BACTERIA THAT WRITE NEW GENES

Variability is fundamental to evolving (= surviving). It is provided by mutations and, in almost all eukaryotes, by sexual reproduction that generates infinite new gene combinations.

New genes can emerge almost out of nowhere, but such examples are extremely rare. The production of new genes generally occurs through gene duplication and independent evolution. In some animals (such as elephants) duplication of the olfactory receptor genes has reached a number of ~2,000.

When significantly higher variability is needed, different and exclusively somatic strategies are exploited. An example is the generation of millions and millions of different antibodies by complex somatic recombination. Such a large number of genes would be too heavy to store in a germline genome.

Bacteria use several defense mechanisms to fight invaders (essentially phages). Restriction enzymes and the adaptive CRISPR-Cas systems are among them.

In a paper in the bioRxiv (1), Tang et al. report an unprecedented workflow some bacteria use to create new defensive genes. The title of the paper reads: “De novo gene synthesis by an antiviral reverse transcriptase”; the sub-title of the comment to this paper which appeared in Nature (2) is: “Bacterial defensive systems scramble the standard workflow of life”.

RNA viruses use the reverse transcriptase (RT) to convert their RNA genomes into DNA, enabling them to integrate into the genome of the host cell and replicate.

Tang et al. found that in some bacteria (*Klebsiella pneumoniae* was the one that was studied) an unprecedented RT produces, from a non-coding RNA, a very long in-frame cDNAs consisting of repeated, concatenated sequences with precise junction sequences, without a stop codon.

This newly created gene has been named Neo gene (**N**ever **e**nding **o**pen reading frame). The absence of a stop codon explains how the Neo gene has always gone unnoticed (annotators should annotate this detail).

Upon phage infection, this cDNA becomes double-stranded and triggers the expression of a novel protein, which induces cell dormancy, thus protecting the larger bacterial population from the spread of the phage.

This work “adds another layer of complexity to the ways in which protein-coding sequences can be stored in the genome”.

The identities and functions of the Neo products remain largely unknown.

The last sentence of Darwin’s The Origin of Species says "… **from so simple a beginning, endless forms most beautiful and most wonderful have been, and are being, evolved”**

1. https://www.biorxiv.org/content/10.1101/2024.05.08.593200v1
2. https://www.nature.com/articles/d41586-024-01477-8