SPOTLIGHT ON GENETICS

# Speed-reading the genome

Cheaper methods of sequencing are opening up doors for new research and new career paths.

BY PAUL SMAGLIK

*This year, Evan Eichler's lab completed a gorilla sequence for about $70,000. "That, to me, is a big deal," he says.*

**OVER THE** last ten years, a set of powerful tools has been reshaping the playing field of the modern genetics researcher. These three — high-throughput, long-read, and portable sequencing — are changing the kinds of questions scientists can ask, expanding the application of genomics to other fields such as energy and climate science, and are even creating new career paths.

High-throughput sequencing machines, like ones developed by San Diego-based Illumina, provide the means to read the genomes of hundreds, or even thousands, of organisms in a relatively short time. This capability has created several new subdisciplines, such as metabonomics, the study of how gut bacteria genes work with the body's own.

Long-read sequencing technologies, like those developed by Pacific Biosciences (PacBio) of Menlo Park, California, can't compete with that speed, but they make up for it in resolution. Records of genomes sequenced with these technologies have fewer data gaps. Therefore, comparing them gives a better picture of genetic variance, which is important in understanding gene function and disease progression.

Meanwhile, portable machines, like those developed by Oxford Nanopore Technologies in the UK, can sequence small samples of DNA rapidly, making them great for field work. They are being used to characterize mutations in viruses, locate potential bioweapons, and sample microorganisms at sea, among other applications.

Whatever technology a researcher embraces, it is a far cry from the early days of the human genome project, when few centres had access to large, expensive sequencing machines that were relatively slow by today's standards. As a result, the role of those large centres is shifting (see **whatever happened to all those sequencing centres?**) Back then, scientists wanting abundant, high-quality genomic information either had to partner with a centre, or pool their institution's resources to build a shared core facility.

Now, with size and cost coming down and speed increasing, sequencing and genome assembly is, if not egalitarian, certainly more democratic, says Evan Eichler, a geneticist at the University of Washington, Seattle. For example, it took more than 50 people, around a dozen centres, $50 million and half a decade to generate a draft chimpanzee genome, published in 2005. This year, Eichler's lab completed a gorilla sequence for about $70,000. "That, to me, is a big deal," he says.

Also a big deal, says Eichler, is the quality of their sequences. An earlier version of a gorilla genome was published in 2012 but that was done with shorter pieces of DNA, and therefore left hundreds of thousands of gaps. His team used long-read technology, closed 90 percent of those gaps, and was able to complete many genes that were only partially sequenced in the first attempt.

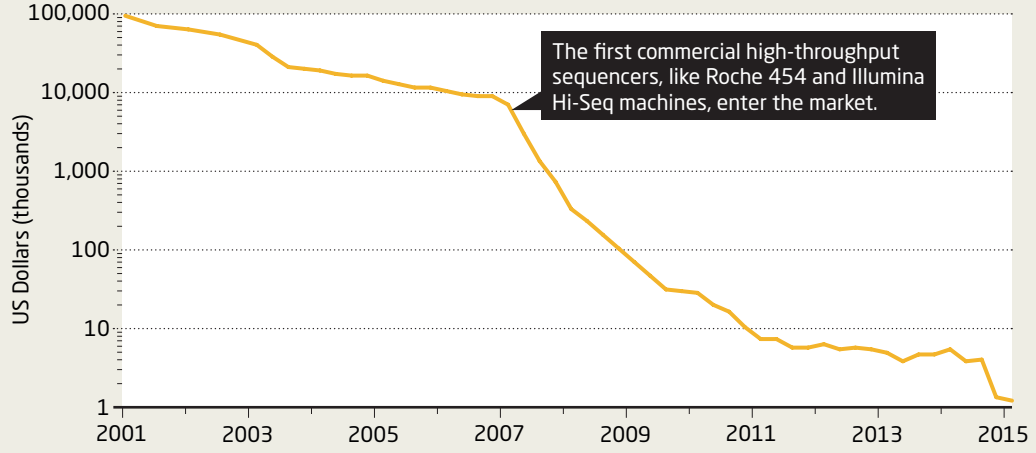
The cost, speed, and accuracy now available can provide a valuable career foundation for any young scientist studying a single organism in great detail. He tells his students to consider developing a high-quality "reference" genome of their own, to use as a basis for the rest of their careers. "If you really care about the genome and you're going to invest 20 years of your life into it and you don't have a good reference, you're missing a lot of functional information important to the biology of the organism," Eichler says.

