Identification of Transposable Elements in Conifer and Their Potential Application in Breeding

Junhui Wang, Nan Lu, Fei Yi and Yao Xiao

State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China.

Evolutionary Bioinformatics
Volume 16: 1–4

DOI: 10.1177/1176934320930263
SAGE

ABSTRACT: Transposable elements (TEs) are known to play a role in genome evolution, gene regulation, and epigenetics, representing potential tools for genetics research in and breeding of conifers. Recently, thanks to the development of high-throughput sequencing, more conifer genomes have been reported. Using bioinformatics tools, the TEs of 3 important conifers (*Picea abies, Picea glauce*, and *Pinus taeda*) were identified in our previous study, which provided a foundation for accelerating the use of TEs in conifer breeding and genetic study. Here, we review recent studies on the functional biology of TEs and discuss the potential applications for TEs in conifers.

KEYWORDS: Transposable elements, conifer breeding, mutant library, molecular marker

RECEIVED: March 16, 2020. **ACCEPTED:** May 4, 2020.

TYPE: Commentary

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Fundamental Research Funds of Chinese Academy of Forestry (CAFYBB2019ZX001); Thirteenth Five-Year Plan for Key & Research Project of China (2017YFD0600606-09).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Junhui Wang, State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of State Forestry Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China. Email: wangjh808@sina.com

COMMENT ON: Yi F, Ling J, Xiao Y, Zhang H, Ouyang F, Wang J. ConTEdb: a comprehensive database of transposable elements in conifers [published online ahead of print January 1, 2018]. *Database (Oxford)*. 2018;2018:bay131. doi:10.1093/database/bay131. PubMed PMID: 30576494. PubMed Central PMCID: PMC6301336. https://academic.oup.com/database/article/doi/10.1093/database/bay131/5255192

Transposable Element Insertions Influence Plant Traits

Transposons are DNA sequences that are autonomously duplicated and can "jump" out from one location on the genome and be reinserted at another place. Transposons randomly shift to different loci in the genome, which can alter the original gene structure and lead to gene silencing or inactivation, consequently changing the traits under the regulation of the functional genes.[1] For example, Zhang et al[2] found a 27.2 kb transposable element insertion in the promoter of the *stiff1* gene of maize that represses the transcription of *stiff1* and, consequently, enhances stalk strength by increasing the cellulose and lignin contents in the cell wall. Sometimes, when a transposon is inserted into the coding region of a gene, it modifies an important functional domain or splice site, consequently changing the transcript. Although the gene could still be translated into a protein, the original function may become weakened or lost. Moreover, some transposable elements (TEs) directly regulate a functional gene as a promoter.[3]

Recent studies have also suggested that plant transposons have important biological functions in gene expression regulation through the generation of noncoding RNAs.[4,5] Most long non-coding RNA (lncRNAs) originated from intergenic regions where transposons were abundantly distributed, and some microRNA (miRNAs) and short interfering RNAs (siRNAs) were directly or indirectly derived from transposons. These noncoding RNAs are known to regulate genes in plants. Because TE insertions could influence gene expression and function and eventually cause phenotypic variances between individuals, they have received attention for their potential value in the study of genetics.

Transposable Elements Participate in Genome Evolution

In addition to genetic association studies, TEs have also been used the study of evolutionary processes. For example, miniature inverted-repeat elements (MITEs) were considered the driving force of genomic evolution, and the number of MITEs was significantly and positively correlated with genome size. Because of their abundant variety of copy numbers and insert sites and highly species-specific characteristics, transposons have great potential in the study of gene function and genome evolution.[6] Tang et al identified an active MITE "*mJing*" that is associated with the *high-tillering dwarf* mutant in rice; moreover, they also found that the copy number of *mJing* dramatically varies among *Oryza sativa* (Asian cultivated rice), *Oryza rufipogon* (wild ancestor of Asian cultivated rice), and *Oryza glaberrima* (African cultivated rice), suggesting that the amplification or contraction of *mJing* elements may play an important role in genome evolution and species diversification in rice.[7]

Interestingly, when plants experience biotic and abiotic stresses, many TEs in the plant genome are specifically activated and probably transposed under the regulation of methylation or transposons themselves, which has been found in tobacco, tomato, and other plants.[8] Some researchers have suggested that the activation of TEs under stress may accelerate genome recombination, which may also reflect the environmental adaptation process of plants during the particular lengthy evolutionary history of species.

