

THE *IN SILICO* PREDICTON OF THE CHLOROPLAST MATURASE K GENE POLYMORPHISM IN SEVERAL BARLEY VARIETIES

KAMIL M. MUSTAFA¹, MUFEED J. EWADH², MOHAMMED BAQUR S. AL-SHUHAIB^{3*},
HAMID G. HASAN⁴

¹Dept. of Field Crops, Faculty of Agriculture Science, Al-Sulaimanya University-Iraq

²Dept. of Clinical Biochemistry, College of Medicine, Babylon University, Hilla-Iraq

³Dept. of Animal Production, College of Agriculture, Al-Qasim Green University, Al-Qasim, Babil-Iraq

⁴University College of Humanity Studies, Al-Najaf-Iraq

MUSTAFA, K.M. – EWADH, M.J. – AL-SHUHAIB, M.B.S. – HASAN, H.G.: The *in silico* prediction of the chloroplast maturase K gene polymorphism in several barley varieties. *Agriculture (Poľnohospodárstvo)*, vol. 64, 2018, no. 1, pp. 3–16.

This study was conducted to describe the role of the chloroplast maturase K (*matK*) genetic polymorphism in the reciprocal crossing between five barley varieties using several *in vitro* / *in silico* tools. Besides, the final consequences of the *matK* gene polymorphism on its protein structure, function, and interactions were predicted computationally. Five parental varieties were crossed to each other by full reciprocal crossing design, DNA was extracted from seeds and two different primers' pairs were designed to scan *matK* gene. Then, polymerase chain reaction – single-stranded conformation polymorphism (PCR-SS-CP) were performed. Two distinct haplotypes in both parents and artificial F1 hybrids in the *matK* gene were observed in both amplified fragments. This finding indicated that the studied gene had no participation in the reciprocal crossing performed. Three SNPs were identified; two of them are non-synonymous (nsSNPs), namely G387V and L459M. The effect of these missense mutations on the *matK* protein was analyzed by several *in silico* tools. It was shown that the coding SNP, L459M was predicted to have much more effective consequences on *matK* protein structure and function. While the I-Mutant 2.0 prediction tool showed a decrease in stability for these two nsSNPs, which may destabilize the protein interactions to some extent. In conclusion, though the observed missense mutations in the *matK* gene have no suggestive role in the reciprocally crossed barley varieties, they caused dramatic alterations in several *matK* protein moieties, which may lead to potential subsequent changes in the *matK* protein-mediated RNA splicing mechanisms.

Key words: chloroplast, *in silico*, *in vitro*, *Hordeum vulgare*, maturase k

Barley is highly variable in adaptation and utilization alone has caused barley to be subjected to more physiological and genetic studies (Von Bothmer *et al.* 2003). The high diversity of barley phenotypes, true diploid nature, and ease of hybridization and cultivation has made barley species the favorite genetic organism. Moreover, mutations can more easily be identified in barley than in its polyploid relatives, such as wheat and oats (Scholtz *et al.* 2001). The reciprocal crossing analysis provides a unique opportunity to obtain a rapid and overall

picture of the genetic control of a set of parents in the early generation (Crusio 1987). One of the symptoms of revolutionizing angiosperm molecular genomics is the development of new sets of genetic markers to genotype barley as long as other essential plant breeds (Yu *et al.* 2011). Several candidate markers are targeted on the level of DNA architecture, one of these candidate genes is chloroplast maturase K (*matK*) gene. It was reported that *matK* gene might be contributed to plant molecular systematics and evolution (Johnson & Soltis

Mohammed Baqur S. Al-Shuhaib (*Corresponding author), Department of Animal Production, College of Agriculture, Al-Qasim Green University, 8 Al-Qasim, Babil governorate 51001, Iraq. E-mail: mohammed79@agre.uoqasim.edu.iq; baquralhilly_79@yahoo.com

1995; Liang & Hilu 1996). The *matK* gene, 1536 bp long, is located within the intron of the chloroplast gene *trnK*, on a large single-copy section adjacent to the inverted repeat (Hilu & Liang 1997). Maturases contain three domains: a reverse-transcriptase domain, domain X (the RNA binding and splicing domain), and a zinc-finger-like domain (Mohr *et al.* 1993; Barthet *et al.* 2015). It's reported that maturases are enzymes that catalyze non-autocatalytic intron removal from premature RNAs (Zoschke *et al.* 2010). Therefore, the *matK* gene may be a strong candidate marker for evaluation of genetic polymorphism. The *matK* gene with its underlying features represents a molecule that has substantial potential in providing insight into evolutionary and systematic problems at variable levels (Hilu & Liang 1997). Moreover, the ability to use *matK* gene in the nucleotide polymorphism is reported (Barthet & Hilu 2007). However, very little genetic data is currently available about the *matK* gene polymorphism in barley varieties around the world since no sufficient data that dealt with this hotspot genetic locus in *H. vulgare* varieties are available at least in terms of their genetic variability. Besides, no association study has highlighted this segment as a candidate genetic marker about its relationship with the main variable identities of *H. vulgare* variations. Thus, employing *matK*, as a gene included in the chDNA in the genetic polymorphism could be considered as an interesting genetic marker. In crops, the detection of the single nucleotide polymorphisms (SNPs) permits a more accurate approach to the analysis of sequence differences between alleles (Rafalski 2002). To manifest this practically, several techniques are employed to exploit the portions of this gene to detect the extent of polymorphisms in a way that could be used in routine labs, such as SSCP. Although PCR-SSCP is not the only reliable technique to genotype breeds of different organisms and other available methods are used instead, its performance is one of the simplest and the most sensitive techniques in the detection of unknown DNA mutation (Gasser *et al.* 2006). Hence, the utilization for SSCP polymorphisms could lead to the finding of useful genetic markers of agricultural populations (Kuhn *et al.* 2008). Thus, much more attention is gained in the *matK* gene to be used as an attractive tool to study the genetic polymorphism in this aspect. On

the other hand, it is apparently demonstrated that many SNPs have functional effects on their corresponding protein structure especially by a single change in the amino acid (Liao & Lee 2010). There is an increasing awareness that other types of genetic variations can affect several aspects of protein expression and function. Synonymous SNPs, (sSNPs) or silent' mutations, are now widely acknowledged to be able to cause changes in protein expression, conformation, and function (Sauna & Kimchi-Sarfaty 2011). However, sSNPs do not alter amino acid sequences. Conversely, a non-synonymous single nucleotide polymorphism (nsSNP), which is present within the exon of a gene, is responsible for the incorporation of other amino acid and known to be one of the leading causes for the possible alteration of the encoded protein. However, tolerant nsSNPs are not deleterious and are not involved in any change, whereas deleterious nsSNPs have a profound influence on protein structure or its interaction (Pauline & Steven 2003). Therefore, it is important to differentiate deleterious nsSNPs from tolerant nsSNPs to characterize the genetic basis of these biochemical pathways. By using bioinformatics prediction tools, it would be possible to segregate neutral SNPs from SNPs of likely functional effect and could also be useful to explain the structural basis of the desired and undesired variations in *matK* protein characterizations. Thus, employing several computational tools in this gene could be considered as interesting genetic markers to give a perspective bioinformatics insight to the nature of the genotypes within this population. Therefore, two main objectives of this study are pursued. The first one is to see the potential participation of the *matK* gene polymorphism in the reciprocal crossing between five different barley varieties. As well, to predict the potential functional effect of the observed nsSNPs and to speculate their possible influence on protein structure and function using several state-of-art *in silico* computational tools.

