



## Journal of Experimental Biology and Agricultural Sciences

<http://www.jebas.org>

ISSN No. 2320 – 8694

### COMPARATIVE ANALYSIS OF TRANSPOSABLE ELEMENTS FROM *Glycine max*, *Cajanus cajan* and *Phaseolus vulgaris*

Gaurav Singh<sup>1\*</sup>, Milind Ratnaparkhe<sup>2</sup>, Anil Kumar<sup>1</sup>

<sup>1</sup>School of Biotechnology, Devi Ahilya University, Khandwa Rd., Indore-452001, India

<sup>2</sup>Indian Institute of Soybean Research (ICAR-IISR), Khandwa Rd., Indore-452001, M.P. India

Received – February 20, 2019; Revision – March 23, 2019; Accepted – March 29, 2019

Available Online – April 10, 2019

DOI: [http://dx.doi.org/10.18006/2019.7\(2\).167.177](http://dx.doi.org/10.18006/2019.7(2).167.177)

#### KEYWORDS

Transposable elements

*Copia*, *gypsy*

Blast

Phylogenetic analysis

*Glycine max*

*Cajanus cajan*

*Phaseolus vulgaris*

#### ABSTRACT

Transposable elements (jumping genes) are the DNA sequences that jumps (or move) from one genomic region to another. These are found in all living species and are genomically diverse due to their mobility. These are active and can relocate themselves to different locations of the genome. Here, a study has been carried out to identify transposable elements in *Glycine max* (soybean), *Cajanus cajan* (pigeon pea) and *Phaseolus vulgaris* (common bean/green bean/French bean) using previously known elements. In present study homology method of transposable elements identification was used. The identified transposable elements have also been characterized using phylogenetic analysis. Different tools such as Uniprot, BLAST, MSA and PHYLIP were used for this study. A total 12 TEs were taken in *G. max*, *C. cajan* and *P. vulgaris*; these were selected from vicinity of Rpg1b on chromosome 13 of *G. max*. NCBI BLAST was done for all 12 TEs. Out of 12 TEs 6 were copia, 5 are gypsy and 1 LINE. It is found that *copia* and *gypsy* are closely related transposable elements. It is revealed that these TEs have high sequence similarity with the transposons from *C. cajan* and *P. vulgaris* indicating that these elements were present before the divergence of these three legumes.

\* Corresponding author

E-mail: [singhgauravbiotech@gmail.com](mailto:singhgauravbiotech@gmail.com) (Gaurav Singh)

Peer review under responsibility of Journal of Experimental Biology and Agricultural Sciences.

Production and Hosting by Horizon Publisher India [HPI]  
(<http://www.horizonpublisherindia.in/>).  
All rights reserved.

All the article published by [Journal of Experimental Biology and Agricultural Sciences](#) is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#) Based on a work at [www.jebas.org](http://www.jebas.org).



## 1 Introduction

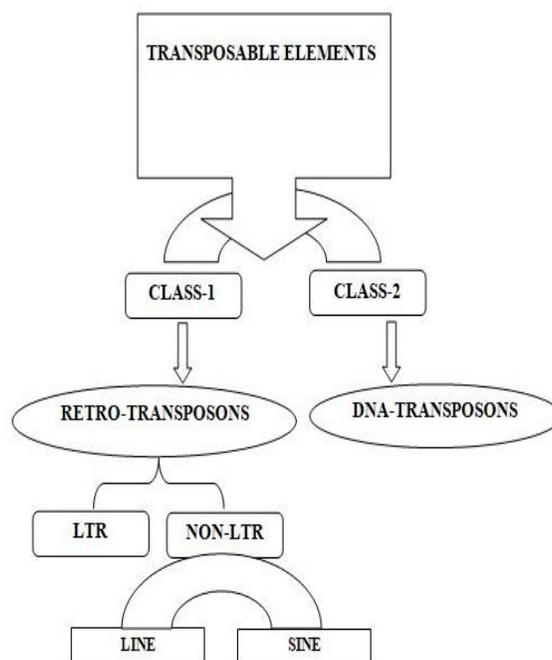
Transposable elements (TEs), also known as transposons or mobile genetic elements, that can move from one position to another in the genome. These are widely found in prokaryotes as well as eukaryotes including both animals and plants (Kolade et al., 2015; Grzebelus, 2018). Due to their mobility, they create genomic diversity and therefore are useful for phylogenetic analysis (Baranek et al., 2012; Nemli et al., 2015). They are not only contributing in genome organization but also play an important role in the regulation of gene expression (Kashkush et al., 2003). These also generate raw materials for evolution of novel genes and new genetic functions (Du et al., 2010). Transposable elements have been used for crop improvement in many plants *viz.* rice, tomatoes, maize and sorghum due to their mobile nature (Staginnus et al., 2001; Wessler 2006). There are various groups of TEs and their classification is the subject of an ongoing debate. Finnegan (1989) proposed the classification of TEs on the basis of transposition intermediate and their mechanisms of transpositions in eukaryotes, accordingly TEs have been classified into two types (Figure 1) namely Class 1 and 2, in which Class 1 elements (Retrotransposons) moves by an RNA intermediate, generated through reverse transcription and moves in a “copy and paste” manner before it is inserted into the genome (Wessler, 2006; Kolade et al., 2015; Jiang et al., 2015). The Retrotransposons are traditionally divided into two sub- families; the long terminal repeats (LTRs) and Non-LTR retrotransposons

(Patil et al., 2015). In plants LTR retrotransposons especially those consist of Ty1-copia and Ty3-gypsy superfamilies generally constitute the major fraction of all plant TEs (Zhang et al., 2017; Grzebelus, 2018). The Non-LTR retrotransposons are classified into LINES (Long Interspersed Elements) and SINEs (Short Interspersed Elements) (Gao et al., 2014).

