Genome **Biology**

RESEARCH HIGHLIGHT

Why rodent pseudogenes refuse to retire

Manyuan Long* and Li Zhang

See related research article http://genomebiology.com/2012/13/11/R102

Abstract

A new study in this issue of *Genome Biology* sheds light on why some pseudogenes persist in rodent, and other mammalian, genomes.

A remarkable difference between mammalian genomes and those of other organisms, such as *Drosophila* or mosquito, is the high prevalence of pseudogenes, which are genetic elements that have lost protein-coding or transcriptional potential by accumulating disruptive mutations. For example, many hemoglobin homologs in the human genome have been shown to be pseudogenes. Irrespective of whether you have a taste for wine or not, your genome will harbor at least one alcohol dehydrogenase homolog that does not encode any dehydrogenase, but is instead a pseudogene [1]. In fact, scattered in our genomes there are as many as 20,000 pseudogenes, two-fifths having been created by RNA-based duplication [2,3]. Given that by definition these pseudogenes do not produce competent proteins, they might be expected to accept the evolutionary fate of shrinking by negative selection, which would ultimately result in a complete disappearance from the genome [4]. It is curious, therefore, that many mammalian pseudogenes have persisted for a long time in evolutionary terms, seemingly shirking evolution's purging of genomic waste. Sporadic findings of protein-coding function for some annotated pseudogenes have suggested that some of the large number of mammalian pseudogenes could be accounted for by mis-identification (see, for example, [3,5,6]). But the remaining large number of apparently genuine pseudogenes means that the problem remains unanswered: why do pseudogenes persist in mammalian genomes? In a new article published in *Genome Biology* [7], Marques and colleagues seek to answer this key question of mammalian genome evolution.

*Correspondence: mlong@uchicago.edu
Department of Ecology and Evolution, The University of Chicago,
1101 E. 57th Street, Chicago, IL 60637, USA

BioMed Central

Competing hypotheses

An obvious hypothesis to explain the pseudogene puzzle might be that the death and disappearance of pseudogenes are balanced with the arrival of new pseudogenes that are created by new mutations, which would sustain the large number seen in the genome. But this hypothesis is contradicted by observations of many old pseudogenes that have stayed for a long time in the genome, suggesting that the primary driver of the mammalian genome's high load of pseudogenes is not novel pseudogenization. An alternative hypothesis might be that mammalian pseudogenes have a long half-life. Indeed, with a half-life of 884 million years, they survive much longer than the paltry 20 million year total lifespan measured for *Drosophila melanogaster* pseudogenes [8]. But a final possibility is that mammalian genomes are enriched for pseudogenes because some of these pseudogenes are functional, and so protected from evolutionary erosion by selective pressure. However, while some previous individual cases of pseudogene function in gene regulation have been reported [9], a general mechanistic regulatory role is yet to be discovered.

The last of these hypotheses is given weight by the study of Marques and colleagues [7] that provides evidence that many transcribed rodent pseudogenes produce functional non-coding RNAs. The first clue of residual function was found by analyzing the expression profiles of rodent-specific unitary pseudogenes, with the observation that at least 65% are expressed. Unitary pseudogenes are of particular interest because they are not associated with gene duplication events, and so any functional role for unitary pseudogene transcripts is less likely to be made expendable by increased transcript dosage.

To investigate what the function of these expressed unitary pseudogenes might be, Marques and colleagues turned to recent work on competitive endogenous RNAs (ceRNAs) [10], which act as decoys, or sponges, for microRNAs (miRNAs). Transcripts that share miRNA-binding sites (also known as miRNA response elements (MREs)) with ceRNAs experience a decrease in miRNA-mediated suppression and hence an increase in gene expression. Marques and colleagues formulated a cellular regulatory model in which pseudogene transcripts might be a major player, based on the cross-regulation of target RNAs by pseudogene ceRNAs with common MREs.

Deconstructing the decoy

A putative ceRNA function for unitary pseudogenes is in fact far more straightforward to study than for protein-coding RNAs, where the function of a transcript's protein product can be difficult to separate from ceRNA-mediated regulation. Similarly, as protein-coding genes that behave as ceRNAs have separate functions at the transcript and protein levels, an evolutionary advantage gained from the loss of protein function may not apply to the loss of transcriptional potential. Therefore, while pseudogenization may occur when there is a negative selective pressure on a transcript's protein product, the transcript's ceRNA function may remain under positive selection, and so retain both its expression and MRE sequences. The separation of protein-coding and ceRNA function make unitary pseudogenes an excellent model for studying how transcripts might act as miRNA decoys, and how the sequestration of miRNAs serves as a regulation mechanism for target MREs and their host transcripts.

The miRNA decoy hypothesis predicts that, were a gene to retain its ceRNA behavior following pseudo-genization, conservation of the ancestral protein-coding gene's expression pattern would be observed. One way to infer such conservation is to compare the expression patterns of a pseudogene and orthologous protein-coding genes in related species. With this prediction in mind, Marques and colleagues designed an elegant system to test the miRNA decoy hypothesis by searching for unitary pseudogenes created in a lineage-specific manner, enabling a comparison between homologous transcripts with or without a protein-coding function [7].

A comparative analysis of human, mouse and rat identified 48 rodent-specific unitary pseudogenes, by searching for disruptive mutations (frame-shifting indels or premature stop codons) specific to mouse-rat in regions exhibiting syntenic conservation with the human genome [7]. The dog genome was used as a non-pseudo-genized outgroup, providing confirmation that protein-coding function had been lost in the mouse-rat lineage. As mentioned previously, 65% of these rodent-specific pseudogenes were found to be expressed, as determined by examining two RNA-seq datasets encompassing 6 and 19 tissue/cell types. One-third of the 48 pseudogenes were expressed in both datasets; a further analysis of these cross-validated transcribed pseudogenes supported the loss of protein-coding function. Although transcribed unitary pseudogenes tended to be expressed at a lower level than protein-coding genes, the tissue distribution profile was not significantly different between the two sets of transcripts.

