

## BIOINFORMATICS

# The rise of the genome bloggers

Hobbyists add depth to ancestry trawls.

BY EWEN CALLAWAY

Hours after Joseph Pickrell put his genome on the internet, an anonymous blogger took the data and concluded that he came from Ashkenazi Jewish stock. Pickrell, a genetics graduate student at the University of Chicago, Illinois, was sceptical about the claim. But after talking to relatives, he discovered that he had a Jewish great-grandfather who had moved to the United States from Poland at the turn of the nineteenth century. “It was a part of my ancestry I was totally unaware of,” he says.

The blogger, who writes under the pseudonym Dienekes Pontikos at <http://dodecad.blogspot.com>, had commandeered Pickrell's DNA as part of the Dodecad Ancestry Project, an ambitious project in which cutting-edge genomic analysis meets Web 2.0. Pontikos analyses genetic data submitted by followers

of his blog to reconstruct personal ancestry and human population history — and reports his findings online. He is part of a small but growing group of ‘genome bloggers’, a mix of professional scientists and hobbyists proving that widely available tools for computational biology could enable recreational bioinformaticians to make new discoveries.

“They are not amateurs. They are far from being amateurs,” says Doron Behar, a population geneticist at Rambam Health Care Campus in Haifa, Israel, who studies human history. “I cannot stress enough the level of appreciation I have for their efforts.”

Pontikos has so far analysed several hundred thousand single-letter DNA variations from more than 2,200 individuals. That includes more than 200 submitted to him by readers of his blog, who had had their genomes analysed by genetics testing firms such as 23AndMe, based in Mountain View, California, with the remainder coming from publicly available datasets. The

readers volunteering their genomes (identities stay private) are mostly keen to delve into their own ancestry. But Pontikos, who is from Greece and describes himself as an “anthropology dilettante”, is more interested in unfurling the history of populations that tend to be overlooked by human-population geneticists. For instance, his analysis of genomes from people living in northern Eurasia reveals a genetic connection between populations in northern Finland and central Siberia (see ‘Meet the ancestors’).

David Wesolowski, a 31-year-old Australian who runs the Eurogenes ancestry project (<http://bga101.blogspot.com>), also focuses on understudied populations. “It’s a response, in a way, to the lack of formal work that’s been

**“They are not amateurs. They are far from being amateurs.”**

done in certain areas, so we’re doing it ourselves,” he says. Wesolowski and a colleague have drilled into the population history of people living in Iran and eastern Turkey who identify as

descendants of ancient Assyrians, and who sent their DNA for analysis. Preliminary findings suggest their ancestors may have once mixed with local Jewish populations, and Wesolowski plans to submit these results to a peer-reviewed journal.

But Pontikos sees little point in formally publishing his findings. “I can bypass them entirely, and have the entire world review what I write,” he wrote in an e-mail. Indeed, comments on his blog — “could you please provide the eigenvalues for the principal component analysis”, for instance — read like the niggling recommendations of a manuscript reviewer.

Pickrell notes that Dodecad and Eurogenes use cutting-edge techniques and open-source software developed by geneticists studying population history. The methods — which involve modelling past mixing between populations and distilling vast quantities of genotype data — still stir debate in the peer-reviewed literature because they can be difficult to interpret unambiguously, says John Novembre, a population geneticist at the University of California, Los Angeles. Behar, whose data on Jewish ancestry have been used by both projects, cautions that the techniques are more robust when applied to the history of an ethnic group, rather than the ancestry of an individual.

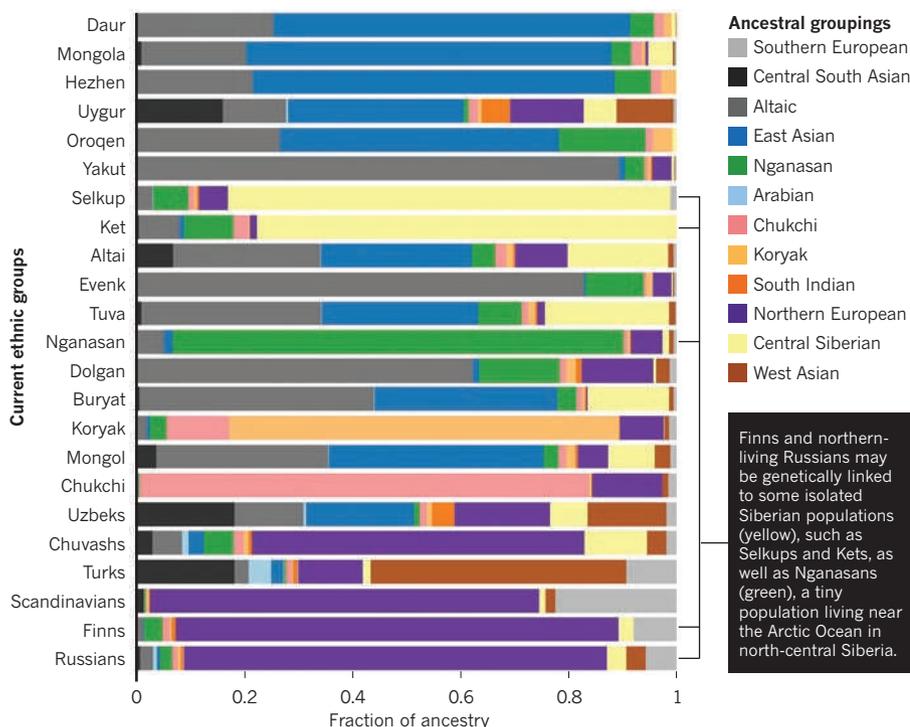
In response to concerns about the genetic privacy of those offering their genomes for analysis, “I don’t think this is too worrisome,” says Hank Greely, director of the Center for Law and the Biosciences at Stanford University in California. Both projects provide adequate privacy protection, he says, although they both could do a slightly better job at disclosing the risk of a release.