Class 2 elements (DNA transposons) transpose in a “cut and paste” manner without an RNA intermediate (Kolade et al., 2015). On the basis of their structural characteristics and transposase resemblance, DNA transposons have been classified into 7 groups namely Tc1/Mariner, Mutator, CACTA, Hat, miniature inverted repeat TEs (MITES), Pong and PIF/Harbinger (Wicker et al., 2007; Fedoroff, 2012; Zhao & Ma, 2017).

Soybean (*Glycine max*) is an important legume crop globally due to its nutritional value and industrial importance. The current availability of genome sequence of soybean (Schmutz et al., 2010) has provided a remarkable chance for discovery and characterization of transposable elements in this crop. In the last few decade, functional characterization of soybean genes was done successfully by utilizing TEs (Sandhu & Bhattacharyya, 2017).

Various endogenous TEs in soybean have been identified and characterized (Zabala & Vodkin, 2008; Zabala & Vodkin, 2014, Xu et al., 2009). Among various identified TEs, Tgm9 is the only active transposon used in gene tagging experiments



**Figure 1** Transposable Elements and their types

(LTR, long terminal repeat; Non-LTR, non-long terminal repeat; LINE, long interspersed elements; SINE, short interspersed elements)

(Xu et al., 2009; Sandhu et al., 2017).

Similar to *Glycine max*, current approaches in sequencing of *Cajanus cajan* whole genome also have resulted in immense increase in genomic resources for this crop (Singh et al., 2011; Varshney et al., 2012). Sequencing of *C. cajan* genome unveiled a large number of TEs that have contributed in its genome organization (Wicker et al., 2007). It was reported that the total size of transposable elements in pigeon pea is 63.95% of total available genome sequence (Singh et al., 2011). In previous studies, various strategies were practiced for identification of retrotransposon sequences in un-sequenced plant genomes. Pearce et al. (1999) designed a method based on PCR to identify Ty/copia- LTR sequences in higher plants and isolation of new LTR sequences from *Pisum sativum*, *Vicia faba* and *Picea abies* (Patil et al., 2015). Only one such Panzee Ty/copia like element was described in pigeon pea before its whole genome sequence was available (Lall & Upadhyaya, 2002). Patil et al., (2015) showed that *in-silico* homology searching is much more efficient for the identification and characterization of both homologous and novel retrotransposon families in *C. cajan* genome.

*Phaseolus vulgaris* (Common bean) is also an essential legume crop along with *Glycine max* (Soybean) and *Cajanus cajan* (Pigeonpea) which is grown globally. Its plant genome contains plenty of TEs that affect genetic variation and genome evolution (Gao et al., 2014). *P. vulgaris* is considered to be closely related to *G. max* and thus utilized for the study of domestication, investigating polyploidy effects and evolutionary studies in legumes (Lin et al., 2010; McClean et al., 2010; Gao et al., 2014).

Despite of the fact that various computational tools like BLASTER suit, Censor, HMMER, LTR\_FINDER, MAK, REPuter, TRANSPO etc. have been developed which have advanced the process of identification of transposons (Bergman & Quesneville 2007), annotation of these abundant elements is still a time taking and challenging task (Gao et al., 2014). Transposon databases are significant resources for plants fundamental as well as applied research (Ouyang & Buell 2004; Du et al., 2010; Anderson et al., 2019). These databases are beneficial for the annotation of genomes and also for genome evolution study (Panaud, 2009). A whole genome transposon annotation in *P. vulgaris* by using homology methods was performed by Gao et al. (2014). A soybean transposable element database (SoyTEdb) has been established in the last decade (Du et al., 2010), that gives the information and resources related to transposable elements in soybean genome. Recently Jiang et al. (2015) developed a novel bioinformatics tool, ITIS (identification of transposon insertion sites) for localizing transposon insertion sites in a genome.

Here, in present study, author have identified few TEs in few legumes crops such as soybean, pigeon pea and common bean

using the homology methods and characterized these with the help of phylogenetic (evolutionary) studies. These TEs are near nucleotide binding – leucine rich repeats (NB-LRR) genes, which can be used for developing disease resistance varieties in legume plants and consequently helps to enhance yield as well as quality of legumes. TEs also contribute in epigenetic changes, which response to stress factors generating phenotypic plasticity that can be subject to selection resulting in adaptation to changed environment so this study will be useful for creating more tolerant legume varieties to a range of abiotic stresses.

## 2 Materials and Methods

Different tools were used for this study namely UniProt for the collection of functional information on proteins, UniProt BLAST (<https://www.uniprot.org/blast/>) for finding regions of local similarity between sequences, ClustalW version 2.1 (<http://www.clustal.org/clustal2/>) is used for multiple sequence alignment (MSA) and PHYLIP 3.696 version (<http://evolution.genetics.washington.edu/phylip.html>) was used for the work. The calculation of the Jones-Taylor-Thornton distance matrix, the phylogenetic analysis and bootstrapping (1000 cycles) were performed with the programs PROTDIST, NEIGHBOR and SEQBOOT, respectively, of the program package PHYLIP (3.696).

