

BRIEF REPORT

Open Access



Amplification of LTRs of extrachromosomal linear DNAs (ALE-seq) identifies two active *Oryco* LTR retrotransposons in the rice cultivar Dongjin

Hyunjin Koo¹, Soomin Kim¹, Hyun-Seung Park¹, Sang-Ji Lee¹, Nam-Chon Paek¹, Jungnam Cho² and Tae-Jin Yang^{1*}

Abstract

Long terminal repeat retrotransposons (LTR-RTs) make up a considerable portion of plant genomes. New insertions of these active LTR-RTs modify gene structures and functions and play an important role in genome evolution. Therefore, identifying active forms of LTR-RTs could uncover the effects of these elements in plants. Extrachromosomal linear DNA (eclDNA) forms during LTR-RT replication; therefore, amplification LTRs of eclDNAs followed by sequencing (ALE-seq) uncover the current transpositional potential of the LTR-RTs. The ALE-seq protocol was validated by identification of *Tos17* in callus of Nipponbare cultivar. Here, we identified two active LTR-RTs belonging to the *Oryco* family on chromosomes 6 and 9 in rice cultivar Dongjin callus based on the ALE-seq technology. Each *Oryco* family member has paired LTRs with identical sequences and internal domain regions. Comparison of the two LTR-RTs revealed 97% sequence identity in their internal domains and 65% sequence identity in their LTRs. These two putatively active *Oryco* LTR-RT family members could be used to expand our knowledge of retrotransposition mechanisms and the effects of LTR-RTs on the rice genome.

Keywords: Long terminal repeat retrotransposon, Transposition, ALE-seq, *Oryza sativa*, Dongjin

Background

Transposable elements (TEs) are mobile DNA sequences that contribute to the size variation and evolution of all eukaryotic genomes, especially plant genomes. Most plant species have a wide range of TEs, which make up more than 85% of the genome in plants with large genomes such as wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*) [1–4]. TEs increase their copy numbers and insert into new regions of the genome. The

transposition of TEs can induce structural rearrangements, generating mutations and affecting genome stability. In addition, these movements affect gene expression and function, ultimately influencing evolution and genome adaptation [5].

TEs are divided into two classes: retrotransposons (class I), which mobilize via a copy-and-paste mechanism; and DNA transposons (class II), which transpose via a cut-and-paste mechanism [6]. Long terminal repeat retrotransposons (LTR-RTs) are ubiquitous class I retrotransposons that are interspersed in all eukaryote genomes [7]. In most plants, LTR-RTs contribute to genome size expansion via their rapid proliferation [8, 9]. For example, LTR-RTs account for more than 70% of the maize (*Zea mays*) and wheat (*Triticum aestivum*)

*Correspondence: tjyang@snu.ac.kr

¹ Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 08826 Seoul, Republic of Korea
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

genomes [8, 10]. Intact LTR-RTs contain two similar or identical LTRs that flank an internal region containing a functional domain, target site duplications (TSDs), a polypurine tract (PPT), and a primer-binding site (PBS) [6]. Most LTR-RTs are classified into two superfamilies, *Copia* and *Gypsy*, which differ in their sequences and the composition of the internal conserved domains [6, 11]. The *Gypsy* superfamily is the most highly represented superfamily in most plant genomes [12–14].

Since transposition of LTR-RTs is a major driving force for genome evolution, various studies have aimed to identify active LTR-RTs [15–18]. The transposition cycle of LTR-RTs begins with the transcription of genomic copies, followed by reverse transcription of the LTR-RT transcripts to form cDNA, also known as extrachromosomal linear DNA (eclDNA). These eclDNAs integrate into the nucleus and influence strand transfer [19]. Therefore, although the production of eclDNAs does not necessarily represent transposition events directly, it is considered a strong indication of transposition potential [20]. The transpositional potential of LTR-RTs has been revealed by identifying eclDNA via amplification of LTRs of eclDNAs followed by sequencing (ALE-seq) [15].

Rice (*Oryza sativa*) is an important cereal crop, providing a major source of food worldwide [21]. 40% of the rice genome consists of most known types of TEs, including LTR-RTs and DNA transposons [22]. Among these, LTR-RTs make up ~22% of the rice genome [23]. ALE-seq analysis was successfully used to detect eclDNAs, such as heat stress-activated *Go-on* and activated *Tos17*, in *O. sativa* ssp. *japonica* cv. Nipponbare callus [15]. Although active LTR-RTs have been identified in this model rice variety, little is known about active transposition in non-reference rice varieties. To close this gap, we conducted ALE-seq analysis to identify the transpositionally competent LTR-RTs during callus culture in the non-reference Korean rice variety Dongjin. To facilitate the identification of putatively active LTR-RTs, we interrogated the rice genome to search for structurally intact LTR-RTs, which would likely possess strong mobility. Importantly, ALE-seq detected eclDNAs derived from two *Oryco* family LTR-RTs, which are unique in Dongjin callus. Our work unveils previously unknown mobile DNAs in a non-reference rice genome and will help broaden the breadth of our knowledge on crop genome plasticity and stability.

