JMBFS
Journal of Microbiology, Biotechnology and Food Sciences
International peer-reviewed scientific online journal
Published by Faculty of Biotechnology and Food Sciences

MOLECULAR ANALYSIS OF A PUTATIVE EN/SPM-RELATED TRANSPOSON PROTEIN IN *BRASSICA JUNCEA*

Ahmed A. M. Yassein, Gehad R. Eid, Nada F. Hemeda, Zaki A. El Fiky*

Address(es): Ahmed A.M. Yassein,
Fayoum University, Faculty of Agriculture, Genetics Department, 63514 Fayoum, Egypt.

*Corresponding author: aay00@fayoum.edu.eg

https://doi.org/10.15414/jmbfs.2825

ARTICLE INFO

Received 24. 3. 2020
Revised 24. 11. 2020
Accepted 3. 12. 2020
Published 1. 4. 2021

Regular article

OPEN ACCESS

ABSTRACT

Open reading frame (ORF) of Putative En/Spm transposon in *Brassica* potentially helps to understand the relationship between various eukaryotic transposable elements. The current study was initially conducted to isolate and analyze the putative En/Spm-related transposon gene from *Brassica juncea*. PCR products (750 bp) from *B. juncea* (accession PI 649105 and PI 271442) were cloned, sequenced, and analyzed. Results of BLAST showed identical sequences between two accessions with 100% similarity. The amplified DNA transposon and conserved domain compared to the GenBank database to evaluate the genetic diversity relationships. Sequence of this putative transposon gene from *B. juncae* was a 98% similar to *B. rapa* subsp pekinensis at the nucleotide level, and 94% with the En/Spm-related transposon protein of *B. oleracea* at the amino acid level. Conserved domain architecture was related to transposase_21_pfam0299 and transposase family tnp2 and had a relationship with space outside the cell structure and/or to space outside the plasma membrane.

Keywords: *Brassica juncea*, Putative En/Spm transposon protein, Transposase_21, Phylogenetic analyses

INTRODUCTION

Brassica juncea is a species of allotetraploid brassica (2n = 36, AABB) formed by hybridization between the diploid ancestors of *B. rapa* (2n = 20, AA) and *B. nigra* (2n = 16, BB), followed by spontaneous chromosome doubling (**Monteiro and Lunn, 1999**). Subsequent diversifying selection then gave rise to the vegetable and oil use subvarieties of *B. juncea*. These subvarieties include vegetable and oilseed mustard in China, oilseed crops in India, canola crops in Canada and Australia, and condiment crops in Europe and other regions (**Chen *et al*., 2013**).

The transposable element system Enhancer (En) of *Zea mays* was originally identified by **Peterson (1953)** at the pale green locus as a mutable allele. Subsequently it was shown to be homologous to the Suppressor-Mutator (Spm) system both genetically (**Peterson, 1965**) and molecularly (**Pereira *et al*., 1985**). According to the broadest classification system, TEs are divided into two main categories on the basis of their transposition intermediate: Class I transposons or retrotransposons, transpose via an RNA intermediate and Class II transposons or DNA transposons, transpose directly from DNA with the help of transposase (**Kapitonov and Jurka, 2008**). Elements of both classes can be found in families of similar elements, though retrotransposons tend to form larger families due to their replicative nature. Within each class of transposon, one can categorize them further into superfamilies according to broad features such as the structure of encoded proteins or non-coding regions, or target site duplication (TSD) length (**Sun *et al*., 2016**; **Orozco-Arias *et al*., 2019**).

DNA-mediated transposons adopted “cut and paste” mechanism of transposition and are characterized by a transposase encoded by autonomous copies and with few exceptions by the presence of terminal inverted repeats (TIRs). Elements belong to the same superfamily can be linked to transposases that are significantly related in sequence. Typically, transposases from the same superfamily can be confidently aligned in their core catalytic region and a monophyletic ancestry can be inferred from phylogenetic analysis (**Capy, *et al*., 1998** and **Robertson, 2002**). Differences in the transposase motifs as well as the TIR sequences and the size and sequence of the TSD, allow the classification of DNA transposons into 6 main superfamilies: PIF/Harbinger, hAT, Tc1/Mariner, CACTA, MULE, and Helitron, which are common in plants (**Wicker *et al*., 2007; Kapitonov and Jurka, 2008**). As identified in other plants, the genome of Brassica also harbor transposable elements (TEs) such as LTR retrotransposons (**Nouroz *et al*., 2015a**), DNA transposons like Mutator (**Nouroz and Noreen, 2015**), hATs (**Nouroz *et al*., 2015b**) and Harbingers (**Zhang and Wessler, 2004; Nouroz *et al*., 2016**).

