

Scientists build a 'Hubble Space Telescope' to study multiple genome sequences

December 2 2019



Credit: CC0 Public Domain

A new tool that simultaneously compares 1.4 million genetic sequences can classify how species are related to each other at far larger scales than previously possible. Described today in *Nature Biotechnology* by researchers from the Centre for Genomic Regulation in Barcelona, the technology can reconstruct how life has evolved over hundreds of millions of years and makes important inroads for the ambition to

understand the code of life for every living species on Earth.

Protecting Earth's biodiversity is one of the most urgent global challenges of our times. To steward the planet for all [life forms](#), humanity must understand the way animals, fungi, bacteria and other organisms have evolved and how they interact amongst millions of other [species](#). Sequencing the genome of life on Earth can unlock previously unknown secrets that yield fresh insights into the evolution of life, while bringing new foods, drugs and materials that pinpoint strategies for saving species at risk of extinction.

The most common way scientists study these relationships is by using Multiple Sequence Alignments (MSA), a tool that can be used to describe the [evolutionary relationships](#) of living organisms by looking for similarities and differences in their biological sequences, finding matches among seemingly unrelated sequences and predicting how a change at a specific point in a gene or protein might affect its function. The technology underpins so much biological research that the original study describing it is one of the most cited papers in history.

"We currently use multiple sequence alignments to understand the family tree of species evolution," says Cédric Notredame, a researcher at the Centre for Genomic Regulation in Barcelona and lead author of the study. "The bigger your MSA, the bigger the tree and the deeper we dig into the past and find how species appeared and separated from each other.

"What we've made lets us dig ten times deeper than what we've been able to do before, helping us to see hundreds of millions of years into the past. Our technology is essentially a time machine that tells us how ancient constraints influenced genes in a way that resulted in life as we know today, much like how the Hubble Space Telescope observes things that happened millions of years ago to help us understand the Universe

we live in today."

Researchers can use MSA to understand how certain species of plants have evolved to be more resistant to climate change, or how particular genetic mutations in one species makes them vulnerable to extinction. By studying a living organism's evolutionary history, scientists may come up with and test new ideas to stave off the collapse of entire ecosystems.

Technological advances have made sequencing cheaper than ever before, resulting in increasingly large datasets with more than a million sequences for scientists to analyze. Some ambitious endeavors, like the Earth BioGenome Project, may run to the tens of millions. Researchers have not been able to take full advantage of these enormous datasets because current MSAs cannot analyze more than 100,000 sequences with accuracy.

To evaluate the scale-up potential of MSA, the authors of the paper used Nextflow, a cloud-computing software developed in-house at the Centre for Genomic Regulation. "We spent hundreds of thousands of hours of computation to test our algorithm's effectiveness," says Evan Floden, a researcher at the CRG who also led on developing the tool. "My hope is that in combining high-throughput instrumentation readouts with high-throughput computation, science will usher in an era of vastly improved biological understanding, ultimately leading to better outcomes for consumers, patients and our planet as a whole."

"There is a vast amount of '[dark matter](#)' in biology, code we have yet to identify in the unexplored parts of the genome that is untapped potential for new medicines and other benefits we can't fathom," concludes Cédric. "Even seemingly inconsequential organisms may play a pivotal role in furthering human health and that of our planet, such as the discovery of CRISPR in archaea. What we have built is a new way of finding the needles in the haystack of life's genomes."

More information: Large multiple sequence alignments with a root-to-leaf regressive method, *Nature Biotechnology*, [DOI: 10.1038/s41587-019-0333-6](https://doi.org/10.1038/s41587-019-0333-6) , [nature.com/articles/s41587-019-0333-6](https://www.nature.com/articles/s41587-019-0333-6)

Provided by Center for Genomic Regulation

Citation: Scientists build a 'Hubble Space Telescope' to study multiple genome sequences (2019, December 2) retrieved 27 April 2024 from <https://phys.org/news/2019-12-scientists-hubble-space-telescope-multiple.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.