



THE SEQUENCERS

Rapid technological developments have spurred big changes in the requisite genome-sequencing jobs. **Kelly Rae Chi** assesses the sequencing assembly line.

Omar Jabado spent his first few years of graduate school at the bench, working on molecular assays to identify gene-expression patterns in viruses. As the data piled up, he needed a method for quick analysis, so he taught himself how to program and created his own software. In 2007, the lab acquired a next-generation sequencing machine and he began identifying viruses from clinical samples. By the end of his studies, he was a full-time bioinformatician, helping to develop programs that target viral genomes. So when he was offered the opportunity to lead a genome sequencing unit at Mount Sinai Medical Center in New York last year, it was an easy decision. "This field is growing fast, so I wanted to be a part of it," he says.

Jabado's experience shows how quickly the sequencing arena is changing; those eager to learn the latest techniques, even if not part of their chosen field, stand a good chance of excelling. The drive to sequence more plant and animal genomes, and to do so faster and more comprehensively, has brought growth to genome sequencing and its ancillary professions. Jabado had entered a field that at every level — from sample preparation to data analysis — remains rich with opportunities, even nine years after the completion of the human genome. Fast-paced and competitive, sequencing has become an integral part of

industries from agriculture to medicine. A lot has changed since the milestone achievement of the draft sequence of the human genome, both in terms of the technology and the jobs associated with its use.

Changing fast

Higher speeds and lower costs have revolutionized the field. Next-generation sequencing instruments introduced in the past five years make it possible to gather data thousands of times faster than was possible with the older Sanger method, which was used for some 30 years. The first wave of next-generation sequencers, which use pyrosequencing, generated data equivalent to those of more than 50 Applied Biosystems capillary sequencers, which use Sanger chemistry, at less than 20% of the cost (see *Nature Methods* 5, 16–18; 2008). Faster, more cost-effective sequencing means many more projects, so many more researchers are entering the field.

But the job of a 'sequencer' is changing fast within core facilities — the university hubs dedicated to sequencing — as well as in individual labs and companies.

"The actual sequencing process is becoming more automated and technical,"

says Julian Parkhill, head of pathogen genomics and sequencing at the Wellcome Trust Sanger Institute in Cambridge, UK, which has the largest sequencing centre in Europe, boasting around 600 gigabytes of genomic data per week. But the flood of data requires software tools and skilled computational biologists to make sense of it.

"Many of the labs getting into next-gen sequencing are thinking, 'Oh, if we just had our own programmer, we would be analysing faster and better,'" says Jabado.

The main growth areas are bioinformatic analysis of large-scale data and applying sequence-based technologies to specific biological questions, says Parkhill — for example, matching therapies to tumours of a particular genotype to gauge their effectiveness.

Last summer, the Intramural Sequencing Center (NISC) run by the US National Institutes of Health discarded six of its 14 older Sanger sequencers, and replaced them with nine much faster Illumina Genome Analyzers and one Roche 454 sequencing instrument. The change worried workers on the production line, recalls Robert Blakesley, director of the NISC's sequencing group. Some were concerned that they would be made redundant as the machines conducted

"Flexibility in the workflow is really important. It takes the routineness out of the day."

sequencing reactions and detection steps automatically.

But, if anything, the new machines mean they can rest easy on that score. “Now it just means we have increased capability,” says Blakesley. As a result, the demand has risen substantially. “Since other investigators across campus heard we have this capability, people we’ve never heard of before have started to knock on our door,” he says, adding that they now have more projects than they can handle.

Blakesley, though, has had to reconfigure the production line and retrain his team to adjust to the latest technologies. For the past 10 years, sequencers at the NISC have been split into four distinct groups. One group handles the first preparation step: making cloned sequence libraries. A second and third group purify subclones and DNA, and load the samples onto pipetting robots that conduct the sequencing reactions, and later onto the sequencing instruments. The last group, called the sequence finishers, checks the quality of data from each run, identifying gaps and determining ways to generate replacement data.