Many families of the En/Spm superfamily are not readily recognize by computer assisted database searches (**Wang *et al*., 2003; Wicker *et al*., 2003**). The molecular research of plant transposable elements (TEs) focused on two areas: (a) gene isolation and TE-induced alleles and subsequent characterization of TE families, and (b) understanding the mechanisms underlying the diversity of unstable phenotyptypes (**Zhang and Wessler 2004; Wessler, 2001**). We aimed here to isolate, cloning and analyze a putative En/Spm-related transposon from two accessions of *B. juncea* and to compare the amplified DNA transposon and conserved domain to GenBank database in order to observe and evaluate the diversity relationships.

MATERIALS AND METHODS

Plant Materials

Rapeseeds (*B. juncea*) accessions PI 649105 and PI 271442 was kindly provided by the North Central Regional Plant Introduction Station (USDA), USA kindly and the cultivar Serw 4 of oilseed rape *B. napus* (L.) was provided by the Agricultural Research Centre, Giza, Egypt.

Microorganisms, Plasmid Vector, and Culture Conditions

Escherichia coli DH5α was obtained from Prof. Dr. Maria Mercedes Bonfill Baldrich, Facultat de Farmàcia, Universitat de Barcelona, Spain. RBC T&A cloning vector, vector System (*Cat. No. RC001*) was used for cloning PCR products using thermostable DNA polymerases (Taq) which add a single terminal 3'-dA nucleotide overhang. Bacteria were grown in the LB medium with shaking (200 rpm) at 37°C (**Bertani, 1951**).

DNA Extraction

Genomic DNA was extracted from 0.2 gm grained leaves using a protocol described (**El Fiky *et al*., 2019**). The high purity and quantity of extracted DNA was stored at (-20 °C) until it was used in the following experiments.

PCR Amplification of a Putative En/Spm-Related Transposon Gene

The specific primers FU2-F: 5′-AAAACTCTCGTCGCAAGCAC-3′ and FU2-R: 5′-CCTCAGCCAGCGGTTAGATT-3′ were designed from the GenBank accession number A62414. The primer sequences were synthesized by

Invitrogen, Biotechnology Co. Ltd. (USA). PCR amplification was performed according to **Schelfhout *et al.*, (2004)**

Purification, A-Tailing and Cloning of PCR Products

In accordance with the protocol defined by (**El Fiky *et al.*, 2019**), PCR products were purified. Six µl from PCR product, 1 µl BD buffer 1x, 0.5 µl firepol, 2 µl dATP and 1 µl MgCl2 were mixed in microfuge tube and incubated at 70ºC for 20 min for A-Tailing overhangs using FIREpolR DNA polymerase (Solis BioDyne) Cat# 01-01-0000S. The RBC T&A cloning Kit (Real Biotech Corporation RC001 RBC T&A Cloning Kit / RC013 RBC T&A Cloning Vector) was used to ligate Atailing of PCR products into cloning vector as a manufacturer's description.

Preparation of *E.coli* Competent Cells, Transformation, and Plasmid Extraction

Competent cells were prepared using the protocol mentioned (**Sambrook, 2001**). Pellets resuspended in 80 µl ice cold 85 Mm $CaCl_2$, 15% glycerol and saved at -80 ºC. To transform the competent cells, *E. coli* was transformed using Rapid Transformation Procedure included in One Shot TOP10 Chemically Competent Cells (Life Technologies Corporation) was used as a user guide. One hundred µl of the resulting culture was spread on LB plates with ampicillin antibiotic (100 µg/ml) and colonies picked after about 12-16 hours to extract recombinant plasmid. Purified recombinant plasmid DNA was stored at -20 °C until used in sequencing.