Data for the analysis of TEs were taken from already published paper of Wawrzynski et al. (2008). A total twelve transposable elements of *G. max* were taken for study and compared with *C. cajan* and *P. vulgaris*. To find the best similar hits for these TEs, Uniprot BLAST was performed. The best hits were taken for further analysis except LINE. The best hits of BLAST include 13 and 8 unique sequences for copia and gypsy respectively. These unique sequences were used for phylogenetic analysis using PHYLIP. For making phylogenetic tree, combination of unique sequences of copia and gypsy with *Arabidopsis thaliana* copia sequence (Uniprot id: Q9C536) as outgroup was used to make phylogenetic tree. The tree was constructed using Neighbor joining (NJ) method and bootstrap analysis was performed using 1000 replicates. All the combinations are mentioned in Table 1.

Table 1 Combination for making Phylogenetic tree

S.No.	Combination
Tree 1	<i>Glycine max</i>
Tree 2	<i>Phaseolus vulgaris</i>
Tree 3	<i>Cajanus cajan</i>
Tree 4	<i>G. max</i> + <i>P. vulgaris</i>
Tree 5	<i>G. max</i> + <i>C. cajan</i>
Tree 6	<i>P. vulgaris</i> + <i>C. cajan</i>
Tree 7	<i>G. max</i> + <i>P. vulgaris</i> + <i>C. cajan</i>

### 3 Results

These transposable elements (TEs) were selected from vicinity of Rpg1b gene on chromosome number 13 of *G. max*. BLAST was carried out for all the 12 transposable elements that were taken from the literature. The results of UniProt Blast are shown in Table 2. The results showed that 13 unique sequences for copia and 8 unique sequences for gypsy are the best hits; which demonstrate that these sequences are similar to already reported

TEs. Out of 12 TEs, 6 are copia, 5 are gypsy and 1 LINE.

The results revealed that these TEs have high sequence similarity with the transposons from *C. cajan* and *P. vulgaris*; further indicating that these elements were present before the divergence of these three legumes. The best hits of BLAST were used for further analyses. The multiple sequences were aligned using CLUSTALW and trees were drawn using different combinations as mentioned in Table 1.

Table 2 Details of different transposable elements in few legumes

S. No.	Accession no. with name	Type	Protein	Organism	Uniprot ID
<i>Glycine max</i>					
1.	FJ198019.1 Retrotransposon gmw1-52d1-re-3	Ty3/gypsy	Polyprotein	<i>Glycine max</i>	Q84ZV5
			Transposon Ty3-I Gag-Pol polyprotein	<i>Cajanus cajan</i>	A0A151S548
				<i>Phaseolus vulgaris</i>	NOT FOUND
2.	FJ198017.1 gmw1-27d20-re-1	Ty3/gypsy	Polyprotein	<i>G. max</i>	Q84ZV5
			Retrotranspos Tf2	<i>C. cajan</i>	A0A151S3E7
				<i>P. vulgaris</i>	NOT FOUND
3.	FJ198015.1 gmw2-129e12-re-3	Copia	Copia-type polyprotein	<i>G. max</i>	C0JJI2
			Retrovirus-related Pol polyprotein from transposon TNT 1-94	<i>C. cajan</i>	A0A151QUV8
			Gag-pol polyprotein	<i>P. vulgaris</i>	Q69F89
4.	FJ198011.1 gmw2-10n21-re-5	Copia	Gag-protease-integrase-RT-RNaseH polyprotein	<i>G. max</i>	B5U9F7
			Retrovirus-related Pol polyprotein from transposon TNT 1-94	<i>C. cajan</i>	A0A151QU04
			Putative Ty-1 copia retrotransposon	<i>P. vulgaris</i>	A0A077SK66
5.	FJ402890.1 gmp1-95h18-re-1	LINE-1	Reverse transcriptase family member	<i>G. max</i>	Q1AKH8
			Retrovirus-related Pol polyprotein LINE-1	<i>C. cajan</i>	A0A151SQ16
				<i>P. vulgaris</i>	Not available
6.	FJ197993.1 gmp1-71h23-re-7	Ty3/gypsy	Polyprotein	<i>G. max</i>	Q84ZV5
			Transposon Ty3-I Gag-Pol polyprotein	<i>C. cajan</i>	A0A151SMN1
				<i>P. vulgaris</i>	NOT FOUND
7.	FJ197985.1 gmp1-125b19-re-9	Ty3/gypsy	Reverse transcriptase	<i>G. max</i>	Q946S9
			Transposon Ty3-I Gag-Pol polyprotein	<i>C. cajan</i>	A0A151QTE7
				<i>P. vulgaris</i>	NOT FOUND

