COMPARATIVE ANALYSIS OF TRANSPOSABLE ELEMENTS FROM

Glycine max, Cajanus cajan and Phaseolus vulgaris

Gaurav Singh1*, Milind Ratnaparkhe2, Anil Kumar1

1School of Biotechnology, Devi Ahilya University, Khandwa Rd., Indore-452001, India
2Indian 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

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.

KEYWORDS

Transposable elements
Copia, gypsy
Blast
Phylogenetic analysis
Glycine max
Cajanus cajan
Phaseolus vulgaris

* Corresponding author
E-mail: singhgauravbiotech@gmail.com (Gaurav Singh)

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

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.

Production and Hosting by Horizon Publisher India [HPI] (http://www.horizonpublisherindia.in/), All rights reserved.
Comparative analysis of Transposable Elements from *Glycine max*, *Cajanus cajan* and *Phaseolus vulgaris*

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)
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 Panzeea 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>
<thead>
<tr>
<th>S.No.</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree 1</td>
<td><em>Glycine max</em></td>
</tr>
<tr>
<td>Tree 2</td>
<td><em>Phaseolus vulgaris</em></td>
</tr>
<tr>
<td>Tree 3</td>
<td><em>Cajanus cajan</em></td>
</tr>
<tr>
<td>Tree 4</td>
<td><em>G. max + P. vulgaris</em></td>
</tr>
<tr>
<td>Tree 5</td>
<td><em>G. max + C. cajan</em></td>
</tr>
<tr>
<td>Tree 6</td>
<td><em>P. vulgaris + C. cajan</em></td>
</tr>
<tr>
<td>Tree 7</td>
<td><em>G. max + P. vulgaris + C. cajan</em></td>
</tr>
</tbody>
</table>

Table 1 Combination for making Phylogenetic tree

Journal of Experimental Biology and Agricultural Sciences
http://www.jebas.org
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

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Accession no. with name</th>
<th>Type</th>
<th>Protein</th>
<th>Organism</th>
<th>Uniprot ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FJ198019.1</td>
<td>Ty3/gypsy</td>
<td>Polypeptide</td>
<td><em>G. max</em></td>
<td>Q84ZV5</td>
</tr>
<tr>
<td></td>
<td>gmw1-52d1-re-3</td>
<td></td>
<td>Ty3-I Gag-Pol polypeptide</td>
<td><em>C. cajan</em></td>
<td>A0A151S548</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Phaselus vulgaris</em> NOT FOUND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>FJ198017.1</td>
<td>Ty3/gypsy</td>
<td>Polypeptide</td>
<td><em>G. max</em></td>
<td>Q84ZV5</td>
</tr>
<tr>
<td></td>
<td>gmw1-27d20-re-1</td>
<td></td>
<td>Retrotransposons Tf2</td>
<td><em>C. cajan</em></td>
<td>A0A151S3E7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>P. vulgaris</em> NOT FOUND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>FJ198015.1</td>
<td>Copia</td>
<td>Polypeptide</td>
<td><em>G. max</em></td>
<td>C0JJI2</td>
</tr>
<tr>
<td></td>
<td>gmw2-129e12-re-3</td>
<td></td>
<td>Retrovirus-related Pol polypeptide from transposon TNT I-94</td>
<td><em>C. cajan</em></td>
<td>A0A151QUV8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>P. vulgaris</em> Q69F89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>FJ198011.1</td>
<td>Copia</td>
<td>Polypeptide</td>
<td><em>G. max</em></td>
<td>B5U9F7</td>
</tr>
<tr>
<td></td>
<td>gmw2-10n21-re-5</td>
<td></td>
<td>Retrovirus-related Pol polypeptide from transposon TNT I-94</td>
<td><em>C. cajan</em></td>
<td>A0A151QU04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>P. vulgaris</em> A0A077SK66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>FJ402890.1</td>
<td>LINE-1</td>
<td>Polypeptide</td>
<td><em>G. max</em></td>
<td>Q1AKH8</td>
</tr>
<tr>
<td></td>
<td>gmp1-95h18-re-1</td>
<td></td>
<td>Retrovirus-related Pol polyprotein LINE-1</td>
<td><em>C. cajan</em></td>
<td>A0A151SQ16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>P. vulgaris</em> Not available</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>FJ197993.1</td>
<td>Ty3/gypsy</td>
<td>Polypeptide</td>
<td><em>G. max</em></td>
<td>Q84ZV5</td>
</tr>
<tr>
<td></td>
<td>gmp1-71h23-re-7</td>
<td></td>
<td>Ty3-I Gag-Pol polypeptide</td>
<td><em>C. cajan</em></td>
<td>A0A151SMN1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>P. vulgaris</em> NOT FOUND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>FJ197985.1</td>
<td>Ty3/gypsy</td>
<td>Polypeptide</td>
<td><em>G. max</em></td>
<td>Q946S9</td>
</tr>
<tr>
<td></td>
<td>gmp1-125b19-re-9</td>
<td></td>
<td>Ty3-I Gag-Pol polypeptide</td>
<td><em>C. cajan</em></td>
<td>A0A151QTE7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>P. vulgaris</em> NOT FOUND</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