MATERIAL AND METHODS

Plants and experiment materials

The field study was conducted in Kurdistan region – Iraq, Qilyasan Agricultural Research Station,

T a b l e 1

The primer BLAST system that is employed to cover most of *matK* gene sequences by designing two pairs of primers

No.	Primer sequence	Position within <i>matK</i> genbank accession number: AB078138	Amplicon length
1	FP: 5'- AGGGGTTTTCGATTGTTGTG -3' RP: 5'- TGCCCCGATCCAGTACAAAA -3'	Start 302, Stop 321 Start 1169, Stop 1150	868 bp
2	FP: 5'- ACTGGATCGGGGCATCCTAT -3' RP: 5'- CCGAACCCAATCGTTGCATA -3'	Start 1156, Stop 1175 Start 1384, Stop 1365	229 bp

Faculty of Agricultural Sciences, University of Sulaimani (35°34'307" N, 45°21'992" E and 765 masl), 2 km Northwest of Sulaimani city, during the autumn growing seasons 2010–2011(crossing between parents at first location only to produce the first filial at the three locations at 2011–2012). Five varieties and pedigrees of two-rowed *H. vulgare* were used in this study, which includes MORA, ABN, Arabi Aswad, Clipper, and Bohoth H1. These varieties were crossed in full diallel mating design to form 20 Artificial F1 hybrids (Online Suppl. Table 1). All the Artificial F1 hybrids along with their parents were grown in the following growing season. Seeds of 20 F1s with their 5 parents (25 entries) were sown in the field experiments; they were conducted in a randomized complete block design (RCBD) with three replications. Each treatment was one row of 2-meter length, 40 cm between rows and 20 cm between plants within a row. The major features of these five varieties were mentioned in concomitance with their main sources (Online Suppl. Table 2).

DNA Samples

Total genomic DNA was obtained from twenty-five varieties (parents and artificial F1 hybrids) of barley seeds. DNA was isolated from about 25 mg seed using genomic DNA mini kit – Plant (Cat # GP100, Geneaid Biotech – Taiwan) following manufacturer's instructions. The purity of DNA was assessed by NanoDrop spectrophotometer, version BioDrop µLITE (Biodrop – UK). The integrity of isolated DNA was checked by ethidium bromide (0.5 mg/ml) pre-stained 0.8% (w/v) agarose gel electrophoresis in 1X TAE (40 mM Tris-acetate; 2 mM EDTA, pH 8.3) buffer. The isolated DNA was used as a template for PCR.

PCR Primers Designing and Amplification

Two fragments were designed for the *matK* gene

(1536 bp length) from GenBank accession number: AB078138 of NCBI website (www.ncbi.nlm.nih.gov), using Primer³ online software (<http://www.simgene.com/Primers>). These two fragments cover most of the *matK* gene (Table 1). The lyophilized primers were purchased from Bioneer (Bioneer, Daejeon, South Korea). PCR reaction was performed using the *AccuPower* PCR premix (Cat # K-2012, Bioneer – South Korea). Each 20µl of PCR premix was contained 1 U of *Top* DNA polymerase, 250 µM of four dNTPs, 10 mM of Tris-HCl (pH 9.0), 30 mM of KCl, 1.5 mM of MgCl₂. The PCR reaction mixture was completed with 10 pmol of each primer and 50 ng of genomic DNA. The following program was applied in gradient PCR thermocycler (Mastercycler-nexus, Eppendorf, 22331Hamburg). The optimum annealing temperature for both designed PCR amplicons was determined by gradient PCR. Then, the amplification was began by initial denaturation (94°C for 5 min), followed by 30 cycles of denaturation (94°C for 30 sec), annealing (52°C for 30 sec), and elongation (72°C for 30 sec), and was finalized with a final extension (72°C for 10 min). Amplification was verified by electrophoresis on an ethidium bromide (0.5 mg/ml) pre-stained 1.5% (w/v) agarose gel in 1 × TBE buffer (2 mM of EDTA, 90 mM of Tris-Borate, pH 8.3). It was made sure that all PCR resolved bands were distinct and consisted of only one clean and sharp band to be submitted into SSCP successfully (Online Suppl. Figure 1).

SSCP analysis

The SSCP experiments were initially based on Al-Shuhaib SSCP protocol (Al-Shuhaib 2017), with some modifications. Different gel concentrations, sample volumes, electrophoresis temperatures were tested to optimize separation power of SSCP experiments. Briefly, 7 µl of each amplification product

was mixed with 10 µl of SSCP denaturing loading buffer (95% formamide, 20 mM EDTA pH 8, 0.05% xylene cyanol and 0.05% bromophenol blue). The samples were heat-denatured at 95°C for 10 min and chilled on ice for at least 5 min, and loaded polyacrylamide gel. The preliminary SSCP experiments were performed in the commercially available mini gels (10 × 10 × 1) mm version; OminiPage (Clever Scientific – UK). After optimization, PCR amplicons were separated onto mini-high resolution aluminum plates vertical gel format, gel size (W × L) cm: 12 × 14.5, and gel thickness: 1 mm (model JY-SCZ9, Junyi-Dongfang Electrophoresis Equipment – China). Denatured PCR products were loaded into the wells of 8% acrylamide/bis (37.5:1), containing 7% glycerol, and 1X TBE buffer. The gel was run under 225 V / 112 mA at room temperature until the tracking dye reached the end of the gel. Each SSCP gel was stained by a rapid silver staining technique (Byun *et al.* 2009).

DNA sequencing and sequencing analysis

Each template DNA sample that showed a different pattern on the SSCP gel for the amplified *matK* gene fragments was sequenced from both ends (Macrogen Inc. Geumchen, Seoul, South Korea). Only clear chromatographs obtained from ABI sequence files were further analyzed, ensuring that the annotation and variations are not because of PCR or sequencing artifacts. The referring *Hordeum vulgare matK* database (GenBank acc. AB078138.1) was retrieved from the NCBI website (<https://www.ncbi.nlm.nih.gov/nuccore/AB078138.1>). The sequencing results of the PCR products of different SSCP patterns were edited, aligned, and analyzed as long as with the respective sequences in the reference database using *BioEdit* Sequence Alignment Editor Software Version 7.1 (DNASTAR, Madison, WI, USA).

Constructing the primary structure of the altered matK protein

The primary protein structure building of each SSCP genotype was begun by mutating the available reference NCBI DNA sequences of the *matK* gene, by substituting each observed SNP from the *BioEdit* /Lasergene software into its accurate position in the whole reference the *matK* DNA sequence to represent each altered variant. The observed vari-

ation was translated into amino acids in a reading frame corresponds to the reference *matK* amino acid sequences using the *Expasy* online program (<http://web.expasy.org/translate/>). Multiple amino acid sequence alignment was made between the reference *matK* amino acid sequences and its observed variations using the Clustal Omega program from the *UniProt* website (<http://www.uniprot.org/align/>).

Finding the deleterious effect of the matK protein nsSNP using SIFT

The amino acid sequence of *matK* along with non-synonymous SNPs (nsSNPs) with corresponding amino acid positions was submitted using SIFT (sorting intolerant from tolerant) program (Pauline & Steven 2003). The SIFT prediction tolerance index (TI) score was ranged from 0.0 to 1.0, which denoted the normalized probability that the amino acid change was tolerated. Substitutions at each position with normalized probabilities less than a tolerance index of 0.05 were predicted to be deleterious or intolerant; those greater than or equal to 0.05 were predicted to be tolerated (http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html).