Having more reference-quality genomes allows comparisons, which helps scientists understand genetic variability. For Erin Price, a researcher at Menzies School of Health Research in Darwin, Australia, such information was essential in understanding antibiotic resistance mechanisms in the bacterium *B. pseudomallei*, which causes melioidosis, a severe infectious disease that is predicted to affect about 165,000 people annually in tropical and subtropical regions a year, killing about half.

There are sequences of the bacterium out there, but that doesn't necessarily tell you about the mutations, says Price. "Bacteria evolve and they evolve very quickly." Her group was looking

## COST OF SEQUENCING

With rapid advances in gene-sequencing technology, the cost to determine a full genome sequence has dropped dramatically in the last 15 years.



at three "very scary isolates" that proved particularly difficult to treat. The bacteria didn't respond to a critical drug in all three cases, causing two deaths.

Her lab wanted to identify why these three strains were particularly resistant, so they needed a complete sequence, without gaps. Price applied for a grant through a PacBio-supported contest and won. The company sequenced and assembled all three genomes; and the Menzies group compared them to reference genomes of other strains, which allowed them to identify the mechanism that had rendered the antibiotics ineffective.

Price says the emerging tools helped change her career from microbiologist to applied geneticist. Her group's work demonstrates how several different techniques can attack the same problem from different angles. The long-read approach helped them identify mechanisms to help the virus fight treatment, short-read methods will allow them to sequence, then compare, thousands of the bacteria, and, eventually, portable technologies could aid public health officials track down the bacteria in water and soil, ultimately preventing further infections.

Portability isn't necessary for genomics to move into the clinic, though. Although the day when every patient has a copy of their own genome sequence hasn't yet arrived, the clinical centre at the US National Institutes of Health (NIH) in Bethesda, Maryland, comes close. Here, every patient with a particular drug-resistant infection has that pathogen sequenced.

The purpose is two-fold, says Julie Segre, head of the NIH's microbial genome section. First, the staff needs to understand what kind of strain the clinicians are dealing with — just as Price's group required. Second, they need to detect, then sequence, plasmids — circular strands of DNA that can replicate independently — because plasmids can "jump" to other organisms and, in some cases, confer antibiotic resistance.

Segre went the opposite way to Price; she began her career in genomics, and then moved into microbiology. She says that microbiology can provide a good bridge for clinicians to learn about genomics. Infection control is a growing concern at hospitals and clinics, so institutions that have the capacity to identify pathogens as well as mutations and plasmids are ahead in that infection control battle.

Her group is providing infection control defence by conducting long-read sequencing of drug-resistant bacteria and creating a database, in collaboration with the US Centers for Disease Control and Prevention and the Food and Drug Administration. That way, when clinicians at other hospitals come across a patient with a drug-resistant bacterium, they can have the sample sequenced with faster short-read technology, and then compare it to a reference genome in the database to help guide treatment.

Outside of infection control, sequencing is moving directly into discovery science. Federico Lauro, an environmental scientist at Nanyang Technical University (NTU) in Singapore, aims to contribute to mapping the microbial biodiversity in the ocean. The health of this marine microbiome serves as a maritime canary in the coalmine, says Lauro.

But collecting the samples isn't easy, and the associated costs are a challenge. The ocean is vast and its inhabitants varied, so there is a need to collect and sequence microorganisms from as many parts as possible. Lauro has been doing so by himself, funded by a nonprofit organization he created — often bringing samples back to sequence in his NTU lab. But this is not always ideal; sample quality can deteriorate en route, and sometimes Lauro doesn't know if data collection has been adequate until he starts sequencing information. This can be especially crucial in protected areas, where researcher access is restricted, or where it can be costly to return, he says.

Portable sequencing will soon make it possible for him to check his results and email them back to the lab, rather than transport samples and hope for the best. It would also boost his vision of citizen science on the seas, where sailors voluntarily collect samples, sequence them, and share the data online. "If we had a way for them to sequence the samples at sea and send us the data, that would be stellar," says Lauro.

Another group is demonstrating that mobile sampling and sequencing can work — and could be life-saving. Nuno Rodrigues Faria, a researcher at the University of Oxford, spent two weeks in June travelling Brazil in a mobile sequencing center tracking the Zika virus.

The sequencing equipment helped the travelling scientists understand how the virus varied by region. "The portability was key," says Faria. Their ability to get samples, then data, from many parts of Brazil helped Faria and his team understand how pathogens spread and how their genomes vary over time.

