SPOTLIGHT ON BIOINFORMATICS

Biology goes digital

A new species of biologist is beginning to thrive in the niche created by recent genomic and computational advances.

"It's important to enjoy your job and be motivated: the best ideas come whilst I'm in bed, or walking my dog, or having a coffee with colleagues."

Federico Abascal

THERE ARE two paths to careers in bioinformatics, both of which require learning a new language. Computer scientists must become fluent in the life science terminology of genetics, genomics and cellular biology. Biologists must pick up skills in data analysis, including statistics, logic and programming. When the field was developing, fledgling bioinformaticians often taught themselves. Now, more institutions are offering formal training, and the field is maturing rapidly.

The skill set needed by a bioinformatician continues to evolve. In the early days of the human genome project, it was sufficient for scientists to find homologous genes of one organism in the genome of another. Now, bioinformaticians routinely compare multiple genomes, analyse regions that don't code for DNA, and incorporate a host of proteomic information in their analysis. Both the type and amount of information continues to expand, as biological techniques continue to improve.

As a result, the proficiency bar in bioinformatics continues to rise, along with the demand for talented bioinformaticians (see **new mobility: a case study**). A few decades ago the ability to scour databases to find a single gene provided at least a plank in the platform for a career in bioinformatics. Now, that skill is a basic part of a molecular biologist's toolkit, as essential as fundamental wet lab techniques. In response, bioinformaticians need to keep improving their skill set. And to really make a mark, they need to develop new tools that others in the field consider valuable.

"The learning curve is both bigger and steeper now," says George Asimenos, director of strategic projects at DNAnexus in Mountain View, California. DNA sequencing was relatively slow and expensive when he started graduate school 13 years ago. As speed has gone up and costs have come down, demands on bioinformaticians have grown. They need to be comfortable with mining much larger data sets, and looking for relationships between them.

Asimenos had to first get comfortable with the language of biology before he could dive into the data. He remembers hearing terms like "3 Prime" and "downstream" and thinking "What do these things even mean? How do I find out?"

He gave himself a crash course by reading textbooks, going to conferences and hanging out with biologists. "I had to overcome the vocabulary barrier," he says. During his undergrad, Asimenos had taken courses in statistics, engineering and computer science. Acquiring those skills earlier gave him time during grad school to bone up on his biology, literally, with an anatomy class that included human dissection; and figuratively, with stints in molecular biology wet labs.

Still, he had some fraught moments. His advisor tapped him to lecture a course on algorithms for biology, where Asimenos explained genetics, genomics and biology to a class of computer scientists. That experience in the biological deep end helped him, because DNAnexus makes software for biologists. He needs to know the language of the company's clients, so he can help create software to meet their needs. Bridging the gap between biology and computer science remains one of his biggest challenges. "That vocabulary is really deeply rooted in every single discussion," he says. But even more skills are necessary as the technology improves. Knowledge in machine learning and artificial intelligence might be needed for the next generation of bioinformaticians, Asimenos says.

After graduate school at Lund University, Sweden, Jean-Baptiste Cazier translated his knowledge of applied mathematics into fluency in statistical genetics analysis at deCode Genetics in Iceland. He focused on how statistics can be used to find areas in the human genome that contribute to increased risk of disease. Now, as the director of the Centre for Computational Biology, University of Birmingham, he has been tasked with teaching bioinformatics to scientists and clinicians at the UK's National Health Service (NHS). Part of the country's 100,000 Genome Project — which aims to sequence and understand the genome of 100,000 patients — his centre was the first of eight to start this educational programme in October last year.

DNANEXUS

George Asimenos

Bioinformatics can often equate to some form of programming. Although many people are initially intimidated by the prospect of learning programming languages, Cazier comforts them. His first lesson is to impart confidence. "Researchers — biologists, clinicians, whatever — can learn mathematics and programming," Cazier says. To get researchers comfortable, he uses data from a few patients to show how mutations are identified, and then asks them if the mutations are new and if the information is statistically reliable. "I am talking to their research brains, so it works," he says. »

EVELYN KIING

Sibon Li

New mobility: A case study

Bioinformatics offers a two-way career street. Once, people trained in maths, statistics and computer science could apply their skills to biological data, thus broadening their job prospects, whilst biologists were stuck firmly within their discipline. Now, scientists trained in bioinformatics are finding they can begin to transfer their skills into disciplines outside the life sciences.

