MADS-box transcription factors in potato (*Solanum tuberosum* L.): molecular cloning, sequence analysis, comparison, prediction of the crucial motifs and three dimensional (3-D) structures

A

DISSERTATION

SUBMITTED

FOR

PARTIAL FULFILLMENT FOR THE AWARD OF DEGREE OF

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

BY

Ravneet Kaur

(Regd. No. 302001012)

Under the Supervision of

Dr. N. Das

Professor

DEPARTMENT OF BIOTECHNOLOGY

Thapar Institute of Engineering & Technology, Patiala

June 2022

CANDIDATE'S DECLARATION

I, hereby declare that the work which is being presented in the dissertation entitled, **MADS-box transcription factors in potato (*Solanum tuberosum* L.): molecular cloning, sequence analysis, comparison, prediction of the crucial motifs and three dimensional (3-D) structures** in the partial fulfillment of the requirement for the award of degree of Master of Science in Biotechnology, Thapar Institute of Engineering & Technology, Patiala, is an original record of my own research work carried out under the guidance and supervision of **Dr. N. Das**, Professor, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India. The content in the dissertation has not been submitted to any other university or institute for award of any other degree.

Date: 29 June, 2022

Place: TIET, Patiala

Ravneet Kaur.

(Ravneet Kaur)

CERTIFICATE

TO WHOM IT MAY CONCERN

This is to certify that the dissertation entitled **MADS-box transcription factors in potato (*Solanum tuberosum* L.): molecular cloning, sequence analysis, comparison, prediction of the crucial motifs and three dimensional (3-D) structures** comprises research work carried out by **Ms. Ravneet Kaur** (Regd. No. 302001012) under my supervision and guidance during the period between January 2022 to June 2022 for the partial fulfillment of the requirement for the award of the degree of Master of Science in Biotechnology, submitted to Thapar Institute of Engineering & Technology, Patiala. The report has not been submitted for the award of any other degree or certificate in this or any other university or institute.

27/06/2022

(Dr. N. Das)
Professor
Department of Biotechnology
Thapar Institute of Engineering & Technology
Patiala

Ravneet Kaur.

(Ravneet Kaur)

ACKNOWLEDGEMENT

I am grateful to each and every one who has helped me throughout the entire project for its successful completion. First of all thanks to Almighty for giving me strength and support so that the project could be completed peacefully.

I find myself privileged to acknowledge my guide **Dr. N. Das**, Professor, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala, Punjab for his guidance, kindness, motivation and splendid supervision during my work. I express my heartfelt thanks for his patient support and excellent advice. It was his constant encouragement, constructive criticism and ability to handle the obstacles has helped me to gain a lot from him during this period.

I express my sincere thanks to **Dr. M.S. Reddy, HOD, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala** for giving me this opportunity to do this project.

A special thanks to all the faculty members for their support and invaluable suggestions throughout the project. I owe my gratitude to **Ms. Yadveer Kaur** and **Mrs. Gurpreet Kaur** for their continuous efforts and support without whom it would have been difficult to complete the project. I also sincerely thank **Ms. Parul Sharma** and **Mr. Davinder Singh** for their help. I would like to acknowledge my lab friend **Kriti Sharma** who helped me a lot to complete the project. I am thankful to the help rendered by the non-teaching staff and the lab attendants.

I would like to thank **Mr. Ram Newal Yadav, Mr. Phool Chand, Mr. Lallan Yadav, Mr. Prabhat Bailey, Mr. Mohinder** and **Mr. Surinderpal Kumar.**

With my heart, I specially thank my parents, **Mr. Avtar Singh** and **Mrs. Davinder Kaur** for their affection and faith. Also I would like to thank my sibling **Akshdeep Singh** for his motivation.

I would like to acknowledge my close friends **Kanika Kalra** and **Mehakpreet Kaur** who has stood by my side during the tough times. The whole credit goes to all the people who had their unshakeable faith in me which has always motivated me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Abbreviations	Name
µg	Microgram
µL	Microliter
ABA	Abscisic acid
AG	AGAMOUS
AtCO	*Arabidopsis thaliana*
Bp	Basepair
$CaCl_2$	Calcium chloride
cDNA	Complementary DNA
DEPC	Diethyl pyrocarbonate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
EDTA	Ethylenediaminetetraacetic acid
GA	Gibberellic acid
HCL	Hydrochloric acid
Kb	Kilo basepair
kDa	Kilo Dalton
KNOX	Knotted1-like homeobox
miR172	Micro RNA 172
mM	Millimolar
NaCl	Sodium chloride
NCBI	National Center for Biotechnology Information
OD	Optical density
PCR	Polymerase chain reaction
Ph	Potential of Hydrogen
POTM1	Potato MADS-box 1
POTH1	Potato Homobox 1
RNA	Ribonucleic acid
Rpm	Rotation per minute
StBEL5	*Solanum tuberosum*
StPHYB	*Solanum tuberosum* phytochrome B
StSUT4	*Solanum tuberosum* sucrose transporter 4

TALE	Three amino acid loop extension
TBE	Tris-borate-EDTA
TE	Tris-EDTA buffer
TF	Transcription factors
TPS	True Potato Seeds
IPTG	Isopropylthio-β-galactoside
CK	Cytokinin
RNA Pol II	RNA Polymerase II
CCT	CONSTANS, CONSTANS-like, TOC 1 region
St-SP6A	*Solanum tuberosum* Self-Pruning 6A
St-SP5G	*Solanum tuberosum*
St-BEL5	*Solanum tuberosum*
AG	Agamous
AP-1	Activator protein-1
RIN	Ripening Inhibitor
PFG	Petunia Flowering Gene
KC-1	Kufri Chipsona-1
UTR	Untranslated region
CDS	Coding sequence
GC	Guanine – cytosine
UV	Ultraviolet

ABSTRACT

In terms of production and consumption rates, potato is the most significant vegetable crop in the world. It is a major contributor to the global food security. It is a significant source of nourishment and income for people of various communities. Agriculture is facing more and more difficulties due to global climate change. Potato plants are extremely susceptible to high temperatures. Tuberization is a key survival mechanism for potato plants. Understanding factors that regulate the process of tuber induction is critical for developing tuber yield and quality improvement methods. MADS-box genes encode transcription factors that play an important role in plant growth and development, particularly in floral organ determination. In eukaryotes, the MADS-box family of transcription factors has been extensively studied for their critical involvement in transcriptional control. A cDNA of potato MADS-box gene (POTM1-1) from superior cultivar has already been identified and described from an early tuber cDNA library, but very little research has been done using Indian potato cultivars. So, in this study Indian potato cultivar was used as a source to isolate and characterize MADS-box gene. RNA was isolated from various tissues of potato plant and RT-PCR was performed using the total RNA. Further molecular cloning was performed using total RNA from tuber only. The cDNA clones obtained were further characterized by restriction digestion and PCR. Then, sequencing of these cDNA was done. The sequence obtained, designated as *POTM1-1A*, was submitted to GenBank (Accession No. ON804239). This sequence was compared with other MADS-box genes of various plant species on both nucleotide and protein level. Sequence analysis and motif prediction was done using various bioinformatics tools. The 3-D protein structures of our predicted protein and some homologous sequences from other plant species were also predicted in order to understand the structural and functional relationship.

Keywords: Potato (*Solanum tuberosum* L.), Tuberization, MADS-box transcription factors, POTM-1, cDNA clones

1.1 Potato (*Solanum tuberosum* L.)

Potato is one of the most significant crops, ranking third after wheat and rice in terms of human consumption (Dutt et al. 2017). About 21% of the nation's entire vegetable production and 26% of its total vegetable area are accounted by the potato alone (Singh et al. 2020). It belongs to the *Solanaceae* family and genus *Solanum*. It is heterozygous, tetraploid, and has twelve chromosomes in its genome (Abassi et al. 2018). Around 5000 varieties of potato along with 200 wild species and sub species have been identified all around the world. Out of the 1500 species of the *Solanum* L. genus, 150 are known to produce tubers (Abelenda et al. 2011). The nightshade family *Solanaceae* is the family of flowering plants having 102 genera and almost 2500 species. These food and drug plants are considered economically very important. Some other members of *Solanaceae* family are tobacco (*Nicotiana tabacum* and *Nicotiana rustica*), belladonna *(Atropa belladonna*), and the poisonous jimsonweed (*Datura stramonium*). Potato is a short duration crop cultivated in temperate, tropical and subtropical climates (Reddy et al. 2018). It provides more dry matter, edible energy and protein in short period of time than other cereals. The carbohydrate content is high in potato whereas fat content is relatively low. Basically, it fulfills all the major nutritional requirements. It also has the potential for high productivity and can be easily marketed. It can be vegetatively propagated through true potato seeds (TPS) (Sarkar 2008). Plants grown using tubers are the clones of parent plant, however, plants grown using seeds produce different varieties. Potato plants that are grown in the fields of TIET are shown in **Fig 1.1.**

Fig. 1.1 *Solanum tuberosum* plants grown in the fields of TIET, Patiala,Punjab, India

1.2 History of Potato

About 10000 years ago, potato was domesticated in the Andes Mountain of Southern America and it originated from the wild species of the *Solanum breviculae* complex (Abelenda et al. 2011). Potato travelled from South America across the continent. It was brought to Europe by Spanish conquerors in between 1570 and 1593 (Reddy et al. 2018). Now it has become the staple food all over the world as it makes up the dominant part of the diet. Potatoes were grown in Japan, in parts of Africa and West Indies in around late seventeenth century. A French explorer introduced potato into New Zealand in 1769 (Hawkes 1992). In the late seventeenth century, Portuguese and Dutch traders introduced potato in India but it became prominent after British missionaries came to India. The British East India Company introduced potato to hills in Northern India and Sri Lanka. They thought that potato would be a strong rival of rice that was grown abundantly in India but it was readily accepted in India (Singh and Rana 2014). With an output of 51.3 million tonnes from an area of roughly 2 million ha in 2017–18, India is the second-largest potato producing country in the world.

1.3 Morphology and Anatomy of potato

Potato is a perennial and herbaceous plant that can grow up to 24 inches in height. It is mainly grown for the edible tubers. Potato plant consists of a branched stem and leaves that are alternately arranged. Leaves are usually of unequal size. They may be oval or oblong shaped. The potato leaf is composed of two to four pairs of primary leaflets that are arranged on the mid-rib with a terminal leaflet present on the end. One major leaf is present on one node in a potato plant (Struik 2007). The flowers of the plant can be of various colors like white, pink, red, blue or purple. Potato tubers are modified underground stems with nodes and internodes. Tuber initiation starts in the subapical region of the stolon. When the swelling takes place, the meristematic activity in the apex of stolon is ceased (Ewing 1995). Potato tubers vary in shape and color. Tubers can be compressed, round, oval and long shaped. Skin color of potato varies from yellow to pink or purple. Several main stems are derived from the seed tuber as a viable shoot is produced by the various buds. **Fig 1.2** shows the various organs of the potato plant.

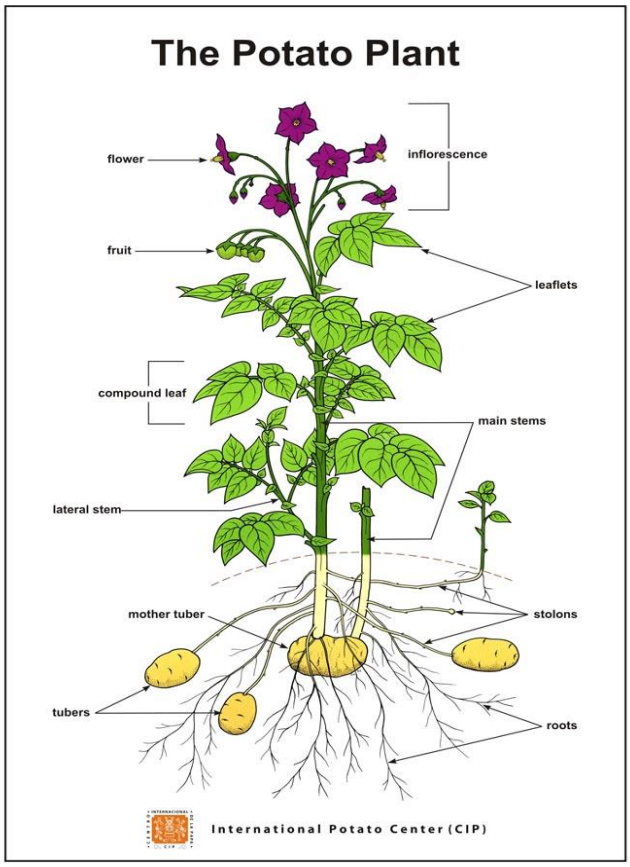

Fig. 1.2 Different organs of the potato plant (https://cipotato.org/potato/how-potato-grows/)

1.4 Nutritional aspects of Potato

Potatoes are known as starchy vegetables with carbohydrate being the main nutrient. According to the report of United States Department of Agriculture, raw potato has 79% water, 17% carbohydrates, 2% protein and small amount of fats. Potatoes are high in vitamins C and B6, as well as potassium, magnesium, and iron, among other nutrients. Riboflavin, thiamin, and folate are all vitamins found in potatoes (Beals 2019). When compared to cereals, potato tubers are high in the amino acids methionine and cysteine (Abelenda et al. 2011). Phytonutrients are also present in potatoes like carotenoids and phenolic acids (Brown et al. 2005; Liu 2013; McGill et al. 2013). The nutrients and nutritional components present in potatoes play an important role in health promotion by reducing the risk of chronic diseases (Beals 2019). Consumption of potatoes is a good source of energy and other nutrients, but occasionally potato can also cause undesirable health conditions like indigestion or other allergic reactions (Zaheer and Akhtar 2016). Along with the nutrients there are also some toxic compounds known as glycoalkaloids present in

leaves, flowers, sprouts and fruits of potato plant. They help in the protection of potatoes from the predators (Woolfe 1987; Friedman 2006). Additionally, potatoes contain dietary fibre, which promotes health, and dietary antioxidants, which help prevent diseases associated with ageing (Singh et al. 2020).

1.5 Tuberization in potatoes

Tuberization is the developmental process in which the stolon, also known as underground stem develops into potato (Viola et al. 2001).The tubers bear buds that are arranged spirally and sprout to form clones of the parent potato plant. Carbon partitioning, starch metabolism, growth correlation, and signal transduction are all significant biological processes involved in tuberization (Gregory 1965). Underground lateral shoots (stolons) with extended internodes, hooked apical ends, and diageotropic development form during the growth of the main potato stem (Cutter 1978). Tuber formation is initiated when the extensional growth of stolon is shifted towards radial growth (Dutt et al. 2017). During tuber initiation, many modifications have been seen in stolon tips as shown in **Fig. 1.3.**

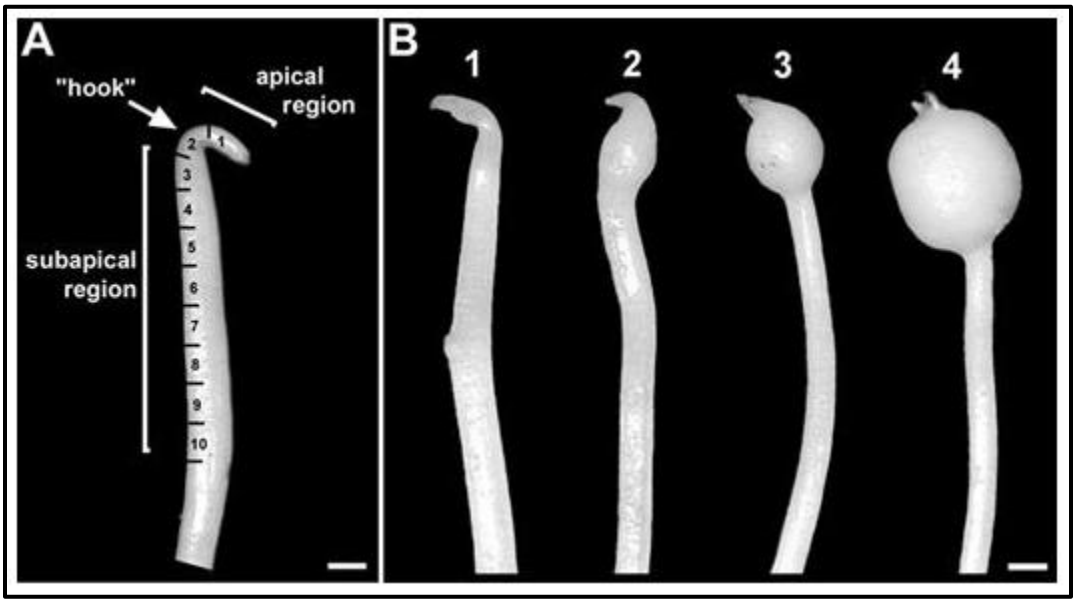

Fig. 1.3 A. Prominent stolon hook, **B.** Different stages of tuberization : 1. Hooked stolon tip with no swelling, 2. Slight swelling below apex that results in straightening of apical hook, 3. Stolon tip swells until the hook is completely open and develops into tuber, 4. Apical bud is incorporated into growing tuber (Viola et al. 2001).

When tuber initiation occurs, stolon growth stops and mitotic activity in the apical meristem ceases as well (Leshem and Clowes 1972; Xu et al. 1998; Vreugdenhil et al. 1999). As developing tubers become the largest sink present, the change of stolons into tubers has a

significant impact on the physiology of the entire plant (Oparka and Davies 1985). Tubers can also form on various sections of the plant above ground, usually from axillary nodes on the stem and flowers under some circumstances (Ewing and Struik 1992). These airborne tubers are normally developed on wounded or sick plants usually when nutrient transfer below ground is blocked (Jackson 1999). Potato development involves both morphological and biochemical processes. Morphological process involves the development of stolon and induction of tuber at the tip of stolon. On the other hand, biochemical process comprises synthesis of starch and accumulation of storage proteins (Taylor et al. 1992). Both cell division and enlargement characterize tuber induction at cellular level (Vreugdenhil et al. 1999).

- **Factors affecting tuberization**

Many factors influence tuber production, including microorganisms in the root zone, but nitrogen levels, temperature, and light have the largest impact (Jackson 1999). **Fig 1.4** shows various factors that regulate tuberization. Temperature has been observed to affect all stages and phases of tuberization including tuber induction, tuber set, tuber bulking, tuber number, size, and yield (Dutt et al. 2017). As the temperature increases, the rate of tuberization decreases and it diminishes at high temperatures (above 17°C). Low temperature (below 0°C) can also cause damage to the potatoes (Reynolds and Ewing 1989). Short days and cool nights are favorable for the tuberization (Chapman 1958). Photoperiod regulates a variety of developmental processes in diverse species, including flower induction and tuberization in potatoes. Photoperiodic regulation of both tuberization and flowering in plants has various common elements. Under favorable conditions the vascular bundles in leaves produce systematic signal that is carried by the phloem to the vegetative shoot apex to induce flowering and to the tips of underground stolon for induction of tuberization. The photoreceptor PHYB was shown to inhibit tuberization under long day conditions (Dutt et al. 2017). When it comes to photoperiodic responses, the length of the dark period is crucial (Jackson 1999). It was demonstrated that tuberization could be manipulated if the nitrogen supply to plant was varied and high levels of nitrogen inhibited this process (Krauss 1985). Along with these environmental factors, plant growth regulators like GA, CK, auxins and ABA also influence tuberization (Xu et al. 1988). The potential role of GAs in tuberization has been extensively investigated, primarily in trials using this chemical exogenously. It was noticed that GA treatment improves stolon extension while inhibiting tuber development (Smith and Rappaport 1969; Kumar and Wareing 1972). On the other hand, under

tuber inducing conditions, CK has been found to favour tuberization (Dutt et al. 2017). ABA is generally believed to be a plant development regulator that inhibits GA-promoted activities and promote tuberization (Hammes and Nel 1965; Marschner et al. 1984). Auxins are also known to play a promoting role in tuberization (Roumeliotis et al. 2013).