He envisions using similar technology to monitor and track other tropical diseases that often infect people in areas far from the lab, like the dengue and chikungunya viruses. More distributed mobile sequencing and related equipment could help in detection and surveillance of these diseases, he says.

Portable technology may well create more applications. The UK military is developing a system to look for potential bioterror pathogens, then analyze them quickly, according to Claire Lonsdale, of the UK's Defence Science and Technology Laboratory. She and others are working with companies so that soldiers can sample and sequence air particles with one hand, while wearing a biohazard suit. "The user wants 'Red/Green. Live/Die. Respirator/No respirator. Evacuate/Don't Evacuate'," Londsdale said in a recent talk during the London Calling sequencing conference, hosted by Oxford Nanopore. Existing technology can now answer those questions within 20 minutes, but the goal is for near-instantaneous results, now within the realm of possibility as sequencing science improves.

The spread of sequencing out of the lab and into other fields, as well as applications unheard of a decade ago, means that genomics is becoming as much a tool as a field of its own. Scientists who take advantage of the appropriate sequencing technologies in their research may find themselves in the vanguard of their discipline, rather than playing catch-up.

*This content was commissioned and edited by the Naturejobs editor*

## Whatever happened to all those sequencing centres?

As the speed and throughput of genomic sequencing has increased, so too has the role of large sequencing centres. Places like the Sanger Institute in Hinxton, UK and the US Department of Energy's (DOE) Joint Genome Institute (JGI) in Walnut Creek, California, generated the lion's share of data in the Human Genome Project and many projects that followed. Now since a single lab can sequence an organism with a simple genome themselves, such projects hardly warrant the attention of large sequencing centres, says Tanja Woyke, microbial genomics leader at the JGI.

"The big centres are shifting their focus to big science," says Woyke. For instance, the Sanger is leading an effort to sequence the genomes of 100,000 people (see **http://go.nature.com/2dmkLDB**). At the JGI, which provides sequencing, DNA synthesis, metabolomics and bioinformatics support for the DOE user community outside the institute, an approved project can include up to 1,000 bacterial genomes and encompasses functions like transcriptomics or metabolomics to address energy and environmental challenges.

This shift has changed the kinds of projects the institute takes on, with a focus on function rather than mere sequencing, Woyke says. "For decades we've been generating a lot of sequence, but we don't know what a lot of the sequence does, even in *E. coli*, which is our model organism." Woyke estimates that about 30-40 percent of *E. coli* gene functions remain unknown.

The renewed focus on function is also changing the skills the JGI looks for both internally and externally. Bioinformatics and sequencing skills are still valued. But to understand gene function, the institute is drawing on people with wet lab experience. "We're definitely looking for strong experimentalists who can do more with the data other than just mine it informatically," Woyke says.

WELLCOME GENOME CAMPUS

# ADVANCED COURSES AND SCIENTIFIC CONFERENCES 2016-2017

## CONFERENCES

**Computational RNA Biology**
17–19 October

**Epigenomics of Common Diseases**
1–4 November

**EMBL–Wellcome Genome Campus Conference: Target Validation using Genomics and Informatics**
4-6 December (Heidelberg, Germany)