Sequencing and Analyses of a Putative Transposon- Gene and Protein

Recombinant plasmid was sequenced by automated DNA sequencing with M13 forward and reverse primers using a sequencing ready reaction kit (Life Technologies) in combination with ABI-PRISM and ABI-PRISM big dye terminator cycler. Sequences of cloning DNA were subjected to alignment with sequences of the GenBank from Brassica using the BLAST 2.2.18 (Basic Local Alignment Search Tool) algorithm at https://blast.ncbi.nlm.nih.gov. The DNA sequences were subjected to the SIB Bioinformatics Resource Portal (ExPASy) which provides access to translate open reading frames (https://web.expasy.org/translate/) and the protein aligned using the BLASTP 2.2.18. MEGA, version 5.2 (**Tamura *et al.*, 2011**) was used to produce a phylogenetic tree of a putative gene and protein mediated the UPGMA method according to (**Sneath and Sokal, 1973**). The evolutionary relationships were calculated using Maximum Composite Likelihood method (**Tamura *et al.*, 2004**). For each query protein sequence, LocTree2 applies machine learning (kernel SVM profile) was used to predict the native sub-cellular localization in 18 classes for eukaryotes (https://rostlab.org/services/loctree3/).

RESULTS

Cloning of PCR product:

PCR amplification of a putative transposon gene, at the nucleotide level, showed one sharp band with a molecular weight 750 bp in each of the two *B. juncea* accessions, whereas no bands were detected in *B. napus* cultivar Serw 4 (Figure 1). The fragments with A-Tailing overhang PCR products were cloned into RBC T&A cloning vector and resulted in RBC- a putative transposon gene. Recombinant plasmids extracted from transformed *E.coli* (DH5α) were sequenced.

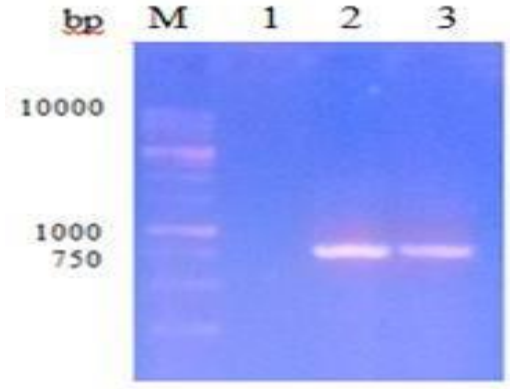

Figure 1 PCR amplification from *B. napus* and *B. juncea*. M: DNA marker, 1 kb DNA ladder, 1: *B. napus* cultivar Serw4, 2: *B. juncea* accession number PI 649105, 3: *B. juncea* accession number PI 271442

Analysis of Putative Transposon Gene Sequence

Sequence analysis using BLASTN software showed 100% similarity and 0.0% gaps between the two *B. juncea* sequences (Figure 2). *Brassica juncea* putative transposon gene (MH674328) subjected to the GenBank. This sequence showed 98% similarity and 93% query covered with *B. rapa* subsp pekinensis. UPGMA topology of the tree of this putative gene from *B. juncae* with 14 accession numbers of *B. rapa* subsp Pekinensis in the GenBank database was a monophyletic group (Figure 3). Successful grouping of DNA sequences into two major clusters. The first cluster was highly diverse and composed of putative gene from *B. juncae* (MH674328). The second cluster had 14 accession numbers of *B. rapa* and divided into two closely related accessions sub clusters.