Although people may be happy to part with their genomes to

**NATURE.COM**  
The human genome at ten:  
[www.nature.com/humangenome](http://www.nature.com/humangenome)

## MEET THE ANCESTORS

An analysis of data from the Dodecad Ancestry Project highlights genetic links between different ethnic groups.



learn more about their ancestry, the genetic and trait data needed for biomedical applications are much harder, if not impossible, for amateurs to come by. Public repositories, such as the US National Institutes of Health's database of Genotypes and Phenotypes, tightly restrict access.

One effort to change that is the Personal Genome Project, which is spearheaded by George Church, a geneticist at Harvard Medical

School in Boston. The project aims to make the complete genome sequences and traits of 100,000 people freely available to anyone, with no strings attached. So far it has enrolled 1,000 participants and published near-complete genomes for 10 of them. Pickrell and 11 other scientists and genomics experts also added to the trove of freely available genomic data recently when they released their genetic data as part of a project called Genomes Unzipped.

Church argues that better access to high-quality data could help this kind of informal bioinformatics to flourish, enabling computer-savvy people to make important contributions to genomics, just as they have with online businesses such as Facebook. "It didn't take that much training to become a social-networking entrepreneur. You just had to be a good coder," he says. With bioinformatics, "I think we're in a similar position." ■

## INFORMATION TECHNOLOGY

# Developers call for handy lab aids

*Macmillan hopes to partner with scientists to turn software into commercial products.*

BY DECLAN BUTLER

It's an unfortunate truism that scientists often have better software tools for managing their music or family photos than they do for tracking their experiments and data. Major software companies tend to focus on much larger consumer and business markets, and offer little software for researchers.

As a result, others are rushing to occupy the niche, among them Digital Science ([www.digital-science.com](http://www.digital-science.com)), launched last week by *Nature's* parent company, Macmillan Publishers. Digital Science's strategy is not only to develop its own products — independently and with partner companies — but also to tap into the innovation of researchers themselves. With few existing products available to satisfy their needs, a growing number of scientists have developed their own tools with which to better organize their research lives.

The company has launched an open call for researchers who have written promising software to submit proposals for turning it into a commercial product. Researchers understand their colleagues' needs best, but often "don't have resources to turn their software into a polished product", says Timo Hannay, Digital Science's managing director and former publishing director of *nature.com*. The company aims to partner with researchers, or their start-ups, to provide them with the financial, developer and business resources they need.

The goal is to offer researchers tools that are as intuitive and user-friendly as well-designed consumer software. The company will initially focus on text-mining software, metrics-based tools to help institutions and funders better assess the performance of their funding and researchers, and lab-management software to help to keep tabs on anything from experiments to reagents.

There is certainly a massive need for better software to increase productivity at every stage of the complex scientific workflow, says Alexander Griekspoor, who founded



Got any nifty apps for your lab?

the company Mekentosj, based in Aalsmeer, the Netherlands, which produces software for molecular-biology applications. Sriram Kosuri, a bioengineer at the Wyss Institute for Biologically Inspired Engineering at Harvard University in Boston, Massachusetts, agrees. "Every bench scientist I know has their own, and self-admitted non-optimal way to organize their information," he says. Kosuri is a founder of OpenWetWare ([openwetware.org](http://openwetware.org)), a wiki for sharing lab protocols and data between biology groups worldwide. "I think there is an increasing willingness to pay for useful tools," he says.

Scientists struggling to organize thousands of article PDFs strewn across their hard drives have already embraced services such as Griekspoor's Papers ([mekentosj.com/papers](http://mekentosj.com/papers)), and London-based Mendeley ([www.mendeley.com](http://www.mendeley.com)), which bring the simplicity of iTunes to managing papers. Mendeley also offers social-networking facilities and other features.

Publishing giant Elsevier has recently entered the research-services market with its SciVal suite of metrics tools. It is one of several publishers that view providing institutions

with performance-measuring applications as a strategic move, says David Bousfield, the London-based vice-president and lead analyst at Outsell, a publishing and information consultancy. Elsevier has also launched SciVerse, a platform for searching and sharing content from Elsevier's own databases and the web. The product provides programming interfaces that allow researchers to build their own customized applications on top of content from Elsevier and other data sources, such as government databases.

Kosuri thinks that Digital Science's strategy of refining software created by researchers makes sense. "Getting the most important sets of tools developed more extensively would be tremendous," he says. But Michael Eisen, a geneticist at the University of California, Berkeley, and co-founder of the Public Library of Science, notes that many researchers have little enthusiasm for devoting time to perfecting software they have written. "I expect few would be interested," he says.

"We're not trying to convince anyone to create a commercial product if the desire isn't already there," Hannay counters. "We're trying to tap into the small but significant proportion of researchers who have identified an unmet need and are trying to do something about it."

Freely available code isn't an ideal alternative — most open-source software for scientists "sucks", says Eisen, although "a lot of it is really good and essential to what we do". But many lab-management tasks can be done using generic consumer open-source tools, he says. "The people I know mostly use wikis to keep track of stuff in the lab: they're free, flexible, and easy to set up and use."

Eisen also warns against thinking of software as a panacea. "Everyone has aspirations to be better organized in the lab," he says, "they think there's a magic piece of software out there that will solve all their problems for them, but then they realize the problem is really that they're disorganized." ■