

# Genome comparison tools found to be susceptible to slip-ups

May 26 2010, by Hannah Hickey

---

(PhysOrg.com) -- You might call it comparing apples and oranges, but lining up different species' genomes is common practice in evolutionary research. Scientists can see how species have evolved, pinpoint which sections of DNA are similar between species, meaning they probably are crucial to the animals' survival, or sketch out evolutionary trees in places where the fossil record is spotty.

But the tools used to align genomes from different species have serious quality-control issues, according to a study published online this week in the journal [Nature Biotechnology](#).

"We discovered that there's a disturbingly low level of agreement between genome alignments produced by different tools," said corresponding author Martin Tompa, a UW professor of computer science and engineering and of genome sciences. "What this should suggest to biologists is that they should be very cautious about trusting these alignments in their entirety."

This is especially true when comparing distantly related species, and in regions of the genome that do not code for a protein, he said.

Aligning genomes, while simple in theory, is difficult in practice. Aligning more than two sequences becomes much harder with every additional sequence. At the scale of a mammal's entire genome, all of its [genetic code](#), finding the optimal alignment of many genomes is far beyond the capabilities of any computer, Tompa said.

Various software tools instead use strategic shortcuts.

"At a high level the tools are very similar," Tompa said. "They make different decisions at the lower, more detailed levels, and those decisions seem to have widespread effect on the outcome."

The new paper compared the alignments from a previous study in which four research teams each took the same 1 percent of the human genome and aligned it to the genomes of 27 other vertebrate animals, ranging from mouse to elephant.

"This is a marvelous dataset," Tompa said. "It's a very large-scale multiple sequence alignment, done by four expert teams using four different tools, all of them working on the same input sequences."

However, the new study found that the resulting alignments were quite different. The authors also compared the coverage of each tool, meaning how much of the human DNA it was able to match to each other species, as well as what fraction of alignments were suspiciously close to a random match.

The best-performing tool was the newest one, Pecan, developed by the European Bioinformatics Institute.

"Our study pretty clearly points to Pecan as being the highest-quality alignment of the four tools we compared," Tompa said. It aligned as much of the human [genome](#) to other species as any of the other tools, and its matches were considerably more reliable, especially between more distantly related species.

The other tools in the study were Threaded Blockset Aligner (or TBA), Multiple Limited Area Global Alignment of Nucleotides (or MLAGAN) and Mavid. All four are free programs developed by academic

institutions, Tompa said.

"I'm hoping that the designers of these tools will take a very close look at our paper and might be able to improve their tools as a result," he said. "I think we're all interested in having a better understanding of which methods work the best and how to make them better."

**More information:** [www.nature.com/nbt/journal/vao...nt/abs/nbt.1637.html](http://www.nature.com/nbt/journal/vao...nt/abs/nbt.1637.html)

Provided by University of Washington

Citation: Genome comparison tools found to be susceptible to slip-ups (2010, May 26) retrieved 24 April 2024 from <https://phys.org/news/2010-05-genome-comparison-tools-susceptible-slip-ups.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.