Results

Identification and annotation of intact LTR-RTs

We tried to detect extrachromosomal linear DNA (eclDNA) forms of active LTR-RTs in Korean rice cultivar Dongjin by ALE-seq. The transpositionally competent LTR-RT should contain an intact structure, including two identical LTRs, a PBS, a PPT, a TSD, and

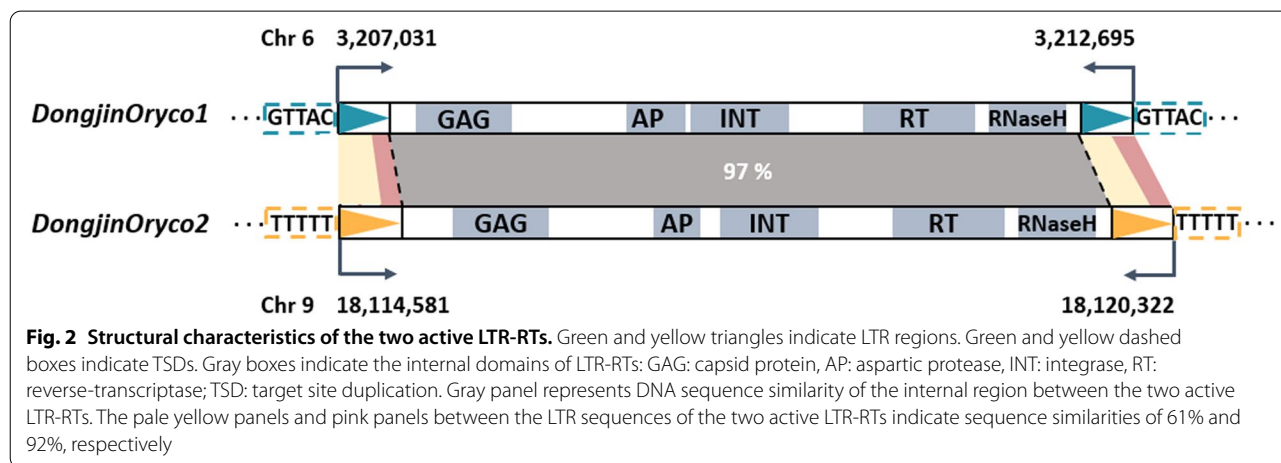
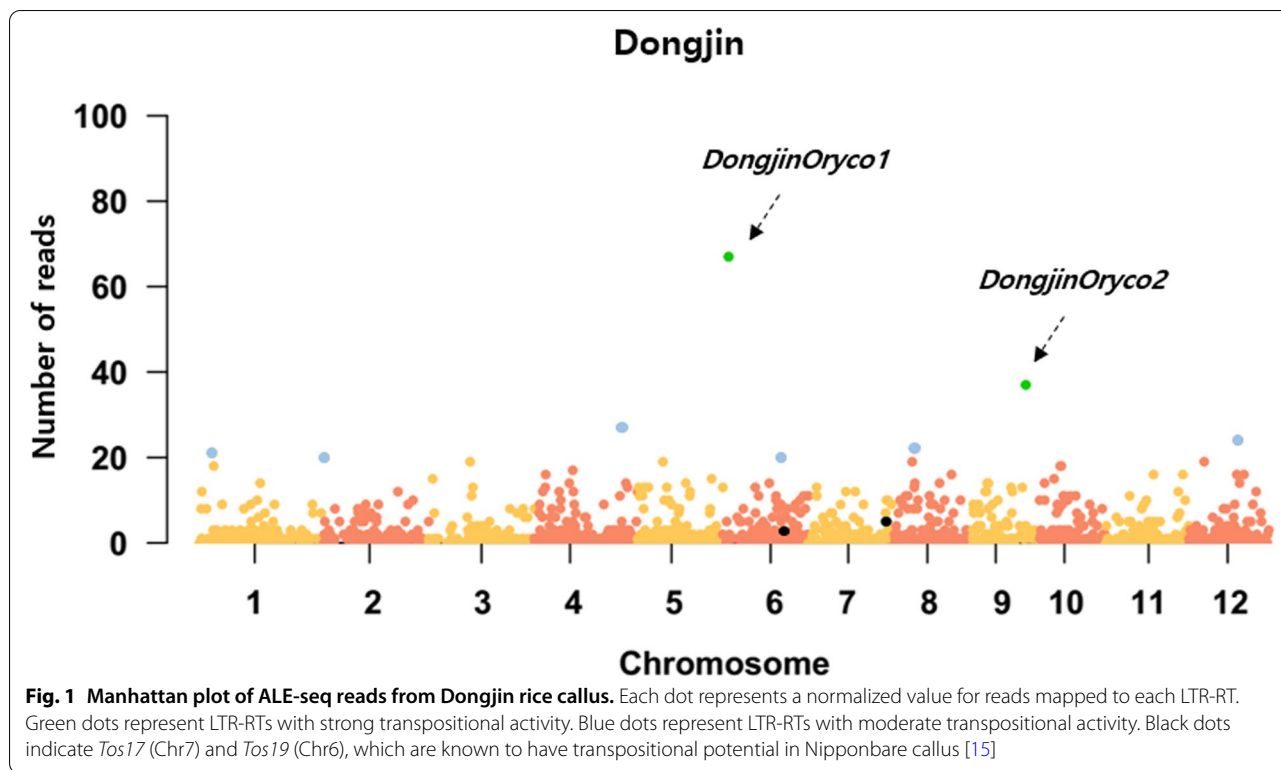
conserved internal domains. Taking all these criteria into consideration, we discovered intact LTR-RTs using several LTR-RTs detection tools and custom python scripts (Additional file 2: Fig S1). We identified 1,783 intact LTR-RTs, including 1,226 *Gypsy*-type (68.8%) and 557 *Copia*-type (31.2%) LTR-RTs, in the *O. sativa* reference genome sequence. The *Del* family as well as the *Tat* family were prominent in the *Gypsy* superfamily, and the *Tork* family was the most abundant in the *Copia* superfamily (Additional file 1: Table S1-3).

