In silico analysis of the core promoters and other regulatory sequence motifs of the patatin genes in potato (*Solanum tuberosum L.*)

A Dissertation Report
Submitted in partial fulfilment of the requirement
For the award of degree of

Masters of Science

In

Biotechnology

Under the guidance of

Dr. N. Das
Professor

Dr. Vikas Handa
Assistant Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Submitted by

Gurparkash Singh Thind
Roll no. 301601009

DEPARTMENT OF BIOTECHNOLOGY
Thapar Institute of Engineering & Technology
PATIALA
June 2018

CANDIDATE'S DECLARATION

I hereby declare that the work being presented in the M.Sc. Dissertation entitled **"*In silico* analysis of the core promoters and other regulatory sequence motifs of the patatin genes in potato (*Solanum tuberosum L.*)"** has been carried out by me during the period of Jan 2018 to June 2018, under the guidance of Dr. Niranjan Das, Professor and Dr. Vikas Handa, Assistant Professor, Department of Biotechnology, Thapar Institute of Engineering & Technology, Patiala. Further, I declare that I have not submitted the matter embodied in this dissertation for the award of any degree or any other qualification of any university or examining body in India/elsewhere.

Gurparkash Singh

Date: 28 June 2018
Place: Patiala

Gurparkash Singh Thind
M.Sc. Biotechnology
Roll no. 301601009

CERTIFICATE

This is to certify that dissertation entitled **"*In silico* analysis of the core promoters and other regulatory sequence motifs of the patatin gene in potato (*Solanum tuberosum* L.)"** submitted by **Gurparkash Singh Thind (301601009)** in partial fulfillment of the requirements for the award of Master in Science in the Department of Biotechnology, Thapar Institute of Engineering & Technology Patiala is the record of candidate's own independent original work carried out by him under my supervision and guidance. The matter embodied in this thesis has not been submitted in part of full to any other university or institute for the award of any degree.

Dr. N. Das
Professor
DBT,TIET
Patiala, Punjab.

Gurparkash Singh
Gurparkash Singh Thind
M.Sc Biotechnology
301601009
TIET, Patiala

Dr. Vikas Handa
Assistant Professor
DBT, TIET
Patiala, Punjab

ACKNOWLEDGEMENT

I would like to acknowledge my sincerest debt of gratitude to the ***ALMIGHTY GOD*** *who always supports me in my endeavors. I would also like to thank* ***Head of the Department, Dr. Moushmi Ghosh*** *and my guide* ***Dr. Niranjan Das, Professor*** *and* ***Dr. Vikas Handa, Assistant Professor, Department of Biotechnology, Thapar University, Patiala*** *for their valuable support, constant encouragement and guidance and faculty members of the department for their advices and help to complete my research work.*

On the **home front**, *I would like to thank my family members, colleagues for their immense help, valuable suggestions and necessary guidance. I would also like to thank my lab-mates,* ***Ms. Kirti Pal*** *and* ***Mr. Vardhan Chabra*** *for their support and help.*

Gurparkash Singh

Dated: 28 June 2018

Gurparkash Singh

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviations	Full Forms
A	Adenine
ARE	Adenine Rich Elements
C	Cytosine
CaMV	Cauliflower Mosaic Virus
CDS	Coding DNA Sequence
DNA	Deoxyribose Nucleic Acid
DPE	Downstream Promoter Element
G	Guanine
IR	Inverted Repeats
LAH	Lipid Acyl Hydrolase
MSA	Multiple Sequence Alignment
NCBI	National Center for Biotechnology Information
RNA	Ribonucleic Acid
T	Thymine
TFBS	Transcription Factor Binding Site
TSS	Transcription Start Site
UTR	Untranslated Region

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Patatin, a family of glycoproteins, constituting about 45% of the total soluble proteins found in tubers of potato (*Solanum tuberosum L.*) is considered to be encoded by a large multigene family of two classes. To study the regulation of plant gene expression, Class I patatin gene promoters are meant as excellent molecular tools to study the regulation of transcription. DNA sequence characteristics such as the presence of *cis*-elements/regulatory motifs are known to be present in promoter/regulatory sequences. Therefore, *in silico* analysis of the promoters and other regulatory sequence motifs of the isoforms of patatin gene was done to search for novel putative elements based on sequence conservation by multiple sequence alignment and other additional characteristics such as presence of palindromic sequences. Presumably, results of such analysis can explain or predict the differential expression patterns of the patatin isoforms at holistic level in potato.

Keywords- Potato (*Solanum tuberosum L.*), Patatin genes, Sequence comparison analysis, Predicted *cis* regulatory motifs, Palindromes

CHAPTER 1 INTRODUCTION

Potato (*Solanum tuberosum L.*), is one of the major tuber crops that is grown in moderate regions. It is essential tuber crop regarding production as it accounts for majority (45%) of crops of tuber that are grown all over the world. Tubers of potato are actually obtained from swollen stems, though they are usually subterranean. In 1986, Campbell and Osborne described that the most important protein in tubers of potato known to be globulin and was designated as 'tuberin'. In 1980, it was investigated by Foote and Racusen that relative molecular mass of glycoprotein is considered to be around 45 000 Da that possessed 20 % of the total number of proteins of tuber in potato and the substitutive name was proposed to be 'patatin', based on 'patata' being the original word for potato.

The main protein of tuber in potato i.e. patatin, consist of group of glycoproteins of 40 kDa that represented about 45 % of the proteins in tubers, but is normally unnoticeable in other tissues. In potato, formation of tubers is a very complicated process of development and is governed by a number of factors which may be environmental, genetic or biochemical in its development (Aksenova *et al*., 2012). Patatin, consisting of a glycoprotein of ∼40 kDa, is driven by a family of multigenes and is the most abound protein of tuber. The isoforms of patatin are immunologically not able to be distinguished and are restricted principally in the vacuolar cells of plant (Foote and Racusen 1980; Park *et al.*, 1983). Patatin also contains special property of both wax synthase and lipid acyl hydrolase activity which is not present in other proteins of storage (Galliard and Dennis et al., 1985). Patatin in large number may possibly assemble in stems and petioles if stolons or tubers are removed (Paiva *et al.*, 1984).

The patatin family of multigenes mainly contain copies which are in the range of 10 to 18 per haploid genome, which depends on the genotype of potato and is further distributed into two classes, I and II (Twell and Ooms 1988; Pikaard *et al.*, 1987). The presence of a 22-bp insertion in the 5′-UTR of class II members is the main feature which distinguish the two classes, significant divergence of sequence which is upstream of the −87 position along with differential patterns of expression in disparate tissues is another difference. The 5′-flanking regions that appear within the same class are said to be well conserved with the exception of some deletions or insertions to a certain extent than point mutations which are cumulative (Liu *et al*. 1991). Class I patatin gene expression take place essentially in tubers (Sosa and Rocha *et al.,* 1989), whereas only certain types of roots and tubers are expressed by class II members (Töpfer and Koster *et al.,* 1989) but its level is very low as compared to the members of class I. Mapping of all the patatin genes has been done with chromosome 8 and are organized in the genome of potato as a single cluster (Ganal *et al*., 1991).

A number of reports have been published on functionality of the 5′-flanking regions which are based on sequence analysis of genomic or cDNA patatin clones. A small number of promoter-GUS fusions of class I patatin gene were detected to be mainly tuber inducible. Since, they were not strictly specific to tuber, as a result, the GUS gene expression may possibly be highly induced to a great extent in cultured explants of stems and leaves by enhancing the concentration of sucrose (Liu and Wenzler *et al*., 1991). Thus, 5′-flanking regions are considered to be excellent tools of research as they are expressed significantly to a large extent in the tubers and are highly sucrose-inducible (Park and Bevan et al., 1993). These promoter activities were observed to be more well-known in Arabidopsis cotyledons under elevated level of concentration of sucrose level (Naumkina *et al*., 2008). Promoters of class I patatin gene are considered to be highly efficient in regulating and expressing genes which are desired in potato and other plants (Zhu *et al*., 2007). Identification of number of sequence motifs of cis-regulatory elements and their related factors which are trans-acting have already been done (Grierson and Jefferson *et al*., 1994).

There is possiblility that the diverse members of the patatin family of multigenes have different types of *cis*-regulatory sequence motifs which may be *cis*-regulatory and typical features of sequence that may result in expressing gene with different functionality in different types of tissues. A small number of members are there which have been analyzed till now. Number of useful patatin genes which belongs to any of the classes are so far needed to be isolated first and their characterization is necessary. Different sequence motifs of *cis*-regulatory were found in 5′-flanking regions. In different tissues of potato, the promoter activities were studied under field conditions using the constructs of GUS gene. The promoters that were under research were also confirmed to be induced by high exogenous concentrations of sucrose.

1.1 Promoter

A promoter is a regulatory region of DNA in a gene that aids in activating the process of transcription Promoters are located upstream towards the 5’ region of a gene near the start sites of transcriptional region of genes. A typical promoter length in base pairs is approximately 100-1000 base pairs.

The promoter contains sequences of DNA which are specific and recognizable by proteins and are known as factors of transcription. These factors process by binding to the sequences of promoter, recruits RNA polymerase enzyme, which is meant to synthesize the RNA from the region which codes for the gene.

Promoter Elements

(i) Core promoter – It is the least portion of the promoter which is required to commence the process of transcription

- Transcription Start Site (TSS)
- RNA polymerase site for binding
- Binding sites for transcription factor

(ii) Proximal promoter – The upstream gene sequence that have primary regulatory elements containing particular binding sites for transcription factor is known as proximal promoter

1.1.1 Function of promoter

- It has RNA polymerase binding site
- It initiates transcription
- It controls the expression of genes

1.2 Gene Regulation: Potato (*Solanum tuberosum*) contains tubers as important storage organs which have two classes of rich proteins, 40 kDa Patatin proteins and 20 kDa inhibitors of proteinase (Foote and Park, 1983; Stiekema *et al.*, 1989). As part of defence mechanism in potato, the patatin protein which consists of lipid acyl hydrolase, releases fatty acids. Now, as Patatin comprises of two classes of genes: Class 1 patatin genes are highly expressed in tubers whereas Class 2 are expressed at very low levels in roots and peripheral cells of the tuber as compared to Class 1 patain genes.

For instance, Ipomoea batatas, a sweet potato, has several abundant proteins such as sporamin and J-amylase which are deposited in the storage root (Nakamura *et al*., 1991). It is explained that these somatic proteins of storage accumulated due to increase in constant levels of their mRNAs (Paiva and Hattori, 1983).

By increasing exogenous sucrose concentration, the constant levels of mRNA which encodes for sporamin, patatin, inhibitor for proteinase and amylase increases in large amount in tissues which are non storing such as stems and leaves. Finally, It has also been found that sucrose suppress the transcription of genes that encodes enzymes of photosynthetic fixation of carbon in maize (Sheen, 1990). These results specify significant role for sucrose levels and its metabolites in integrating and governing the activity of different sets of genes having function of storage.

Sucrose plays very essential role in influencing expression of genes. By analyzing the patatin promoter function, it would assist in describing that sequences of *cis*-elements are responsible for transcriptional regulation (Humphreys, 1988).

