

New genome alignment tool empowers large-scale studies of vertebrate evolution

November 11 2020



Comparative genomics sheds new light on the diversity of birds and other vertebrates. Credit: Jon Fjeldsa/Josefin Stiller/University of Copenhagen

Three papers published November 11 in *Nature* present major advances in understanding the evolution of birds and mammals, made possible by new methods for comparing the genomes of hundreds of species.

Comparative genomics uses genomic data to study the evolutionary relationships among species and to identify DNA sequences with essential functions conserved across many species. This approach requires an alignment of the genome sequences so that corresponding positions in different genomes can be compared, but that becomes increasingly difficult as the number of genomes grows.

Researchers at the UC Santa Cruz Genomics Institute developed a powerful new genome alignment method that has made the new studies possible, including the largest genome alignment ever achieved of more than 600 vertebrate genomes. The results provide a detailed view of how species are related to each other at the genetic level.

"We're literally lining up the DNA sequences to see the corresponding positions in each genome, so you can look at individual elements of the genome and see in great detail what has changed and what's stayed the same over evolutionary time," explained Benedict Paten, associate professor of biomolecular engineering at UC Santa Cruz and a corresponding author of two of the new papers.

Identifying DNA sequences that are conserved, remaining unchanged over millions of years of evolution, enables scientists to pinpoint elements of the genome that control important functions across a wide range of species. "It tells you something is important there—it hasn't changed because it can't—and now we can see that with higher resolution than ever before," Paten explained.

The previous generation of alignment tools relied on comparing everything to a single reference genome, resulting in a problem called "reference bias." Paten and coauthor Glenn Hickey originally developed a reference-free alignment program called Cactus, which was state-of-the-art at the time, but worked only on a small scale. UCSC graduate student Joel Armstrong (now at Google) then extended it to create a

powerful new program called Progressive Cactus, which can work for hundreds and even thousands of genomes.

"Most previous alignment methods were limited by reference bias, so if human is the reference, they could tell you a lot about the human genome's relationship to the mouse genome, and a lot about the human genome's relationship to the dog genome—but not very much about the mouse genome's relationship to the dog genome," Armstrong explained. "What we've done with Progressive Cactus is work out how to avoid the reference-bias limitation while remaining efficient enough and accurate enough to handle the massive scale of today's genome sequencing projects."

Armstrong is a lead author of all three papers, and first author of the paper that describes Progressive Cactus and presents the results from an alignment of 605 genomes representing hundreds of millions of years of vertebrate evolution. This unprecedented alignment combines two smaller alignments, one for 242 placental mammals and another for 363 birds. The other two papers focus separately on the mammal and bird [genome](#) alignments.

This international collaborative effort was coordinated by an organizing group led by coauthors Guojie Zhang at the University of Copenhagen and China National GeneBank, Elinor Karlsson at the Broad Institute of Harvard and MIT, and Paten at UCSC. The genomic data used in these analyses were generated by two broad consortia: the 10,000 Bird Genomes (B10K) project for avian genomes and the Zoonomia project for mammalian genomes.

Scientists have been making plans for years to sequence and analyze the genomes of tens of thousands of animals. Coauthor David Haussler, director of the UCSC Genomics Institute, helped initiate the Genome 10K project in 2009. Related efforts include the Vertebrate Genome

Project and the Earth BioGenome Project, and all of these projects are now gathering steam.

"These are very much forward-looking papers, because the methods we've developed will scale to alignments of thousands of genomes," Paten said. "As sequencing technology gets cheaper and faster, people are sequencing hundreds of new species, and this opens up new possibilities for understanding [evolutionary relationships](#) and the genetic underpinnings of biology. There is a colossal amount of information in these genomes."

More information: Progressive Cactus is a multiple-genome aligner for the thousand-genome era, *Nature* (2020). [DOI: 10.1038/s41586-020-2871-y](https://doi.org/10.1038/s41586-020-2871-y) , www.nature.com/articles/s41586-020-2871-y

Provided by University of California - Santa Cruz

Citation: New genome alignment tool empowers large-scale studies of vertebrate evolution (2020, November 11) retrieved 27 April 2024 from <https://phys.org/news/2020-11-genome-alignment-tool-empowers-large-scale.html>

<p>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.</p>
