CHARACTERIZATION OF AP-4 COMPLEX PROTEIN IN GUAR AND OTHER PLANTS

Ph.D THESIS

by

NISHU MITTAL

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247667 (INDIA)
JULY, 2018

CHARACTERIZATION OF AP-4 COMPLEX PROTEIN IN GUAR AND OTHER PLANTS

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

NISHU MITTAL

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247667 (INDIA)
JULY, 2018

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**CHARACTERIZATION OF AP-4 COMPLEX PROTEIN IN GUAR AND OTHER PLANTS**" in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee, Roorkee is an authentic record of my own work carried out during a period from July, 2013 to July, 2018 under the supervision of Dr. G.S. Randhawa, Professor, Department of Biotechnolgy, Indian Institute of Technology Roorkee, Roorkee.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

(**NISHU MITTAL**)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

(G.S. RANDHAWA)
Supervisor

Date: _____

ABSTRACT

AP-4 complex plays an important role in intracellular protein transport carried out by vesicular trafficking mechanism. The first comprehensive study in plants on the characterization of genes encoding AP-4 complex has been presented here. These genes were characterized for introns, CDS, GC content, CpG islands, conserved motifs, *cis* regulatory elements (CREs) and SNPs. Two conserved motifs, one in the mu subunit for binding to the sorting signal of the cargo protein and the other in the beta subunit for binding to the accessory protein, were observed in each studied AP-4 complex. CREs that regulate stress response, cellular development and hormonal regulation were found in the promoters of AP-4 complex genes. Out of about 1376 SNPs observed in these genes, 31 were found in the functional binding pockets of proteins of *Zea mays* and soybean. Out of these 31 SNPs 6 were showing damaging effects. Transcriptome analysis of *Arabidopsis thaliana, Z. mays* and *Glycine max* showed that the genes coding for four subunits of AP-4 complex express in various tissues, however, the expression was significantly more in seed and other reproductive tissues. The RNA-Seq analysis of genes coding for ARFs, Epsins and VSR1 accessory proteins in all studied tissues of *A. thaliana* and *Z. mays* showed expression in all studied tissues, however, their expression like AP-4 complex genes was high in seed. The expression of AP-4 complex genes in different tissues of guar, *Z. mays* and *G. max* while ARFs, epsin, and VSR1 in studied tissues of *Z. mays* was further confirmed by real time PCR analysis. Among the 15 studied genes, genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1E, ARFA1F, epsin2 and VSR1 were found to co-express in anther while AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1E, ARFA1F, ARFB1B, epsin 1 and epsin 2 proteins were found to co-express in the seed tissue of *A. thaliana*. In *Z. mays*, genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARF3, epsin 1 and VSR1 proteins were observed to co-express in the anther tissue whereas whereas the genes encoding AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARF3, epsin 1, epsin 2 and VSR1 proteins were the co-expressing genes in seed. The protein-protein interaction studies of the AP-4 complex and its accessory proteins generated a protein network containing 39 interacting combinations. Molecular docking revealed that Ile 49 residue of ARFA1A is critical for interaction with epsilon subunit in maize. These results indicated a role of the AP-4 complex and its accessory proteins in the development and functioning of various tissues, especially reproductive tissues.

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my supervisor, Prof. G.S. Randhawa, Professor, Department of Biotechnology respectively, IIT Roorkee for his support, motivation and continuous guidance in completion of this research work. I express my deep sense of indebtedness for Prof. G. S. Randhawa for his scholarly guidance, inspiring suggestions and constructive criticism. Words are inadequate to express my deep sense of gratitude for his invaluable guidance. I express my deep sense of indebtedness for Prof. G. S. Randhawa for his scholarly guidance, inspiring suggestions and constructive criticism. Words are inadequate to express my deep sense of gratitude for his invaluable guidance. I would like to take this chance to also thank Dr. Surinder Pal Kaur Randhawa who has been supportive throughout my Ph.D.

I would like to acknowledge my deep sense of gratitude to the present Heads of the Department Prof. A.K. Sharma and former Heads Prof. R. Prasad ad Prof. Partha Roy for their vision and support. I extend my sincere thanks to my committee members, Prof R. Prasad (Chairman SRC), Prof. A.K. Sharma (Internal Expert) and Dr. Anuj Sharma (External Expert) for their valuable and constructive criticism on my research work. I also express my gratitude to rest of the faculty members of Department of Biotechnology, IIT Roorkee for their support, guidance and timely help in research.

I would like to extend my thanks to my seniors Dr. Pallavi Gahlot, Dr. Swati Verma, Dr. Navneet Kaur, Dr. Shalini pareek, Dr. Shilpi Kumari, Dr. Umesh Tanwar, Dr. Deepa Dewan for giving me valuable suggestions and emotional support during the entire course of my work. I would like to give special thanks to my friend and labmate Poonam S. Jaiswal for working with me and creating a friendly environment in our lab. A special thanks to Dharmendra Singh, Shipra and Mamtesh Kumari for their critical analysis on my research work and help in thesis writing. I would like to stretch my appreciation to my labmates Omika Thakur, Manjeshree Shail and Neha who incented me to strive towards my goal. I am thankful to M.Sc and B.Tech students Tsering Angmo, Parvendu Pradeep and Santosh for helping me in my research project.

I am grateful to my colleagues and friends Anjali Malik, Benazir Fatma Khan, Dr. Khusboo Gulati for providing me homely environment in hostel and standing with me during my tough time. I am thankful to my batchmates Jyoti Singh, Sonam Gupta, Deepa Teotia, Mukta Singh and Timsy Bhando for their support and help.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future, for supporting me in every walk of life. My heart is full of love for my sister Kaffi Mittal and brother Ankush Mittal who are the source of my happiness. I am very much thankful to my in-laws family, specially my husband Mr. Dhananjay Singh for his constant love, care and understanding, for being my strength when I needed it the most to complete this research work.

I would also like to acknowledge the financial support provided by Counsil of scientific & industrial research (CSIR), Human resource development group in the form of JRF and SRF during the course of this research.

Finally, I would like to praise and thank the Almighty, for his showers of blessings throughout my research work to complete the research successfully.

Date NISHU MITTAL

ABBREVIATIONS

°C	degree centigrade
Å	Ångström
A.	*Arabidopsis*
APS	ammonium per sulphate
ATP	adenosine triphosphate
AP	adaptor protein
BLAST	basic local alignment search tool
bp	base pair
C.	*Cyamopsis*
CAZRI	central arid zonal research institute
cDNA	complementary deoxyribonucleic acid
CLUSTAL O	clustal omega
cm	centimetre
CRE	*cis*-regulatory elements
D	dimensional
Da	Dalton
DAF	days after flowering
DEPC	diethylpyrocarbonate
DNA	deoxyribose nucleic acid
DNase	deoxyribose nuclease
E. coli	*Escherichia coli*
e.g.	example
EDTA	ethylenediaminetetracetic acid
et al.	*et alia*
FASTA	FAST All
Fig	figure
FOR	forward
FTIR	Fourier transform infrared spectroscopy
g	gram
G.	*Glycine*
GDP	guanosine diphosphate
Gln	glutamine
Glu	glucose
GRAVY	grand average hydropathicity
h	hours
His	histidine
IEP	isoelectric point
IMAC	immobilized metal ion affinity chromatography
IPTG	isopropyl β-D-thiogalactoside
IARI	Indian agricultural research institute
Kbp	kilo bases
kDa	kilo Daltons
l	liter
L.	Linnaeus
Lys	lysine
m	meter
M	molar
MCMC	Markov chain Monte Carlo
M/G ratio	mannose/galactose ratio

MEGA6	molecular evolutionary genetics analysis tool 6
MEME	multiple em for motif elicitation
min	minutes
ml	millilitre
mM	millimolar
mRNA	messenger RNA
MYA	million years ago
Nacl	sodium chloride
NCBI	national center for biotechnology information
NJ	neighbour joining
nm	nanometer
ORF	open reading frame
PCR	polymerase chain reaction
PDB	protein data bank
Phe	Phenylalanine
pI	isoelectric point
PyMol	python enhanced molecular graphics tool
qRT-PCR	quantitative real time PCR
REV	reverse
RNA	ribose nucleic acid
RNaseA	ribonucleaseA
rpm	revolutions per minute
s	seconds
SAVES	structure analysis and verification server
Ser	serine
SNP	single nucleotide polymorphism
Taq	*Thermus aquaticus*
TBE	tris borate EDTA
TE	tris EDTA
Tm	melting temperature
Tris	*tris*-(hydroxymethyl)-aminomethane
Trp	tryptophan
Tyr	tyrosine
U	unit
U.S.A./USA	United States of America
UK	United Kingdom
UV	ultravoilet
V	volt
v/v	volume / volume
V_e	elution volume
V_o	void volume
w/v	weight / volume
w/w	weight / weight
Z.	*Zea*
α	alpha
β	beta
γ	gamma
µ	mu
δ	delta
ζ	zeta
ε	epsilon
µl	microlitre

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Proteins are important biomolecules which play structural and functional roles in living cells. Several proteins, after synthesis on endoplasmic reticulum (ER), are transported to their respective destinations in a cell. Some specific proteins are secreted out of a cell (exocytic/secretory pathway) whereas some others are taken up from outside (endocytic pathway) and transported to their respective destinations in a cell. Like other eukaryotes, plant cells are compartmentalized into different organelles. Interaction among different organelles in a cell takes place mostly by vesicular trafficking process which includes recognition and encapsulation of the cargo protein, recognition of destination organelle membrane, and the fusion of vesicle with target membrane to release the cargo protein into the destination organelle [40]. The mechanism of vesicular trafficking involves the participation of several proteins, like, COPI, COPII, adaptor protein (AP) complexes and other accessory proteins, namely, tepsin/epsin(s), ADP-ribosylation factors (ARFs) and vacuolar sorting receptors (VSRs).

The AP complexes play a significant role in sorting of proteins and vesicle membrane formation. Five AP complexes, namely, AP-1, AP-2 [129], AP-3 [2, 129], AP-4 [60, 107] and AP-5 [106] have been identified in the living cells. The AP complexes are heterotetramers; each consisting of two large (γ/α/δ/ε/ζ and β1-β5) one medium (μ1-μ5) and one small (σ1-σ5) subunits. Each AP complex has three domains, namely, a core, two appendages (ears) and two hinges. The core domain is made up of the N-terminal domains of two large subunits and the complete mu and sigma subunits. The two appendage domains are constituted by C-terminal domains of two large subunits. A core domain and an appendage domain are linked by a hinge domain.