Phylogenetic analysis of the LTR-RTs based on their reverse transcriptase (RT) domain distinguished the *Copia* and *Gypsy* superfamilies. The *Copia* superfamily was grouped into the *Oryco*, *Tork*, *Retrofit*, and *Sire* families; the *Gypsy* superfamily was grouped into the *Tat*, *Del*, *Reina*, and *Crm* families (Additional file 2: Fig S2). The *Tat* family accounted for the largest number of *Gypsy* superfamily members and had more variants than the other families. We plotted the pattern of full-length LTR-RTs on the phylogenetic tree to examine the length distribution of the LTR-RTs across the entire family (Additional file 2: Fig S3). A comparison of the LTRs on the left versus the right side for each element revealed that most elements were active during the past one million years (Myr), except for the *Sire* family (about 1–2 Myr ago) (Additional file 1: Table S4, Additional file 2: Fig S4).

Identification of putatively active LTR-RTs in Dongjin callus

The ALE-seq method was recently developed to identify candidate active LTR-RTs that have retrotranspositional activity and thus produce eclDNAs. Two LTR-RTs, *Tos17* and *Tos19*, which are known to be active in rice, were successfully detected as active LTR-RTs by ALE-seq using *O. sativa* cultivar Nipponbare callus [15]. In this study, we conducted ALE-seq to identify active LTR-RTs in callus of another rice variety, Dongjin. Mapping of ALE-seq reads from Dongjin callus identified signatures of eclDNA production derived from two LTR-RTs located on chromosomes 6 and 9 (Fig. 1). These putatively active LTR-RTs belong to the *Oryco* family and were therefore named *DongjinOryco1* and *DongjinOryco2* (Fig. 2, Additional file 1: Table S2, Additional file 2: Fig S5).

BLAST searches revealed that *DongjinOryco1* and *DongjinOryco2* are transcribed in various tissues, such as shoots, roots, and flowers, in diverse rice accessions (Additional file 1: Table S5-6). However, *DongjinOryco1* and *DongjinOryco2* were not detected in the callus of Nipponbare rice using ALE-seq. The ALE-seq reads were uniquely mapped to 5' LTR regions and primer binding sites of the internal domains of the active LTR-RTs *DongjinOryco1* and *DongjinOryco2* in Dongjin callus (Fig. 3a and b) and *Tos17* in Nipponbare callus (Fig. 3c),

