

New genes are more likely than expected to emerge full-fledged from a genome's noncoding regions

April 26 2017, by Daniel Stolte



First, do no harm: For a new gene to be "born," it must have the structure of an "uber gene," to avoid its protein product from accumulating and harming the cell. Credit: University of Arizona

New genes are more likely to appear on the stage of evolution in fullfledged form rather than gradually take shape through successive stages of "proto genes" that become more and more refined over generations.



This is the surprising upshot from research led by Benjamin Wilson and Joanna Masel at the University of Arizona, published as an Advance Online Publication by the scientific journal *Nature Ecology & Evolution* on April 24.

Evolutionary biologists have long pored over the question of where new genes come from, which poses something of a chicken-and-egg problem. Conventional wisdom has it that new genes—DNA sequences that code for a protein molecule—evolve from existing genes through duplication and divergence. This happens when DNA copying mechanisms accidentally leave behind an extra copy of a particular gene. Naturally occurring mutations subsequently introduce changes that alter the DNA sequence such that the new gene assumes a function previously not found in the organism's lineage.

Previous studies by other researchers suggested that new genes also emerge from non-coding DNA sequences, via primitive "proto-genes" that become refined over generations, resulting in an "adult," fully functional gene.

Masel and her team found the opposite to be more likely, based on the fact that non-coding DNA sequences are likely to give rise to highly ordered proteins. Proteins, which consist of <u>amino acids</u> chained together into so-called polypeptides, tend to fold into three-dimensional structures that range from simple to mindbogglingly complicated. And while "ordered" may sound like a good thing, Masel is quick to point out that a healthy dose of disorder is key to success when it comes to evolution coming up with new genes that serve as blueprints for new proteins.

For the study, the researchers compiled data on full-genome DNA sequences downloaded from yeast and mouse databases.



"We take all the known mouse genes and yeast genes and query them against everything that's ever been sequenced and see what they're related to," explains Masel, a professor in the Department of Ecology and Evolution and a member of the UA's BIO5 Institute, "and based on that, we assign each gene an age that tells us when it was born."

In the next step, the team used statistical analyses to create a model revealing the average degree of order that would be present in each gene's product.

"We found that the youngest genes are the least ordered of all, which is what you would expect to get if you birthed a gene," Masel says.

The key to a protein that can contribute a useful function for its organism while not harming it is a healthy mix between regions that are soluble because they consist of hydrophilic, or "water-loving," amino acids and stretches that are insoluble because of their hydrophobic, or "water-repelling," amino acids.

If a protein consists of too many water-loving amino acids, it will remain largely unfolded, floating around inside the cell as an unorganized chain incapable of performing biological tasks. If too much of its length is water-repelling, the amino acids will clump together, rendering the protein unusable, and even dangerous, because when such misfolded proteins bump into each other, they tend to stick to each other and accumulate.

"Now think about the most highly ordered proteins we know—amyloids," Masel says, referring to the infamous piles of proteins found in the brain of Alzheimer's patients. "Because of this, the first order of business for any prospective gene is: 'Do no harm. Do not misfold.'"



This has profound implications for the evolution of new genes from noncoding DNA sequences. Because such sequences are likely to give rise to highly ordered proteins, they are likely to be deleterious to the organism. In this scenario, any prospective new gene must start out as some kind of "super gene," in contrast to a "proto gene." Rather than making its debut in the gene pool as an unrefined gene that still bears many similarities to the non-coding DNA sequences it came from, the protein it encodes must start with a higher-than average degree of disorder to prove itself before evolution would allow it becoming a permanent member of the gene pool.

"Instead of gradually working up to having more hydrophilic regions, young genes work their way down from being more hydrophilic and disordered, to more hydrophobic regions," Masel says. "In other words, when it comes to structural disorder, a polypeptide has the highest chance of being born if it is 'extra gene-like,' rather than 'sort of genelike.'"

The probability that a gene could arise from a random, non-coding sequence—also known as "junk DNA," on the other hand, used to be considered negligible, based on the premise that in the vast majority of cases, a random sequence does more harm than good. This may not be so, argues a second paper in the same issue by Rafik Neme, one of the co-authors of the study discussed here. Neme, currently a postdoctoral researcher at Columbia University Medical Center in New York, found the first experimental evidence that non-coding, "silent" stretches of DNA are anything but that.

"Until now, nobody knew whether a randomly sequence could immediately have any effect that would result in a function, or whether function was slowly acquired over time," Neme says. "It's similar to the idea of having a monkey typewriting at random, and expecting it to produce meaningful work."



Neme's experiments show that many sequences exhibit relevant activities immediately, some good and some bad. This, in turn, suggests a discrete transition between non-genes and genes and would favor certain kind of sequences and functions over others.

Based on their findings, Neme and Masel point out, the pool from which genes are born might be more conducive to birthing new genes than one might expect.

"In our scenario, a gene precursor would be a transcript that happened to be translated into a protein sometimes but has no function," she says. "These things come up in evolution all the time, and mutation will quickly destroy it unless that polypeptide provides the organism with some advantage. There either is an advantage that natural selection can act on, or there isn't, so we don't think the would-be genes stick around for very long."

This in turn suggests that gene birth is a sudden transition, rather than a gradual process involving many intermediate steps.

More information: Benjamin A. Wilson et al. Young genes are highly disordered as predicted by the preadaptation hypothesis of de novo gene birth, *Nature Ecology & Evolution* (2017). DOI: 10.1038/s41559-017-0146

Provided by University of Arizona

Citation: New genes are more likely than expected to emerge full-fledged from a genome's noncoding regions (2017, April 26) retrieved 27 April 2024 from <u>https://phys.org/news/2017-04-genes-emerge-full-fledged-genome-non-coding.html</u>



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.