Plant Pseudogenes: Decoding the Hidden Regulators in Genome Evolution

Hussainbi[1], Swathi P.[2] and Ranjitha[3]

[12]Department of Genetics and Plant Breeding, University of Agricultural Sciences, Dharwad
[3]Department of Agronomy, University of Agricultural Sciences, Dharwad
*Corresponding Author: husnaang@gmail.com

Abstract

The genome, dominated by non-coding regions, was historically deemed as non-functional or junk DNA, encompassing transposable elements, regulatory elements, simple and tandem repeats, functional non-coding RNAs, and pseudogenes. Pseudogenes, discovered in Xenopus laevis as compromised copies of functional genes, are genes that have lost the ability to produce functional proteins due to critical defects in their sequences. Initially considered genomic fossils, pseudogenes are now acknowledged as ubiquitous and abundant in genomes, evolving neutrally. Despite their non-functional protein-coding status, pseudogenes play essential roles in gene regulation, transcribing into RNA, forming small interfering RNA, or influencing cellular miRNA concentration. Bioinformatics tools like Pseudo Pipe and PSF aid in pseudogene prediction. Systematic plant pseudogene studies uncover their lineage-specific expression patterns, demonstrating their crucial role in genome evolution and the potential contribution of pseudogenes to regulatory noncoding RNAs and transcription factor binding sites in plants.

Introduction

In genome, non-coding regions are more abundant than the coding region. These non-coding regions in general were thought to be nonfunctional or junk or no purpose DNA. The non coding regions are transposable elements, regulatory elements, simple and tandem repeats, functional noncoding RNAs, pseudogenes etc. In 1977, Jacq *et al*. discovered in *Xenopus laevis* a truncated copy of a 5S rRNA gene with a compromised function, which they termed a "pseudogene".

Pseudogenes are defined as the gene that have lost its ability to produce functional proteins, which contain critical defects in their sequences, such as lacking a promoter, having a premature stop codon or frameshift mutations. Pseudogenes are ubiquitous and abundant in genomes. They have highest homology to parental functional gene. Most arise as superfluous copies of functional genes, either directly by DNA duplication (unprocessed pseudogenes) or indirectly by reverse transcription of an mRNA transcript (processed pseudogenes). Pseudogenes were once called "genomic fossils" and treated as "junk DNA" several years. They are expected to be evolving neutrally (Xiao *et al*., 2016).

Nevertheless, it has been recognized that some pseudogenes play essential roles in gene regulation of their parent genes, and many pseudogenes are transcribed into RNA. Pseudogene transcripts may also form small interfering RNA or decrease cellular miRNA concentration. Pseudogenes are usually identified when genome sequence analysis finds gene-like sequences that lack regulatory sequences needed for transcription or translation, or whose coding sequences are obviously defective due to frameshifts or premature stop codons.

Types and Origin of Pseudogenes

Pseudogenes are classified based on functionality, birth, and death of gene families. Zheng and Gerstein's classification includes living genes (functional genes), dead pseudogenes (non-transcribed and evolving neutrally), and ghost pseudogenes (with intermediate functionality). Another classification based on origin distinguishes unprocessed pseudogenes (formed by duplication and mutation), processed pseudogenes (formed by retrotransposition), and unitary pseudogenes (rare, resembling unprocessed pseudogenes). Pseudogenes can originate through duplication events or by accumulating mutations in the parental DNA sequence. Duplicated pseudogenes arise when a duplicated copy of a functional gene acquires deleterious mutations, resulting in the loss of its original protein-coding capacity. Unitary pseudogenes form through the accumulation of mutations in a single gene without prior duplication.

AgriTech Today

Pseudogenes in Plants

While pseudogenes are predominantly studied in mammals, their presence in plants is gaining attention. Pseudogenes have been identified and predicted in various plant species, including Rice, Arabidopsis, Rye, Barley, Soybean, *Medicago truncatula*, *Populus trichocarpa*, Sorghum, and *Brachypodium distachyon*. Mascagni *et al*. (2021) identified and characterized the pseudogene complements of five plant species: *Arabidopsis thaliana*, *Vitis vinifera*, *Populus trichocarpa* and *Phaseolus vulgaris*) and *Oryza sativa* based on sequence homology to functional loci.

Prediction of Pseudogenes

Bioinformatics approaches play a crucial role in pseudogene prediction due to the high sequence similarity between functional genes and pseudogenes. Bioinformatics tools for pseudogene prediction have been developed, among which Pseudo Pipe, PSF (Pseudogene Finder), Shiu's pipeline are publicly available (Xie *et al*.,2019). Some other tools such as PPFinder, and Plant pseudogene use homology-based approaches to identify pseudogenes based on parent-pseudogene homologous pairs.

