

# Software, evolution and micro-inversions -- improving the building of phylogenetic trees

December 19 2006

---

Biologists will be able to reconstruct the process of evolution, determine relationships between species and build phylogenetic trees with greater accuracy thanks to a new method for identifying “microinversions,” which are extremely short strings of inverted nucleotides.

This new work from researchers at UC San Diego and Brown University will appear in the online version of PNAS on December 18, 2006.

Microinversions – usually tens to thousands of base pairs in length – can only be detected if you have the exact nucleotide sequence of the same genomic region for all the species you are considering. Many recent studies have pointed to microinversions as large sources of genetic diversity that have not previously been characterized, and the new research from UCSD provides a more careful and accurate approach to identifying microinversions.

“As more fine-grained genomic data becomes available, microinversions will be increasingly important in understanding genetic diversity both between and within species,” said Mark Chaisson, the first author on the paper and a Bioinformatics Ph.D. student from UCSD’s Jacobs School of Engineering.

“This method might be able to provide evidence for the entire mammalian phylogeny, such as the presence of an afrotheria clade,” he said.

Using data from their microinversion detection technique – an open-

source software system called InvChecker – the researchers reconstructed the phylogenetic tree for 15 mammals. This work largely confirmed the existing phylogenetic tree that connects these mammals.

“Three years ago, we didn’t know microinversions existed,” explained Pevzner. “When they were discovered, there was a lot of skepticism. In the last year, scientists have discovered just how common they are in evolution – even in variation between humans, which is why they are such a hot topic today.”

“We’ve only looked for microinversions in 0.1 percent of the genomic sequence from several mammals, and we can already confirm many of today’s ideas about the history of evolution. When similar analyses extend to one percent of the genomes under investigation, we’ll have a 10 fold increase in data. This should shed light on splits between species that have been debated in molecular evolution,” explained Pavel Pevzner, the senior authors on the paper, a computer science and engineering professor at UCSD’s Jacobs School of Engineering, and director of the newly-established Center for Algorithmic and Systems Biology (CASB) at the UCSD Division of Calit2.

“This microinversion detection method could be used for detecting human structural variants once we have the necessary data,” explained Ben Raphael, a professor of computer science at Brown University. Raphael is the second author on this paper and a former postdoctoral researcher at UCSD.

To create InvChecker, the researchers modified an existing software system created at UCSD by Glenn Tesler, in order to make it better at detecting microinversions and differentiating microinversions from other genomic rearrangements. Such false positives are generally not useful in understanding the history of evolution and can introduce error to the reconstruction of phylogenetic trees.

With InvChecker, the researchers analyzed the CFTR region in a collection of mammal species. CFTR is a heavily studied and highly conserved, gene rich area of human chromosome 7 that is home to the cystic fibrosis gene.

“It’s quite a subtle problem to find microinversions. Our goal is to use these tiny inversions to develop a history of species,” said Pevzner.

The researchers also used InvChecker to study the specific differences between humans and chimpanzees. They found that 80 percent of the microinversions between humans and chimps that were proposed last year are, in fact, repeat-induced artifacts and not microinversions. The researchers also uncovered 167 human-chimp microinversions recently missed by scientists using software other than InvChecker.

“This finding doesn’t change the conclusions between humans and chimps, but it does say that the detection of microinversion needs to be done carefully,” said Chaisson. “InvChecker does a more careful job of comparing sequences than previous attempts to find microinversions.”

With InvChecker, you can take the same genomic region from two species sequences, partition them into regions that are unique to one species or common to both (orthologous), and find how the order of these regions relates between the two species.

“We are looking for orthologous sequences in reverse order that are surrounded by elements in forward order. That’s a microinversion,” Chaisson explained.

Microinversions have certain advantages over other evolutionary signals used for studying evolution such as amino acid changes, Chaisson explained. “With microinversions, it’s easy to develop evolutionary relationships between species and difficult to debate whether one species

is inverted relative to another species.”

With InvChecker and microinversions, researchers are not limited to comparing species that are evolutionarily close, as is the case when using other genomic features like repetitive sequences and deletions for phylogenetic analysis. The new process can also detect microinversions that are the result of convergent evolution and thus do not play a role in tracking evolution and defining phylogenies.

Once the researchers have the microinversion data, they use it to reconstruct phylogenies using an algorithm that attempts to move “back in time” by iteratively undoing microinversions and bringing the existing species closer to the ancestral mammalian genome.

Source: University of California - San Diego

Citation: Software, evolution and micro-inversions -- improving the building of phylogenetic trees (2006, December 19) retrieved 24 April 2024 from <https://phys.org/news/2006-12-software-evolution-micro-inversions-phylogenetic.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.