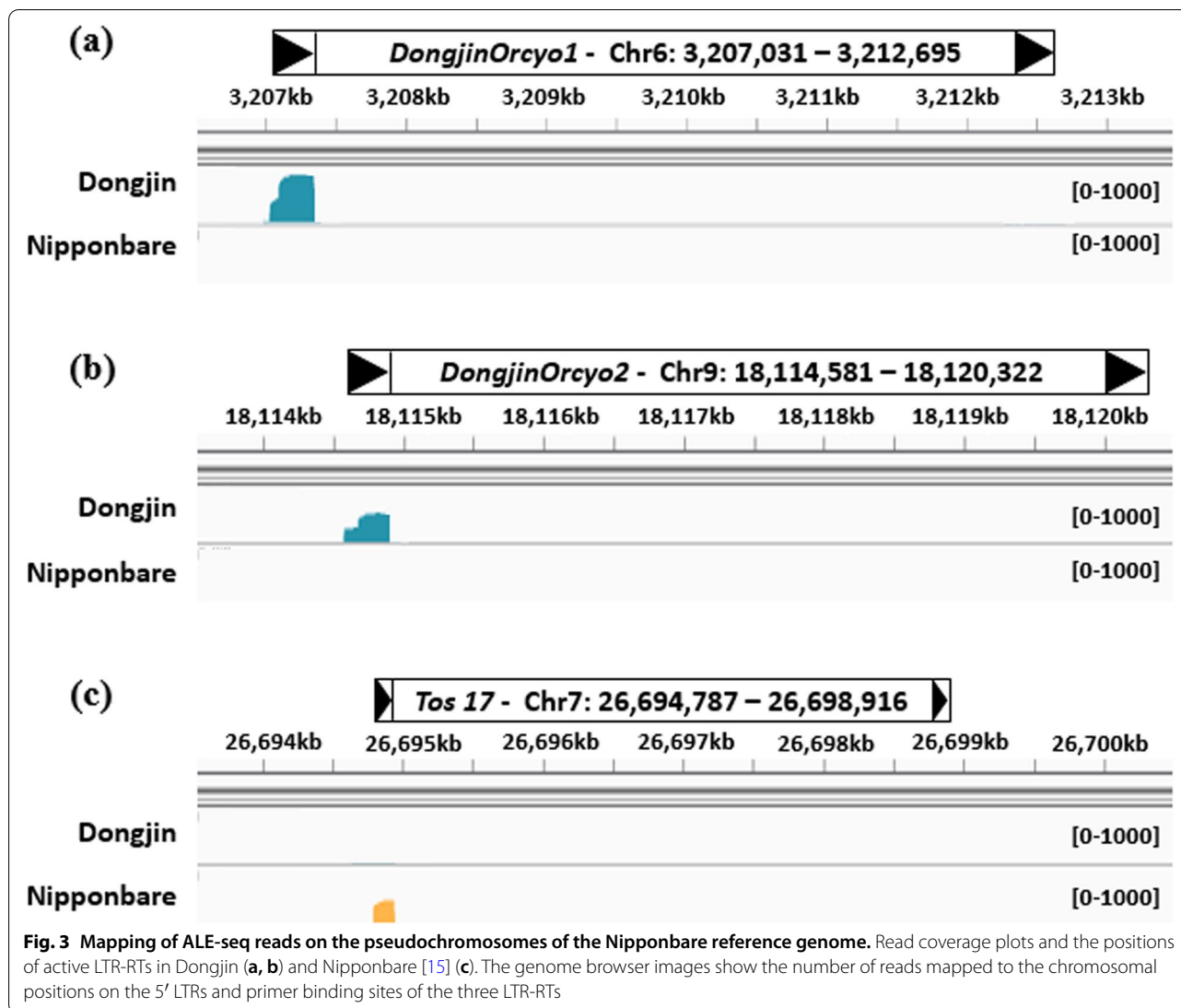


respectively. The other six *Tat* family members showed moderate transpositional potential, and their map positions were identified on chromosomes 1, 2, 4, 6, 8, and 12 (Fig. 1, Additional file 1: Table S2).

Structural characteristics of *DongjinOryco1* and *DongjinOryco2*

The putative active *Oryco* elements *DongjinOryco1* and *DongjinOryco2* have intact structures, including 100% identical sequences between the left and right LTR

structures (Additional file 2: Fig S6) and the conserved internal domain region (GAG, AP, INT, RT, and RNaseH) (Additional file 2: Fig S7). The *Oryco* elements share 97% identical conserved internal sequences. However, the LTR sequences show diversity between *DongjinOryco1* and *DongjinOryco2*, with 61% and 92% sequence similarity for the 5' and 3' regions of the LTRs, respectively (Fig. 2, Additional file 2: Fig S5). Six other *Tat* family members showed relatively high (moderate) ALE-seq read depth, although the depth was much lower than



