

BMB Reports – Manuscript Submission

Manuscript Draft

Manuscript Number: BMB-18-010

Title: New Role of LTR-Retrotransposons for Emergence and Expansion of Disease-Resistance Genes and High-Copy Gene Families in Plants

Article Type: Perspective (Invited Only)

Keywords: NLR; retroduplication; LTR-retrotransposon; disease-resistance gene; genome evolution

Corresponding Author: Doil Choi

Authors: Seungill Kim¹, Doil Choi^{1,*}

Institution: ¹Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea,

New Role of LTR-Retrotransposons for Emergence and Expansion of Disease-Resistance Genes and High-Copy Gene Families in Plants

Seungill Kim, & Doil Choi

Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea.

Running Title: Massive Emergence of Disease-Resistance Gene Family by Retroduplication in Plants

Keywords: NLR, retroduplication, LTR-retrotransposon, disease-resistance gene, genome evolution

Perspective to: Seungill Kim et al. (2017), New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication, *Genome Biology*, 18(1):210, doi: 10.1186/s13059-017-1341-9

Corresponding author

Doil Choi

Tel: +82-2-880-4568

Fax: +82-2-873-2056

Email: doil@snu.ac.kr

ABSTRACT

Long terminal repeat retrotransposons (LTR-Rs) are major elements creating new genome structure for expansion of plant genomes. However, in addition to the genome expansion, the role of LTR-Rs has been unexplored. In this study, we constructed new reference genome sequences of two pepper species (*Capsicum baccatum* and *C. chinense*), and updated the reference genome of *C. annuum*. We focused on the study for speciation of *Capsicum spp.* and its driving forces. We found that chromosomal translocation, unequal amplification of LTR-Rs, and recent gene duplications in the pepper genomes as major evolutionary forces for diversification of *Capsicum spp.* Specifically, our analyses revealed that the nucleotide-binding and leucine-rich-repeat proteins (NLRs) were massively created by LTR-R-driven retroduplication. These retoduplicated NLRs were abundant in higher plants, and most of them were lineage-specific. The retroduplication was a main process for creation of functional disease-resistance genes in Solanaceae plants. In addition, 4-10% of whole genes including highly amplified families such as MADS-box and cytochrome P450 emerged by the retroduplication in the plants. Our study provides new insight into creation of disease-resistance genes and high-copy number gene families by retroduplication in plants.

Transposable elements (TEs) are key regulatory elements involved in shaping of genome structures, gene expression changes and creation of new genes for evolution and diversification of species. Specifically, LTR-retrotransposons (LTR-Rs) are the most abundant TEs especially in plants occupying more than 70% of many plant genomes. To date, several studies have reported evolutionary roles of the LTR-Rs for lineage-specific genome expansion generating species barriers by extreme genome size variations. Except the genome expansion, few studies revealed that LTR-Rs drive emergence of new genes via retroduplication events. However, most of the new genes generated by the LTR-Rs-mediated retroduplication were known as uncharacterized or pseudo genes.

Since the first pepper genome project was completed in 2014, we have constructed two more *de novo* genomes of hot pepper (*Capsicum baccatum* and *C. chinense*, hereafter Baccatum and Chinense) to make references representing the genus *Capsicum*. We improved qualities of gene annotation and pseudomolecule chromosomes for the pre-existing pepper genome (*C. annuum*, hereafter Annum). Gene annotation quality influences further functional and comparative studies but most of plant gene annotations remained as initial version, because the researchers that implemented the plant genome project did not further update the annotation. When we compared the previous and the updated version of annotations, 9,000 genes mainly containing hypothetical proteins and TE-related genes were eliminated and 10,000 genes related to various functions were newly identified in the updated gene model.

We had two main biological questions for speciation and diversification of *Capsicum* spp., and massive emergence of high-copy genes in the pepper genomes. The divergence of peppers was generated first between Baccatum and an ancestor species of other two peppers at 1.7 million years ago (MYA), and subsequently between Annum and Chinense at 1.1 MYA. The origin of *Capsicum* species was in western South America: Peru, Ecuador, Colombia and Bolivia. However, the habitats of Annum, Baccatum and Chinense had distinctly diverged to North America, south to west of South America, and west to north-east of South America, respectively. Considering that North and South America had been connected at Pliocene (5.3 – 2.6 MYA), Annum might be moved from west of South America to North America after the connection of the two continents and Chinense was

relocated in north-east of South America, after the speciation at 1.1 MYA.

