Each bespoke knockout in the Sanger group's library contains an added 'conditional allele'. This allows scientists to disrupt gene function in a living mouse at any body site and at any point in the animal's development by the timely addition of enzymes that recognize the inserted allele. By this means, the effects of the missing gene do not kill the mouse before the researchers have a chance to study it.

"It is truly a feat of genius," says Geoff Hicks, a geneticist at the University of Manitoba in Winnipeg who leads the Canadian contribution to the IKMC. "This paper really pushed the technology in an extremely innovative way and met a challenge that seemed unattainable."

Various groups in the international effort are using other, non-conditional techniques to inactivate thousands more genes. Researchers in Texas, Canada and Germany have mutated close to 12,000 genes using an untargeted approach called gene trapping, and Regeneron Pharmaceuticals, a company based in Tarrytown, New York, has specifically targeted around 3,500 genes using a technology that works well in smaller genes but results in mice that are less flexible for research than conditional knockouts. "The approaches are complementary," says Aris Economides, Regeneron's senior director of genome engineering technologies. "This is going to play out well for the end user."

To date, nearly 17,000 different genes have been knocked out, leaving only around 3,000 more to go. The Sanger team, however, hopes to replace most of the genes hit by gene trapping with conditionally targeted knockouts, because targeting allows individual genes to be manipulated with greater precision.

***"It is one of the most significant biological resources in the past century of science."***

Already, mutant mice have been generated from almost 1,000 of the embryonic stem-cell lines obtained, and the IKMC repositories in the United States, Canada and Europe receive hundreds of new orders every month. The next challenge is to study the function of each missing gene. To this end, the US National Institutes of Health last year committed $110 million over the next five years to characterize around 2,500 of the IKMC's mutant mice through the International Mouse Phenotyping Consortium, with plans for another $110 million to define 5,000 more if the first phase is successful.

"Knocking out the mice is simple relative to the huge task of finding out what all those genes do," says Richard Finnell, a geneticist at the Texas A&M Health Science Center in Houston. ■

BOTANY

# Species spellchecker fixes plant glitches

*Online tool should weed out misspellings and duplications.*

BY JOHN WHITFIELD

R. EVANS/PHOTOLIBRARY

**Would it smell as sweet by any other name?**

Brian Enquist and his collaborators were delighted with their freshly compiled data set of 22.5 million records on the distribution and traits of plants in the Americas. But their delight turned to horror when they realized that the data set contained 611,728 names: nearly twice as many as there are thought to be plant species on Earth.

Completed in December 2010, the records were intended to help Enquist and his colleagues to discern trends in how forest trees in a wide variety of environments respond to climate change. But the data were clearly full of bogus names, making it impossible to count the species in a particular area, or their relative abundance. "I started to question our ability even to compare something as basic as species diversity at two sites," says Enquist, a plant ecologist at the University of Arizona in Tucson.

This month, Enquist's team will unveil a solution that could help botanists and ecologists worldwide. The Taxonomic Names Resolution Service (TNRS) aims to find and fix the incorrect plant names that plague scientists' records.

"It looks really good," says Gabriela Lopez-Gonzalez, a plant ecologist at the University of Leeds, UK, who curates a database of forest plots. Fixing species lists by hand is arduous, she says. "This should save us a lot of time".

She and others agree that the problem is widespread in botanical databases. "Digitization has made the problem worse," says TNRS co-leader, botanist Brad Boyle, also at the University of Arizona. Boyle explains that as more data are added to digital records, the chance of introducing errors also increases. Even in herbarium specimens, which ought to be the gold standard for plant identification, about 15% of the names are misspelt, he says.

Many of the errors seem to arise because biologists are not as careful as they should be when entering data into digital records. The TNRS team estimates that about one-third of the names entered into online repositories — such as GenBank, the US National Institutes of Health collection of DNA-sequence data, or the Ecological Society of America's VegBank database of plant-plot data — are incorrect.

The other problem is that names change. Old names can be abolished when experts reclassify plants as ideas about evolutionary relationships change, or when they realize the species already had a name — an occurrence almost as old as taxonomy itself. The result is that the same plant can have many names, and not everyone knows which one to use. Such synonyms are a particular problem in the study of medicinal plants, says Alan Paton, a plant taxonomist and bioinformatician at Kew Gardens in London.

The TNRS was built with financial and technical support from iPlant, a project run by the US National Science Foundation to fund cyberinfrastructure for plant science. It corrects names by comparing lists that users feed into it with the 1.2 million names in the Missouri Botanical Garden's Tropicos database, one of the most authoritative botanical databases. If the TNRS cannot find a name in Tropicos, it uses a fuzzy-matching algorithm, similar to a word-processor's spellchecker, to find and correct misspellings. It also hunts through Tropicos's lists of alternative names and supplies the one that is most up to date. When Enquist ran the 611,728 names through the system, just 202,252 came back, showing that two-thirds of them were invalid.

Because Tropicos is less comprehensive for plants outside the Americas, the team hopes to link the TNRS with The Plant List (www.theplantlist.org), a collaborative compilation of databases from Kew and other sources. Launched online in December 2010, it aims to become a global record of plants. The scientists are also working on a tool to correct geographical data — one that knows, for example, that Brazil, Brasil and Brésil are the same place, and can recognize when someone has muddled up longitude and latitude. ■