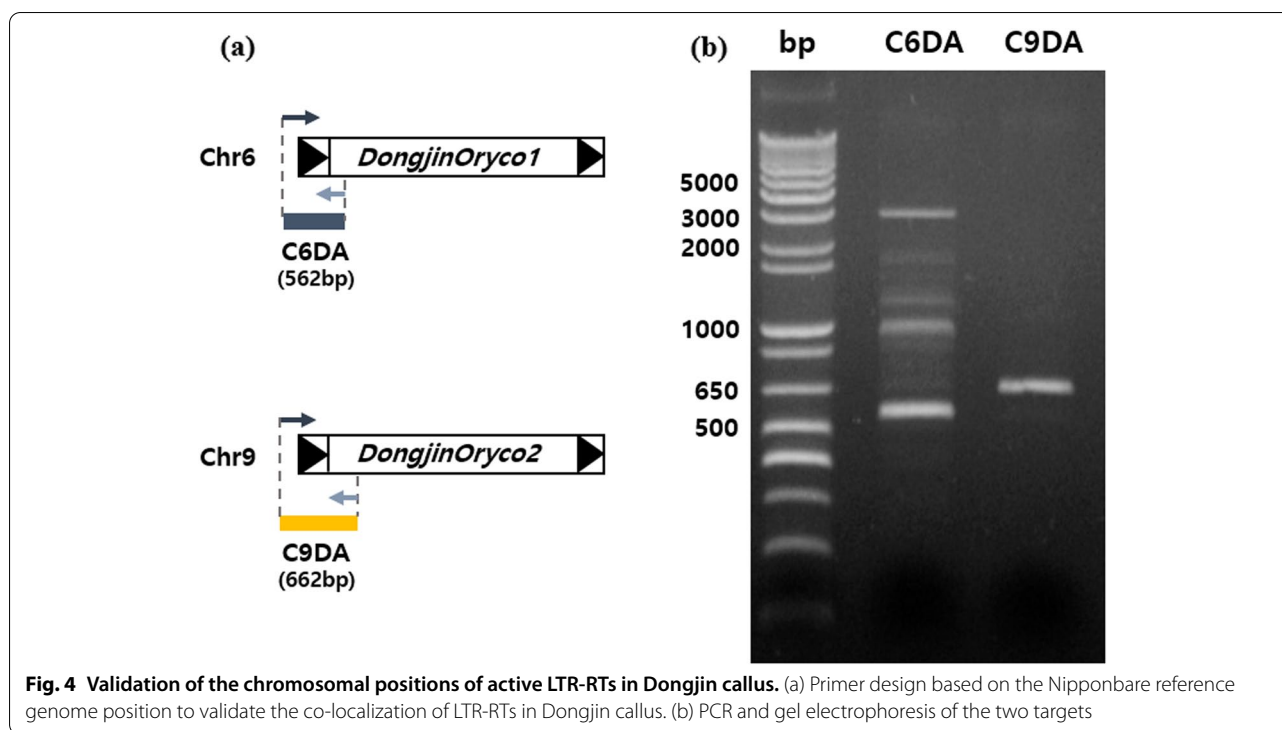
that of the two *Oryco* family members. The *Tat* family members shared 59–96% sequence similarity for the entire internal region and 97.73–100% sequence similarity between the left and right LTR pair of each element (Additional file 1: Table S7, Additional file 2: Fig S7-9).

We annotated and identified the intact LTR-RTs using the Nipponbare reference genome. Since the positions of LTR-RTs can differ among varieties, we validated the chromosomal positions of *DongjinOryco1* and *DongjinOryco2* by PCR amplification using primers designed based on the flanking site and internal LTR-RTs located on chromosomes 6 and 9 (Fig. 4).

Discussion

Retrotranspositional activity and structural integrity of the active LTR-RTs

ALE-seq analysis discovered a total of eight potentially active LTR-RTs in Dongjin callus including two *Oryco* family members and six *Tat* family members. The putative transposition time and retrotranscriptional activity of an LTR-RT can be estimated by comparing its LTR sequences [24] and the conservation of its internal domain regions [25]. The two *Oryco* family members were more conserved than the six *Tat* family members in terms of their overall structure, suggesting that the two *Oryco* LTR-RTs transposed recently and maintain higher transpositional potential than the six *Tat* family members.



Gypsy elements are more abundant in rice than *Copia* elements, which is similar to findings for *Solanum*, *Populus*, and *Capsicum* species [12–14]. *Copia* elements are usually more transcriptionally active than *Gypsy* elements [26, 27]. Elements that are abundant throughout the genome are more likely to be silenced by the host genome, whereas elements that are less abundant in the genome reduce the efficiency of the host's silencing mechanisms [26–28]. *Tat* elements belonging to the *Gypsy* superfamily and *Oryco* elements belonging to the *Copia* superfamily accounted for 28.49% and 4.77% of the rice LTR-RTs, respectively (Additional file 1: Table S1). Therefore, the strong retrotranspositional potential of the *Oryco* elements belonging to the *Copia* superfamily appears to be supported by their strong structural integrity as well as their lower copy numbers in the rice genome [15].

ALE-seq identifies new active LTR-RTs in Dongjin callus

LTR-RTs are transcriptionally activated when plants are exposed to environmental stress, such as heat, cold, salt, or tissue culture conditions [15, 18, 29–32]. However, it is difficult to directly confirm the transpositional potential of an LTR-RT due to post-transcriptional control [33]. The ALE-seq method was recently introduced to identify the direct retrotransposition potential of an LTR-RT by detection of eclDNA. Using this method, *Tos17*, which

is known to be a major factor in tissue culture–induced mutations, was detected as having transpositional activity in the callus of *O. sativa* variety Nipponbare [15]. The *Tos17* system has greatly contributed to functional genomics studies of rice. However, this technique has been limited to Nipponbare, which was selected as a reference cultivar among japonica rice accessions [34]. Because functional genomics tools are also needed for non-reference varieties, we performed ALE-seq analysis using the callus of Korean rice variety Dongjin.

Notably, *Tos17* was highly active in Nipponbare callus but not in Dongjin callus (Fig. 3). *Tos17* elements are activated in most rice varieties, but not in the Moritawase variety, which has higher DNA methylation levels than other varieties, although the *Tos17* sequence in Moritawase is identical to that of Nipponbare [35]. Epigenetic variation might induce the silencing of *Tos17* in the Moritawase variety [35]. Similarly, the *Oryco* family members showed high transpositional potential in Dongjin callus but not in Nipponbare callus, which may have been influenced by the various epigenetic or genetic mutations that exist between the Dongjin and Nipponbare varieties.

