Iraqi Journal of Biotechnology, 2024, Vol. 23, No. 3, 231-236

Polymorphism of Kunitz Trypsin Protease Inhibitor (KTI-2) in Some *Glycine Max* Var

[1]Aya R. Gomaa, [2]Osamah A., [1]Majid Sh. Hamdallah

[1]Institute of Genetic Engineering and Biotechnology for Postgraduate Studies, University of Baghdad
[2]Plant Protection Department, Faculty of Agriculture, University of Kufa

Received: November 7, 2023 / Accepted: December 20, 2023 / **Published: December 30, 2024**

Abstract: Kunitz Trypsin Protease Inhibitor enzyme is a type of protein found in legume seeds that works as a protease inhibitor. Kunitz-type soybean trypsin inhibitors are usually specific for trypsin, which is secreted by the pancreas, leading to a lack of trypsin in the body and resulting in dyspepsia of proteins, pancreatic expansion and other health problems. The aim of the study to evaluate the glycine max genome sequences have been retrieved from GenBank-NCBI for cultivars Wm82, Lee, and ZH13. The targeted trypsin sequence (KTI-2) was searched through the assembled genomes using map to reference tool. The KTI-2 sequence was about 872 bp and existed in the three assembled genomes with 100% match. Searching through mRNA sequences of the three genomes revealed that three copies of KTI-2 were found in Wm82 and Lee cultivars each, while only one copy found in ZH13 cultivar with different consistency percentages and lengths. The amino acid lengths ranged between 91 and 216 amino acids, however, 210 amino acids. corresponded to 100% pairwise identity. It was concluded the convergence of mRNA sequences and amino acid lengths among the studied cultivars refers to appear some modifications in the structural gene. The results revealed that the KTI2 gene is a stable and conserved feature in the studied soybean cultivars.

Keywords: Kunitz Trypsin, gene(KTI-2), (KTI), *Glycine max*, Soybean**,** protease inhibitor

Corresponding author: (Email: ayaradhi7@gmail.com).

Introduction

Soybean (*Glycine max L.*) is an annual legume crop with major economic significance, contributing to more than half of the global oilseed production (1). It is rich in oil (20%) and protein (35-40%) and is used for human and animal consumption as well as for industrial purposes, such as biofuels (2).Soybeans are also a significant source of polysaccharides, soluble fibers, phytosterol, lecithins, saponins, and phytochemicals mainly is flavones which either individually or collaboratively help in promoting health by reducing the incidence of some diseases like diabetes, hypertension, dyslipidemia, obesity, inflammation, cancer, etc. (15). The lack of starch in soybeans makes them a good source of protein for diabetics. However, soybean protein contains trypsin inhibitors and may make up to 6% of the seed's protein content (3). Soybean trypsin inhibitors affect growth and basal metabolism of different animal species and human upon consuming it. It thereby reduces food intake by diminishing their digestion and absorption, besides reducing retention of nitrogen absorbed (4-5). Trypsin is a digestive enzyme found in the small intestine that helps

break down proteins. The pancreas secretes trypsin in the form of trypsinogen. Research indicates that trypsin inhibitors hinder the normal functioning of the trypsin enzyme, which in the process damages the pancreas and prevents protein absorption (6). Trypsin is secreted into the human body in an inactive form as trypsinogen. The enzyme trypsin digests proteins by hydrolyzing the peptide bond found after the amino acid arginine or amino acid lysine in proteins. A lack of trypsin in the body may result in indigestion of nutrients, especially proteins, pancreatic enlargement, anemia, delayed wound healing, kidney inflammation, or kidney failure in the case of pancreatitis. Soybean trypsin inhibitor accumulates in the soybean seeds primarily as Kunitz trypsin inhibitor (KTI) and to a lower extent as Bowman-Birk trypsin inhibitor(BBTI) (7) KTI is a monomeric and non-glycosylated protein weighing 21.0 kDa (8). Soybean trypsin inhibitor (Kunitz) family inhibits proteases by binding with high affinity to their active sites. Kunitz Trypsin Inhibitor gene (KTI2) is one of the most genes which play a major role in production trypsin inhibitor enzyme, which site on chromosome 09 and has 2 exons. The trypsin inhibitor enzyme consists of 210-211 amino acids with 20000-25000 Dalton, and is expressed during embryogenesis and in mature plant in the seeds, and at low levels in leaf, root, and stem (16). This study intends to identify kti -2 in three soybean cultivars using a map to reference. It also aims to estimate the similarity between DNA and RNA sequences of KTI-2 in Wm82, Lee and ZH13 cultivars.