Each AP complex has a specific location(s) and function(s). The AP-1 and AP-3 complexes are localized on *trans*-Golgi network (TGN) and endosomes. AP-1 complex helps in the synthesis of secretory granules [37] and bidirectional movement of proteins between TGN and endosomes [246, 275 whereas AP-3 complex transports proteins from early to late endosomes [206, 257]. The AP-2 complex is located on plasma membrane and mediates endocytosis of several proteins including receptors [109]. The AP-4 complex is located on TGN and it transports proteins from TGN to endosomes [38]. The AP-5 complex is present on late endosomes and it mediates transport of proteins from endosomes to other membrane-bound compartments [106]. The AP-1, AP-2 and AP-3 complexes associate with accessory protein clathrin whereas AP-4 and AP-5 complexes make association with other accessory proteins (ARFs, VSRs and Tepsin/Epsins).

Based on their amino acid sequences, protein sizes, gene structures and phylogenetic analysis, the ARFs are subdivided into three classes: class-I (ARF1, ARF2, and ARF3), class-II (ARF4 and ARF5), and class-III (ARF6) [262]. Each ARF has two conserved regions which have been named as switch I and switch II regions. They control the assembly of various vesicle coat components, such as COPI and adaptor complexes, on the Golgi apparatus [224, 116, 26]. The ARFs recruit AP-1 and AP-3 complexes to TGN and/or endosomal membranes [64], and AP-2 complex to plasma membrane [135]. The ARF1 regulates the localization of AP-4 complex onto TGN via interacting with its epsilon (ε) and mu (μ) subunits [26].

In plant cells, vacuoles are essential multi functional organelles which store water, ions, proteins and several secondary metabolites. Two types of vacuoles, protein storage vacuoles (PSVs) and lytic vacuoles (LVs) are present in plant cells. The transport of a vacuolar protein to its destination vacuole takes place via vesicle mediated pathway. A main step of this pathway at the destination stage is the recognition of a specific signal on the protein by a vacuolar sorting receptor (VSR) [285]. These VSRs, belonging to the well studied receptor family [57], help in sorting of storage proteins and lytic cargos to PSVs [236, 81] and LVs [303], respectively. In *Arabidopsis thaliana*, VSR1 interacts with epsin protein [124]. The epsin protein possesses a well conserved epsin N-terminal homology (ENTH) domain and disorganized C-terminal end [46, 56]. It binds to phosphatidylinositols (Ptdlns) and cargo proteins with its ENTH domain [180] and ubiquitin-interacting motifs [191], respectively and plays a role in transport of proteins to lysosomes [124, 245].

The presence of genes coding for AP-4 complex in plants was first reported in 2001 in *A. thaliana* [24]. The mu4 subunit of AP-4 complex has been found to interact with the tail of vacuolar sorting receptor (VSR) in tobacco [86]. Fuji *et al*. [82] identified the four subunits of AP-4 complex of *A. thaliana* and demonstrated that mu subunit of this complex recognizes the vacuolar sorting receptor 1 (VSR1) which is a receptor for sorting of seed storage proteins to target specific proteins to the protein storage vacuoles (PSVs). The mutations in the genes coding for each of the four subunits of AP-4 complex resulted in defective vacuolar sorting of the major storage protein 12S globulins, suggesting a role of this complex in transport of proteins to vacuoles.

The above reports show that a comprehensive study of AP-4 complex genes in higher plants, especially legumes and cereals, has not been done so far. Hence, the present work was undertaken to carry out the structural, evolutionary, SNP genotyping and expression analyses of genes coding for AP-4 complex in various plant species with special emphasis on cultivated crop plants namely, guar, soybean and maize. The genes coding for proteins involved in a pathway in an organism are expected to express simultaneously or almost simultaneously. The

co-expression analysis of genes has emerged as an important tool to understand biological pathways at molecular level. Still no report is available on the co-expression analysis of plant genes coding for proteins involved in AP-4 complex-mediated vesicular trafficking. The interactions among various proteins involved in this pathway have also not been studied in plants so far. Therefore the present work was undertaken with the following objectives:

- To identify the genes coding for AP-4 complex proteins of guar
- Characterization of genes coding for AP-4 complexes of guar, maize and soybean
- To study the expression of genes coding for AP-4 complex subunits of guar, maize and soybean
- To study co-expression analysis of genes involved in AP-4 complex- mediated vesicular trafficking of maize
- *In silico* analysis of AP-4 complex proteins

2. REVIEW OF LITERATURE

The literature on work done on AP-4 complex and its accessory proteins in clusterbean, maize and soybean has been presented below:

2.1. Origin of clusterbean

Gillet [89] reported that Africa was probably the center of origin of clusterbean (*Cyamopsis tetragonoloba*). Poats [211] and Hymowitz [115] reviewed the cultivation history of clusterbean. In India, domesticated *C. tetragonoloba* was developed from the *C. senegalensis*, a drought tolerant wild African species [115]. *C. senegalensis* species was brought as fodder to the Asian countries from the African desert by Arab traders between 9^{th} -13^{th} century A.D. [115]. The clusterbean was introduced in USA from India in 1903 [114]. During World War II, guar crop came into notice as the gum produced in its seed was found to be an alternative of commercial industrial gum. Gillett [89] recognized three *Cyamopsis* species, namely, *C. tetragonoloba*, *C. senegalensis*and and *C. serrate*. A further boost to the clusterbean cultivation came from the increased demand of clusterbean gum in oil and methane extraction. Guar crop is cultivated in India, Pakistan, USA and in some stretches of Italy, Morocco, Spain, Greece, France and Germany [214, 97].

2.2. Genetic studies in clusterbean

Clusterbean has not been well characterized like other legumes. The work on the genetics and breeding of clusterbean has been reviewed in two reports [138, 9]. The diploid chromosome number of guar is 7 [13]. Two new crossing techniques in guar have been reported [44, 88], but hybridization is still very tricky as flower is small and cleistogamous in nature [138]. Chaudhary and Lodhi [43] studied the inheritance of five characters, namely branching behaviour, clustering pattern, growth habit, leaf size and hairiness in clusterbean. Except branching behaviour remaining four characters were controlled by a single pair of genes. Total 68 genotypes of guar have been characterized based on morphology and biological study by Gresta *et al*. [96]. Pathak *et al*. [203] investigated genetic diversity and interrelationship studies among 40 genotypes of *C. tetragonoloba*. A significant amount of variability for 7 characters was found in the genotypes. These researchers showed that the genotypic variation in the studied guar material was due to high supplement genetic effects.

In genetic diversity study of guar, five molecular markers, namely, randomly amplified polymorphic DNA (RAPD), ribosomal DNA (rDNA), inter simple sequence repeats (ISSR), simple sequencing repeats (SSR) and sequence characterized amplified region (SCAR) markers have been used in the study of genetic diversity in guar [141, 142, 203, 214]. RAPD genetic

markers were introduced in 1990 [281]. RAPD markers have been used at large scale to study polymorphism in plants [91]. Several workers have studied the genetic diversity by RAPD markers in commercially important cultivars of guar [214, 204, 141] These workers found a significant polymorphism among the cultivars and indicated a significant genetic diversity among the tested genotypes [204]. In a separate study, 35 genotypes of guar collected from different states of India were analyzed for genetic diversity with RAPD, ISSR and SCAR markers. SSRs genetic markers are commonly highly polymorphic, genome specific, abundant and co-dominant in nature [92]. The genetic diversity in 3 commercial varieties and two wild species of guar was studied using 224 EST-SSR markers [142]. The results indicated very less genetic variation in varieties of guar. Similar study conducted on 32 genotypes of guar using EST-SSR markers showed very little genetic diversity among guar varieties [137]. Tanwar *et al.* [253] reported potential 3594 high quality SNP for the first time in guar.

2.3. Guar gum and its applications

Guar seed consists of endosperm, hull and germ. The germ and hull of the guar seeds together also called as guar meal. Guar meal is protein rich meal formed as a by-product during the synthesis of guar gum. Galactomannan and seed storage proteins (albumins) build up in mature guar seeds and form a major part of dry seed weight. More than 90% (w/w) of the endosperm of guar seed contained galactomannan, a commercial gum [61]. At the time of guar seed development galactomannan accumulate as secondary wall thickenings in the endosperm [265]. Galactomannan forms about 80% of the total seed endosperm. Guar gum is most commonly used commercial hetero-polysaccharide with an M/G ratio of approximately 2 among the other galactomannan [165]. Several uses of guar gum and its derivatives have been reported in petroleum, textile, explosive, paper, coal mining, analytical, food and pharmaceutical industries [138,140].

2.4. Origin and evolution of maize

The history about the origin of maize was reviewed by Galinat [84]. The commonly used maize was selected either from its close relative *Zea maxicana* [18] or from an ancestor [217]. Maize was originally grown in the mesoamerican region of Mexican highlands and spread briskly to other parts of the world. Sleper and Poehlman [241] revealed that after wheat and rice, maize is the third leading cereal crop in the world. The study, based on phylogenetic analysis showed that maize was domesticated about 6,000 years ago [210].

2.5. Genetic studies in maize

The *Zea mays*, corn or maize belongs to family Poaceae. The haploid chromosome number is 10 [259]. The size of maize genome is approximately 2.3-2.7 Gbp [10] and consists of 42,000 to 56,000 genes. Maize flowers are monoecious in nature. Hake and Walbot [103] revealed that the maize genome consists of high amount of repetitive sequences, including both transposons and retrotransposons [158]. Brunner *et al*. [35] studied the noncolinearities phenomenon by analyzing the DNA sequences of two maize inbred lines Mo17 and B73. The genetic diversity in maize accessions was studied using isoenzymes [146] and RFLP [151, 20], RAPD [41] microsatellite [159] and SSR [225] markers. The RAPD markers were used in maize by several workers to find the genetic distance between segregant lines [170] explore genetic diversity [182] and find out the best crossing material for crop breeding programme [147]. The SNPs and InDels in maize genome have been identified by Several workers [48, 80, 36]. These SNPs were used to depict arrays/chips for genotyping. Unterseer and co-workers [263] developed high density 600 K Affymetrix® Axiom® Maize Genotyping. The arrays/chips designed from SNPs have been used for specifying the genetic properties of diverse maize populations, fingerprinting important germplasms [160, 258], executing genome wide association studies and QTL mapping [51,157,192].

2.6. Applications of maize seeds

Maize seeds consist of endosperm and embryo which formed a major part of the dry weight of whole seed [76]. Starch and proteins are stored in endosperm while oil in embryo. Protein forms about 10% part of endosperm and categorized as seed storage proteins [76]. The endospermic proteins based on their solubility in various solvents were classified into different groups: albumins, globulins, glutamines, and prolamins. Maize starch is commonly used in several industries like paper, fermentation and for the production of methanol, and fuel. During starch synthesis several byproducts like oil and proteins are also synthesized and used in food industry [30]. Pena-Rosas and co-workers [207] explained that to improve the nutritional value, fortification of maize was done in several countries.