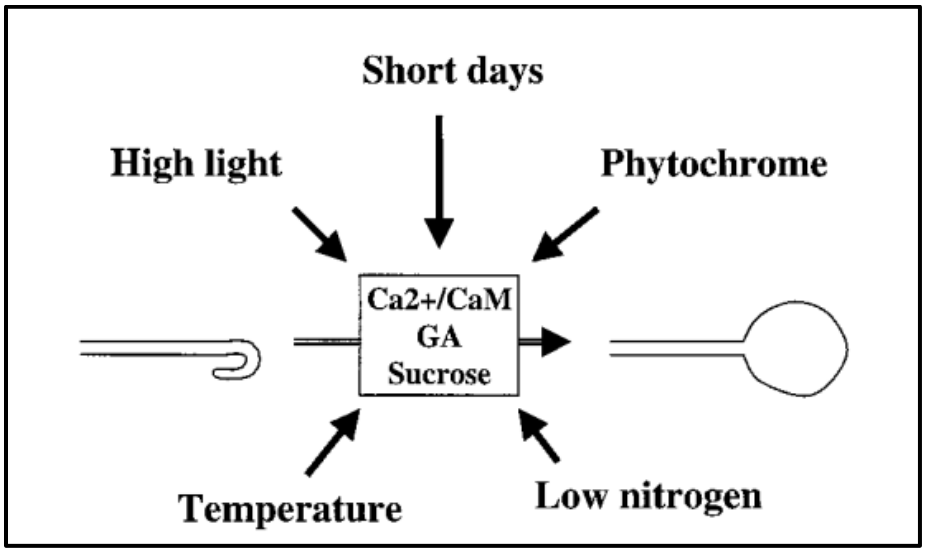

Fig. 1.4 Various factors affecting the process of tuberization in *Solanum tuberosum*(Jackson 1999)

1.6 Transcription factors

Development of plants and their reaction to the changes in environment are controlled by various transcription factors, regulatory RNAs and enzymes (Muñoz et al. 2019). Transcription factors are the proteins that bind to DNA, and then interact with RNA Pol II/other proteins in order to modulate the efficiency of transcription. TFs must be able to bind to DNA and then regulate transcription positively or negatively in order to have an effect. They can attach to a wide range of DNA sequences within target genes due to a diversity of DNA binding domains (Latchman 1997). The formation of potatoes is controlled by various TFs including POTM1, POTH1, StBEL5 and a potato orthologue of AtCO (Sarkar 2008). AtCO is a transcriptional regulator with a B-box domain at the N terminus and a CCT region at the C terminus (Putterill et al. 1995; Griffith et al. 2003). It is believed to inhibit tuberization in potatoes (Inui et al. 2010). Some other key proteins involved are StSP6A, Sucrose transporter StSUT4, StSP5G, StPHYB. Four molecules that act as the mobile signals for tuberization are StSP6A protein, StBEL5 RNA, miR172 and GAs. In potato phloem cells, the RNA of multiple BEL1-like transcription factors has been discovered. According to RNA detection methods and grafting tests, StBEL5 transcripts

migrate across a graft union to localize in stolon tips, the site of tuber induction (Hannapal 2010). The StBEL5-like transcription factor and its Knox protein partner, POTH1, regulate tuber development in potatoes by modulating hormone levels in the stolon tip (Rosin et al. 2003a; Chen et al. 2004). *KNOX* genes belong to the TALE superclass transcription factors (Btirglin 1997). They have been linked to the regulation of gibberellin levels in several studies. By binding a particular promoter sequence, a cooperative interaction between the TALE TFs StBEL5 and POTH1 directly represses *ga20ox1* (a gene that encodes a crucial enzyme in GA biosynthesis pathway) promoter activity (Chen et al. 2004). Activity of specific target genes is enhanced or repressed by the DNA-binding proteins which in turn mediate the process of tuberization in potato (Hannapal et al. 2004). Ca^{2+} and calmodulin also known as Ca-binding modulator proteins act as signaling molecules for induction of tubers (Jena et al. 1989).

1.7 MADS-box transcription factors

MADS domain transcription factors are encoded by MADS-box genes that are found in almost all eukaryotes including yeasts, plants, insects and mammals (Shore et al. 1995). The MADS in the MADS-box gene refers to the four founding members that are MINICHROMOSOME MAINTENANCE1 (MCM1, *Saccharomyces cerevisiae*; Passmore et al. 1988), AGAMOUS (*Arabidopsis* thaliana; Yanofsky et al. 1990), DEFICIENS (*Antirrhinum majus*; Sommer et al. 1990), and SERUM RESPONSE FACTOR (SRF, *Homo sapiens*; Norman et al., 1988). MCM1 plays an important role in transcription of cell-type specific genes and mating in *Saccharomyces cerevisiae* by controlling the pheromone response. SRF coordinate the proto-oncogene c-*fos* transcription. AG and DEFA are involved in the development of flower (Shore et al. 1995). MADS-box gene contains a conserved DNA binding region known as MADS-box and a conserved protein interaction domain known as K-box (Schwarz-Sommer et al. 1990; Ma et al. 1991). These genes are highly conserved transcription factors that play an important role in development of plants (Dutt et al. 2017). They are crucial for determination of floral organ identity and regulating vegetative development. The MADS-domain proteins play different roles in different organisms. They are responsible for transmission of neural signals, development of muscles, and occurrence of tumor in humans (Cao et al. 2016). In plants, they play an important role during osmotic stress response and survival of cell (Nadal et al.2003). On the basis of their sequence, expression patterns and function they are divided into two classes namely type I and type II (Alvarez-Buylla et al. 2000; Becker and Theißen, 2003). The MADS-box sequence

consists of 56 amino acids, out of which 9 amino acids are identical in all the family members. Within the MADS-box two functional regions have been identified, the N-terminal region; which determines the DNA binding specificity and the C-terminal region; that is mainly responsible for dimerization (Shore et al. 1995).

1.8 *POTM1*

POTM1 (Potato MADS-box 1) belongs to SQUAMOSA like family of MADS-box genes and is isolated from early stage tuber cDNA library. It was isolated by Kang and Hannapel in 1996. *POTM1* is expressed during early phase of tuberization. The amino acid sequence of *POTM1* encodes 250 amino acids which consists of a MADS box domain and a K box domain (Kang et al. 1996). POTM1 amino acid sequence matches with that of a MADS-box gene in petunia known as *PFG* which helps in transition from vegetative to reproductive development (Hart et al. 2002). This gene is also known to be a homolog of *AP-1* gene family as its conserved domains share high homologies to *TM4* of tomato and *AP1* of *Arabidopsis* (Kang and Hannapal 1996). It has been found out that the levels of *POTM1* were high in the newly synthesized tubers as compared to the mature tubers and the axillary bud development is controlled by *POTM1* as it regulates the growth of vegetative meristems. Vegetative meristems of potato contains RNA of *POTM1* in abundance, mostly in tunica and corpus layers of meristem, the procambium, the lamina of new leaves and axillary meristems that are newly formed (Rosin et al. 2003b). *POTM1* is expressed highly in underground stolons and axillary buds. The gene expression patterns of *POTM1* gene in both early and late flower development as well as tuber development have also been studied. It was deduced that the distribution levels of *POTM1* in both these were very different. CK levels in potato meristems and branching of axillary shoots is regulated by *POTM1* (Hannapal et al. 2004).

1.9 Other members of *Solanaceae* family

- **Tobacco (*Nicotiana tabacum*)**

Nicotiana tabacum is primarily an American native, although it is now commercially grown all over the world (Kishore 2014). It belongs to *Solanaceae* family and genus *Nicotiana*. *Nicotiana tabacum* L. is an herbaceous perennial plant. It is the most widely grown of all *Nicotiana* genus plants, and its leaves are commercially farmed in several countries for use in tobacco production (Agyare et al. 2013). It is grown all over the world, from tropical and subtropical areas to

temperate latitudes (Hanelt et al. 2001; Randall 2012. Tobacco is an allopolyploid (n = 24) species that shares its fundamental chromosome number of n = 12 with many other solanaceous plants like tomato, potato, pepper, and eggplant (Bindler et al. 2011).

- **Tomato (*Solanum lycopersicum*)**

The cultivated tomato, *Solanum lycopersicum* L., is a member of the *Solanaceae* family, which includes over 3000 species that thrive in a variety of environments (Knapp 2002). The vivid yellow blooms of tomatoes and their pinnate; non-spiny leaves set them apart from other *Solanum* species (Knapp and Peralta 2016). In the early 16th century, the Spanish introduced the tomato to Europe (Harvey et al. 2002). The Andean region, which includes sections of Chile, Colombia, Ecuador, Bolivia, and Peru, is home to all related wild tomato species (Sims 1980). As tomatoes are easy to cultivate, have a short life cycle, and are easy to alter, they are frequently used as a 'model crop' for various physiological, cellular, biochemical, molecular, and genetic studies (Kinet 1997). They are deployed as model plants in studies of seed development, germination, dormancy physiology and biochemistry. It has a 950 Mb genome which is diploid with 12 chromosomal pairs (Barone et al. 2008).

- **Capsicum (*Capsicum annum*)**

Sweet pepper (*Capsicum annum*) is a non-pungent kind of pepper that is a major cash crop in India and the world's second most significant solanaceous vegetable after tomato (Vengaiah and Pandey 2007). It belongs to family *Solanaceae* and is largely used as a spice and vegetable in a variety of cuisines around the world (Singh et al. 2021). It is high in beta carotene, capsaicin, and vitamins A and C. All of these nutrients help to avoid ailments like liver disease and impotency. This spicy vegetable is utilized in cooking not only for its fiery flavor, but also with its medicinal benefits. The chromosome number for Capsicum annum is 12 like other members of *Solanaceae* family.

2.1 MADS-box genes in plants

MADS-box genes code for a group of transcription factors that regulate everything from floral development to root growth in flowering plants (Aswath et al. 2005). The fact that MADS-box genes have been identified from yeast, mammals and plants, exemplify their diversity (Alvarez-Buylla et al. 2000; Henschel et al. 2002). They has also been discovered to play a role in stress response (Chen et al. 2019). The role of MADS-box genes in development of flowers and fruits has been rapidly discovered in the last decade. Evolutionary scientists can now analyze the homology of body parts and the evolution of body plans by looking at the expression pattern of MADS-box genes throughout the development of different plant species (Ng and Yanofsky 2001). Recently, considerable progress had been made in the area of MADS-box transcription factors with regard to their structure, function and roles in the process of transcription. Along with MADS-box genes, various other genes and transcription factors responsible for the process of tuberization have also been studied by various biologists. DEF from Antirrhinum (Schwarz-Sommer et al. 1990; Sommer et al. 1990) and AG from *Arabidopsis* (Yanofsky et al. 1990) were the first MADS-box genes that were isolated from plants. It was established that the MADS domain evolved from a region of topoisomerases IIA subunit A using whole genome sequencing information and distant homology detection methods (Groamzw et al. 2010).

On the basis of amino-acid sequences in the MADS-box domain, two types of MADS-box gene have been identified namely type I and type II (Masiero et al. 2011) as shown in **Fig. 2.1.** Type I proteins are more closely related to animal SRF like MADS-box proteins, whereas type II proteins are more closely related to MEF2-like MADS-box proteins (Alvarez-Buylla et al. 2000). In view of the M domain of the encoded protein, type I MADS-box transcription factors can be divided into four subclasses (Mα, Mβ, Mγ, and Mδ), but only a few have been defined for their biological function (Smaczniak et al. 2012). Type I genes are also known as M-type MADS-box genes. Their structure consists of N-terminal MADS domain and a C-terminal domain. Type II MADS-box genes are named as MIKC-type genes. In addition to the MADS domain, they contain three other domains that include intervening (I), kertain-like (K), and C-terminal (C) domains (Gao et al. 2018). The MADS-box domain (M), located at the N-terminal region of the protein is characteristic of all the MADS-box genes having 58-60 amino acid

residues. Its main functions include binding the DNA to CC(A/T)GG boxes in their target genes' regulatory regions, dimerization with other MADS box proteins, and nuclear localization (Zhu et al. 2005). The intervening (I) domain contains around 30 amino acids and aids in MADS-box protein dimerization (Masiero et al. 2002). The Kertain-like (K) domain is roughly 70 amino acids long and has a more conservative structure than the intervening (I) domain. The C-terminal (C) domain of MADS-box proteins is a highly variable region involved in transcriptional activation and protein complex formation (Krammer et al. 2006).

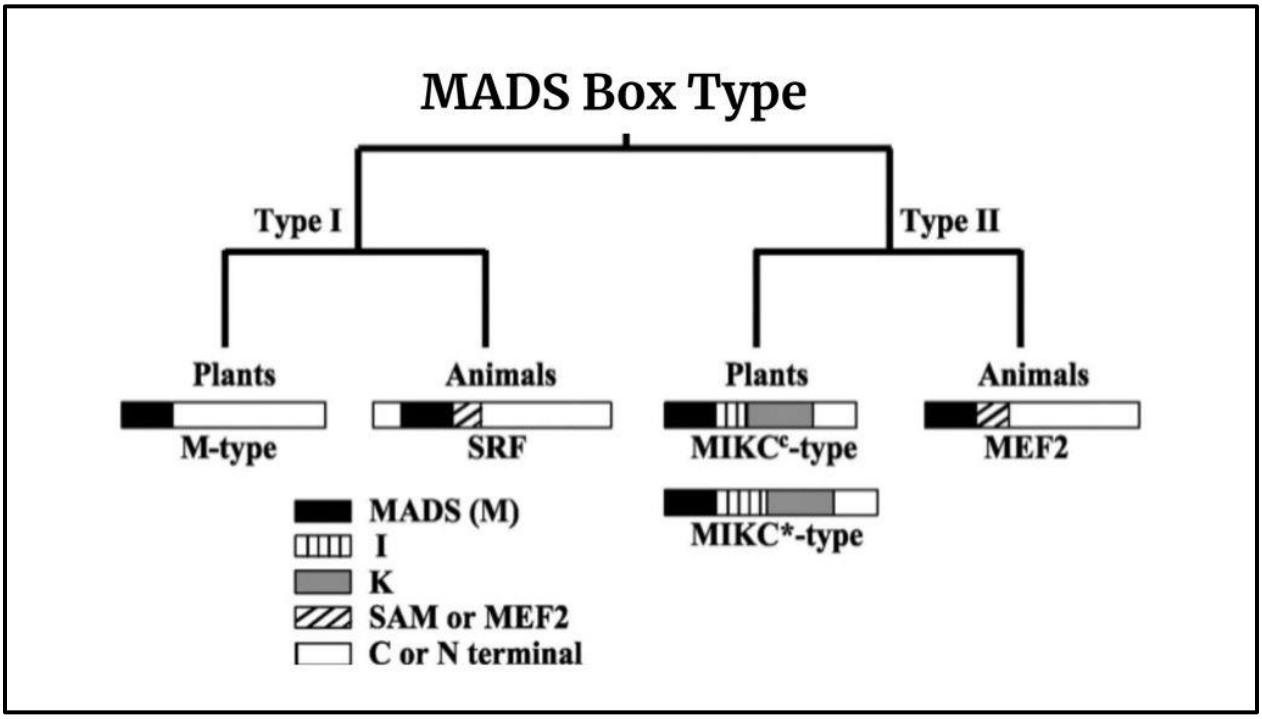

Fig. 2.1 Types of MADS-box genes present in plants and animals. (https://slideplayer. com/slide/14037956/)

In *Arabidopsis*, the first comprehensive analysis of the MADS gene family was conducted, yielding 107 MADS-box genes, divided into five subfamilies—25 Mα genes, 20 Mβ genes, 16 Mγ genes, six MIKC* genes, and 39 MIKC^c genes with *AGL33* being the only unassigned exception (Parenicová et al. 2003). Furthermore, according to phylogenetic research, the *Arabidopsis* type II family has 39 MIKC^c genes that may be classified into 13 groups (Díaz-Riquelme et al. 2009).

On the basis of sequence similarity, expression patterns and function, plant MADS proteins have been divided into several families. *Arabidopsis APETALA* (*AP1)* and *Antirrhinum* SQUA, responsible for identity of floral meristem and floral organ, belong to SQUA family. The AP3/PI and AG family genes that include *APETALA3* (*AP3*), *PISTTILATA* (*PI*) and *AGAMOUS* from

Arabidopsis; *GLOBOSA* (*GLO*), *DEFICIENS* (*DEF*) and *PLENA* (*PLE*) from *Antirrhinum*, regulate floral organ identity (Jang et al. 2002).

MADS-box genes in Potato

Based on their evolutionary ties to the *Arabidopsis* and rice MADS-box genes, a total of 153 MADS-box genes were found and classified into the MIKC subfamily (MIKCC and MIKC*) and the M-type subfamily (Mα, Mβ, and Mγ). The potato M-type subfamily comprised 114 members, almost three times the number of MIKC members (39), implying that M-type MADS-box genes have a higher rate of duplication and/or loss during potato genome evolution. All 12 potato chromosomes had MADS-box genes, with M-type members accounting for the majority of the clustering (Gao et al. 2018). In potato, three MADS-box genes have been previously reported in MIKC* subfamily, that are *POTM1*–1, *StMADS11*, and *StMADS16* (Kang et al. 2003; Guo et al. 2013). In both vegetative and floral organs, *POTM1*–1 gene expression is temporally and spatially controlled, and transcriptional repression of *POTM1*–1 promotes axillary meristem development by raising levels of cytokinin (Kan et al. 2002; Rosin et al. 2003b; Guo et al. 2013). *StMADS11* is found in all vegetative tissues of the potato plant, primarily in the stem, but not in the flower organs (Carmona et al. 1998). The inflorescence structure is altered by ectopic expression of *StMADS16*, which increases the internode length and flower proliferation of the inflorescence meristems and provides vegetative characteristics to the flower (García-Maroto et al. 2000).

MADS-box genes in Tomato

Till date, 131 MADS-box genes including 81 type I and 50 type II genes has been discovered in tomato. 15 tomato MADS-box genes were known to be involved in floral organ identification and 5 tomato MADS-box genes were involved in fruit development (Wang et al. 2019). The MADS-box gene in tomatoes known as *RIN* is a key component in fruit ripening as it controls a variety of ripening-related activities, including both ethylene-dependent and ethylene-independent ripening (Vrebalov et al. 2002).

MADS-box genes in Capsicum

By screening a cDNA library with the *OsMADS1* rice MADS-box gene as a probe, the cDNA clone *CanMADS1* was obtained from immature flower buds of the hot pepper (*Capsicum annuum* L.).*CanMADS1* transcript was found in flower buds and fruits, with the greatest transcription signal during the fruit set stage. *CanMADS6* is a MADS-box gene that was

extracted from immature flower buds utilizing the K domain and 15 amino acid residues of *CanMADS1*'s C-terminal region as bait. Based on the phylogenetic analysis, it was concluded that the *CanMADS1* gene belongs to the *AGL2* subfamily (Sung et al. 2001).

MADS-box genes in Tobacco

Due to the complicated and partial assembly of the tobacco genome, 111 NtMADS-box genes were found on 20 chromosomes, with 57 NtMADS-box genes found on unanchored scaffolds (Bai et al.2019). By using *OsMADS1*, a rice MADS-box gene as a probe, a tobacco MADS-box gene, *NtMADS4* was isolated from the cDNA library of flower bud (Jang et al. 2002). Yeast two-hybrid screening identified *NtMADS11* as an interaction partner of *NtMADS4*. While *NtMADS4* transcripts were only found in reproductive organs, *NtMADS11* transcripts were found in both reproductive and vegetative organs (Jang et al, 2002).

2.2 Expression pattern of *POTM1* in different tissues

In situ hybridization and RNA blotting analysis were done by Kang et al. (2003) to study *POTM1-1* gene expression patterns in development of flowers and tubers in potato. It was concluded that *POTM1–1* transcripts were plentiful in the developing reproductive organs of early flowers, including the placentae of carpels and the pollen sacs of stamens. However, distribution of *POTM1–1* during late floral development differed from that during early flower development. Sepals and petals of late flowers had a lot of *POTM1–1* transcripts in comparison to stamens and carpels (Dutt et al. 2017). *POTM1* is found to have the highest sequence match with *SCM1* from *Solanum commersonii* (a wild potato) and *PFG* from petunia (*Petunia hybrida*) having 97% and 91% similarity over the entire protein length, respectively (Hart and Hannapel 2002). Despite the significant sequence similarity and expression patterns, *POTM1* appears to play a different role in development than PFG and other SQUA-like family members (Rosin et al 2003b). The *POTM1* suppression phenotype is similar to mutants in *Arabidopsis* (van der Graaff et al. 2001), tobacco (Li et al. 1992), tomato (Groot et al. 1995), and potato (Gális et al. 1995; Machácková et al. 1997) in many ways as the *A. tumefaciens ipt* gene was introduced in these plants to overproduce CK. The expression of the *ipt* gene affects tuberization in potatoes, with high levels of CK limiting tuber development and moderate levels favoring tuber formation (Gális et al. 1995; Romanov et al. 2000).

2.3 Origin of problem

Potato tuber development and growth is a complicated process governed by various environmental signals and plant hormones. Although the relevant literature has been revised from time to time, a complete understanding of the signal transduction pathways leading to tuber induction remains unclear. Because of the economic importance of this growth process, environmental tuberization management has been a topic of concern for decades. While key components of day-length perception and the inhibitory consequences of a night interruption are well recognized, the molecular underpinnings of this developmental transition are still a topic of debate. Understanding and utilizing tuberization becomes even more important and necessary in light of changing global climate and expectations for the potato as a primary crop to meet the food and nutritional needs of a growing human population. As recently discussed, many biotechnological and genomics approaches may be employed to manipulate tuberization, regulate biomolecules, and pathways with the goal of enhancing potato productivity. TFs play an important role in regulation of tuberization, so it is important to understand their functions and expression in various tissues of potato plant. *POTM1* that belongs to MADS-box family of TFs, is known to be crucial for tuberization process. Although gene expression of *POTM1* gene from Superior cultivar has been studied but research on Indian potato cultivar has not been done. Due to significant environment and genetic interactions a gene can exist in more than one form. Keeping all these aspects in view, efforts have been made to explore *POTM1*, a MADS-box gene from Indian cultivar of potato, KC-1 and following objectives have been proposed.

2.4 Objectives of study

- Molecular cloning and characterization of *POTM1* cDNA from potato
- Sequence analysis and comparison between the MADS-box transcription factors of different plant species
- To identify the crucial motifs, and prediction of three dimensional (3-D) structures

3.1 Experimental Analysis

3.1.1 Plant materials and reagents

We used Kufri Chipsona-1 (KC-1), a high-yielding and economically important Indian potato cultivar obtained from the Central Potato Research Institute (CPRI) in Shimla, India. It is routinely micropropagated on MS-basal medium with 2.5 percent sucrose in our laboratory under controlled conditions (16 h light/8 h dark, light intensity approximately 40-42 μmol m^{-2} s^{-1} , 25-27°C, 70% relative humidity). The micropropagated potato plantlets were acclimatized before being farmed for 16 weeks in the field from mid-November to mid-March. At various phases of development, the tubers, leaves and various propagating organs were harvested. They were flash frozen in liquid nitrogen and stored at -80°C for further studies.