Predicting the functional effect of nsSNPs using PolyPhen-2

PolyPhen or Polymorphism Phenotyping is a tool which predicts possible impact of an amino acid substitution on the structure and function of the protein using straightforward physical and comparative considerations. PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>), is a new development of the PolyPhen tool for annotating coding nonsynonymous SNPs (Adzhubei *et al.* 2010). Prediction outcomes could be classified as probably damaging or benign according to the score ranging from 0–1 respectively.

Prediction of functional impact of the matK nsSNP using PROVEAN

Further confirmation of the effect of nsSNPs on protein was done using PROVEAN (Protein Variation Effect Analyzer) tool. PROVEAN tool can predict the impact of an amino acid substitution on the biological function of the protein (<http://provean.jvvi.org/index.php>). The default threshold of PROVEAN tool is –2.5, *i.e.*, variants with a score equal to or below –2.5 are considered “deleterious,”

while variants with a score above -2.5 are considered “neutral” (Choi *et al.* 2012).

Prediction of the severity effect of the matK nsSNP using SNAP2

SNAP2 is a tool that can distinguish between the effect and the neutral nsSNPs (Smigielski *et al.* 2000). Predicting a score (ranges from -100 strong neutral prediction to $+100$ strong effect prediction), the analysis prediction scores were potentially correlated with the severity of effect (<https://www.predictprotein.org>).

Investigation of altered protein stability of the matK nsSNPs using I-Mutant 2.0

To have a better insight to the stability of the protein caused by mutation, the altered positions were analyzed using I-Mutant 2.0 (Capriotti *et al.* 2005). The study of the effect of a mutation may alter the stability of the matK protein of interest and might lead to a change in the main characteristics of this protein. I-Mutant 2.0 (<http://folding.biofold.org/cgi-bin/i-mutant2.0>) is a tool for the prediction of protein stability changes upon missense mutations.

Three-dimensional matK structure modeling using RaptorX and PyMol

The whole amino acid sequences of barley matK were retrieved from NCBI (GenBank acc. No. AB078138.1). The UniProtKB/Swiss-Prot accession number for barley matK protein was P17158, and no matching PDB entries were found in this protein (<http://www.uniprot.org/uniprot/P17158>). The three-dimensional (3-D) structure of the matK gene was constructed from the online 3-D model prediction software; namely RaptorX (<http://raptorx.uchicago.edu/StructurePrediction/predict/>). The observed mutations were inserted in the native sequence of the entire 511 amino acid residues matK protein to represent the observed variation. Then, the 3-D structures of the native matK protein, as well as the observed alteration were predicted using RaptorX tool. For given FASTA query sequence, RaptorX predicted its secondary and tertiary structures as well as solvent accessibility and disordered regions (Kallberg *et al.* 2012). The virtually proposed changes within its corresponding altered proteins were performed by using PyMol-v1, 7.0.1 software (www.shrodinger.com).

Prediction of some physio-chemical properties of the matK protein

A novel META-Disorder prediction method, or MD tool, molds various sources of information predominantly obtained from orthogonal prediction methods, to significantly improve in performance over its constituents (Schlessinger *et al.* 2009). Profisis tool, which is a web server that can predict protein-protein and protein – polynucleotide binding sites, was utilized to identify the interacting residues from sequence alone (Ofrañ & Rost 2007). Both tools that utilized for these predictions were available on (<https://www.predictprotein.org/>).

Post-translation modification sites present on matK protein

Glycation sites of ϵ amino groups of lysine residues were predicted using a NetGlycate1.0 server (<http://www.cbs.dtu.dk/services/NetGlycate/>). In NetGlycate a score of $N0.5$ was considered glycosylated. Phosphorylation sites were predicted using a NetPhos3.0 server (<http://www.cbs.dtu.dk/services/NetPhos/>). In NetPhos3.0, serine, threonine, and tyrosine residues with a score of $N0.5$ were considered phosphorylated. Ubiquitylation sites were predicted using UbPerd (www.ubpred.org). In UbPerd, lysine residues with a score of ≥ 0.62 were considered ubiquitylated. Sumoylation sites were predicted using SUMOplot (<http://www.abgent.com/sumoplot>). For SUMOplot, high probability motifs having a score of 0.5 were considered sumoylated (Jenna & Stephen 2014).

RESULTS

To identify the reason that lies behind the variations of barley, it was deduced to study the chloroplast matK protein as a possible candidate in this regard. Accordingly, we have artificially created variable entries of barley in this study to see – if any – possible participation of the matK gene in this reciprocal crossing between five different parental varieties (Figure 1, A). By designing two specific PCR primers pairs, 390 out of 511 amino acid residues were covered, which constitute about 76% of the whole amino acid sequences of the chloroplast matK protein in twenty-five artificially crossed *H.*

vulgare varieties. The genetic polymorphism of each observed haplotypes was predicted computationally. Two sets of results were obtained in this

study; *in vitro* genotyping results through PCR-SSCP-sequencing experiments, and *in silico* prediction results through several bioinformatics tools.