2.7. Origin of soybean

The soybean is an annual crop and belongs to Fabaceae family. The seeds of soybean are a rich source of proteins and have been used as food products. The morphological, cytogenetic and molecular studies have revealed that the present variety of soybean, *Glycine max* was domesticated in China from wild variety *G. soja* [32]. The *G. soja* is distributed geographically in East Asia, Russia, the Korean Peninsula, Japan and extensive regions of China [240]. *G. max* has been mostly grown in warmer regions of the world. The germination of seed has

positive correlation with growth of planted seedling in a field [90]. The cultivars of soybean have been propagated in USA and Canada [58, 228], China [55], Brazil [181], India [230], and Japan [298] have low genetic diversity. Morse [185] reported that initially soybean was used as fodder, vegetable and grain. In 1949, focus was shifted to the development of high yielding soybean varieties having high oil and protein content. Soybean alone can fulfill the need of vegetable oil for humans. The oil obtained from seeds of soybean is used in several industries like food, cosmetics, textiles, pharmaceutical, leather, plastic, petroleum and detergents [184].

2.8. Genetic studies in soybean

The genome of soybean was sequenced by using a whole genome shotgun approach [231]. The genome size of soybean (*Glycine max* L. Merr.) is 1.1–1.15 Gb [10]. The diploid chromosome number is 20. Inbreeding and domestication leads to the reduction of genetic diversity. The genetic diversity found in any organisms is adaptive sometime [122]. The level of genetic diversity in germplasm of soybean was usually determined by analyzing various morphological characters [14, 167]. Although, the majority of these characters are tremendously affected by environmental conditions. SSRs markers have been used for the analysis of genetic diversity in many legume crops such as soybean [34, 45], finger millet [200] pigeonpea [70], cowpea [11], mungbean [101], common bean [39], barley [188]. Appiah-Kubi and co-workers [8] reported SSR markers in 36 accessions of soybean grown in different geographical regions. The genetic diversity among 90 Indian soybean cultivars grown in geographical different regions has been studied by SSR markers. RAPD markers were used to study the genetic diversity among soybean accessions selected for breeding technique to develop high protein yielding varieties [63,113,166]. The first genetic linkage map of molecular markers was developed by Keim *et al.* [130] in soybean. Total 20 molecular linkage groups in soybean were developed by RFLP and SSR markers [244]. The analyses of genetic diversity and relatedness among the studied plant materials are key components for the development of novel germplasm [3].

2.9. Transport of proteins by vesicular trafficking

Proteins, one of the main constituents of living cells, play important structural and functional roles in the cellular machinery. In a living cell different types of proteins are present. After synthesis on ribosomes present in the cytoplasm, proteins are transported to their respective destinations in the cell which may be cell wall, vacuole or other compartments of the endomembrane system [269]. Like other eukaryotes, plant cells are compartmentalized into different organelles. Membrane trafficking comprises of endocytosis and exocytosis (secretory) pathways. The endocytosis takes place at the plasma membrane (PM) while exocytosis from the endoplasmic reticulum (ER) via the Golgi apparatus to the plasma membrane [198]. The

degradation of cargos and receptor (proteins) takes place via endosomes to lysosomes. The protein synthesized on ribosomes recognizes specific sorting signals found at the membrane, then move into the endoplasmic reticulum for proper folding and transported to its final destination through *trans* Golgi network (TGN) [268]. One of the mechanisms for this intracellular protein transport is vesicular trafficking. Interaction among different organelles in a cell takes place mostly by vesicular trafficking; process which includes the formation of vesicle bud from a donor organelle/ membrane (vesicle budding) in which addition of selective cargos to be transported occurred while the resident protein remain in donor compartment (protein sorting). These filled vesicles further targeted to acceptor organelle/ membrane where vesicle fuses with target membrane and release cargo protein into the desitnation organelle (vesicle fusion) [40]. In vesicular trafficking mechanism the transport of cargo proteins take place in specialized structures known as vesicles. Size of these vesicles ranged from 40nm - 100nm in a cell. Vesicles are membrane bound and coated with proteins on their cytosolic surface.

Vesicular trafficking mechanism involves the participation of several proteins, like, clathrin, COPI, COPII, adaptor protein (AP) complexes and accessory proteins, namely, ADP-ribosylation factors (ARFs), tepsin/epsin(s), vacuolar sorting receptors. Kirchhausen [133] reported that COPI, COPII (COP stands for coat protein complex) and clathrin coat proteins are involved in the encapsulation of the vesicles. COPII-coated vesicles are involved in anterograde transport from the early Golgi [15] whereas the COPI-coated vesicles take part into both intra-Golgi cisternal transport and retrograde traffic from the Golgi to the ER [155]. Clathrin is used at different locations, including the *trans* Golgi network (TGN), endosomes and the plasma membrane in combination with specific adaptor proteins and thus may contributes to both the secretory and endocytic pathways.

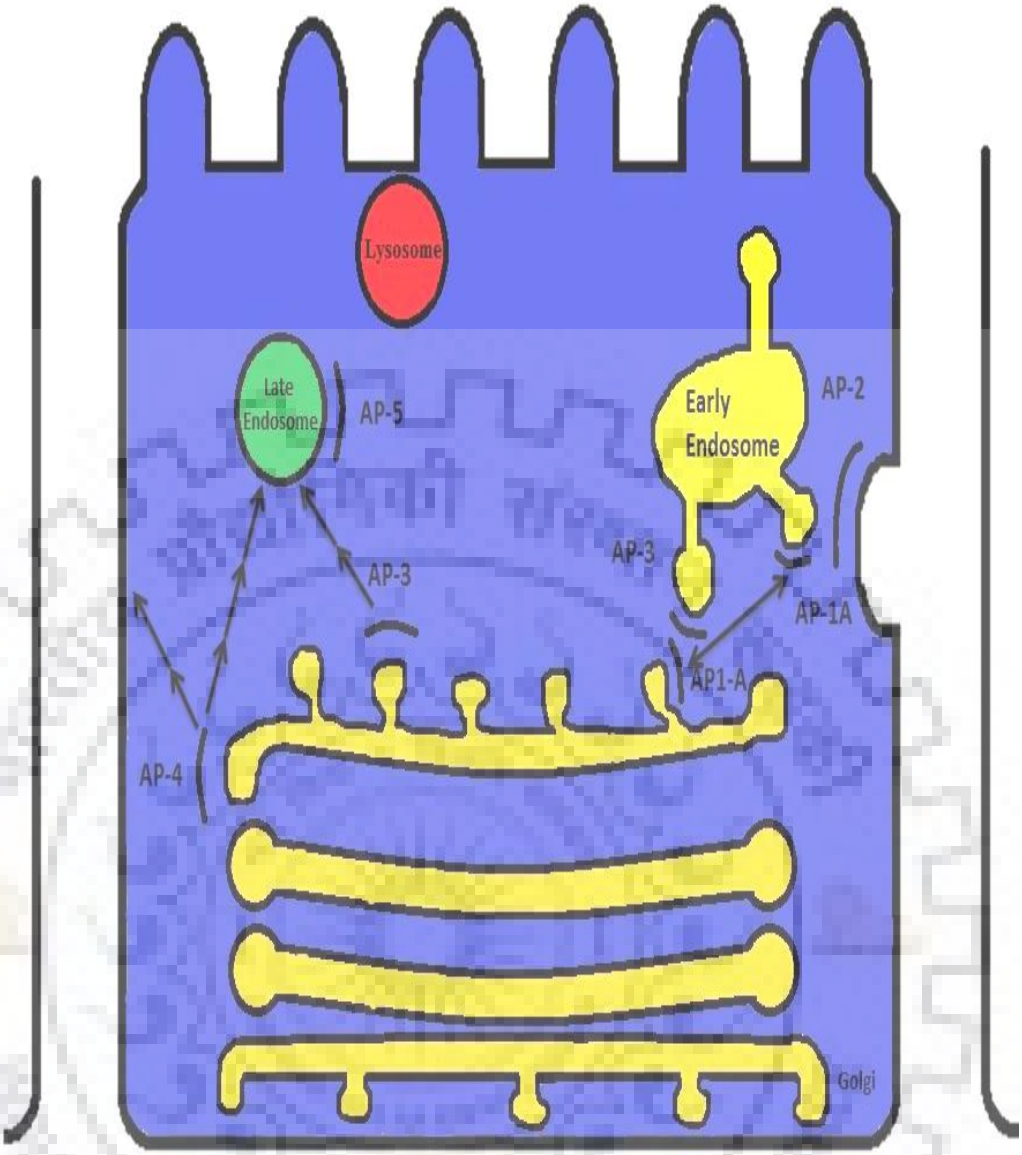


Fig. 2.1. Scheme of vesicular transport pathway in a living cell by adaptor complexes.

2.10. Clathrin Coat

In receptor regulated endocytosis clathrin coated vesicles (CCV) play an important role. Kirchhausen [133] reviewed that CCV takes part in the trafficking of cargos from TGN to endosome/ lysosomes. Pearse in 1975 purified clathrin for the first time [205]. Gupta and Gray in 1999 identified a protein in *A. thaliana*, which showed sequence similarity with the N-terminal end of clathrin assembly AP180 protein [100]. The CCV structure is three layered in nature and made up of an outer layer of clathrin, an inner layer of membrane and middle layer of adaptor proteins. Clathrin is made up of three heavy (192kDa) and three light chains (30kDa). Each heavy chain is attached to any of two light chains and forms a triskelion structure [222]. Fotin and co-workers [78] gave the crystal structure of clathrin using cryo-EM

and revealed the assemblies of heavy and light chains triskelia to form the clathrin lattice. A vesicle coated with clathrin must lose its coat to fuse properly with target membrane like other vesicles, coated with COPI and COPII.