3.1.2 Designing of oligonucleotide primers

Oligonucleotides, also known as oligomers, are small DNA or RNA molecules that have a variety of uses in genetic testing, research, and forensics. Oligonucleotides are most commonly used as primers in PCR. The primers were constructed using the available *POTM1* gene sequence found in the GenBank database (GenBank ID: U23757). The length of the whole sequence of *POTM1* is 1068bp. The CDS region of *POTM1* starts at 58^{th} nucleotide and ends at 810^{th} nucleotide making the overall length 753bp. This CDS region encodes 251 codons which includes the stop codon. The length of 5′-UTR and 3′-UTR is 57bp and 257bp.

For effective PCR primers, the following properties are desirable:

- A complementary sequence should be 18–22 nucleotides long.
- The melting temperatures (Tm) of paired primers must be identical.
- The ideal composition of GC content is 40–60%.
- It's best to avoid long repeats of a single base.
- There should be few or no secondary structures like hairpins.
- There should be no self-dimers in a primer; a primer should not be homologous to itself.
- There should be no cross dimers, and primer pairs should not dimerize.

Keeping all of the above criteria in mind and using cDNA sequence of *POTM1* found in the database, the following primers were made.

Table 3.1 Details of designed primers of *POTM1*

Sr. No.	Name of Oligomers	Description of oligomers	Length
1.	PMF-0057	5′-AATGGGAAGAGGAAGAGTCC-3′	20 bp
2.	PMR1-0935	5′-ATTGTGGCATCAATATAGAG-3′	20 bp
3.	PMR2-0982	5′-ATCCATCTATATTGTGGCATG-3′	21 bp

3.1.3 Isolation of total RNA from tissues of potato cultivar KC-1

GeNei^{TM} Plant RNA Isolation kit was used to extract total RNA from the leaves and freshly harvested tubers of the potato cultivar KC-1. The Plant RNA isolation kit is a solution-based RNA isolation method. The first step includes tissue disruption using SDS which is followed by acid-phenol extraction and selective RNA precipitation with lithium chloride. The kit includes enough reagents to make 10 preparations from 1 g of plant tissue and it is stored at 4°C.

Materials required: Equipment: Dry bath, water bath, microcentrifuge, spectrophotometer, agarose gel apparatus

Reagent: Distilled ethanol, isopropanol, chloroform, nuclease free water

Other requirements: 1.5 ml vials, sterile mortar and pestle, liquid nitrogen, polypropylene tubes. Materials that are supplied in the kit are listed below:-

Table 3.2 Working solution for isolation of total RNA

Materials	Quantity 21090800011730 (10 expts.)
Cell Lysis Solution (Solution A)	35 ml
Precipitation Solution (Solution B)	20 ml
Wash Buffer (Solution C)	50 ml
Water Saturated Phenol	35 ml
Nuclease Free Water	3 ml

Before starting the experiment, Solution A was kept at room temperature to prevent precipitation. Fresh mixture of Solution A+ Phenol in the ratio 1:1 was prepared and heated to

80°C before use. For the better quality and quantity of RNA, extraction of RNA was done from actively growing tissues.

Procedure:

Using liquid nitrogen, 1 g of tissue was homogenized in a cooled pulverizer. Pre-warmed Solution A+ Phenol mixture was added in 1:4. Then equal volume of chloroform was added and the mixture was incubated at room temperature for 30 minutes. Centrifugation at room temperature was done for 15 minutes at 5000 rpm. We removed the upper phase and transferred it into a fresh tube. Then one-third volume of solution B was added and mixed well. The mixture was incubated at 4°C overnight. Centrifugation was done at 4°C for 20 minutes at 7000 rpm. Solution C was used to dissolve the precipitate obtained after centrifugation, resuspended completely and transferred to a fresh vial. Again centrifugation was done at 4°C for 20 minutes at 7000 rpm. The washing of pellet was done with 70% ethanol followed by 100% ethanol. The obtained pellet was air-dried and dissolved in 150-200 µL nuclease free water. RNA quality was checked through agarose gel electrophoresis and quantification was done using nanodrop.

3.1.4 Agarose gel electrophoresis

Materials required: Agarose (Hi media), 0.5 X TBE buffer, sterile water, Ethidium bromide dye (0.5 µg mL^{-1}), DNA samples and amplicons, bromophenol blue dye, gel electrophoresis instrument, UV transilluminator, gel documentation system (BIO-RAD)

Procedure: Standard methods (Sambrook- a laboratory manual) were used to perform Agarose gel electrophoresis. 1.0 % agarose gel was made in 0.5X TBE buffer followed by addition of Ethidium bromide dye (0.5 µg mL^{-1}). After that, the gel was poured into the casting tray. After the gel had solidified, the DNA/RNA samples were loaded into wells. Electrophoresis was performed at 2-5 volts per cm in 0.5X TBE (running buffer) until the tracking dye covered two-thirds of the gel length. Finally, the bands were examined under ultraviolet (UV) light in UV transilluminator or gel documentation system.

3.1.5 Purification of total RNA

Materials required: RNase-free DNase enzyme, Phenol: Chloroform, ethanol, RNase-free deionized water

Procedure: RNA purification is a technique for separating pure RNA from tissues and DNA or protein combinations. Treatment with RNase-free DNase, solvent extraction, and ethanol precipitation were used to purify the crude RNA. RNase-free deionized water was used to

dissolve the RNA, and aliquots were kept at -70°C for further use. The quality of the RNA samples was further checked by spectrophotometric analysis using the A_{260}/A_{280} ratio.

3.1.6 Reverse-Transcription Polymerase Chain Reaction (RT-PCR)

Materials required: DEPC water, Oligo dT primer, 10 mM dNTPs, RNase inhibitor, Reaction buffer, RNA sample.

Procedure: Reverse transcription (RT) was performed with M-MuLV reverse transcriptase using the RevertAid^{TM} H Minus First Strand cDNA Synthesis Kit (Fermentas Life Sciences). The gene-specific forward and reverse primer, PMF-0057 and PMR-0935 were used respectively to perform RT. In RNA-DNA hybrids, the enzyme lacks ribonuclease H activity unique to RNA. As a result, RNA is not degraded during first-strand cDNA synthesis, yielding better yields of full-length cDNA from lengthy templates up to 13 kb. Approx. 2.0 µL of total RNA from tuber or leaf of Indian potato cultivars was used as template for each RT reaction. Manufacturer's instructions were followed to carry out the process of reverse transcription. Using the individual RT result as a template, PCR was used to obtain full length cDNA. The RNA sample was added to 6 µL of DEPC water in an eppendorf. After that, the oligo dT primer were added and thoroughly mixed, and the eppendorf was maintained at 65°C for 5 minutes and then at room temperature for 5 minutes. Then 4 µL of reaction buffer, 1 µL RNase inhibitor and 10 mM dNTP were added. The tubes were incubated for 1 hour at 42 ºC.

The composition of reaction mixture for 30 cycles of PCR reaction is given in **Table 3.3**:

Table 3.3 Composition of reaction mixture (volume 50µL) of PCR reaction

PCR Component	50 µL reaction	Final Concentration
Template	Variable	<1,000 ng
10X Taq Reaction Buffer	5 µL	1 X
10 µ mole Forward Primer	2.5 µL	0.2 µM (0.05-1 µM)
10 µ mole Reverse primer	2.5 µL	0.2 µM (0.05-1 µM)
10 mM dNTPs	2.5 µL	200 µM
Taq polymerase	0.4 µL (1 unit)	1.2units/50µl PCR
Distilled water	To make up volume 50 µL	

Al lreagents followed by template were added in a tube and gentle mixing was done with the help of vortex. The Thermal cycling parameters of set as follows: initial denaturation at 94ºC for

1 min, denaturation at 94ºC for 1 min, annealing at 55ºC for 2 min and polymerization at 72ºC for 2 min for 30 cycles. Final extension was set at 72ºC for 5 min.

3.1.7 cDNA cloning

Klenow treatment

Klenow Fragment is the large fragment of DNA Polymerase I, which preserves its 5'→3' polymerase and 3'→5' exonuclease activity. The Klenow treatment is done to fill in 5' overhangs or remove 3' overhangs to make blunt ends.

Material required: PCR product, taq polymerase enzyme, buffer, dNTPs, millipore water

Procedure: Appropriate amount of all the components for Klenow treatment were taken in an eppendorf and mixed thoroughly. Volume was made upto 50 µL and the mixture was spinned down using microcentrifuge. Then it was incubated at 28°C for 1 hour.

PCR kit purification

Purification of the PCR product after Klenow treatment is done using the QIAquick® PCR Purification Kit. This kit can be stored for one year at room temperature. This kit can purify up to 10 µg PCR products, 100 bp to 10 kb in length. Purification of DNA from a PCR reaction is usually required for downstream application, as it allows enzymes, nucleotides, primers, and buffer components to be removed.

Materials required: PCR product, ethanol, buffer PE, buffer PB, buffer EB, agarose gel apparatus, microcentrifuge, QIAquick column, microcentrifuge tube, collection tube

Notes:

- Ethanol (96-100%) was added to buffer PE before use.
- All centrifugation steps were done at 13000 rpm at room temperature.
- To buffer PB, 1:250 volume of pH indicator was added.

Procedure: 5 volumes PB buffer was added to 1 volume of PCR reaction and mixed thoroughly. If orange or violet color appeared then 10 µL 3M Sodium Acetate was added at pH 5.0 and mixed. The mixture turns yellow. QIAquick column was placed in a 2 ml collection tube provided in the kit. The sample was applied to QIAquick column for binding DNA and centrifuged for 30-60 s. Flow through was discarded and column was placed in same tube. 0.75

ml Buffer PE was added to the column for washing and centrigugation was done for 30-60 s. Flow through was discarded and column was placed in same tube. Centrifugation was done again for 1 minute to remove the residual buffer and the column was placed in clean 1.5 ml microcentrifuge tube. 50 µL Buffer EB was added to the centre of QIAquick column for dilution of DNA and it was centrifuged for 1 minute. To analyze purified DNA 1 volume of loading dye was added to 5 volumes of purified DNA and the mixture was loaded in the gel.

Ligation

Ligation is the joining of two nucleic acid segments by the action of an enzyme. In molecular cloning experiments, DNA ligation is routinely employed to physically bind a DNA vector to the gene of interest. Ligation is performed by T4 DNA ligase enzyme.

Materials required: Millipore water, pUC19 Vector, T4 DNA Ligase Buffer, T4 DNA Ligase enzyme, PCR Product, Polyethylene glycol

Procedure: Appropriate amount of Millipore water and the components mentioned in the table were added to the sterile eppendorf tube. The mixture was incubated overnight at 25°C. T4 DNA Ligase buffer that contains PEG 4000 was used at 1X concentration. Overall volume of reaction mixture was made 15 µL.

Table 3.4 Main components of ligation reaction

Components	Quantity
Linearized vector	~ 0.3 µg
Insert	~ 0.5 µg
T4 DNA Ligase enzyme	4 units
Ligation buffer	0.5 µL
Ligase	1 µL

Transformation

Bacterial transformation is the acquisition of a gene by a recipient cell from free DNA molecules in the surrounding media. $CaCl_2$ treatment generates a transitory condition of "competence" in recipient bacteria such as E. coli cells, allowing them to take up various DNA molecules such as plasmid DNA, bacteriophage DNA, chromosomal DNA fragments, and so on.

Materials required: E. coli DH5α strain, pUC19, LB medium, 100 mM $CaCl_2$, microfuge tubes and micro tips, microcentrifuge

***Procedure*:** Freshly grown single colony of E. coli DH5α was transferred to 20 ml of LB broth in a 250ml flask. The culture was then incubated at 37°C for 16-20 hrs along with vigorous shaking. The 200 ul of overnight grown culture was inoculated into fresh 25mL of LB media. The media was incubated until OD_{590} reaches between 0.3-0.5. After this, the culture was transferred to sterile, disposable, ice-cold 50ml polypropylene tubes and stored on ice for 10 minutes. Cells were recovered by centrifugation at 4°C for 10 minutes at 5000 rpm. The supernatant was discarded and the tubes were allowed to stand in an inverted position for about 1 minute. The pellet was resuspended in 10ml of 0.1 M $CaCl_2$ and stored on ice for 10-15 minutes and third step was repeated to recover the cells. Cell pellet was again resuspended in 1 ml of 0.1 M $CaCl_2$ and stored on ice for 12-24 hrs. 100 µL of the suspension mixture of competent cells was transferred to fresh microfuge tube. Ligation product (~100 ng in a volume of 5 µL or less) was added and mixed before incubating on ice for 30 minutes. Tubes were incubated in a water bath at 42°C for 2 minutes and then rapidly transferred on ice for 2 minutes. 1 ml LB media was added and incubated at 37°C for 1 hour. After that, centrifugation was done for 5 minutes at 6000 rpm. Lastly, extra media was decanted and the pellet was dissolved.

Blue-white screening

Blue-white screening is a fast and effective method for detecting recombinant bacteria. It is based on the activity of an E. coli enzyme, β-galactosidase that breaks down lactose into glucose and galactose.

Materials required: LA ampicillin plates, X-gal, IPTG, millipore water

Procedure: Appropriate amount of X-gal, IPTG and millipore water were added in the eppendorf and mixed thoroughly. 80 µL of the mixture was spreaded on LA ampicillin plates.

Table 3.5 Components for blue-white screening

Components	**Quantity/plate**
X-gal (20mg/ml)	30 µL
IPTG (100mg/ml)	10 µL
Millipore water	40 µL

Then 100 µL of the transformed cells were spread on the plates with X-gal and IPTG. These plates were incubated overnight at 37°C and growth of blue-white colonies was observed next day. Recombinant white colonies were taken and patches were made. Again single white colonies were obtained and inoculated in the LB media containing ampicillin and incubated overnight at 37°C.

Plasmid DNA isolation

Material required: E. coli DH5α (pUC 19), LB , STET Buffer (Sucrose : 8.0% (w/v), Triton X-100:0.5% (w/v), EDTA:50 mM (pH= 8.0), Tris-HCl:10.0 mM (pH= 8.0), lysozyme 10 mg/ml, 3 MSodium acetate pH 3.0, isopropanol, TE buffer pH 8.0, microfuge tubes, microtips, boilingwater bath, agarose gel electrophoresis apparatus

Procedure: The white recombinant colonies were inoculated in 5 ml of LB media and incubated overnight at 37°C with vigorous shaking.1.5-2.0 ml of above saturated culture was poured in microfuge tube and centrifuged in a microcentrifuge at 8000 rpm for 5 minutes. The supernatant was discarded. The pellet was dried and resuspended in 750 µL of STET buffer using vortex. Then 30µL of freshly prepared 10 mg/ml lysozyme was added and the caps were sealed before keeping in water bath for 90 seconds. Bacterial lysate was centrifuged at 12000 rpm for 15 minutes and supernatant was poured in fresh microfuge tube. 40 µl of 3M sodium acetate and equal volume of Isopropanol were added to precipitate nucleic acids. Tubes were kept at 4°C for 20 minutes. Centrifugation at 12000 rpm for 15 minutes was done to recover precipitated nucleic acids and pellet was rinsed in 70% ethanol. Again centrifugation was done at 11000 rpm for 5 minutes. The pellet was air dried and dissolved in 50 µl TE buffer. The quality of the plasmid DNA was checked by Agarose gel electrophoresis.

Plasmid DNA Purification

For purification 370 µl of sterile water was added to 30 µl of crude plasmid. 2 µl of RNase was added and incubated at 37°C for 30-45 mins. Equal volume of Phenol and Chloroform was added and mixed for 5 minutes. Centrifugation at 8000 rpm for 10 mins was done and to the upper aqueous layer $1/10^{th}$ volume of 3M Sodium acetate was added. Then double volume of dehydrated alcohol was added and kept overnight at –20°C. Centrifugation was done at 12000 rpm for 15 minutes and the supernatant was discarded. After addition of 70% ethanol

centrifugation was done once again, the pellet was air dried before dissolving in 30 µl TE buffer. The quality of purified plasmid DNA was checked by agarose gel electrophoresis.

Clone derived PCR

The same steps and parameters were used for clone derived PCR as for RT-PCR. However, instead of using a template, recombinant plasmid DNA was used.

Restriction Digestion

Materials required: DNA samples-pUC19, Plasmid DNA, bacteriophage lambda DNA, DNA molecular wt. markers such as 500 bp ladder, 1.0 kb DNA ladder; Restriction enzymes-BamHI, EcoRI, SmaI, along with respective 10xassay buffer; agarose gel electrophoresis apparatus, microfuge tubes, micro tips, disposable gloves, UV torch, UV transilluminator, gel documentation System

Procedure: In a sterile microfuge tube, DNA solution was taken and sterile water was added to make up a volume of 17 µL. Then 2 µL of 10x restriction enzyme assay buffer was added and mixed. Next, 2-5 units of the restriction enzyme were added followed by incubation for 1-2 hrs at appropriate temperature. This reaction was stopped by adding 0.5 M EDTA (pH 8.0) to a final conc. of 10 mM. Lastly, the size of DNA band(s) was compared with molecular weight markers using agarose gel electrophoresis.

3.2 *In-silico* analysis

3.2.1 Nucleotide Sequence of the cDNA from Indian potato cultivar KC-1 was obtained

↓

BLASTn at NCBI (http://www.ncbi.nlm.nih.gov)

↓

BLASTp at NCBI (http://www.ncbi.nlm.nih.gov)

↓

The sequences were retrieved and selected for further analysis

↓

Sequence analysis, prediction of motifs and 3-dimensional structure of cloned gene

3.2.2 Multiple Sequence Alignment

After performing BLASTn and BLASTp, different nucleotide and protein sequences were selected to perform Multiple Sequence Alignment (MSA) using multalin server (http://multalin.toulouse.inra.fr/multalin/). This server was established by Florence Corpet in 1988. MSA at both nucleotide and protein level was performed using nucleotide and protein sequences of different species of *Solanaceae* family.

3.2.3 Identification of the conserved domains

Prediction of motifs was done by uploading the amino acid sequence of the cloned gene to MY HITS server (https://myhits.sib.swiss/cgi-bin/motif_scan)

3.2.4 Hydropathyplot

The Protscale-Expasy (https://web.expasy.org/protscale/) tool was used to create hydropathy plots for each query protein. It helps to analyze if the amino acids present in the protein are hydrophobic or hydrophilic. Then the amino acid composition was analyzed using ProtParam Expasy tool (https://web.expasy.org/protparam/).

3.2.5 Three dimensional structure prediction

It is critical to understand a protein's structural information in order to predict its biological activity. For the prediction of protein 3-D structure I-TASSER server was used (https://zhanggroup.org/I-TASSER/). This server is a web-based platform that uses I-TASSER algorithms to predict protein structure and function. Then the quality of the 3-D structures was verified using QMEAN SWISS-MODEL tool (https://swissmodel.expasy.org/qmean/). X-ray/NMR analysis was also done using ProSa-web programme (https://prosa.services.came.sbg.ac.at/prosa.php).

2.2.6 Ramachandran Plot

Lastly the Ramachandran plots were made using PROCHECK server (https://saves.mbi.ucla.edu/).

4.1 Molecular cloning and characterization of *POTM1-1A*

Understanding the structure and function of MADS-box TF in potato, as well as its expression patterns are key areas of research in order to comprehend the complex tuberization mechanism. This study focused on molecular cloning of MADS-box TF i.e., *POTM1-1A* and to compare the expression patterns of different MADS-box TFs in various other plant species of *Solanaceae* family. The outcomes of this study are discussed as follows.

4.1.1 Isolation of total RNA from tissues of potato cultivar KC-1: Total RNA was extracted from various tissues i.e., leaf, stem, tuberizing stolon, Stage 1 tuber (ST-1), Stage 2 tuber (ST-2), Stage 3 tuber (ST-3), Stage 4 tuber (ST-4) and flower. Agarose gel electrophoresis was used to verify the crude RNA samples as shown in **Fig. 4.1.** Ribosomal RNA bands were evident, indicating that whole RNA preparations were intact. Impurities of genomic DNA were found in some of the RNA samples.

Fig 4.1 Total RNA from various potato tissues. Lane 1- Leaf, Lane 2- Stem, Lane 3- Tuberizing stolon Lane 4- ST-1 tuber, Lane 5-ST-2 tuber, Lane 6-ST-3 tuber, Lane 7- ST-4 tuber, Lane 8- Flower

4.1.2 Purification of total RNA: RNase-free DNase treatment and solvent extraction were used to purify total RNA. Agarose gel electrophoresis was used to verify the purified RNA samples as shown in **Fig. 4.2.** The presence of distinct bands in lane 3 and 5 indicates the purity of the RNA. In lane 2 and 4, bands indicate presence of DNA. The A_{260}/A_{280} ratio was calculated using a nanodrop spectrophotometer to determine the quality and amount of the produced RNA samples.