With these data in hand, Marques and colleagues set out to test the conservation of expression between rodent unitary pseudogenes and their human protein-coding orthologs, and found the correlation between pseudogene-protein-coding ortholog pairs to be far stronger than between randomly selected pairs of protein-coding transcripts [7]. They then tested the conservation of gene expression networks in which unitary pseudogenes might interact, as ceRNAs, with other genes. The ceRNA hypothesis predicts that the expression levels of target transcripts will be positively correlated with their regulatory ceRNAs [10]. Marques and colleagues reasoned that such a pattern of conserved positive correlation would be apparent if rodent unitary pseudogenes had indeed retained the ceRNA function of their protein-coding ancestors. An analysis of the mouse RNA-seq data identified pseudogene-protein-coding gene pairs that were positively correlated. The human orthologs of the identified pairs were significantly enriched for positively over negatively correlated expression, suggesting that some of the pseudogene transcripts may indeed serve as ceRNAs. Further support for a ceRNA function came from the observation that mouse-rat unitary pseudogenes tended to preserve the cognate MREs of their human orthologs.

The aforementioned computational analyses of expression and sequence data provided several lines of strong statistical evidence for a hypothesis in which a ceRNA function drives the conservation of rodent pseudogenes; but building a highly mechanistic model from statistical tests alone is perhaps not the most convincing way to arrive at biological insights. Marques and colleagues therefore used the statistical conclusions not as an endpoint, but as a basis for testing their hypothesis with laboratory bench experiments. They chose *Pbcas4*, a mouse pseudogene that has a well-characterized, high, ubiquitous expression pattern in adult mouse, and the N2a neuroblastoma cell line as a model. Knocking down *Pbcas4* resulted in the differential expression of 165 genes, most of which were downregulated, as would be expected for ceRNA targets upon ceRNA knockdown [7]. In further support of a conserved ceRNA function, the authors noted that the human orthologs of these putative *Pbcas4* cross-regulated targets were statistically significantly enriched and showed positively correlated expression with the protein-coding human ortholog of *Pbcas4*. Furthermore, the functional role of an MRE in the *Pbcas4* transcript was detected by transfection of the corresponding miRNA (miR-185) in both mouse and human neuroblastoma cells. As would be expected for upregulation of a miRNA that mediates ceRNA regulation, miR-185 significantly reduced the abundance of mouse *Pbcas4* and each of the five predicted protein-coding transcript targets. These experiments suggest that *Pbcas4* does indeed share functional MREs with its putative cross-regulation targets, and that it is indeed a conserved ceRNA.

The answer to the pseudogene puzzle?

A ceRNA function may explain why older pseudogenes have been retained in vertebrates for a long period of evolution. To test this conjecture, it may be valuable to establish the time frame during which pseudogenization occurred, using a combinatory approach of substitution-based statistical estimation and phylogenetic distribution. Unitary pseudogenes in mouse and rat may be as old as the mammalian radiation (90 million years) or as young as the divergence between *Mus* and *Rattus*, two genera that split just 12 million years ago. It is likely that unitary pseudogenes lost their coding potential at different time points during rodent evolution; if this assumption holds true, the distribution of pseudogenization times can be extrapolated by identifying lineages in which a protein-coding ortholog is still present. For example, the naked mole rat appears to contain a protein-coding gene that is an ortholog of *Pbcas4*, which sets the pseudogenization time as within the last 60 million years or so. Such an approach should widen our appreciation of pseudogene functionality and help determine the generality of the observations discussed here.

From Marques and colleagues' [7] beautifully combined computational and laboratory experiments, seeking to test a mechanistic model for pseudogene function, it can be learned that the role of miRNA decoys might make some expressed pseudogenes resistant to evolutionary erosion. Given the generality of miRNA regulation in mammalian cells, it is conceivable that many long-living pseudogenes may play such a role.

Abbreviations

ceRNA, competitive endogenous RNA; miRNA, microRNA; MRE, miRNA response element.

References

1. Matsuo Y, Yokoyama S: Cloning and sequencing of a processed pseudogene derived from a human class III alcohol dehydrogenase gene. *Am J Hum Genet* 1990, **46**:85-91.
2. Zhang Z, Harrison PM, Liu Y, Gerstein M: Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 2003, **13**:2541-2558.
3. Zhang YE, Landback P, Vibranovski M, Long M: New genes expressed in human brains: Implications for annotating evolving genomes. *Bioessays* 2012, **34**:982-991.
4. Li W-H: *Molecular Evolution.* Sunderland, Massachusetts: Sinauer Associates; 1997.
5. Long M, Langley CH: Natural selection and the origin of jingwei, a chimeric processed functional gene in *Drosophila*. *Science* 1993, **260**:91-95.
6. Betrán E, Wang W, Jin L, Long M: Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. *Mol Biol Evol* 2002, **19**:654-663.
7. Marques AC, Tan J, Lee S, Kong L, Heger A, Ponting CP: Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. *Genome Biology* 2012, **13**:R102.
8. Petrov DA, Hartl DL: High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. *Mol Biol Evol* 1998, **15**:293-302.
9. Zheng D, Gerstein MB: The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? *Trends Genet* 2007, **23**:219-224.
10. Marques AC, Tan J, Ponting CP: Wrangling for microRNAs provokes much crosstalk. *Genome Biol* 2011, **12**:132.