

Scientists explore mysteries behind diversity of DNA composition among species

January 3 2018



A depiction of the double helical structure of DNA. Its four coding units (A, T, C, G) are color-coded in pink, orange, purple and yellow. Credit: NHGRI

To make the iconic, twisted double helix that accounts for the diversity of life, DNA rules specify that G always pairs with C, and A with T.

But, when it's all added up, the amount of G+C vs A+T content among species is not a simple fixed percentage or, standard one-to-one ratio.

For example, within single-celled organisms, the amount of G+C content can vary from 72 percent in a bacteria like *Streptomyces coelicolor* while the protozoan parasite that causes malaria, *Plasmodium falciparum*, has as little as 20 percent.

In single-celled eukaryotes, yeast contain 38 percent G+C content, plants like corn have 47 percent, and humans contain about 41 percent.

The big question is, why?

"This has been one of the long-standing problems in genome evolution, and prior attempts to explain it have involved considerable arm waving," said Michael Lynch, who leads a new Center for Mechanisms of Evolution at Arizona State University's Biodesign Institute.

Is there something within the chemical nature of DNA itself that favors one nucleotide over the other, or does the bias of mutation pressure vary, and if so, why would this be different among species?

"In the absence of key observations on the mutation process, there has been a struggle to fathom what the mechanism is," said Lynch.

Michael Lynch's group has now experimentally demonstrated that G+C composition is generally strongly favored, whereas this is often opposed by mutational pressure of various strengths in the opposite direction.

"On average, natural selection or some other factor (possibly associated

with recombinational forces) favors G+C content, regardless of the class of DNA, size of a species' genome, or where the species is found on the evolutionary tree of life," said Lynch.

The study was published in the journal *Nature Ecology and Evolution*.

To err is universal

Driving evolution are DNA [mutations](#), errors in the genome that are introduced and passed along to the next generation, so that over time, providing the fuel for the invention of new adaptations or traits.

To get to the heart of the matter, the scientists wanted a way to quantify the full spectrum of DNA mutations in the lab across a wide swath of species.

This can now be done due in part by new technologies to make DNA sequencing faster and cheaper. It has fueled a golden age of evolutionary experimental biology.

"We started with knowledge of the mutational spectrum that occurs at the genome level in about 40 species examined in my lab," said Lynch. "You can use such information to calculate what the GC composition would be in the absence of selection. And then we can compare this null expectation with the the actual genome content, the difference being due to selection."

In a tour de force experiment that is the largest survey to date, they examined every single DNA mutation across different species, sequencing billions of DNA chemical bases.

"This represented a very substantial work load, effort and cost that was necessary to test different evolutionary models with high statistical

power," said Hongan Long, a postdoctoral researcher who led the experiments.

They also took advantage of an analysis of 25 current datasets of mutations and 12 new mutation-accumulation (MA) experiments (many from their own lab), including bacteria and a menagerie of multicellular organisms including yeast, worms, fruit flies, chimpanzees and humans.

During each MA experiment, they performed complete genome sequencing of about 50 different bacterial lines that had been passaged through severe, single-cell bottlenecks for thousands of cell divisions.

"This single-cell passage of each line acts like a filter, eliminating the ability of natural selection to modify the accumulation of all but the most severe and deleterious mutations, giving us an effectively unbiased view of the mutation process," said Long.

With each generation, they carefully measured the mutation rate, or every occurrence of when just a single DNA letter is changed.

This can happen in two ways: a single G or C DNA base pair being converted to the A+T direction; or the opposite can happen, with an A or T base switching in the G+C direction.

After all the number and data crunching, a striking pattern emerged between G+C content and the expectations based on DNA mutations.

"It turns out, they are correlated," said Lynch.. "The G+C composition is always higher than you expect, based on neutrality. That tells us that there is pervasive selection. So mutation drives the overall pattern, but selection for G's and C's over A's and T's boosts the genome content above the neutral mutational expectation.

This seems to be almost universally true."

The end of the beginning

Now that they've shown the G+C composition correlation, it has opened up the door to many more questions, and answers that remain elusive.

"One question is, 'why does the mutation spectrum change so dramatically across species'?" asked Lynch. "Species don't have the same mutation spectrum. There are species whose mutation profiles are more AT rich and others more GC rich. We still don't know the mechanisms behind such divergence in the mutational spectrum."

They may be due to simple differences in chemistry and biophysics.

One general force that may be of relevance is DNA stability, driven by the chemistry of the DNA letters. The forces that keep the DNA ladder intact are called hydrogen bonds. G:C pairs involve three hydrogen bonds, whereas, A:T pairs involve only two.

"The prevailing thought is that more G:C content adds to genome stability," said Lynch.

Another possibility is during reproduction, when the DNA strands intertwine from each parent to make a fertilized egg, mismatches can occur in the base pairing, leading to mistakes that DNA proofreading enzymes have to fix later on. Sometimes, a G can get changed to an A, or a T becomes a C, converting genes during this mismatch repair process.

"That's generally thought to be biased towards Gs and Cs," said Lynch.

Now, with their experimental setup in place, Lynch's team is poised to

further explore the mechanisms of evolution and fundamental forces behind this great mystery.

More information: Hongan Long et al, Evolutionary determinants of genome-wide nucleotide composition, *Nature Ecology & Evolution* (2017). [DOI: 10.1038/s41559-017-0425-y](https://doi.org/10.1038/s41559-017-0425-y)

Provided by Arizona State University

Citation: Scientists explore mysteries behind diversity of DNA composition among species (2018, January 3) retrieved 8 May 2024 from <https://phys.org/news/2018-01-scientists-explore-mysteries-diversity-dna.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.