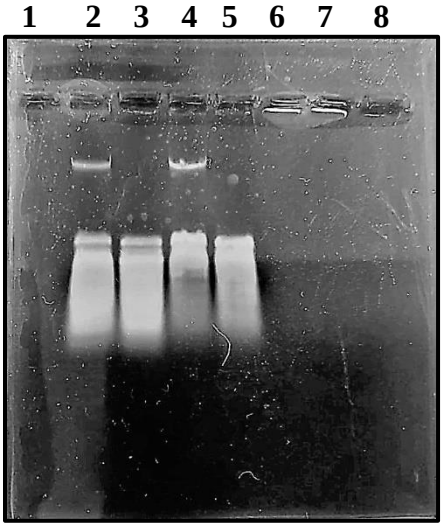

Fig 4.2 Purified RNA from different Potato organs. Lane 2 and 4- Unpurified RNA, Lane 3 and 6- Purified RNA

4.1.3 RT-PCR: The purified total RNA was used as a template for cDNA synthesis. PCR was carried out using forward and reverse primers as mentioned in material and methods. Agarose gel electrophoresis was then carried out using the PCR product. The bands were observed as given in **Fig. 4.3.**

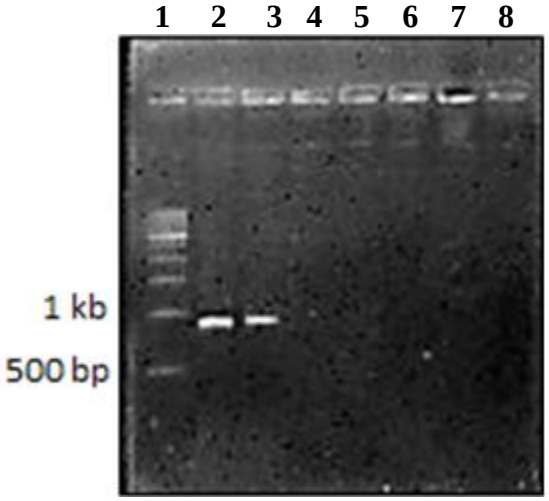

Fig 4.3 Agarose gel electrophoresis of RT-PCR products. Lane 1- 1 kb ladder, Lane 2- tuber, Lane 3- leaf

As seen in the **Fig. 4.3**, intact bands for cDNA amplification corresponding to *POTM1-1A* from tuber and leaf were observed. By comparing these bands with those of 1 kb ladder, the size of each amplicon was found to be less than 1kb (~0.9 kb). However a slight difference between bands intensity in lane 1 and 2 showed that the expression of *POTM1-1A* differs among various tissues of a potato plant.

4.1.4 Purification of the RT-PCR product: Firstly, the RT-PCR product (~0.9 kb) of tuber was treated with Klenow fragment to create blunt ends and then purified using PCR Purification kit. After purification the product was analyzed by agarose gel electrophoresis and intact band was observed as shown in **Fig. 4.4.**

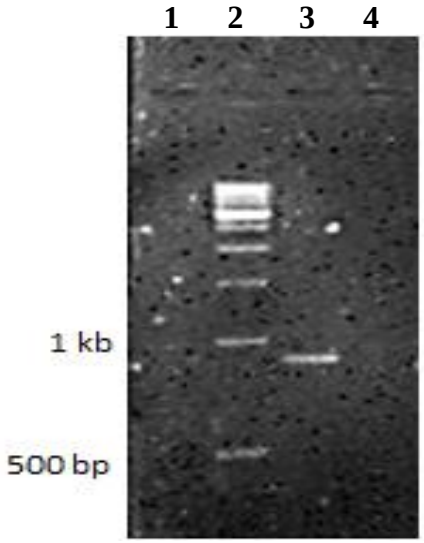

Fig 4.4 Purified RT-PCR product. Lane 1- 1 kb ladder, Lane 2- tuber

4.1.5 Blue-white screening: After the transformed cells were spread on X-gal/IPTG plates, blue white colonies were formed. Then patches were made from recombinant white colonies (i.e., putative clones) and blue-white patches were observed on the plates next day as shown in **Fig. 4.5.** White colonies were formed from the transformed cells that contain recombinant vector. On the other hand the blue colonies were formed from the non-transformed cells.

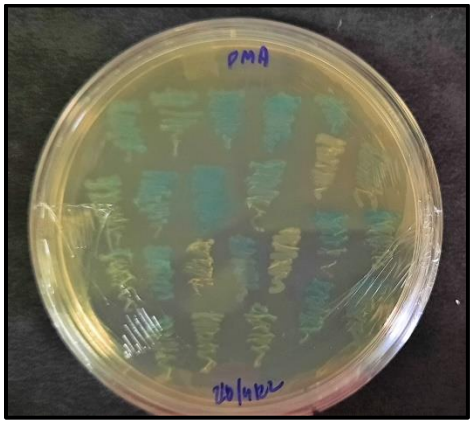

Fig 4.5 Blue and white patches on X-gal/IPTG plates

4.1.5 Isolation and RNase treatment of plasmid DNA: White colonies were inoculated in LB media containing ampicillin and isolation of plasmid DNA was done using boiling lysis method.

The crude plasmid DNA was treated with RNase for purification and was resolved by Agarose gel electrophoresis. The bands observed are shown in the **Fig. 4.6.**

Fig4.6 Agarose gel electrophoresis of purified putative clones. Lane 1,3,5,7,8-Unpurified putative clones; Lane 2,4,6- Purified putative clones

4.1.6 Preliminary characterization of putative clones: Restriction digestion (EcoR1 and BamH1) and clone mediated PCR was used for preliminary analysis of the putative clones. PCR reaction was set using the same set of primers as in RT-PCR. The samples after analysis were resolved by Agarose gel electrophoresis. The bands observed are shown in **Fig. 4.7.**

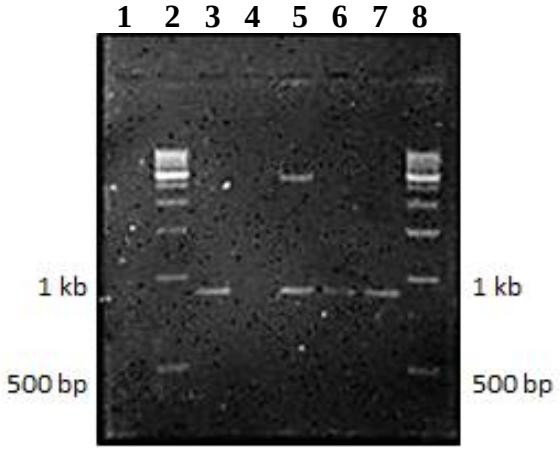

Fig 4.7 Preliminary characterization of putative clone. Lane2- 500 bp ladder, Lane 3- Mother PCR, Lane 5- Restriction digested PCR, Lane 6- PCR, Lane 7- Mother PCR, Lane 8- 500 bp ladder

The clones that were derived from the total RNA samples were then sent for sequencing to obtain the nucleotide sequence of *POTM1-1A*.

4.2 *In-silico* Analysis

The 821 bp long gene sequence named *POTM1-1A* (GenBank accession no. ON804239) corresponding to *POTM1-1* was obtained through RT-PCR mediated cloning as discussed in earlier sections. As MADS-box gene plays significant role during the process of tuberization. *In silico* characterization of this crucial gene *POTM1-1A* will enable us to explore various sequence features, biochemical properties and conserved motifs. Also, a comparison with other *Solanaceae* family members through multiple sequence alignment could provide us more clues on conserved regions across the family. To speculate the biological function of any protein, it is important to understand the structural and functional relationship. Hence, to understand the structure of *POTM1-1A,* 3D-model was built. These 3D-structures were further validated through Ramachandran plot analyses. A detailed explanation of the results obtained through bioinformatics analysis is discussed in following sections.

4.2.1 Salient sequence features

The nucleotide and amino acid sequences of the *POTM1-1A* are shown in **Fig. 4.8.** Using ExPASy ProtParam tool, the molecular weight and pI of *POTM1-1A* was found to be 28.92 kDa and 8.77 respectively. GRAVY value was -0.920, as provided by the server.

POTM1-1A gene and protein: As shown in the **Fig. 4.8**, *POTM1-1A* is the 821 bp cloned *POTM1* cDNA derived from potato cultivar KC-1. The CDS starts from 1st bp and and ends at 753rd bp. So, the total length of ORF is 753 bp including the stop codon. The region from 754th bp to 821st bp is the 3' flanking region. This sequence of cloned gene was compared with the nucleotide sequence of *POTM1* gene of superior cultivar. The nucleotides of the cloned gene differed from that of *POTM1* at 6 positions, which are marked red and highlighted with grey color. The polypeptide of POTM1-1A consists of 250 amino acids. The comparison of the polypeptide sequences of the cloned gene and the *POTM1* gene revealed only one variation, which is highlighted with blue color.

From the above observations it can be concluded that although there are 6 variations in nucleotide sequences of both the genes, the variation among amino acid sequences is only at one place i.e., leucine is replaced by isoleucine in *POTM1-1A*. While mutations usually alter the DNA sequence, they don't necessarily result in a change in the protein or a noticeable effect on the organism. This is due to the fact that most amino acids can be coded by two or more codons. These types of mutations are conservative mutations. The amino acid that is replaced in a

conservative missense mutation has a function and form that is similar to the amino acid that is being replaced.

```
ATGGGTAGAGGTAGAGTCCAGTTGAAGCGTATAGAGAACAAAATTAACCGTCAAGTGACTTTCTCGAAAC   70
M  G  R  G  R  V  Q  L  K  R  I  E  N  K  I  N  R  Q  V  T  F  S  K  R   24

GTCGATCTGGTTTGCTGAAGAAAGCTCATGAGATCTCTGTGCTTTGTGATGCCGAGGTTGGTTTGATTGT  140
  R  S  G  L  L  K  K  A  H  E  I  S  V  L  C  D  A  E  V  G  L  I  V    47

CTTTTCCACTAAAGGAAAACTCTTTGAATATGCCAATGATTCATGCATGGAGAGGTTACTTGAAAGATAT  210
 F  S  T  K  G  K  L  F  E  Y  A  N  D  S  C  M  E  R  L  L  E  R  Y    70

GAAAGATACTCATTTGCTGAGAGGCAGCTTGTTCCTACTGATCATACATCCCCGGGAAGCTGGACTCTGG  280
E  R  Y  S  F  A  E  R  Q  L  V  P  T  D  H  T  S  P  G  S  W  T  L  E   94

AACATGCAAAACTTAAGGCCAGACTTGAGGTTCTGCAGAGGAACCAAAAGCATTATGTGGGAGAAGATTT  350
  H  A  K  L  K  A  R  L  E  V  L  Q  R  N  Q  K  H  Y  V  G  E  D  L   117

GGAGTCGTTAAATATGAAAGAACTTCAGAATCTTGAACACCAGCTTGATTCTGCTCTTAAACACATTCGA  420
 E  S  L  N  M  K  E  L  Q  N  L  E  H  Q  L  D  S  A  L  K  H  I  R    140

TCAAGAAAGAACCAATTGATGCATGAATCCATTTCTGTGCTTCAAAAACAGGACAGAGCATTGCAGGAGC  490
S  R  K  N  Q  L  M  H  E  S  I  S  V  L  Q  K  Q  D  R  A  L  Q  E  Q  164

AAAATAACCAGCTTTCCAAGAAGGTGAAGGAGAGGGAGAAAGAGGTGGCACAGCAAAATCAGTGGGATCA  560
  N  N  Q  L  S  K  K  V  K  E  R  E  K  E  V  A  Q  Q  N  Q  W  D  Q   187

GCAGAACCATGAAATCAACTCATCTACATTTGTTTTGCCACAACAATTGGACTCTCCTCACCTTGGGGAA  630
 Q  N  H  E  I  N  S  S  T  F  V  L  P  Q  Q  L  D  S  P  H  L  G  E    210

GCATACCAGAATACTAATGTAGTAGATAATGGAGAAGTGGAAGGAGGTAATTCTTCACAGCAGCAAGGTG  700
A  Y  Q  N  T  N  V  V  D  N  G  E  V  E  G  G  N  S  S  Q  Q  Q  G  A  234

CAGCTAATAATACTGTGATGCCACAATGGATGCTTCGCCATATTAATGGTTAAAAGTACCACTAATACTT  770
  A  N  N  T  V  M  P  Q  W  M  L  R  H  I  N  G  *                    250

GATGGTCACTATCAATATAATTACCTAGGTATATATATATATATATCTATA                     821
```

Fig. 4.8: 821-bp *POTM1* cDNA from the potato cultivar KC-1 (designated *POTM1-1A*). The predicted POTM1 transcription factor i.e., 250-aa POTM1-1A is also presented. Dark red with grey highlight represent nucleotide mismatches as compared with the *POTM1* of the potato cultivar 'Superior' and a single amino acid substitution at 248th position is highlighted blue.

4.2.2 Sequence analysis of *POTM1-1A* by BLASTn

To compare the nucleotide sequence of *POTM1-1A* with MADS-box genes in other plant species, 753 nucleotide ORF of *POTM1-1A* was taken and BLASTn was performed. The result showed significant *POTM1-1A* sequence similarity with various plant species belonging to both *Solanaceae* and non *Solanaceae* family members. From all the sequences, we selected 8 nucleotide sequences belonging to *Solanaceae* family for further analysis. The 753 bp long gene

sequence showed 100% query coverage and 99.20% sequence identity with nucleotide sequence of (*St-SRPM1*) *Solanum tuberosum* (U23757.1), 100% query coverage and 98.67% sequence identity to nucleotide sequence in *Solanum tuberosum* Agamous like MADS-box protein (NM_001288213.1), 100% query coverage and 97.34% sequence identity to nucleotide sequence in *Solanum commersonii* MADS-box transcription factor (AF002666.1), 99% query coverage and 93.04% sequence identity to nucleotide sequence in *Solanum lycopersicum* Agamous like MADS-box protein (AF002666.1), 99% query coverage and 92.90% sequence identity to nucleotide sequence in *Lycopersicon esculentum* TD4 transcription factor (AY098732.1), 100% query coverage and 89.75% sequence identity to nucleotide sequence in Predicted: *Capsicum annum* Agamous like MADS-box protein AGL8 homolog (XM_016700410.2), 100% query coverage and 88.10% sequence identity to nucleotide sequence in Predicted: *Nicotiana attenuata* Agamous like MADS-box protein AGL8 homolog (XM_019381005.1), 100% query coverage and 87.32% sequence identity to nucleotide sequence in Predicted: *Nicotiana tabbacum* Agamous like MADS-box protein AGL8 homolog (XM_016593472.1).

4.2.3 Sequence analysis of *POTM1-1A* by BLASTp

To compare the polypeptide sequence of *POTM1-1A* with that of other plant species, 250 aa protein sequence of POTM1-1A (GenBank protein ID: UVH36213.1) was taken and BLASTp was performed. The result showed various plant species belonging to *Solanaceae* and non *Solanaceae* family having homologous sequences with POTM1-1A. The 250-amino acid sequence was found to be identical (99.60% sequence identity with 100% query coverage) to amino acid sequence in *Solanum tuberosum* (GenBank protein ID: AAA92839.1), 99.60% sequence identity with 100 % query coverage to the amino acid sequence in *Solanum tuberosum* agamous-like MADS-box protein AGL8 homolog (GenBank protein ID: NP_001275142.1), 95.60% sequence identity with 100% query coverage to the amino acid sequence in *Solanum commersonii* agamous-like MADS-box protein AGL8 homolog (GenBank protein ID: AAB65161.1), 93.98% sequence identity with 99% query coverage to the amino acid sequence in (Sl-AFUL1) *Solanum lycopersicum* agamous-like MADS-box protein AGL8 homolog (GenBank protein ID: NP_001234173.2), 93.57% sequence identity and 99% query coverage to nucleotide sequence in *Lycopersicon esculentum* TD4 transcription factor (GenBank protein ID: AAM33098.1), 89.64% sequence identity and 100% query coverage to nucleotide sequence in Predicted: *Capsicum annum* Agamous like MADS-box protein AGL8 homolog (GenBank protein ID: AAM33098.1), 85.66% sequence identity and 100% query coverage to nucleotide sequence in Predicted: *Nicotiana attenuata* Agamous like MADS-box protein AGL8 homolog

(GenBank protein ID: XP_019236550.1), 83.53% sequence identity and 100% query coverage to nucleotide sequence in Predicted: *Nicotiana tabbacum* Agamous like MADS-box protein AGL8 homolog (GenBank protein ID: XP_016448958.1). Details of some homologous sequences of other non-*Solanaceae* members are discussed in the **Table. 4.1.**

Table 4.1 Details of some homologous MADS-box protein sequences as available in the database

Species name	**Accession number**	**Max. score**	**Identity %**	**Query coverage %**	**Amino Acids**
Solanum tuberosum	AAA92839.1	515	99.60%	100%	250
Solanum tuberosum Agamous like	NP_001275142.1	515	99.60%	100%	250
Solanum commersonii	AAB65161.1	493	95.60%	100%	250
Solanum lycopersicum	NP_001234173.2	482	93.98%	99%	245
Lycopersicon esculentum	AAM33098.1	480	93.57%	99%	245
Capsicum annum	XP_016555896.1	457	89.64%	100%	251
Nicotiana attenuate	XP_019236550.1	434	85.66%	100%	248
Nicotiana tabacum	XP_016448958.1	415	83.53%	100%	241
Petunia X hybrid	AAF19721.1	419	83.00%	100%	246
Physalis pubescens	AGT21642.1	411	84.40%	99%	237
Cestrum nocturnum	QNU41459.1	396	79.12%	99%	234
Brugmansia suaveolens	QNU41472.1	395	79.37%	100%	247
Lycianthes glandulosa	ASL04652.1	390	78.09%	100%	243
Streptosolen jamesonii	QBL14536.1	388	78.37%	97%	238
Dunalia spinosa	QBL14489.1	379	77.96%	97%	238
Coffea Arabica	AHW58040.1	347	71.20%	99%	242
Carpinus fangiana	KAE8124485.1	344	72.29%	99%	243
Quercus lobata	XP_030937847.1	343	72.69%	99%	243
Hydrangea macrophylla	BAG68947.1	340	70.59%	100%	250
Castanea mollissima	KAF3976599.1	338	72.29%	99%	243

Vitis vinifera	NP_001268210.1	337	73.71%	85%	241
Betula pendula	CAA67969.1	337	71.60%	99%	244
Morella rubra	KAB1216810.1	336	70.68%	99%	245
Vitis riparia	XP_034699499.1	336	73.24%	85%	241
Jatropha curcas	XP_012080420.1	336	70.80%	99%	246