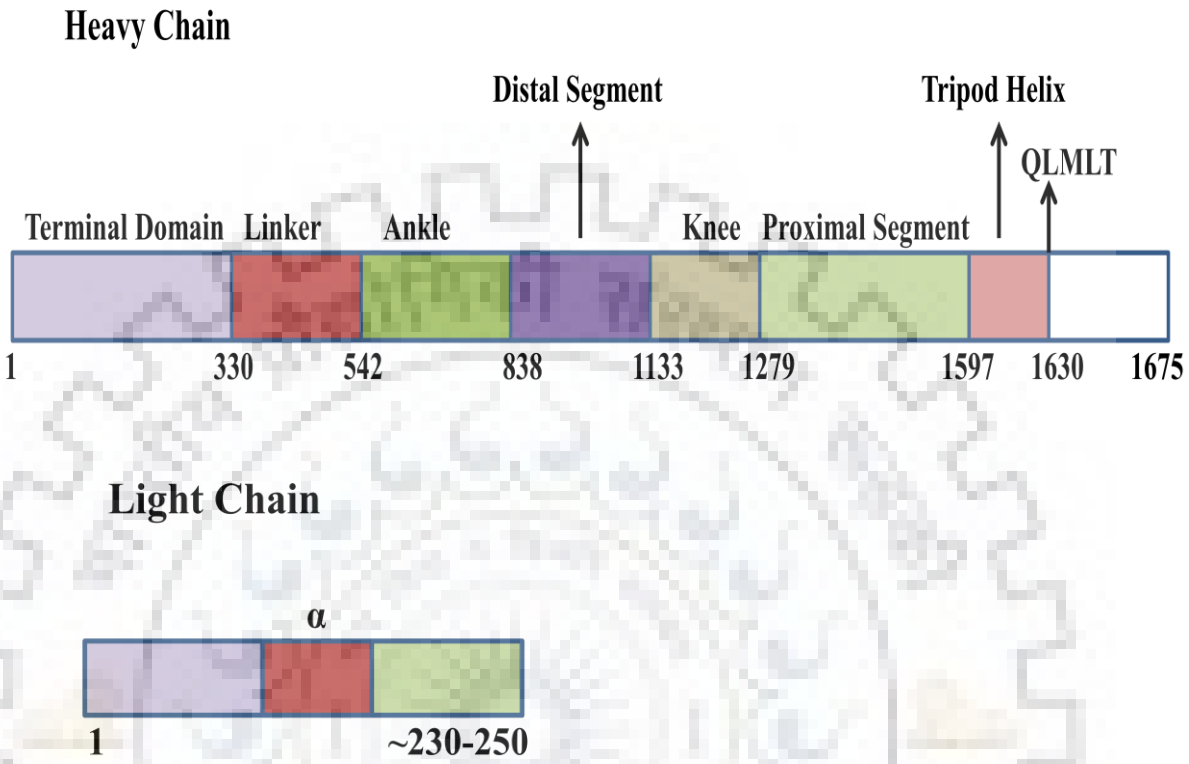


Fig. 2.2. Structural organization of the heavy and light chains of the clathrin protein. Modified from Kirchhausen *et al*., 2014 [134].

2.11. Coat protein complex II (COPII)

The COPII conserved protein helps in the formation of vesicles from the membrane of ER. Barlowe *et al*. [15] reported that COPII coats are made up of five cytosolic subunits, namely, Sec23/24, Sec13/31 and Sar1. The COPII coated vesicle synthesis required GTPase in its active form which leads to addition of structural elements of the vesicle coat to the donor membrane. Formation of vesicle at the donor membrane initiates with the recruitment of the Sar1 a small GTPase along with the Sec12p substituting as a guanine nucleotide exchange factor (GEF) at ribosome-free ER membranes in the cytosol [15, 229]. Eventually other coat proteins assembled and leads to the formation of coated vesicle which bud from mother membrane and reach to its destination organelle. COPII-coated vesicles are involved, especially in the transport of secretory proteins from their site of synthesis i.e ER to Golgi [153, 21]. The distribution of COPII isoforms and role in the transport of proteins in plants was reviewed

by Marti and co-workers [171]. These workers showed the distribution of COPII isoforms with the help of confocal imaging.

2.12. Coat protein complex I (COPI)

Coat protein complex I (COPI) also known as coatmers, composed of two subcomplexes, namely, FCOPI and BCOPI. The FCOPI is consists of β, δ, γ and ζ subunits while BCOPI contains α, β' and ε subunits [177]. These subunits of both subcomplexes show structural similarity with the clathrin and adaptor proteins (AP). The domains of β and γ subunits of FCOPI are similar to the appendage domain of α subunit of AP2 [108, 278]. Andag and Schmit [6] suggested the similarity between the δ and μ subunits of COPI and adaptor protein complex, respectively. The COPI vesicles mostly regulate protein transport within the Golgi network. Coatmers sometime also involved in retrograde transport from Golgi complex to ER [52, 177]. The ARF1 member of class I in its active form, regulates the localization of COPI like Sar1 required for the recruitment of COPII coat protein. The β and γ subunits of COPI associate directly with the membrane bound active form of ARF. The dissociation of COPI from ARF at the time of vesicle uncoating depends upon the hydrolysis of GTP [93]. Both Sar1 and ARF1 acts as stimuli for the maturation of both COPII and COPI coated vesicles in a cell, respectively [152].

2.13. Adaptor proteins

Pearse (1975) gave the term "adaptin" to the 100kDa proteins which was purified along with clathrin coated vesicles [205]. These 100kDa proteins were found to be part of adaptor protein (AP) complexes. The term adaptin was finally expanded to five AP complexes. Each AP complex plays a significant role in sorting of proteins and vesicle membrane formation. Five AP complexes, namely, AP-1, AP-2 [129], AP-3 [2, 129], AP-4 [60,107] and [106] AP-5 [106] have been identified in the living cells. Each AP complex has a heterotetrameric structure consisting of two large subunits (~90-130kDa), one medium subunit (~50kDa) and one small subunit (~20kDa) [156, 133].

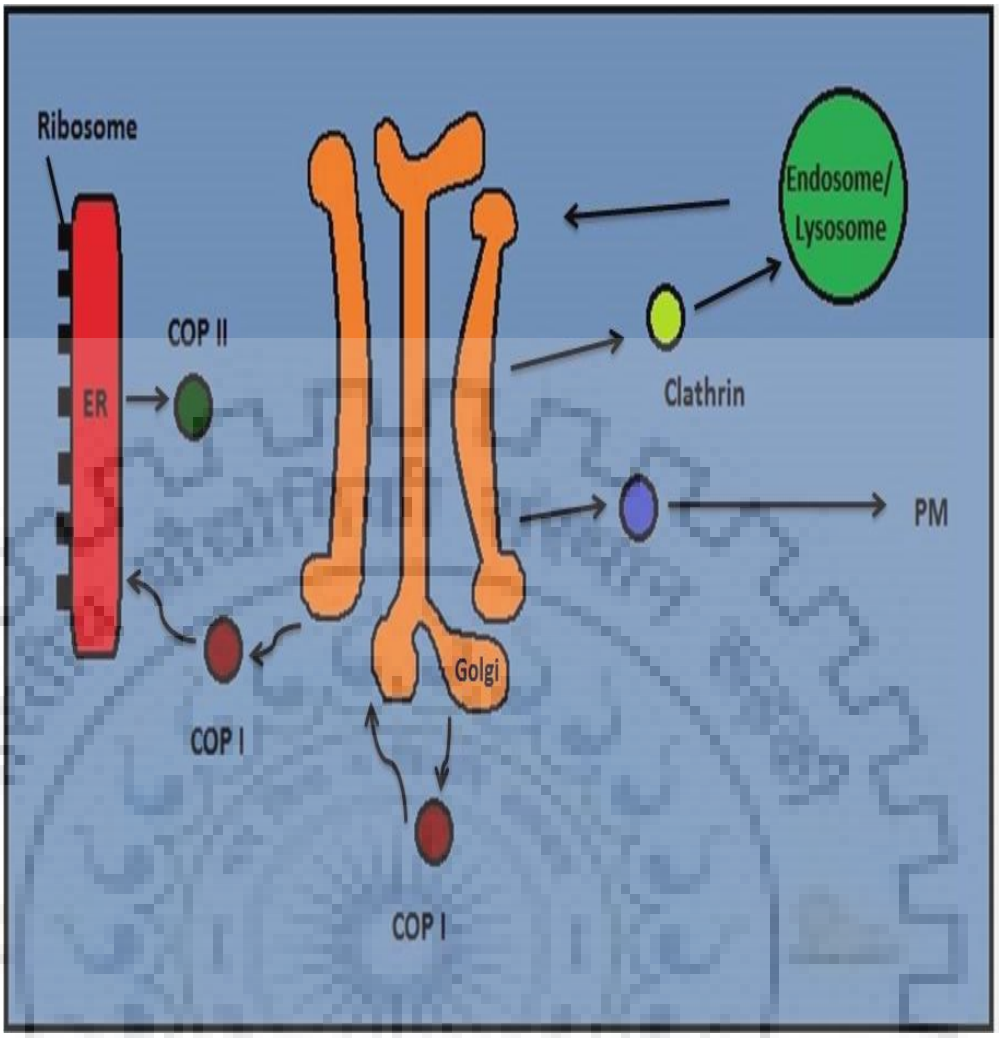


Fig. 2.3. The transport between ER and Golgi network regulated by COPI and COPII.

. One large subunit is made up a γ/α/δ/ε/ζ polypeptide whereas the second large subunit contains a β1/β2/β3/β4/β5 polypeptide. The medium and small subunits are composed of μ1/μ2/μ3/μ4/μ5 and σ1/σ2/σ3/σ4/σ5 polypeptides, respectively. The five AP (1-5) complexes are predominantly expressed in eukaryotes including human, mice and *A. thaliana* [25,104]. In *Saccharomyces cerevisiae*, *S. pombe*, *Drosophila. melanogaster*, *Caenorhabditis elegans* expression of only AP-1, AP-2 and AP-3 [24, 104], recommending that AP-4 and AP-5 have some specific roles in higher eukaryotes. AP-4 and AP-5 are not linked to clathrin unlike to AP-1, AP-2, and AP-3. In AP-1, AP-2 and AP-3, isoforms of subunits encoded by different genes are also found such as two γ (γ 1 and γ 2), two μ 1 (μ 1A and μ 1B) and three σ1 (σ 1A, σ 1B and σ 1C) for AP-1; two α (αAand αC) for AP-2; two β3 (β3A and β3B), two μ3 (μ3A and μ3B), and two σ3 (σ3A and σ3B) for AP-3 [24]. Each AP complex has three domains, namely, a core, two appendages (ears) and two hinge regions. The core domain is made up of the N-terminal domains of two large subunits and the complete mu and sigma subunits. The

two appendage domains are constituted by C-terminal domains of two large subunits. A core domain and an appendage domain are linked by a hinge domain. The AP complexes are one of the component of protein coats of vesicles and interact with the cytoplasmic face of organelles involved into the secretory and endocytic pathways. These complexes participate in the synthesis of coated vesicular carriers and assortment of cargo molecules for addition into their respective carriers.