Potential Application of TEs in Conifer Breeding

Conifers (Coniferales) are the largest and most important group of gymnosperms, as well as the most archaic branch. They have

a total of 6 families, 69 genera, and 605 species. Although the number of species is limited, most of the important tree species belonging to terrestrial ecosystems evolved from conifers, placing conifers in an important taxonomic status and were usually used for the study of plant evolution history.[9] Moreover, conifers have been regarded as dominant species in the forest for more than 200 million years, not only for supporting large industries by providing wood, fiber, and energy but also for making great contributions to the global ecosystem. Due to their high ecological significance and economic importance, the genetic breeding of conifers has become an important research subject worldwide. With the rapid development of genome sequencing technology, an increasing number of high-quality plant genomes have been reported, from which researchers could get genetic data that could be used for the study of genetic mechanisms and for identifying key genes associated with important economic traits. These genomic resources could be a great resource when making breeding plans and, in addition, accelerate the process of genetic breeding in plants. Although several plant whole-genome sequences have been reported, the sequences available from conifers are limited due to their extremely large genomes and assembly difficulty. Examples of such conifers include *Pinus lambertiana* (~31 Gb), *Pinus taeda* (~22 Gb), and *Picea abies* (~20 Gb).[10-12] Their genomes contain considerable amounts of repetitive DNA, which contributes to the large size of conifers' genomes and, especially, a high number of TEs, which provide vast potential genetic resources for further study.

In a previous study, we reported our work on the identification and classification of TEs in 3 important conifers, *P abies, Picea glauce*, and *P taeda*, at the whole-genome level.[13] In total, 413 423 TEs were identified from the 3 conifers, and further classification suggested a great variety among the 3 species (Table 1). Further identification of the genetic functions or association with the economic characteristics will facilitate breeding. Identifying functional genes associated with interesting traits through the construction of transposon-based mutation libraries and the development of TEs as molecular markers that could be used for assisted breeding are 2 main methods by which TEs could be used in plant breeding. Related research has been carried out in other economic plants.[14,15]

Transposon-based mutation libraries are important tools for the study of functional genomics in economic crops, including *Oryza sativa*[16] and *Zea maize*.[17] Currently, studies have been conducted in which exogenous TEs have been inserted into the plant genome to create mutants using genome walking.[18] *Tnt1, Activator/Dissociation* (*Ac/Ds*), and other TEs are traditional transposon systems used in the construction of mutation libraries. For example, *Tnt1* was originally found in *Nicotiana tabacum* and is usually not active in common plant tissues; however, it can be activated in other plant species via tissue culture. Moreover, it can be stably inherited and re-transpose during tissue culture.[19] In a recent study, Sun et al[20] found that *Tnt1* insertions are basically randomly distributed throughout the genome and can be identified on both arms of chromosomes in a *Tnt1*-based *Medicago truncatula* mutation library. Moreover, they found that the *Tnt1* insertion is positively correlated with gene methylation and negatively correlated with CG content in *M truncatula*, suggesting a special tendency of *Tnt1* insertions. Zhou et al[21] identified 2 mariner-like elements (*MLEs*), *Ppmar1* and *Ppmar2*, from moso bamboo that contain terminal inverted repeats. Furthermore, heterologous transfer of the 2 *MLE*s to *Arabidopsis thaliana* suggested that they could also transpose in other plants. However, *Tnt1* must be activated under in vitro tissue culture[22]; therefore, a stable and effective tissue culture system in conifers must be constructed first. Another *Activator/Dissociation* (*Ac/Ds*) element comes from maize and is a classic maize transposon method that is used for constructing mutation libraries. Both elements can duplicate and transpose throughout the plant genome. Previous studies have suggested that *Ac/Ds* can insert into gene-rich regions and alter the expression of genes or the genome through the formation of unstable insertion alleles, stable derivatives, or excision alleles.[23–25] Until now, *Ac/Ds* transposons have been shown to be effective in approximately 20 plant species and have become the most widely used TEs for gene tagging and functional genomics studies in plants. For example, Carter et al[25] produced activation-tagged and knockout mutants in the processing of tomato cultivar M82 using the *Ac/Ds* system and agrobacterium-mediated method. Finally, a population of 509 independent transposed lines with only *Ds* insertions throughout all 12 chromosomes was developed. Wang et al[26] established an *Ac*-based transposon system with the donor *Ac* tightly linked with sugary1 (su1) on maize chromosome 4S, and a total of 208 independent long-distance-transposed *Ac* lines were identified. This *Ac*-based system combined with 2 other *Ac*-based regional mutagenesis systems in maize could be integrated into a new maize mutation library with a more even and high-density distribution of *Ac* elements throughout the genome for functional genomics studies.