4.2.4 Multiple Sequence Alignment using nucleotide sequences

Multiple sequence alignment was done using cDNA nucleotide sequences from 9 different species of *Solanaceae* family as shown in the **Fig. 4.9.**

```
POTM1-1A  ATGGGTAGAG GTAGAGTCCA GTTGAAGCGT ATAGAGAACA AAATTAACCG TCAAGTGACT TTCTCGAAAC   70
St-SRPM1  ATGGGAAGAG GAAGAGTCCA GTTGAAGCGA ATAGAGAACA AAATTAACCG TCAAGTGACT TTCTCGAAAC   70
St-AFUL1  ATGGGAAGAG GAAGAGTCCA GTTGAAGCGA ATAGAGAACA AAATTAACCG TCAAGTGACT TTCTCGAAAC   70
Sc-SCM1   ATGGGAAGAG GAAGAGTCCA GTTGAAGCGA ATAGAGAACA AAATTAACCG TCAAGTGACT TTCTCGAAAC   70
Sl-AFUL1  ATGGGAAGAG GAAGAGTCCA GTTGAAGCGA ATAGAGAACA AAATTAACCG TCAAGTTACC TTCTCGAAAC   70
Le-TDR4   ATGGGAAGAG GAAGAGTCCA CTTGAAGCGA ATAGAGAACA AAATTAACCG TCAAGTTACC TTCTCGAAAC   70
Ca-AGL8   ATGGGGAGAG GAAGAGTACA ATTGAAGAGA ATAGAGAACA AAATAAACAG GCAAGTGACT TTCTCTAAAC   70
Na-AGL8   ATGGGAAGAG GAAGGGTACA GTTGAAGAGA ATTGAGAACA AAATTAATAG ACAAGTTACT TTCTCAAAAC   70
Nt-AGL8   ATGGGAAGAG GAAGGGTGCA GTTGAAGAGA ATTGAGAACA AAATTAATAG GCAAGTTACT TTCTCAAAAC   70
Consensus *****a**** ****a**.** g******a** **a******* ****t**ca* .*****t**t *****.****

POTM1-1A  GTCGATCTGG TTTGCTGAAG AAAGCTCATG AGATCTCTGT GCTTTGTGAT GCCGAGGTTG GTTTGATTGT  140
St-SRPM1  GTCGATCTGG TTTGCTGAAG AAAGCTCATG AGATCTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
St-AFUL1  GTCGATCTGG TTTGCTGAAG AAAGCTCATG AGATCTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
Sc-SCM1   GTCGATCTGG TTTGCTGAAG AAAGCTCATG AGATCTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
Sl-AFUL1  GTCGATCTGG TTTGCTGAAG AAAGCCCATG AGATCTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
Le-TDR4   GTCGATCTGG TTTGCTGAAG AAAGCCCATG AGATCTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
Ca-AGL8   GTCGATCTGG TTTATTGAAG AAAGCTCATG AGATCTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
Na-AGL8   GTCGATCTGG TTTGCTTAAG AAAGCTCATG AGATTTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
Nt-AGL8   GTCGATCTGG TTTGCTTAAG AAAGCTCATG AGATCTCTGT GCTTTGTGAT GCTGAGGTTG GTTTGATTGT  140
Consensus ********** ***gc*g*** *****t**** ****c***** ********** ********** **********

POTM1-1A  CTTTTCCACT AAAGGAAAAC TCTTTGAATA TGCCAATGAT TCATGCATGG AGAGGTTACT TGAAAGATAT  210
St-SRPM1  TTTTTCCACT AAAGGAAAAC TCTTTGAATA TGCCAATGAT TCATGCATGG AGAGGTTACT TGAAAGATAT  210
St-AFUL1  TTTTTCCACT AAAGGAAAAC TCTTTGAATA TGCCAATGAT TCATGCATGG AGAGGCTACT TGAAAGATAT  210
Sc-SCM1   TTTTTCCACT AAAGGAAAAC TCTTTGAATA TGCAACTGAT TCATGCATGG AGAGGTTACT TGAAAGATAT  210
Sl-AFUL1  TTTTTCTACT AAAGGAAAAC TCTTTGAATA TGCCAACGAT TCCTGCATGG AGAGGATACT TGAAAGATAT  210
Le-TDR4   TTTTTCTACT AAAGGAAAAC TCTTTGAATA TGCCAACGAT TCCTGCATGG AGAGGATACT TGAAAGATAT  210
Ca-AGL8   TTTTTCTACT AAAGGCAAAC TCTTTGAATA TGCCAATGAT TCCTGCATGG AAAGGATACT TGAAAGATAT  210
Na-AGL8   TTTTTCTACT AAAGGCAAAC TCTTTGAATA TGCCACTGAT TCTTGCATGG AGAGGATCCT TGAAAGATAT  210
Nt-AGL8   TTTTTCTACA AAAGGCAAAC TCTTTGAATA TGCCACTGAT TCTTGCATGG AGAGGATCCT TGAAAGATAC  210
Consensus ******t**t *****c**** ********** *****at*** **.******* *g***a*a** *********t

POTM1-1A  GAAAGATACT CATTTGCTGA GAGGCAGCTT GTTCCTACTG AT...CATAC ATCCCCGGGA AGCTGGACTC  277
St-SRPM1  GAAAGATACT CATTTGCTGA GAGGCAGCTT GTTCCTACTG AT...CATAC ATCCCCGGGA AGCTGGACTC  277
St-AFUL1  GAAAGATACT CATTTGCTGA GAGGCAGCTT GTTCCTACTG AT...CATAC ATCCCCGGGA AGCTGGACTC  277
Sc-SCM1   GAAAGATACT CATTTGCTGA GAAGCAGCTT GTTCCTACTG AT...CATAC ATCCCCGGGA AGCTGGACTC  277
Sl-AFUL1  GAAAGATACT CATTTGCTGA GAAACAGCTT GTTCCTACTG AT...CATAC CTCCCCGGTA AGCTGGACCC  277
Le-TDR4   GAAAGATACT CATTTGCTGA GAAACAGCTT GTTCCTACTG AT...CATAC CTCCCCGGTA AGCTGGACCC  277
Ca-AGL8   GAAAGATACT CGTATGCTGA GAGGCAGCTT GTTCCTAATG ATGATCATAC CTCCCCGGGA AGCTGGACTC  280
Na-AGL8   GAAAGATACT CATATGCTGA GAGGCAGCTT GTTGCTACTA CTGATCATAG CTACCCGGGA AGCTGGACTC  280
Nt-AGL8   GAAAGATACT CATATGCTGA GAGGCAACTT GTTGCTACTA CTGATCATAG CTGCCCGGGA AGCTGGACCC  280
Consensus ********** *a*a****** **gg**g*** ***c***c*g a*gat****c c*c*****g* ********t*
```

Fig. 4.9 contd…

```
POTM1-1A  TGGAACATGC AAAACTTAAG GCCAGACTTG AGGTTCTGCA GAGGAACCAA AAGCATTATG TGGGAGAAGA  347
St-SRPM1  TGGAACATGC AAAACTTAAG GCCAGACTTG AGGTTCTGCA GAGGAACCAA AAGCATTATG TGGGAGAAGA  347
St-AFUL1  TGGAACATGC AAAACTTAAG GCCAGACTTG AGGTTCTTCA GAGGAACCAA AAGCATTATG TGGGAGAAGA  347
Sc-SCM1   TTGAAAATGC AAAACTTAAG GCCAGACTTG AGGTTCTGCA GAGGAACGAA AAGCTTTATG TGGGAGAAGA  347
Sl-AFUL1  TTGAACATGC AAAACTTAAG GCCAGACTTG AGGTTCTGCA GAGGAACCAA AAGCATTATG TGGGAGAAGA  347
Le-TDR4   TTGAACATGC AAAACTTAAG GCCAGACTTG AGGTTCTGCA GAGGAACCAA AAGCATTATG TGGGAGAAGA  347
Ca-AGL8   TGGAACATGC AAAACTTAAG GCCAGACTTG AAGTTTTGCA GAGGAACCAA AAGCATTATG CAGGAGAAGA  350
Na-AGL8   TGGAACATGC AAAACTTAAG GCTAGACTTG AGGTTTTGCA GAGAAACCAA AGGCATTATA TGGGAGAAGA  350
Nt-AGL8   TGGAACATGC AAAACTTAAG GCTAGACTTG AGGTTTTGCA GAGAAACCAA AGGCATTATA TGGGAGAAGA  350
Consensus *g******** ********** **c******* *g***t**** ***g****** *a*******g tg********

POTM1-1A  TTTGGAGTCG TTAAATATGA AAGAACTTCA GAATCTTGAA CACCAGCTTG ATTCTGCTCT TAAACACATT  417
St-SRPM1  TTTGGAGTCG TTAAATATGA AAGAACTTCA GAATCTTGAA CACCAGCTTG ATTCTGCTCT TAAACACATT  417
St-AFUL1  TTTGGAGTCG TTAAATATGA AAGAACTTCA GAATCTTGAA CACCAGCTTG ATTCTGCTCT TAAACACATT  417
Sc-SCM1   TTTGGAGTCG TTAAATATGA AAGAACTTCA GAATCTTGAA CACCAGCTTG CTTCTGCTCT TAAACACATT  417
Sl-AFUL1  TTTGGAGTCC TTAAGTATGA AGGAACTTCA GAATCTGGAG CACCAGCTTG ATTCAGCTCT TAAACACATT  417
Le-TDR4   TTTGGAGTCC TTAAGTATGA AGGAACTTCA GAATCTGGAG CACCAGCTTG ATTCAGCTCT TAAACACATT  417
Ca-AGL8   ATTAGATTCG TTAAATATGA AAGAACTTCA GAATCTGGAG CACCAGCTTG ATTCTGCTCT TAAACACATT  420
Na-AGL8   TCTGGACTCG TTAAGCATGA AAGAACTTCA GAATTTGGAG CAGCAGCTGG ATTCTGCTCT TAAACACATT  420
Nt-AGL8   TTTGGACTCG TTAAGTACGA AGGAACTTCA GAATTTGGAA CACCAGCTGG ATTCTGCTCT TAAACACATT  420
Consensus tt*g**.**g ****gt*t** *a******** ****c*g**g **c*****t* ****t***** **********

POTM1-1A  CGATCAAGAA AGAACCAATT GATGCATGAA TCCATTTCTG TGCTTCAAAA ACAGGACAGA GCATTGCAGG  487
St-SRPM1  CGATCAAGAA AGAACCAATT GATGCATGAA TCCATTTCTG TGCTTCAAAA ACAGGACAGA GCATTGCAGG  487
St-AFUL1  CGATCAAGAA AGAACCAATT GATGCATGAG TCCATTTCTG TGCTTCAAAA ACAGGACAGA GCATTGCAGG  487
Sc-SCM1   CGATCAAGAA AGAACCAATT GATGCATGAG TCCATTTCTG TGCTTCAAAA ACAGGACAGA GCATTGCAGG  487
Sl-AFUL1  CGATCAAGAA AGAATCAATT GATGCATGAG TCCATTTCTG TGCTTCAAAA AAAGGACAGA GCATTGCAGG  487
Le-TDR4   CGATCAAGAA AGAATCAATT GATGCATGAG TCCATTTCTG TGCTTCAAAA AAAGGACAGA GCATTGCAGG  487
Ca-AGL8   CGATCAAGAA AGAACCAATT GATGCATGAG TCCATATCTG TGCTTCAAAA AAAGGACAAA GCATTGCAGG  490
Na-AGL8   CGCTCAAGAA AGAACCAATT GATGCATGAG TCTATTTCTG AGCTTCAAAA AAAGGACAAA GCACTGCAGG  490
Nt-AGL8   CGCTCAAGCA AGAACCAATT GATGCATGAG TCTATTTCTG AGCTTCAAAA AAAGGACAAA GCACTGCAGG  490
Consensus **a*****a* ****c***** *********g **c**t**** t********* *a******a* ***t******

POTM1-1A  AGCAAAATAA CCAGCTTTCC AAGAAGGTGA AGGAGAGGGA GAAAGAGGTG GCACAGCAAA ATCAGTGGGA  557
St-SRPM1  AGCAAAATAA CCAGCTTTCC AAGAAGGTGA AGGAGAGGGA GAAAGAGGTG GCACAGCAAA ATCAGTGGGA  557
St-AFUL1  AGCAAAACAA CCAGCTTTCC AAGAAGGTGA AGGAGAGGGA GAAAGAGGTG GCACAGCAAA ATCAGTGGGA  557
Sc-SCM1   AGCAAAACAA CCAGCTTTCC AAGAAGGTGA AGGAGAGGGA GAAAGAGGTG GAACAGCAAA ATCAGTGGGA  557
Sl-AFUL1  AGCAAAACAA CCAGCTTTCG AAGAAGGTGA AGGAGAGGGA GAAAGAGGTG GCACAGCAAA ATCAGTGGGA  557
Le-TDR4   AGCAAAACAA CCAGCTTTCG AAGAAGGTGA AGGAGAGGGA GAAAGAGGTG GCACAGCAAA ATCAGTGGGA  557
Ca-AGL8   AGCAAAACAA CCAGCTTTCC AAGAAGGTGA AGGAGAGGGG GAAAGAACTG GCCCAGCAAA ACAATTGGGA  560
Na-AGL8   AGCAAAACAA CCAGCTTTCC AAGAAGGTGA AAGAGAGGGA GAAAGAGCTA GCCCAGCAAA ATCAGTGGGA  560
Nt-AGL8   AGCAAAACAA CCAGCTTTGC AAGAAGGTGA AGGAGAGGGA GAAAGAGTTG GCCCAGCAAA ATCAGCGGGA  560
Consensus *******c** ********cc ********** *g*******a ******g.*g **c******* *tc*gt****

POTM1-1A  TCAGCAGAAC CATGAAATCA ACTCATCTAC ATTTGTTTTG CCACAACAAT TGGACTCTCC TCACCTTGGG  627
St-SRPM1  TCAGCAGAAC CATGAAATCA ACTCATCTAC ATTTGTTTTG CCACAACAAT TGGACTCTCC TCACCTTGGG  627
St-AFUL1  TCAGCAGAAC CATGAAATCA ACTCATCTAC ATTTGTTTTG CCACAACAAT TGGACTCTCC TCACCTTGGG  627
Sc-SCM1   TCAGCAGAAC CATGAAATCA ACTCATCTAC ATTTGTTTTG CCACAACAAT TGGACTCTCC TCACCTTGGG  627
Sl-AFUL1  .......... .....AATCA ACTCATCTTC ATTTGTTTTG CCACAACAAC TGGACTCTCC TCACCTTGGG  612
Le-TDR4   .......... .....AATCA ACTCATCTTC ATTTGTTTTG CCACAACAAC TGGACTCTCC TCACCTTGGG  612
Ca-AGL8   TCAACAGAAT AATGAAATCA GTTCATCTTC ATTTGTTTTG CCACAGCAAT TGGACTCTTC CCTCCCTAGG  630
Na-AGL8   GCAGCAGAAC CAAGAACTCA ACTCATCTTC ATTTGTTTTG CAACAACAAT TAGACTCTCC CAACCTTGGG  630
Nt-AGL8   GCAGCAGAAC CAAGAACTCA ACTCACCTTC ATTTATTTTT CAACAGCAAT TGGACTCTCC TAACCTTGGG  630
Consensus .cagcagaac ca.ga*a*** ac***t**t* ****g****g *c***a***t *g******c* tca**t*g**

POTM1-1A  GAAGCATACC AGAATACTAA TGTAGTAGAT AATGGAGAAG TGGAAGGAGG TAATTCTTCA CAGCAGCAAG  697
St-SRPM1  GAAGCATACC AGAATACTAA TGTAGTAGAT AATGGAGAAG TGGAAGGAGG TAATTCTTCA CAGCAGCAAG  697
St-AFUL1  GAAGCATACC AGAATACTAA TGTAGTAGAT AATGGAGAAG TGGAAGGAGG TAATTCTTCA CAGCAGCAAG  697
Sc-SCM1   GAAGCATCCC AGAATACTAA TGTAGTAGAT AATGGAGAAG TGGAAGGAGG TAATTCTTCA CAGCANCAAG  697
Sl-AFUL1  GAAGCATACC AGAGTACTAA TGTAATAGAT AATGGGGAAG TGGAAGGAGG TAGTTCTTCA CAGCAGCAAG  682
Le-TDR4   GAAGCATACC AGAGTACTAA TGTAATAGAT AATGGGGAAG TGGAAGGAGG TAGTTCTTCA CAGCAGCAAG  682
Ca-AGL8   GAAGCATACC AGAATACTGC AGGAGTAGAT AATGGTGAAG TAGAAGGAGC TAATTCACAG CAGCAGCAGC  700
Na-AGL8   GAAGCATATC AGAGTACT.. .GCAGAAGAA AATGGAGAAG TAGAAAGAGG TTCGAATTCA CAGCAGCAAA  697
Nt-AGL8   GAAGCATATC AGAGTACT.. .GCAGAAGAA AATGGAGAAG TAGAAGGAGG TTCG...... CAGCAGCAAA  691
Consensus ********c* ***g****.. .*.*gt***t *****a**** *a***g***g *a.ttcttca ********a.
```

Fig. 4.9 contd…

```
POTM1-1A   GTGCAGCTAA TAATACTGTG ATGCCACAAT GGATGCTTCG CCATATTAAT GGTTAA  753
St-SRPM1   GTGCAGCTAA TAATACTGTG ATGCCACAAT GGATGCTTCG CCATCTTAAT GGTTAA  753
St-AFUL1   GTGCAGCTAA TAATACTGTG ATGCCACAAT GGATGCTTCG CCATCTTAAT GGTTAA  753
Sc-SCM1    GTGCAGCTAA TAATACTGTG ATGCCACAAT GGATGGTTCG CCATCTTAAT GGTTAA  753
Sl-AFUL1   GTGCAGCTAA TAATACTGTG ATGCCACAAT GGATGCTTCG TCATCTTAAT AATTAA  738
Le-TDR4    GTGCAGCTAA TAATACTGTG ATGCCACAAT GGATGCTTCG TCATCTTAAT AATTAA  738
Ca-AGL8    AAGCTGCACC TAATACTGTG ATGCCACAAT GGATGCTTCG CCATCTTAAT GGCTAA  756
Na-AGL8    CTGC...... TAATACTGTG ATGCCACCAT GGATGATTCG CCATCTTAAT GGCTAA  747
Nt-AGL8    CTGC...... TAATACTGTG ATGCCACCAT GGATGATTCG CCATCTTTAA ......  735
Consensus  .t**.gc... ********** *******a** *****c**** c******a*t gg.taa
```

Fig. 4.9 Multiple sequence alignment using MADS-box cDNA nucleotide sequences from different species of *Solanaceae* family. *POTM1-1A* (ON804239, *Solanum tuberosum* KC-1 Cultivar), *St-SRPM1* (U23757, *Solanum tuberosum* Superior cultivar), *St-AFUL1* (NM_001288213, *Solanum tuberosum* Agamous like), *Sc-SCM1* (AF002666, *Solanum commersonii*), *Sl-AFUL1* (NM_001247244, *Solanum lycopersicum*), *Le-TDR4* (AY098732, *Lycopersicon esculentum*), *Ca-AGL8* (XM_016700410, *Capsicum annum*), *Na-AGL8* (XM_019381005, *Nicotiana attenuata*), *Nt-AGL8* (XM_016593472, *Nicotiana tabacum*). Dashes indicate gaps that arise during alignment; Asterisks '*' indicate the conserved nucleotides; Lowercase letters refers to the almost conserved or conservative substitutions.

MSA revealed that most of the nucleotides were highly conserved. However, at some places species-specific insertion/deletion mutations could be seen. The observations were made using *POTM1-1A* as the reference sequence. From the figure it can be seen that the nucleotides from 253-255 are missing in first six nucleotide sequences that belong to species of potato and tomato. It means that insertion mutation has taken place in other three sequences of capsicum and tobacco with respect to *POTM1-1A*. Further nucleotides from 558-572 are missing in both species of tomato which means deletion mutation has taken place in both these sequences. Gaps can also be seen from nucleotides 646-648 in nucleotide sequences of tobacco and from 682-697 in *Nicotiana tabacum* indicating that deletion mutation has occurred. Further, gaps are observed from nucleotides 702-707 in both tobacco species and from 748-753 in *Nicotiana tabacum*. This again indicated that deletion mutation must have occurred in these sequences with respect to *POTM1-1A*. From these observations it can be concluded that potato and tomato are more closely related as compared to other species of *Solanaceae* family.

4.2.5 Multiple Sequence Alignment using amino acid sequences

Amino acid sequences from 9 different species of *Solanaceae* family were taken and multiple sequence alignment was performed. The results were obtained as shown in **Fig. 4.10.** All the observations were made with respect to amino acid sequence of POTM1-1A. Various species-specific insertion/deletion mutations were observed. From the figure it can be seen that amino acid at 85thaa is missing in first six sequences belonging to potato and tomato species indicating that insertion mutation has occurred in sequences of capsicum and tobacco species.

```
                                    MADS_BOX domain
          <------------------------------------------------------------
 POTM1-1A  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYAND   60
 St-SRPM1  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYAND   60
 St-AFUL1  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYAND   60
  Sc-SCM1  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYATD   60
 Sl-AFUL1  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYAND   60
  Le-TDR4  MGRGRVHLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYAND   60
  Ca-AGL8  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYAND   60
  Na-AGL8  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYATD   60
  Nt-AGL8  MGRGRVQLKR IENKINRQVT FSKRRSGLLK KAHEISVLCD AEVGLIVFST KGKLFEYATD   60
Consensus  ******q*** ********** ********** ********** ********** ********n*

                                                K-Box domain
          >              ==================================================
 POTM1-1A  SCMERLLERY ERYSFAERQL VPTD.HTSPG SWTLEHAKLK ARLEVLQRNQ KHYVGEDLES  119
 St-SRPM1  SCMERLLERY ERYSFAERQL VPTD.HTSPG SWTLEHAKLK ARLEVLQRNQ KHYVGEDLES  119
 St-AFUL1  SCMERLLERY ERYSFAERQL VPTD.HTSPG SWTLEHAKLK ARLEVLQRNQ KHYVGEDLES  119
  Sc-SCM1  SCMERLLERY ERYSFAEKQL VPTD.HTSPG SWTLENAKLK ARLEVLQRNE KLYVGEDLES  119
 Sl-AFUL1  SCMERILERY ERYSFAEKQL VPTD.HTSPV SWTLEHAKLK ARLEVLQRNQ KHYVGEDLES  119
  Le-TDR4  SCMERILERY ERYSFAEKQL VPTD.HTSPV SWTLEHAKLK ARLEVLQRNQ KHYVGEDLES  119
  Ca-AGL8  SCMERILERY ERYSYAERQL VPNDDHTSPG SWTLEHAKLK ARLEVLQRNQ KHYAGEELDS  120
  Na-AGL8  SCMERILERY ERYSYAERQL VATTDHSYPG SWTLEHAKLK ARLEVLQRNQ RHYMGEDLDS  120
  Nt-AGL8  SCMERILERY ERYSYAERQL VATTDHSCPG SWTLEHAKLK ARLEVLQRNQ RHYMGEDLDS  120
Consensus  *****i**** ****%**r** *ptdd*ts*g *****h**** *********# kh*.**#*#*

                                K-Box domain contd..
          ============================================================
 POTM1-1A  LNMKELQNLE HQLDSALKHI RSRKNQLMHE SISVLQKQDR ALQEQNNQLS KKVKEREKEV  179
 St-SRPM1  LNMKELQNLE HQLDSALKHI RSRKNQLMHE SISVLQKQDR ALQEQNNQLS KKVKEREKEV  179
 St-AFUL1  LNMKELQNLE HQLDSALKHI RSRKNQLMHE SISVLQKQDR ALQEQNNQLS KKVKEREKEV  179
  Sc-SCM1  LNMKELQNLE HQLASALKHI RSRKNQLMHE SISVLQKQDR ALQEQNNQLS KKVKEREKEV  179
 Sl-AFUL1  LSMKELQNLE HQLDSALKHI RSRKNQLMHE SISVLQKKDR ALQEQNNQLS KKVKEREKEV  179
  Le-TDR4  LSMKELQNLE HQLDSALKHI RSRKNQLMHE SISVLQKKDR ALQEQNNQLS KKVKEREKEV  179
  Ca-AGL8  LNMKELQNLE HQLDSALKHI RSRKNQLMHE SISVLQKKDK ALQEQNNQLS KKVKERGKEL  180
  Na-AGL8  LSMKELQNLE QQLDSALKHI RSRKNQLMHE SISELQKKDK ALQEQNNQLS KKVKEREKEL  180
  Nt-AGL8  LSTKELQNLE HQLDSALKHI RSSKNQLMHE SISELQKKDK ALQEQNNQLC KKVKEREKEL  180
Consensus  *sm******* h**d****** **r******* ***v***k*k *********s ******e**l


 POTM1-1A  AQQNQWDQQN HEINSSTFVL PQQLDSPHLG EAYQNTNVVD NGEVEGGNSS QQQGAANNTV  239
 St-SRPM1  AQQNQWDQQN HEINSSTFVL PQQLDSPHLG EAYQNTNVVD NGEVEGGNSS QQQGAANNTV  239
 St-AFUL1  AQQNQWDQQN HEINSSTFVL PQQLDSPHLG EAYQNTNVVD NGEVEGGNSS QQQGAANNTV  239
  Sc-SCM1  EQQNQWDQQN HEINSSTFVL PQQLDSPHLG EASQNTNVVD NGEVEGGNSS QXQGAANNTV  239
 Sl-AFUL1  AQQNQW.... .EINSSSFVL PQQLDSPHLG EAYQSTNVID NGEVEGGSSS QQQGAANNTV  234
  Le-TDR4  AQQNQW.... .EINSSSFVL PQQLDSPHLG EAYQSTNVID NGEVEGGSSS QQQGAANNTV  234
  Ca-AGL8  AQQNNWDQQN NEISSSSFVL PQQLDSSLPR EAYQNTAGVD NGEVEGANSQ QQQQAAPNTV  240
  Na-AGL8  AQQNQWEQQN QELNSSSFVL QQQLDSPNLG EAYQSTAE.E NGEVERGSNS QQQTA..NTV  237
  Nt-AGL8  AQQNQREQQN QELNSPSFIF QQQLDSPNLG EAYQSTAE.E NGEVEGGS.. QQQTA..NTV  235
Consensus  a***#w.qqn .*in*ss*!l p*****p.lg **y*s*a..# *****ggsss *q*.*a.***


 POTM1-1A  MPQWMLRHIN G  250
 St-SRPM1  MPQWMLRHLN G  250
 St-AFUL1  MPQWMLRHLN G  250
  Sc-SCM1  MPQWMVRHLN G  250
 Sl-AFUL1  MPQWMLRHLN N  245
  Le-TDR4  MPQWMLRHLN N  245
  Ca-AGL8  MPQWMLRHLN G  251
  Na-AGL8  MPPWMIRHLN G  248
  Nt-AGL8  MPPWMIRHL. .  244
Consensus  **q**l***n g
```