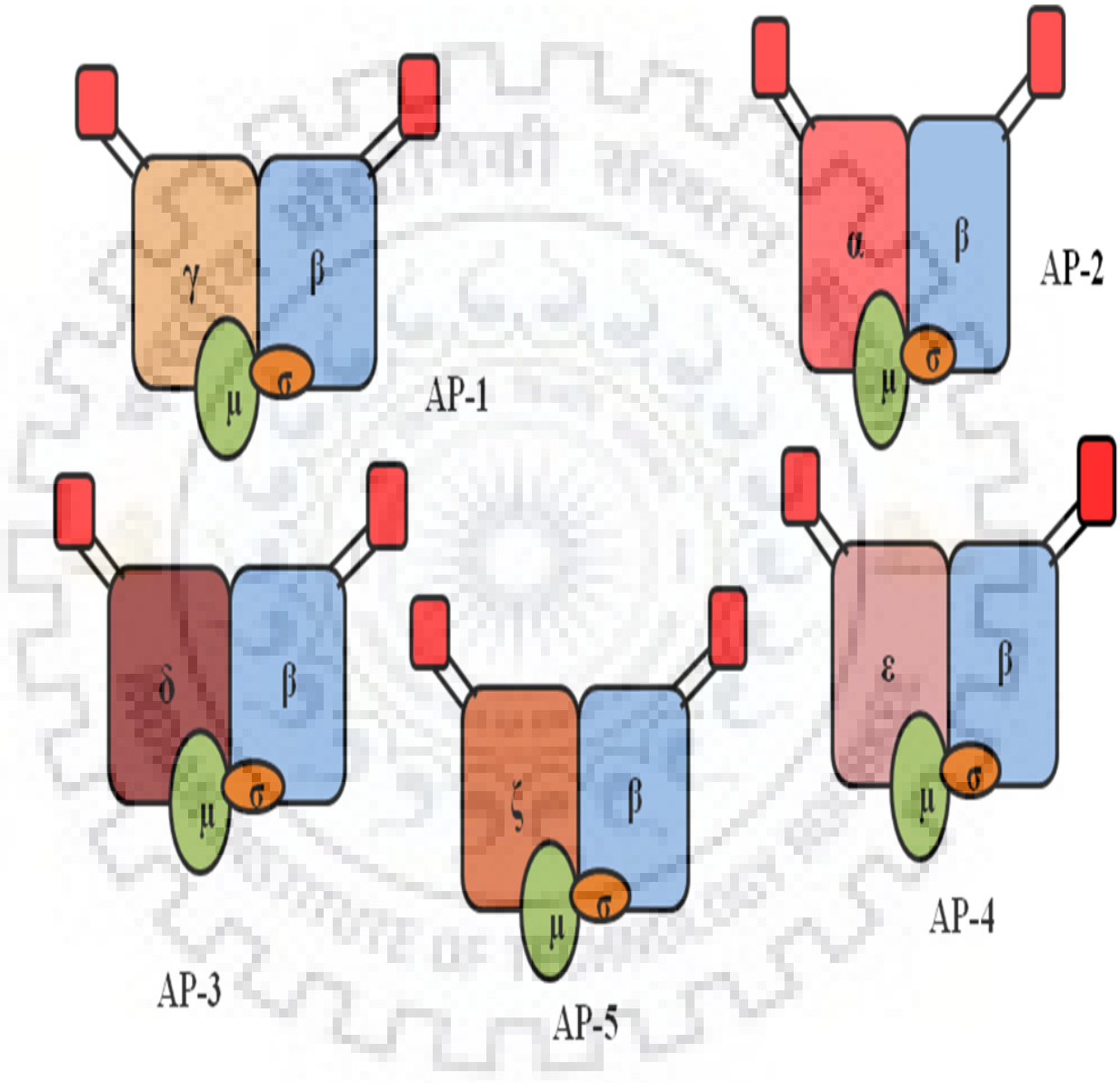


Fig. 2.4. The structures of heterotetarmeric AP complexes (AP 1-5). Each AP complex has two large (γ, α, δ, ε, ζ and β1-β5), one medium (μ1-μ5) and one small (σ1- σ5) subunits.

2.13.1 AP-1 complex

Adaptor protein 1 (AP1) is heterotetrameric complex consists of two large, one medium and one small subunit (γ, β, μ and σ). AP-1 possess two γ (γ1A and γ2B), μ (μ1A and μ1B) and three σ isoforms (σ1A, σ1B and σ1C). AP-1 is localized to the TGN and endosomes. AP-1 A is found at TGN and endosome through interaction with ARF1 and PI4P phosphatidylinositol 4 phosphate (PI4P) [302, 275, 219]. This complex attributes to synthesis of secretory granules [37] and transport of cargo between TGN to endosomes in both directions [246, 275]. AP-1 B is located predominantly at recycling endosomes (REs), perhaps by interacting with ARF6 and PI(3,4,5) P3 (phosphatidylinositol 3,4,5-trisphosphate) [74, 238] to performe basolateral sorting in epithelial cells [77]. The mutation in AP-1A shows disruption in sorting of polarized cells [95]. These studies revealed that both AP-1A and AP-1B are responsible for the regulation of cell polarity by helping in polarized sorting of cells in biosynthetic and endocytic pathways.

2.13.2. AP-2 complex

The AP-2 complex was first adaptin to be identified. This complex is located at the plasma membrane of the cell. The localization of AP-2 complex at plasma membrane is regulated by α subunit. The specific positioning of AP-2 complex on membrane takes place due to the interaction of α and μ_2 subuits with phosphotidalinositol (PIP2) [83, 117] and ARF6 [135, 199]. Positively charged residues of hinge region are involved in the interaction with PIP2. The AP-2 complex plays a role in clathrin-mediated endocytosis of proteins, such as receptors, adhesion molecules and viral proteins[125, 127, 218,111, 109]. Traub and co-workers [260] crystallized the α subunit appendage and suggested that it had a molecular platform domain which organizes several protein protein interactions which are required to assemble a CCV.

2.13.3. AP-3 complex

The AP-3 complex like AP-1 complex, localized at TGN and endosomes. The recruitment of AP-3 complex on membrane is controlled by ARF1 [194, 26]. Immuno-electron microscopy reveals the presence of AP-3 and AP-1 complex on discrete endosomal bud [206, 257]. The AP-3 regulates the transport of cargo from tubular to late endosomes and participates in the formation of lysosome-related organelles (LROs). For the synthesis and detachment of exocytic organelles like, large dense core and synaptic vesicles in neurons, AP-3B isoform is required [72, 23, 94]. In *A. thaliana*, AP-3 complex along with clathrin and dynamin proteins play a role in sorting of vacuolar proteins to PSV and LVs [304].

2.13.4. AP-4 complex

The first identification and characterization of the AP-4 complex was reported in the mammalian (human and mouse) cells in the year 1999 [60, 107]. The AP-4 complex (ε-β4-μ4-σ4), like AP-1 complex, is located on TGN and is involved in transport of cargo proteins from TGN to early endosomes, sorting of basolateral membrane in epithelial cells, and somatodendritic domain in neurons [239, 174]. Bohem and co-workers [26] revealed in their study that the localization of AP-4 complex to TGN was regulated by the interaction of epsilon (ε) and mu (μ4) subunits with ARF1, a small GTP binding protein. The sequence analysis revealed that one large subunit of AP-4 complex, β4 had a high similarity with the N-terminal trunk region of β1, β2 and β3 subunits of other AP complexes. The beta subunit interacts with other subunits through the conserved trunk region of N-terminal end [175, 233]. Hirst *et al.* [107] investigated that hinge and ear regions of β4 subunit do not contain any binding site for the clathrin protein. This study showed that β4 subunit might be interacting with some coat proteins other than the clathrin [107]. In another study, Hirst and co-workers in 2013, revealed the interaction of AP-4 complex with spastic paraplegia type 11 (SPG11) and SPG15. AP-5, similar to AP-4 is also located at TGN and not associated with clathrin [105].

2.14. AP-4 complex in plants

The presence of genes coding for AP-4 complex in plants was first reported in 2001 in *A. thaliana* [24]. The μ4 subunit of the AP-4 complex has been found to interacts with the tail of the vacuolar sorting receptor (VSR) in tobacco [86]. Fuji *et al.* [86] identified the four subunits of AP-4 complex of *A. thaliana* and demonstrated that mu subunit of this complex recognizes the vacuolar sorting receptor 1 (VSR1) which is a receptor for sorting of seed storage proteins to target specific proteins to the protein storage vacuoles (PSVs). The mutations in the genes coding for each of the four subunits of AP-4 complex resulted in defective vacuolar sorting of the major storage protein 12S globulins, suggesting a role of this complex in transport of proteins to vacuoles. In plants, the physiological roles of AP-4 complex and its involvement in vesicular trafficking mechanism is still unknown [86].

2.15. Sorting signal recognition

The transport of proteins from their site of synthesis to final destination like endosomes, lysosomes and the basolateral plasma membrane of polarized epithelial cells is controlled by the some specific signals present on the cytosolic domains of the proteins. The adaptor proteins, components of membrane coats, recognize sorting signals which are short stretches of amino acid residues [24, 221 ,261]. The most studied sorting signals are based on dileucine

[DE]XXXL[LIM] and tyrosine YXXØ signal (where Ø is bulky hydrophobic amino acid: leucine, isoleucine, methionine, valine or phenylalanine) and X can be any amino acid).

In AP complexes (AP1-3), the combination of large (γ, α and δ) and small (σ1, σ2 and σ 3) subunits together form a binding site for a signal [DE]XXXL[LIM] [121, 65]. The tyrosine based signal YXXØ attaches to the AP complexes via their C-terminal domain of the μ1, μ2 and μ3 subunits [189, 190]. The interaction of sorting signals [DE]XXXL[LIM] to α and σ 2 [131]and YXXØ to μ2 [197,117], and μ3A [169] on structural level have been described with the help of X-ray crystallography. In AP-2 complex, binding sites for signals were observed in the core region consisting of amino terminal of large subunits and full length of μ2 and σ2 subunits. Two hydrophobic pockets in μ2 of AP-2 complex reside the Y and Ø residues of sorting signal YXXØ [197].

The AP-4 complex does not show efficacious interactions with [DE]XXXL[LIM] as compared to other AP complexes [1, 120]. The interaction of μ4 with YX[FYL][FL]E motif of signal YXXØ, in the cytosolic tail of amyloid precursor protein (APP) is stronger [38] as compared to signal found in transmembrane lysosomal proteins [247, 1]. Yap *et al*. [290] reported that μ4 subunit interacts with the glutamate receptor *δ*2 with the help of di-aromatic residue (FXF) and phenylalanine-based motifs (FGSV), and FR motifs found at the C-terminal end. The μ subunit of AP-5 complex binds to some other sorting signals as conserved amino acid residues required for binding to signal YXXØ were not found [106].

2.16. Mutational study of AP-4 complex

The knockouts of β4 subunit of AP-4 complex in mice have been reported to cause mild symptoms of motor neuron impairment. These mutant mice showed mislocalization of α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptors and transmembrane AMPA receptor regulatory proteins (TARPs) in the neuronal axons [174]. The dissociation of the AP-4 complex with amyloid precursor protein (APP) increases the breakdown of APP to amyloid-β peptide and elevate the chances of Alzheimer's disease [38]. Jamra and co-workers [119] found that human being suffereing from critical intellectual disability, absence of speech, spastic paraplegia, microcephaly, deformation of foot, decreased in muscle weight of the lower limbs, difficulty in walking, and stunted growth had mutations in sigma, beta and mu subunits of AP-4 complex. In another study, mutation in any subunit leads to malfunction of whole AP-4 complex and cause autosomal recessive disorder known as AP-4 deficiency syndrome [183]. Verkerk *et al*. [266] investigated that mutation in A4PM subunit in brain cell leads to congenital spastic tetraplegia (CST). Bauer and co-workers [17] revealed that mutation in AP-4 complex subunits, results in hereditary spastic paraplegia (HSP), a group of neurodegenerative

disorders characterized by progressive lower limb spasticity and weakness [17]. Fuji *et al*. [86] reported that mutation in mu subunit of AP-4 complex leads to inadequate distribution of seed storage proteins by VSR1 regulated vacuolar protein sorting.

