequencing are located in China and it is not surprising that genome databases for both the Chinese staple crop, rice, and the iconic giant panda are based there. It may be rather more surprising that China is becoming known as a centre for high-throughput human genome sequencing and analysis of human genomic variation. One indicator of how far China has come in the bioinformatics field is the establishment by the BGI (formerly the Beijing Genomics Institute) of a bespoke cloud computing network to handle the dizzyingly vast amount of data generated by its 150 or so next-generation sequencers and potentially to offer a service to genomics researchers worldwide[2].

Rice (*Oryza sativa*) is both one of the world's most important cereal crops and a model organism for the study of other cereal genomes. The RiceMap genome browser, located at the Center for Bioinformatics in Peking University, holds genomic data for the two cultivated rice subspecies: *japonica* and *indica*. The genomes are fully annotated, the source code is freely available and useful videos are available to introduce the resource to new users. A new paper describing this database[3] is one of the most highly cited papers from 2011 in *BMC Genomics*.

The taxonomic classification of the giant panda (*Ailuropoda melanoleuca*, or 'black-and-white cat-foot') was under dispute for many years, and it was only molecular genetic analysis that enabled it to be firmly classified in the family Ursidae, as a true bear. Its genome was sequenced in 2008 by the Beijing Genomics Institute and the data are held on a server there. This has many

Seven-month old panda cub in the Wolong Nature Reserve in Sichuan, China (Sheila Lau, Wiki)

characteristics in common with the genomes of other mammals, including ourselves: pandas have 20 pairs of autosomes, two fewer than humans, and a rather large mammalian gene set of about 31 000 genes. There are fewer than 3000 pandas left in the wild, and the more we can learn about the biology of this unique animal, the better placed we will be to help it survive.

In 2003, China and other East Asian countries suffered from an epidemic of a new respiratory disease that became known as SARS (severe acute respiratory syndrome). Once the cause of this disease had been established as a coronavirus, a consortium including many leading Chinese scientists was set up to sequence the genomes of different viral strains and analyse its variability. This has provided some important insights into both the large variations observed in patient susceptibility and possible treatments for the disease. Very recently, when infection by a novel strain of *Escherichia coli* broke out in Germany, eventually killing over 50 people, a Beijing-based consortium sequenced its genome and revealed the unusual features that made it so deadly[4].

And you may, just possibly, have used Chinese-made bioinformatics programs without realizing it. Many groups have released free Web-based interfaces to bioinformatics tools that make them simpler for non-specialists to use. WebLab, from the Center for Bioinformatics, is one of these, and it is straightforward and reliable. It provides interfaces to many commonly used tools: CLUSTAL for multiple sequence alignment, GLIMMER for identifying genes in microbial genome sequences, and the entire EMBOSS package. I use it with students very often, and have never had any complaints. ■

Selected online resources

www.cbi.pku.edu.cn/ (Center for Bioinformatics, Peking University)
www.ricemap.org/ (RiceMap, rice genome browser)
www.genomics.cn/en/index.php (Beijing Genomics Institute)
http://panda.genomics.org.cn/page/panda/index.jsp (BGI's panda genome database)
http://weblab.cbi.pku.edu.cn/index.jsp (WebLab bioinformatics tools from CBI)

References

1. Wei, L. and Yu, J. (2008) PLoS Comput. Biol. **4**, e1000020
2. Callaway, E. (2011) Nature **475**, 435–437
3. Wang, J., Kong, L., Zhao, S. et al. (2011) BMC Genomics **12**, 165
4. Rohde, H., Qin, J., Cui, Y. et al. (2011) N. Engl. J. Med., doi:10.1056/NEJMoa1107643

Best of the Web

The writing on the wall

Mark Burgess (Executive Editor)

In a small alley, called 'Colonnade' off Herbrand Street, London WC1, is a faded, painted sign reading 'The Horse Hospital'. In Barbon Close, a cul-de-sac off Great Ormond Street, is a sign for 'C. Bailey & Sons. Horse and Motor Contractors'. These are two examples of ghost

Picture by Katy Stoddard

signs; painted announcements of businesses long departed. I see these two on the way to work every day. London has quite a few such signs, but fewer than other conurbations because of its constant programme of refurbishment and rebuilding.

However, it was in London that the idea of documenting theses signs began. Sam Roberts became interested in the ghost ads of Stoke Newington as a personal hobby. It was later that he found that his interest was shared with a number of urban historians in London and elsewhere in the UK.

Of course a website, a blog and a Twitter account swiftly followed. The History of Advertising Trust became interested too, and managed to secure some money from Rank Hovis to set up a permanent archive. The original website is www.ghostsigns. co.uk and this is the home of the Ghostsigns Project, where you can find out more about hand-painted wall advertising and how to contribute your own photographs. There is a Google map, a Flickr map and a reading list that includes books from America and France, as well as a guide for those wishing to become a signwriter. Is also has links to articles and to The History of Advertising Trust Ghostsigns Archive (www.hatads.org. uk), which went online last year.

The archive of over 600 images can be browsed through categories such as Alcohol and Tobacco, Food and Drink, Medical and Health, Tradesmen and DIY, and Animals and Agriculture. The archive has a search function, but it's a bit quirky. Altogether, an excellent web resource for a neglected chapter of social history. ■