Table 2 Details of different transposable elements in few legumes

S. No.	Accession no. with name	Type	Protein	Organism	Uniprot ID
<i>Phaseolus vulgaris</i>					
8.	FJ402928.1 pva1-144m6-re-2	Copia	Copia-type polyprotein	<i>G. max</i>	C0JJI2
			Retrovirus-related Pol polyprotein from transposon TNT 1-94	<i>C. cajan</i>	A0A151QZS0
			Polyprotein	<i>P. vulgaris</i>	A0A059QBK0
9.	FJ402929.1 pva1-144m6-re-5	Copia	Copia-type polyprotein	<i>G. max</i>	C0JJI2
			Retrovirus-related Pol polyprotein from transposon TNT 1-94	<i>C. cajan</i>	A0A151RYL1
			Putative Ty-1 copia retrotransposon	<i>P. vulgaris</i>	A0A077SK66
10.	FJ402924.1 pva1-47b16-re-5	Ty3/gypsy		<i>G. max</i>	NOT FOUND
			Transposon Ty3-I Gag-Pol polyprotein	<i>C. cajan</i>	A0A151SQK2
			Phytoalexin-deficient 4-2 protein	<i>P. vulgaris</i>	I7A4H7
<i>Glycine tomentella</i>					
11.	FJ402921.1 gtt1-296j12-re-1	Copia	Gag-pol polyprotein	<i>G. max</i>	Q84VI4
			Retrovirus-related Pol polyprotein from transposon TNT 1-94	<i>C. cajan</i>	A0A151SK87
			Phytoalexin-deficient 4-2 protein	<i>P. vulgaris</i>	I7A4H7
12.	FJ402920.1 gtt1-158p11-re-2	Copia		<i>G. max</i>	Not available
			Copia protein	<i>C. cajan</i>	A0A151RG27
			Uncharacterized protein	<i>P. vulgaris</i>	V7CY32

To differentiate copia and gypsy transposable elements, trees were drawn without combination. Figure 2 (a, b, c) depicts the phylogenetic tree of transposable elements (copia and gypsy) for *Glycine max*, *Phaseolus vulgaris* and *Cajanus cajan*, respectively.

### 3.1 Phylogenetic tree for *Glycine max*

The Neighbor Joining (NJ) method was used for constructing tree. For that, tree bootstrapping was done, and for each sample, *A. thaliana* was taken as an out group. All the five sequences were compared 500 times to outgroup. As shown in Figure 2a, C0JJI2\_SOYBN was compared to all the four sequences. The results showed that all the four sequences were grouped 969 times together. It means this group may share common ancestors. In the same way, the Q84VI4\_SOYBN was compared with all the three remaining sequence which showed that these sequences were grouped (557 times) together. When the comparison of

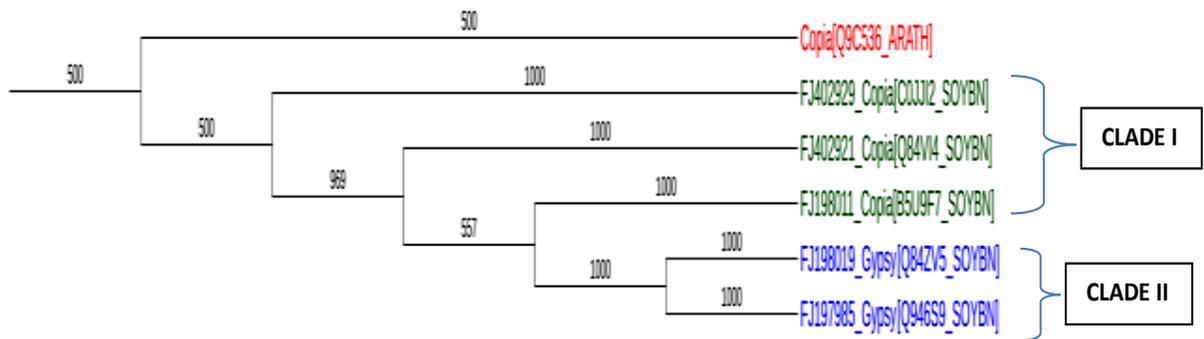
B5U9F7\_SOYBN was done with the remaining two sequences, the outcome showed that these were arranged together. Similarly the remaining sequences were compared. The figure 2a depicts that three sequences viz., B5U9F7\_SOYBN, Q84ZV5\_SOYBN and Q946S9\_SOYBN are closely related. It is to be noted that in bootstrapping only those group whose results were between 75-95% were considered.

There are two major clades, clade I and clade II; clade I includes copia while clade II includes gypsy elements. Within clade I, three copia while in clade II two gypsy elements are present.

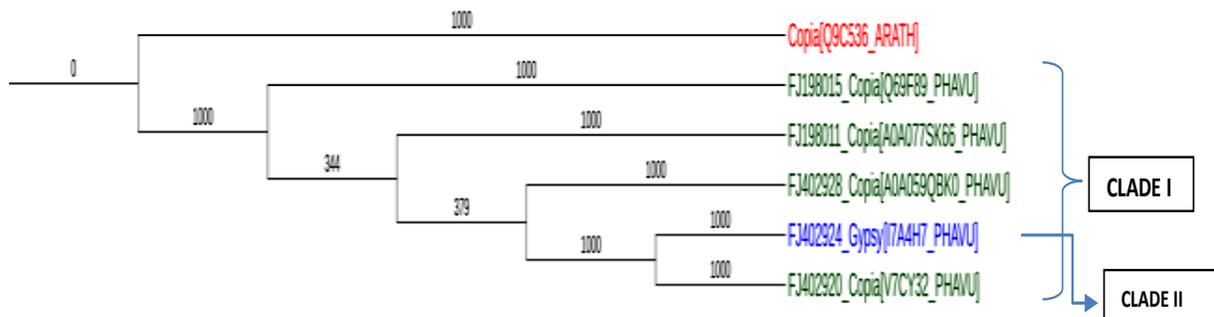
### 3.2 Phylogenetic tree for *Phaseolus vulgaris*

For *P. vulgaris* also, NJ method was used for constructing phylogenetic tree. Bootstrapping was done in the similar way as mentioned for *Glycine max*. It has been shown in the figure 2b