2.17. Structure of AP-4 complex

The crystal structure of C-terminal end of μ subunit of AP-4 complex was given by Burgos *et al*. [38]. The 1.6 Å X- ray crystal structure of μ subunit of AP-4 from human revealed the interaction of YX[FYL][FL]E- signal from APP with μ4. The μ4 C-terminal domain has an immunoglobulin-like b sandwich fold consisting of 17 strands organized into two subdomains, similar to the structure of the C-terminal domain of μ2 [197]. The binding site in μ4 for tyrosine based signal YXXØ motif is different from found in μ2 [197] and μ3 [169] subunits. Breyan *et al*. [31] gave 1.8 Å X crystal structure of mutated D190A C-terminal of μ4 complex with a peptide of APP. Breyan and coworkers predicted that L261K or E265R mutations completely hinder the binding of μ4 with sorting signal found on APP. The mutation in amino acid residue D190A present at the binding site for YXXØ sorting signals also break off the interaction between the YKFFE signal and μ subunit [31]. This study revealed that binding site of μ4 subunit for the YKFFE signal is similar to the binding site of μ2 and μ3 subunits for YXXØ sorting signal. Both binding sites of μ4 subunit are located on opposite direction to each other and mutation in D190A might cause changes in stability of structure of μ4 subunit of AP-4 complex [38].

2.18. Accessory proteins involved in vesicular trafficking

2.18.1. ADP-ribosylation factor (ARF)

The ARFs (about 21-kDa) are ubiquitous and nucleotide (Guanosine triphosphate) binding proteins and share more than 60% sequence similarity. Like ARFs, several ARF like proteins (ARLs) have been identified in mammals. ARF proteins differ from ARF like proteins based on their sequences and functions [49]. Based on their amino acid sequences, protein sizes, gene structures and phylogenetic analysis, the ARFs are subdivided into three classes: class-I (ARF1, ARF2, and ARF3), class-II (ARF4 and ARF5), and class-III (ARF6) [262]. Each ARF has two conserved regions which have been named as switch I and switch II regions. The cycling of ARFs between membrane and cytosol is controlled by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs). In switch I and switch II effector binding regions of ARFs conformational changes have been observed during the replacement of GTP with GDP which plays an important role in ligand binding [224, 93,179]. The gene structures of members of class I (ARF1) are conserved in nature [235]. The Class-I ARFs are well studied among all known ARFs. The ARF1 control the assembly of various vesicle coat

components, such as COPI and adaptor complexes, on the Golgi apparatus [224,116,26] and effectors molecules including phospholipase D1 [33,50], phosphotidylinositol 4-phosphate 5-kinase [110], Arfaptin [126] and recruitment of GDAP1 to membrane [173]. The ARF1 regulates the localization of AP-4 complex onto TGN via interacting with the trunk region ε and signal binding specific region of μ4 subunits [26]. Bohem and co-workers [26] reported that the interaction of ε adaptin with ARF1 is dependent on binding of nucleotide and sensitive to mutation in switch I and switch II regions. The interaction between ARF and μ4 is nucleotide independent and does not include switch regions [26].

The localization and function of class II ARFs were found to be identical to class I ARFs [42]. The members of class III ARF play a role specifically in endosomal plasma membrane system and participate in endocytosis and exocytosis [215].

2.18.2. Vacuolar sorting receptors (VSRs) and epsins

In plant cells, vacuoles are essential multi functional organelles which store water, ions, proteins and several secondary metabolites. Two types of vacuoles, protein storage vacuoles (PSVs) and lytic vacuoles (LVs) are present in plant cells. The proteins stored in PSVs are used as a source of carbon and nitrogen during germination of seeds [272]. The transport of a vacuolar protein to its destination vacuole takes place via vesicle mediated trafficking pathway. A main step of this pathway at the destination stage is the recognition of a specific signal on the protein by a vacuolar sorting receptor (VSR) [285,223]. The VSRs belonging to type I membrane proteins, contained luminal domain along with a cytosolic motif. These VSRs, belonging to the well studied receptor family [57], help in sorting of storage proteins and lytic cargos to PSVs [236,81] and LVs [303], respectively. In *A. thaliana*, VSR1 interacts with mu subunit of AP-4 complex by the tyrosine motif present in the tail region [82]. The binding of ligand to the VSR is pH dependent process so in an acidic environment, VSR dissociates from its ligand. Kalthoff *et al*. [124] reported that VSR interacts with epsin protein in *A. thaliana*.

The epsin protein possesses a well conserved epsin N-terminal homology (ENTH) domain and disorganized C-terminal end [46,56,196]. It binds to phosphatidylinositols (Ptdlns) and cargo proteins with its ENTH domain [180] and ubiquitin-interacting motifs [191] respectively and plays a role in transport of proteins to lysosomes [124,245]. In mammals, tepsin (tetra epsin), a member of epsin family interacts with C-terminal ear domain of β4 and ε subunits with their conserved amino acid motifs [GS]LFXG[ML]X[LV] and S[AV]F[SA]FLN, respectively, found in their disorganized C-terminal end [176]. These associations played an important role in the localization of tepsin protein to the TGN in a human cell. In another study

Frazier and co-workers [79] reported that mutation in either β4 or LFXG[ML]X[LV] peptide leads to reduction in amount of tepsin that associate with AP-4 complex subunits. Borner *et al*. [29] revealed that tepsin protein like AP-4 complex has been lost from various organisms including yeast, worms and flies. This study suggested that both AP-4 complex and accessory protein tepsin depend on each other for their functions. The tepsin protein showed interaction with β subunit of AP-4 complex, but not with any other adaptor protein.

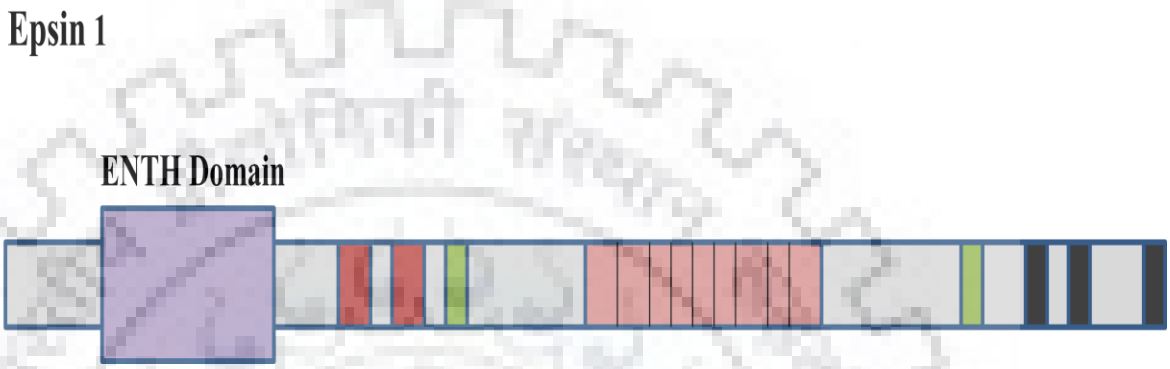


Fig. 2.5. Structural organization of Epsin 1 protein. Red boxes show ubiquitin binding regions, pink boxes are APs binding regions, green boxes are clathrin binding regions, black boxes are NPF motif for EH binding.

2.19. Co-expression analysis

The gene co-expression analysis emerged as an important tool to explore the function and/or identification of gene(s) that are related to concerned genes at molecular level. Covington and coworkers [53] revealed that co-expression study is based on the concept that genes having same functions must have similar expression patterns in different conditions [53]. The co-expression study of genes helps in to explore the transcriptomic data and study the gene expression at the molecular level in plants [172]. Aoki *et al*. [7] in *Escherichia coli* and yeast revealed gene regulatory mechanism by co-expression analysis. The gene co-expression has been studied in several species including, *A. thaliana [12, 208,168,274]*, *Gossypium arboretum* [293], *Oryza sativa* [154, 73], *Z. mays* [73], *H. sapiens* [150], *M. musculus [164]*. Stuart and co-workers in 2003 investigated that sometime co-expressed genes might express together by chance, but not based on their functions and produce false positive results [248]. To avoid false positive results, several workers explained that co-expression study of orthologus genes should be performed in other species [248,19,161]. Co-expression studies helps in identification of genes involved in phenolic pathway in pollen development, flavonoid pathways in *Arabidopsis* [291] and regulation in circadian rhythm [53].

2.20. Protein protein interaction network

The protein-protein interactions (PPIs) are necessary in the living cell to perform its cellular processes comprising of DNA replication, mRNA synthesis, signal transduction and metabolic pathways. PPI analysis has been studied in various species of plants and animals [201,232,254,27,234,267]. Von Mering and coworkers [271] reported that PPIs played important roles in the developmental process at the molecular level, such as transcriptional co-factor recruitment, post translational modifications, protein phosphorylation, assembly of cytoskeleton, transporter activation, and many other [62,297,250,112,66, 54] in organisms. Recently, the computational based studies have gained much attention in the detection of PPIs. Wass *et al.* [277] reported that PPI could be found out by docking of three dimensional structure of proteins. The structure-based analysis gave more information as compared to non-structure based method by providing details about the amino acid residues at the interface.

2.21. Transcriptomic analysis

Auffray in 1996 gave the term trnascriptome for the first time. Several researchers in the last few years have performed the transcriptome analysis technique in different organisms [187,132, 284,69,253]. The advancement in sequencing techniques leads to development of association mapping in plants [216]. The analysis of transcriptome is known as expression profiling, which helps to study the level of expression of mRNA in a given biological sample or tissue [283]. The transcriptome analysis can be conducted via either microarray or next generation sequencing (NGS) methods [264]. Transcriptomic study by next generation sequencing (NGS) is an effective technique to examine gene expression in plants. RNA Sequencing (RNA-Seq) based on NGS has the potential to identify differentially expressed genes (DEGs) because it has an effective range of expressions [123]. The RNA-seq era began in 2008 with the publication of three papers on a new short-read technology developed by Solexa (now Illumina) [186, 249, 280]. RNA-Seq experiment executed in different steps such as cDNA is synthesized from mRNA to construct RNA-Seq library. The cDNA sequences were mapped to reference genome and expression level of genes were estimated. The mapped data were normalized statistically and DEGs were identified [195,274].

The above literature review showed that guar, maize and soybean are economically important crops. The seeds of these crops are used as food and in various industries. In legumes and cereals very little information is available regarding the AP-4 complex mediated vesicular trafficking process. As AP-4 complex plays an important role in transport of seed storage proteins so detailed analysis of the AP-4 complex and its accessory proteins was done to explore the vesicular trafficking mechanism in guar, maize and soybean.

3. MATERIALS AND METHODS

3.1. Materials

3.1.1. Chemicals/Biochemicals

All the chemicals/biochemicals used in this study were brought from standard companies and belong to molecular biology and analytical grade.