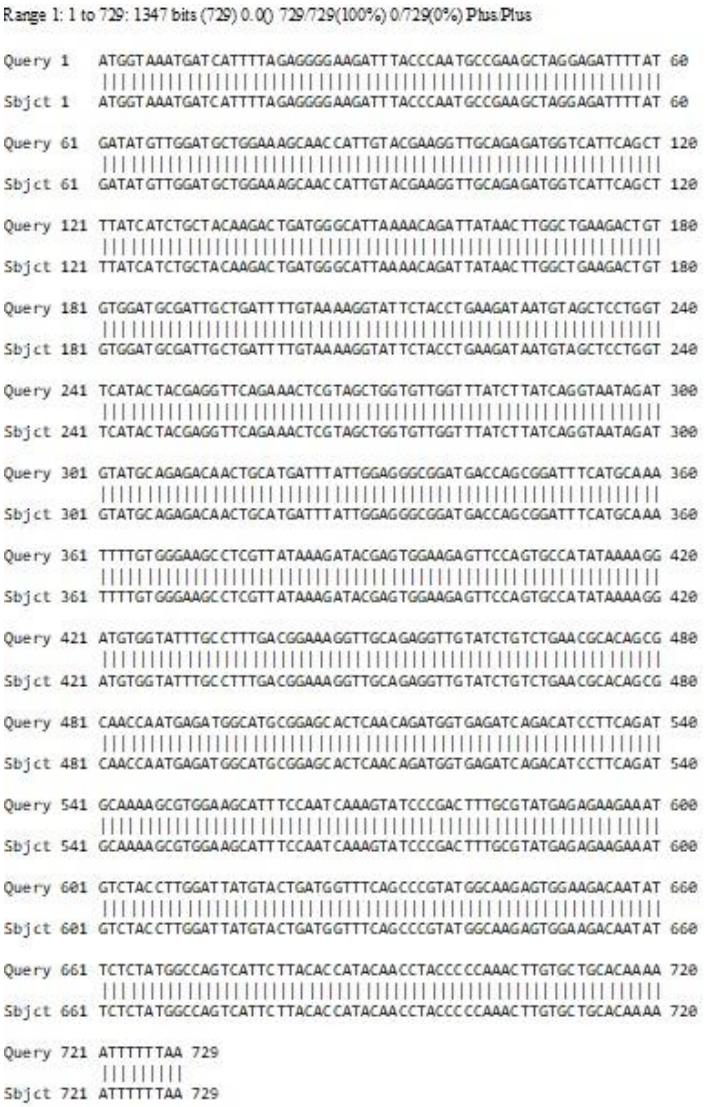

Figure 2 Alignment of nucleotide sequences of a putative gene amplified from *B. juncea*, accessions PI 649105 and PI 271442.

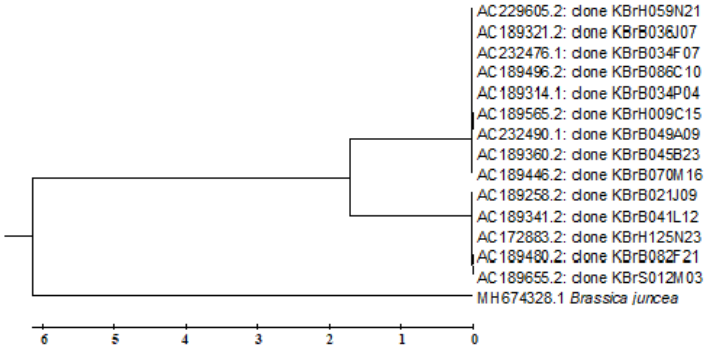

Figure 3 UPGMA dendrogram generated based on Sneath and Sokal distances of *B. juncea* putative transposon gene sequence (accession number MH674328) and 14 *B. rapa* subsp Pekinensis.

Analysis of Putative Transposon Protein Sequence

DNA transposon is a recurrent source of coding sequences to advent new genes. ExPASY is a translate tool which allows the translation of a nucleotide (DNA) sequence to a protein sequence (Amino acid). Results of the translation showed that the 5'3' frame 1 had one open reading frame (ORF), and contains 242 amino acids and the sequence of this protein was subjected to the GenBank, EMBL, DDBJ and PDB (Accession number QAU19549). This open reading frame (ORF) was subjected to BLASTP software to determine the alignment between this ORF and protein database collections. Results showed a 94% similarity and 100% queries covered between this protein and En/Spm-related transposon protein. Putative conserved domain with transposase_21 has been detected between nucleotide176-250. UPGMA tree of putative transposon protein with 16 protein accession numbers in the GenBank database represented a monophyletic group (Figure 4). The tree had successfully grouped into two main clusters, the putative transposon protein sequence (accession number QAU19549) closely related with En/Spm- related transposon protein (accession number ACG60686) in the first cluster. The second cluster is highly diverse and composed of 15 uncharacterized proteins from Brassica.