Sibon Li studied bioinformatics at the University of Auckland in New Zealand, and did postdocs at the University of Southern California and the University of California, Los Angeles. That training prepared him for a job at Google in 2014.

What prompted you to get into bioinformatics?

As a teenager, I was always interested in computers and knew that I wanted to do something with them. I would buy PC magazines, and play with text-based video games. At the same time, at school, there were two things that were fascinating to me — I didn't care about my biology classes until they taught me about evolution, which was really exciting. And the other was probability theory in statistics, where I enjoyed the problem solving.

At the end of high school, I had no idea what I wanted to do at university. I was flicking through a university prospectus and stumbled across the bioinformatics program they were offering for the first time at the University of Auckland. I had no idea what it was or the job opportunities in the field but it seemed like the perfect union between all of the things that I was interested in.

What's your research background?

In academia my research focus was in developing algorithms for understanding the variation in rate of molecular evolution. As part of my research, I worked on the BEAST project, which my PhD and postdoc advisors Alexei Drummond and Marc Suchard had developed at Oxford.

Currently, I work at Google as a software engineer on the Knowledge Graph team, focusing on natural language understanding. Our technology is used in Google Search, as well as a range of other products.

How did you interact with more biologically-minded people to solve problems?

Frankly, there was very little interaction between the groups I worked with and traditional biologists. Computational biologists know enough about the biology to find some problems to solve, but often fail to address the biologically relevant and interesting problems.

On the other hand, I feel biologists believe that they have all the tools necessary to get good results. Of course, this is inefficient in many cases. This is a general problem in the bioinformatics community — there needs to be more communication and collaboration across the spectrum.

How do you bridge the communication gap between computer scientists and biologists?

Attend general conferences rather than going to those that are specialised to your area of research. You get to interact with others outside of your field of work and the feedback can be valuable. In addition, other researchers may see the significance of your research towards their own, and might want to collaborate.

Also, try to interact with colleagues across departments. When I first started my graduate work, my desk was in the biology department. I made a lot of friends there and attended biology seminars. Often, I would see how my research would benefit others and assisted some of my peers in their computational work.

What advice do you have for biologists or computer scientists considering bioinformatics?

I think an understanding of bioinformatics and proficiency in computational analysis is essential towards being a biologist in this day and age. The impression that I get from biologists is that learning these sorts of techniques is difficult and outside of their comfort zone. In reality, it's fairly easy to understand and just requires a change in mindset.

For computer scientists, I would say that there are a ton of interesting problems inside biology that are worth solving. The problems in biology are no different to the traditional problems that computer scientists generally focus their efforts on, in the sense that they are complex, intangible and challenging. If anything, the benefit to the world is potentially much greater than many of these other fields.

Finally, what does a degree in bioinformatics get you?

A degree in bioinformatics provides you with a diverse skill set that opens doors for a range of career options. I myself transitioned to working at Google on pure computer science problems such as natural language understanding and developing infrastructure. Peers from my program at university have found work in areas like biostatistics, pure statistics and biology. Outside of academia, there are plenty of careers for bioinformatics graduates in companies doing things like software and biotech.

» That breaks the ice for basic programming — especially when he demonstrates how code can help them ask and answer scientific questions. He has been surprised at the response. "I was quite worried about the teaching course, but they embraced it and asked for more," says Cazier.

Basic bioinformatics skills can empower biologists to make use of their own data: after all, they have the best understanding of biological processes. However, because the field is advancing so quickly, they need to keep in touch with the "hardcore" bioinformaticians to have any hope of keeping abreast with the latest developments.

Vicky Schneider

Vicky Schneider, associate professor and deputy director of the EMBL Australia Bioinformatics Resource, at The University of Melbourne, Australia, says one way each side can learn the language of the other is by having more conversations. "You have to have a minimal common vocabulary," Schneider says.