2.1 General features of eukaryotic gene promoters

Gene promoters that are generally found in upstream of gene are allowed to begin transcription from presence of RNA polymerase II binding sites. This enzyme is involved during sequence binding of specific proteins, called transcription factors thus resulting in initiation of transcription (Russell 1996; Porto *et al.*, 2014)

Eukaryotic gene consist of three types of RNA Polymerases

1. RNA polymerase I transcribes ribosomal RNA genes
2. RNA polymerase II transcribes messenger RNA genes
3. RNA polymerase III transcribes transfer RNA genes

The structure of eukaryotic gene promoters which includes plant promoters, can be further divided as- core promoter and distal promoter region consisting of silencers and enhancers as shown in Fig. 1. Both of them contain regions of *cis*-elements which bind to trans factor proteins (Cummings and Klug 2003; Porto *et al.*, 2014)

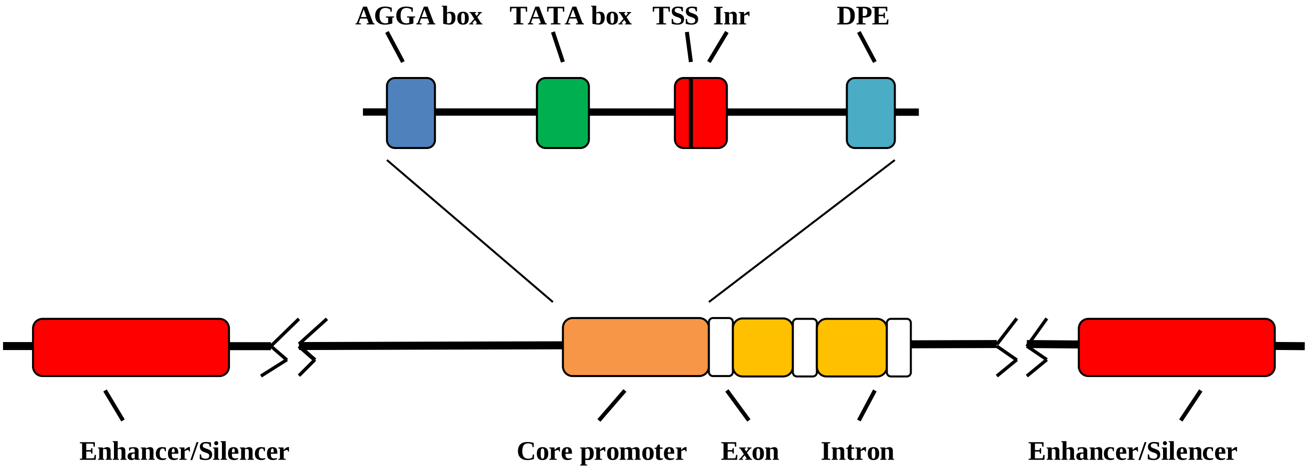


Fig. 1 Model of plant gene organization. In this figure, it may not be necessary that each and every element which is depicted in this figure is universal. TATA box, AGGA box, Inr and DPE are parts of Core promoter. Transcription start site is TSS, Downstream promoter element is shown as DPE and Inr is the initiator This figure is based on reference of Cummings and Klug (2003).

2.2 *Cis* regulatory elements

Cis regulatory element is a DNA sequence of non-coding DNA which is present within a gene generally required for the expression of the gene. Various *cis*-regulatory sequences such as promoters and enhancers play very important role in controlling and regulating gene expression as shown in Fig. 2.

Cis regulatory elements regulate gene transcription by binding to transcription factors.

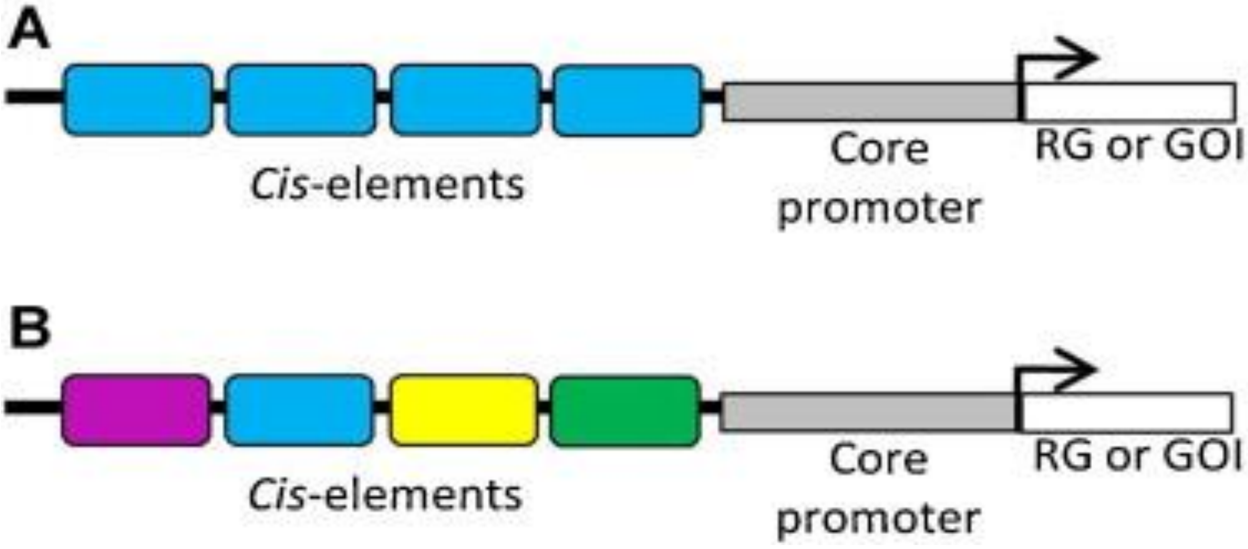


Fig. 2 This structure represents Core promoters. (A) Expression of either a reporter gene (RG) or a gene of interest (GOI) is driven by the repeats of same regulatory elements. (B) Example of a core promoter element consisting different elements of regulation which controls a reporter gene (RG) or a gene of interest (GOI). (Hernandez-Garcia *et al*., 2014)

2.2.1 *Cis* elements in regulating gene expression

Regulation of expression of a gene in promoter is primarily controlled by the *cis*-acting elements that are localized at transcriptional start site (TSS) upstream region.

Gene expression in plants is a very comprehensively controlled process as it occur at different levels, transcription being one of the most essential. Expression of genes depend on interaction of various types of mediators by the promoter area of the nucleus, along with sequence-specific proteins which bind to DNA and its related elements of promoter.

According to the PLACE database, approximately 469 *cis*-regulatory elements are present in plants. Central ACGT sequence that has been established in many promoters as a functionally important *cis*-element that responds to various stimuli such as light, enerobiocytes, jamonic acid, salicylic acid and hormones such as abscisic acid and auxin. All these factors often regulates gene expression along with other *cis*-elements.

2.2.2 *R*esearch on *cis*-acting elements

Presently, two major approaches for researching regulation of *cis*-elements are to be found which stressed on element localization and orientation. It led to finding many more novel applications related to *cis*-regulation. The main objective is to obtain methods which are not dependent on the coding sequences of the gene, which are primarily based on the organization of promoters, because some orthodox forms such as the conserved motifs have lost their previous work and, as well, some special elements have appeared during the development process (Dashtpagerdi-Shamloo *et al*., 2015). Sequence specific DNA-protein interaction are very frequently used methods for studying transcriptional regulation of genes. At present, many important factors of transcription are already validated and featured. It has expanded the methods for identifying *cis*-regulatory elements, for example: test of electrophoretic mobility, surface plasmon resonance , DNASE I footprinting (Allen 2008), chip microchips and hybrid system of yeast (Dare *et al*., 2008). Other central approaches uses silico algorithms and databases to determine the probability of precise regulatory elements (Agrawal et al., 2014). With this method, different sets of estimated elements are collected together and the classification and verification is done by using the tests which are statistical to lower the amount of elements most likely (Sharma et al., 2011). In 2015, Dashtpagerdi-Shamloo *et al*, explained the entire research process in Silico. Their main purpose is to describe the possible event of *cis*-regulatory elements, which can be statistically identifiable in research, and to determine the difference between the characteristics of the promoters and the level of gene expression which have the same function.

2.2.3 Enhancer sequences

Plant activators are generally positioned at a very different, often large distance, which are downstream or upstream from the sequence of promoter. Their role is to stimulate expression of gene through collaboration using transcription factors which are specific (Mehrotra *et al*., 2011). First theory suggested that a protein will specifically bound to sequence of the enhancer. Thus, a loop is formed with the help of connection between both the transcription factors and the enhancer. Core promoter interaction is independent of distance between promoter and enhancer. This property is compatible to looping of the DNA. It has been reported that a relationship has been mentioned between the chromatin state and the enhancers, which includes the histone modification dependence. It is well known that with the DNA availability and duplication increases by means of enhancers, these processes increase to accelerate. (Barrett *et al*., 2012). The interaction between promoters and enhancers activity depended on the distance between them, if it is important as base pairs in thousands, then the activator stops the activity of enhancers is stopped. Thus, the correlation between the number of enhancer copies and the level of expression of gene was analyzed. Initially, an additive effect was shown in many copied structures of enhancers. However, with the increase in the number of copies, the effect was often synergic, or there was also a general decline in activity of gene (Bao and Liu, 2009).

A synthetic basal ocs promoter was constructed by Bao and Liu (2009) by the help of enhancer which is obtained from gene ocs of octopine synthase. Its main aim was to study the effects of the distance from the TATA box on the number of inverted repeats (IR) of enhancers and impact on the expression of the gus reporter gene in leaves of tobacco. Level of enhancement was shown to be determined by its distance from the TATA-box and its low growth (Bao and Liu 2009). It was also confirmed by these results that the increase in the number of copies is due to enhancement of gene activity. Every copy which seemed to be additional diluted the whole effect. It indicated that each and every pair of it contains best possible number of copies of enhancer which needs to be analyzed before doing necessary studies.

There are enhancers that are obtained from active genes which are relative to promoters which are specific, resulting in variety of outcomes. CAMV 35S enhancer stimulates promoters which are specific to tissue to increase its effect in those tissues in which they are not activated originally, or can be active often (Yang *et al*., 2011). In 2007, Zheng studied three types of constructs, each of which includes the CMV 35S promoter to analyze: AAP2 with specific vascular promoter gus gene, AGL5 ovary promoter which is specific to iaaM gene, early embryonic and sexual reproduction PAB5 with typical promoter barnase gene. The enhancers act to enhance and thus trigger activity of each and every gene with different levels of expression. Though, the enhancers used are obtained from promoters in constructs consisting of promoters which are specific resulting in decreased efficiency to transform by activating expression together in vegetative and in tissues of generation (Yang *et al*., 2011).

2.2.4 5′-UTR sequence

The 5′-UTR sequence which is generally a untranslated region, known as the leader sequence, is a mRNA transcript fragment which is positioned on its 5′ end. It is programmed in the DNA strand and is submitted to transcription but not translation, though, it also regulates the subsequent process (Barrett *et al*., 2012). The 5′-UTR sequence includes 5′ cap, an upstream open reading frame (uORF), a guanine-rich fragment and IRES (internal ribosome entry site). A number of plant UTRs comprises of a pyrimidine-rich fragment which is accountable for generating high transcriptional level (Barrett *et al.*, 2012).

All of these 5′-UTR elements shows very important role in influencing the process of mRNA. Though, each and every nature of 5′-UTR is essential. As the content of nucleotides and its arrangement affects the formation of secondary loop structures, it is important. Those fragments having high GC content are meant to be good inhibitors of translation for the reason that they lead to formation of very stable hairpin loop and thus obstruct to find the start codon in nucleotides (Barrett *et al*., 2012).

There are present two different types of fragments of 5′-UTR which are distinguished. First fragment, which is present in the transcripts that are obtained from genes that are expressed in all the cells, is quite short and moreover its structure is quite simple due to small GC content. As additional AUG codons are not present so initiation of translation occurs in a very simple way and thus maintaining a steady level. In other case, 5′-UTRs are quite longer which is present in genes in developmental stage. In contrast to the first type, the latter one has more complicated secondary structure which is due to higher G+C content (Barrett *et al.*, 2012).

5′-UTR helps in stabilizing the transcript in transportation to cytoplasm from nucleus where it protects from the activity of endonucleases. Moreover, proper binding of transcript to the subunits of ribosomes depends significantly on each and every character of regions of 5′-UTR.