### Table 2 Details of different transposable elements in few legumes

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Accession no. with name</th>
<th>Type</th>
<th>Protein</th>
<th>Organism</th>
<th>Uniprot ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>FJ402928.1</td>
<td>Copia</td>
<td>Copia-type polyprotein</td>
<td>G. max</td>
<td>C0JJI2</td>
</tr>
<tr>
<td></td>
<td>pva1-144m6-re-2</td>
<td></td>
<td>Retrovirus-related Pol polyprotein from transposon TNT 1-94</td>
<td>C. cajan</td>
<td>A0A151QZS0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Polyprotein</td>
<td>P. vulgaris</td>
<td>A0A059QBK0</td>
</tr>
<tr>
<td>9</td>
<td>FJ402929.1</td>
<td>Copia</td>
<td>Copia-type polyprotein</td>
<td>G. max</td>
<td>C0JJI2</td>
</tr>
<tr>
<td></td>
<td>pva1-144m6-re-5</td>
<td></td>
<td>Retrovirus-related Pol polyprotein from transposon TNT 1-94</td>
<td>C. cajan</td>
<td>A0A151RYL1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Putative Ty-1 copia retrotransposon</td>
<td>P. vulgaris</td>
<td>A0A077SK66</td>
</tr>
<tr>
<td>10</td>
<td>FJ402924.1</td>
<td>Ty3/gypsy</td>
<td>Transposon Ty3-I Gag-Pol polyprotein</td>
<td>C. cajan</td>
<td>A0A151SQK2</td>
</tr>
<tr>
<td></td>
<td>pva1-47b16-re-5</td>
<td></td>
<td>Phytoalexin-deficient 4-2 protein</td>
<td>P. vulgaris</td>
<td>I7A4H7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glycine tomentella</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. FJ402921.1</td>
</tr>
<tr>
<td>gtt1-296j12-re-1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>12. FJ402920.1</td>
</tr>
<tr>
<td>gtt1-158p11-re-2</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
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.
that all the five sequences were compared with *A. thaliana* sequence and were grouped together. Sequence of Q69F89_PHAVU was compared 1000 times to all the four sequences, results of study showed that all the four sequences were grouped 344 times together. When A0A077SK66_PHAVU was compared 1000 times with the remaining three sequences, these were grouped 479 times together. Further, A0A058QBKO was compared 1000 times with the remaining two sequences; these were grouped 1000 times together. From these data, it looks that these three sequences (A0A059QBKO_PHAVU, I7A4H7_PHAVU and V7CY32_PHAVU) were evolved together and may share a common ancestor. It is clearly visible from the tree (Figure 2b) that within clade I, four copia and in clade II one gypsy elements are present.

### 3.3 Phylogenetic tree for Cajanus cajan

NJ method was used for constructing tree of *C. cajan* using *A. thaliana* as an outgroup (Figure 2c). When all the 11 sequences of *C. cajan* were compared 1000 times together with the outgroup, all the 11 sequences were grouped 1000 times together. The data indicates that all these sequences which were compared 1000 times with the remaining sequences and also grouped together 1000 times are closely related sequences. For instance, A0A151QU04_CAJCA and A0A151RYL1_CAJCA; A0A151RG27_CAJCA and A0A151RG27_CAJCA; A0A151SMN1_CAJCA and A0A151S548_CAJCA. The phylogenetic tree showed that there are six copia elements in clade I and five gypsy in clade II (Figure 2c). These are present on the different chromosome in *C. cajan*.

### 3.4 Phylogenetic tree for *G. max* and *P. vulgaris*

As shown in Figure 3, for further characterization, a phylogenetic tree for combination of *G. max* and *P. vulgaris* was drawn. The tree shows the evolutionary relationship of copia and gypsy. The study was carried out with 10 protein sequences of copia and gypsy along with Q9C536- *Arabidopsis thaliana* (copia). The tree has been constructed using the NJ method, with the bootstrap of 1000 replicates. Q9C536-Arabidopsis thaliana was kept as an outgroup. Phytoalexin-deficient 4-2 protein “I7A4H7” showed a close relationship with V7CY32, while Q69F89 of PAVU showed close relationship to Q946S9 and Q842V5 of SOYBN. It is clearly visible in the tree that clade I contains six copia and in clade II four gypsy transposons are present.

### 3.5 Phylogenetic tree for *G. max* and *C. cajan*

A phylogenetic tree of *G. max* and *C. cajan* was drawn (Figure 4). Here, the evolutionary relationship of copia and gypsy was studied with 16 protein sequences taking Q9C536- *A. thaliana* (copia) as an outgroup. The tree was constructed using NJ method with the bootstrap of 1000 replicates.

FJ198015 Copia (A0A151QUV8_CAJCA) has been found closely related to the gypsy in CAJCA and SOYBN. In addition, FJ402928_Copia (A0A151QZS0_CAJCA) was found closely related to FJ402921_Copia (Q84VH4_SOYBN), FJ402920_Copia (A0A151RG27_CAJCA), FJ402921_Copia (A0A151RG27_CAJCA), FJ402929_Copia (A0A151RYL1_CAJCA), FJ198011_Copia (A0A151QU04_CAJCA) and FJ198011_Copia (B5U9F7_SOYBN). It is clearly visible from the figure 4 that clade I contains nine copia and in clade II seven gypsy elements.
Comparative analysis of Transposable Elements from *Glycine max*, *Cajanus cajan* and *Phaseolus vulgaris*

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 “T7A4H7” 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.
In the Figure 6, numbers on the branches refer the bootstrap values. The “Q9C536” was kept as an outgroup. It was found that C0JJI2 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 adaption 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, Cajanus 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**


**Used Online Resource**

https://www.uniprot.org/blast/
http://www.clustal.org/cluster2/