

# Solving puzzles without a picture: New algorithm assembles chromosomes from next generation sequencing data

January 10 2013

---

One of the most difficult problems in the field of genomics is assembling relatively short "reads" of DNA into complete chromosomes. In a new paper published in *Proceedings of the National Academy of Sciences* an interdisciplinary group of genome and computer scientists has solved this problem, creating an algorithm that can rapidly create "virtual chromosomes" with no prior information about how the genome is organized.

The powerful [DNA sequencing methods](#) developed about 15 years ago, known as next generation sequencing (NGS) technologies, create thousands of short fragments. In species whose genetics has already been extensively studied, existing information can be used to organize and order the NGS fragments, rather like using a sketch of the complete picture as a guide to a [jigsaw puzzle](#). But as [genome](#) scientists push into less-studied species, it becomes more difficult to finish the puzzle.

To solve this problem, a team led by Harris Lewin, distinguished professor of evolution and ecology and vice chancellor for research at the University of California, Davis and Jian Ma, assistant professor at the University of Illinois at Urbana-Champaign created a [computer algorithm](#) that uses the known chromosome organization of one or more known species and NGS information from a newly sequenced genome to create virtual chromosomes.

"We show for the first time that chromosomes can be assembled from NGS data without the aid of a preexisting genetic or physical map of the genome," Lewin said.

The new algorithm will be very useful for large-scale sequencing projects such as G10K, an effort to sequence 10,000 vertebrate genomes of which very few have a map, Lewin said.

"As we have shown previously, there is much to learn about phenotypic evolution from understanding how [chromosomes](#) are organized in one species relative to other species," he said.

The algorithm is called RACA (for reference-assisted chromosome assembly), co-developed by Jaebum Kim, now at Konkuk University, South Korea, and Denis Larkin of Aberystwyth University, Wales. Kim wrote the software tool which was evaluated using simulated data, standardized reference genome datasets as well as a primary NGS assembly of the newly sequenced Tibetan antelope genome generated by BGI (Shenzhen, China) in collaboration with Professor Ri-Li Ge at Qinghai University, China. Larkin led the experimental validation, in collaboration with scientists at BGI, proving that predictions of chromosome organization were highly accurate.

Ma said that the new RACA algorithm will perform even better as developing NGS technologies produce longer reads of DNA sequence.

"Even with what is expected from the newest generation of sequencers, complete chromosome assemblies will always be a difficult technical issue, especially for complex genomes. RACA predictions address this problem and can be incorporated into current NGS assembly pipelines," Ma said.

Provided by UC Davis

Citation: Solving puzzles without a picture: New algorithm assembles chromosomes from next generation sequencing data (2013, January 10) retrieved 11 May 2024 from <https://phys.org/news/2013-01-puzzles-picture-algorithm-chromosomes-sequencing.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.