Material and methods

Three Soy bean assembled genome sequences were obtained from Blast Database Options (soybase.org) cultivar Williams 82 under accession Wm82.a2, cultivar Lee under accession Lee.a1 and cultivar Zhuonghuang 13 under accession ZH13.a1. The transcripts of Wm82, Lee and ZH13 were retrieved under same accessions above.The kti-2 sequence of genome and mRNA was downloaded from the Gene Bank under accession number NM_001254177 and NM_001250857 respectively.The assembled sequences were imported into Geneious Prime 2023 software(http://www.geneious.com/)(Kearse *et al.,* 2012). The kti-2 and its mRNA sequence mapped through the three genomes and transcripts using map to reference tool following settings; Geneious mapper and Geneious RNA mapper with Sensitivity: Medium-Low respectively and variant call was determined using find variation/SNPs tool in Geneious with minimum coverage 1 and default genetic code. The mapped sequence was extracted and then aligned using pairwise alignment tool to exactly determine the pairwise identity ratio.

Results and discussion

The size of examined genomes was 978, 1016 and 1020 Mb for Wm82, L33 and ZH13respectively.However, the assemblies of L33 and ZH13 were more representative than Wm82, and nearly covered the whole genomes Table (1).

Table (1): The sequence numbers and the entire length of the three examined genomes

Assembled genome	Sequences No.	Total length of genome (Mb)
Wm82.a2.v1.genome	1190	978
Lee.a1.genome	265	1016
ZH13.a1.genome	570	1020

The enzyme existed in one position in Ch.09 (Gm09) of Wm82, L33 with identical length 872 bp and 100% identity. While in ZH13, it's found in two copies, one was identical in length and identity in Gm09, and the other was shorter with 785 bp and 70% identity, interestingly existed in Gm16 Table (2), Further, the ch.09 was assembled with similar lengths so far in the three genomes, and it was reported that WM82 has 2049 genes, however the other two have not been studied their gene counts yet, in addition to ch.16 in ZH13 Table (3).

Table (2): Mapped trypsin sequence to three soybean genomes

Soy bean cultivar	Mapped sequence to DNA	Sequence length	pairwise identical
WM82	Gm09	872	100%
Lee	Gm09	872	100%
ZH13	Gm09 Gm16	872 785	100% 70%

Table (3): Size and GC content of chromosome 9 for the studied genotypes.

		WM82	Lee	ZH13		ZH13
Chr 09	**Genbank acc.**	CM000842.4	CM009354.1	CM010417.2	**Chr 16**	CM010424.2
	Size (bp)	50,572,668	50,397,291	51,981,488		38,219,656
	Genes N.	2049	-	-		-
	GC%	34.5%	34.5%	34.8%		34.4%

At the level of mRNA, three sequencings have been aligned for both Wm82 and Lee cultivars, and one sequencings for ZH13 Figure (1), The sequencings lengths for Wm82 were 273, 648 and 630 bp with 71.2%, 69.6% and 100% pairwise identity respectively. The aligned sequencings for Lee cultivar were 636, 648 and 630 bp with 72.5%, 72.8% and 100% pairwise identity respectively. While, the aligned for sequencings ZH13 cultivar was 630 bp with 100% identity. The sequence Glyma.WmG163800.1 of Wm82 cultivar, Glyma.Lee. G127900.1 of Lee cultivar, and G1yma ZH13.G146800.1 of ZH13 cultivar had equal length (630 bp) which showed complete identical with reference enzyme sequence Table (4),Three mRNA sequences were found in Wm82 and Lee cultivars, despite the current study proving that there is only one sequence identical to the gene for both cultivars. This is due to the RNA variation Ontology (VariO). This term descripts the variations (alteration) in DNA, RNA and protein that related to numerous cellular processes and functions ranging from catalysis to regulatory processes, from information transfer for protein synthesis to function in cellular machineries, from RNA base and sugar modification to RNA interactions, and so on (17). There are many types of RNA chain variations, i.e. RNA deletion, RNA Indel, RNA insertion, RNA subsstitution, RNA invesion, RNA translocation termination code change, etc (18). The amino acids length ranged between 91and 216amino acid, but 210 amino acids corresponded to 100% pairwise identity. The convergence of mRNA sequence length and amino acids length among the studied cultivars refers to appear some modifications in the structural gene.