(a)



(b)



(c)

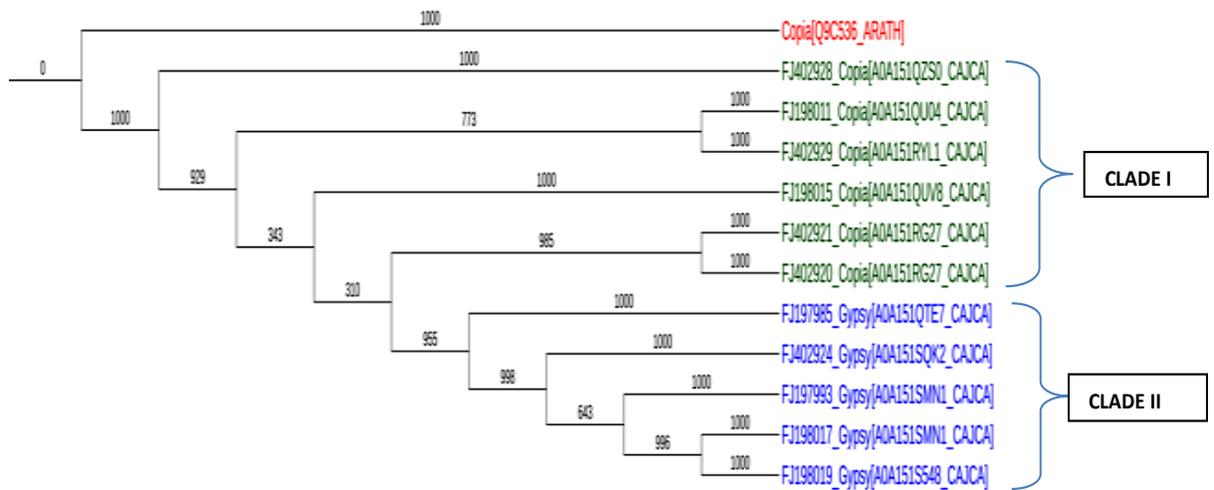
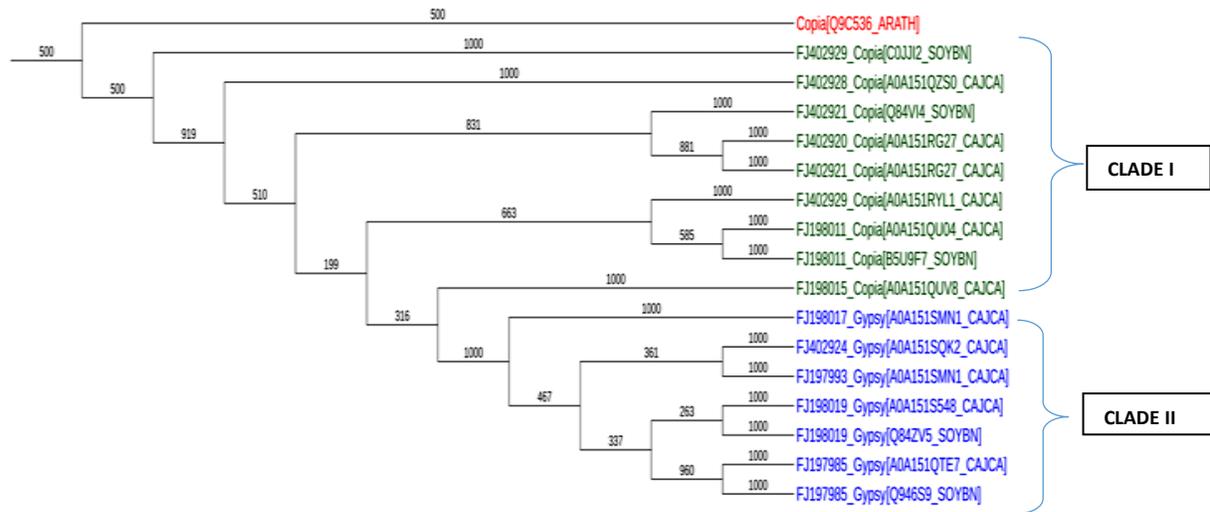
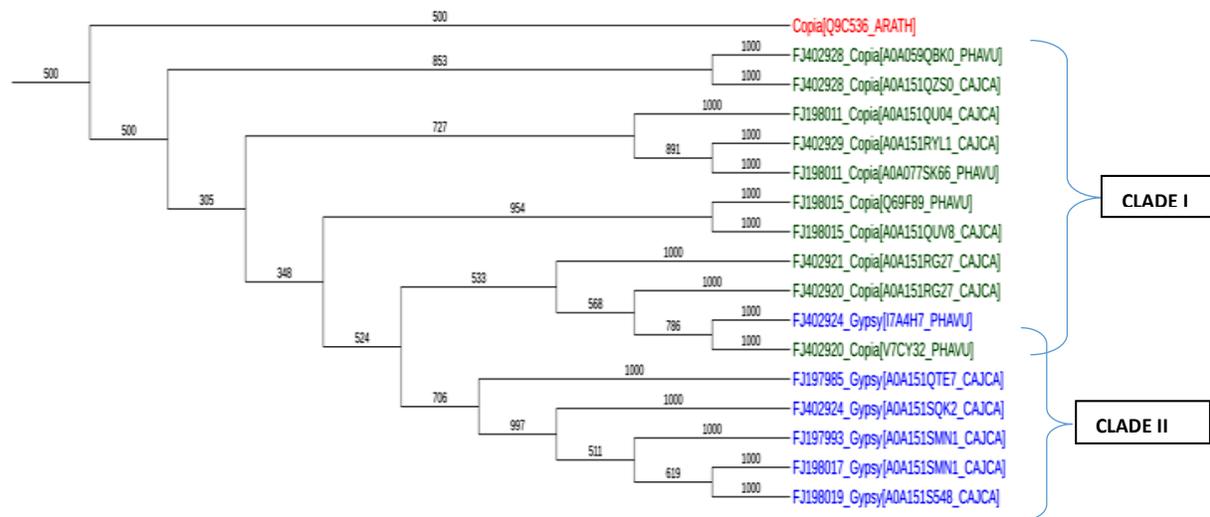