3.1.2. Plant material and plant growth conditions

The seeds of guar variety RGC-1066 were obtained from CAZRI (central arid zone research institute, Jodhpur). Seeds of *Zea mays* Trimurti 826K variety were purchased from a local supplier and those of *Glycine max* DS9712 were obtained from Indian Agricultural Research Institute, New Delhi. Fifty seeds of each variety were surface sterilized with 0.01% $HgCl_2$ for 30 seconds, washed thrice with distilled water. Seeds were sown in Feb-June session in soil at Indian Institute of Technology Roorkee. The plants were grown under normal field conditions.

3.1.3. Composition of buffers

3.1.3.1. Composition of 10X TBE buffer [229]

Sr. No.	Component	Amount
1	Tris	10.9 g
2	Boric acid	5.56 g
3	EDTA	0.98 g
4	Distilled H_2O	100 ml

3.1.3.2. Composition of 1X TE buffer [229]

Sr. No.	Reagent Name	Amount
1	1M Tris- HCl (pH=8)	1 ml
2	0.5M EDTA (pH=8)	200 µL
3	Sterile milliQ H_2O	100 ml

3.1.3.3. Composition of DNA extraction buffer [229]

Sr. No.	Reagent Name	Concentration
1	Tris-HCl (pH=8.0)	100 mM
2	EDTA (pH= 8.0)	20 mM
3	NaCl	1.4 M
4	CTAB	5%
5	β-mercaptoethanol	0.01%

3.1.4. Preparation of 6x DNA gel loading dye

The DNA loading buffer was prepared by mixing both 0.25 g bromophenol blue and 0.25 g xylene cyanol in 6 ml of 50 % glycerol, final volume made up to 10 ml after adding 4 ml nuclease free water.

3.1.5. Composition of polymerase chain reaction (PCR) used in amplification of genes

Sr. No.	Content	Amount per 20 µl
1	10X Master mix (Thermofischer)	10 µl
2	Forward primer (20 pM)	0.5 µl
3	Reverse primer (20 pM)	0.5 µl
4	DNA (100 ng/µl)	1 µl
5	Nuclease free H_2O	8 µl

3.1.6. Composition of single stranded complementary DNA (cDNA) synthesis

Sr. No.	Content	Amount per 20 µl
1	5X Reverse transcriptase buffer	4.0 µl
2	dNTP mix (10mM)	0.5 µl
3	OligodT primer (20 pM)	0.5 µl
4	Reverse transcriptase enzyme (10 U/ µl)	0.3 µl (of 10000 unit enzyme)
5	DEPC treated H_2O	4.7 µl

3.2. Methods

3.2.1. Isolation of genomic DNA

Leaves were collected from 15 days old guar, maize and soybean plants and genomic DNA was isolated by the CTAB method [67] with some changes. Around 0.5 g leaves were crushed in liquid nitrogen with sterile, pre-chilled mortar and pestle to make a fine powder. The fine crushed powder was transferred to a 2 ml Eppendorf tube having 1 ml of pre-warmed DNA extraction buffer. The components were mixed well. The mixture was incubated in a water bath at 65 °C for 1 h. In the Eppendrof tube an equal volume of chloroform: isoamyl alcohol (24:1) solution was added and incubated for 10 mins at room temperature. The mixture was centrifuged (5810 R) at 8,000 rpm for 10 min at room temperature. The upper aqueous layer containing DNA particles was transferred to another fresh 1.5 ml Eppendorf tube and an equal volume of chilled isopropanol was added to precipitate the DNA. The isopropanol and DNA mixture was incubated at -20 °C for 30 min. The pellet was collected after centrifugation of precipitated DNA at 8000 rpm for 10 min and washed with 70 % ethanol, air dried and dissolved in 40 µl TE buffer. The contamination of RNA was removed by adding 2µl of RNase A (10 mg/ml) and incubated at 37 °C in a dry bath for 1 h. Heat inactivation of RNase A was done at 65 °C for 10 min in a dry bath. Recovery of DNA was carried out by addition of an equal volume of chloroform: isoamyl alcohol (24:1) and the contents were mixed and centrifuged at 8000 rpm for 10 min at room temperature. Supernatant was transferred to a 1.5 ml fresh Eppendorf tube. DNA was precipitated by the addition of 500 µl of 100% ethanol and further incubated at 4 °C for 2 h. The mixture was centrifuged at 8000 rpm for 10 min. The supernatant was discarded and the pellet was air dried and dissolved in 40 µl TE buffer.

3.2.2. Quantification and dilution of DNA

The quality of the genomic DNA was analyzed using gel electrophoresis on 0.8 % agarose gel. The isolated DNA from leaves of guar, maize and soybean was quantified by checking absorbance at 260 nm in a nano drop (Thermo scientific) and diluted with TE buffer. The diluted DNA sample was stored at -20 °C for further use.

3.2.3. Agarose gel electrophoresis

The genomic DNA was run on 0.8% agarose gel to check its integrity. The 0.3g of agarose (SeaKem LE agarose-Lonza) was added to 30 ml of 1X tris borate EDTA (TBE) buffer and mixture was heated until the solution become clear. The EtBr (10 mg/mL) prepared in sterile milliQ water was added after cooling down the solution. The solution was discharged in a gel

casting tray where the comb was placed already. The comb was removed from solidified gel. The DNA sample was mixed with 6X gel loading buffer in a ratio of 5:1 and loaded onto the gel. The DNA sample (5 µl) was loaded into a well of gel and placed in the electrophoresis unit already filled with 1X TBE buffer. The composition of TBE and TE buffers [229] have been presented in Tables 3.1.3.1 and 3.1.3.2, respectively. A 1 kb DNA ladder solution (5 µl) (Biolab, Bangalore) was loaded to determine the approximate size of DNA. The power supply was set at 70 V and DNA sample was run until the dye covered 75-80% of the distance down the gel. The gel was visualized under UV light and the results were documented by gel documentation system (Bio-Rad).

3.2.4. Primer designing for amplification of genes coding for AP-4 complexes of guar, maize and soybean

The primers for genes coding for epsilon (AP4E), beta (AP4B), mu (AP4M) and sigma (AP4S) subunits of AP-4 complex of guar, maize and soybean were designed. The primers for the amplification of AP-4 complex genes of guar and soybean were designed from the CDS, and genomic sequences of soybean, respectively. The full length sequences of AP-4 complex of *Glycine max* genes were downloaded from Phytozome 10 database. The orthologous of genes coding for AP-4 complex were identified in *Medicago truncatula* and *Phaseolus vulgaris* using BLASTn tool (Altschul *et al*., 1990). To identify the conserved nucleotide stretches of these genes, the multiple sequence alignment of the CDS sequences was performed with Clustal O tool. The primers were designed manually from the conserved regions and analyzed by OligodtAnalyzer tool. For the amplification of AP-4 complex genes of maize, primers were designed from the genomic sequences spanning the exon-intron regions. The genomic sequences of maize were retrieved from Phytozome 9 database.

Primer sets were designed after fulfilling the basic conditions as listed below:

1)The length of primers were ranged from 18-22bp,

2) Tm was selected from 55 to 60 °C

3) The 3’ end of primer does not contain three continuous C or G

4) GC content was kept upto 45-70 %

Table 3.1. Primer sequences designed from genes coding for epsilon, beta, mu and sigma subunits AP-4 complexes of guar, maize and soybean.

Sr. No.	Gene Name	Forward Primer	Reverse Primer	Amplicn Length
1	Zm AP4E	GCAGAGCAATTTGCGCCTAGC	GCACCTTCATCC TCCTCTCCTG	156bp
2	Zm AP4B	GCGAAGTCTGCCTCTCCTTCC	GCAGGAGAT CACCCTCTTGAAG	141bp
3	Zm AP4M	GCGACCACATCG TCTTCCG	CCAAGAGAAGGGATGGGGAGAC	240bp
4	Zm AP4S	TCGTCAACAAGCAGGGCCAG	CGTAGCGCCTGTAAACGACC	170bp
5	Gm AP4E	GAGGCTGCTGCGGATGTAATCG	CTCCAAGCAGTCAATGACGGCA	147bp
6	Gm AP4B	CAGAGGAAGCAGCCAGGA	TCCAGCACAAGACACTGAGCC	108bp
7	GmAP4M	CGTTCTGTCGCAGCGTGG	TCCTGCAGTCCTCCG TCTTCC	120bp
8	Gm AP4S	GCAAGGCCAAACTCTCTTGC	GCGATCTCAACGAAGGAAC	130bp
9	Ct AP4E	CCACGACCTCATCATCCTA	GCCATAACGCCTTCTTCCTAAC	191
10	Ct AP4B	GGGCTTGCTGATTTTCACA	CTCAAATGTCCCTCGGTGT	243
11	Ct AP4M	TGCAAGGGACCAAACGAATG	AGCAAGTCGAATCTCTGGGT	212
12	Ct AP4S	CGAAATCGTTCGCAAATGC	ACATTGCCGAAATGACGATCCATAG	195
13	Ct AP4EF	ATGAGTTCAAAGGTGGTTTTAC	TTAGCCAAGCAAGTCACTGAAC	2904
14	Ct AP4BF	CCGCCACCGCAGTCTCTC	CAAATTTGGACAAGGCTGATTGG	2535
15	CtAP4MF	ATGATATTGCAGTTCTTGTG	TTATAACCGTGCTACGTATGA	1323

16	Ct AP4S F	ATGGGGATCCGATTCGA	TCAAGATGTTTTATCCAT	432

3.2.5. Polymerase chain reaction (PCR) amplification of AP-4 complex genes

The genes coding for AP-4 complex were amplified from the total genomic DNA of guar, maize and soybean using the Applied Biosystems® thermal cycler. The different range of temperature (°C) used for optimization of annealing temperature (Ta) for genes were: AP4E (55-67), AP4B (54-65), AP4M (57-71), AP4S (54-63). The sample was initially denatured for 4 min at 94 °C followed by 30 cycles (1 cycle: 1 min at 94 °C, 1 min at respective Ta, 1 min at 72 °C) of amplification and a final extension at 72 °C for 10 min (Fig. 3.1).