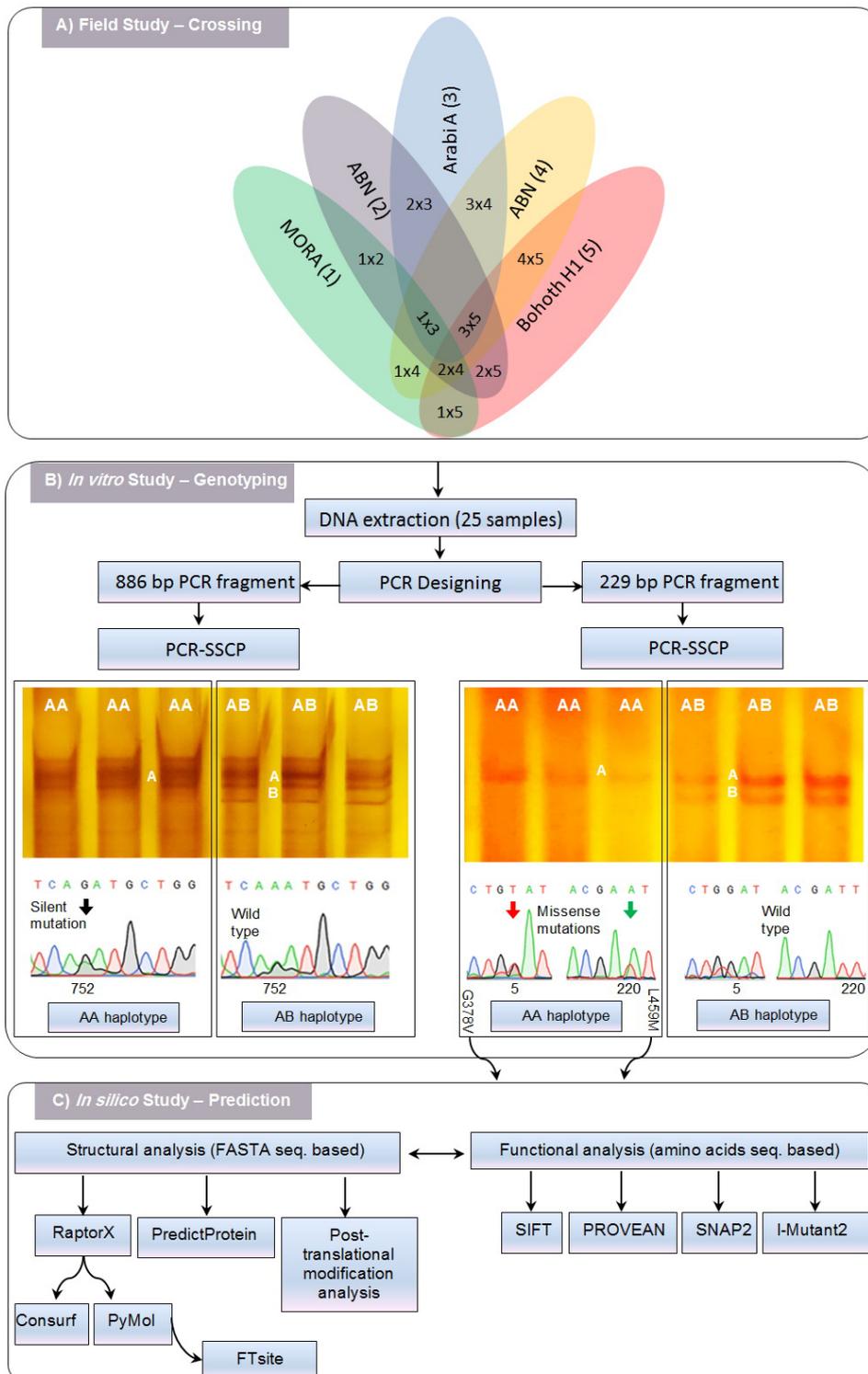


Figure 1. A full schematic workflow of the *matK* gene polymorphism study. This study consists of three main partitions; A) field study (crossing of the five different parents), B) *in vitro* genotyping through PCR-SSCP-sequencing of both designed amplicons of 868 bp and 229 bp, in which two genotypes were resolved in each case, and C) *in silico* computational analysis of the observed missense mutations.

T a b l e 2

Nucleotide substitutions and types among haplotype A, and haplotype B, and the reference NCBI sequence of *H. vulgare matK* gene

The size of amplicon	Nucleotide position in the NCBI reference <i>matK</i> gene	Nucleotide position in the PCR fragment	SSCP Haplotypes of <i>Hordeum vulgare matK</i> gene			Amino acid change	Type of SNPs
			Reference	A haplotype	B haplotype		
868 bp PCR fragment	1053	752	A	G	–	non	Synonymous
<i>matK</i> 229 bp PCR fragment	1160	5	G	T	–	G387V	Non-synonymous
	1375	220	T	A	–	L459M	Non-synonymous

Genotyping analysis

The polymorphism of *matK* gene was detected by PCR-SSCP and DNA sequencing methods, and the observed variations of SSCP gels indicate the detection of two different variations. In both studied PCR fragments, two haplotypes were resolved, AA and AB (Figure 1, B). The pattern and nature of each SNP that was detected by sequencing indicated that the AA variation has one novel sSNP (A1053G), while in the variation AB no SNP was observed (Table 2). The haplotype distribution is slightly predominated by AB haplotype. Despite the fact that the multitude of SSCP bands that we observed in this studied *matK* gene fragment, this is usually evident in some SSCP configurations that exceed 600 bp (Gasser *et al.* 2006). However, both haplotypes were available in both parents and offspring, suggesting a nonparticipation of *matK* large gene fragment in the reciprocal crossing that performed between parents. In this aspect, the situation was the same in *matK* the 229 bp gene fragment, in which two haplotypes were revealed. It was revealed that the haplotype AA has much more frequencies over variant AB. The variant AA was the predominant one since the number of variant AA in this study was fifteen samples in both parents and offspring, which constituted about 60%, while the variant AB was less common since it was included only ten samples, whereas no other samples showed any other noticeable variant even after repeating SSCP electrophoresis for several times. Sequencing results confirmed these two different electrophoretic SSCP patterns since two SNPs were detected between the two re-

solved variations and NCBI reference sequences. Both nsSNPs, (G1160T and T1375A), were found in AA variants. However, the functional nsSNPs were apparently shown only in the 229 bp PCR fragment out of the whole *matK* amino acid sequences. According to PCR-SSCP-Sequencing, two conformational patterns of both *matK* PCR fragments were observed in this study as they determined in parental and artificial F1 barley samples. However, it was revealed that both variants were the only genetic polymorphisms that were available for both parents and their artificial F1 hybrids.

Computational analysis

To get better exploration to the detailed genetic variation(s) within the *matK* gene, both SSCP-DNA sequencing variations of both PCR fragments were computationally analyzed by several *in silico* tools (Figure 1, C). While only one non-coding SNP is positioned in the AA variant of the 868 bp PCR fragment, two coding SNPs that were available only in the AA variants of the 229 bp PCR fragment. In both nsSNPs, the hydrophobic aliphatic amino acids Gly and Leu in the positions 387 and 459 of the whole *matK* proteuren (Figure 2, A) are substituted into another amino acids of the same group, which is Val and Glu respectively (Figure 2, B). The consequences of these nsSNPs were evaluated using different publicly available computational algorithms, namely, SIFT, PolyPhen-2, SNAP2, PROVEAN, and I-Mutant 2 bioinformatics tools (Table 3). However, by comparing the prediction of these methods for each observed nsSNP, vari-

able results were observed. Usually, the SIFT tool is utilized first, which predicts whether an amino acid substitution affects protein function among related genes and domains over evolutionary time (Ng & Henickoff 2006). SIFT predicts that G387V nsSNP

was found to be tolerated, which showed a highly acceptable tolerance index score (more than 0.05), which couldn't affect the protein function. While in the case of L459M, SIFT predicts an effect on protein function, which showed an unacceptable

