Send Orders for Reprints to reprints@benthamscience.net

Current Genomics, 2019, *20*, 000-000

OPINION ARTICLE

The Role of Sequence Duplication in Transcriptional Regulation and Genome Evolution

Luis M. Vaschetto[1,2,*] and Natalia Ortiz [1,3]

[1]*Instituto de Diversidad y Ecología Animal, Consejo Nacional de Investigaciones Científicas y Técnicas (IDEA, CONICET), Av. Vélez Sarsfield 299, X5000JJC Córdoba, Argentina;* [2]*Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, (FCEFyN, UNC), Av. Vélez Sarsfield 299, X5000JJC Córdoba, Argentina;* [3]*Cátedra de Genética de Poblaciones y Evolución, Facultad de Ciencias Exactas, Físicas y Naturales, UNC, Córdoba, Argentina*

ARTICLE HISTORY

Received: January 06, 2019
Revised: January 22, 2019
Accepted: January 24, 2019

DOI:
10.2174/1389202920666190320140721

Abstract: Sequence duplication is nowadays recognized as an important mechanism that underlies the evolution of eukaryote genomes, being indeed one of the most powerful strategies for the generation of adaptive diversity by modulating transcriptional activity. The evolutionary novelties simultaneously associated with sequence duplication and differential gene expression can be collectively referred to as duplication-mediated transcriptional regulation. In the last years, evidence has emerged supporting the idea that sequence duplication and functionalization represent important evolutionary strategies acting at the genome level, and both coding and non-coding sequences have been found to be targets of such events. Moreover, it has been proposed that deleterious effects of sequence duplication might be potentially silenced by endogenous cell machinery (*i.e.*, RNA interference, epigenetic repressive marks, *etc*). Along these lines, our aim is to highlight the role of sequence duplication on transcriptional activity and the importance of both in genome evolution.

Keywords: Sequence duplication, genome evolution, transcriptional regulation, gene duplication, transposable elements.

1. INTRODUCTION

In the last years, comparative genomics has greatly shed light on our understanding of how sequence duplication impacts on genome function, especially because it was found to be associated with both the generation of potentially adaptive variation and the alteration of gene expression [1, 2]. In higher organisms, there are many different genomic rearrangements associated with sequence duplication and functionalization (see Box 1), including polyploidy, chromosome duplication, unequal crossover, *etc* [3-5]. All these events may potentially alter the transcriptional activity and thus affect fundamental biological processes such as development and differentiation. It is well known that duplication of coding sequences alters transcriptional activity through dosage effects [6, 7]. Similarly, duplication of non-coding sequences has been associated with the control of transcriptional activity in animals ranging from arthropods to vertebrates. For instance, in the spider *Parasteatoda tepidariorum*, approximately one-third of the annotated microRNAs (miRNAs) are present in two or more copies, suggesting thereby that duplication of non-coding RNAs (ncRNAs) is an important molecular mechanism underlying the evolution of these arachnids [8]. In vertebrates, miRNA duplication has shown to be essential for accelerating the recruitment of new target sites during the evolution of higher organisms [9].

Box 1 - Mechanisms of sequence duplication associated with functionalization

Neofunctionalization: functional divergence associated with the origin of a novel function in one of the duplicated sequences.

Subfunctionalization: division of the ancestral function between duplicated sequences.

Recent genome-wide studies have highlighted the importance of the link between gene duplication and transcriptional regulation [10-12]. Moreover, computational analysis of high-throughput RNA-Seq data has also recently shed light on the mode in which functional genes emerge through duplication and undergo functionalization [13]. This is especially true for Hox genes, which play essential roles in the development of the anterior-posterior body axis in metazoans. In mammals, Hox genes are classified into 13 sets of paralogous families arranged into four linkage groups [14]. Hox gene clusters contain regulatory non-coding sequences that modulate gene expression at these loci and orchestrate downstream developmental programs [15, 16]. Recent reports reveal that Hox duplication had an important impact during the evolution of distantly related animal lineages, including spiders and mammals. In mammals, the expression of paralogous Hox genes is responsible for the development and differentiation of cells and tissues as well [17, 18].

