

## TECHNOLOGY

# No rest for the bio-wikis

*Biologists' collaborative data repositories come of age.*

BY EWEN CALLAWAY

Most Tuesdays, a group of scientists at the Wellcome Trust Sanger Institute in Hinxton, UK, meets over lunch to edit Wikipedia pages. But there is no obsessing over the minutiae of Britney Spears's career to be found here — instead, they are building the next generation of global biological databases.

"Yesterday, we created 18 microRNA articles," says Alex Bateman, a computational biologist at the Sanger Institute who helped to found Rfam, a database that includes a set of Wikipedia articles covering about 1,500 families of RNA molecules, maintained by more than 2,000 editors. In the past few years, community-curated biological websites such as this have multiplied, and on 29 November scientists will gather for a first-of-its kind conference called Biological Wikis, in Naples, Italy, to take stock of the 'bio-wiki' approach and plan future expansion.

Wikis, collaboratively edited web pages named after the Hawaiian word for quick, offer a solution to the growing data glut in biology. Conventional databases are struggling to keep up with the flood of information about genes and proteins that labs are amassing by the terabyte ( $10^{12}$  bytes). "The old model of annotation, where the central database handles that information, doesn't work," says Dan Bolser, a computational biologist at the University of Dundee, UK, who is involved in curating a protein-structure database called PDBWiki.

So biologists increasingly maintain and update web pages focused on particular genes or proteins — or any concept or object of interest (see 'The bio-wiki boom'). The success of Wikipedia, the ubiquitous online encyclopaedia compiled using the same technique, proves that community annotation works, say bio-wiki enthusiasts. "There's a genuine generational and

technological change that's happening," says Ewan Birney, a bioinformatician at the European Bioinformatics Institute in Hinxton.

At the Naples meeting, scientists will discuss the lessons of the handful of bio-wikis that are beginning to assemble a critical mass of readers and contributors. These bio-wikis are now attracting more contributions from the community than from the developers themselves, and advocates say that the sites are becoming indispensable tools in some areas of biology.

Gene Wiki, for instance, has more than 10,000 Wikipedia pages, each devoted to a single gene, and draws some 4 million views and 1,000 edits every month. Many scientists come to the site after an experiment identifies a laundry list of genes that are of interest in their work — and which they know little about or have never heard of, says Andrew Su, a bioinformatician at the Genomics Institute of the Novartis Research Foundation in San Diego, California, and one of the driving forces behind Gene Wiki. "It's a great way to go in and get up to speed," he says.

Bio-wikis that are hosted by Wikipedia benefit from the contributions of its existing altruistic community of 'Wikipedians', says Bateman. His team will soon launch a protein-family wiki that will also be hosted on Wikipedia. "One of the big surprises for me in all of this is that the contributions we're getting are as much from non-scientists as they are from scientists," he says. Non-specialists with enough interest and expertise to add information about the activity of microRNAs may be rare, but others can provide help with page formatting and standardization, which are "important, valid contributions", says Bateman.

But Wikipedia comes with its own rules and idiosyncrasies, which limit its usefulness for some kinds of biological data. To merit a page on Wikipedia, a subject — whether a gene, ▶



ENGEL, UKB

After more than a year of uncertainty, the mood is upbeat. "Germany has not been competitive in health research in the past decades," says Oliver Eickelberg of the Comprehensive Pneumology Centre in Munich, a member of the new lung-research centre. Even sceptical universities are coming round to the idea. Clinician-researcher Andreas Zeiher, from the Goethe University in Frankfurt, who is a member of the cardiovascular-research centre, says they now realize "that it is better to have something than nothing".

One project that the lung centre is discussing is the identification of molecular signatures of lung fibrosis in order to develop targeted therapies. A single clinical centre would typically have only some 150 such cases to work with, but by pooling patients from around Germany the lung centre expects to be able to recruit up to 1,000. This number would make analysis — using the sequencing and systems-biology platforms the centre intends to establish — statistically feasible. "The new centres allow us to concentrate our activities with the security of long-term investment and increase our visibility internationally," says Eickelberg. "That's important if we want to attract outside investigators."