Fig. 4.10 Multiple sequence alignment using the full-length MADS-Box transcription factors from different plant species. 250 aa protein POTM1-1A (UVH36213.1, *Solanum tuberosum* KC-1 Cultivar), 250 aa protein St-SRPM1(AAA92839.1, *Solanum tuberosum* Superior cultivar), 250 aa protein St-AFUL1(NP_001275142.1, *Solanum tuberosum* Agamous like), 250 aa protein Sc-SCM1(AAB65161.1, *Solanum commersonii*), 245 aa protein Sl-AFUL1(NP_001234173.2, *Solanum lycopersicum*), 245 aa protein Le-TDR4 (AAM33098.1, *Lycopersicon*

esculentum), 251 aa protein Ca-AGL8 (XP_016555896.1, *Capsicum annum*), 248 aa protein Na-AGL8 (XP_019236550.1, *Nicotiana attenuata*), 244 aa protein Nt-AGL8(XP_016448958.1, *Nicotiana tabacum*). Various motifs and domains were shown with regard to 250-aa POTM1-1A as identified by MyHits tool. The segment 1–61 refers to MADS-BOX domain denoted by '<---- >'. The entire domain is also known as MADS-BOX_2 which includes MADS-BOX_1 domain (3−57) and the DNA-binding and dimerization domain (9–59); the segment 75–174 refers to K-Box domain denoted by '===='. Asterisks '*' indicate the conserved amino acids. Lowercase letters refers to the almost conserved or conservative substitutions. '!' is anyone of IV, '$' is anyone of LM, '%' is anyone of FY and '#' is anyone of NDQEBZ.

A gap from 186thaa to 190th aa in both species of tomato showed that deletion has occurred in these sequences. Gaps can be seen at the positions of the 218th, 236th, and 237th amino acids in the sequences of both tobacco species. Additionally, *Nicotiana tabacum* also lack four amino acids at positions 228, 229, 243, and 244. This indicated that deletion must have occurred in these sequences as compared to that of POTM1-1A. All these observations can help in study of the evolutionary relationship between these species.

In **Fig. 4.10** various important motifs have also been marked and it can be seen that the MADS-box domain is highly conserved in all the sequences and K-box domain is also mostly conserved. However, the C-terminal region shows a lot of variation as compared to the N-terminal region.

4.2.6 Motif Scan

A motif is a section of a protein sequence that has a particular structure and can serve as the basis for classifying proteins. They also help to predict the function of proteins. Important protein motifs that were identified using amino acid sequences of chosen MADS-box proteins from *Solanaceae* family, with the help of MY HITS program are described as follows:

- *MADS box domain:* The MADS domain (MCM1, AGAMOUS, DEFICIENS, and SRF, serum response factor) is a DNA-binding/dimerization area found in several transcription factors from other kingdoms. In vascular plants, MADS box genes form a vast multigene family.. The DNA-binding MADS domain is encoded by the MADS box. The MADS domain interacts to DNA sequences that are highly comparable to the CArG-box motif CC[A/T]$_6$GG.
- *Bipartite nuclear localization signal:* Two clusters of 2–3 positively charged amino acids separated by a 9–12 amino-acid linker region, which contains numerous proline (P) residues, describe Bipartite nuclear localization signal (BP NLS). In particular, in BP NLS, the upstream and downstream clusters of amino acids are dependent on each other and important. They work together to establish the protein's location in the cell.

Table 4.2 Position and motif sites of MADS-box proteins in different species

Site	MADS_BOX_1 *MADS-box domain signature.*	NLS_BP *Bipartite nuclear localization signal.*	K_BOX *K-box domain profile*	MADS_BOX_2 *MADS-box domain profile*	SRF-TF *SRF-type transcription factor (DNA-binding and dimerisation domain)*
Species					
Solanum tuberosun **(POTM1-1A)**	3-57	9-26	88-178	1-61	9-59
Solanum tuberosum **(St-SRPM1)**	3-57	9-26	88-178	1-61	9-59
Solanum tuberosum **(St-AFUL1)**	3-57	9-26	88-178	1-61	9-59
Solanum Commersonni **(Sc-SCM1)**	3-57	9-26	88-178	1-61	9-59
Lycopersicon esculentum **(Le-TDR4)**	3-57	9-26	88-178	1-61	9-59
Solanum lycopersicum **(Sl-AFUL1)**	3-57	9-26	88-178	1-61	9-59
Capsicum annum **(Ca-AGL8)**	3-57	9-26	89-179	1-61	9-59
Nicotiana attenuata **(Na-AGL8)**	3-57	9-26	89-179	1-61	9-59
Nicotiana tabacum **(Nt-AGL8)**	3-57	9-26	89-179	1-61	9-59

- *K-box domain profile:* The 80-amino-acid K-box domain was discovered as a region with low but considerable homology to a keratin area, a part of the coiled-coil sequence which is responsible for core rod-shaped domain of keratin. The K-box protein-protein interaction domain, which enables MIKC-type MADS protein heterodimerization, has many heptad repeats in which hydrophobic amino acids occupy the first and fourth positions, implying that the K-box domain forms three amphipathic α-helices known as K1, K2, and K3.

- *SRF-type transcription factor (DNA-binding and dimerisation domain):* The serum response factor (SRF) is a widely expressed transcription factor that binds to the CArG box, a DNA cis element located in the proximal regulatory regions of more than 200 experimentally verified target genes. SRF binds to the CArG box and makes particular major groove contacts with the G. C base pairs and the core A+T-rich region.

4.2.7 Hydropathy Plot

The ProtScale tool, which is based on the Kylie-Doolittle scale, was used to construct the hydropathy profile for POTM1-1A protein. The hydropathy plot of the MADS-box protein revealed that all four members of the family had an alternate pattern of hydrophobicity and hydrophilicity.

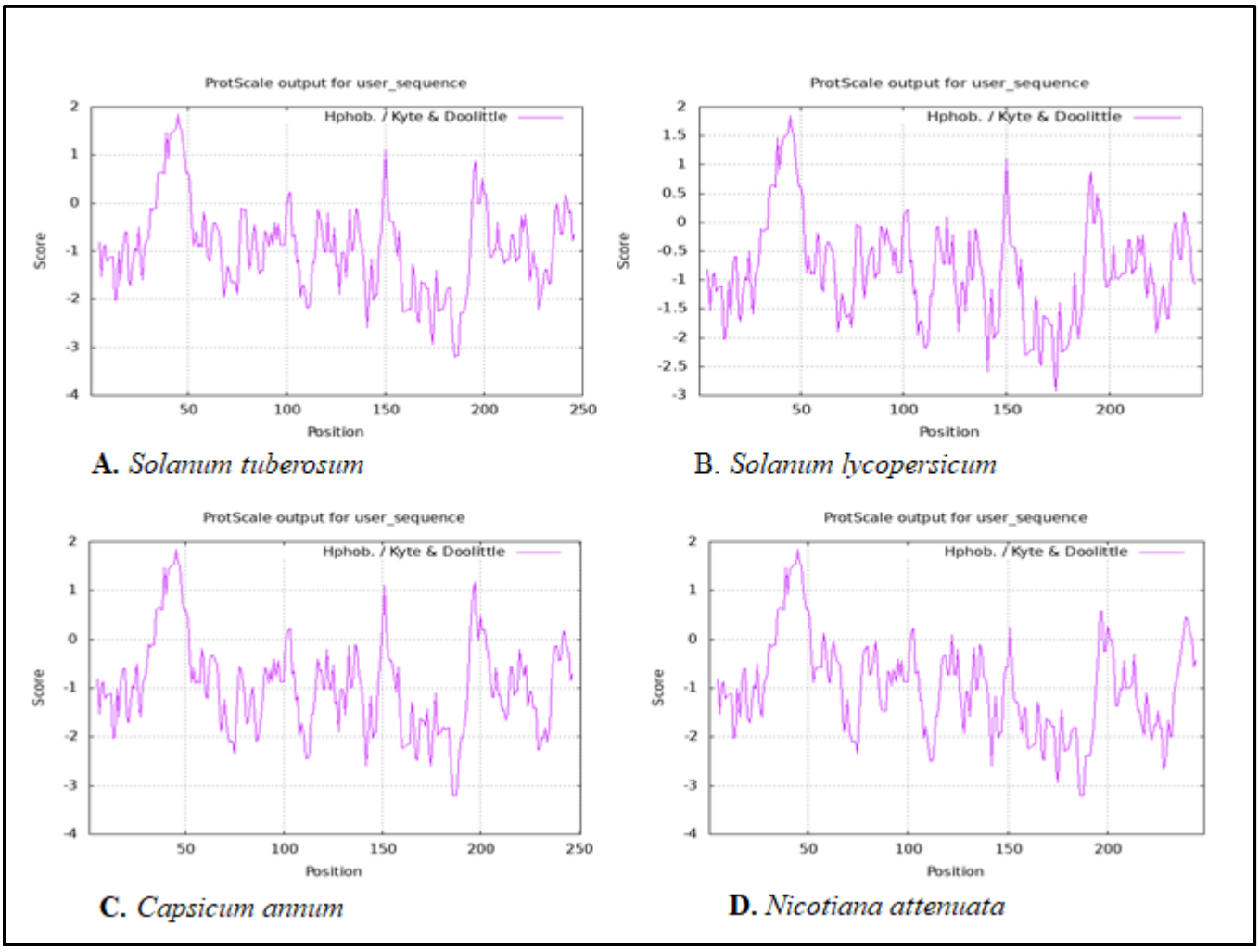

Fig. 4.11 Hydropathy plots of MADS-box proteins in 4 members of *Solanaceae* family

Using the scale Hphob. / Kyte & Doolittle, the individual values for the 20 amino acids are:

Ala: 1.800 Arg: -4.500 Asn: -3.500 Asp: -3.500 Cys: 2.500 Gln: -3.500

Glu: -3.500 Gly: -0.400 His: -3.200 Ile: 4.500 Leu: 3.800 Lys: -3.900

Met: 1.900 Phe: 2.800 Pro: -1.600 Ser: -0.800 Thr: -0.700 Trp: -0.900

Tyr: -1.300 Val: 4.200: -3.500: -3.500: -0.490

The **Fig. 4.11** shows the hydropathy plots of four different species of *Solanaceae* family. A protein's potential structure can be determined using Kyte-Doolittle hydropathy plots. The shape of the plot can be used to deduce information about the protein's partial structure. The amino acid sequence is represented on the x-axis, and the degree of hydrophobicity and hydrophilicity on the y-axis. The amino acids are said to be hydrophobic if the value is positive and, hydrophilic if the value is negative. As majority of amino acids in the sequence of POTM1-1A have negative values, it means that the protein will be hydrophilic in nature.

4.2.8 Amino Acid composition of MADS-box protein

Amino acid analysis helps to find out the number of the individual residues present in a polypeptide. It was done using ProtParam Expasy tool and the results are shown in **Table 4.3.**

As it can be seen in the table, all the amino acids except Pyl (O) and Sec (U) are present on variable amount in different plant species. In case of *Solanum tuberosum*, *Solanum lycopersicum* and *Nicotiana attenuata*, amino acids like glutamine, glutamic acid and leucine are present in greater concentrations than cysteine, phenylalanine, tryptophan and proline. But in case of *Capsicum annum* valine, proline, arginine and leucine are present in high concentration. The structure and function of the MADS-box proteins are influenced by these alterations in amino acid concentrations.

Table 4.3 Analysis of amino acid composition in MADS-box proteins in different plant species

Amino acid	***Solanum tuberosum***		***Solanum lycopersicum***		***Capsicum annuum***		***Nicotiana attenuata***	
	Repetition	**%occurrence**	**Repetition**	**%occurrence**	**Repetition**	**%occurrence**	**Repetition**	**%occurrence**
Ala (A)	12	4.8%	12	4.9%	7	4.0%	13	5.2%
Arg (R)	17	6.8%	16	6.5%	15	8.6%	18	7.3%
Asn (N)	19	7.6%	16	6.5%	12	6.9%	15	6.0%
Asp (D)	9	3.6%	8	3.3%	10	5.7%	8	3.2%
Cys (C)	2	0.8%	2	0.8%	3	1.7%	2	0.8%
Gln (Q)	25	10.0%	22	9.0%	4	2.3%	26	10.5%
Glu (E)	23	9.2%	23	9.4%	7	4.0%	26	10.5%
Gly (G)	13	5.2%	11	4.5%	12	6.9%	11	4.4%
His (H)	10	4.0%	9	3.7%	4	2.3%	7	2.8%
Ile (I)	8	3.2%	9	3.7%	9	5.1%	8	3.2%
Leu (L)	27	10.8%	27	11.0%	14	8.0%	28	11.3%
Lys (K)	18	7.2%	20	8.2%	3	1.7%	19	7.7%
Met (M)	6	2.4%	6	2.4%	4	2.3%	7	2.8%
Phe (F)	5	2.0%	5	2.0%	9	5.1%	4	1.6%
Pro (P)	5	2.0%	5	2.0%	15	8.6%	4	1.6%
Ser (S)	19	7.6%	23	9.4%	9	5.1%	22	8.9%
Thr (T)	8	3.2%	7	2.9%	10	5.7%	9	3.6%
Trp (W)	3	1.2%	3	1.2%	2	1.1%	3	1.2%
Tyr (Y)	5	2.0%	5	2.0%	7	4.0%	7	2.8%
Val (V)	16	6.4%	16	6.5%	19	10.9%	11	4.4%
Pyl (O)	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Sec (U)	0	0.0%	0	0.0%	0	0.0%	0	0.0%

4.2.9 Prediction of three-dimensional protein structure

The tertiary structure of a polypeptide refers to its overall three-dimensional structure. To annotate a protein's biological activity, a precise understanding of its 3D structure is required. The three-dimensional arrangement of a protein's atoms determines its biological function. A protein structure helps us understand how a protein works, allowing us to come up with new ways to influence, regulate, or modify it. It also helps to determine which compounds bind to that protein and gain a better understanding of different biological interactions. The server I-TASSER was used to model the entire protein sequences of *Solanum tuberosum*, *Solanum lycopersicum*, *Capsicum annum* and *Nicotiana attenuata*. I-TASSER is a hierarchical protocol for automatically predicting protein structure and annotating structure-based functions.

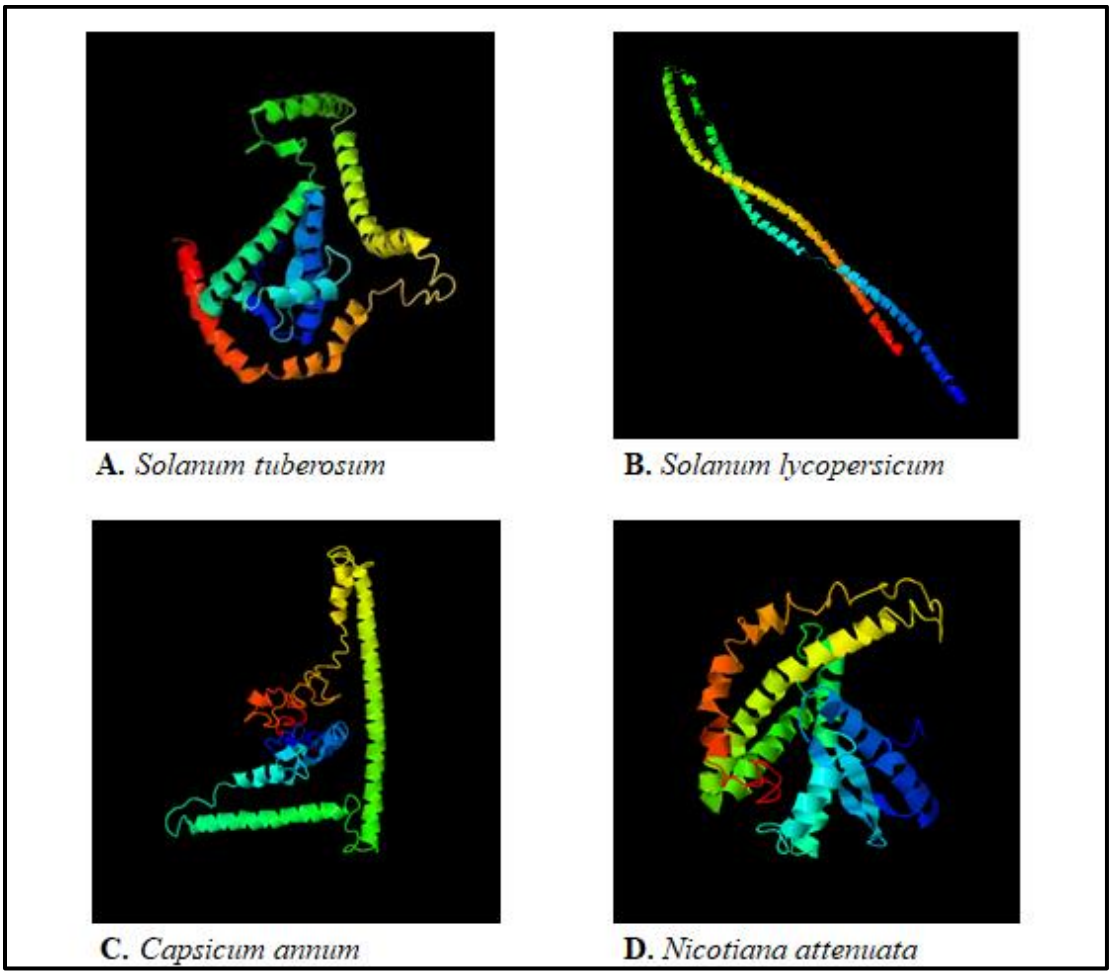

Fig. 4.12 3-D protein structure of 4 different species of *Solanaceae* family.

- ***Confidence score (C-score)***

The C-score is a confidence score used by I-TASSER to estimate the quality of predicted models. The C-score is usually in the range of [-5, 2], with a higher C-score indicating a more confident model and vice versa. Greater the value of C-score better is the model.

- ***Quality estimation***

Quality estimation of the 3-D structure is necessary to determine the applications of the protein. This is done by studying the Z-score values provided by I-TASSER. The Z-score of the alignment is commonly used to judge the quality of threading alignment. It is the difference in standard deviation between the raw and average scores. The energy score's standard deviation is determined as a percentage of the statistical average of all alignments. The superior alignment is represented by a normalized Z-score <1.

The normalized Z- score of the models obtained is shown in the **Tables 4.4 (A, B, C and D)**:

Table 4.4 Normalized Z scores for models of **A.** *Solanum tuberosum,* **B.** *Solanum lycopersicum,* **C.** *Capsicum annum* and **D.** *Nicotiana attenuata.*

A.

Rank	PDB hit	Norm. Z score
1.	1n6jA	1.84
2.	7ogtB	1.09
3.	1hbx	4.94
4.	4ox0A	1.61
5.	1hbx	3.56
6.	1n6j	5.70
7.	4ox0A	1.27
8.	1n6jA	1.48
9.	4ox0A	1.79
10.	1n6jA	2.57

B.