Figure 2 Phylogenetic tree for *Glycine max* (a), *Phaseolus vulgaris* (b) *Cajanus cajan* (c). Organism symbols used for making phylogenetic tree: SOYBN for *Glycine max*; PHAVU for *Phaseolus vulgaris*; CAJCA for *Cajanus cajan* and ARATH for *Arabidopsis thaliana*



Figure 4 Phylogenetic tree for combination of *Glycine max* and *Cajanus cajan*Figure 5 Phylogenetic tree for combination of *Phaseolus vulgaris* and *Cajanus cajan*

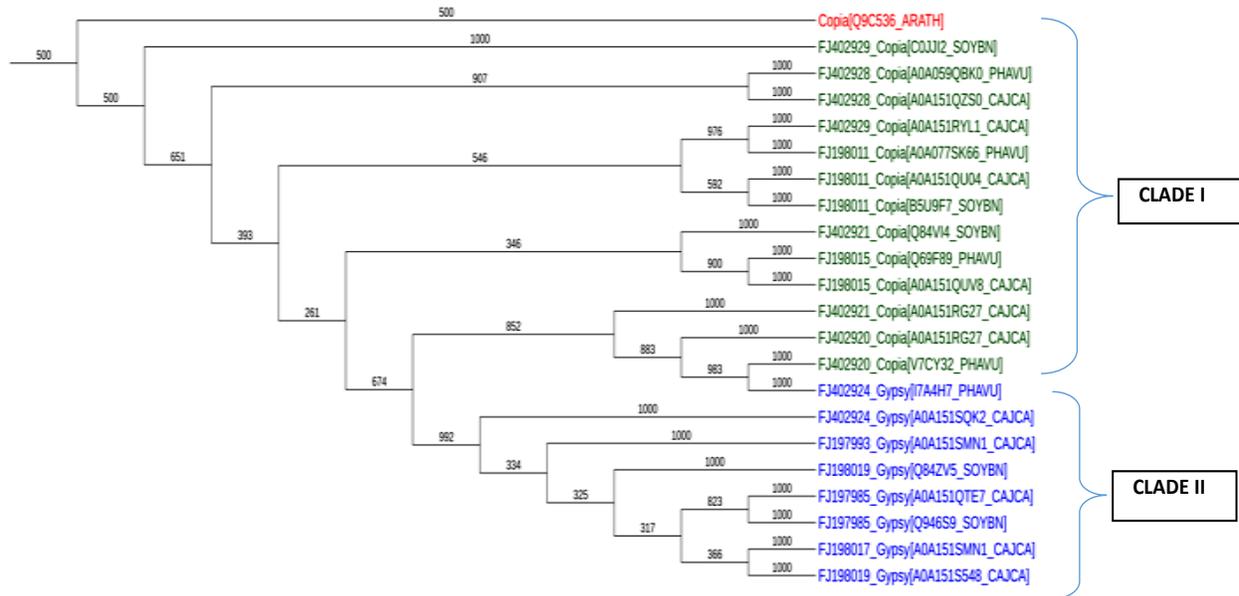
### 3.6 Phylogenetic tree for *P. vulgaris* and *C. Cajan*

Another phylogenetic tree was constructed for the combination of *P. vulgaris* and *C. Cajan* (Figure 5). The evolutionary relationship of copia and gypsy was studied using 16 protein sequences along with Q9C536 *A. thaliana* (copia). Phylogenetic tree was constructed using NJ method with the bootstrap of 1000 replicates, where Q9C536 *A. thaliana* was taken as an outgroup. Here, Phytoalexin-deficient 4-2 protein "I7A4H7" revealed a close relationship with V7CY32. It is clearly visible in the figure 5 that clade I contains ten copia and in

clade II six gypsy transposable elements (both of *P. vulgaris* and *C. cajan*) are present.

### 3.7 Combined phylogenetic tree for *G. max*, *P. vulgaris* and *C. cajan*

For further characterization, a combined phylogenetic tree for all the three, *G. max*, *P. vulgaris* and *C. cajan* was constructed (Figure 6). The evolutionary relationship of copia, gypsy and *A. thaliana* (copia) was inferred using the NJ method. The analysis involved resulting protein sequences from the blast search of 12 transposable elements.



**Figure 6** Phylogenetic tree for combination of *Glycine max*, *Phaseolus vulgaris* and *Cajanus cajan*.

In the Figure 6, numbers on the branches refer the bootstrap values. The "Q9C536" was kept as an outgroup. It was found that COJJ12 was more distantly related to all others. It is clearly visible in the tree that clade I contains thirteen copia and in clade II eight gypsy transposable elements are present.