More dialogue between users and developers results in better tools, she says. For instance, a computer science-trained developer might create a powerful tool. But if someone with a biology background doesn't know how to use it, that tool is useless. "The two sides need to work together to develop user interfaces," she says.

There are formal efforts in place to achieve this. For instance, the Global Organisation for Bioinformatics Learning, Education & Training (GOBLET) helps each side learn the others' science. But even that can only go so far. It is unrealistic for specialists from each side of the field to be completely fluent in the other's field. "Each side has to understand their own limitations," Schneider says.

David Martin, a senior lecturer in bioinformatics at the University of Dundee, agrees with Cazier that biologists need more familiarity with bioinformatics. "The core skills for modern data-rich biology are not always there," he says. If he could, he would teach every biology grad student enough skills so that they could do some basic programming, read data into a file, then be able to manipulate and process it — "not enough to be a computer scientist, but enough to have the tools to work with the data," he says (see **skills spectrum**). But money and time is always a problem. "These skills take time to develop and craft, much like lab skills take time to develop and craft," Martin says.

However, it can be done, if one is willing to put in the work, says Joseph Mullen, a PhD student at Newcastle University in the United Kingdom. After an undergraduate degree in biology and with little computational experience, he decided that the career opportunities bioinformatics would open up would be worth the effort. "It was an incredibly steep learning curve," Mullen says. "I jumped into it with everything I had." Indeed, between coursework, working three part-time jobs to fund his education and putting in the hours to learn multiple programming languages, he had time for little else. Mullen estimates he averaged about five hours of sleep during his MSc year.

The sacrifice paid off, though. The Engineering and Physical Sciences Research Council (EPSRC) and GlaxoSmithKline are funding his PhD research. In return, Mullen contributes to drug discovery work for the company. He has already been offered a government-funded joint postdoc position with Newcastle and Prozomix — a biotech company based in Northumberland, United Kingdom — even though he hasn't yet written up his dissertation.

Federico Abascal was similarly computationally illiterate when he completed his undergraduate degree in 1998. Now he works as a bioinformatician at the Wellcome Trust Sanger Institute in the United Kingdom — one of the world's most renowned bioinformatics hubs. When he finished his undergrad work, he remained interested in biology, but knew he didn't want to perform experiments.

He took a course in the programming language C, then went on to graduate school at the Spanish National Biotechnology Centre in Madrid. His drive to solve problems in evolutionary biology led him to learn more programming languages. "Once you know one language, it is easier and easier to learn others."

He advises would-be bioinformaticians to get out of the lab as much as they can to avoid losing perspective. "In my case, the best ideas never come in front of a computer," Abascal says. "It's important to enjoy your job and be motivated: the best ideas come whilst I'm in bed, or walking my dog, or having a coffee with colleagues."

The next generation of bioinformatician may well find the lab and the computer indistinguishable. Dual training in both fields, as early as undergrad education, may well become the norm, says Atul Butte, director of the University of California, San Francisco's Institute for Computational Health Sciences.

Butte is a pioneer in training for bioinformatics. In high school, he was fascinated by *National Geographic* covers displaying MRI and CT scan images. He thought combining computers and medicine would prepare him for a career in radiology.

ELISABETH FALL

Atul Butte

MATTHEW WAKEFIELD

He pursued that career by enrolling in an eight-year programme at Brown University, studying medicine and computer science. Towards the end of his studies, gene expression microarray chips were invented, the human genome project took off, and the era of big data in biology was born. He emerged with a skill set training him in both worlds.

He may have been the exception then, but Butte sees dual training as the new norm. "More and more people come up with both," he says.

In fact, it is harder now to get into top bioinformatics graduate programs conversant in only biology, or only computer science. "You have to demonstrate you know more than a little of both," Butte says.

But knowing enough to use the software may not be enough to excel, Butte says. "The point of being in this field is to develop new tools, new methods — you have to innovate. You have to write new code." Focusing too much on one technique or one problem could be career limiting, he says.

He advises constant learning — the amount of data keeps growing and the nature of it keeps changing. "Treat the field with respect," he says. "If you want to thrive in biomedical informatics it can't be a casual thing. You have to be here to stay."