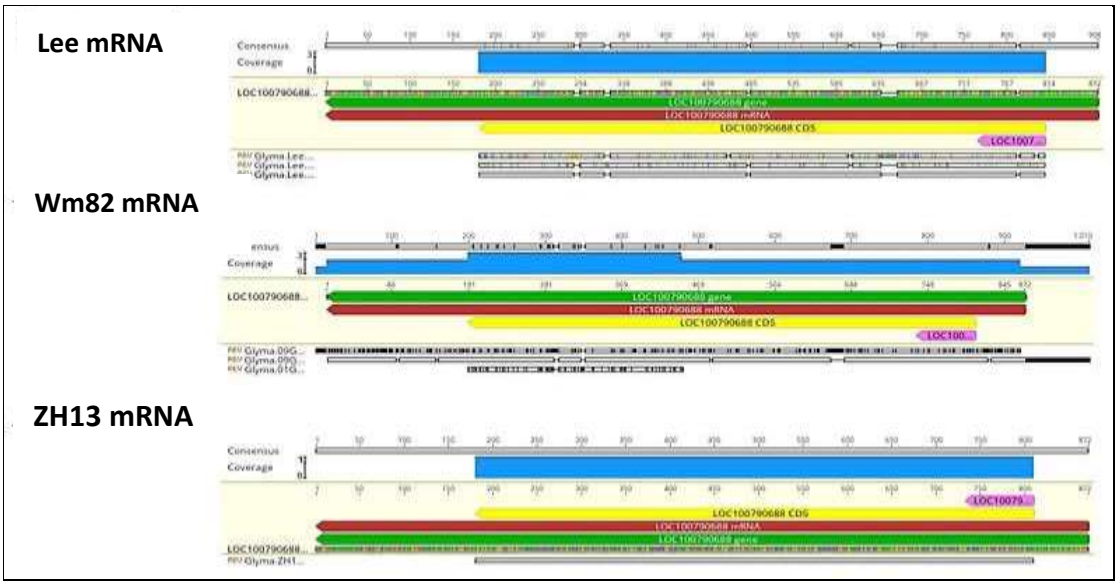

Figure (1): The KTI-2 sequence (NM_001254177) mapped over three genomes mRNA using map to reference tool in the Geneious prime software.

Table (4): Mapped trypsin sequence to mRNA of three soybean genomes.

Soy bean cultivar	Mapped sequence to RNA	Nucleotide Sequences Length/ mRNA	Pairwise identical	Amino Acid Sequences Length	GC content %	BLAST Protein
WM82	Glyma.Wm G01G117700.1	273	71.2%	91	38.8%	Trypsin and protease inhibitor
	Glyma.WmG163300.1	648	69.6%	216	46.0%	
	Glyma.WmG163800.1	630	100%	210	45.2%	
Lee	Glyma.Lee.G127500.1	636	72.5%	212	46.0%	Trypsin and protease inhibitor
	Glyma.Lee.G127600.1	648	72.8%	216	45.2%	
	Glyma.Lee.G127900.1	630	100%	210	43.7%	
ZH13	G1yma ZH13.G146800.1	630	100%	210	45.2%	Trypsin and protease inhibitor

The size of examined genomes was 978, 1016 and 1020 Mb for Wm82, L33 and ZH13 respectively (Table1), and those assemblies were approximately similar to the published soybean genome size 1.1–1.15 Gb (9). Kunitz Trypsin Inhibitor gene (KTI2) sits in chromosome 09 with close size between soybean cultivars and equal GC content, and interestingly, the assembled lengths of Ch.09 over the three genomes was approximately close Table (3), The association of trypsin inhibitor sequence as a reference sequence clearly indicated that this enzyme is specifically present in the studied cultivars. After mapping sequence to the DNA genomes has been done for each cultivar, KTI-2 sequence was found in all the studied cultivars with 100% pairwise identical and 872 bp in length on chromosome 09. Another sequence was found in ZH13 with 785 bp and 70% identical on chromosome 16. The KTI2 gene was found to be specifically present in all the studied cultivars. This suggests that the gene is a common feature in these soybean cultivars. After mapping the KTI2 sequence to the DNA genomes of each cultivar, it was observed that the KTI2

sequence in all studied cultivars on ch. 09 was 872 base pairs (bp) in length. Moreover, the sequence in all cultivars was 100% pairwise identical. This high sequence identity implies that the KTI2 gene in these cultivars is conserved and stable. In contrast to the uniform KTI2 sequence in the other cultivars, the ZH13 cultivar exhibited a sequence variation. The KTI2 sequence in ZH13 was found on chromosome 16 and was 785 bp in length. It also displayed 70% sequence identity compared to the reference KTI2 sequence. This suggests that ZH13 has a different variant or allele of the KTI2 gene compared to the other cultivars. Mapping to the extant reference RNA for three cultivars are illustrating in Figure (1), Three sequences were found for Lee cultivar with 5, 4 and 5 gaps for each conting; and three sequences for Wm82 cultivar which had 1,7 and 1 gaps; while one sequence was found for ZH13 cultivar with no gaps. The polymorphisms for each sequence in the figures are corresponding to pairwise identity that showed in the table 4. Multiple sequences for both Wm32 and Lee cultivars refer to existence duplicate genes.Gene duplication generates two gene copies; this theoretically allows one or both to evolve under reduced selective constraint and, on some occasions, to acquire novel gene functions that contribute to adaptation (10).