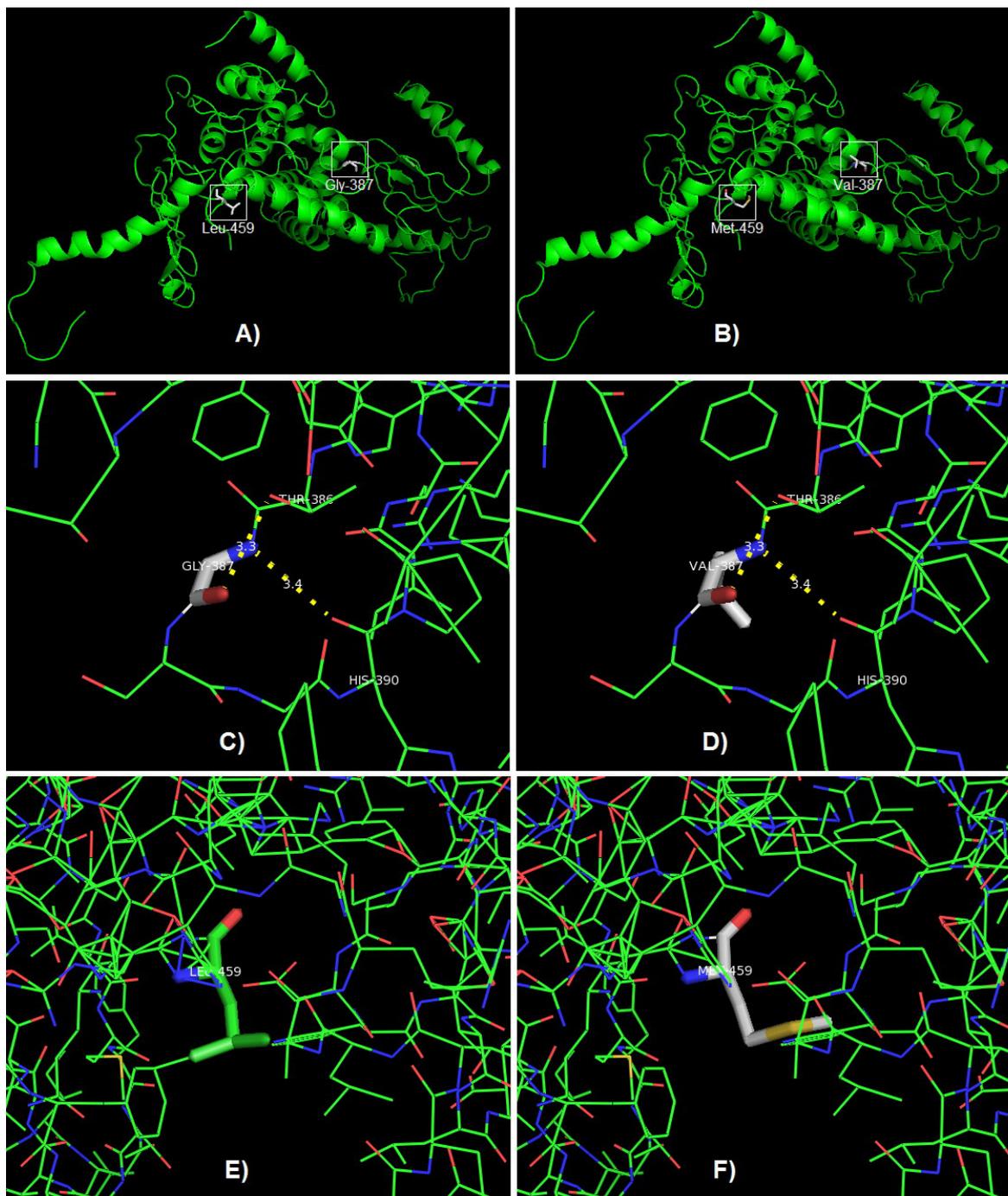


Figure 2. The proposed 3-D structure of matK protein of *H. vulgare* species with its polar interactions. A) The native protein is shown with two highlighted amino acids (G387 and L459). B) The altered protein is shown with two altered amino acids (V387 and M459). C) The interaction of S387 with both T386 and H390 before mutation. D) The interaction of V387 with both T386 and H390 after mutation. E) The absence of any polar connections in the L387 before mutation. F) The absence of any polar connections in the M387 after mutation.

T a b l e 3

List of the possible effects and consequences of the observed nsSNPs on matK protein structure and function as determined by four bioinformatics tools

nsSNP	SIFT	PolyPhen-2	PROVEAN	SNAP2	I-Mutant 2.0	
G387V	1	0.002 (sensitivity: 0.99) (specificity: 0.30)	4.470 ^{CF*}	-50	-1.22 ^{DDG*} (reliability index: 3)	Score
	Tolerated	Benign	Neutral	Neutral	Decrease	Prediction
L459M	0.00 ^{LC*}	0.98 (sensitivity: 0.75) (specificity: 0.96)	-0.730 ^{CF*}	-21	-1.14 ^{DDG*} (reliability index: 8)	Score
	Affect protein function	Probably damaging	Neutral	Neutral	Decrease	Prediction

^{CF*}; cutoff = -2.5, ^{LC*}: (LC; low confidence); this substitution may have been predicted to affect function just because the sequences used were not diverse enough. ^{RI*}; reliability index score. ^{DDG*}; DG(NewProtein)-DG(WildType) in Kcal/mol. ^{S.S*}; (Sensitivity: 0.99, Specificity: 0.30)

tolerance index score (less than 0.05), which could affect the protein function. However, the last one SIFT prediction had little confidence because the submitted *H. vulgare* matK protein sequence wasn't diverse enough for this program to give high confidence prediction. To predict the possible impact of these nsSNPs on both structure and function of the matK protein, the PolyPhen-2 program was used (Adzhubei *et al.* 2010). In agreement with SIFT results, PolyPhen-2 has predicted the same outcomes of these two SNPs. Thus, based on both SIFT and PolyPhen-2, the missense mutation L459M would be of prime importance concerning both its structural and functional effect on the matK protein. The SIFT/ PolyPhen-2 predicted characterization of G387V, and L459M nsSNPs were further validated using PROVEAN tool (Choi *et al.* 2012). The default threshold of PROVEAN tool is -2.5, *i.e.*, variants with a score equal to or below -2.5 are considered "deleterious", while variants with a score above -2.5 are considered "neutral". In contrast to SIFT/ PolyPhen-2 variable scores, PROVEAN revealed that both these nsSNPs weren't deleterious. It predicts that G387V and L459M nsSNPs are neutral (below -2.5). The SNAP2 program was also used to assess the effect of both SNPs on the matK protein sequence. In agreement with PROVEAN program, SNAP2 was found that both G387V and L459M were neutral (less than 0). Nevertheless, in the last two programs, there is a slight tendency of L459M to be more effective. However, it's interesting to note that when these four programs were used

together, we observed 100% agreement of the SIFT/ PolyPhen-2/PROVEAN/SNAP2 results and 50% of agreement of these programs on the G387V program. To add another layer of confirmation, these two variants were submitted to the I-Mutant-2.0 web server to predict the DDG stability and reliability index (RI) upon mutation. If the DDG value is >0, protein stability decreases and when DDG value is <0 protein stability increases. The I-Mutant 2.0 program showed a reduced stability of the matK protein for both two nsSNPs. Thus, there are higher chances that protein stability might get affected in mutation at two amino acid positions; 378 (DDG score -1.22), and 459 (DDG score -1.14).

The 3-D structure of the native matK model that generated through RaptorX (Kallberg *et al.* 2012) was revealed that all 511 (100%) residues were modeled, and only 1 (0%) positions were predicted as disordered, while the mutated protein has 25 (4%) positions were predicted as disordered. Secondary structures of the native matK model revealed 50% helix, 9% beta sheet, and 40% loop structures, while the secondary structure of the altered protein model showed 42% helix, 10% beta sheet, and 47% loop structures. The protein solvent accessibility was divided into three states by 2 cut-off values: 10% and 42%. A value less than 10% was identified as buried, larger than 42% value was identified as exposed and if the value was between 10% and 42% was designated as a medium. Regarding solvent access, noticeable differences were found between native and altered protein. The proportions of

exposed, medium and buried regions in the native matK protein were 37%, 15%, and 47% respectively, while in the altered protein was 31%, 37%, and 30% respectively. The nature of the polar interaction of the observed three nsSNPs was tested by PyMol to unravel its possible role in the conversion of the nature of the native amino acid binding with its surrounding residues in the matK protein.

It was found that G387 in the native protein has two polar interactions with Thr386 and His390 of

3.3 Å and 3.4 Å lengths respectively (Figure 2, C). While the altered amino acid 387V forms the same polar interactions with the same residues and distances (Figure 2, D). In the case of L459M nsSNP, it was found that L459 in the native protein doesn't have any polar interactions (Figure 2, E); and the situation is the same for the altered 459M (Figure 2, F).