Rank	PDB hit	Norm. Z score
1.	4ox0A	1.84
2.	7x1nB	1.15
3.	7ogtB	1.06
4.	1hbx	5.01
5.	4ox0A	1.65
6.	1hbx	3.56
7.	1mnm	5.67
8.	4ox0A	1.33
9.	1n6jA	1.47
10.	4ox0A	1.31

C.

Rank	PDB hit	Norm. Z score
1.	1n6jA	1.78
2.	7ogtB	1.12
3.	7x1nB	1.15
4.	4ox0A	1.72
5.	1hbx	4.96
6.	1hbx	3.55
7.	4ox0A	1.24
8.	1n6j	5.83
9.	4ox0A	1.69
10.	1n6jA	1.48

D.

Rank	PDB hit	Norm. Z score
1.	1n6jA	1.75
2.	4ox0A	1.72
3.	7x1nB	1.10
4.	1hbx	5.05
5.	4ox0A	1.31
6.	1hbx	3.54
7.	4ox0A	1.73
8.	1n6jA	1.03
9.	4ox0A	1.35
10.	1n6j	5.68

- ***Structural similarity***

The structural similarity of two structures can be compared with the help of TM-score. **Tables 4.5 (A, B, C and D)** lists the top PDB proteins with the greatest TM-score and the most structural similarity to the predicted I-TASSER models.

Table 4.5 TM scores for protein models of **A.** *Solanum tuberosum*, **B.** *Solanum lycopersicum*, **C.** *Capsicum annum* and **D.** *Nicotiana attenuata.*

A.

Rank	PDB hit	TM score
1	7ogtB	0.896
2	6z9lA	0.706
3	6yvuA	0.700
4	5xg2A	0.691
5	5nnvA	0.654
6	7nyxA	0.635
7	6z6fC	0.629
8	1ciiA	0.607

B.

Rank	PDB hit	TM score
1	7ogtB	0.885
2	6yvuB	0.738
3	5xg2A	0.708
4	6z9La	0.663
5	5nnvA	0.655
6	1ciiA	0.633
7	6z6fC	0.630
8	7nyxA	0.620

C.

Rank	PDB hit	TM score
1	7ogtB1	0.850
2	6yvuB	0.773
3	5xg2A	0.720
4	7nyxA	0.694
5	5nnvA	0.686
6	6z9lA	0.686
7	6z6fC	0.652
8	6ek7A	0.603

D.

Rank	PDB hit	TM score
1	1g8xA	0.447
2	1h7wD	0.442
3	2uvcG	0.437
4	6qq5A	0.434
5	6uz0A	0.431
6	6m5aA	0.431
7	5gupl	0.429
8	7w9kA3	0.429
9	4he8L	0.429
10	5xtcl	0.427

In **Tables 4.5 A, B, C and D** all the possible PDB structures close to the target proteins are mentioned. TM-score of >0.5 suggests a valid topology model, while a TM-score of 0.17 implies random similarity. The length of the protein has no impact on these cutoffs. The goal of the TM-score is to overcome the RMSD problem, which is susceptible to local inaccuracy. RMSD is the average distance between all residues pairs in two structures, a local inaccuracy will result in a large RMSD number even though the global topology is accurate. However, with TM-score, the little distance is weighted more heavily than the large distance, making the result insensitive to local modeling mistake.

- ***Model quality***

The QMEAN i.e., Quality model energy analysis model, quality assessment was used to estimate the better model quality. It is a single model method that linearly combines statistical potentials and agreement terms. The QMEAN Z-score is calculated by comparing the absolute quality of a model to reference structures of similar size deposited in the PDB, i.e. model size +/- 10%. It ranges between 0 and 1 and higher the QMEAN Z-score, better the model. **Fig. 4.13** shows the

estimation of the quality of model that is obtained by QMEAN server. In the figure, all the structures have Z score between 0 and 1 which means the predicted models are of better quality.

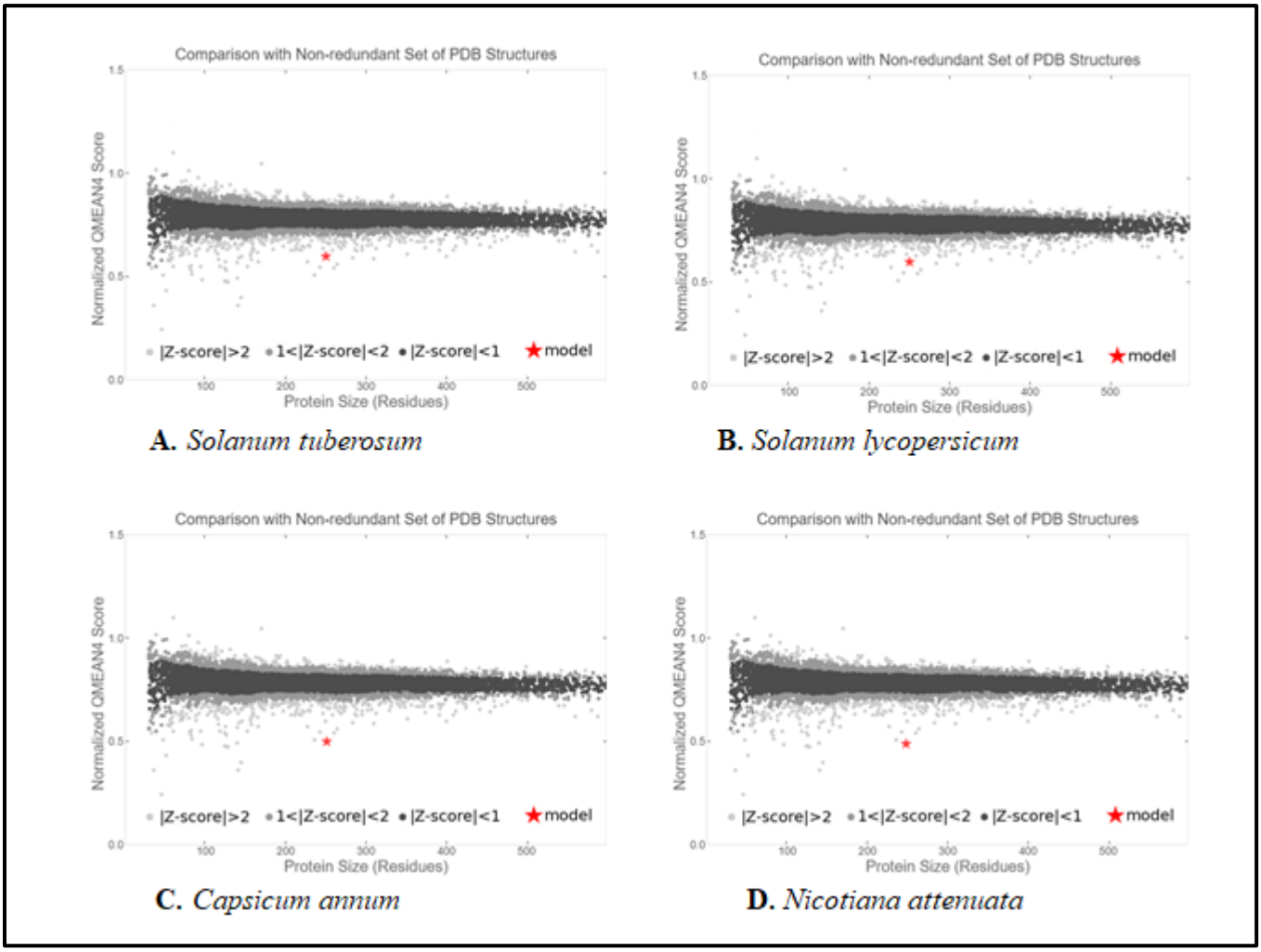

Fig. 4.13 The estimation of quality of models obtained by QMEAN server.

- ***B-factor profile***

The B-factor is a number that indicates how much thermal mobility residues/atoms in proteins have. The raw B-factor data are normalised using z-scores for a target protein's normalised B-factor. A mix of template-based assignment and profile-based prediction is used to forecast the normalised B-factor (called B-factor profile, BFP). In experimental structures, residues having BFP values greater than 0, are less stable. The best model is one in which the energy level of the sequences is smaller than zero. **Fig 4.14** shows the energy levels for the validation of given proteins. For all the protein models, positive normalized B-factors are projected for the N- and C-terminal sections as well as the majority of the loop regions, suggesting that all these regions

are more flexible structurally than other regions. The forecasted normalized B-factors are negative or nearly zero for the alpha and beta regions, indicating that these regions are structurally more stable.

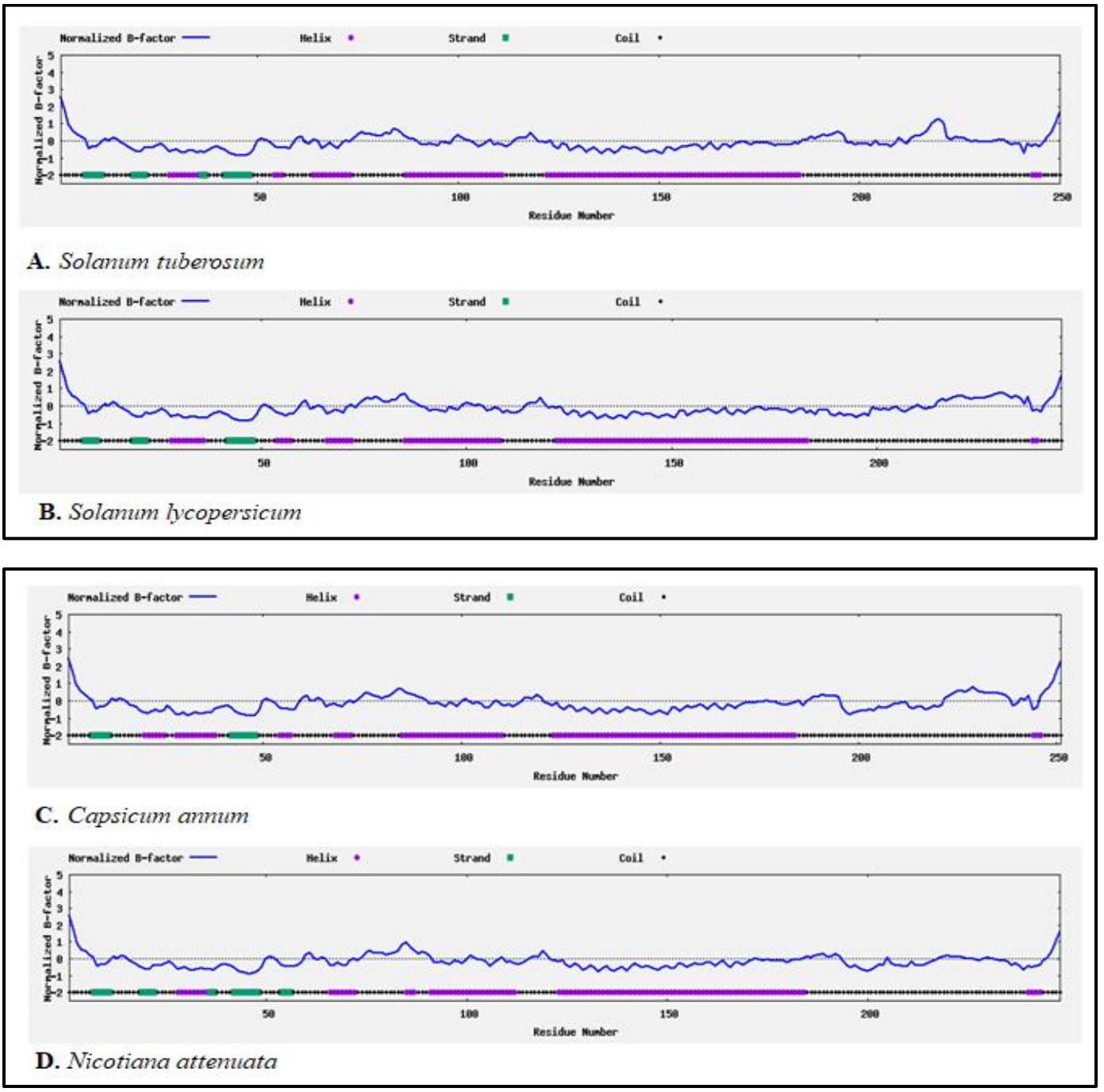

Fig. 4.14 B-factor profile for predicted models. (1) *Solanum tuberosum,* (2) *Solanum lycopersicum,* (3) *Capsicum annuum,* (4) *Nicotiana attenuata*

- ***X-ray/NMR structure***

The ProSA-web programme was used to validate the protein model. The structure of proteins in the form of pdb file was uploaded on the server and the results were obtained as shown in the **Fig. 4.15.** The z-score of the protein in query can be compared to the range of scores commonly observed for the proteins of same size. In the figure, the black colored dot (**.**) in the blue region represents the protein.

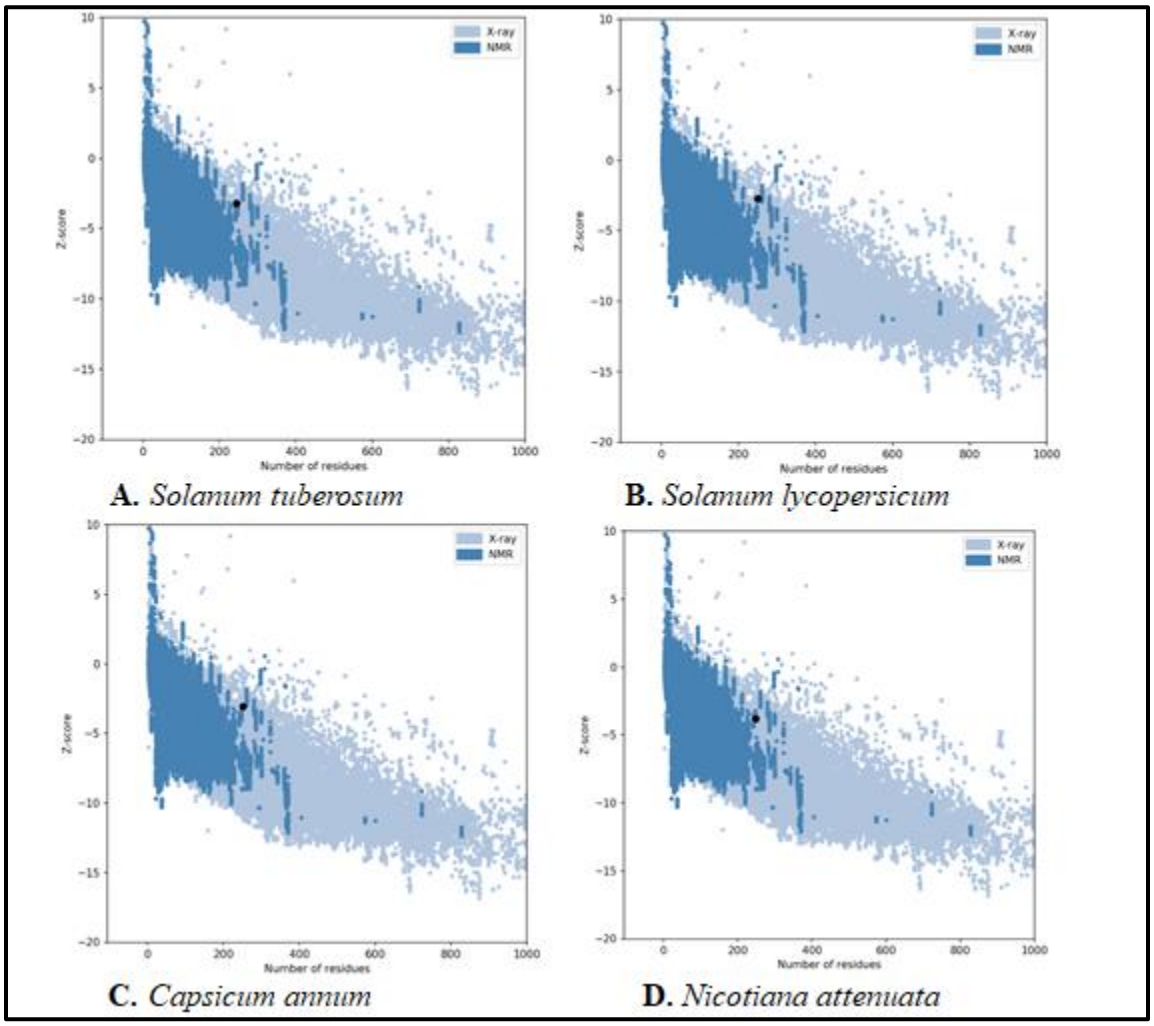

Fig. 4.15 X-ray/NMR structure of predicted 3-D models of proteins.

- ***Ramachandran Plot***

The Ramachandran principle is used to measure angles in amino acids (proteins) as well as in other kind of molecules. This plot helps to determine the secondary structure of proteins. The conformation of the backbone of any peptide is governed by the values of two torsional angles

which are Psi (ψ) and Phi (φ) angles. The protein sequence was uploaded on the PROCHECK server and the Ramachandran plots as shown in **Fig. 4.16** were obtained. The regions that are shown in red color i.e., A, B and L contain most favored residues. Residues in additional and generously allowed regions are also shown.

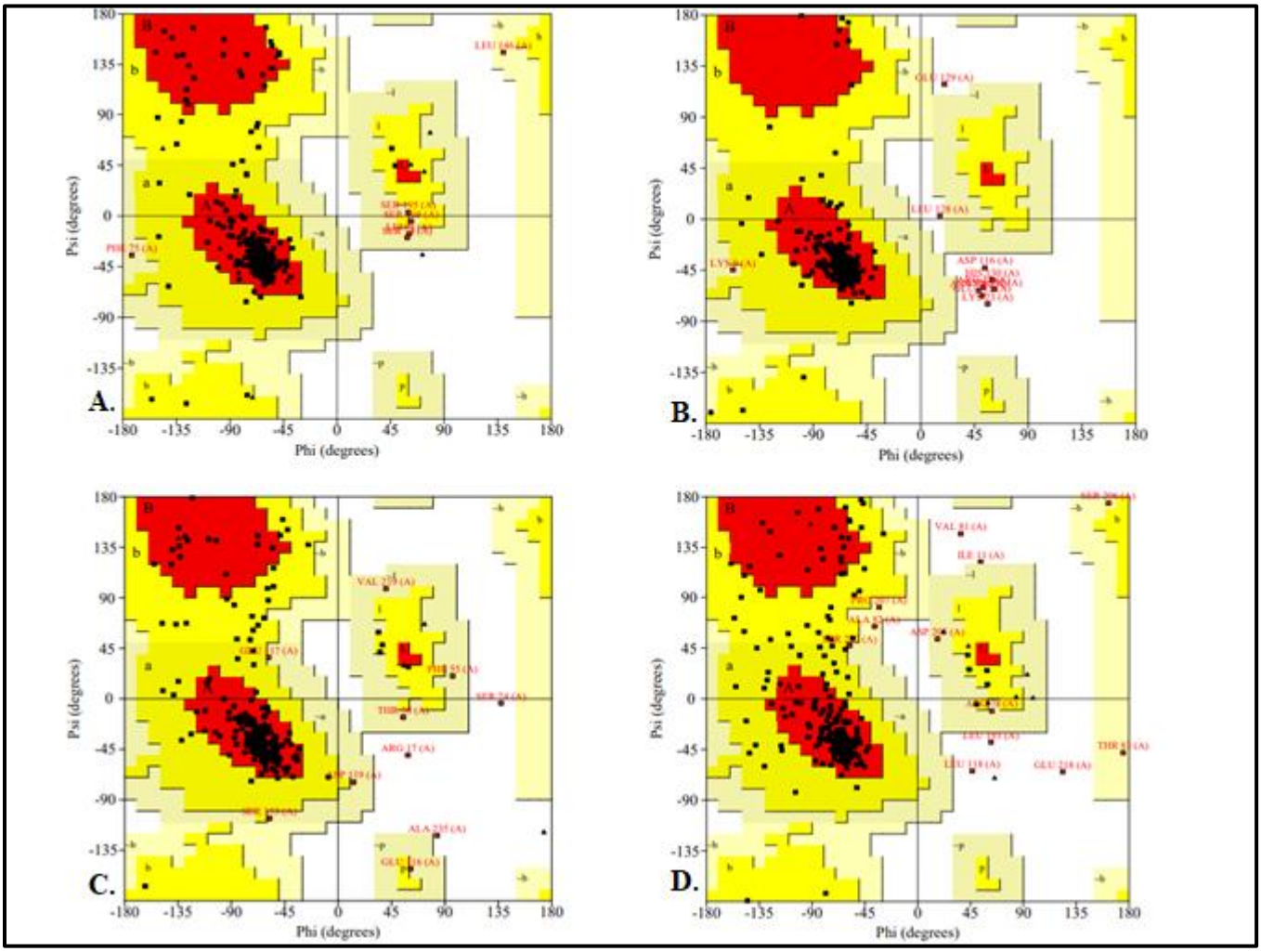

Fig. 4.16 Ramachandran plot of MADS-box proteins indifferent plant species of *Solanaceae* family **A.** *Solanum tuberosum,* **B.** *Solanum lycopersicum,* **C.** *Capsicum annum* and **D.** *Nicotiana attenuata*

Table 4.6 shows the statistics of the plot for different plant species. The % of various amino acids in different regions is given in the table. It can be observed that in case of all the four predicted protein structures majority of the amino acids are present in the most favored regions and additional allowed regions. This helps to validate the 3-D protein structures. These statistic values help in the better understanding of the predicted 3-D protein structures.