"Germany is making a smart move," says David Warburton, a physician-scientist at the Children's Hospital Los Angeles at the University of Southern California, and a member of the expert advisory panel for the lung-disease centre. He believes that synergies between groups involved with the new centres will add value to the "well-funded, well-organized but decentralized research activities in Germany".

Participants agree that the effort will succeed only if it is backed by stable funding, and in a statement to *Nature* the ministry confirmed that this is its intention. Oncologist Otmar Wiestler, head of the German Cancer Research Centre in Heidelberg, is confident that the government will not allow the centres to fail. "This is a new era for biomedical research in Germany — nothing less." ■



## MORE ONLINE

### STAY CURRENT

- Demonic device converts information into energy [go.nature.com/g9oqpr](http://go.nature.com/g9oqpr)
- ITER fusion project begins hunt for budgetary savings [go.nature.com/djzau8](http://go.nature.com/djzau8)
- Beating heart cells controlled by light [go.nature.com/zaxrhz](http://go.nature.com/zaxrhz)

### VIDEO



Caught on camera: the secret of how cats lap milk [go.nature.com/8ka2wt](http://go.nature.com/8ka2wt)

IDAMINI/LAWRY

► a protein or any other biological entity — must be considered noteworthy by the Wikipedia community. Important data, such as protein crystal structures and genetic variants, do not always qualify, says Su.

The rub is that many bio-wikis not housed within Wikipedia struggle to attract readers and editors. But Alexander Pico, a bioinformatician at the Gladstone Institute of Cardiovascular Disease in San Francisco, California, thinks that these problems will fix themselves. “The vision going forward is that more and more scientists will be involved in the curation and consumption of data and they won’t need to accidentally stumble on it through Wikipedia,” he argues. His team’s WikiPathways site, which characterizes and visualizes biological pathways and is independent of Wikipedia, thrives because the systems biologists it attracts are already avid consumers of other people’s data and therefore see the benefits of a wiki, says Pico.

One challenge to bio-wikis that will be addressed at the Naples meeting is their text-based default layout, says Su. Written entries devoted to individual genes and proteins fit well within Wikipedia. But the format is a poor match to the highly structured, searchable data

sets favoured by computational biologists, which include the precise relationships between genes, proteins and other factors. A number of bio-wikis, including Su’s Gene Wiki, are adopting a software package called Semantic MediaWiki. This will bring them closer to working like true databases: for instance, the software could allow scientists to search for all the proteins phosphorylated by a specific kinase enzyme expressed in a particular tissue, rather than having to look up each interaction individually.

Despite such innovations, bio-wikis might not truly take off until scientists can get career-advancing credit for contributing to them.

## THE BIO-WIKI BOOM

Collaboratively edited biological databases help the community keep up with a flood of information.

Wiki	Topic	URL	Pages
EcoliWiki	Genes and proteins in <i>Escherichia coli</i>	ecoliwiki.net	63,784
PDBWiki	Protein structures	pdbwiki.org	64,071
Wikigenes	Genes, proteins and biochemicals	www.wikigenes.org	123,332
Rfam wiki	RNA families	en.wikipedia.org/wiki/User:Ppgardne/Rfam10.0	824
Gene Wiki	Genes	en.wikipedia.org/wiki/Portal:Gene_Wiki	10,118

“Editing your wiki is not going to get you your grant, it’s not going to get you promoted,” says Jim Hu, a molecular biologist at Texas A&M University in College Station and one of the founders of EcoliWiki, a repository of information about the model bacterium *Escherichia coli*. One database trying to solve the attribution problem is Wikigenes, a site devoted to annotating 120,000 genes and other biomedical concepts, which meticulously records and displays the individual contributions of its 1,800 active editors.

Persuading funding agencies and tenure committees to take those contributions seriously would mark a major milestone for bio-wikis. Until then, says Bolser, “it’s not clear to scientists why they should spend time editing a wiki article if it just gets them kudos from a few geeks on Wikipedia”. ■

### CORRECTION

The Editorial ‘A painful remedy’ (*Nature* **468**, 6; 2010) misspelt the name of physicist Jan Hendrik Schön as ‘Hendrick’ and incorrectly gave his nationality as Austrian. He was born in Germany.