It is necessary to identify the disordered regions in proteins from their amino acid sequences since

T a b l e 4

Protein disorder possibility prediction of the S387V and L459M nsSNPs in the matK protein

Number	Residue	NORSnet	NORS2st	PROFbval	Ucon	Ucon2st	MD_raw	MD_rel	MD2st
387	G	0.31	–	0.55	D	0.19	–	0.263	8
459	L	0.08	–	0.55	D	0.14	–	0.263	8

Number – residue number; Residue – amino acid type; NORSnet – raw score by NORSnet (prediction of unstructured loops) NORS2st – two-state prediction by NORSnet; D = disordered; PROFbval – raw score by PROFbval (prediction of residue flexibility from sequence); Bval2st – two-state prediction by PROFbval; Ucon – raw score by Ucon (prediction of protein disorder using predicted internal contacts); Ucon2st – two-state prediction by Ucon; MD_raw score by MD (prediction of protein disorder using orthogonal sources); MD_rel – reliability of the prediction by MD; values range from 0–9. 9 = strong prediction; MD2st – two-state prediction by MD

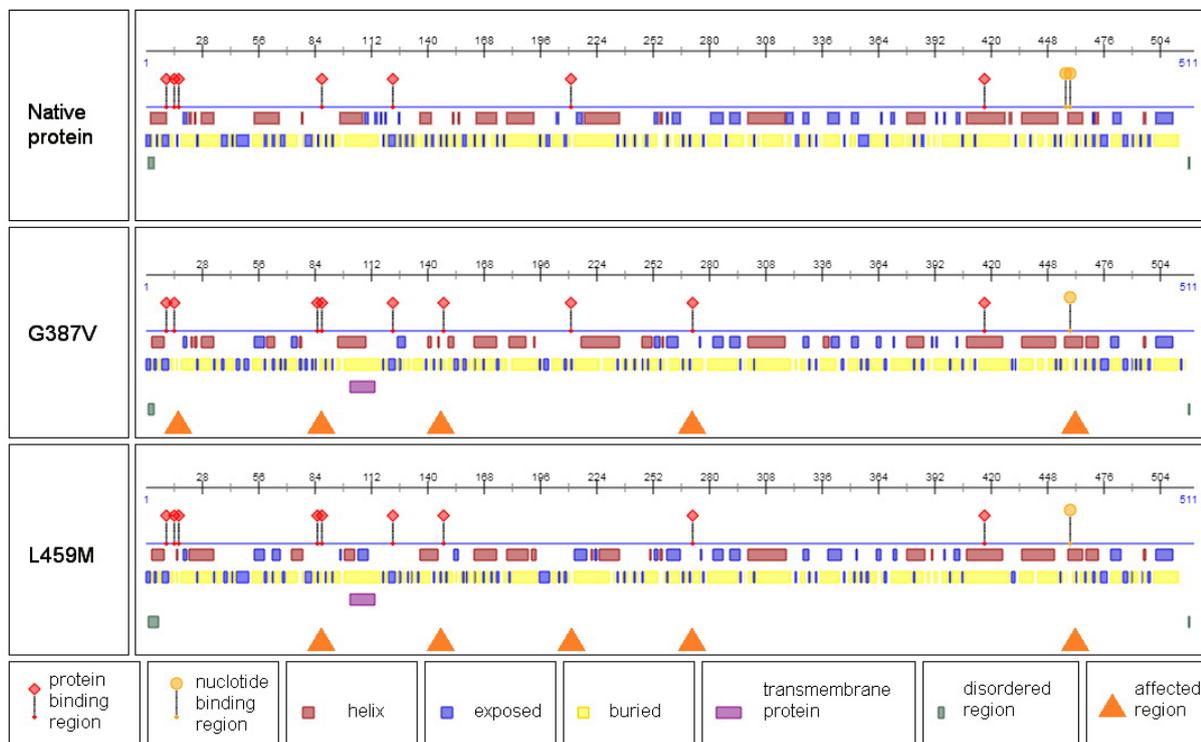


Figure 3. Prediction of protein-protein and protein-polynucleotides binding sites for native matK protein A), and its altered G387V mutant B), and L459M mutant proteins C)

the disordered regions are also important for the functional annotation of proteins. The available prediction MD tool has enabled us to use the “meta approach” to predict disordered regions (Schlessinger *et al.* 2009). Interestingly, MD program predicted that both of G387V and L459M occupied readily disordered positions in the native *matK* protein (Table 4). Since the disordered regions that have no stable structures without their partner molecules are often found in functional sites of proteins (Ward *et al.* 2004), a potential participation of these two nsSNPs in the binding with other proteins, such as transcription factors or polynucleotide binding proteins was predicted. This prediction was confirmed using proFISIS, which has shown noticeable differences between the native and the altered *matK* protein. In comparison to the native protein, 4 protein binding sites were altered in the mutant G37V and L459M proteins. Besides, one additional polynucleotide binding site and a transmembrane protein were observed (Figure 3). This observation indicates the possible interaction of G387V and L459M with the binding with other proteins and polynucleotides. Several *in silico* post-translational modification prediction tools, such as NetGlycate1.0, NetPhos3.0, UbPerd, and SUMOplot were utilized to identify the potential participation of G387V and L459M in the post-translational modification. No participation was observed for both nsSNPs in all studied post-translational glycation, phosphorylation, ubiquitylation, and sumoylation modification sites (Online Suppl. Table 3).

DISCUSSION

Initially, we were led to engage the *matK* gene polymorphism in this study because of several reasons; 1) The chloroplast *matK* DNA sequence, a chloroplast-encoded locus that has been reported to relatively have higher substitution rates at both nucleotide and amino acid level than other genetic loci (Barthet *et al.* 2007). 2) The *matK* gene has recently emerged as an invaluable gene in plant systematics because of the high phylogenetic signal it has compared with other genes used in this field (Muller *et al.* 2006). 3) Its protein is the only candidate for splicing introns from many genes in chloro-

plast (Zoschke *et al.* 2010). 4) Its worth mentioning that there has not been any sufficient data regarding the effect of *matK* genetic polymorphisms on the protein structure. Thus, for its essential and irreplaceable function in the plant makes the *matK* gene as an excellent marker that may deserve to deeply explored in genetic breeding.

Genotyping analysis

The sequencing that performed in both 686 bp and 229 bp *matK* gene fragments were confirmed the discrete genetic differences between them in the SSCP gel. These SNPs are novel and have not been previously reported in the *matK* gene of *H. vulgare* (<https://www.ncbi.nlm.nih.gov/projects/SNP/>). It deserves to note that there is no significant relationship between the identified SNPs patterns in *matK* gene and diallel and reciprocal crossing that performed among five different parental varieties. Although the SSCP results of *matK* 686 bp and 229 bp gene fragments are shown significant variations in barley varieties, no obvious consistent results were obtained regarding the pattern of crossing among these varieties. Nonetheless, distinct variable forms of polymorphisms were obtained in parents concerning both *matK* gene PCR amplified fragments. These distinct SSCP results might be referred to the presence of a wide ratio of hybridization, selection, or genetic diversity for barley (Schneider *et al.* 2001). The presence of the polymorphism in the *matK* gene does not imply that this gene has a role in the variations among the studied diallel or reciprocal crossed five varieties. This is evident since we found that both genetic varieties existed in both parental and F1 hybrids varieties. Therefore, the SNPs detected here may not be a causal mutation. Although no comparable standards were available in these local varieties, the results of this study in the barley samples have yielded an initial molecular marker that can be applied to identify the sort of polymorphism in other nearby barley samples, through relying on the amplification of both 686 bp and 229 bp *matK* gene PCR fragments. However, it is evident from this study that the use of 229 bp PCR *matK* gene fragment might be more informative than the 686 bp PCR fragment *matK* gene as the two observed nsSNPs of this study is readily detected using this fragment.