*Address correspondence to this author at the Instituto de Diversidad y Ecología Animal, Consejo Nacional de Investigaciones Científicas y Técnicas (IDEA, CONICET), Av. Vélez Sarsfield 299, X5000JJC Córdoba, Argentina; Tel: +54 03572 15407205; E-mail: luisvaschetto@conicet.gov.ar

1389-2029/19 $58.00+.00

Moreover, in arachnids, differential expression of duplicated Hox genes may have played important roles in the speciation of this group. In the embryonic stage of *P. tepidariorum*, Hox factors emerged by duplication which exhibit differential spatiotemporal expression have been identified [19]. In the scorpion *Centruroides sculpturatus*, it was reported that differential expression of a duplicated Hox-gene cluster is associated with the segmental identity [20]. Similarly, the differential expression of the duplicated dachshund (dac) developmental gene has been linked with the emergence of the patella, a leg segment only present in spiders [21]. These works, as well as others, show that arachnids can be used as model organisms for the elucidation of the evolutionary mechanisms underlying transcriptional regulation, especially when they involve the study of developmental genes. Likewise, further research into the mechanisms of gene duplication could also help to understand how different gene expression patterns are associated with developmental signaling pathways.

The multicopy ribosomal RNA genes (rDNA) have also been found to be associated with duplication and transcriptional regulation. In eukaryotes, the major family of ribosomal genes is constituted by housekeeping (constitutively expressed) genes organized in tandem arrays and transcribed together as a large precursor RNA molecule, which is then processed to generate mature 28S, 18S and 5.8S ribosomal RNAs (rRNAs). Functional diversification of rDNA genes occurred independently across different lineages during eukaryotic evolution. In chaetognaths, which constitutes a small phylum of marine organisms, an interesting case of functionalization involving rDNA genes has been reported. In *Spadella cephaloptera* (Chaetognatha), two different classes of 18S rDNA paralogous genes have been identified: 18S Class I, constitutively expressed in all tissues, and 18S Class II genes, differentially expressed in oocytes, being the last one of the cellular types that exhibit higher translational activity [22, 23]. Consequently, this suggests that 18S and 28S paralogous genes have important roles during specific cellular processes and they are key players in the evolution of this group [23].

The amplification and duplication of Transposable Elements (TEs) have also been recognized as important mechanisms underlying the evolution of eukaryotic genomes [24-26]. Until recently, TEs were believed to be only selfish/parasitic sequences, but now we know that they may also play important roles in host genome evolution and gene function [27-29]. According to their mechanisms of transposition, TEs can be classified into retrotransposons and DNA transposons, transposed *via* copy-and-paste and cut-and-paste mechanisms, respectively. A particular group of non-autonomous TEs referred to as miniature inverted-repeat transposable elements (MITEs), abundant in many plant and animal genomes, has recently been associated with the regulation of metabolism-related gene clusters that probably emerged through duplication [30]. In grasses, it has been suggested that MITEs could play a role in the mechanism of regulation of master developmental genes found in conserved syntenic blocks [31]. Interestingly, it has been proposed that subfamilies of MITEs derived from amplification of one or a few elements that generated hundreds to thousands of copies [32, 33]. Insights into biological mechanisms associated with MITE activity (*e.g.*, amplification, insertion, *etc*.) will certainly improve our understanding about the roles of TEs in eukaryotic genome evolution. Another interesting example of the association between TEs and transcriptional regulation is the case of Short Interspersed Transposable Elements (SINEs), which are non-autonomous retrotransposons involved in gene expression variation [34]. In the zebrafish (*Danio rerio*) model system, a miRNA family that has undergone multiple duplications and segmental replication during the evolution of vertebrates has been associated with the regulation of genes predicted to bear SINE elements [35]. Moreover, in the elephant shark *Callorhinchus milii* (Elasmobranchii:Callorhynchidae), transcriptomics analysis suggests that SINE retroelements could be involved in the biogenesis of miRNAs and be targets of miRNA regulation [36].

The evolutionary synergy among TEs and target DNA sequences might also shed light on the mechanisms involving transcriptional regulation. Integration of TEs into the genome can be a targeted and site-specific mechanism, as has been described for the insertion of retrotransposons into specific rDNA genes, microsatellite and telomere sequences [37, 38]. These targeted transposition events represent itself a coevolutionary strategy since they may potentially increase the rate of TE insertion without affecting the fitness of the individual (*e.g.*, by preventing the deleterious consequences caused by single-gene disruption). In animals, R2 retrotransposons insert into target 28S rDNA gene sequences, ultimately contributing to their diversification [38]. R2 is a group of retrotransposons that would have diverged more than 850 million years ago before the split of major bilaterian lineages, and they have been identified in rotifers [39], insects [40, 41], crustaceans [42], and chordates [38, 43]. Regarding the importance of R2 TEs in gene regulation, it has been shown that these elements encode an enzyme required for the cleavage of rRNA-R2 co-transcripts [44], being thus involved in the modulation of rDNA transcription. In addition, R2 TEs and target rDNA genes are both epigenetically silenced by the formation of heterochromatin blocks in the nucleus of germline cells [45, 46]. Therefore, since redundant rDNA arrays are present in eukaryotic cells, targeted insertion of R2 elements might be beneficial to the host by generating transcriptional variation at the redundant rDNA gene clusters whilst R2 TEs would exploit this mechanism to increase their inheritance rate. It is important to note that chromosomal localization can also influence rDNA expression [47], thereby the mechanism of insertion of R2 retroelements has not the same effect on different rDNA arrays.