As evolutionary forces for speciation and diversification among the three genomes, we detected translocations among chromosome 3, 5, 9 between *Baccatum* and the other pepper genomes, and unequally expanded LTR-Rs and genes around and after speciation of the three pepper genomes. The predominantly duplicated genes at the speciation time were nucleotide-binding and leucine-rich-repeat proteins (NLRs) known as disease resistance gene families in plants. We devoted attention to possibility for rapid duplication of NLRs by LTR-Rs due to expansion of LTR-Rs and NLRs. On average, 105 NLRs (13%) were inside LTR-Rs, having reduced the number of exons compared to their parental genes in each pepper genome. Specifically, 70% of them were included in specific subgroup CNL-G2. This indicates that the unusual expansion of specific NLRs in specific chromosomes were caused by LTR-Rs. In principle, retrogenes should have a single-exon but, 32% of the retroNLRs had introns. When we analyzed sequence structures of retroNLRs with introns and compared them to their parental sequences, the intron numbers of the retroNLRs were reduced. This suggests that the retroNLRs with introns were created via alternative splicing mechanisms such as exon skipping or intron retention. In addition, 5%-18% of NLRs in tomato, potato and rice genomes were retrogenes. Like the peppers, NLRs in CNL-G9 were particularly expanded in potato via LTR-Rs-driven retroduplication.

We investigated retroduplication events for functional disease-resistance genes in plants. We analyzed *L* genes in pepper, which provides resistance to *Tobamoviruses*, and *R3a* of potato, a resistance gene to the late blight pathogen *Phytophthora infestans*. Our analyses revealed that recent ancestral genes of the *L* and *R3a* derived from retroduplication and subsequently, *L* genes diversified by sequence mutations and *R3a* emerged by tandem duplications (Fig. 1). In addition to the retroduplication events for the NLRs and the disease-resistance genes, we observed that 4%-10% (1,398 to 3,898 genes) of whole genes in the three peppers, tomato, potato and rice were located inside LTR-Rs as retrogene candidates. We confirmed 42% of them were expressed through transcriptome analyses and 45% of those genes contained functional domains. Especially, MADS-Box in *Baccatum* and cytochrome P450 genes in potato were highly expanded by the retroduplication.

These results revealed that the LTR-Rs are main evolutionary elements for emergence and lineage-specific amplification of the high-copy gene families including NLRs, MADS-Box, and cytochrome P450 genes through retroduplication in plants.

In this study, we demonstrated the importance for improvement of gene annotation quality using updated materials and method based on accumulated knowledge. Specifically, many essential genes were newly identified through the gene annotation update process. Because gene annotation is a prediction process relying on former resources and methods, the annotated gene models would be always fallacious and thus should be continuously improved. Our results revealed new evidence for massive emergence of NLRs and other many functional gene families by LTR-Rs-driven retroduplication. This phenomenon was generally observed in higher plants considering that 4%-10% of whole genes were classified as the retrogenes in peppers, tomato, potato and rice. A remarkable feature of the retroduplication event was that the specific genes explosively expanded and they were contained in distinct subgroups of different plants. Taken together, our study provides insight into new role of LTR-Rs for evolution of disease resistance gene families and other high-copy number gene families in plant kingdom.

ACKNOWLEDGEMENTS

This study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1A6A3A04004014) to S.K. and by a grant from the Agricultural Genome Center of the Next Generation Biogreen 21 Program of RDA (Project No. PJ013153) to D.C.

UNCORRECTED PROOF

Figure 1. Evolutionary history of *L* and *R3a* genes. The diagram shows the *L* and *R3a* were first emerged by retroduplication and subsequently evolved by unequal sequence mutation and tandem duplication, respectively. This figure is modified from Kim et al., *Genome Biology* (2017).

UNCORRECTED PROOF