TE amplification generates abundant genetic diversity in domesticated cultivars of rice and many other crops [1, 36–40]. The Dongjin and Nipponbare varieties exhibit large genetic variations [41]. Here, we identified two LTR-RTs, *DongjinOryco1* and *DongjinOryco2*, which

are active in Dongjin callus but not in Nipponbare using ALE-seq method. These results suggest that the differential activation of diverse LTR-RTs might have influenced the diversification of rice varieties, which will require further investigation. In conclusion, ALE-seq is a robust approach to unearth hidden mobile DNAs and can be applicable to other plant species, varieties, and tissues to dissect the genome dynamics and plasticity.

Methods

Identification and characterization of LTR-RTs

De novo annotation of LTR-RTs in the rice reference genome (MSU7.0; <http://rice.plantbiology.msu.edu>) was performed with LTRharvest using the following options: `minlenlr=100`, `maxlenlr=5000`, and `similar=80`. The plant-specific tRNAs used to screen PBS regions were obtained from plantRNA (<http://seve.ibmp.unistra.fr/plantrna/>). To filter out strictly intact LTR-RTs, we also scanned for the presence of tRNA and internal domains using LTRdigest with the following options: `pptlen=c[8,40]`, `pbsalilen=c[8,40]`, and `pbsoffset=c(0,10)`. The flanking 5-bp regions at the start and end of each LTR-RT with more than 60% sequence homology were considered to be TSDs. Conserved domains (such as gag and pol proteins) were investigated using the TESorter tool, which uses the gydb database containing GAG, AP, INT, RT, and RNaseH domain sequences [42] (Additional file 2: Fig S1).

To confirm the relatedness of the annotated LTR-RT families, RT amino acid sequences for each family were aligned using MAFFT [43]. The phylogenetic tree was constructed using the aligned RT amino acid sequences based on the maximum-likelihood method via IQ-TREE [44].

The pair of LTR sequences was aligned with MAFFT, and sequence divergence (k) between the pair of LTRs was estimated using Kimura's two-parameter *moDel* [45]. The insertion ages of the LTR-RTs were estimated using the formula $T = k/2r$ (r : substitution rate of 1.3×10^{-8} substitutions per site per year) [46].

Plant materials and ALE-seq analysis

Rice callus was generated from mature embryos of *O. sativa* ssp. *japonica* cv. Dongjin seeds. Dehulled seeds were sterilized with 2% sodium hypochlorite for 30 min and washed three times in sterile water. The sterilized seeds were inoculated on N6D medium containing 2 mg/L 2,4-D and cultured under continuous light at 30 °C for 2 weeks [47].

Total genomic DNA was extracted from the callus using a GeneAll Plant SV Midi Kit (GeneAll Biotechnology Co. Ltd., Seoul, Korea) following the manufacturer's protocol. Genomic DNA (300 ng) was used to generate the ALE-seq library. The first step in ALE-seq was to

connect an adapter containing the T7 promoter sequence to the end of 5' end of the ecDNA, followed by *in vitro* transcription using T7 RNA polymerase. Reverse transcription was then performed using a primer that binds to the transcript of the PBS region, and LTR-RTs showing transposition potential were extracted [15]. The ALE-seq library was analyzed on the Illumina MiSeq platform with 30-bp paired-end reads by Macrogen (Seoul, Korea). The raw ALE-seq data for *O. sativa* ssp. *japonica* cv. Dongjin were deposited in the National Center for Biotechnological Information (NCBI) Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra>) (accession number: SRR17128798).

The adapter sequences and poor-quality reads were filtered from the raw reads using Trimmomatic-0.33 [48]. Paired-end sequences were mapped to the rice reference genome (MSU7.0) using Bowtie2 (-X 3,000) [49]. The number of reads mapped throughout each LTR-RT was counted with the FeatureCounts tool using the customized annotation file generated by *de novo* LTR-RT prediction pipelines [50]. The IGV genome browser was used to visualize the read depth pattern [51].

Validation of active LTR-RTs in Dongjin rice

The hit positions were manually inspected to identify the complete structures of active LTR-RTs. PCR verification was conducted to confirm that the active LTR-RTs in Dongjin were present at the same location in the Dongjin genome as in the reference genome. PCR primer pairs were designed using Primer 3 (<http://bioinfo.ut.ee/primer3-0.4.0/>). Forward primers were located in the flanking region of LTR-RTs, whereas reverse primers were located in the internal parts of LTR-RTs. The PCR products were separated on 1% agarose gels (Additional file 1: Table S8).