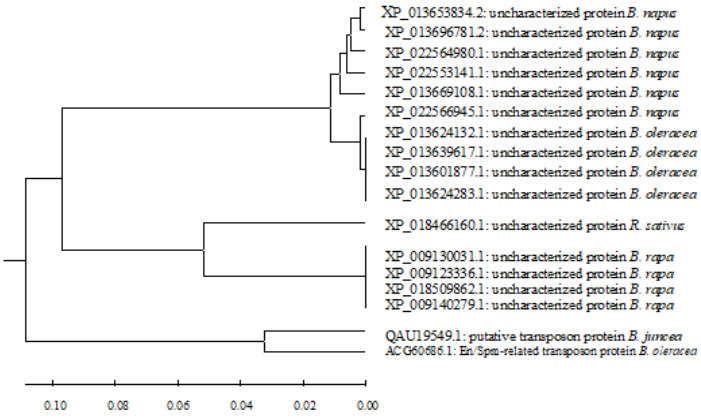

Figure 4 UPGMA dendrogram generated based on Sneath and Sokal distances of *B. juncea* putative transposon protein (accession number QAU19549) with 16 Brassica proteins.

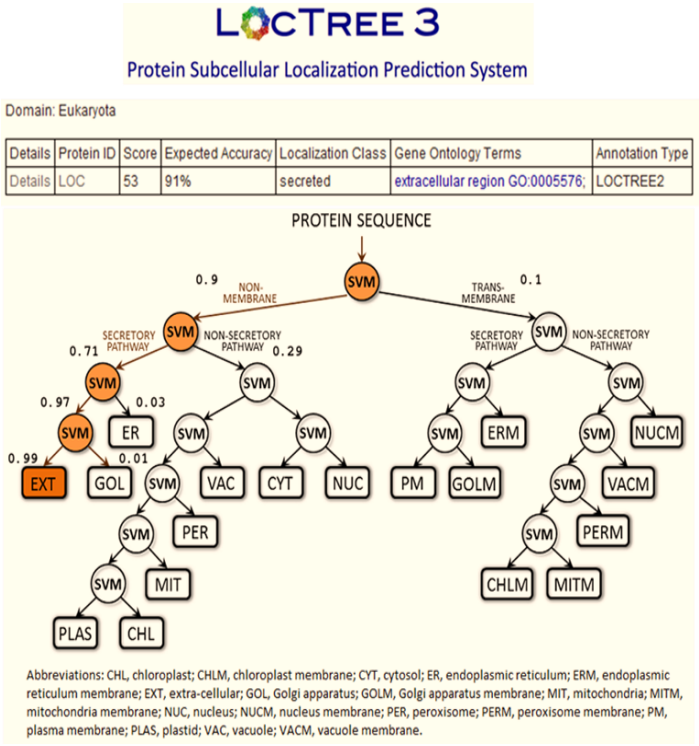

Figure 5 The LocTree3 prediction result output for input protein sequence.

DISCUSSION

Analysis of Putative Transposon Gene Sequence

As a result of cytogenetic and breeding studies, the genetic relationship between Brassica oilseed species was largely established. **Morinaga, (1934)** suggested that natural amphidiploid hybrid developed from haybridisation of the diploid species *B. nigra* (2n=16) and *B. rapa* (syn. campestris) (2n=20) is *B. juncea* (2n=36). The homology within En/Spm transposon sequences remained consistent among *Brassicas* (*B. juncea* and *B. rapa*). These results are consistent with **Zhang and Wessler (2004)** research, which studied the evolutionary relationship between *Brassica* and *Arabidopsis* CACTA transposons and showed a high intra-family homology of *B. oleracea* CACTA with a close relation to *Arabidopsis*. For this reason, the obtained results clearly indicate that the *B. juncea* and *B. rapa* subsp Pekinensis genome sequences provide an valuable resource for analyzing the evolution of amphidiploid genomes through its transposon genes.

