

Recipes for supertrees

New Zealand is at the international forefront of phylogenetic reconstruction algorithms, which attempt to build a genetic data set from overlapping species into a single evolutionary tree. Charles Semple spoke with Jenny Rankine.

Creating a best-fit evolutionary tree for 20 species is extremely difficult, yet biologists now regularly try to build supertrees of hundreds or thousands of species. There are often conflicts in the genetic data. Sometimes these are errors, but they may represent the combination of different species into a new hybrid, lateral gene transfers or recombinations of DNA strands.

"These events are more common than people first thought," says Associate Professor Charles Semple, who is based at the University of Canterbury. "Sometimes two genes simply evolve in different ways." In these cases, instead of seeking a single tree, mathematicians can create a network that highlights the conflicting genetic signals.

Semple is a co-director of the NZIMA programme in algorithmics with Professor Mike Atkinson at the University of Otago and associate director Dr Mark Wilson at the University of Auckland. The programme aims to link local and overseas algorithmic researchers, and raise awareness of the field with high school, undergraduate and postgraduate students.

An algorithm is a list of well-defined instructions for completing a task from an initial state through well-defined successive states to an end-state. "A cake recipe is an algorithm," says Semple. "You have to do things in a certain order otherwise the cake may sink in the middle or fail to rise."

Computer games are a more common example. Algorithms are also behind packet routing on

the Internet. Another application is the use of phylogenetic algorithms by Russell Gray's group at the University of Auckland to analyse the development of Pacific languages. This has challenged existing theories about Pacific settlement patterns.

Many problems are so intractable that even the best algorithms may be unsolvable with a supercomputer, or take days to run. Semple gives the example of timetabling problems. "Universities need to schedule a lot of exams in a very short time, but want to avoid students having to attend two exams at the same time. The best solution is trial and error, but there are so many possibilities it can take a computer a long time to run through them." Semple will use examples like these as part of the programme's outreach to high school students in term four.

The programme's first event, a Workshop on Algorithms in Napier in February, enabled programme Masters student Josh Collins to reduce computer time on a problem from two days to two hours.

Algorithms are playing the role for many sciences that formulae play for physics, says Semple. "While faster computers enable far more extensive use of algorithms than we imagined a decade ago, the great leaps forward come from better algorithms, not better hardware."

Above: A phylogenetic tree of 42 representative mammals reconstructed by Frédéric Delsuc and Nicolas Lartillot using a Bayesian approach, concatenating 12 protein coding mitochondrial genes.

Below: Participants in the Workshop on Algorithmics in Napier, February 2008.

while, if, else, input, compute..