Sequence duplication and transcriptional regulation collectively represent a strategy to generate potentially adaptive variation at the genome level. The combination of these events is known to produce the fixation of sequence variants and to be associated with different mechanisms that are capable to generate genome instability (*i.e.*, conditions leading to genetic rearrangements). Since transcriptional regulation can involve duplication of both protein (coding) and regulatory (non-coding) sequences, it is expected that insights into this adaptive strategy will not only give us a better understanding of the genome organization and sequence evolution but will also shed light on the underlying biological factors that control gene expression and developmental programs.

LIST OF ABBREVIATIONS

miRNAs	=	microRNAs
ncRNAs	=	non-coding RNAs
rDNA	=	ribosomal RNA genes
rRNA	=	ribosomal RNA
TEs	=	Transposable Elements
MITEs	=	Miniature Inverted-repeat Transposable Elements
SINEs	=	Short Interspersed Transposable Elements

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors wish to thank Susana Cervi for the language corrections in a previous version of this manuscript.

REFERENCES

[1] Tang, M.; Chen, Z.; Grover, C. E.; Wang, Y.; Li, S.; Liu, G.; Ma, Z.; Wendel, J.F.; Hua, J. Rapid evolutionary divergence of Gossypium barbadense and G. hirsutum mitochondrial genomes. *BMC genomics*, **2015**, *16*(1), 770.

[2] Hermansen, R.A.; Hvidsten, T.R.; Sandve, S.R.; Liberles, D.A. Extracting functional trends from whole genome duplication events using comparative genomics. *Biol Proced Online*, **2016**, *18*(1), 11.

[3] McElroy, K.E.; Denton, R.D.; Sharbrough, J.; Bankers, L.; Neiman, M.; Gibbs, H.L. Genome expression balance in a triploid trihybrid vertebrate. *Genome Biol Evol*., **2017**, *9*(4), 968-980.

[4] Bittel, D.C.; Kibiryeva, N.; Dasouki, M.; Knoll, J. H.; Butler, M.G. A 9-year-old male with a duplication of chromosome 3p25. 3p26. 2: Clinical report and gene expression analysis. *Am J Med Genet A*, **2006**, *140*(6), 573-579.

[5] Andersson, D.I.; Jerlström-Hultqvist, J.; Näsvall, J. Evolution of new functions de novo and from preexisting genes. *Cold Spring Harb Perspect Biol.*, **2015**, *7*(6), a017996.

[6] Espinosa-Cantú, A.; Ascencio, D.; Barona-Gómez, F.; DeLuna, A. Gene duplication and the evolution of moonlighting proteins. *Front Genet.*, **2015**, *6*, 227.

[7] Bianconi, M.E.; Dunning, L.T.; Moreno-Villena, J.J.; Osborne, C.P.; Christin, P.A. Gene duplication and dosage effects during the early emergence of C4 photosynthesis in the grass genus Alloteropsis. *J Exp Bot.*, **2018** *69*(8), 1967-1980.

[8] Leite, D.J., Ninova, M.; Hilbrant, M.; Arif, S.; Griffiths-Jones, S.; Ronshaugen, M.; McGregor, A.P. Pervasive microRNA duplication in Chelicerates: insights from the embryonic microRNA repertoire of the spider Parasteatoda tepidariorum. *Genome Biol Evol*, **2016**, *8*(7), 2133–44.

[9] Luo, J; Wang, Y; Yuan, J; Zhao, Z; Lu, J. MicroRNA duplication accelerates the recruitment of new targets during vertebrate evolution. *RNA*, **2018**, rna-062752.

[10] Innan, H.; Kondrashov, F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*, **2010**, *11*(2), 97-108.

[11] Lan, X.; Pritchard, J.K. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science*, **2016**, *352*(6288), 1009-1013.

[12] Guschanski, K.; Warnefors, M.; Kaessmann, H. The evolution of duplicate gene expression in mammalian organs. *Genome Res*, **2017**, *27*(9), 1461-1474.