2.2.5 3′-UTR sequence and poly-A tail

Sequence of 3′-UTR also known as the trailer sequence which is generally positioned in a transcript that lies between the signal of polyadenylation and stop codon. 3′-UTR which is added after transcription is considered to be very significant in regulating the next stage of expression of gene i.e. translation. The 3′-UTRs perform its role as stabilizers, enhancers and silencers. 3′-UTRs transcripts which were obtained from seed storage protein (SSP) were compared with Nos terminator regarding the level of translation of protein and its accumulation in Oriza sativa seeds (Li *et al*., 2012). The mRNA transcripts level was found to be over two times greater after using the construct which contains SSP 3′-UTR whereas Nos. presented results confirmed that 3′-UTR impacted highly on enhancing the process of transcription. The degree of translation intensification was demonstrated and shown to vary but invariable for each 3′-UTR either application of promoter (CaMV 35S, GluC or Ubi-1). Those elements which are rich in adenine (AREs) are commonly found to be present within 3′-UTRs. They perform role in binding proteins which are involved in transcriptional degradation further contributing to decreased efficiency of translation or its macro-environmental conditions can also lead to strengthen the process, which depends on the cellular position (Yang *et al.*, 2011).

The 3′-UTR along with its content considerably influence the expression of final gene (Papadakis *et al.*, 2004). During the analysis of content of β-carotene, this property was confirmed .The compound that is mentioned is an essential vitamin A precursor used industrially. Inbred maize lines studies were reported and it was found that there were variations in the storage level of β-carotene.To begin with, the crtRB1 sequencing was conducted from 11 different lines of maize. The crtRB1 gene is the most important gene which is accountable for accumulation of β-carotene. This analysis has given away differences in 3′UTRs which confirmed preceding assumptions regarding the effects of sequences on variety.

The sequence of poly-A tail consists of polynucleotides containing about adenines upto 250. During termination of transcription, poly-A as well as 3′-UTR were merged with the transcript. They aided in stabilizing the structure of m RNA and thus affects the transcript transport (Li *et al.*, 2012). For determining the number of transcripts and regulating the efficiency of gene expression, these sequences are used as targets for a variety of small RNA (sn RNA) types and hence very important. (Barrett *et al.*, 2012).

2.3 Patatin- a major tuber protein, encoded by multigene family

Patatin, a family of multigenes comprises of two classes. Promoters of class I patatin gene are considered to be very attractive molecular tools to study various aspects of regulation of transcription. Merely a small number of members of this class are quite well understood at the molecular level. At this point, it was reported that two novel partial class I patatin genes were isolated with its related 5′-flanking regions, which were named as StPK01 and StPN02 from two economically effective and significant cultivars of Indian potato, ‘Kufri Chipsona-1’ and ‘Kufri Jyoti’, respectively. Its 5′-flanking regions were found to differ considerably due to insertions/deletions and also found to have point mutations. Insertion of 487-bp was a main feature of the 5′-flanking region of StPN02 which was distinguishing from other members of family. Most of the varied sequence motifs of cis-regulatory elements that were estimated in this research study have not been studied earlier. The 5′-flanking regions of 1,215 and 2,225 bp of StPN03 and StPK01, respectively, are used to generate fusions of promoter–GUS gene of patatin to assess the activities of promoter in potato. Both StPK01 as well as StPN03 promoters are found to be specific to tuber in conditions of Under field. As estimated by β glucuronidase assays of fluorometry and histochemical assays, the former was considerably more effective than the latter. Under conditions of *in vitro,* these promoters were found to be extremely inducible via sucrose in stems and leaves.

Expression of Class I patatin genes being principally expressed in parenchymatic tissues of tubers (Sosa-Rocha and Wenzler, 1989), whereas members of class II are generally restricted to root tips, rhizodermis and only in specific layers of tubers in cell (Töpferetal and Koster, 1989) but at a level which is very small in comparison to the class I members of patatin gene. The isoforms of patatin which are immunologically related are expressed during various stages of the development of plants, stating that the patatin protein needs to provide another role other than its storage protein function. According to this hypothesis, it was found that patatin produced by tubers contain enzymatic functions in relation to the lipid metabolism activity in addition to acyl transferase (wax synthase) activity.

2.4 Regulation of patatin gene synthesis

The patatin which is present in tubers has been encoded by transcripts of class I whereas transcripts of class II were present in approximately 50–100 times much smaller extent (Mignery *et al.,* 1989). Though, roots containing small number of an immunological divergent form of patatin appearing to be encoded by transcripts of class II Patatin (Mignery and Pikaard, 1987).

Tuber formation usually take place from stolons that are underground but a disease, injury or excision of stolons or tubers can lead to formation of tubers above ground from auxillary buds. The Patatin accumulation in these tubers are at same levels as the tubers, i.e. about 40–45 % of total proteins which are soluble (Paiva *et al.*, 1983). Thus, excision of auxiliary buds and tubers may determine in the storage of patatin, other proteins of tuber and level of starch in stems and petioles, which are devoid of formation of tuber (Paiva *et al*., 1983). The genes of class I patatin are expressed and patatin accumulation are also induced in leaves which are incubated with high concentration of sucrose (Sosa-Rocha and Jefferson *et al*.,1990), however, expression of patatin gene can be inhibited by wounding in tubers (Logeman *et* *al.*,1988) or cuttings of stem which is to be induced by treating with gibberellic acid. The 5′ upstream sequences studies of a patatin gene have been reported by (Grierson and Holdsworth, 1982) which intended to identify particular sequences and factors that are *trans*-acting factors as it may lead to determination of developmental regulation and inducibility of sucrose. A protein which binds to DNA, known as Storekeeper (STK) has been identified which is considered to control the regulation and expression of patatin gene (Zourelidou *et al.*, 2002).

2.4.1 Regulation of patatin genes using GUS as reporter gene

In order to examine the regulation of patatin genes , a gene was constructed consisting of 2.5 kb from sequence of 5 ' flanking region obtained by using the genomic clone PS20 of class I patatin by means of transcription and it was integrated to 3 – glucuronidase (GUS) and was introduced into plants of potato by using vector Ti-plasmid Agrobacterium tumefaciens. As the gene was expressed at elevated levels in tubers and in stolons which are attached to growing tubers, it was not generally expressed in leaves, stems, roots, or in stolons before the process of formation of tubers. However, construct of class I Patatin-GUS gene was not strictly "tuber-specific" as leaf and stem explants that were cultured on the medium contained high levels of sucrose upto 300 to 400 mM showing that activity of GUS is almost similar to or larger than tubers . The activity of GUS by means of sucrose induction in stem and leaf explants were assisted by the storage of the patatin protein and large amount of starch , but not through the morphological changes which are usually associated with tuberization . On the Contrary, ‘cauliflower mosaic virus’ (CaMV) consist of 35S promoter which is under the control of GUS reporter gene did not show any "tuber-specific" pattern of expression in transgenic potato , neither it was expressed in transgenic stem and leaf explants by the means of sucrose.

The intensity of activity of GUS in tubers which contains the construct of class I Patatin-GUS gene was about 50 times high than the one that we see in tubers of construct of class II patatin-GUS (Park-Mignery *et al.*, 1989).

Stolons which are extracted from the tuber plant also found to possess high levels of expression of GUS (10-30% in comparison to tubers). This activity found to vary which depends on the physiological status and age of tissue. Though activities of GUS were 2 to 3 times higher in leaves and stem extracts in plants which are tuberizing to plants in comparison to plants which are non-tuberizing, they were responsible for lower than 0.5% of the activity seen in the extracts of tuber. These results tend to oppose the research studies of Paiva *et al*., 1989 who stated that protein levels in patatin were found to be greater in tubers and different in comparison to stolons. On the contrary ,it was very small in stolons from non-tuberizing plants and in leaves , stems , and root from either extracts of tuberizing or non-tuberizing plants.

Although activity of GUS found in the extracts of stem and leaves in general was quite very low. The level of GUS expression in the leaf extracts was found to be maximum in the central plant part after the process of tuberization , but was consistently very low in the extracts of leaf of such plants before the process of tuberization . On the contrary, the activity of GUS in the extracts of stem were quite high in the shoot region of plant before and after the process of tuberization.

2.4.2 Tissue specificity of the promoters

Those promoters are specific to tissue which express in particular types of tissue, which have already been characterized, the residual tissues which do not have the heterologous protein is a important feature. Production of exogenous protein leads to a variety of disorders and developmental deformity, for example, programmed cell death or dwarfism. As plant metabolism hazard is much lower, so the promoters which are constitutive are used because of the presence of heterologous peptide in the exact organ.

The Class I patatin family members of Solanaceae are PAT 21 and B33 which are expressed highly in early stage of development of tuber in vascular tissues with later stages of development, both in vascular as well as parenchymal tissues. Both being specific to tuber, but inducible in leaves through exogenous sucrose production. Thus, the sequence homology is quite extensive in the PAT 21 and B33 promoters. For instance, the cytochrome P450 gene taken from rat (CYP1A1) provided in potatoes to increase the process of detoxifying the residual herbicides in soil. The tubers which were developing showed very elevated levels of the CYP1A1 m RNA and were rich in proteins and the concentration of these herbicides were comparatively very less than tubers which were not transformed. (Opabode & Akinyemiju, 2015)

2.4.3 Functional characteristics of patatin

Potatoes are considered to be an important starch source for its food and industrial uses, as its tuber proteins forms a by-product. These proteins limit their use for low value feed for livestock due to their recovery in a separate state. However, potato proteins which are not denatured are capable of functional properties However, potato proteins which are not denatured have capable properties of functionality (e.g. stabilization and development of emulsions and foams) (Yada and Jackman, 1988; Ralet and Gueguen and Ralet, 2000) and excellent quality of nutrition (Liedl-Kapoor *et al.*, 1987). As a result, various studies have been conducted based on the composition and properties of patatin, principally on its sustainability and thermal aggregation in relation to the production of functional proteins on industrial scale.

According to Pots (1999), patatin contained 10 peaks by reverse-phase high-pressure liquid chromatography, which were separated further into four pools by ion exchange chromatography. These pools are responsible for 7% (D), 5% (C), 26% (B) and 62% (A) of the total, containing about 40,400 and 41,600 in each fraction of isoform which were due to glycosylational differentiation. No other differences have been observed in the properties or conformational stability of the pools. In different studies, the same group has investigated the results of temperature and pH on the aggregation of total fractions and thus stability of patatin (Pots *et al.*, 1998).

CHAPTER 3 OBJECTIVES

- To find the motifs conserved in the core promoters and regulatory regions of the patatin gene homologues .
- To analyse the promoter and regulatory regions and identify the subsets of the above sequences showing highly conserved sequence motifs.
- To find some unique *cis*-elements/regulatory motifs in some of the patatin gene homologues.

CHAPTER 4 MATERIALS AND METHODS

MATERIALS

Data Source

The patatin gene promoter sequence (accession numbers JX124227.1 and JX124228.1) data were collected from the published work of Aminedi and Das (2014). Complete genomic DNA sequence of *Solanum tuberosum* was searched from National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). BLAST (Basic Local Alignment Tool) was then run for both the sequences. In BLAST, considering JX124227.1 and JX124228.1 as query sequences which comprised of 23 hits and 35 hit sequences respectively of Patatin genes from potato (*Solanum tuberosum*).

Table 1 :- A total of 60 sequences were selected which were downloaded in FASTA format.