Molecular markers are an important tool for genetic studies and breeding, the construction of linkage maps, and population genetics analysis. Transposable elements constitute most of the repetitive sequences and are distributed throughout the genome in plants; thus, TE could have great potential to be explored as molecular markers. Liang et al[27] identified 336 long terminal repeats (LTR) effective primer pairs that could amplify DNA from *P virginiana*, 117 of which could produce amplification in other *Prunus* species; in addition, 59 of the LTR markers were used to construct a linkage map after qualification. Sequence-characterized amplified region (SCAR) markers derived from transposons are another TE-based molecular markers that are notorious for being low time consuming and highly efficient. Roy et al[28] developed 108 CACTA transposable element-derived SCAR markers in maize, 32 of which could be integrated into the genetic map of a Recombinant Inbred Lines (RIL) population; in addition, 76 markers could

Table 1. Summary of identified TEs in the *Picea abies, Picea glauce*, and *Pinus taeda* genomes.[13]

CLASS	ORDER	SUPERFAMILY	*P ABIES* MEMBERS/FAMILIES	*P GLAUCE* MEMBERS/FAMILIES	*P TAEDA* MEMBERS/FAMILIES
Retrotransposons	LTR	*Caulimovirus*	51/17	196/43	315/35
		Copia	7304/78	35 826/89	40 645/92
		Copia(*Xen1*)			26/5
		DIRS	76/51	226/113	287/143
		ERV			7/3
		ERV1	299/124	749/231	1392/182
		ERV4	25/21	44/39	145/46
		ERVK	153/88	563/187	584/169
		ERVL	28/22	105/71	115/75
		Gypsy	12 267/129	64 831/396	58 349/113
		Ngaro	29/24	63/41	126/63
		Pao	241/84	832/229	874/179
		RUnknown	9225/941	42 750/3435	43 951/1147
	LINE (long interspersed nuclear element)	*L1*	5230/83	11 150/40	24 553/27
		PTE-X	4/4		
		Tad1	4/4		
Subtotal			34 936/1670	157 335/4914	171 369/2279
DNA transposons	TIR (terminal inverted repeat)	*hAT*	8/8	12/12	7/7
		TcMar	3/3		3/3
		PIF-Harbinger		3/3	4/4
		CMC			3/3
		DUnknown	7/7	12/12	2/2
	MITE	*MITE*	378/277	287/261	390/297
	Helitron	*Helitron*	6575/609	21 869/359	20 220/403
Subtotal			6971/904	22 183/647	20 629/719
Total			41 907/2574	179 518/5561	191 998/2998

Abbreviations: LINE, long interspersed nuclear element; LTR, long terminal repeats; MITE, miniature inverted-repeat element; TE, transposable elements; TIR, terminal inverted repeat.

be used for diversity analysis of different corn lines. Inter-retrotransposon amplified polymorphisms (IRAP) are another important type of molecular marker that were developed from polymorphisms in the conserved region of retrotransposon sequences from the same or allied species. Cui et al[29] developed 29 stable and polymorphic amplifications in *Pinus massoniana*, according to the conservation region of the reverse transcriptase sequences of *Ty1-copia* and *Ty-gypsy*-type retrotransposons. These IRAP markers were effectively used in the identification of germplasm and genetic relationships, as well as further genetic breeding in *P massoniana*. Moreover, because LTRs are usually conserved and widely distributed in the plant genomes, they can be used for developing TE junction-based markers. The study of LTRs has provided a great example of molecular marker development in conifers.

Conclusions

TEs have great potential to be used in conifer breeding; however, a limited number of related studies have been reported. In future studies, mutant libraries of conifers could be constructed

to identify essential genes associated with target traits, according to the genotype and phenotype variations in the library. Moreover, TE-based molecular markers could be developed for use in molecular breeding in conifers. It is important to mention that the massive genome sizes and relatively long life cycles of conifers make the use of TEs in conifer breeding more challenging than in annual crop breeding and much foundational work is needed.

Acknowledgements

The authors would like to thank Thirteenth Five-Year Plan for Key & Research Project of China (2017YFD0600606-09) for the financial support.

Author Contributions

JW, NL, FY, and YX composed, reviewed, and approved the final article.