[13] Gao, D.; Ko, D.C.; Tian, X.; Yang, G.; Wang, L. Expression divergence of duplicate genes in the protein kinase superfamily in Pacific oyster. *Evol Bioinform Online*, **2015,** *11*(1), 57-65.

[14] Chen, F.; Capecchi, M.R. Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy. *Proc Natl Acad Sci USA*, **1999**, *96*(2), 541-546.

[15] Brena, C.; Chipman, A.D.; Minelli, A.; Akam, M. Expression of trunk Hox genes in the centipede Strigamia maritima: sense and anti-sense transcripts. *Evol Dev*., **2006**, *8*(3), 252-265.

[16] Lemons, D; McGinnis, W. Genomic evolution of Hox gene clusters. *Science*, **2006**, *313*(5795), 1918-1922.

[17] Hrycaj, S.M.; Dye, B.R.; Baker, N.C.; Larsen, B.M.; Burke, A.C.; Spence, J.R.; Wellik, D.M. Hox5 genes regulate the Wnt2/2b-Bmp4-signaling axis during lung development. *Cell Rep*, **2015**, *12*(6), 903-912.

[18] Ptaschinski, C.; Hrycaj, S.M.; Schaller, M.A.; Wellik, D.M.; Lukacs, N.W. Hox5 paralogous genes modulate Th2 cell function during chronic allergic inflammation *via* regulation of Gata3. *J. Immunol*., **2017**, *199*(2), 501-509.

[19] Schwager, E.E.; Sharma, P.P.; Clarke, T.; Leite, D.J.; Wierschin, T.; Pechmann, M.; Akiyama-Oda, Y.; Esposito, L.; Bechsgaard, J.; Bilde, T.; Buffry, A.D.; Chao, H.; Dinh, H.; Doddapaneni, H.; Dugan, S.; Eibner, C.; Extavour, C.G.; Funch, P.; Garb, J.; Gonzalez, L.B.; Gonzalez, V.L.; Griffiths-Jones, S.; Han, Y.; Hayashi, C.; Hilbrant, M.; Hughes, D.S.T.; Janssen, R.; Lee, S.L.; Maeso, I.; Murali, S.C.; Muzny, D.M.; Nunes da Fonseca, R.; Paese, C.L.B.; Qu, J.; Ronshaugen, M.; Schomburg, C.; Schönauer, A.; Stollewerk, A.; Torres-Oliva, M.; Turetzek, N.; Vanthournout, B.; Werren, J.H.; Wolff, C.; Worley, K.C.; Bucher, G.; Gibbs, R.A.; Coddington, J.; Oda, H.; Stanke, M.; Ayoub, N.A.; Prpic, N.M.; Flot, J.F.; Posnien, N.; Richards, S.; McGregor, A.P. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol*, **2017**, *15*(1), 62.

[20] Sharma, P.P.; Schwager, E.E.; Extavour, C.G; Wheeler, W.C. Hox gene duplications correlate with posterior heteronomy in scorpions. *Proc R Soc Lond B Biol Sci*, **2014**, *281*(1792), 20140661.

[21] Turetzek, N.; Pechmann, M.; Schomburg, C.; Schneider, J.; Prpic, N.M. Neofunctionalization of a duplicate dachshund gene underlies the evolution of a novel leg segment in arachnids. *Mol. Biol. Evol*., **2015**, *33*(1), 109-121.

[22] Barthélémy, R.M.; Grino, M.; Pontarotti, P.; Casanova, J.P.; Faure, E. The differential expression of ribosomal 18S RNA paralog genes from the chaetognath Spadella cephaloptera. *Cell. Mol. Biol. Lett.*, **2007**, *12*(4), 573.

[23] Barthélémy, R.M.; Grino, M.; Pontarotti, P.; Casanova, J.P.; Faure, E. A possible relationship between the phylogenetic branch lengths and the chaetognath rRNA paralog gene functionalities: ubiquitous, tissue-Specific or pseudogenes. In *Evolutionary Biology from Concept to Application*, **2008**, (pp. 155-164), Springer, Berlin, Heidelberg.

[24] Elbarbary, R.A.; Lucas, B.A.; Maquat, L.E. Retrotransposons as regulators of gene expression. *Science*, **2016**, *351*(6274), aac7247.

[25] Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet*, **2017**, *18*(2), 71-86.

[26] Trizzino, M.; Kapusta, A.; Brown, C.D. Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics*, **2018**, *19*(1), 468.

[27] Wang, K.; Huang, G.; Zhu, Y. Transposable elements play an important role during cotton genome evolution and fiber cell development. *Sci China Life Sci*., **2016**, *59*(2), 112-121.