Computational analysis

The dramatic revolution in the computational prediction of the individual nsSNPs made it possible to generate an overview for the final destination for each nsSNP on the protein structure and function (Patel *et al.* 2015). As long as *in vitro* genotyping analysis, we have utilized *in silico* prediction tools to get a final estimation of the possible role of each observed nsSNP on the modulating of the main characteristic features of the *matK* protein. Three SNPs were discovered in this study; one is sSNP, two are nsSNPs. It was found through several computational tools that the effect of these SNPs is ranged from non-effective into mild-effective on protein structure and function. As it was known, the understanding the roles of some of the nsSNPs can significantly help to understand the barley genotype/phenotype variations. We opted to use *in silico* tools based on a combination of different algorithms that previously reported for other genes for the analysis of the observed amino acid variations in the *matK* gene. As genomic variations view among barley is common sense (Hilu & Liang 1997), the nsSNPs that exist through the coding region of this genome can, therefore, become necessary since they modulate the amino acid composition (Liao & Lee 2010). Such changes can have an impact on protein structure, function, stability, and subcellular localization. The nsSNPs may initiate unique structures between population members in conferring different protein biological activities. Likewise, nsSNPs may affect the *matK* gene expression by inactivating the active sites of the *matK* protein through which it interacts with the exon-intron junctions during splicing. Therefore, the observed nsSNPs was given special attention in this study. To determine the functional effect of these two nsSNPs on *matK* protein, several widely used *in silico* tools were employed in this study, specifically SIFT, PolyPhen-2, SNAP2, PROVEAN, and I-Mutant 2.0 (Pauline & Steven 2003; Capriotti *et al.* 2005; Adzhubei *et al.* 2010; Smigielski *et al.* 2000; Choi *et al.* 2012) respectively. However, the utilized bioinformatics tools had given different predictions on the deleterious effect of both nsSNPs on the *matK* protein. The differences in the results of these several prediction tools are due to the differences in features utilized by the methods; therefore we would expect the outcomes

to occur dissimilar at some point (De Alencar *et al.* 2010). SIFT, PolyPhen-2, SNAP2, and PROVEAN tool had all given parallel results about the non-effectiveness of S387V on the *matK* protein function. Meanwhile, SIFT and PolyPhen-2 showed that L459M was effective and damaging, whereas both SNAP2 and PROVEAN had shown a neutral effect of this nsSNP. However, taking these notes into account, L459M is an effective missense mutation in comparison with S387V. Concerning the stability of the altered protein, I-Mutant 2.0 showed a decrease in stability for these two nsSNPs upon mutation. Thus, data this suggests that S387V and L459M variants of the *matK* gene could directly or indirectly modulate the amino acid interactions causing functional deviations of protein to some extent. In addition to the previously mentioned bioinformatics tools, the polar interaction of these nsSNPs with other residues was shown no modification to the binding pattern with other residues. This suggestion was validated through MD prediction tool (Schlessinger *et al.* 2009). It is important to identify the disordered or the ordered regions from their amino acid sequences since the prediction of such configuration is also important for the functional annotation of proteins. In addition to the involvement in many biological processes, such as transcription, translation, signaling and cell cycle control, the disordered regions are also involved in alternative splicing (Romero *et al.* 2006). Hence, this finding may have some effect on the *matK* protein splicing mechanism. Upon binding with ligands, disorder-to-order transitions are frequently observed, where the flexibility of the disordered regions may be necessary to specifically interact with multiple partners (Ishida & Kinoshita 2007). Though no internal amino acids connection were changed according to PyMol prediction, the differences observed between the native and altered *matK* protein in the external binding with other proteins and polynucleotides as determined by ISIS program may be reflected in its function in the splicing of exon-intron junctions too (Zoschke *et al.* 2010), which may lead to a change in the genetic expression. This observation may become evident during splicing when the *matK* protein domain X binds to intron to form a lariat structure. It does remain bound to the excised RNA to form a ribonucleoprotein particle

(Saldanha *et al.* 1999). The formation of this particle may be modulated as a result of the role of these two nsSNPs in the alteration of the binding of the matK protein with polynucleotides. Though the precise mechanism by which matK protein potentially interacts with other factors still not fully understood, this finding may pinpoint some initial data which may help in the accurate understanding of these motifs which may change the substrates with which they are interacting. Eventually, the bioinformatics tools helped us to characterize the impact of nsSNP on the *matK* gene and suggest that the *in silico* analysis may be a useful tool to predict the effect of DNA variation on gene function.

Eventually, it had been hoped that the genotyping analysis would exploit some sections of the genome which are highly variable, such as the *matK* gene, to be used for varieties differentiation particularly in diallel and reciprocal crossing. Nonetheless, this genetic segment doesn't differentiate between the different varieties since we have chosen a region that despite its confirmed variability it might not have a relation with the pattern of the genealogical variations among these diallel or reciprocal crossed varieties. Although the two resolved variants may not be connected with the nature of the crossed varieties, both *matK* amplified fragments have exerted only two distinct variants in all of these varieties. This study has computationally found a noticeable reduction in the matK protein stability upon mutation with both S378V and L459M. In addition to MD/ISIS prediction, a concomitant alteration in the binding with the external proteins and polynucleotides were predicted in these two observed nsSNPs too.

CONCLUSIONS

This study predicts that both S378V and L459M missense mutations of the *matK* gene may participate in the alteration of the matK protein splicing of intron-exon splicing mechanisms. This pilot study may determine the pattern of polymorphism within the *matK* gene to conceive its consequences within these highly important barley varieties.

Appendix. Supplementary data

Supplementary data associated with this article

can be found, in the online version, at doi: 10.2478/agri-2018-0001

REFERENCES

- ADZHUBEI, I.A. – SCHMIDT, S. – PESHKIN, L. – RAMENSKY, V.E. – GERASIMOVA, A. – BORK, P. – KONDRASHOV, A.S. – SUNYAEV, S.R. 2010. A method and server for predicting damaging missense mutations. In *Nature Methods*, vol. 7, no. 4, pp. 248–249. DOI: 10.1038/nmeth0410-248
- AL-SHUHAIB, M.B. 2017. A Universal, rapid, and inexpensive method for genomic DNA isolation from the whole blood of mammals and birds. In *Journal of Genetics*, vol. 96, no 1, pp. 171–176. DOI: 10.1007/s12041-017-0750-6
- BARTHET, M.M. – HILU, K.W. 2007. Expression of matK: functional and evolutionary implications. In *American Journal of Botany*, vol. 94, no. 8, pp. 1402–1412. DOI: 10.3732/ajb.94.8.1402
- BARTHET, M.M. – MOUKARZEL, K. – SMITH, K.N. – PATEL, J. – HILU, K.W. 2015. Alternative translation initiation codons for the plastid maturase MatK: unraveling the pseudogene misconception in the Orchidaceae. In *BMC Evolutionary Biology*, vol. 15, pp. 210. DOI: 10.1186/s12862-015-0491-1
- BYUN, S. – FANG, Q. – ZHOU, H. – HICKFORD, J. 2009. An effective method for silver-staining DNA in large numbers of polyacrylamide gels. In *Analytical Biochemistry*, vol. 385, pp. 174–175. DOI: 10.1016/j.ab.2008.10.024
- CAPRIOTTI, E. – FARISELLI, P. – CASADIO, R. 2005. I-Mutant 2.0: predicting stability changes upon mutation from the protein sequence or structure. In *Nucleic Acids Research* vol. 33, pp. 306–310. DOI: 10.1093/nar/gki375
- CHOI, Y. – SIMS, G.E. – MURPHY, S. – MILLER, J.R. – CHAN, A.P. 2012. Predicting the functional effect of amino acid substitutions and indels. In *PLOS ONE*, vol. 7, pp. e46688. DOI: 10.1371/journal.pone.0046688
- CRUSIO, W. 1987. A note on the analysis of reciprocal effects in diallel crosses. In *Journal of Genetics*, vol. 66, pp. 177–185. DOI: 10.1007/BF02927711
- DE ALENCAR, S.A. – LOPES, J.C.D. 2010. A comprehensive *in silico* analysis of the functional and structural impact of SNPs in the IGF1R gene. In *Journal of Biomedical Biotechnology* ID 715139, 8 pages. DOI:10.1155/2010/715139
- GASSER, R.B. – HU, M. – CHILTON, N.B. – CAMPBELL, B.E. – JEX, A.J. – ONTRANO, D. – CAFARCHIA, C. – BEVERIDGE, I. – ZHU, X. 2006. Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. In *Nature Protocols*, vol. 1, no. 8, pp. 3121–8. DOI:10.1038/nprot.2006.485
- HILU, K.W. – LIANG, H. 1997. The *matK* gene: sequence variation and application in plant systematics. In *American Journal of Botany*, vol. 84, pp. 830–839. url: <https://www.ncbi.nlm.nih.gov/pubmed/21708635>
- ISHIDA, T. – KINOSHITA, K. 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. In *Nucleic Acids Research*, vol. 35, pp. W460–464. DOI: 10.1093/nar/gkm363
- JENNA, N.K. – STEPHEN, D.B. 2014. *In silico* analysis of functional single nucleotide polymorphisms in the human TRIM22 gene. In *PLoS One*, vol. 9, no. 7, pp. e101436. DOI: 10.1371/journal.pone.0101436
- JOHNSON, L.A. – SOLTIS, D.E. 1995. Phylogenetic inference in saxifragaceae sensu stricto and gilia (Polemoniaceae) using matK sequences. In *Annals of the Missouri Botanical*