This content was commissioned and edited by the Naturejobs editor

SKILLS SPECTRUM

There are **three** essential skill sets bioinformaticians need. Here's where to start.

1. COMMAND

Understand how **Unix** commands work.

2. PROGRAM

Learn **Python**, a basic language. Then consider **R**, a useful language for handling statistics.

3. DATA

Understanding what type of data is in different kinds of **databases**, and how to mine it, is essential. Learning relational database techniques is another **plus**.

EMBL

EMBL-EBI

The European Bioinformatics Institute (EMBL-EBI) is one of the world's leading bioinformatics institutes, employing around 570 staff with a focus on both computational biology research and bioinformatics service delivery. Located on the beautiful Wellcome Genome Campus near Cambridge, EMBL-EBI offers a strongly collegiate working environment as part of the world-renowned European Molecular Biology Laboratory (EMBL). We are looking for outstanding individuals to advance EMBL-EBI's research and bioinformatics services in the following faculty roles. EMBL-EBI is committed to achieving gender balance in its leadership and strongly encourages applications from women, who are currently under-represented at faculty level.

Head of Research

This is an exciting opportunity for a successful scientist with high-calibre leadership skills to provide strategic direction for EMBL-EBI's research and influence the broader European computational biology research area. As Head of Research you will provide leadership for the entire EMBL-EBI research portfolio – similar to a Head of Department in a University setting, but without teaching and with minimal administration commitments. You will also run your own research group. We are looking for a leader with an impressive record in delivering world-class research who can inspire the next generation of bioinformatics researchers. Communication and collaborating, both within EMBL and with the global scientific community, will be a key aspect of this role.

Informal enquiries are welcome – please contact Ewan Birney, Director of EMBL-EBI **birney@ebi.ac.uk** or Nick Goldman, Research Group Leader **goldman@ebi.ac.uk**.

Research Group Leaders

We are looking for enthusiastic, motivated computational biologists to lead independent research groups in EMBL-EBI's blue skies research programme. We offer successful applicants a unique opportunity to pursue their own research direction in computational biology, and provide generously for the recruitment of students and postdocs. Research Group Leaders at EMBL-EBI enjoy world-class facilities and technical infrastructure, and are free from teaching requirements. We encourage collaboration with other groups at EMBL, with the co-located Wellcome Trust Sanger Institute and in the broader Cambridge area in the UK. We are particularly interested in hiring young investigators, subsequent to their first postdoctoral position, or even direct from a PhD programme. We focus on the potential demonstrated by aspiring group leaders to develop over time. These research leader roles are open to all areas of computational biology, from methods development through to data discovery, and from electron microscopy data analysis through to ecological modelling, including genomics, transcriptomics, metabolomics and chemoinformatics.

More information is available at
www.ebi.ac.uk/research/research-group-leaders-at-embl-ebi

Service Team Leaders

EGA AND ARCHIVE INFRASTRUCTURE

Help shape the sequence archive infrastructure of the European Genome-phenome Archive (EGA) database and lead the team responsible for data presentation of this archive along with the European Nucleotide Archive (ENA) and the European Variation Archive (EVA). Leading a faculty level team of around 15 staff members you will be responsible for ensuring data flow and infrastructure development are coordinated across these high-value archive resources, which are designed to serve as a platform for research into molecular medicine, disease, and connections between genome variation and phenotype for human and other species. As well as providing leadership and management for existing grants, your work will involve establishing external collaborations and raising funds via competitive peer-reviewed grants to supplement core funding allocated for the team leader position, staff members and computational infrastructure.

With an MD or a PhD in Genetics, Molecular Biology, Computer Science or other relevant field and postdoctoral experience, you will also be able to demonstrate practical experience with bioinformatics, genome-wide computational analysis and database infrastructure, preferably in a production environment.

VARIATION ANNOTATION

We are looking for a visionary Team Leader to head up EMBL-EBI's newly-formed faculty level Variation Annotation Resources team of around 15 staff members, responsible for the annotation, curation and distribution of variation data from human and other species. Working closely with other components of the Ensembl project, you will also form external collaborations and raise funds via competitive peer-reviewed grants to supplement core funding allocated for the team leader position, staff members and computational infrastructure. A key task will be the continued development of EMBL-EBI's high-value resources designed to serve as a platform for research into molecular medicine, disease and connections between genome variation and phenotype for human and other species.