BLAST searches were conducted for the putatively active LTR-RTs to examine their transcriptional activity in various tissues or cultivars. BLASTN analysis was performed using the two candidate LTR-RTs as queries against the expressed sequence tag (EST) database using the parameter settings $E\text{-value} < e^{-5}$ and $\text{identity} > 95\%$ (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Abbreviations

TE: transposable element; LTR-RT: long terminal repeat retrotransposon; ecDNA: extrachromosomal linear DNA; ALE-seq: amplification of LTRs of ecDNAs followed by sequencing; EST: expressed sequence tag.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13100-022-00274-2>.

Additional file 1.

Additional file 2.

Acknowledgements

We appreciate Dr. Jee Young Park for assistance of research funding and management.

Authors' contributions

Conceptualization: HK and T-JY; Data analysis: HK; Validation: HK; Resources: S-JL; Writing-original draft: HK, JC and T-JY; Supervision: N-CP, JC and T-JY. All authors read and approved the final manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A4A1032888).

Data Availability

Raw read ALE-seq data are publicly available in the NCBI sequence read archive under accession number SRR17128798.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, College of Agriculture and Life Sciences, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 08826 Seoul, Republic of Korea. ²CAS-JIC Centre of Excellence for Plant and Microbial Science, 200032 Shanghai, China.

Received: 25 January 2022 Accepted: 27 May 2022

Published online: 13 June 2022

References

- Chénais B, Caruso A, Hiard S, Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*. 2012;509(1):7–15.
- Jedlicka P, Lexa M, Vanat I, Hobza R, Kejnovsky E. Nested plant LTR retrotransposons target specific regions of other elements, while all LTR retrotransposons often target palindromes and nucleosome-occupied regions: in silico study. *Mobile DNA*. 2019;10(1):1–14.
- Lanciano S, Cristofari G. Measuring and interpreting transposable element expression. *Nature Reviews Genetics*. 2020;21(12):721–36.
- Zhang Q-J, Gao L-Z. Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* species. *G3: Genes, Genomes, Genetics*. 2017;7(6):1875–85.
- Shahid S, Slotkin RK. The current revolution in transposable element biology enabled by long reads. *Current Opinion in Plant Biology*. 2020;54:49–56.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*. 2007;8(12):973–82.
- Gao L, McCarthy EM, Ganko EW, McDonald JF. Evolutionary history of *Oryza sativa* LTR retrotransposons: a preliminary survey of the rice genome sequences. *Bmc Genomics*. 2004;5(1):1–18.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. *science*. 2009;326(5956):1112–5.
- Kumar A, Bennetzen JL. Plant retrotransposons. *Annual review of genetics*. 1999;33(1):479–532.
- Novoselskaya-Dragovich AY, Fisenko AV, Konovalov FA, Mitrofanova OP, Shishkina AA, Kudryavtsev AM. Analysis of genetic diversity and evolutionary relationships among hexaploid wheats *Triticum L.* using LTR-retrotransposon-based molecular markers. *Genetic resources and crop evolution*. 2018;65(1):187–98.
- Hsu C-C, Chen S-Y, Lai P-H, Hsiao Y-Y, Tsai W-C, Liu Z-J, et al. Identification of high-copy number long terminal repeat retrotransposons and their expansion in *Phalaenopsis* orchids. *BMC genomics*. 2020;21(1):1–13.
- Jouffroy O, Saha S, Mueller L, Quesneville H, Maumus F. Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC genomics*. 2016;17(1):1–15.
- Natali L, Cossu RM, Mascagni F, Giordani T, Cavallini A. A survey of Gypsy and Copia LTR-retrotransposon superfamilies and lineages and their distinct dynamics in the *Populus trichocarpa* (L.) genome. *Tree Genetics & Genomes*. 2015;11(5):1–13.
- Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proceedings of the National Academy of Sciences*. 2014;111(14):5135–40.
- Cho J, Benoit M, Catoni M, Drost H-G, Brestovitsky A, Oosterbeek M, et al. Sensitive detection of pre-integration intermediates of long terminal repeat retrotransposons in crop plants. *Nature plants*. 2019;5(1):26–33.
- Dooner HK, Wang Q, Huang JT, Li Y, He L, Xiong W, et al. Spontaneous mutations in maize pollen are frequent in some lines and arise mainly from retrotranspositions and deletions. *Proceedings of the National Academy of Sciences*. 2019;116(22):10734–43.
- Griffiths J, Catoni M, Iwasaki M, Paszkowski J. Sequence-independent identification of active LTR retrotransposons in *Arabidopsis*. *Molecular Plant*. 2018;11(3):508–11.
- Esposito S, Barteri F, Casacuberta J, Mirouze M, Carputo D, Aversano R. LTR-TEs abundance, timing and mobility in *Solanum commersonii* and *S. tuberosum* genomes following cold-stress conditions. *Planta*. 2019;250(5):1781–7.
- Cho J. Transposon-derived non-coding RNAs and their function in plants. *Frontiers in plant science*. 2018;9:600.
- Fan W, Wang L, Chu J, Li H, Kim EY, Cho J. Tracing Mobile DNAs: From Molecular to Population Scales. *Frontiers in Plant Science*. 2022;13:837378.
- Campo S, Sánchez-Sanuy F, Camargo-Ramírez R, Gómez-Ariza J, Baldrich P, Campos-Soriano L, et al. A novel Transposable element-derived micro-RNA participates in plant immunity to rice blast disease. *Plant biotechnology journal*. 2021;19(9):1798–811.
- Akakpo R, Carpentier MC, le Hsing Y, Panaud O. The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytologist*. 2020;226(1):44–9.
- Ma J, Devos KM, Bennetzen JL. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome research*. 2004;14(5):860–9.
- Anderson SN, Stitzer MC, Brohammer AB, Zhou P, Noshay JM, O'Connor CH, et al. Transposable elements contribute to dynamic genome content in maize. *The Plant Journal*. 2019;100(5):1052–65.
- Mascagni F, Giordani T, Ceccarelli M, Cavallini A, Natali L. Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.). *BMC genomics*. 2017;18(1):1–16.
- Meyers BC, Tingey SV, Morgante M. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Research*. 2001;11(10):1660–76.
- Qiu F, Ungerer MC. Genomic abundance and transcriptional activity of diverse gypsy and copia long terminal repeat retrotransposons in three wild sunflower species. *BMC plant biology*. 2018;18(1):1–8.
- Vukich M, Giordani T, Natali L, Cavallini A. Copia and Gypsy retrotransposons activity in sunflower (*Helianthus annuus* L.). *BMC Plant Biology*. 2009;9(1):1–12.
- Kirov I, Omarov M, Merkulov P, Dudnikov M, Gvaramiya S, Kolganova E, et al. Genomic and Transcriptomic Survey Provides New Insight into the Organization and Transposition Activity of Highly Expressed LTR Retrotransposons of Sunflower (*Helianthus annuus* L.). *International journal of molecular sciences*. 2020;21(23):9331.
- Nie Q, Qiao G, Peng L, Wen X. Transcriptional activation of long terminal repeat retrotransposon sequences in the genome of pitaya under abiotic stress. *Plant Physiology and Biochemistry*. 2019;135:460–8.
- Tiwari SK, Srivastava Y, Singh VK, Singh M, Singh B. LTR retroelement in genes related to abiotic stress in *Capsicum annum* L. *Vegetable Science*. 2017;44(2):1–7.