- Garden*, vol. 82, pp.149–175. DOI: 10.2307/2399875
- KALLBERG, M. – WANG, H. – WANG, S. – PENG, J. – WANG, Z. – LU, H. – XU, J. 2012. Template-based protein structure modeling using the RaptorX web server. In *Nature Protocols*, vol. 7, no. 8, pp. 1511–1522. DOI: 10.1038/nprot.2012.085
- KUHN, D. – MOTAMAYOR, J. – MEEROW, A. – BORRONE, J. – SCHNELL, R. 2008. SSCP markers provide a useful alternative to microsatellites in genotyping and estimating genetic diversity in populations and germplasm collections of plant specialty crops. In *Electrophoresis*, vol. 29, no. 19, pp. 4096–5108. DOI: 10.1002/elps.200700937
- LIANG, H. – HILU, K.W. 1996. Application of the *matK* gene sequences to grass systematics. In *Canadian Journal of Botany*, vol. 74, pp. 125–134. url: <http://www.nrcresearchpress.com/doi/pdf/10.1139/b96-017>
- LIAO, P.Y. – LEE, K.H. 2010. From SNPs to functional polymorphism: The insight into biotechnology applications. In *Biochemical Engineering Journal*, vol. 49, pp. 149–158. <https://doi.org/10.1016/j.bej.2009.12.021>
- MOHR, G. – PERLMAN, P. – LAMBOWITZ, A. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. In *Nucleic Acids Research*, vol. 21, pp. 4991–4997. url: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC310608/>
- MULLER, K.F. – BORSCH, T. – HILU, K.W. 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F* and *rbcL* in basal angiosperms. In *Molecular Phylogenetics and Evolution*, vol. 41, pp. 99–117. DOI: 10.1016/j.ympev.2006.06.017
- NG, P.C. – HENICKOFF, S. 2006. Predicting the effects of amino acid substitutions on protein function. In *Annual Review of Genomics and Human Genetics*, vol. 7, pp. 61–80. DOI: 10.1146/annurev.genom.7.080505.115630
- OFRAN, Y. – ROST, B. 2007. ISIS: interaction sites identified from sequence. In *Bioinformatics*, vol. 23, pp. e13–e16. DOI: 10.1093/bioinformatics/btl303
- PATEL, S.M. – KORINGA, P.G. – REDDY, B.B. – NATHANI, N.M. – JOSHI, C.G. 2015. In silico analysis of consequences of non-synonymous SNPs of *Slc11a2* gene in Indian bovines. In *Genomics Data*, vol. 5, pp. 72–79. DOI: 10.1016/j.gdata.2015.05.015
- PALINE, C.N. – STEVEN, H. 2003. SIFT: predicting amino acid changes that affect protein function. In *Nucleic Acids Research*, vol. 31, pp. 3812–73814. DOI: 10.1093/nar/gkg509
- RAFALSKI, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. In *Current Opinion in Plant Biology*, vol. 5, pp. 94–100. DOI: 10.1016/S1369-5266(02)00240-6
- ROMERO, P.R. – ZAIDI, S. – FANG, Y.Y. – UVERSKY, V.N. – RADIVOJAC, P. – OLDFIELD, C.J. – CORTESE, M.S. – LEGALL, T. – OBRADOVIC, Z. – DUNKER, A.K. 2006. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. In *Proceedings of the National Academy of Sciences USA*, vol. 103, pp. 8390–8395. DOI: 10.1073/pnas.0507916103
- SALDANHA, R. – CHEN, B. – WANK, H. – MATSUURA, M. – EDWARDS, J. – LAMBOWITZ, A.M. 1999. RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. In *Biochemistry*, vol. 38, pp. 9069–9083. DOI: 10.1021/bi9827991
- SAUNA, Z.E. – KIMCHI-SARFATY, C. 2011. Understanding the contribution of synonymous mutations to human disease. In *Nature Review Genetics*, vol. 12, pp. 638–691. DOI: 10.1038/nrg3051
- SCHLESSINGER, A. – PUNT, A.M. – YACHDAV, G. – KAJAN, L. – ROST, B. 2009. Improved disorder prediction by combination of orthogonal approaches. In *PLoS One*, vol. 4, no. 2, pp. e4433. DOI: <http://dx.doi.org/10.1371/journal.pone.0004433>
- SCHNEIDER, K. – WEISSHAAR, B. – BORCHARDT, D.C. – SALAMINI, F. 2001. SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. In *Molecular Breeding*, vol. 8, pp. 63–74. DOI: 10.1023/A:1011902916194
- SCHOLTZ, S. – LORZ, H. – LUTTICKE, S. 2001. Transposition of maize transposable element *Ac* in barley (*Hordeum vulgare* L.). In *Molecular and General Genetics*, vol. 264, pp. 653–661. DOI: 10.1007/s004380000351
- SMIGIELSKI, E.M. – SIROTKIN, K. – WARD, M. – SHERRY, S.T. 2000. dbSNP: a database of single nucleotide polymorphisms. In *Nucleic Acids Research*, vol. 28, no. 1, pp. 52–355. url: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102496/pdf/gkd114.pdf>
- VON BOTHMER, R. – SATO, K. – KOMATSUDAM, T. – YASUDA, S. – FISCHBECK, G. 2003. The domestication of cultivated barley. In von BOTHMER, R. – VAN HINTUM, T. – KNÜPFER, H. – SATO, K. (Eds.) *Diversity in barley (Hordeum vulgare)*. Amsterdam: Elsevier Science BV, pp. 9–27.
- WARD, J.J. – SODHI, J.S. – MCGUFFIN, L.J. – BUXTON, B.F. – JONES, D.T. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. In *Journal of Molecular Biology*, vol. 337, pp. 635–645. DOI: 10.1016/j.jmb.2004.02.002
- YU, J. – XUE, J.H. – ZHOU, S.L. 2011. New universal *matK* primers for DNA barcoding angiosperms. In *Journal of Systematic Evolution*, vol. 49, no. 3, pp. 176. DOI: 10.1111/j.1759-6831.2011.00134.x
- ZOSCHKE, R. – NAKAMURA, M. – LIERE, K. – SUGIURA, M. – BORNER, T. – SCHMITZ-LINNEWEBER, C. 2010. An organellar maturase associates with multiple group II introns. In *Proceedings of the National Academy of Sciences USA*, vol. 107, no. 7, pp. 3245. DOI: 10.1073/pnas.0909400107

Received: October 4, 2017