

Technical advances in reading long DNA sequences have ramifications in understanding primate evolution, human disease

March 31 2016



A new assembly of the gorilla genome offers important biological insights into their evolution and how these primates differ from humans. Credit: Alice C. Gray

Technical advances in reading long DNA sequences have ramifications in understanding primate evolution and human disease.

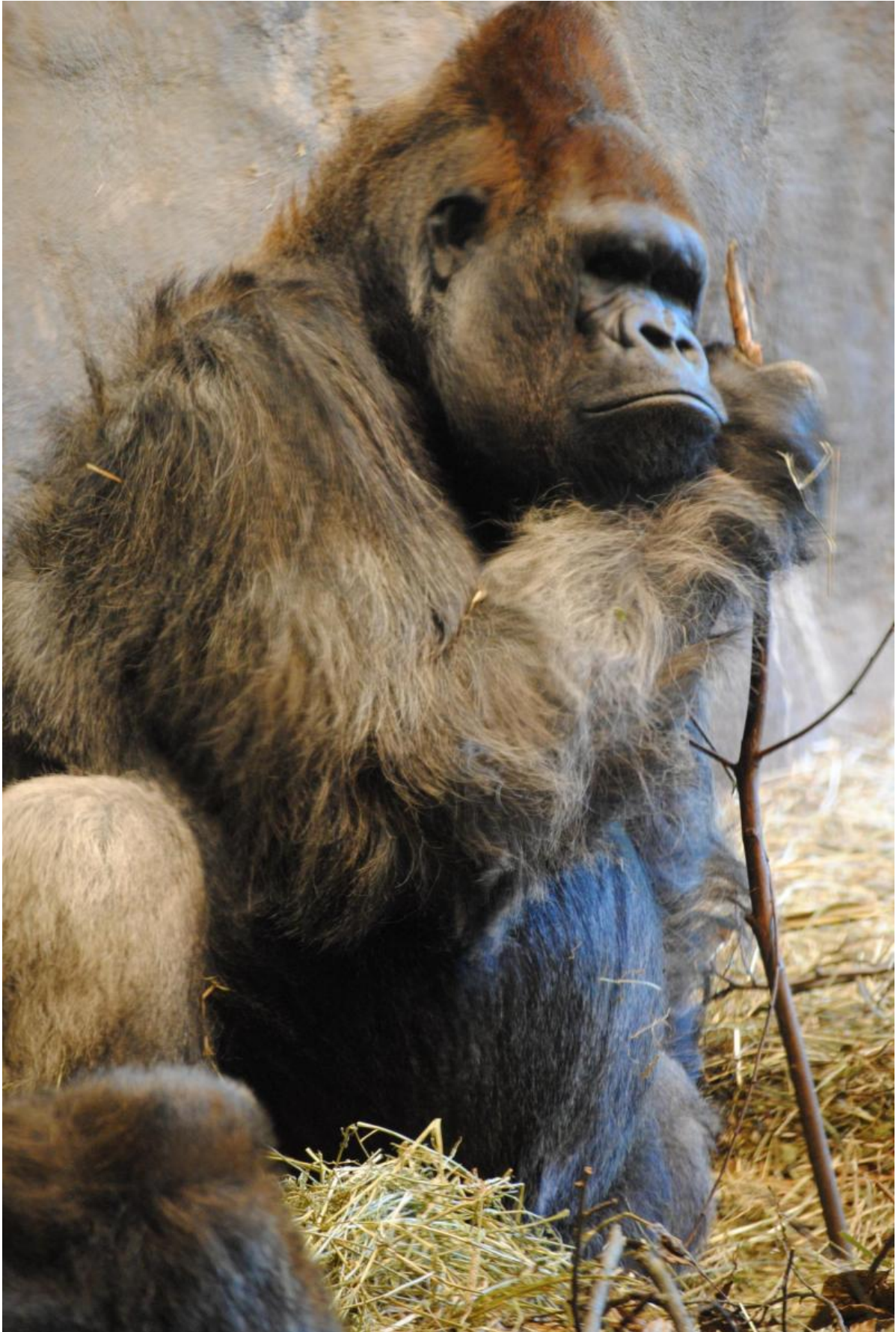
The [genome](#) of the Western lowland gorilla has now been sequenced and assembled at a high level of quality beginning to approach that of the mouse and human genome.

A new sequencing technology based on longer sequence reads allows missing genes and missing forms of genetic variation to be discovered for the first time. This assembly offers new biological insights into a living species that is second only to chimps in its evolutionary closeness to humans.

Reporting in the April 1 edition of *Science*, researchers led by Evan Eichler, University of Washington professor of genome sciences, explained why previous genome assemblies for the gorilla and other mammals have been fragmented, incomplete, and potentially misleading:

Massively parallel sequencing technologies, while increasing the speed, improving the accuracy, and reducing the cost of [genome sequencing](#), typically produce only short stretches of sequences called "reads" After sequencing, the reads are pieced back together with [genome assembly](#) software.