Genome sequence data have yielded paralog frequencies as high as approximately 75% in soybean based on mRNA data (11).Gene duplication remains of specific interest both because of the abundance of plant gene duplicates and their potential to contribute to plant novelties.Over time, if multiple cultivars independently develop the same trait (e.g., disease resistance) due to similar selective pressures, their mRNA sequences and amino acid lengths in the structural genes responsible for that trait may converge. In other words, they may show similarities or even identical changes at the genetic and protein level. The convergence can result from specific mutations or alterations in the DNA sequence of the structural genes.

These changes may lead to similar or identical amino acid sequences in the corresponding proteins. Cultivars with these convergent genetic and protein changes would exhibit the same trait, even if they originated from different genetic backgrounds (12,13 and 14).

Conclusion

In conclusion, the findings indicate that the KTI2 gene is a stable and conserved feature in the studied soybean cultivars, with a 100% identical sequence on chromosome 09 in all of them. However, ZH13 stands out with a different variant of the KTI2 gene on chromosome16, showing some sequence variation. These results can be valuable for understanding the genetic diversity of KTI2 in different soybean cultivars and may have implications for soybean breeding and research.

References

1. Wilson, R. F. (2008). Soybean: market driven research needs. In Genetics and genomics of soybean (pp. 3-15): New York, NY: Springer New York.
2. Maranna, S.; Verma, K.; Talukdar, A.; Lal, S. K., Kumar, A. and Mukherjee, K. (2016). Introgression of null allele of Kunitz trypsin inhibitor through marker-assisted backcross breeding in soybean (*Glycine max L. Merr.*). BMC genetics, 17(1): 1-9.
3. Rashed, N. A., MacDonald, M. H., and Matthews, B. F. (2008). Protease inhibitor expression in soybean roots exhibiting susceptible and resistant interactions with soybean cyst nematode. Journal of Nematology, 40(2): 138.
4. Armour, J. C.; Perera, R. C.; Buchan, W. C.; and Grant, G. (1998). Protease inhibitors and lectins in soya beans and effects of aqueous heat-treatment. Journal of the Science of Food and Agriculture, 78(2): 225-231.

5. Krishnan, H. B. (2001). Characterization of a soybean (*Glycine max (L.) Merr.)* mutant with reduced levels of Kunitz trypsin inhibitor. Plant Science, 160(5): 979-986.
6. Liener, I. E. (1995). Possible adverse effects of soybean anticarcinogens. The Journal of Nutrition, 125: S744-S750.
7. Omran, R. (2007). Partial purification and characterization of protease IV from *Psudomonas aeruginosa*. Iraq Journal Biotechnology, 6(1), 94-106.
8. Shawkat, A. A. K. M. S. (2022). Primary and Secondary Screening of *Pseudomonas aeruginosa* for Protease Production. Iraqi Journal of Biotechnology, 21(2).
9. Shultz, J. L.;Kurunam, D.; Shopinski, K.; Iqbal, M. J.; Kazi, S.; Zobrist, K.; ... and Lightfoot, D. A. (2006). The Soybean Genome Database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max. Nucleic Acids Research, 34(:D758-D765.
10. Panchy, N.; Lehti-Shiu, M.; and Shiu, S. H. (2016). Evolution of gene duplication in plants. Plant physiology, 171(4):2294-2316.
11. Schmutz, J.; Cannon, S. B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; ... and Jackson, S. A. (2010). Genome sequence of paleopolyploidsoybean. Nature, 463(7278):178-183.
12. Lavoie, H.; Debeane, F.; Trinh, Q. D.; Turcotte, J. F.; Corbeil-Girard, L. P.; Dicaire, M. J.; *et al.*, (2003). Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. Human Molecular Genetics, 12(22): 2967-2979.
13. Vanderplank, J. E.; (2012). Disease resistance in plants. Elsevier.
14. Stern, D. L. (2013). The genetic causes of convergent evolution. Nature Reviews Genetics, 14(11): 751-764.
15. Tanwar, B.; and Goyal, A. (Eds.). (2021). Oilseeds: health attributes and food applications. Berlin/Heidelberg, Germany: Springer.
16. Mokoena, T. (2010). Establishment of a transformation procedure to study the role of trypsin inhibitors in soybean (Doctoral dissertation, University of Pretoria)
17. Vihinen, M. (2021). Systematics for types and effects of RNA variations. RNA Biology, 18(4): 481-498.
18. Yip, K. L; Chan, S. Y.; Ip, W. K.; and Lau, Y. L.; (2000). Bruton's tyrosine kinase mutations in 8 Chinese families with X-linked agammaglobulinemic. Human Mutation, 15(4): 385-385.