Utilities of Pseudogene

Pseudogenes play crucial roles in gene expression and regulation, contribute to solving biochemical pathways, provide information on splice diversity of RNA, and exhibit a neutral substitution rate. Garewal *et al. (*2022) functionally characterizes diverse roles of the resistance pseudogenes as novel gene footprints and as significant gene regulators in the grapevine genome. They used PlantPseudo pipeline and HMM-profiling to identify whole-genome duplication-derived (WGD)pseudogenes associated with resistance genes (ψ-Rs).

Ali *et al*. (2013) conducted studies in formation and expression of pseudogenes on the B chromosome of rye and results revealed that Bs of rye (*Secale cereale*) are rich in gene-derived sequences. They compared these gene-like fragments of the rye B with their ancestral A-located counterparts and confirmed an A chromosomal origin and the pseudogenization of B-located gene-like fragments. About 15% of the pseudogene-like fragments on Bs are transcribed in a tissue-type and genotype-specific manner. In addition, B-located sequences can cause in trans down or upregulation of A chromosome encoded genic fragments. Phenotypes and effects associated with the presence of Bs might be explained by the activity of B-located pseudogenes. We propose a model for the evolution of B-located pseudogenes.

Xie *et al*. (2019) conducted few systematic studies of plant pseudogenes, hampering comparative analyses. They examined the origin, evolution, and expression patterns of pseudogenes and their relationships with noncoding sequences in seven angiosperm plants. They identified 250,000 pseudogens, most of which are more lineage specific than protein-coding genes, showing tissue or stage specific expression patterns and also found that a large fraction of nontransposable element regulatory noncoding RNAs (miRNAs and lnc RNAs) originate from transcription of pseudogenes proximal upstream regions. They also found that transcription factor binding sites preferentially occur in putative pseudogene proximal upstream regions compared with random intergenic regions, suggesting that pseudogenes have conditioned genome evolution by providing transcription factor binding sites that serve as promoters and enhancers.

Resistance genes associated pseudogenes (pseudo-R genes) derived from whole-genome duplications were identified in the genome of Vitis vinifera and were annotated for their roles in plant defense responses. Pseudo-R genes play by regulating the gene expression either directly by acting as mRNA mimics for miRNA/tasiRNA targeting or indirectly by lncRNA mediated regulation of miRNA/tasiRNAs (Garewal *et al.*, 2022)

Future Aspects

Understanding the mechanisms of pseudogene action may unlock solutions to essential biochemical pathways. Pseudogenes, often considered junk DNA, are emerging as critical players in gene regulation.

Reasons to Take Pseudogenes Seriously in Research

Dysregulation of pseudogenes may contribute to the pathogenesis of various diseases. Pseudogenes can affect functional gene expression, and failure to

distinguish their transcription level from functional genes may impact research data accuracy.

Conclusion

Pseudogenes, once labeled as junk DNA, are gaining prominence in genomic research. Their evolutionary significance, functional roles, and potential contributions to disease pathogenesis highlight the need to study them seriously. Pseudogenes, often overlooked, might hold the key to unraveling intricate aspects of genome regulation and evolution. As research progresses, pseudogenes are likely to become central players in understanding the complexity of genomic landscapes

References

Ali, M., Moghaddam, B., Meier, K., Karimi A., R. and Houben, A., 2013, Formation and expression of pseudogenes on the b chromosome of rye. *Plant Cell*, 25: 2536 - 2544.

Garewal, N., Pathania, S., Bhatia, G. and Singh, K., 2022. Identification of Pseudo-R genes in Vitis vinifera and characterization of their role as immunomodulators in host-pathogen interactions. *Journal of Advanced Research*, 42, pp.17-28.

Mascagni, F., Usai, G., Cavallini, A. and Porceddu, A., 2021. Structural characterization and duplication modes of pseudogenes in plants. *Scientific Reports*, 11(1): 5292.

Xiao, J., Kumar, M., Ragupathy, P., R., Cloutier, S., Wang, X. and Frank, 2016, Pseudogenes and their genome-wide prediction in plants. *Int. J. Mol. sci.*, 17(12): 1991.

Xie, J., Chena, S., Xua, W., Zhao, Y. and Zhang, D., 2019, Origin and function of plant pseudogenes. *Plant signaling behavior*, 14(8).

Xie, J., Li, Y., Liu, X., Zhao, B., Ingvarsson, P., K. and Zhang, D., 2019, Evolutionary origins of pseudogenes and their association with regulatory sequences in plants, *Plant Cell,*31: 563 -578.

* * * * * * * *