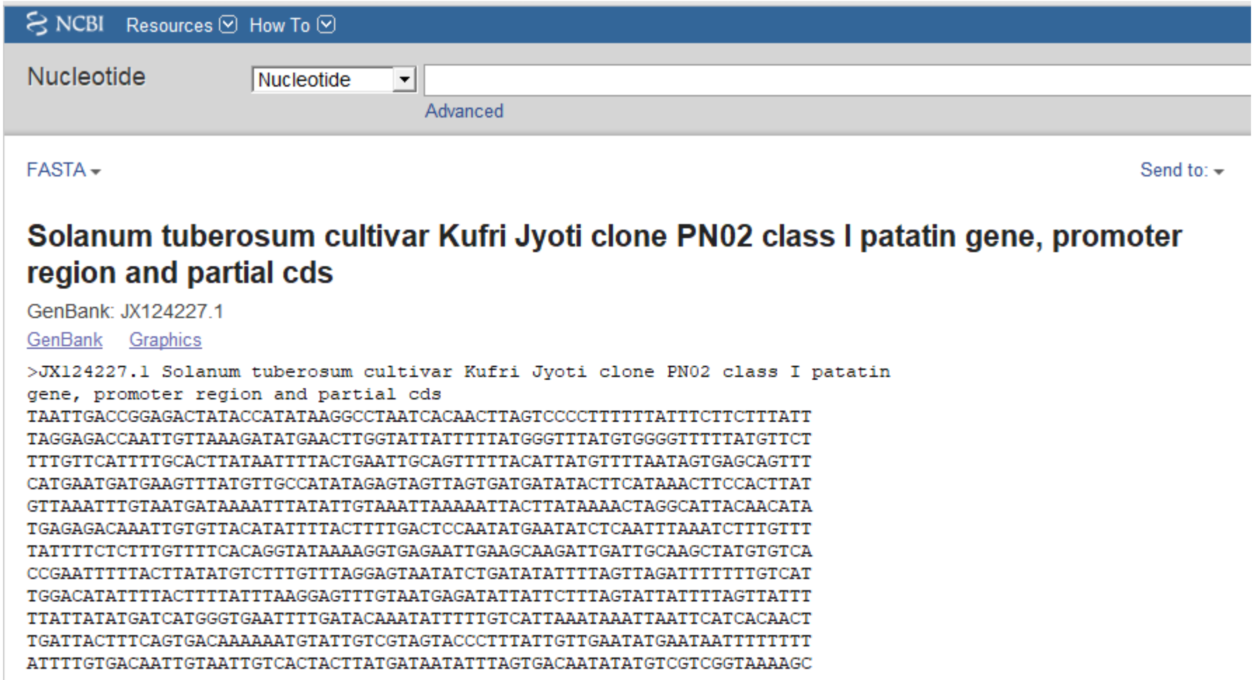
S. No.	Accession number	Length
1	JX124227.1	2747
2	DQ274179.2	154382
3	JX124228.1	2332
4	X03956.1	6094
5	X60399.1	1391
6	M18882.1	1139
7	M18881.1	1138
8	JX124230.1	1454
9	JQ341191.1	1254
10	JX124231.1	1441
11	GQ352473.1	999
12	AY485645.1	987
13	AY778964.1	970

S. No.	Accession number	Length
1	JX124228.1	2332
2	X03956.1	6094
3	DQ274179.2	154382
4	JX124227.1	2747
5	JX124230.1	1454
6	GQ352473.1	999
7	AY485645.1	987
8	JX124229.1	874
9	AY778964.1	970
10	X87216.1	1770
11	JQ341191.1	1254
12	JX124231.1	1441
13	M18880.1	4029

14	X60396.1	1407
15	X14483.1	1975
16	X87216.1	1770
17	JX124229.1	874
18	M18880.1	4029
19	GU906238.1	46060
20	GU906240.1	98148
21	GU906234.1	76152
22	GU168944.1	1060
23	M18883.1	708
24	AC237669.1	120723

14	X60396.1	1407
15	GU168944.1	1060
16	X14483.1	1975
17	X60399.1	1391
18	GU906238.1	46060
19	M18883.1	708
20	M18882.1	1139
21	M18881.1	1138
22	X07029.1	297
23	M37225.1	287
24	M37223.1	454
25	X07032.1	465
26	X04077.1	6799
27	M18885.1	738
28	X17397.1	1356
29	X60397.1	1611
30	X04078.1	7373
31	X03932.1	5313
32	X01125.1	1437
33	M27565.1	217
34	Z27221.1	1357
35	X13178.1	1378
36	M21879.1	1404

Fig. 3 Screenshot of FASTA sequence of JX124227.1.

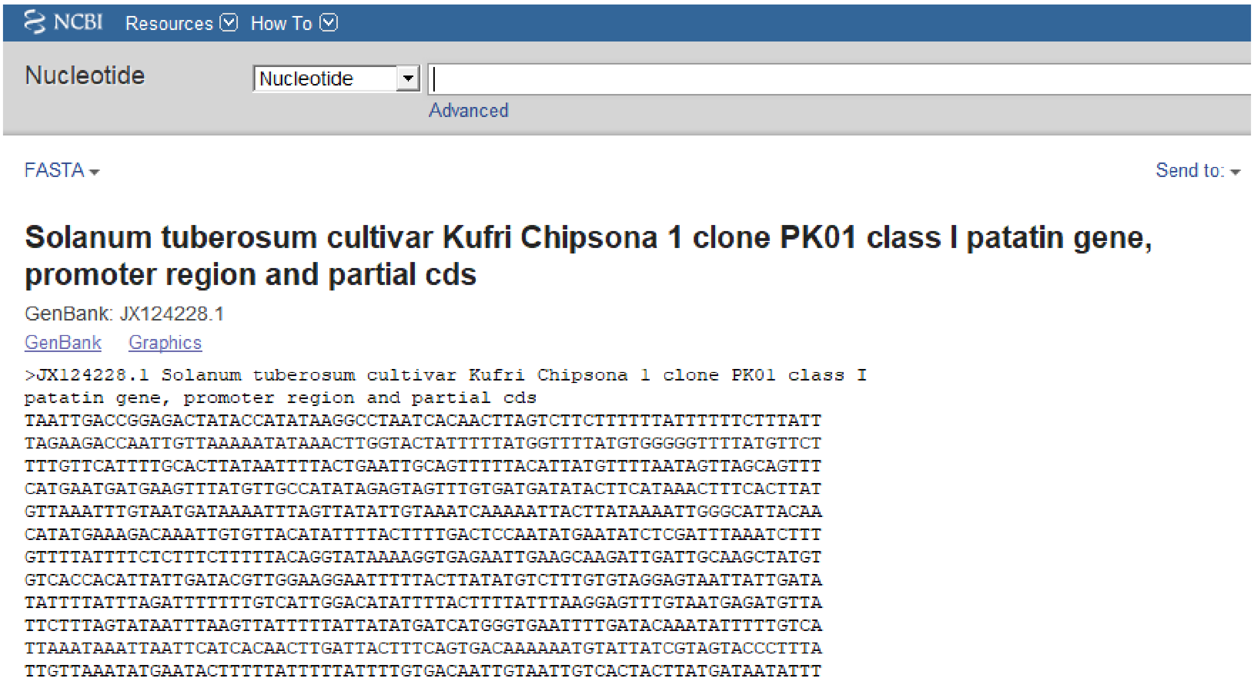


Solanum tuberosum cultivar Kufri Jyoti clone PN02 class I patatin gene, promoter region and partial cds

GenBank: JX124227.1

```
>JX124227.1 Solanum tuberosum cultivar Kufri Jyoti clone PN02 class I patatin
gene, promoter region and partial cds
TAATTGACCGGAGACTATACCATATAAGGCCTAATCACAACTTAGTCCCCTTTTTTATTTCTTCTTTATT
TAGGAGACCAATTGTTAAAGATATGAACTTGGTATTATTTTTATGGGTTTATGTGGGGTTTTTATGTTCT
TTTGTTCATTTTGCACTTATAATTTTACTGAATTGCAGTTTTTACATTATGTTTTAATAGTGAGCAGTTT
CATGAATGATGAAGTTTATGTTGCCATATAGAGTAGTTAGTGATGATATACTTCATAAACTTCCACTTAT
GTTAAATTTGTAATGATAAAATTTATATTGTAAATTAAAAATTACTTATAAAACTAGGCATTACAACATA
TGAGAGACAAATTGTGTTACATATTTTACTTTTGACTCCAATATGAATATCTCAATTTAAATCTTTGTTT
TATTTTCTCTTTGTTTTCACAGGTATAAAAGGTGAGAATTGAAGCAAGATTGATTGCAAGCTATGTGTCA
CCGAATTTTTACTTATATGTCTTTGTTTAGGAGTAATATCTGATATATTTTAGTTAGATTTTTTTGTCAT
TGGACATATTTTACTTTTATTTAAGGAGTTTGTAATGAGATATTATTCTTTAGTATTATTTTAGTTATTT
TTATTATATGATCATGGGTGAATTTTGATACAAATATTTTTGTCATTAAATAAATTAATTCATCACAACT
TGATTACTTTCAGTGACAAAAAATGTATTGTCGTAGTACCCTTTATTGTTGAATATGAATAATTTTTTT
ATTTTGTGACAATTGTAATTGTCACTACTTATGATAATATTTAGTGACAATATATGTCGTCGGTAAAAGC
```

Fig. 4 Screenshot of FASTA sequence of JX124228.1.



Solanum tuberosum cultivar Kufri Chipsona 1 clone PK01 class I patatin gene, promoter region and partial cds

GenBank: JX124228.1

```
>JX124228.1 Solanum tuberosum cultivar Kufri Chipsona 1 clone PK01 class I
patatin gene, promoter region and partial cds
TAATTGACCGGAGACTATACCATATAAGGCCTAATCACAACTTAGTCTTCTTTTTTATTTTTTCTTTATT
TAGAAGACCAATTGTTAAAAATATAAACTTGGTACTATTTTTATGGTTTTATGTGGGGGTTTTATGTTCT
TTTGTTCATTTTGCACTTATAATTTTACTGAATTGCAGTTTTTACATTATGTTTTAATAGTTAGCAGTTT
CATGAATGATGAAGTTTATGTTGCCATATAGAGTAGTTTGTGATGATATACTTCATAAACTTTCACTTAT
GTTAAATTTGTAATGATAAAATTTAGTTATATTGTAAATCAAAAATTACTTATAAAATTGGGCATTACAA
CATATGAAAGACAAATTGTGTTACATATTTTACTTTTGACTCCAATATGAATATCTCGATTTAAATCTTT
GTTTTATTTTCTCTTTCTTTTTACAGGTATAAAAGGTGAGAATTGAAGCAAGATTGATTGCAAGCTATGT
GTCACCACATTATTGATACGTTGGAAGGAATTTTTACTTATATGTCTTTGTGTAGGAGTAATTATTGATA
TATTTTATTTAGATTTTTTTGTCATTGGACATATTTTACTTTTATTTAAGGAGTTTGTAATGAGATGTTA
TTCTTTAGTATAATTTAAGTTATTTTTATTATATGATCATGGGTGAATTTTGATACAAATATTTTTGTCA
TTAAATAAATTAATTCATCACAACTTGATTACTTTCAGTGACAAAAAATGTATTATCGTAGTACCCTTTA
TTGTTAAATATGAATACTTTTTATTTTTATTTTGTGACAATTGTAATTGTCACTACTTATGATAATATTT
```

Sequence Analysis

Microsoft Excel: Microsoft Excel was used for the statistical and computational analysis of the data. Microsoft excel was also used for forming graphs and analyzing graphical representation. The various operations along with the functions used were:

Operation Name	**Class**	**Function**
IF	Logical	It makes logical comparisons and checks whether a condition is met; returns a value depending on whether the test is TRUE or FALSE.
COUNTIF	Statistical	It counts the number of cells within a range that meet a specific given criteria

Table 2: Microsoft excel operations

Multiple Sequence Alignment (MSA)

Multiple Sequence Alignment of all the hit sequences which are non-overlapping with JX124227.1 and JX124228.1 of *Solanum tuberosum* patatin genes were carried out. The tool that was used to carry out alignment of 46 sequences of Patatin gene was “MUSCLE”.

PLACE

PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE/) is a database of motifs which are found in plant *cis*-acting regulatory DNA elements extracted from reports which are previously published.

PLANTPROM

PLANTPROM(http://www.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom) is a database of motifs with annotated, non-redundant collection of proximal promoter sequences for RNA polymerase II with experimentally determined transcription start site(s) (TSS) from various plant species.

Notepad++

Notepad++ is a text editor and a source code editor. It mainly supports tabbed editing, which allows working with multiple open files in a single window. It was used for recording and running macros in order to perform repetitive changes in the text information.

METHODS

Genomic sequence search for Class I patatin genes

The sequence information of Class I patatin genes was submitted to the GenBank database under the Accession Numbers JX124227 and JX124228, respectively. BLAST (Basic Local Alignment Tool) was then run for both the sequences. The nucleotide sequences of the class I patatin genes were analyzed by NCBI BLAST search tools (http://www.ncbi.nlm.nih.gov/BLAST). A total of 60 sequences were selected which were downloaded in FASTA format. Only sequences of patatin genes related to *Solanum tuberosum* were selected.