32. Cavrak W, Lettner N, Jamge S, Kosarewicz A, Bayer LM, Mittelsten Scheid O. How a retrotransposon exploits the plant's heat stress response for its activation. *PLoS genetics*. 2014;10(1):e1004115.
33. Sabot F, Picault N, El-Baidouri M, Llauro C, Chaparro C, Piegou B, et al. Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *The Plant Journal*. 2011;66(2):241–6.
34. Hirochika H. Contribution of the Tos17 retrotransposon to rice functional genomics. *Current opinion in plant biology*. 2001;4(2):118–22.
35. Cheng C, Daigen M, Hirochika H. Epigenetic regulation of the rice retrotransposon Tos17. *Molecular Genetics and Genomics*. 2006;276(4):378–90.
36. Chen J, Lu L, Benjamin J, Diaz S, Hancock CN, Stajich JE, et al. Tracking the origin of two genetic components associated with transposable element bursts in domesticated rice. *Nature communications*. 2019;10(1):1–10.
37. Biémont C, Vieira C. Junk DNA as an evolutionary force. *Nature*. 2006;443(7111):521–4.
38. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*. 2007;41:331–68.
39. Pearce S, Knox M, Ellis T, Flavell A, Kumar A. Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Molecular and General Genetics MGG*. 2000;263(6):898–907.
40. Jiménez-Ruiz J, Ramírez-Tejero JA, Fernández-Pozo N, Leyva-Pérez MdIO, Yan H, Rosa Rdl, et al. Transposon activation is a major driver in the genome evolution of cultivated olive trees (*Olea europaea* L.). *The Plant Genome*. 2020;13(1):e20010.
41. Qin Y, Shin K-S, Woo H-J, Lim M-H. Genomic variations of rice regenerants from tissue culture revealed by whole genome re-sequencing. *Plant Breeding and Biotechnology*. 2018;6(4):426–33.
42. Zhang R-G, Wang Z-X, Ou S, Li G-Y. TESorter: lineage-level classification of transposable elements using conserved protein domains. *Biorxiv*. 2019:800177.
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013;30(4):772–80.
44. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*. 2015;32(1):268–74.
45. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*. 1980;16(2):111–20.
46. Ma J, Bennetzen JL. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences*. 2004;101(34):12404–10.
47. Jeon J-S, Chung Y-Y, Lee S, Yi G-H, Oh B-G, An G. Isolation and characterization of an anther-specific gene, RA8, from rice (*Oryza sativa* L.). *Plant molecular biology*. 1999;39(1):35–44.
48. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357–9.
50. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
51. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*. 2013;14(2):178–92.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