Analysis of Putative Transposon Protein Sequence:

The Conserved domain database (CDD) classifies proteins into functionally different families and sub-families, and throught conserved domain architectures to facilitates comparative studies of protein families (**Marchler-Bauer *et al.*, 2017).** The results showed that conserved protein domain family is transposase_21, pfam02992 and transposase family tnp2 and the conserved domain architecture of this enzyme showed a protein homology with putative transposase, putative retrotransposase and putative CACTA transposon proteins. (**Nouroz *et al.*, 2017**) reported that not only Brassica CACTA or En/Spm transposases conserved in diploid Brassicas but actively proliferate in Brassicas allotetraploid (*B. juncea*, *B. napus*, *B. carinata*) and Arabidopsis sister genera.
Many families of the En/Spm superfamily are not readily recognize by computer assisted database searches (Wang et al., 2003; Wicker et al., 2003). The availability of genomic sequence data as well as sequence search tools has allowed bioinformatic methods to identify DNA transposon families based on sequence similarity to a known class II. because the terminal sequences of DNA transposons are often the only demand for transposase recognition (**Craig *et al.*, 2002**). DNA transposons are grouped into several families, based on structural diversity of transposases, of which 6 (CACTA, hAT, Harbingers, Helitron, Mutator and Mariner) are widespread in plants (**Wicker *et al.*, 2007; Kapitonov and Jurka, 2008**). The genome of Brassica also includes harbour transposable elements (TEs) such as LTR retrotransposons (Nouroz et al., 2015c), DNA transposons such as Mutator (**Nouroz and Noreen, 2015**), hATs (**Nouroz *et al.*, 2015b**) and Harbingers (**Zhang and Wessler, 2004; Nouroz *et al.*, 2016**), as found in other plants.
Functionally, only a few genes have been investigated, so the functional contribution of the transposase domain(s) to the corresponding protein remains a matter of speculation in most cases. However, it is possible to determine some predictions based on the functional analyses of associated transposases. So we selected this ORF amino acid sequence and applied into LocTree3 to predict the results output (Figure 5). These results showed that the protein obtained from this study (id: LOC) had expected the accuracy of the prediction. 91% with the id: GO0005576 which defined as a space external to the outermost structure of a cell. This relates to space outside of the plasma membrane for cells without external protective or external encapsulating structures (Goldberg et al., 2014).

CONCLUSION

Current study detected band of 750 bp and the sequence analysis indicated highly similar to *B. rapa* subsp Pekinensis. Conserved domain architecture is related to transposease_21_pfam0299 and transposease family tnp2 and had a relationship with space outside a cell's outermost structure of a cell and/or to space outside the plasma membrane. Putative En/Spm transposon in Brassica potentially helps in understanding the relationship between various eukaryotic transposable elements.

REFERENCES

Bertani, G., (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology*, 62, 293–300.
Capy, P.; Bazin, C.; Higuet, D.; Langin, T., (1998). Dynamics and Evolution of Transposable Elements. Austin, TX: Springer-Verlag.
Chen, S., Wan, Z., Nelson, M.N., Chauhan, J.S., and … Cowling, W.A., (2013). Evidence from genome-wide simple sequence repeat markers for a polyphyletic origin and secondary centers of genetic diversity of *Brassica juncea* in China and India. *Journal of Heredity*, 104, 416–427. https://doi.org/10.1093/jhered/est015.
Craig, N.L., Craigie R., Gellert M., Lambowitz, A.M., (2002). Mobile DNA II. Washington, DC: *American Society Microbiology Press.*
El Fiky, Z. A.; Yassein, A. A. M.; Hemeda, N. F.; Eid, G.R., (2019). Simple, rapid, efficient and low cost method for mini preparation of DNA from different sources. *Fayoum Journal of Agricultural Research Development*, 33, 12–23.

