A large number of young de novo genes were identified with discernible recent ancestral non-coding sequences and evidence of translation, using high-quality genome sequences of 13 closely related Oryza species and targeted proteomics.
Publishing this work, as it stands today, took a long journey. Since submitting an early version elsewhere for the first time in March 2016, three years had elapsed before we were informed of an acceptance decision from *Nature Ecology & Evolution* in January. With each review cycle, anonymous reviewers presented new challenges, and their insightful comments encouraged us to design new experiments and perform new analyses to improve our manuscript. Remarkably, across the board, all reviewers expressed enthusiasm about the scientific problems and basic observations we presented but required more support for our conclusions, such as those regarding the role of positive selection and the evidence for translation.

The two major problems we tackled in our article have hindered understanding about how a novel protein is created *de novo*. Firstly, a vast majority of previously reported *de novo* genes actually lacked key evidence of ancestral noncoding sequences to distinguish them from orphan genes that have no detectable homologues in other lineages, which can come about via several alternative molecular evolutionary processes. These processes include fast evolution of homologous genes, loss of homologues in other lineages, rapid amplification of short peptides from microsatellites into simple proteins, and horizontal
gene transfer from fast-evolving donors \(^1\,^2\) (e.g., bacteria evolve so rapidly that they would often lose the detectable homologue of a transferred gene in a short period of time). \textit{De novo} origination is just an alternative source of orphan genes, but it is critical for understanding origination of these novel proteins, because it is a mechanism to generate a new protein-coding gene out of noncoding sequences, which was not accepted conventionally. Secondly, most reported \textit{de novo} genes were rarely or not all supported by translation evidence, which made it difficult to distinguish between protein-coding genes or noncoding genes or simply functionless pseudogenes.

These seemingly endless but often constructive criticisms came from reviewers, who also apparently worked hard and provided extensive reviews (some writing more than 10 pages of comments!). Their responses often made us feel uneasy at first, but after reading them carefully, we realized that they were often well thought-out. Following their suggestions, we have considerably improved both science and our presentation of it. What we present in this publication benefited from all these critical reviews, especially the four anonymous reviewers from \textit{Nature Ecology & Evolution}, who showed excellent expertise in the scientific and technical issues related to this study.

Words are not enough to describe the excellent teamwork accomplished by the team. Composed of multiple research groups with distinct expertise, this cross-disciplinary team has put together a piece of outstanding work described in this article. Each group made
unique contributions to solving the important scientific question – are the new genes really created through \textit{de novo} origination? Are they indeed translated into proteins? Seven years ago, my group joined an international consortium organized by collaborator Rod Wing with the goal of creating high-quality genome sequences for 10 \textit{Oryza} species, along with a closely related outgroup species\textsuperscript{3}. The efforts of Rod Wing and other colleagues in the consortium laid down a sound basis for subsequent analyses of new gene evolution.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{Li Zhang (left) identified a large number of \textit{de novo} genes that recently originated in ancestral stages of the focal species, \textit{O. sativa} subspecies \textit{japonica}, and Yan Ren (right) contributed significantly to the finding of a large number of peptides translated from a high proportion of these genes.}
\end{figure}
In particular, I want to praise the perseverance exhibited by two first co-authors, Li Zhang and Yan Ren (Figure 1). Their tireless efforts allowed them to hunt for signals of ancestral processes that shaped de novo genes stepwise, providing critical evidence that the new genes are really de novo genes that evolved from noncoding ancestors and are translated to novel proteins. Five years ago, Li Zhang showed me a surprising result – he had identified a number of de novo genes with discernable ancestral sequences identified in the genome of the focal species, *Oryza sativa ssp japonica*, by comparing it to other *Oryza* and outgroup species. I was astonished by the clear signals indicating that rapid evolution had occurred from their ancestral noncoding sequences to characteristic coding sequences. We predicted that they are translated into novel proteins but did not expect several years of work awaiting us. We were lucky that the experienced team led by Siqi Liu was excited by our observations and bravely joined us in pursuing evidence that they are indeed de novo formed translatable genes. He, with Yan Ren and others, designed protocols specifically working on de novo genes and successfully conducted targeted proteomics experiments. Their powerful results have proved our prediction and thus led us to say farewell to a conventional scavenger approach, which has low efficiency and certainty in finding translational evidence for a de novo gene from often unsatisfactory public databases.

I would also like to share that growing *Oryza* plants in the field in Hainan, an island close to the equator, is not an easy job. This is a place, I observed, where one can boil a chicken-
egg in the sand near the paddy rice field under sun at noon. It is Yidan Ouyang and her team members that took on the unusual challenge to grow the plants in the hot season of 2017 and dissect many tissues for proteomics analysis. For example, they collected 60,000 tiny anthers (<0.5 mm) for ~1 gram of sample, needed for running proteomics analyses. They also ran into trouble trying to deliver the samples to the relatively distant city of Shenzhen, where the team of Siqi Liu was waiting to feed them into their mass-spectrum machines. The airlines refused to transport nitrogen-frozen biological samples. They solved the problem by relocating plants for tissue dissection, allowing the use of a different delivery tool -- I have to omit the details here, as they simply did far more than I can describe in this short blog post (Figure 2).
This article is not the end of the beginning, far from the beginning of the end, but just the beginning of the beginning. Several sets of new problems can be raised to pursue. Firstly, a new standard has been set to define a de novo gene, in order to account for the origination of typical protein-coding genes in eukaryotes. It requires evidence of an ancestral sequence to distinguish from alternative mechanisms that can also create orphan genes. It also asks for evidence for protein translation that can distinguish de novo non-translational RNA genes and pinpoint the process of a protein being generated from
noncoding ancestor. Secondly, with the new standard set, we can ask and explore many questions. How often do de novo genes originate in various organisms? What proportion of orphan genes are de novo genes? What are evolutionary forces involved in the evolution of de novo genes? Another set of questions can be explored now. What do structures of de novo proteins look like? Do they show any distinctive features in comparison to regular gene proteins? Do the de novo genes explain evolution of protein structure? More generally, what are functions of de novo genes from across molecular, genetic, and phenotypic levels? The variety of potential inquiries regarding this topic has been wildly expanded in comparison with what could be asked and explored a dozen years ago, when the concept of de novo origination was initially proposed and discussed4-5.


Acknowledgment: Thanks to David Kudrna for preparing the poster image of *O. sativa* subspecies *japonica*.

Manyuan Long
Edna K Papazian Distinguished Service Professor, The University of Chicago

I joined the Chicago faculty in November 1997 in the Department of Ecology and Evolution and the College, after postdoctoral researches at Harvard with Walter Gilbert and Richard Lewontin. I received Ph.D. in December 1992 at the University of California at Davis in the laboratory of Charles Langley. A serendipitous finding of jingwei in African Drosophila, the first new gene ever known that was generated by retroposition and exon shuffling, in my doctoral research ignited my passion in a new and big problem: how does a new gene originate and evolve? I have been working on this problem since my years at Davis and found more and more interesting and challenging with the new gene problem. After years indulged in the rates, patterns and mechanisms of new gene evolution, I am attracted to the...
phenotypic evolution and expression network evolution driven by new genes. I am examining the evolutionary forces underlying new gene evolution and related phenotypic evolution, including evolutionary force 1 -- the natural selection for adaptation and evolutionary force 2 -- the sexual selection for reproductive success, proposed by Charles Darwin, and other forces such as the neutrality and the conflict in evolution of genes and genomes.

1 Comments

Yidan Ouyang
16 days ago

A charming journey, a team-work collaboration, and a fascinating work!

Reply  Share

You may also be interested in...