#### 4 Discussion

The evolution of TEs and their effects on genome evolution is currently a dynamic field of study. Various studies have been performed on the TEs evolution and their contribution (Anderson et al., 2019). TEs are genetic mobile elements that have the ability to alter their position within the genome and in few cases, to produce new copies of them. As a result, TEs are valuable source of mutations and they reported an essential part of all sequenced genomes (Tenaillon et al., 2010). Genetic variability can be generated within gene by using TEs. They can modify the function of genes if they are inserted within the gene due to which gene can become susceptible or resistant as per requirement. TEs have various important functions, as they are a major component of plant genomes and also affect genome structure. By altering genes, TEs act as an essential source of mutations and thus having important impact on genome evolution. It influences the gene regulation as they can change gene expression by providing their own regulatory elements (Contreras et al., 2015).

There are various tools that have been developed for the identification and annotation of TEs in assembled genomes (Bergmann & Quesneville, 2007 ; Jiang et al., 2015). This study has provided information about the evolutionary history of TEs in

*G. max*, *P. vulgaris* and *C. cajan* as described in table 2. A multiple sequence alignment of soybean TEs and other legume species followed by phylogenetic analysis revealed the genetic relationship between different copia and gypsy elements. This research aims the identification and characterization of TEs from *G. max*, *P. vulgaris* and *C. cajan*.

The protein BLAST result of the TEs showed that they encode for different products. The Ty3/gypsy codes for a polyprotein in *G. max* while in *C. cajan* it codes for transposons Ty3-1 Gag-Pol polyprotein. Similarly one of the copia (with Uniport ID-B5U9F7) codes for a Gag-protease-integrase-RT-RNase H polyprotein in *G. max*, is also found in *C. cajan* (A0A151QU04) and *P. vulgaris* (A0A077SK66) where it codes for Retrovirus-related Pol polyprotein from transposon TNT 1-94 and Putative Ty-1 copia retrotransposons respectively. One of the copia in *P. vulgaris* is V7C432 was found to code for an uncharacterized protein. These finding will constitute a useful source of information to the legume genome annotation.

For a bulky genome like soybean, in which genes composed only a small part of whole genome, a transposon that has a preference for a gene rich region is required for a successful mutagenesis experiment. Tnt1 insertion site study showed that it transposes to gene containing regions (Cui et al., 2013; Sandhu & Bhattacharyya, 2017). Here in this analysis, a Tnt1 copia transposons (as shown in table 2) is found, which would be beneficial for tagging genes. Most of the TEs identified in this study are near NB-LRR gene that can be utilized for construction of disease resistance plants.

The basic knowledge and genetic variation developed using TEs will give means to enhance yield and quality in soybean for acquiring a feasible supply of nutritionally essential crop. The lifestyle of plants makes them susceptible to environmental stress, the capacity for genetic adaptation and development of novel mechanism of resistance is thus essential for them to survive (Grzebelus, 2018). As discussed above, TEs contribute in epigenetic changes in response to stress factors, creating phenotypic plasticity that can result in adaptation to changed environment. Confronting the undeniable problems imposed by global warming, this approach would be useful in creating more tolerant legume varieties to a range of abiotic stresses.

### Conclusion

The result indicates that these transposable elements (TEs) of *G. max* have high similarity with the transposons from *C. cajan* and *P. vulgaris*; further indicating that these were present before the divergence of these three legumes (*Glycine max*, *C. cajan* and *P. vulgaris*). All the copia elements grouped together from *G. max*, *C. cajan* and *P. vulgaris* in clade I, similarly all the gypsy elements grouped together in clade II. Our results will contribute significantly to the improvement of legumes by providing important genomic resource, as TEs are mobile elements so they can be inserted within gene and thus will be beneficial in creating genetic variability or genetic improved crops.

### Acknowledgements

Authors acknowledge the facilities of the Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi (DBT) under the Bioinformatics Sub Centre as well as the golden jubilee research fellowship (GJRF) provided by School of Biotechnology, Devi Ahilya University Indore. M.P. India.

### Conflict of Interest

The authors confirm that they have no conflict of interest.

### References

Anderson SN, Stitzer MC, Brohammer AB, Zhou P, Noshay JM, Hirsch CD, Ross-Ibarra J, Hirsch CN, Springer NM (2019) Transposable elements contribute to dynamic genome content in maize. *BioRxiv* 547398.

Baránek M, Meszáros M, Sochorová J, Čechová J, Raddová J (2012) Utility of retrotransposon-derived marker systems for differentiation of presumed clones of the apricot cultivar Velkopavlovická. *Scientia Horticulturae* 143: 1-6.

Bergman CM, Quesneville H (2007) Discovering and detecting transposable elements in genome sequences. *Briefings in Bioinformatics* 8: 382-392.

Contreras B, Vives C, Castells R, Casacuberta JM (2015) The impact of transposable elements in the evolution of plant genomes: from selfish elements to key players. In *Evolutionary biology: Biodiversification from genotype to phenotype*. Springer, Cham, Pp. 93-105.

Cui Y, Barampura S, Stacey MG, Hancock CN, Findley S, Mathieu M, Zhang Z, Parrott WA, Stacey G, (2013) Tnt1 retrotransposon mutagenesis: a tool for soybean functional genomics. *Plant Physiology*, 161:36-47.