[28] Macias, F.; Afonso-Lehmann, R.; Lopez, M.C.; Gomez, I.; Thomas, M.C. Biology of Trypanosoma cruzi Retrotransposons: From an Enzymatic to a Structural Point of View. *Curr Genomics*, **2018**, *19*(2), 110-118.

[29] González, J.; Petrov, D.A. The adaptive role of transposable elements in the Drosophila genome. *Gene*, **2009**, *448*(2), 124-133.

[30] Boutanaev, A.M.; Osbourn, A.E. Multigenome analysis implicates miniature inverted-repeat transposable elements (MITEs) in metabolic diversification in eudicots. *Proc Natl Acad Sci USA*, **2018**, *115*(28), E6650-E6658.

[31] Vaschetto L.M. Miniature Inverted-repeat Transposable Elements (MITEs) and their effects on the regulation of major genes in cereal grass genomes. *Mol breed*, **2016**, *36*(3), 1-4.

[32] Feschotte, C.; Swamy, L.; Wessler, S.R. Genome-wide analysis of

mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). *Genetics*, **2003**, *163*(2), 747-758.

[33] Yang, G.; Nagel, D.H.; Feschotte, C.; Hancock, C.N.; Wessler, S.R. Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. *Science*, **2009**, *325*(5946), 1391-1394.

[34] Zeng, L.; Pederson, S.M.; Cao, D.; Qu, Z.; Hu, Z.; Adelson, D.L.; Wei, C. Genome-Wide Analysis of the Association of Transposable Elements with Gene Regulation Suggests that Alu Elements Have the Largest Overall Regulatory Impact. *J Comput Biol*, **2018**, *25*(6), 551-562.

[35] Scarpato, M.; Angelini, C.; Cocca, E.; Pallotta, M.M.; Morescalchi, M. A.; Capriglione, T. Short interspersed DNA elements and miRNAs: a novel hidden gene regulation layer in zebrafish?. *Chromosome Res*, **2015**, *23*(3), 533-544.

[36] Luchetti, A.; Plazzi, F.; Mantovani, B. Evolution of Two Short Interspersed Elements in Callorhinchus milii (Chondrichthyes, Holocephali) and Related Elements in Sharks and the Coelacanth. *Genome Biol Evol*, **2017**, *9*(6).

[37] Kojima, K.K.; Fujiwara, H. Cross-genome screening of novel sequence specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol. Biol. Evol*., **2004,** *21*, 207–217.

[38] Kojima, K.K.; Fujiwara, H. Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol. Biol. Evol*., **2005**, *22*(11), 2157-2165.

[39] Gladyshev, E.A.; Arkhipova, I.R. Rotifer rDNA-specific R9 retrotransposable elements generate an exceptionally long target site duplication upon insertion. *Gene*, **2009**, *448*(2), 145-150.

[40] Roiha, H.; Miller, J.R.; Woods, L.C.; Glover, D.M. Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in D. melanogaster. *Nature* **1981**, *290*, 749–753.

[41] Fujiwara, H.; Ogura, T.; Takada, N.; Miyajima, N.; Ishikawa, H.; Maekawa, H. Introns and their flanking sequences of Bombyx mori rDNA. *Nucleic Acids Res*, **1984**, *12*(17), 6861–6869.

[42] Luchetti, A; Mingazzini, V; Mantovani, B. 28S junctions and chimeric elements of the rDNA targeting non-LTR retrotransposon R2 in crustacean living fossils (Branchiopoda, Notostraca). *Genomics* **2012**, *100*(1), 51-56.

[43] Kapitonov, V.V.; Jurka, J. R2 non-LTR retrotransposons in the bird genome. *Repbase Rep*, **2009**, *9*, 1329.

[44] Eickbush, D.G.; Eickbush, T.H. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol. Cell Biol.,* **2010**, *30*(13), 3142-3150.

[45] Eickbush, D.G.; Ye, J.; Zhang, X.; Burke, W.D.; Eickbush, T.H. Epigenetic regulation of retrotransposon within the nucleolus of Drosophila. *Mol. Cell Biol*., **2008,** *28*(20), 6452–6461.

[46] Eickbush, T.H.; Eickbush D.G. Integration, regulation, and long-term stability of R2 retrotransposons. *Microbiol. Spectr*., **2015**, *3*(2), MDNA3-0011-2014.

[47] Ferhat, A.; Bunnik, E.M.; Varoquaux, N.; Bol, S.M.; Prudhomme, J.; Vert, J-P.; Noble, W.S.; Le Roch, K.G. Three-dimensional modeling of the P. falciparum genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res*., **2014**, *24*(6), 974-988.