Sequence analysis from GenBank

Data obtained from annotated GenBank record for Class I patatin genes of *Solanum tuberosum* demonstrated many of the features of the GenBank flat file format. By clicking on the GenBank link of the genomic refseq, the genomic sequence is obtained with detailed portions that correspond to the exons and introns of the genomic locus.

It provided us detailed information about the position of TATA box and TSS. In this way, the regulatory region and thus promoters can be characterized, as, first exon sequence which is not part of the CDS (5′ UTR) starts with the TSS. So, the regions upstream to it correspond to "promoter" sequences.

Sequence selection

In order to remove redundancy, only those blast result sequences were selected which did not have any overlapping regions with the query sequences. As a result 46 sequences were selected which were non-overlapping with the query sequences or each other. These non-overlapping hits in the same accession number based on nucleotide blast are paralogs in series containing rich information of some unique elements present in promoter region.

Multiple Sequence Alignment of genomic sequences

Multiple Sequence Alignment (MSA) of 46 sequences was performed by using MUSCLE tool. It was done to identify regions of similarity or conserved regions. So, in this way sequences were assessed which shared common evolutionary origin.

Analysis in MS Excel

Most of the statistical and computational analysis was done in MS Excel, for that, the data had to be in a format that could be viewed under MS Excel. For this each aligned sequence were copied in Notepad++. Similar procedure was repeated for each 46 sequences and these sequences were then copied to MS Excel worksheet.

Software Used	File Extension Required
PLACE	.htm
Notepad ++	.txt
Microsoft Excel	.xlsx/.xls
PLANTPROM	.htm

Table 3: File extensions required for different software tools

PLANTPROM

Plant gene promoter analysis – Basal promoter and putative *cis*-regulatory sequence motifs were predicted by sequence comparison with the published reports and other databases namely **PLACE** (http://www.dna.affrc.go.jp/PLACE/; Higo *et al.* 1999) and **PLANTPROM** (http://www.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom;Shahmuradov *et al.*, 2003). PlantProm is a database which provides DNA sequence of the promoter regions with TSS on the fixed position for promoter elements: TATA-box, CAAT-box and TSS-motif. Its only purpose was to provide experimentally determined promoter sequences which were very close to the transcription start site. PlantProm software was run for each sequence. A total of 741 regulatory motifs were collected along with the transcription factor binding sites for each promoter at its particular position.

Then, comparison of both these databases was done in MS Excel to analyse the motifs which were found to be common in both the PLACE as well as PLANTPROM. It was aimed to look for the unique elements that were present in the gene.

PLACE

It is a database which is used for searching for the presence of motifs to be found in plant *cis*-regulatory DNA elements**.** A total of 4995 *cis*-regulatory motifs were predicted along with its position in the genomic sequence.

EMBOSS: palindrome (emboss.bioinformatics.nl/cgi-bin/emboss/palindrome) is a tool for identification of palindromes in DNA sequence. This tool that was run to get the palindromic sequences in each one of the 46 sequences.

Its parameters selected are as follows:

Minimum length of palindrome = 4

Maximum length of palindrome = 100

Maximum gap between repeated regions = 1

Number of mismatches allowed = 1

The data was analysed in MS Excel for further attributes such as GC% in the palindrome.

CHAPTER 5 RESULTS

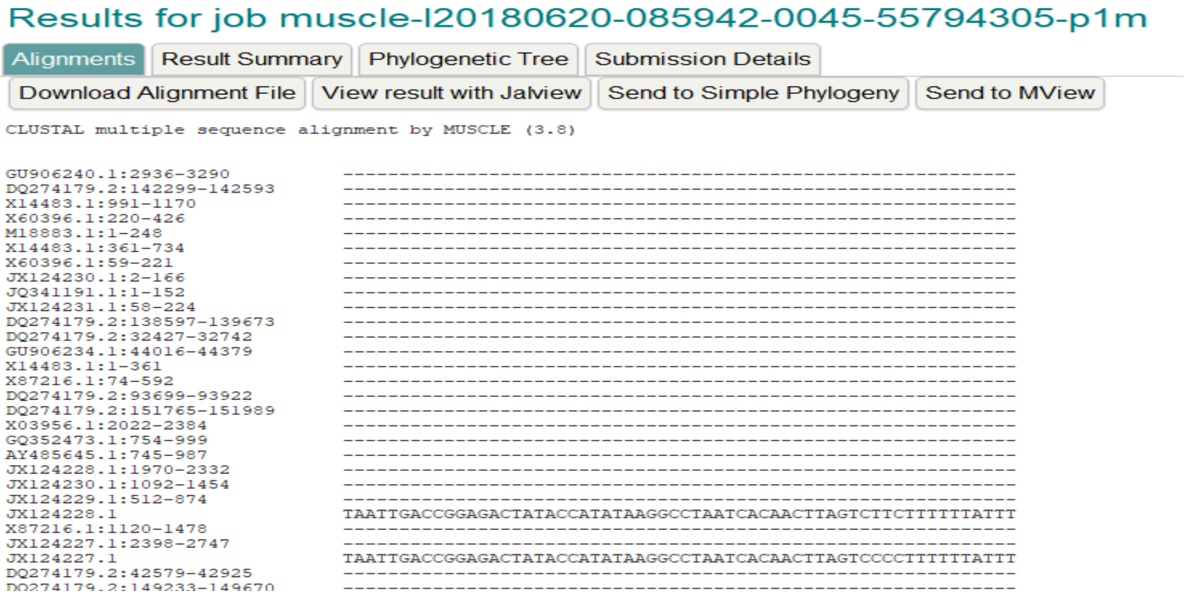
BLAST

To find all the possible homologues of patatin gene, its known sequences (Accession Numbers JX124227 and JX124228) were used as query sequences in nucleotide blast search. The blast of JX124227.1 and JX124228.1 resulted in 23 and 35 hit sequences respectively that had potato origin. A total of 60 sequences (including the two query sequences) were selected and downloaded in FASTA format.

Out of these 60 sequences obtained through nucleotide blast, those sequences along with its sub sequences or hit sequences were selected which were not overlapping with the query sequences *i.e.* JX124227.1 and JX124228.1. Removal of partly or completely overlapping hits sequence resulted in a subset of 46 sequences.

Multiple Sequence Alignment

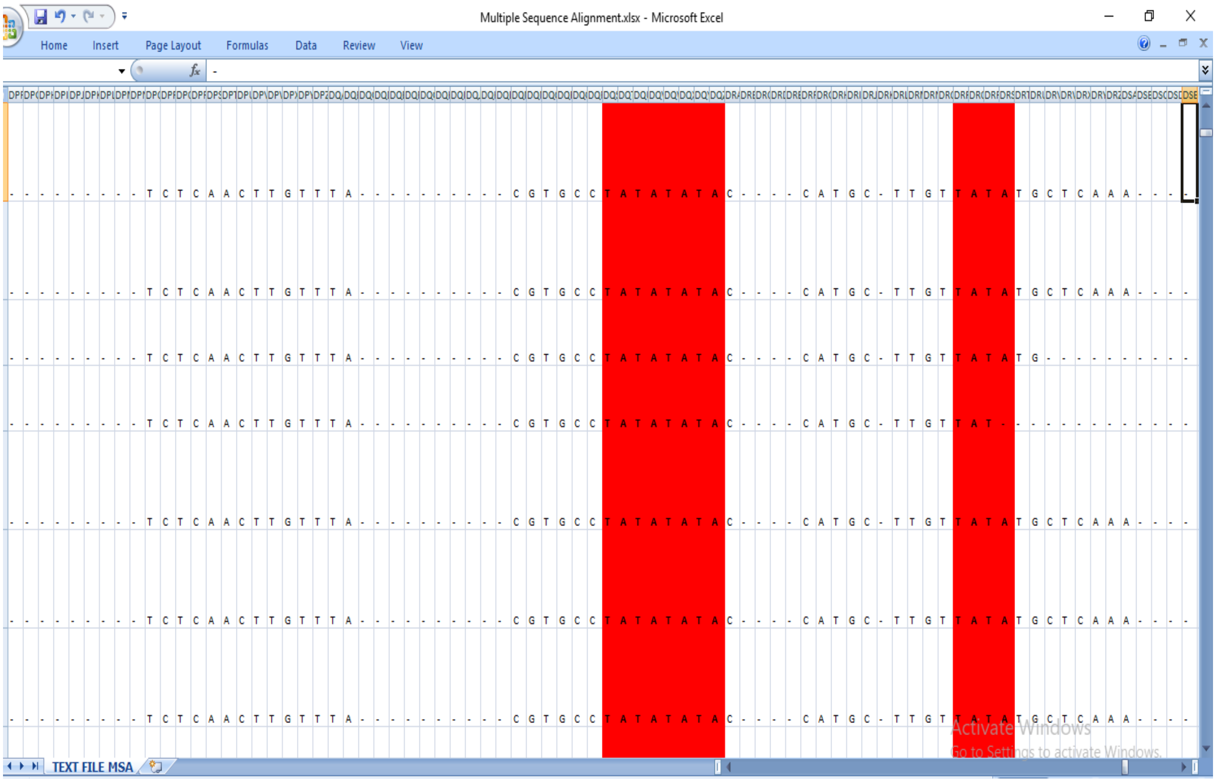
One of the major objective of this work was to compare the promoters and cognate regulatory regions of all the known paralogs of Class I patatin genes in *Solanum tuberosum* to get information about regulatory *cis*-elements. The 46 promoter sequences were subjected to Multiple Sequence Alignment (MSA) for detection of conserved regions as shown in Fig. 5. It was expected that regulatory sequences might be conserved in the paralogs. MSA of 46 homologues of Patatin gene were performed using different tools such as Clustal W, MUSCLE and T-coffee, however it failed to show conservation of motif with exception of TATA box. Best alignment was shown by MUSCLE tool. As a result, MUSCLE was selected to perform MSA of 46 homologues of Patatin gene. However no significant conservation was detected except the TATA box.



```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


GU906240.1:2936-3290          ------------------------------------------------------------
DQ274179.2:142299-142593      ------------------------------------------------------------
X14483.1:991-1170             ------------------------------------------------------------
X60396.1:220-426              ------------------------------------------------------------
M18883.1:1-248                ------------------------------------------------------------
X14483.1:361-734              ------------------------------------------------------------
X60396.1:59-221               ------------------------------------------------------------
JX124230.1:2-166              ------------------------------------------------------------
JQ341191.1:1-152              ------------------------------------------------------------
JX124231.1:58-224             ------------------------------------------------------------
DQ274179.2:138597-139673      ------------------------------------------------------------
DQ274179.2:32427-32742        ------------------------------------------------------------
GU906234.1:44016-44379        ------------------------------------------------------------
X14483.1:1-361                ------------------------------------------------------------
X87216.1:74-592               ------------------------------------------------------------
DQ274179.2:93699-93922        ------------------------------------------------------------
DQ274179.2:151765-151989      ------------------------------------------------------------
X03956.1:2022-2384            ------------------------------------------------------------
GQ352473.1:754-999            ------------------------------------------------------------
AY485645.1:745-987            ------------------------------------------------------------
JX124228.1:1970-2332          ------------------------------------------------------------
JX124230.1:1092-1454          ------------------------------------------------------------
JX124229.1:512-874            ------------------------------------------------------------
JX124228.1                    TAATTGACCGGAGACTATACCATATAAGGCCTAATCACAACTTAGTCTTCTTTTTTATTT
X87216.1:1120-1478            ------------------------------------------------------------
JX124227.1:2398-2747          ------------------------------------------------------------
JX124227.1                    TAATTGACCGGAGACTATACCATATAAGGCCTAATCACAACTTAGTCCCCTTTTTTATTT
DQ274179.2:42579-42925        ------------------------------------------------------------
DQ274179.2:149233-149670      ------------------------------------------------------------
```

Fig. 5 Multiple Sequence Alignment of 46 class I patatin genomic sequences performed by MUSCLE tool.



```
- - - - - - - - - T C T C A A C T T G T T T A - - - - - - - - - - C G T G C C T A T A T A T A C - - - - C A T G C - T T G T T A T A T G C T C A A A - - - -
- - - - - - - - - T C T C A A C T T G T T T A - - - - - - - - - - C G T G C C T A T A T A T A C - - - - C A T G C - T T G T T A T A T G C T C A A A - - - -
- - - - - - - - - T C T C A A C T T G T T T A - - - - - - - - - - C G T G C C T A T A T A T A C - - - - C A T G C - T T G T T A T A T G - - - - - - - - - -
- - - - - - - - - T C T C A A C T T G T T T A - - - - - - - - - - C G T G C C T A T A T A T A C - - - - C A T G C - T T G T T A T - - - - - - - - - - - - -
- - - - - - - - - T C T C A A C T T G T T T A - - - - - - - - - - C G T G C C T A T A T A T A C - - - - C A T G C - T T G T T A T A T G C T C A A A - - - -
- - - - - - - - - T C T C A A C T T G T T T A - - - - - - - - - - C G T G C C T A T A T A T A C - - - - C A T G C - T T G T T A T A T G C T C A A A - - - -
- - - - - - - - - T C T C A A C T T G T T T A - - - - - - - - - - C G T G C C T A T A T A T A C - - - - C A T G C - T T G T T A T A T G C T C A A A - - - -
```

Fig. 6 Some of the Patatin gene homologues showing conserved TATA box in the MSA

However the MSA showed conservation in the TATA box regions of the promoters. TATA box plays major role in the activation of eukaryotic genes transcribed by RNA polymerase II.