REFERENCES

1. Saxena RK, Edwards D, Varshney RK. Structural variations in plant genomes. *Brief Funct Genomics*. 2014;13:296-307.
2. Zhang Z, Zhang X, Lin Z, et al. A large transposon insertion in the stiff1 promoter increases stalk strength in maize. *Plant Cell*. 2020;32:152-165.
3. Butelli E, Licciardello C, Zhang Y, et al. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell*. 2012;24:1242-1255.
4. Wang X, Ai G, Zhang C, et al. Expression and diversification analysis reveals transposable elements play important roles in the origin of Lycopersicon-specific lncRNAs in tomato. *New Phytol*. 2016;209:1442-1455.
5. Petri R, Brattås PL, Sharma Y, et al. LINE-2 transposable elements are a source of functional human microRNAs and target sites. *PLoS Genet*. 2019;15:e1008036.
6. Liu Y, ul Qamar MT, Feng JW, et al. Comparative analysis of miniature inverted-repeat transposable elements (MITEs) and long terminal repeat (LTR) retrotransposons in six Citrus species. *BMC Plant Biol*. 2019;19:140.
7. Tang Y, Ma X, Zhao S, et al. Identification of an active miniature inverted-repeat transposable element mJing in rice. *Plant J*. 2019;98:639-653.
8. Negi P, Rai AN, Suprasanna P. Moving through the stressed genome: emerging regulatory roles for transposons in plant stress response. *Front Plant Sci*. 2016;7:1448.
9. Christenhusz MJ, Reveal JL, Farjon A, Gardner MF, Mill RR, Chase MW. A new classification and linear sequence of extant gymnosperms. *Phytotaxa*. 2011;19:55-70.
10. Gonzalez-Ibeas D, Martinez-Garcia PJ, Famula RA, et al. Assessing the gene content of the megagenome: sugar pine (Pinus lambertiana). *G3*. 2016;6:3787-3802.
11. Neale DB, Wegrzyn JL, Stevens KA, et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol*. 2014;15:R59.
12. Nystedt B, Street NR, Wetterbom A, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 2013;497:579-584.
13. Yi F, Ling J, Xiao Y, Zhang H, Ouyang F, Wang J. ConTEdb: a comprehensive database of transposable elements in conifers. *Database*. 2018;2018:bay131.
14. Mielich K, Shtifman-Segal E, Golz J, et al. Maize transposable elements Ac/Ds as insertion mutagenesis tools in Candida albicans. *G3*. 2018;8:1139-1145.
15. Zhao G, Zou C, Li K, et al. The Aegilops tauschii genome reveals multiple impacts of transposons. *Nat Plants*. 2017;3:946-955.
16. Ram H, Soni P, Salvi P, et al. Insertional mutagenesis approaches and their use in rice for functional genomics. *Plants*. 2019;8:310.
17. Liang L, Zhou L, Tang Y, et al. A sequence-indexed mutator insertional library for maize functional genomics study. *Plant Physiol*. 2019;181:1404-1414.
18. He P, Ma Y, Zhao G, et al. FaRE1: a transcriptionally active Ty1-copia retrotransposon in strawberry. *J Plant Res*. 2010;123:707-714.
19. Cui Y, Barampuram S, Stacey MG, et al. Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. *Plant Physiol*. 2013;161:36-47.
20. Sun L, Gill US, Nandety RS, et al. Genome-wide analysis of flanking sequences reveals that Tnt1 insertion is positively correlated with gene methylation in Medicago truncatula. *Plant J*. 2019;98:1106-1119.
21. Zhou M, Hu H, Liu Z, Tang D. Two active bamboo mariner-like transposable elements (Ppmar1 and Ppmar2) identified as the transposon-based genetic tools for mutagenesis. *Mol Breeding*. 2016;36:163.
22. Courtial B, Feuerbach F, Eberhard S, et al. Tnt1 transposition events are induced by in vitro transformation of Arabidopsis thaliana, and transposed copies integrate into genes. *Mol Genet Genomics*. 2001;265:32-42.
23. Bregitzer P, Brown RH, Dahleen LS, et al. Registration of the barley transposon-tagged population II: sixty-one lines each characterized for *Ds* insertion sites. *J Plant Regist*. 2019;13:109-112.
24. Hayashi H, Czaja I, Lubenow H, Schell J, Walden R. Activation of a plant gene by T-DNA tagging: auxin-independent growth in vitro. *Science*. 1992;258:1350-1353.
25. Carter JD, Pereira A, Dickerman AW, Veilleux RE. An active Ac/Ds transposon system for activation tagging in tomato cultivar M82 using clonal propagation. *Plant Physiol*. 2013;162:145-156.
26. Wang F, Li Z, Fan J, et al. An Ac transposon system based on maize chromosome 4S for isolating long-distance-transposed Ac tags in the maize genome. *Genetica*. 2010;138:1261-1270.
27. Liang Y, Lenz RR, Dai W. Development of retrotransposon-based molecular markers and their application in genetic mapping in chokecherry (Prunus virginiana L.). *Mol Breeding*. 2016;36:109.
28. Roy NS, Park KC, Lee SI, Im MJ, Ramekar RV, Kim NS. Development of CACTA transposon derived SCAR markers and their use in population structure analysis in Zea mays. *Genetica*. 2018;146:1-12.
29. Cui BW, Fan FH, Ding GJ, Yang ZQ, Wen XP. Development of IRAP markers based on genomic LTR retrotransposon sequences in Masson Pine (Pinus massoniana). *For Res*. 2016;29:348-353.