The program attempts to reconstruct the original genome by using the overlap between the sequence reads. Unfortunately, the presence of long, repetitive DNA, which is common in human and other primate genomes, confuses assembly software and causes it to break the genome into very small fragments.



A new assembly of the gorilla genome is offering new insights into differences between humans and this great ape. Above, a zoo gorilla sits and observes his surroundings. Credit: Alice C. Gray

"Such assemblies can be like Swiss cheese," Eichler said, "with a lot of missing biological information in the gaps." The original, published Western lowland gorilla genome, created using the short-read technology, he said, was broken into more than 400,000 pieces.

"These gaps are not random, but are clustered at sites of repeats," he said. "If geneticists can't capture these repeats and determine structural differences in genomes, they have problems understanding the organization of genes and comparing genetic variation within and across species."

His team included UW bioinformatics specialists David Gordon and John Huddleston, as well as postdoctoral fellows, Mark Chaisson, Chris Hill and Zev Kronenberg. The research team analyzed DNA in a blood sample of a female Western lowland gorilla from Chicago's Lincoln Park Zoo.

The researchers used Single Molecule, Real-Time (SMRT) sequencing technology, the assembly tools Falcon and QUIVER, and other techniques to generate long sequence reads. These were more than a hundred times the length of the most popular sequence technologies. The long reads allowed them to traverse most of the repeat regions of the gorilla genome during the assembly.

The result was a new gorilla genome assembly that was larger and had

far fewer pieces. Instead of 400,000 fragments, there are now only 1,800 pieces. The average size of the genome fragments was 800 times larger with approximately 90 percent of all gaps in the original assembly closed.

This additional sequencing information, the researchers observed, greatly improved gene annotation for that species of gorilla. It also led to the discovery of thousands of protein- and peptide-coding segments and new regulatory elements that had been missed as part of the first genome assembly.



A seated gorilla rests his leathery hand on his leg. Gorilla knuckle pads have a hard coating to protect them during walking. Credit: Alice C. Gray

Differences in how genes are controlled, or even the loss or disruption of certain gene regulatory elements, may explain why human ancestors evolved to be so different from their great ape relatives.

The scientists also found tens of thousands of new structural variants, such as deletions or insertions of DNA, that are likely to be more important than the smaller single base pair differences that were cataloged before. (Base pairs are the two chemicals that bond into a rung on the DNA ladder)

"My motivation in studying human and great ape genomes," Eichler said, "is to try to learn what makes us tick as a species. I'd like to see a re-doing of all the great ape genomes, including chimpanzee and orangutan, to get a comprehensive view of the genetic variants that distinguish humans from the great apes. I believe there is far more genetic variation than we had previously thought. The first step is finding it."

Among the areas where the researchers have seen intriguing dissimilarities between humans and gorillas are in genes associated with sensory perception, keratin (a skin protein) production, insulin regulation, immunity, reproduction and cell signaling.

The new genome assembly also provides new clues into the evolutionary history of the lowland gorilla. Prior studies have demonstrated that the gorilla population underwent a bottleneck in the not so distant past, but analyses with the new genome shows that the bottleneck was more severe than previously thought.

Patterns of genetic variation within the gorilla genome can provide evidence of how disease, climate change and human activity affect lowland gorilla populations.

"I think the take home message," Eichler said, "is that the new genome

technology and assembly bring us back to the place we should have been 10 years ago."

"Sequencing technology and computational biology," Eichler and his team wrote in their paper, "have now advanced to the stage where individual laboratories can generate high quality genomes of mammals. This capability has the promise to revolutionize our understanding of genome evolution and species biology."

Eichler added that these advances are also likely to contribute greatly to research on the genetic underpinnings of human disease, especially if more [human genomes](#) are sequenced in this way.

"As medical researchers, if we depend only on short read sequences, there's a chink in our armor. The work on gorilla and other human genomes clearly demonstrates that large swathes of genetic variation can't be understood with the short sequence-read approaches. Long read sequencing is allowing us to access a new levels of [genetic variation](#) that were previously inaccessible ,inaccessible," he said.

However, he added, "At \$80,000 a pop, the price is not yet right today for clinical sequencing of human genomes using the long reads. Given a few years of years of cost reduction and further advances in technology, I am willing to bet this is the way we will sequence human genomes to discover disease-causing mutations in the future. "

More information: "Long-read sequence assembly of the gorilla genome," *Science*, [DOI: 10.1126/science.aae0344](https://doi.org/10.1126/science.aae0344)

Provided by University of Washington

Citation: Technical advances in reading long DNA sequences have ramifications in understanding primate evolution, human disease (2016, March 31) retrieved 7 July 2024 from <https://phys.org/news/2016-03-technical-advances-dna-sequences-ramifications.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.