This region is necessary as TATA binding protein binds specifically to the TATA box promoter element which lies very close to the position of transcriptional initiation. Nearly half of the eukaryotic promoters have TATA box. The aligned sequences were reformatted into MS Excel spreadsheet for further analysis. Fig. 6 depicts conserved regions being analyzed in MS Excel which also helped in finding the position of promoter from Transcription Start Site (TSS).

In silico analysis of promoter sequences of Patatin gene homologues

Since no conserved sequences were detected by MSA, an alternative approach was used in search of *cis* elements. Each promoter sequence was analyzed using web based computational tools. The putative *cis*-regulatory sequence motifs were predicted by database screening tools namely **PLACE** (http://www.dna.affrc.go.jp/PLACE/; Higo *et al.* 1999) and **PLANTPROM**(http://www.softberry.com/berry.phtml?topic=plantprom&group=data&subgroup=plantprom; Shahmuradov et al., 2003)

PLACE

It is a database which is used for searching for the presence of motifs in plant promoters and regulatory regions. A total of 4995 *cis*-regulatory motifs in 46 input sequences were predicted by using PLACE tool out of which 2394 motifs were found on negative strand while 2601 motifs were present on positive strand. There are more elements found on (+) strand than (-) strand. This is an interesting observation. Binomial distribution based analysis shows that there is a bias for elements to be found on the (+) strand with a p-value of 0.0018. It may be inferred that many of these elements are playing important function in the gene expression/regulation.

PLANTPROM

It is a database of motifs with annotated, non-redundant collection of proximal promoter sequences. A total of 741 *cis* regulatory motifs out of 46 sequences were predicted by using PlantProm tool out of which 343 motifs were found on negative strand while 398 motifs were present on positive strand. These motifs are more in (+) strand and less in (-) strand and binomial distribution analysis showed that like PLACE results only, there are significantly more motifs on (+) strand in comparison with (-) strand with a p-value of 0.0236.

Analysis of promoters by comparing PLACE and PLANTPROM

Since same set of DNA sequences were analyzed by two similar yet distinct tools, it was interesting to compare the search results. The goal of this comparison was to look for *cis*-elements that were commonly found by the two tools. The set of *cis*-element sequences were pasted in MS Excel spreadsheet and using '*Remove duplicates'* tool in '*Data*' tab, all the duplicate sequences were removed in each set of results. Removal of duplicate sequences in each set reduced the total number of sequences from 4995 to 128 in PLACE data set while from 741 to 76 in PLANTPROM data set. Then one of the set was copied and special pasted using '*transpose*' option, in the column next to the one having sequence of other set of sequences as shown in Table 4. Using '*if*' function, each possible pair between the two sets of sequences was searched for identical sequences.

PLACE			
S.No.	Accession no.	Length	No. of Motifs
1	JX124227.1	2747	695
2	DQ274179.2	32427-32742	86
3	DQ274179.2	93699-93922	47
4	DQ274179.2	110976-111184	43
5	DQ274179.2	138597-139673	288
6	DQ274179.2	142299-142593	66
7	DQ274179.2	148257-148366	41
8	DQ274179.2	151765-	47

PLANTPROM			
S.No.	Accession no.	Length	No. of Motifs
1	JX124227.1	2747	168
2	DQ274179.2	32427-32742	-
3	DQ274179.2	93699-93922	-
4	DQ274179.2	110976-111184	-
5	DQ274179.2	138597-139673	44
6	DQ274179.2	142299-142593	17
7	DQ274179.2	148257-148366	-
8	DQ274179.2	151765-	-

		151989	
9	JX124228.1	1970-2332	99
10	X03956.1	2022-2384	97
11	JX124230.1	2-166	48
12	JX124230.1	1092-1454	99
13	JQ341191.1	1-152	42
14	JX124231.1	58-224	46
15	JX124231.1	1085-1441	96
16	GQ352473.1	754-999	69
17	AY485645.1	745-987	69
18	AY778964.1	741-970	62
19	X60396.1	429-456	5
20	X14483.1	991-1170	53
21	X14483.1	1177-1477	84
22	X14483.1	1456-1807	84
23	X87216.1	593-1141	174
24	X87216.1	1120-1478	88

		151989	
9	JX124228.1	1970-2332	30
10	X03956.1	2022-2384	30
11	JX124230.1	2-166	-
12	JX124230.1	1092-1454	30
13	JQ341191.1	1-152	-
14	JX124231.1	58-224	-
15	JX124231.1	1085-1441	36
16	GQ352473.1	754-999	-
17	AY485645.1	745-987	-
18	AY778964.1	741-970	-
19	X60396.1	429-456	-
20	X14483.1	991-1170	-
21	X14483.1	1177-1477	-
22	X14483.1	1456-1807	28
23	X87216.1	593-1141	73
24	X87216.1	1120-1478	31

25	JX124229.1	512-874	99
26	M18880.1	1-400	130
27	GU906240.1	2936-3290	70
28	GU906234.1	44016-44379	70
29	GU168944.1	434-734	85
30	JX124228.1	2332	640
31	DQ274179.2	42579-42925	97
32	DQ274179.2	149233-149670	128
33	JX124227.1	2398-2747	99
34	X87216.1	74-592	120
35	JQ341191.1	1015-1254	74
36	X60396.1	59-221	43
37	X60396.1	220-426	64
38	X60396.1	1118-1407	82
39	GU168944.1	3-213	54
40	X14483.1	1-361	88
41	X14483.1	361-734	108
42	X14483.1	752-974	52

25	JX124229.1	512-874	30
26	M18880.1	1-400	-
27	GU906240.1	2936-3290	9
28	GU906234.1	44016-44379	-
29	GU168944.1	434-734	-
30	JX124228.1	2332	96
31	DQ274179.2	42579-42925	30
32	DQ274179.2	149233-149670	-
33	JX124227.1	2398-2747	30
34	X87216.1	74-592	-
35	JQ341191.1	1015-1254	-
36	X60396.1	59-221	-
37	X60396.1	220-426	-
38	X60396.1	1118-1407	29
39	GU168944.1	3-213	-
40	X14483.1	1-361	-
41	X14483.1	361-734	-
42	X14483.1	752-974	-

43	X60399.1	1011-1357	97
44	M18883.1	1-248	63
45	M18882.1	1-350	102
46	M18881.1	1-350	102
		Total	**= 4995**

43	X60399.1	1011-1357	30
44	M18883.1	1-248	-
45	M18882.1	1-350	-
46	M18881.1	1-350	-
		Total	**= 741**

Table 4 Total number of motifs obtained in each sequence by both PLACE and PLANTPROM

Note: where (-) indicated that no promoters/enhancers were predicted in that region of Accession number. In similar way, the common motifs present or occurring no. of times in both PLACE and PLANTPROM were counted to find unique elements responsible for regulating transcription

It was found that 16 motifs were common between the results obtained by the two tools. Out of these 16 motifs, 5 were found to be present in 5 or less number of promoters in the results of either of the two tools (Table 5). It may be inferred that 5 motifs are detected commonly by both the promoter analysis tools and might be playing important regulatory role. On the other hand 11 motifs are more commonly present in the promoters of the patatin gene homologues. It is expected that some of these relatively common *cis*-elements regulate the subset of potato patatin gene homologues.

S.No.	**Name of Motif**	**Motif detected**	**No. of Motifs in PLACE**	**No. of Motifs in PLANTPROM**
1	*REBETALGLHCB21*	*CGGATA*	*1*	*1*
2	*S1FSORPL21*	*ATGGTATT*	*1*	*1*
3	*PALBOXAPC*	*CCGTCC*	*3*	*1*
4	*CIACADIANLELHC*	*CAANNNNATC*	*5*	*1*
5	*SEF3MOTIFGM*	*AACCCA*	*3*	*2*

6	HEXMOTIFTAH3H4	ACGTCA	3	4
7	DRE2COREZMRAB17	ACCGAC	7	1
8	SEF1MOTIF	ATATTTAWW	7	5
9	ACGTTBOX	AACGTT	8	2
10	CGACGOSAMY3	CGACG	11	1
11	CATATGGMSAUR	CATATG	16	2
12	PROLAMINBOXOSGLUB1	TGCAAAG	17	12
13	REALPHALGLHCB21	AACCAA	26	15
14	PYRIMIDINEBOXOSRAMY1A	CCTTTT	47	15
15	RAV1AAT	CAACA	75	17
16	SEF4MOTIFGM7S	RTTTTTR	77	12

Table 5 Regulatory motifs detected which are commonly occurring per no. of times in both PLACE and PLANTPROM

Note: Motifs highlighted with grey background are unique elements which are occurring once or more than once but less than five times as detected by PLACE and PLANTPROM

A total number of 16 motifs were detected which were found to be common in both PLACE and PLANTPROM. Out of those 16 regulatory motifs, 5 unique motifs were selected which were further analyzed in MS Excel. The selection was based on occurrence of the motifs in the promoters. Only those motifs were selected which were found in six or fewer promoters.

Analysis of Unique Regulatory Motifs

(I) CGGATA (REBETALGLHCB21) which was detected in JX124227.1 accession number sequence was also detected in Lemna gibba Lhcb21 gene (Degenhardt *et al*., 1996)

(II) ATGGTATT (S1FSORPL21) which was detected in JX124227.1 accession number sequence was also detected in RPL21 gene (Zhou *et al*., 1992)

(III) CCGTCC (PALBOXAPC) which was detected in JX124227.1, M18882.1 & M18881.1 accession number sequences was also detected in phenylalanine ammonia-lyase gene (Logemann *et al*., 1995)

(IV) CAANNNNATC (CIACADIANLELHC) which was detected in JX124227.1, DQ274179.2, X14483.1, GU168944.1 & JX124228.1 accession number sequences was also detected in tomato Lhc gene (Piechulla et al., 1998).

