SMBE 2017

Molecular innovation
OT-MI14
De novo gene evolution: How do we transition from non-coding to coding?
Jorge Ruiz-Orera , José Luis Villanueva-Cañas , William R. Blevins , Mar Albà *

Abstract:
De novo gene evolution: How do we transition from non-coding to coding?

Jorge Ruiz-Orera[1] , José Luis Villanueva-Cañas[2] , William R. Blevins[1] , M.Mar Albà[1,3]

[1]Evolutionary Genomics Group, Research Programme on Biomedical Informatics, Hospital del Mar Research Institute, Universitat Pompeu Fabra, Barcelona. [2]Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona. [3]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona.

Recent years have witnessed the discovery of protein–coding genes which appear to have evolved de novo from previously non-coding sequences. This has changed the long-standing view that coding sequences can only evolve from other coding sequences. However, there are still many open questions regarding how new protein-coding sequences can arise from non-genic DNA.

Two prerequisites for the birth of a new functional protein-coding gene are that the corresponding DNA fragment is transcribed and that it is also translated. Transcription is known to be pervasive in the genome, producing a large number of transcripts that do not correspond to conserved protein-coding genes, and which are usually annotated as long non-coding RNAs (lncRNA). Recently, sequencing of ribosome protected fragments (Ribo-Seq) has provided evidence that many of these transcripts actually translate small proteins. We have used mouse non-synonymous and synonymous variation data to estimate the strength of purifying selection acting on the translated open reading frames (ORFs). Whereas a subset of the lncRNAs are likely to actually be true protein-coding genes (and thus previously misclassified), the bulk of lncRNAs code for proteins which show variation patterns consistent with neutral evolution. We also show that the ORFs that have a more favorable, coding-like, sequence composition are more likely to be translated than other ORFs in lncRNAs. This study provides strong evidence that there is a large and ever-changing reservoir of lowly abundant proteins; some of these peptides may become useful and act as seeds for de novo gene evolution.

Disclosure of Interest : None Declared

Keywords : None