Du J, Grant D, Tian Z, Nelson, RT, Zhu L, Shoemaker RC, Ma J (2010) Soy TEdb: a comprehensive database of transposable elements in the soybean genome. *BMC Genomics* 11: 113.

Fedoroff NV (2012) Transposable elements, epigenetics, and genome evolution. *Science* 338: 758-767.

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends in Genetics* 5: 103-107.

Gao D, Abernathy B, Rohksar D, Schmutz J, Jackson SA (2014) Annotation and sequence diversity of transposable elements in common bean (*Phaseolus vulgaris*). *Frontiers in Plant Science* 5: 339.

Grzebelus D (2018) The functional impact of transposable elements on the diversity of plant genomes. *Diversity* 10: 18.

Jiang C, Chen C, Huang Z, Liu R, Verdier J (2015) ITIS, a bioinformatics tool for accurate identification of transposon insertion sites using next-generation sequencing data. *BMC Bioinformatics* 16: 72.

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* 33: 102.

Kolade O, Raji A, Fawole I, Ingelbrecht I (2015) Molecular Characterization of Type II Transposable Elements in Cowpea [*Vigna unguiculata* (L.) Walp]. *American Journal of Plant Sciences* 6: 767.

Lall I, Upadhyaya K (2002) Panzee, a copia-like retrotransposon from the grain legume, pigeonpea (*Cajanus cajan* L.). *Molecular Genetics and Genomics* 267: 271-280.

Lin JY, Stupar RM, Hans C, Hyten DL, Jackson SA (2010) Structural and functional divergence of a 1-Mb duplicated region in the soybean (*Glycine max*) genome and comparison to an orthologous region from *Phaseolus vulgaris*. *The Plant Cell* 22:2545-2561. doi: 10.1105/tpc.110.074229.

McClellan PE, Mamidi S, McConnell M, Chikara S, Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics* 11: 184.

- Nemli S, Kianoosh T, Tanyolac MB (2015) Genetic diversity and population structure of common bean (*Phaseolus vulgaris L.*) accessions through retrotransposon-based interprimer binding sites (iPBSs) markers. *Turkish Journal of Agriculture and Forestry* 39: 940-948.
- Ouyang S, Buell CR (2004) The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Research* 32: D360-D363.
- Panaud O (2009) The molecular bases of cereal domestication and the history of rice. *Comptesrendus Biologies* 332: 267-272.
- Patil PG, Byregowda M, Agbagwa IO, Shashidhar HE (2015) Characterization of Ty1/copia-like retrotransposon families from pigeonpea genome. *Genetics and Molecular Research* 14: 5812-5822.
- Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis TH, Flavell AJ (1999) Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. *The Plant Journal* 19: 711-717.
- Sandhu D, Bhattacharyya MK (2017) Transposon-Based Functional Characterization of Soybean Genes. In: *The Soybean Genome* Springer, Cham, Pp. 183-192.
- Sandhu D, Ghosh J, Johnson C, Baumbach J, Baumert E, Cina T, Grant D, Palmer RG, Bhattacharyya MK (2017) The endogenous transposable element Tgm9 is suitable for generating knockout mutants for functional analyses of soybean genes and genetic improvement in soybean. *PloS one* 12: e0180732.
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178.
- Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S, Dogra V, Singh BP, Kumawat G, Pal JK (2011) The first draft of the pigeonpea genome sequence. *Journal of Plant Biochemistry and Biotechnology* 20: 15.
- Staginnus C, Huettel B, Desel C, Schmidt T, Kahl G (2001) A PCR-based assay to detect En/Spm-like transposon sequences in plants. *Chromosome Research* 9: 591-605.
- Tenaillon MI, Hollister JD, Gaut BS (2010) A triptych of the evolution of plant transposable elements. *Trends in Plant Science*, 15: 471-478.
- Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, Farmer AD (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology* 30: 83.
- Wawrzynski A, Ashfield T, Chen NW, Mammadov J, Nguyen A, Podicheti R, Cannon SB, Thareau V, Ameline-Torregrosa C, Cannon E, Chacko B, (2008) Replication of nonautonomous retroelements in soybean appears to be both recent and common. *Plant Physiology*, 148: 1760-1771.
- Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences* 103: 17600-17601.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: 973.
- Xu M, Brar HK, Grosic S, Palmer RG, Bhattacharyya MK (2009) Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in soybean [*Glycine max (L.) Merr.*] *Genetics* 184:53-63. doi: 10.1534/genetics.109.107904.
- Zabala G, Vodkin L (2008) A putative autonomous 20.5 kb-CACTA transposon insertion in an F3'H allele identifies a new CACTA transposon subfamily in *Glycine max*. *BMC plant biology* 8: 124.
- Zabala G, Vodkin LO (2014) Methylation affects transposition and splicing of a large CACTA transposon from a MYB transcription factor regulating anthocyanin synthase genes in soybean seed coats. *PLoS One* 9: e111959.
- Zhang Y, Fan C, Li S, Chen Y, Wang RR, Zhang X, Han F, Hu Z (2017) The diversity of sequence and chromosomal distribution of new transposable element-related segments in the rye genome revealed by FISH and lineage annotation. *Frontiers in Plant Science* 8:1706.
- Zhao M, Ma J (2017) Transposable Elements. In *The Soybean Genome*, Springer, Cham, pp. 171-181.

#### **Used Online Resource**

<https://www.uniprot.org/blast/>

<http://www.clustal.org/clustal2/>

<http://evolution.genetics.washington.edu/phylip.html>