(V) AACCCA (SEF3MOTIFGM) which was detected in JX124227.1, DQ274179.2 & JX124228.1 accession number sequences was also detected in beta-conglycinin (7S globulin) gene (Allen *et al*., 1989)

Analysis of Palindromic sequences

To compare promoter and regulatory regions of potato patatin gene homologs, their sequences were also analyzed by an unconventional approach. It has been reported that palindromic sequences and potential secondary structures arising because of them are related to regulation of gene expression (Cattani et al., 2016 ; Brázda, *et al.,* 2017 ; Jerabek *et al*., 2017 ; Miura *et al*., 2018). Each of the 46 promoter sequences was analyzed for palindromic sequences using '*Emboss: palindrome*' tool with the parameters mentioned in the methods. The search resulted in detection of 1633 palindromic sequences (both with loop size 0 or 1 and having <2 mismatches)**.** The 1633 palindromes consisted of 568 different palindromic sequences which showed that many palindromes were present in more than one sequence. There was one palindrome sequence which was found in 26 different sequences while there were 364 palindrome sequence which were unique. The distribution of different palindrome sequences having varying frequencies is shown in the Fig. 7

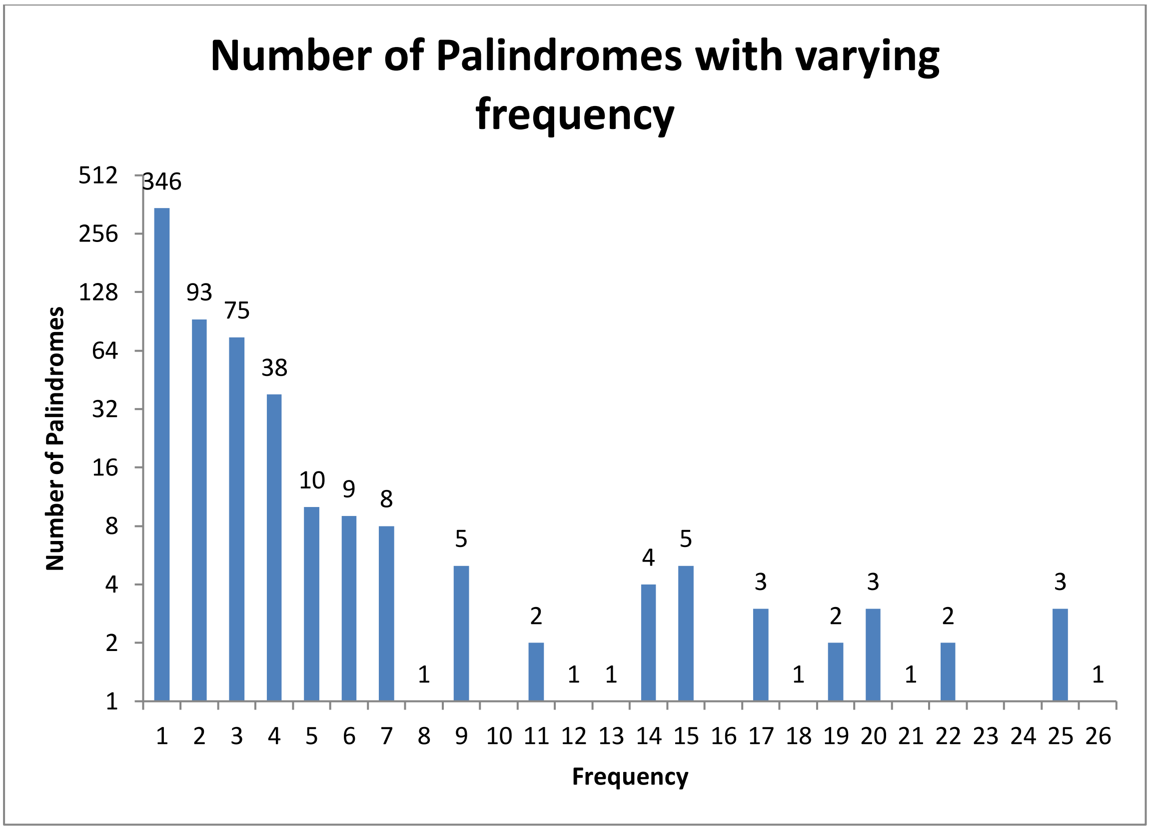


Fig. 7 Frequency of palindromes occurring per number of times. As shown from left: where 346 sequences are occurring only once and from extreme right: where there is 1 sequence which has occurred 26 number of times thus showing variations in the frequency.

The minimum length of detected palindromes was 8 bp (as lower limit set in the search tool) while the longest palindrome was 22 bp long (tttttttatatataaaaaata). The frequency distribution of palindromes with varying length is shown below (Fig. 8). As expected there is a downward trend of number of palindromes with their increasing length. The longest palindrome mentioned above consists entirely of As and Ts only. Most of the palindrome sequences had either 0 or very low G + C content. For this reason G + C content was determined for all the palindrome sequences and their frequency distribution was plotted (Fig. 9). It can be observed that most of the palindromes are indeed consisting of only As and Ts or very low G + C content. However there are many palindromes which have G + C content varying between 30% to 80%. These palindrome might play a direct role as TFBSs or converting into secondary structure such as hairpin or cruciform and thereby influencing gene expression. Further studies are required both at *in silico* as well as wet bench level to probe deeper into the role of motifs found in the promoters of potato patatin homologues and functional significance of the detected palindromes.

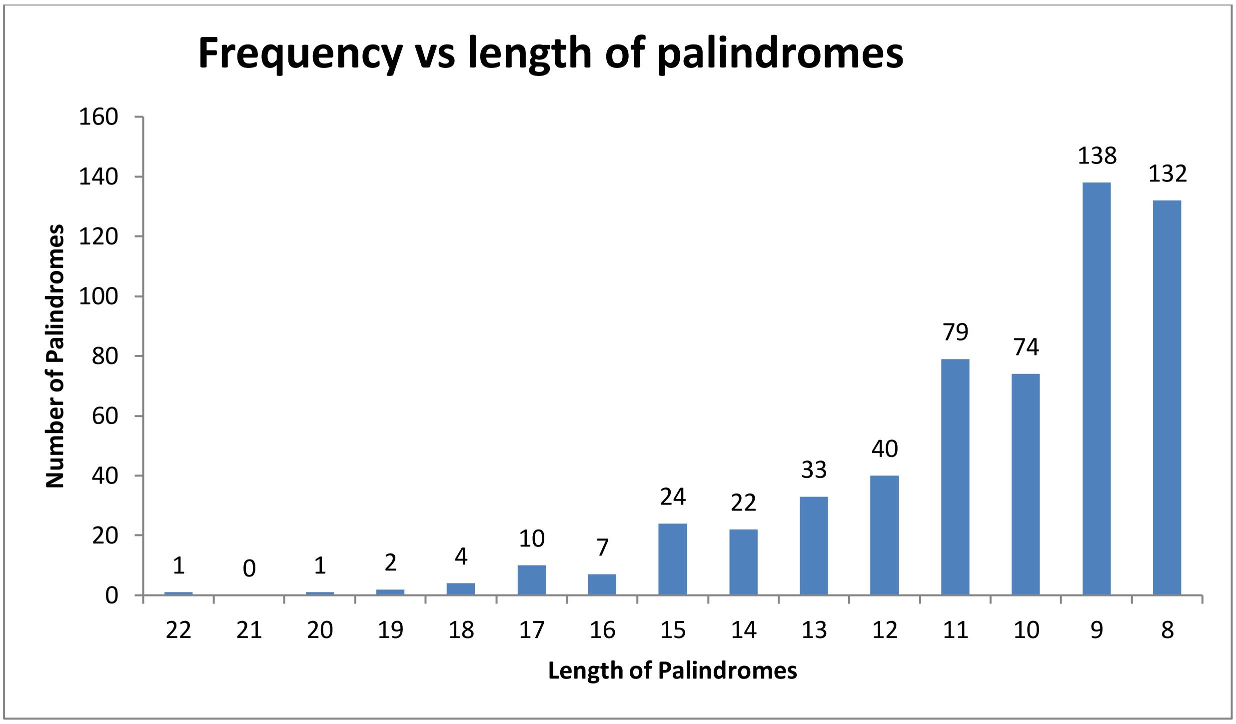


Fig. 8 Frequency of number of palindromes with their length. As shown from left, there is only one palindrome whose length is 22 nucleotides long whereas on the right side, there are palindromes having length of 8 nucleotides which are occurring 132 times.

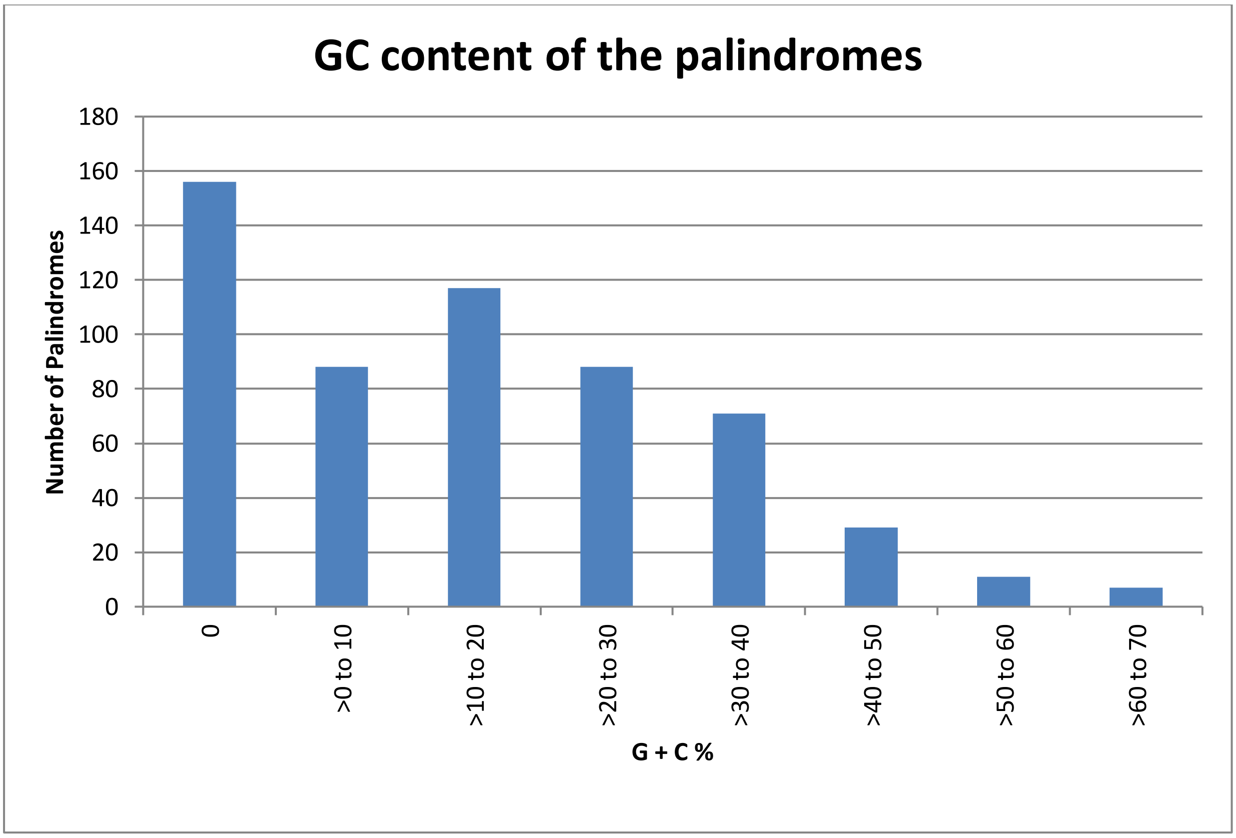


Fig. 9 GC distribution of Palindromes

CHAPTER 6 DISCUSSION

Patatin gene products play very important role during tuber development. Different isoforms exist in the potato genome and expression of some of the patatin isoforms is tuber specific. Therefore, it is interesting to analyze DNA sequence motifs that confer tissue specificity in potato. It is proposed to investigate various DNA sequence attributes such as presence of cis-elements that are known to be present in promoter/regulatory sequences. Novel putative elements may be searched based on sequence conservation by multiple sequence alignment. Additional attributes such as palindromic sequences are also studied. The basic idea of this study is to find common controls of regulation of these homologous genes. It was attempted to find unique regulatory mechanisms of gene expression among the gene homologues. This may improve our understanding of tissue-specific or temporal regulation of homologues of patatin gene expression which might be playing important role in physiology of potato.

In attempt to find the promoter sequences of all the homologues of patatin gene in potato, BLAST was performed. After removal of redundant hits, it yielded 46 distinct promoter sequences of patatin homologues. These sequences were subjected to Multiple Sequence Alignment (MSA) for detection of conserved regions. Following the comparative genomics based method, it was expected to find functionally important motif conserved in subset of the sequences analyzed. However only TATA box exhibited strong conservation and no other conserved sequence could be detected.