Table 4.6 Ramachandran plot statistics of 4 different plant species of *Solanaceae* family

	Number of Residues			
	Solanum tuberosum	***Solanum lycopersicum***	***Capsicum annum***	***Nicotiana attenuata***
In most favoured regions [A,B,L]	194 (84.0%)	127 (86.8%)	183 (78.9%)	146 (62.9%)
In additional allowed regions [a,b,l,p]	31 (13.4%)	20 (8.8%)	39 (16.8%)	57 (24.6%)
In generously allowed regions [~a,~b,~l,~p]	6 (2.6%)	1 (0.4%)	7 (3.0%)	22 (9.5%)
In disallowed regions	0 (0.0%)	9 (4.0%)	3 (1.3%)	7 (3.0%)
Non-glycine and non-proline residues	231 (100%)	227 (100%)	232 (100%)	232 (100%)
End-residues	1	2	1	1
Glycine residues	13	22	12	11
Proline residues	5	5	6	4
Total residues	250	245	251	248

CONCLUDING REMARKS

- Potato is major non-grain food crop as it has high nutritional value.
- Understanding the signaling pathways associated with the growth, development, and most importantly tuberization is currently one of the major goal to improve the desirable traits in potato.
- MADS box gene family plays important role in transcriptional regulation during tuberization in potato. Literature survey revealed that considerable progress has been made to understand the mechanism of action during tuber induction and development.
- Till the date 153 MADS- box genes has been identified in potato but there is no report on isolation and characterization of MADS box genes from Indian potato cultivars.
- Keeping its crucial role in mind, present study focused on cloning and characterization of *POTM1*, a MADS box gene, designated as *POTM1-1A* from Indian potato cultivar KC-1.
- Sequence analysis revealed that our sequence displayed change at 6 positions at nucleotide level from the *POTM1-1* sequence available in the database. But there was only 1 change of amino acid which resulted in conservative mutation at protein level.
- The molecular weight of the predicted protein was found to be 28.92 kDa having isoelectric point of 8.77.
- 3-D structure was predicted using I-TASSER software available online. Further Ramachandaran plot analyses revealed that more amino acids lied in most favourable allowed region which validates the predicted 3D-structures.
- To find out the phylogenetic relationship with other *Solanaceae* family members multiple sequence alignment was performed that revealed species specific insertion and deletion.
- To conclude, our study revealed pivotal features of potato *POTM1-1* gene, which would help in understanding the transcriptional regulation of this gene during tuber development. Further, phylogenetic tree formation can help in studying the phylogenetic relationship with other plant species as well. Hence, this thesis work will be useful in crop improvement.

REFERENCES

Abelenda JA, Navarro C, Prat S (2011) From the model to the crop: genes controlling tuber formation in potato. Curr Opin Biotechnol 22:287–292

Agyare C, Obiri DD, Boakyea YD, Osafob N (2013) 19 - Anti-Inflammatory and Analgesic Activities of African Medicinal Plants. J Med Plant Res 725-752

Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, De Pouplana LR Martinez-Castilla L, Yanofsky MF (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc Natl Acad Sci USA 97:5328–5333

Aswath CR, Kim SH, Mo SY, Kim DH (2005) Transgenic plants of creeping bent grass harboring the stress inducible Gene,*9-cis-epoxycarotenoid dioxygenase*, are highly tolerant to drought and NaCl stress. Plant Growth Regul 47:129–139

Barone A, Chiusano ML, Ercolano, MR, Giuliano G, Grandillo S, Frusciante L (2008) Structural and Functional Genomics of Tomato. Int J Plant Genomics 1–12

Beals KA (2019) Potatoes, Nutrition and Health. Am J Potato Res 96:102–110

Becker A, Theißen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol 29:464–489

Biirglin TR (1997) Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res 25:4173–4180

Bindler G, Plieske J, Bakaher N, Gunduz I, Ivanov N, Hoeven RV, Ganal M, Donini P (2011) A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development. Theor Appl Genet 123:219–230

Brown CR, Culley D, Yang CP, Durst R, Wrolstad R (2005) Variation of anthocyanin and carotenoid contents and associated antioxidant values in potato breeding lines. J Am Soc Hortic Sci 130:174–180

Cao Y, Wang L, Zhao J, Zhang H, Tian Y, Liang H, Ma Q (2016) Serum response factor protects retinal ganglion cells against high-glucose damage. J Mol Neurosci 59:232–240

Carmona MJ, Ortega N, Garcia-Maroto F (1998) Isolation and molecular characterization of a new vegetative MADS-box gene from *Solanum tuberosum* L. Planta 207:181–188

Chapman HW (1958) Tuberization in potato plant. Physiol Plant 11:215–224

Chen H, Banerjee AK, Hannapel DJ (2004) The tandem complex of BEL and KNOX partners is required for transcriptional repression of *ga20ox1*. Plant J 38:276–284

Chen H, Rosin FM, Prat S, Hannapel DJ (2003) Interacting transcription factors from the TALE superclass regulate tuber formation. Plant Physiol 132:1391–1404

Chen R, Ma J, Luo D, Hou X, Ma F, Zhang Y, Meng Y, Zhang H, Guo W (2019) CaMADS, a MADS-box transcription factor from pepper, plays an important role in the response to cold, salt, and osmotic stress. Plant Sci 280:164–174

Cutter EG (1978) Structure and development of the potato plant. In: Harris, P.M. (eds) The Potato Crop. Springer, Boston, pp 70-152

Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ (2009) Genome-wide analysis of MIKCC-type MADS box genes in grapevine. Plant Physiol 149:354-369

Dutt S, Manjul AS, Raigond P, Singh B, Siddappa S, Bhardwaj V, Kawar PG, Patil VU, Kardile B (2017) Key players associated with tuberization in potato: potential candidates for genetic engineering. Crit Rev Biotechnol 37:942–957

Ewing EE (1995) The role of hormones in Potato (*Solanum Tuberosum* L.) tuberization. Plant Hormones 698–724

Ewing EE, Struik PC (1992) Tuber formation in potato: induction, initiation, and growth. Am Soc Hortic Sci 14:89–198

Friedman M (2006) Potato glycoalkaloids and metabolites: roles in the plant and in the diet. J Agric Food Chem 54:8655–8681

Gális I, Macas J, Vlasák J, Ondřej M, Onckelen HAV (1995) The effect of an elevated cytokinin level using the ipt gene and N 6-benzyladenine on single node and intact potato plant tuberization *in vitro* . J Plant Growth Regul 14:143–150

Gao H, Wang Z, Li S, Hou M, Zhou Y, Zhao Y, Li G, Zhao H, Ma H (2018) Genome-wide survey of potato MADS-box genes reveals that StMADS1 and StMADS13 are putative downstream targets of tuberigen StSP6A. BMC Genom 19:726

García-Maroto F, Ortega N, Lozano R, Carmona MJ (2000) Characterization of the potato MADS-box gene STMADS16 and expression analysis in tobacco transgenic plants. Plant Mol Biol 42:499–513

Gregory LE (1965) Physiology of tuberization in plants. In: Ruhland W (ed) Encyclopedia of Plant Physiology vol 15. Springer, Berlin, pp 1328-1354

Griffith S, Dunford RP, Coupland G, Laurie AD (2003) The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis*. Plant Physiol 131:1855–1867

Groamzw L, Ritz MS, Theißen G (2010) On the origin of MADS-domain transcription factors. Trends Genet 26:149–153

Groot SPC, Bouwer R, Busscher M, Lindhout P, Dons HJ (1995) Increase in endogenous zeatin riboside by introduction of the *ipt* gene in wild type and the *lateral suppressor* mutant of tomato. Plant Growth Regul 16:27–36

Guo J, Shi XX, Zhang JS, Duan YH, Bai PF, Guan XN, Kang ZS (2013) A type I MADS-box gene is differentially expressed in wheat in response to infection by the stripe rust fungus. Biol Plant 57:540–546

Hammes PS, Nel PC (1975) Control mechanisms in the tuberization process. Potato Res 18:2562–2574

Hannapal DJ (2010) A Model System of Development Regulated by the Long-distance Transport of mRNA. J Integr Plant Biol 52:40–52

Hannapal DJ, Chen H, Rosin FM, Banerjee AK, Davies PJ (2004) Molecular controls of tuberization. Am J Potato Res 81: 263–274

Hart JK (2002) In situ hybridization of the MADS-box gene *POTM1* during potato floral development. J Exp Bot 53:465–471

Hart JK, Hannapel DJ (2002) In situ hybridization of the MADS-box gene *POTM1* during potato floral development. J Exp Bot 53:465–471

Harvey M, Quilley S, Beynon H (2002) Exploring the tomato: transformations in nature, economy and society. Edward Elgar Publishing Ltd.

Hawkes JG (1992) History of the potato. In: Harris, P.M. (eds) The Potato Crop. World Crop Series. Springer, Dordrecht, pp 1–12

Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theißen G (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. Mol Biol Evol 19:801–814

Inui H, Ogura Y, Kiyosue T (2010) Overexpression of *Arabidopsis thaliana* LOV KELCH REPEAT PROTEIN 2 promotes tuberization in potato (*Solanum tuberosum* cv. May Queen). FEBS Lett 584:2393–2396

Jackson SD (1999) Multiple signaling pathways control tuber induction in potato. Plant Physiol119:1–8

Jang S, An K, Lee S, An G (2002) Characterization of Tobacco MADS-box Genes Involved in Floral Initiation. Plant Cell Physiol 43:230–238

Jena PK, Reddy ASN, Poovaiah BW (1989) Molecular cloning and sequencing of a cDNA for plant calmodulin: signal-induced changes in the expression of calmodulin. Proc Natl Acad Sci USA 86:3644–3648

Kang SG, Hannapel DJ (1996) A novel MADS-box gene of potato (*Solanum tuberosum* L.) expressed during the early stages of tuberization. Plant Mol Biol 31:379–386

Kang SG, Hannapel DJ, Suh SG (2003) Potato MADS-box gene *POTM1-1* transcripts are temporally and spatially distributed in floral organs and vegetative meristems. Mol Cells 15:48–54

Kinet JM (1977) Effect of light conditions on the development of the inflorescence in tomato. Sci Hortic 6:15–26

Kishore K. (2014) Monograph of tobacco (*Nicotiana tabacum*). Indian j drugs dermatol 2:5–23

Knapp S (2002) Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the *Solanaceae*. J Exp Bot 53:2001–2022

Knapp S, Peralta IE (2016) The Tomato (*Solanum lycopersicum* L., *Solanaceae*) and Its Botanical Relatives. In: Causse, M., Giovannoni, J., Bouzayen, M., Zouine, M. (eds) The Tomato Genome. Compendium of Plant Genomes. Springer, Berlin, pp 7-21

Kramer EM, Su HJ, Wu CC, Hu JM (2006) A simplified explanation for the frameshift mutation that created a novel C-terminal motif in the *APETALA3* gene lineage. BMC Evol Biol 6:30

Krauss A (1985) Interaction of nitrogen nutrition, phytohormones, and tuberization. In: Li PH (ed) Potato physiology. Academic Press, Inc. Orlando, pp 209–230

Kumar D, Wareing PF (1972) Factors controlling stolon development in the potato plant. New Phytol 71:639–648

Latchman DS (1997) Transcription factors: An overview. Int J Biochem Cell Biol 29:1305–1312

Leshem B, Clowes FAL (1972) Rates of Mitosis in Shoot Apices of Potatoes at the Beginning and End of Dormancy. Ann Bot 36: 687–691

Li Y, Hagen G, Guilfoyle TJ 1992 Altered morphology in transgenic tobacco plants that overproduce cytokinins in specific tissues and organs. Dev Biol 153:386–395

Liu RH (2013) Health-promoting components of fruits and vegetables in the diet. Adv Nutr 4:384–392

Ma H, Yanofsky MF, Meyerowitz EM (1991) AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev 5:484–495

Machácková I, Sergeeva L, Ondrej M, Zaltsman O, Konstantinova T, Eder J, Ovesná J, Golyanovskaya S, Rakitin Y, Aksenova N (1997) Growth pattern, tuber formation and hormonal balance in in vitro potato plants carrying ipt gene. Plant Growth Regul 21:27–36

Marschner H, Sattelmacher B, Bangerth F (1984) Growth rate of potato tubers and endogenous contents of indolylacetic acid and abscisic acid. Physiol Plant 60:16–20

Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM (2011) The emerging importance of Type I MADS Box Transcription Factors for Plant Reproduction. The Plant Cell 23:865–872

Masiero S, Imbriano C, Ravasio F, Favaro R, Pelucchi N, Gorla MS, Mantovani R, Colombo L, Kater MM (2002) Ternary complex formation between MADS-box transcription factors and the histone fold protein NF-YB. J Biol Chem 277:26429–26435

McGill CR, Kurilich AC, Davignon J (2013) The role of potatoes and potato components in cardiometabolic health: a review. Ann Med 45:467–473

Muñoz NC, Herrera J, Sánchez WC, Arrizubieta M, Trejo C, Ponce BG, de la Paz Sánchez M, ÁlvarezBuylla ER, Arroyo AG (2019) MADS-box genes are key components of genetic regulatory networks involved in abiotic stress and plastic developmental responses in plants. Front Plant Sci 10

Nadal Ed, Casadome L, Posas F (2003) Targeting the MEF2-like transcription factor Smp1 by the stress-activated Hog1 mitogen-activated protein kinase. Mol Cell Biol 23:229–237

Ng M, Yanofsky MF (2001) Function and evolution of the plant MADS-box gene family. Nat Rev Genet 2:186–195

Norman C, Runswick M, Pollock R, Treisman R (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. Cell 55:989–1003

Oparka KJ, Davies HV (1985) Translocation of Assimilates Within and Between Potato Stems. Ann Bot 56:45–54

Parenicova L, Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. Plant Cell 15:1538–1551

Passmore S, Maine GT, Elble R, Christ C, Tye BK (1988) *Saccharomyces cerevisiae* protein involved in plasmid maintenance is necessary for mating of MATα cells. J Mol Biol 204:593–606

Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80:847–857

Reddy BJ, Mandal R, Chakroborty M, Hijam L, Dutta P (2018) A review on potato (*Solanum tuberosum* L.) and its genetic diversity. Genetics 10:360–364

Reynolds MP, Ewing EE (1989) Heat tolerance in tuber bearing*Solanum* species: A protocol for screening. Am Potato J 66:63–74

Romanov GA, Aksenova NP, Konstantinova TN, Golyanovskaya SA, Kossmann J, Willmitzer L (2000) Effect of Indole3Acetic Acid and Kinetin on Tuberization Parameters of Different Cultivars and Transgenic Lines of Potato *In vitro*. Plant Growth Regul 32:245–251

Rosin FM, Hart JK, Horner HT, Davies PJ, Hannapel DJ (2003a) Overexpression of a Knotted-Like Homeobox gene of potato alters vegetative development by decreasing gibberellin accumulation. Plant Physiol 132:106–117

Rosin FM, Hart JK, Onckelen HV, Hannapel DJ (2003b) Supression of a vegetative MADS-box gene of potato activates axillary meristem development. Plant Physiol 131:1613–1622

Roumeliotis E, Kloosterman B, Oortwijn M, Kohlen W, Bouwmeester HJ, Visser RGF, Bachem CWB (2012) The effects of auxin and strigolactones on tuber initiation and stolon architecture in potato. J Exp Bot 63:4539–4547

Sarkar D (2008) The signal transduction pathway controlling in planta tuberization in potato: an emerging synthesis. Plant Cell Rep27:1–8

Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H (1990) Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science 250:931–936

Shore P, Sharrocks AD (1995) The MADS-box family of transcription factors. Eur J Biochem 229:1–13

Sims WL (1979) History of tomato production for industry around the world. Acta Hortic 100:20–26

Singh AP, Singh RS, Singh SRP, Pal M, Singh R, Kumari R (2021) Screening of genotypes against major biotic stresses in chilli (*Capsicum annum* L.). J Pharm Innov 10:502–505

Singh BP, Rana RK (2014) History of Potato and its Emerging Problems in India.In : Souvenir- National Seminar on Emerging Problems of Potato pp 7-21

Singh DK, Raigond P, Kharumnuid P (2020) Potatoes: The Food for Nutritional Security. Agric &Food :e-Newsl 2:498–501

Smaczniak C, Immink RG, Angenent GC, Kaufmann K (2012) Developmental and evolutionary diversity of plant MADS-domain factors: Insights from recent studies. Development 139:3081–3098

Smith OE, Rappaport L (1969) Gibberellins, inhibitors, and tuber formation in the potato (*Solanum tuberosum*). Am Potato J 185

Sommer H, Beltran JP, Huijser P, Pape H, Lo¨nnig WE, Saedler H, Schwarz-Sommer Z (1990*) Deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: The protein shows homology to transcription factors. EMBO J 9:605–613

Struik PC (2007) Above-Ground and Below-Ground Plant Development.Plant Biotechnol J 219–236

Sung SK, Moon YH, Chung JE, Lee SY, Park HG, An G (2001) Characterization of MADS Box Genes from Hot Pepper. Mol Cells 11:352–359

Taylor MA, Arif SA, Pearce SR, Davies HV, Kumar A, George LA (1992) Differential expression and sequence analysis of ribosomal protein genes induced in stolon tips of potato (*Solanum tuberosum* L.) during the early stages of tuberization. Plant Physiol 100:1171–1176

van der Graaff EE, Auer CA, Hooykaas PJJ (2001) Altered development of *Arabidopsis thaliana* carrying the *Agrobacterium tumefaciens* ipt gene is partially due to ethylene effects. Plant Growth Regul 34:305–315

Vengaiaha PC, Pandey JP (2007) Dehydration kinetics of sweet pepper (*Capsicum annum* L.). J Food Eng 81:282–286

Viola R, Roberts AG, Haupt S, Gazzani S, Hancock RD, Marmiroli N, Machray GC, Oparkab J (2001) Tuberization in Potato Involves a Switch from Apoplastic to Symplastic Phloem Unloading. Plant Cell 13:385–398

Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J (2002) A MADS-box gene necessary for fruit ripening at the tomato *Ripening-Inhibitor* (*Rin*) locus. Science 296:343–346

Vreugdenhil D (1999) Initial Anatomical Changes Associated with Tuber Formation on Single-node Potato (*Solanum tuberosum* L.) Cuttings: A Re-evaluation. Ann Bot 84:675–680

Wang Y, Zhang J, Hu Z, Guo X, Tian S, Chen G (2019) Genome-Wide Analysis of the MADS-Box Transcription Factor Family in *Solanum lycopersicum*. Int J Mol Sci 20:2961

Woolfe JA (1987) The potato in the human diet. Cambridge University Press, Cambridge, 231

Xu X, Lammeren AAMV, Vermeer E, Vreugdenhil D (1998) The Role of Gibberellin, Abscisic Acid, and Sucrose in the Regulation of Potato Tuber Formation in Vitro. Plant Physiol 117:575–584

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. Nature 346:35–39

Zaheer K, Akhtar MH (2016) Potato Production, Usage, and Nutrition--A Review. Crit Rev Food Sci Nutr 56:711–721

Zhu C, Perry SE (2005) Control of expression and autoregulation of AGL15, a member of the MADS-box family. Plant J 41:583–594

Curiginal

Document Information

Analyzed document	MSc Dissertation_Ravneet Kaur_Jun 2022.docx (D141431912)
Submitted	6/29/2022 7:57:00 AM
Submitted by	Niranjan Das
Submitter email	ndas@thapar.edu
Similarity	7%
Analysis address	ndas.thapar@analysis.urkund.com

Sources included in the report

SA	**Thapar Institute Of Engineering And Technology / Plagiary Checking_Simran Shrimal.docx** Document Plagiary Checking_Simran Shrimal.docx (D111873774) Submitted by: ndas@thapar.edu Receiver: ndas.thapar@analysis.urkund.com	10
W	URL: https://www.frontiersin.org/articles/10.3389/fgene.2021.750137/full Fetched: 6/29/2022 7:57:23 AM	1
W	URL: http://www.molcells.org/journal/download_pdf.php?spage=48&volume=15&number=1 Fetched: 6/29/2022 7:57:47 AM	1
W	URL: https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/mads-box Fetched: 4/13/2020 12:23:44 AM	1
W	URL: https://pubmed.ncbi.nlm.nih.gov/12692320/ Fetched: 6/29/2022 7:57:21 AM	2
W	URL: https://link.springer.com/article/10.1007/BF00021798 Fetched: 11/10/2021 3:39:31 PM	1
W	URL: https://www.researchgate.net/publication/240359647_Characterization_of_potato_vegetative_MADS-Box_gene_POTM1-1_in_response_to_hormone_applications Fetched: 7/10/2020 6:55:11 AM	3
SA	**Sardar V. B. Patel Uni Of Agriculture & Technology / kratika 4797 full thesis.pdf** Document kratika 4797 full thesis.pdf (D114194621) Submitted by: headdbtcoa@gmail.com Receiver: headdbtcoa.svbpu@analysis.urkund.com	2
W	URL: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-5113-z Fetched: 11/15/2019 5:37:39 AM	6
W	URL: https://www.researchgate.net/figure/The-detailed-information-of-potato-MADS-box-gene-family-Genes-with-an-asterisk-tail_tbl1_328058704 Fetched: 7/10/2020 6:55:46 AM	1
W	URL: https://pubmed.ncbi.nlm.nih.gov/12661760/ Fetched: 6/29/2022 7:57:17 AM	1

Ravneet Kaur.
29/6/2022

29/06/2022