**Immunogenomics of Disease: Accelerating to Patient Benefit**
6–8 February NEW

**The Challenge of Chronic Pain**
1–3 March

**Innate Immune Memory**
14–16 March

**Genomics of Rare Disease**
5–7 April

**The Biology of Regenerative Medicine**
25–27 April

**Applied Bioinformatics and Public Health Microbiology**
17–19 May

**Healthy Aging: From Molecules to Organisms**
24–26 May

**Exploring Human Host-Microbiome Interactions in Health and Disease**
13–15 September

## COURSES

**Molecular Pathology and Diagnosis of Cancer**
9–14 October

**Genomics for Dermatology**
12–14 October NEW

**Next Generation Sequencing Bioinformatics**
23–29 October

**Chromatin Structure and Function**
31 October–9 November

**Proteomics Bioinformatics**
4–9 December

**Translating and Commercialising Genomic Research**
7–9 December NEW

**Derivation and Culture of Human Induced Pluripotent Stem Cells (hiPSCs)**
12–15 December

**Molecular Neurodegeneration**
9–14 January

**Genomics and Clinical Microbiology**
22–27 January

**Genomic Medicine for Clinicians**
25–27 January

**Genomic Practice for Genetic Counsellors**
7–9 February

**Mathematical Models for Infectious Disease Dynamics**
13–24 February

**Immunophenotyping: Generation and Analysis of Immunological Datasets**
19–25 February

**Genetic Engineering of Mammalian Stem Cells**
12–24 March

**Next Generation Sequencing**
21–28 April

**Computational Molecular Evolution**
8–19 May

**Fungal Pathogen Genomics**
11–16 May NEW

**Summer School in Bioinformatics**
26–30 June

**In Silico Systems Biology: Dynamic Modelling of Biological Networks**
9–14 July

**Genetic Analysis of Mendelian and Complex Disorders**
19–25 July

## OVERSEAS COURSES

**Working with Parasite Database Resources**
16–21 October
(Montevideo, Uruguay)

**Genomics and Molecular Epidemiology of Bacterial Pathogens**
11–16 December NEW
(Ho Chi Minh City, Vietnam)

**Malaria Experimental Genetics**
30 April–6 May
(Accra, Ghana)

**Genomics and Epidemiological Surveillance of Bacterial Pathogens**
9–14 July
(San Jose, Costa Rica)

Wellcome Genome Campus
Hinxton, Cambridge, UK
wellcomegenomecampus.org/coursesandconferences

U282463EL

EMBL

*The European Molecular Biology Laboratory offers a highly collaborative, uniquely international culture. It fosters top quality, interdisciplinary research by promoting a vibrant environment consisting of young independent research groups with access to outstanding graduate students and postdoctoral fellows. Significant core funding and limited teaching responsibilities allow EMBL group leaders to undertake ambitious, farsighted research programmes and provide a scientific environment that is particularly attractive for researchers with families.*

# EMBL Group Leader Opportunities

## at EMBL Monterotondo, near Rome, Italy

The Scientific Programme of EMBL emphasises experimental analysis at multiple levels of biological organisation, from the molecule to the organism, as well as Computational Biology, Bioinformatics and Systems Biology. Within this overall structure, the EMBL Monterotondo Unit applies a wide range of modern technologies to diverse problems of whole organism biology. Currently, its research groups address areas of neurobiology, epigenetics, developmental biology and developmental genetics and are supported by state-of-the-art core facilities in histology, recombineering/ gene editing, flow cytometry, microscopy, and mouse transgenesis.

EMBL Monterotondo benefits from close interactions with groups at EMBL Heidelberg in the Developmental Biology, Genome Biology, Cell Biology and Biophysics, and Structural and Computational Biology Units with whom it shares core facilities in high-throughput sequencing, advanced light and electron microscopy, small molecule screening, protein production, and mass spectroscopy. EMBL Monterotondo groups also have access to research activities at the EMBL-EBI (European Bioinformatics Institute) in Hinxton, UK, and structural biology expertise at EMBL Hamburg and Grenoble.

## GROUP LEADER

### NEUROBIOLOGY

We seek dynamic and interactive individuals having recently completed their post-doctoral training with an excellent scientific track record and demonstrated experience or interest in molecular neurobiology.

The successful candidate will have a PhD in Neurobiology or an allied field, show strong potential in setting their own research agenda and show evidence of leadership qualities.

We encourage applications from scientists working on diverse questions relating to nervous system structure, function, and plasticity that would benefit from the mouse as an experimental system and the use of modern genetic and genomic approaches.

## GROUP LEADER

### EPIGENETICS AND MOUSE GENETICS

We seek dynamic and interactive individuals having recently completed their post-doctoral training with an excellent scientific track record and demonstrated experience or interest in the general area of epigenetics and developmental biology.

The successful candidate will have a PhD in Epigenetics or an allied field, show strong potential in setting their own research agenda and show evidence of leadership qualities.

We encourage applications from scientists working on diverse questions relating to mammalian organismal biology that would benefit from the mouse as an experimental system and the use of modern genetic and genomic approaches.

### ADDITIONAL INFORMATION

Further information about the positions can be obtained from the Head of the Monterotondo Unit Philip Avner (**philip.avner@embl.it**).

**Interviews** are planned for January 2017.

### APPLICATION INSTRUCTIONS

Please apply online through **www.embl.org/jobs** and include a cover letter, CV and a concise description of research interests and future research plans. Please also arrange for 3 letters of recommendation to be emailed directly by your referees to **references@embl.de** at the latest by 27 October 2016.

Further information on Group Leader appointments can be found under **www.embl.org/gl_faq**.

For more information please visit our website at:

**www.embl.de/jobs/work-at-embl**

W282470R