Then, *In silico* analysis of promoter of patatin gene homologues was done in search of regulatory motifs, the promoter sequences were analyzed using computational tools. The putative cis-regulatory sequence motifs were predicted by database screening tools namely PLACE and PLANTPROM which are used for searching for the presence of motifs to be found in plant *cis*-regulatory DNA elements. The regulatory motifs data of PLACE and PLANTPROM were analyzed in MS Excel to find the motifs commonly detected by both the tools. This method reduced the number of motifs detected by PLACE (4995 motifs) and PLANTPROM (741 motifs) to a score of 16 common motifs. Further analysis was focused on these 16 motifs. Five of the 16 common motifs were rare and their position on the gene sequence was determined. These motifs were searched in the scientific literature and were found to be present in Lemna gibba Lhcb21 gene (REBETALGLHCB21), RPL21 gene (S1FSORPL21), phenylalanine ammonia-lyase gene (PALBOXAPC), tomato Lhc gene (CIACADIANLELHC), beta-conglycinin (7S globulin) gene (SEF3MOTIFGM).

Since palindromic sequences are known to influence gene expression via various mechanisms such as cruciform structure affecting transcription, palindrome sequence acting as TFBS and formation of hairpin in primary transcript, the promoter sequences were screened for palindromes. In 46 promoter and regulatory regions of patatin homologues, 1633 palindromes were detected. Removal of duplicate sequences ended up in 568 different palindromic sequences. The most frequently present palindrome was occurring at 26 different positions. There were 346 unique panlidromes which were present in single copy only. The longest palindrome was 22 bp long and was consisting of As and Ts exclusively. However there were several palindromes which had much higher G + C content. It is interesting to study if any of these palindromes are actually affecting gene expression and those which do, what is the mechanism involved. Further work on these aspects may provide deeper insight into differential regulation of patatin homologues.

CHAPTER 7 CONCLUSIONS

.

Class I patatin gene promoters are meant as excellent molecular tools to study the regulation of plant gene expression. Current study is to investigate the DNA sequence attributes such as *cis*-elements/regulatory motifs are known to be present in promoter/regulatory sequences. In order to find the promoter sequences of all the homologues of patatin gene in potato, BLAST was performed. After removal of redundant hits, it resulted into 46 distinct promoter sequences of patatin homologues. These sequences were subjected to Multiple Sequence Alignment (MSA) for detection of conserved regions. Following the comparative genomics based method, it was expected to find functionally important motif conserved in subset of the sequences analyzed. However only TATA box exhibited strong conservation and no other conserved sequence could be detected.

In this study, *in silico* analysis of the promoters and other regulatory sequence motifs of the isoforms of patatin gene was done to search for novel putative elements based on sequence conservation by multiple sequence alignment. Five putative *cis*-regulatory sequence motifs were predicted by database screening tools namely PLACE and PLANTPROM to search for motifs found in *cis*-regulatory DNA elements and other additional characteristics such as presence of palindromic sequences. The results of such analysis can help us to predict the differential expression patterns of the patatin isoforms in potato.

Future prospects of the study

The *cis*-acting elements as predicted and analysed in the study will help to comprehend the molecular basis of regulation of patatin gene expression in potato. We can perform experiments by introducing some of the *cis*-acting elements found in the present study to know their effects for functional characterization of the patatin promoters and/or to ensure tuber-specific expression of the heterologous genes. This study would be useful to know the regulation and spatio-temporal nature of expression of patatin genes. Therefore, proper validation of the *cis*-regulatory motifs is an important aspect of future study. In other words, a number of *cis*-acting elements have been predicted but to what extent they are functional, we should know by functional characterization by reporter gene assay based experiments.

1. Aminedi, R., & Das, N. (2014). Class I patatin genes from potato (*Solanum tuberosum L.*) cultivars: molecular cloning, sequence comparison, prediction of diverse cis-regulatory motifs, and assessment of the promoter activities under field and in vitro conditions. *In Vitro Cellular & Developmental Biology-Plant*, *50*(6), 673-687.

2. Stupar, R. M., Beaubien, K. A., Jin, W., Song, J., Lee, M. K., Wu, C., & Jiang, J. (2006). Structural diversity and differential transcription of the Patatin multicopy gene family during potato tuber development. *Genetics*, *172*(2), 1263-1275.
3. Shewry, P. R. (2003). Tuber storage proteins. *Annals of botany*, *91*(7), 755-769.
4. Rocha-Sosa, M., Sonnewald, U., Frommer, W., Stratmann, M., Schell, J., & Willmitzer, L. (1989). Both developmental and metabolic signals activate the promoter of a class I patatin gene. *The EMBO journal*, *8*(1), 23-29.
5. Bárta, J., &Bártová, V. (2008). Patatin, the major protein of potato (*Solanum tuberosum L.*) tubers, and its occurrence as genotype effect: Processing versus table potatoes. *Czech Journal of Food Sciences*, *26*(5), 347-359.
6. Pikkard, C. S., Brusca, J. S., Hannapel, D. J., & Park, W. D. (1987). The two classes of genes for the major potato tuber protein, patatin, are differentially expressed in tubers and roots. *Nucleic acids research*, *15*(5), 1979-1994.
7. Bevan, M. (1991). Gene activity during tuber formation in the potato (*Solanum tuberosum*). In *Developmental regulation of plant gene expression* (pp. 75-93). Springer, Dordrecht.
8. Bevan, M., Barker, R., Goldsbrough, A., Jarvis, M., Kavanagh, T., & Iturriaga, G. (1986). The structure and transcription start site of major potato tuber protine gene. *Nucleic acids research*,*14*(11), 4625-4638.
9. Grierson, C., Du, J. S., Zabala, M., Beggs, K., Smith, C., Holdsworth, M., & Bevan, M. (1994). Separate cis sequences and trans factors direct metabolic and developmental regulation of a potato tuber storage protein gene. *The Plant Journal*, *5*(6), 815-826.
10. Jefferson, R., Goldsbrough, A., & Bevan, M. (1990). Transcriptional regulation of a patatin-1 gene in potato. *Plant molecular biology*, *14*(6), 995-1006.

11. Kloosterman, B., Vorst, O., Hall, R. D., Visser, R. G., &Bachem, C. W. (2005). Tuber on a chip: differential gene expression during potato tuber development. *Plant biotechnology journal*, *3*(5), 505-519.
12. Racusen, D., & Foote, M. (1980). A major soluble glycoprotein of potato tubers. *Journal of Food Biochemistry*, *4*(1), 43-52.
13. Köster-Töpfer, M., Frommer, W. B., Rocha-Sosa, M., Rosahl, S., Schell, J., & Willmitzer, L. (1989). A class II patatin promoter is under developmental control in both transgenic potato and tobacco plants. *Molecular and General Genetics MGG*, *219*(3), 390-396.
14. Mignery, G. A., Pikaard, C. S., & Park, W. D. (1988). Molecular characterization of the Patatin multigene family of potato. *Gene*, *62*(1), 27-44.
15. Naumkina, E. M., Bolyakina, Y. P., & Romanov, G. A. (2007). Organ-specificity and inducibility of patatin class I promoter from potato in transgenic arabidopsis plants. *Russian Journal of Plant Physiology*, *54*(3), 350-359.
16. Biłas, R., Szafran, K., Hnatuszko-Konka, K., &Kononowicz, A. K. (2016). Cis-regulatory elements used to control gene expression in plants. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *127*(2), 269-287.
17. Ibraheem, O., Botha, C. E., & Bradley, G. (2010). In silico analysis of cis-acting regulatory elements in 5′ regulatory regions of sucrose transporter gene families in rice (Oryza sativa Japonica) and Arabidopsis thaliana. *Computational biology and chemistry*, *34*(5-6), 268-283.
18. Priest, H. D., Filichkin, S. A., & Mockler, T. C. (2009). Cis-regulatory elements in plant cell signaling. *Current opinion in plant biology*, *12*(5), 643-649.
19. Hernandez-Garcia, C. M., & Finer, J. J. (2014). Identification and validation of promoters and *cis*-acting regulatory elements. *Plant Science*, *217*, 109-119.
20. Christ, A., Maegele, I., Ha, N., Nguyen, H. H., Crespi, M. D., &Maizel, A. (2013). In silico identification and in vivo validation of a set of evolutionary conserved plant root-specific *cis*-regulatory elements. *Mechanisms of development*, *130*(1), 70-81.
21. Higo, K., Ugawa, Y., Iwamoto, M., & Higo, H. (1998). PLACE: a database of plant *cis*-acting regulatory DNA elements. *Nucleic acids research*, *26*(1), 358-359.
22. Finkler, A., Kaplan, B., & Fromm, H. (2007). Ca2+-responsive cis-elements in plants. *Plant signaling & behavior*, *2*(1), 17-19.

23. Opabode, J. T., & Akinyemiju, O. A. (2015). Tissue-and organ-specific promoters for expression of heterologous genes in transgenic cassava (Manihot Esculenta Crantz) plants. *Gene Technol*, *4*, 125.
24. Wenzler, H., Mignery, G., Fisher, L., & Park, W. (1989). Sucrose-regulated expression of a chimeric potato tuber gene in leaves of transgenic tobacco plants. *Plant molecular biology*,*13*(4), 347-354.
25. Benfey, P. N., & Chua, N. H. (1989). Regulated genes in transgenic plants. *Science*, *244*(4901), 174-181.
26. Twell, D., & Ooms, G. (1987). The 5′ flanking DNA of a patatin gene directs tuber specific expression of a chimaeric gene in potato. *Plant molecular biology*, *9*(4), 345-375.
27. Rosahl, S., Schmidt, R., Schell, J., & Willmitzer, L. (1986). Isolation and characterization of a gene from Solanum tuberosum encoding patatin, the major storage protein of potato tubers. *Molecular and General Genetics MGG*, *203*(2), 214-220.
28. Wenzler, H. C., Mignery, G. A., Fisher, L. M., & Park, W. D. (1989). Analysis of a chimeric class-I patatin-GUS gene in transgenic potato plants: high-level expression in tubers and sucrose-inducible expression in cultured leaf and stem explants. *Plant Molecular Biology*, *12*(1), 41-50.
29. Sugiyama, A., Ikoma, Y., Fujii, H., Endo, T., Nesumi, H., Shimada, T., & Omura, M. (2017). Allelic diversity of phytoene synthase gene influences the transcription level in citrus fruit among a citrus F1 hybrid population. *Breeding science*, *67*(4), 382-392.
30. Zhang, J., Tian, J., Song, T., & Yao, Y. (2014, August). Cloning and characteristics of flavonol synthase promoter from Malus crabapple. In *XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014): VI 1106* (pp. 123-130).
31. Harris, N. N., Luczo, J. M., Robinson, S. P., & Walker, A. R. (2013). Transcriptional regulation of the three grapevine chalcone synthase genes and their role in flavonoid synthesis in Shiraz. *Australian journal of grape and wine research*, *19*(2), 221-229.
32. Cattani, A. M., Siqueira, F. M., Guedes, R. L. M., & Schrank, I. S. (2016). Repetitive elements in Mycoplasma hyopneumoniae transcriptional regulation. *PloS one*, *11*(12), e0168626.

33. Brázda, V., Čechová, J., Battistin, M., Coufal, J., Jagelská, E. B., Raimondi, I., & Inga, A. (2017). The structure formed by inverted repeats in p53 response elements determines the transactivation activity of p53 protein. *Biochemical and biophysical research communications*, *483*(1), 516-521.
34. Jerabek, S., Ng, C. K., Wu, G., Arauzo-Bravo, M. J., Kim, K. P., Esch, D., & Yang, X. (2017). Changing POU dimerization preferences converts Oct6 into a pluripotency inducer. *EMBO reports*, *18*(2), 319-333.
35. Miura, O., Ogake, T., & Ohyama, T. (2018). Requirement or exclusion of inverted repeat sequences with cruciform-forming potential in Escherichia coli revealed by genome-wide analysis. *Current genetics*, 1-14.