Goldberg, T., Hecht, M., Hamp, T., Karl, and …. Rost, B., (2014). LocTree3 prediction of localization. *Nucleic Acids Research*, 42, 350–355. https://doi.org/10.1093/nar/gku396.
Kapitonov, V. V., Jurka, J., (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nature Reviews Genetics*, 9, 411– 412. https://doi.org/10.1038/nrg2165-c1.
Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., and …. Bryant, S.H., (2017). CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45, D200–D203. https://doi.org/10.1093/nar/gkw1129.
Monteiro, A., Lunn, T., (1999). Trends and perspectives of vegetable brassica breeding world-wide. *Acta Horticulturae*, 495, 273–280.
Morinaga, T., (1934). Interspecific Hybridization in Brassica VI. The cytology of F1 hybrids of *B. juncea* and *B. nigra*. *Cytologia (Tokyo)*, 6, 62–67. https://doi.org/10.1508/cytologia.6.62
Nouroz, F., Noreen, S., (2015). Molecular characterization and diversity of a novel non- autonomous mutator-like transposon family in Brassica. *Pakistan Journal of Botany*, 47(4), 1367–1375.
Nouroz, F., Noreen, S., Heslop-Harrison, J.S., (2016). Characterization and diversity of novel PIF/Harbinger dna transposons in brassica genomes. *Pakistan Journal of Botany*, 48(1), 167–178.
Nouroz, F., Noreen, S., Heslop-Harrison, J.S., (2015a). Identification and evolutionary genomics of novel LTR retrotransposons in Brassica. *Turkish Journal of Biology*, 39, 740–757. https://doi.org/10.3906/biy-1501-77
Nouroz, F., Noreen, S., Heslop-Harrison, J.S., (2015b). Identification, characterization and diversification of non-autonomous hAT transposons and unknown insertions in Brassica. *Genes and Genomics*, 37, 945–958. https://doi.org/10.1007/s13258-015-0324-z
Nouroz, F., Noreen, S., Heslop-Harrison, J.S., (2015c). Identification and evolutionary genomics of novel LTR retrotransposons in Brassica. *Turkish Journal of Biology*, 39, 740–757. https://doi.org/10.3906/biy-1501-77
Nouroz, F., Noreen, S., Khan, M.F., Ahmed, S., Heslop-Harrison, J.S.P., (2017). Identification and characterization of mobile genetic elements LINEs from Brassica genome. *Gene*, 627, 94–105. https://doi.org/10.1016/j.gene.2017.06.015.
Pereira, A., Schwarz-Sommer, Z., Gierl, A., Bertram, I., Peterson, P.A., Saedler, H., (1985). Genetic and molecular analysis of the Enhancer (En) transposable element system of *Zea mays*. *EMBO Journal*, 4, 17–23. https://doi.org/10.1002/j.1460-2075.1985.tb02311.x.
Peterson, P.A., (1965). A Relationship between the Spm and En Control Systems in Maize. *American Naturalist*, 99, 391–398. https://doi.org/10.1086/282380
Peterson, P.A., (1953). A mutable pale-green locus in maize. *Genetics*, 38, 682.
Robertson, H. (2002). Evaluation of DNA transposons in eukaryotes. In: Mobile DNA II (Eds. Craig *et al.*). Pp. 1093-1110. ASM Press Washington DC. pp. 1093–1110.
Sambrook, J. and R.D.W., (2001). Molecular cloning - a laboratory manual. Third Edition Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Schelfhout1, C,J., Snowdon, R., Cowling, W.A., Wroth, J.M., (2004). A PCR based B-genome specific marker in *Brassica* species. *Theoretical and Applied Genetics*, 109(5):917-921. DOI: 10.1007/s00122-004-1713-x
Sneath, P.H.A. and Sokal, R.R., (1973). Numerical Taxonomy. The principles and practice of numerical classification. Freeman, San Francisco.
Tamura, K., Nei, M., Kumar, S., (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings National Academy of Science*. U. S. A., 101, 11030–11035. https://doi.org/10.1073/pnas.0404206101
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10), 2731–2739. https://doi.org/10.1093/molbev/msr121.
Wang, G.D., Tian, P.F., Cheng, Z.K., Wu, G., and …. He, Z.H., (2003). Genomic characterization of Rim2 / Hipa elements reveals a CACTA-like transposon superfamily with unique features in the rice genome. *Molecular Genetics and Genomics*, 270(3), 234–242. https://doi.org/10.1007/s00438-0030918-z.
Wessler, S.R., (2001). Plant transposable elements. A hard act to follow. *Plant Physiol*ogy, 125, 149–151. https://doi.org/10.1104/pp.125.1.149
Wicker, T., Guyot, R., Yahiaoui, N., Keller, B., (2003). CACTA Transposons in Triticeae. A Diverse Family of High-Copy Repetitive Elements. *Plant Physiology*, 132, 52–63. https://doi.org/10.1104/pp.102.015743
Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., and … Schulman, A.H., (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8, 973–982. https://doi.org/10.1038/nrg2165
Zhang, X., Wessler, S.R., (2004). Genome-wide comparative analysis of the transposable elements in the related species Arabidopsis thaliana and *Brassica oleracea*. *Proceedings National Academy of Science*, U. S. A., 101, 5589–5594. https://doi.org/10.1073/pnas.0401243101