Ideally you will hold an MD or a PhD in Genetics, Molecular Biology, Computer Science or other relevant field with postdoctoral experience. You will be able to demonstrate practical experience with bioinformatics, genome-wide computational analysis and database infrastructure, preferably in a production environment.

More information about these roles, closing and interview dates and application instructions are available at **www.embl.org/jobs**.

EMBL-EBI offers world class research and computer facilities in addition to highly competitive pay and excellent benefits plus on-site amenities such as gym, conference centre, nursery and free shuttle buses. We also benefit from close ties to the Wellcome Trust Sanger Institute and the University of Cambridge.

W275111R

WELLCOME GENOME CAMPUS

ADVANCED COURSES AND SCIENTIFIC CONFERENCES 2016

CONFERENCES

Mouse Models of Disease: Improving Reproducibility of Pathology Endpoints in Challenge Models
9–11 February

Evolutionary Systems Biology: From Model Organisms to Human Disease
2–4 March NEW

Single Cell Biology
8–10 March NEW

Genomics of Rare Disease: Beyond the Exome
13–15 April

Genomics of Brain Disorders
25–27 April NEW

Mitochondrial Medicine: Developing New Treatments for Mitochondrial Disease
4–6 May

Molecular Biology of Hearing and Deafness
17–20 May

Genomic Epidemiology of Malaria
5–8 June

Virus Genomics and Evolution
8–10 June NEW

Curating the Clinical Genome
22–24 June NEW

Exploring Human Host-Microbiome Interactions in Health and Disease
7–9 September

Single Cell Genomics
14–16 September NEW

EMBL–Wellcome Genome Campus Conference: Proteomics in Cell Biology and Disease Mechanisms
14–17 September NEW

Genome Informatics
19–22 September

EMBL–Wellcome Genome Campus Conference: Big Data in Biology and Health
25–27 September NEW

The Genomics of Common Diseases – in association with Nature Genetics
25–28 September

Computational RNA Biology
17–19 October

Epigenomics of Common Diseases
1–4 November

COURSES

Fundamentals of Clinical Genomics
13–15 January

Genomics and Clinical Microbiology
17–22 January

Working with Pathogen Genomes
17–22 January

Genomic Practice for Genetic Counsellors
3–4 February

Immunophenotyping: Generation and Analysis of Immunological Datasets
21–27 February NEW

Mathematical Models for Infectious Disease Dynamics
22 February–4 March

Next Generation Sequencing
11–17 March

Genetic Engineering of Mammalian Stem Cells
11–23 April

Malaria Experimental Genetics
8–14 May

Bioinformatics Summer School
13–17 June

Mouse Models: Genetics Breeding and Experimental Design
13–17 June

Functional Genomics and Systems Biology
15–24 June

Practical Aspects of Small Molecule Drug Discovery
19–24 June

In Silico Systems Biology
3–8 July

Drosophila Genetics and Genomics
3–10 July

Evolutionary Biology and Ecology of Cancer
11–15 July NEW

Public Engagement Masterclass
20–22 July

Human Genome Analysis: Genetic Analysis of Multifactorial Diseases
20–26 July

Leena Peltonen School of Human Genomics
21–25 August

Design and Analysis of Genetic-based Association Studies
26–30 September

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

Molecular Neurodegeneration
28 November – 4 December

Proteomics Bioinformatics
4–9 December

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

OVERSEAS COURSES

Human and Vertebrate Genomics: Bioinformatics Tools and Resources
7–12 February
(Bangkok, Thailand)

Molecular Approaches to Clinical Microbiology in Africa
5–10 March
(MRC Unit, The Gambia)

Genomics and Epidemiological Surveillance of Bacterial Pathogens
17–22 April
(Buenos Aires, Argentina)

Human and Vertebrate Genomics: Bioinformatics Tools and Resources
12–16 September
(Montevideo, Uruguay)

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)

U274566EL

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