The NISC is phasing out the second and third groups and moving them into the first group — training technicians to prepare samples for next-generation instruments. Sample preparers need to know the nuances of four to six different protocols rather than one or two, because the technology allows investigators to sequence RNA and DNA in several ways. “Right now, we’re subdividing that group into smaller groups that do one or two protocols initially,” says Blakesley.

Production-line work tends to be an entry-level job, usually including those with bachelor’s degrees. For some, the wet-lab experience improves their chances of getting into graduate or medical school. Others move up the sequencing chain into project-management or data-analysis roles. These positions tend to attract more PhD-level scientists.

In many facilities, production-line technicians also have a role in research and development, which for sequencing means optimizing protocols and identifying problems, such as trouble with a particular sequencing chemical. At the Broad Institute in Cambridge, Massachusetts, entry-level sequencers are formally trained for 6 months before they start working in the line. They then typically alternate between 6–8-week sequencing projects and 2–4-week ‘sabbaticals’, during which they can improve or troubleshoot what they were working on, or even design experiments to test the limits of the chemical reactions, says Robert Nicol, director of sequence operations and technology development at the Broad Institute.

The changing technologies and constant upgrades to existing machines mean that the research and development of instruments

and protocols will be important for years to come. For example, Blakesley’s group has just received one of Illumina’s new platforms, the HiSeq 2000, which has a data-production rate four times faster than their older models. In the three months before it is used on real samples, users will need to learn how to operate the software and load and maintain the instrument. They will run practice samples using known sequences, and will compare results with those from older machines.

But big leaps in technology could require intensive, highly specialized training, says Elaine Mardis, co-director of the Genome Center at Washington University in St Louis, Missouri. Some instruments involve highly specific sample-preparation methods, buffers and reagents, and use different software tools, requiring trained sets of individuals to operate each one.

Although many of the roles within sequencing centres are compartmentalized, biotech start-ups and academic labs working on sequencing-technology development offer a chance for employees to take on different positions. LaserGen, a company founded in 2002 and based in Houston, Texas, is working to develop sequencing chemistry that will ultimately allow genomes to be sequenced for US\$1,000, using less-expensive reagents and faster chemical reactions. The backgrounds of the company’s 10 employees run the gamut from organic chemistry to engineering and informatics. “It’s really hard to find individuals who are good at many things,” says chief executive Michael Metzker, who is also a molecular geneticist at the Human Genome Sequencing Center at Baylor College of Medicine in Texas. “We may bring people in with certain skills, but they will jump into other areas and get involved quickly.” There is no typical

day at LaserGen, says Metzker. Chemists make the tweaks to sequencing chemistry, biologists test and work with the reagents in sequencing applications, engineers do the data acquisition, and informaticians perform

the analysis. “Flexibility in the workflow is really important,” says Metzker. “That takes the routineness out of the day.”

As the ones who spearhead the analysis, bioinformaticians are the fastest-growing group of sequencer-related scientists in academia and industry. This role has changed, too. About 15 years ago, bioinformaticians in the sequencing field might have been responsible for anything from IT support to processing the data to annotating genomes and determining what the data

mean. “Sometimes it was the same person who did all of those things,” says Chad Nusbaum, co-director of the sequencing and analysis programme at the Broad Institute. Now, specialization rules.

These layered roles mean that bioinformaticians working with data in a sequencing centre have diverse backgrounds, such as software engineering, database administration, mathematics or even expertise in specific programming languages. Training in biology is not necessary for every role, and it can be picked up on the job. Either way, on-the-job training will continue as a wider range of sequencing applications, from ‘ChIP-seq’ to exome sequencing, means that bioinformaticians will have to understand many different types of data, says Jim Mullikin, acting director of the NISC. Sequencing centres, and even individual labs, will be hiring data-savvy experts for some time to come. “It’s almost that every lab now”, Mullikin says, “needs to have a bioinformatician in their team.”

Kelly Rae Chi is a freelance journalist based in Cary, North Carolina.



Robert Blakesley of the NISC says demand for sequencing has risen considerably.



Staff operating next-generation sequencing machines at the Wellcome Trust Sanger Institute.