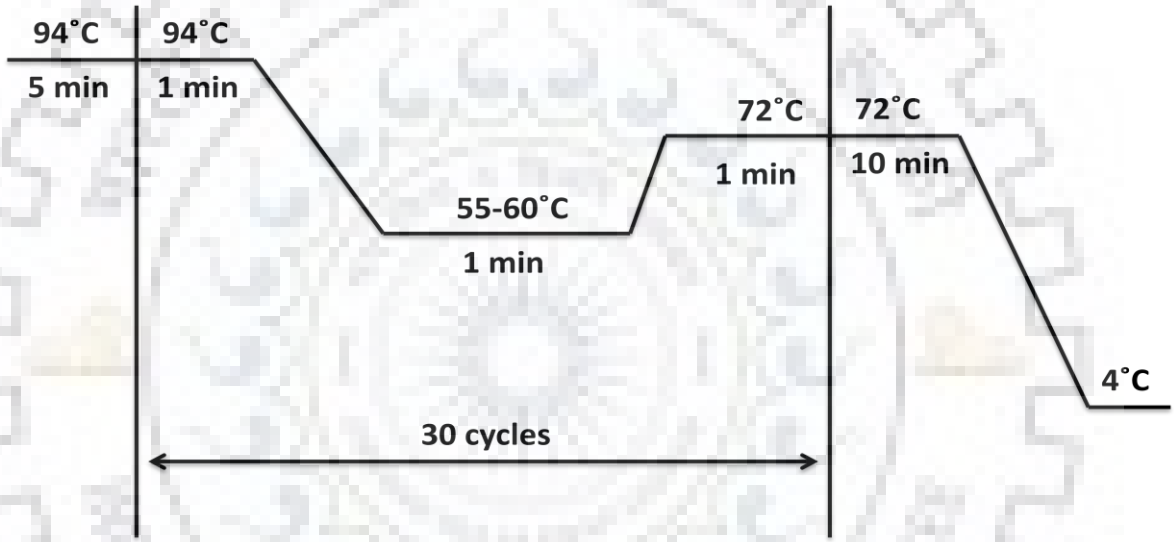


Fig. 3.1. PCR programme used for amplification of different gene reactions

3.2.6. Total RNA isolation

Total RNA was isolated by TriZol (HIMEDIA® RNA-XPressTMReagent) method from different tissues of guar (roots and leaves at 15 days, flower at 30 DAS and seeds at 35 DAF), *Z. mays* (roots at 20 days, bracts and anther at 40 days after sowing (DAS), and seeds at 30 days after flowering (DAF) and *G. max* (roots and leaves at 15 days, flower at 30 DAS and seeds at 30 DAF) following the manufacturer's protocol. Solutions were prepared with dH_2O treated with 0.1 % DEPC overnight. The plant tissue was crushed into a fine powder in a mortar and pestle by using liquid nitrogen and transferred into a fresh 2 ml Eppendorf tube containing 1ml of TRIzol. The mixture was vortexed and centrifuged at 12,000 x g for 5 min at 4 °C. The supernatant was pipetted out to a fresh 1.5 ml Eppendorf tube and incubated for 5 min at room temperature. The amount of 200 µl chloroform was added, vortexed for 15 sec and kept at room

temperature for 5 min, and centrifuged at 12,000 x g for 15 min at 4 °C. The supernatant was discarded and an equal volume of isopropanol was added and incubated for 10 min at room temperature. The sample was then centrifuged at 12,000 x g for 10 min at 4 °C. The pellet was resuspended in 1 ml of 75 % ethanol and centrifuged at 7500 x g for 5 min at 4 °C. The supernatant was disposed of, pellet was air dried and finally dissolved in 40 µl of DEPC treated dH_2O. The concentration of RNA was checked by nano drop.

3.2.7. DNase treatment and precipitation of total RNA

To further increase the purity, isolated total RNA was treated with DNase I enzyme (Himedia). In 1 µg of RNA sample 5 µl of 10X reaction buffer, 2.5 µl of DNaseI was added and final volume was made off up to 20 ml. The reaction mixture was incubated at 37 °C for 30 min, then 75 °C for 10 min. The mixture was transferred immediately to ice. After DNase treatment, purification of RNA was carried out with phenol: chloroform method. The EDTA (20mM) about $1/10^{th}$ volume of total sample volume was added to the reaction. The reaction mixture was kept in water bath (RW 0525G) at 65 °C for 10 min and an equal volume of phenol: chloroform was added. The sample was centrifuged at 12000 x g for 10 min. Supernatant was transferred to a fresh 1.5 ml Eppendorf tube. The 3M Sodium acetate (pH=5.2) $1/10^{th}$ volume of total sample volume along with 2.5 volume of absolute alcohol was added to vial. The sample was centrifuged at 12000 x g for 10 min at room temperature. The supernatant was disposed of, pellet was air dried and resuspended in 40 µl of DEPC treated dH_2O and stored at -80 °C until required.

3.2.8. Single stranded complementary DNA (cDNA) synthesis

The synthesis of cDNA from total RNA isolated from guar, maize and soybean tissues was carried out with semi-quantitative PCR method. In a PCR tube 1 µg (~ 5 µl) of RNA sample, 1 µl oligo dT primer and 4 µl of 0.1 % DEPC dH_2O were added. The mixture was heated at 65 °C for 10 min in a water bath and then incubated for 2 min at room temperature. The 10 µl of master mix (Table 3.1.6) was added to the PCR tube and was kept in a thermocycler (Applied Biosystems®). The RT-PCR temperature cycling conditions were: one PCR cycle (10 min at 25 °C, 1 h at 42 °C and 2 min at 95 °C) for the 20. The synthesized single stranded complementary DNA (cDNA) was quantified using nano drop. The PCR program was run as described in Figure 3.2. The cDNA samples were stored at -40 °C for future use.

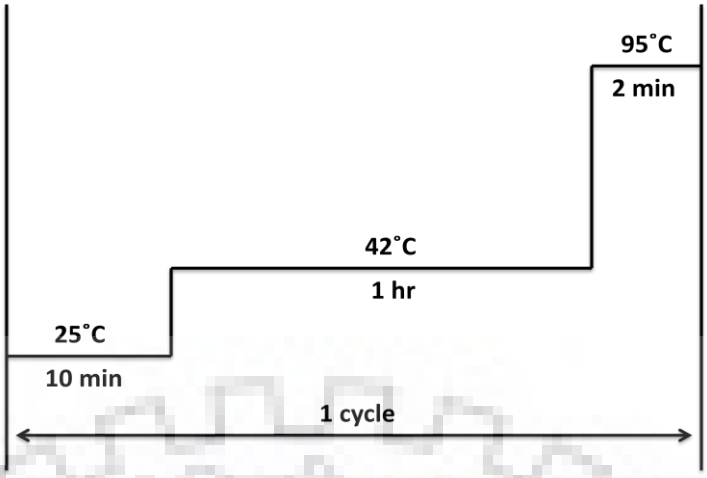


Fig 3.2. The PCR programme used for cDNA synthesis.

3.2.9. PCR amplification of cDNA

For the amplification of genes on cDNA 20 µl PCR reaction mixture was prepared containing 2.5 µl of synthesized cDNA samples, a primer set for the gene, master mix and nuclease free water. A PCR was performed with following cycling conditions: intial denaturation at 94 °C for 4 min, 30 cycles of 1 min at 94 °C, 1 min at respective Ta, 1 min at 72 °C of amplification and a final extension of 10 min at 72 °C. Beta-actin genes of guar, maize and soybean were used as internal control. The integrity of amplified PCR product was checked on 1 % agarose gel.

3.2.10. Sequencing of the amplified products

The PCR products of AP-4 complex genes of guar were purified by using PCR purification kit (Invitrogen®) following the manufacturer's instructions given with kit. The purified products were got sequenced commercially (Eurofines, Bengaluru, India). The confirmation of obtained data was done by the nucleotide tBLASTx tool (Altschul *et al*., 1990).

3.2.11. Evolutionary profiling

3.2.11.1. Retrieval of orthologus/homologous sequences

The gene and protein sequences of epsilon, beta, mu and sigma subunits of AP-4 complex of guar were used to Basic Local Alignment Search Tool (BLAST) (TBLASTN and BLASTP) the NCBI (National Center for Biotechnology Information) and Uniprot databases to retrieve the orthologus /homologous sequences from various plant species. The gene and protein sequences showing similarity with the gene and protein sequences of AP4 complex of *A. thaliana* with predicted BLAST value ≤ 1e-5, sequence identity ≥ 60% for gene and ≥ 40% for protein, and

query coverage more than 70% were selected as candidate genes and proteins. Partial and non redundant sequences were not included in the study.

3.2.11.2 Identification and characterization of exons and introns

The CDS sequences of *G. max*, *Medicago truncatula*, *Eucalyptus grandis*, *Citrus sinensis*, *A. thaliana*, *Brassica rapa*, *Z. mays*, *Oryza sativa* and *Physcomitrella patens* were aligned with the genomic sequences to locate and characterize the exons and introns in genes coding for epsilon, beta, mu and sigma subunits of AP-4 complex.

3.2.11.3. Phylogenetic analysis of protein sequences

The multiple sequence alignments of amino acid sequences of epsilon, beta, mu and sigma were done by using CLUSTAL O (Clustal omega) software [148]. The unrooted phylogenetic trees were constructed by N-J (neighbor-joining) method of MEGA 6.0 [226, 251] software with 1000 replicates of bootstraps. Maximum likelihood and minimum evolution methods of phylogeny were also used to construct trees to confirm tree topologies.

3.2.11.4. Motif analyses of protein sequences of AP-4 complex

The conserved motifs were identified with help of MEME program (http://meme.sdsc.edu) using the following parameters: repetitions per sequence = zero, maximum numbers of motifs=8/12 and optimum motif size ranges between 6 and 50 amino acids.

The protein sequences of beta and mu subunits of AP-4 complexes from *Brachypodium disatchyon*, *Z. mays*, *G. max*, *C. tetragonoloba*, *A. thaliana*, *E. grandis* (along with *Plasmodium falciparum*, *Phytophthora nicotiana*, *Albugo laibachii*, *P. patens* for comparison) were aligned with the protein sequences of beta and mu subunits of AP-4 complexes from *Homo sapiens* and *Mus musculus* by CLUSTAL O software to identify homologous regions recognizing the binding sequences of accessary protein tepsin and sorting signals of the protein to be transported.

3.2.11.5. Estimation of divergence time

The coding sequences of genes of AP4 complex were aligned with the help of CLUSTAL Omega software and results were obtained in NEXUS format. Sequence from *P. patens* was used as an outgroup. The divergence time for a gene was estimated using calibration nodes from monocot and dicot species *O. sativa-Z. mays* (32 ± 6 MYA) and *Malus domestica –C. sinensis* (106 ± 4 MYA) [47], Bayesian Markov Chain Monte Carlo (MCMC) approach and Beast v1.8.4 [68] software. The analysis was performed by using the subsequent conditions: a relaxed molecular clock along with an uncorrected log-normal distribution model, the Hasegawa, Kishino and Yano (HKY) substitution model, a Yule model of speciation and four

gamma categories. Total 3 runs were performed with 20 million MCMC generations and each were sampled after 1000th generation. The data from all the runs were combined and uploaded in Tracer v1.5 (http://beast.bio.ed.ac.uk/tracer). The final tree was graphically visualized and produced using FigTree v1.3.1 software [68].

3.2.11.6. Estimation of functional divergence

The functional divergence of epsilon, beta, mu and sigma subunits was estimated by the DIVERGE v3. 0 software using the GU99 method [98, 99]. This method describes the coefficient of Type I (significant changes in site specific shifts in evolutionary rate) and Type II (changes in site specific shifts in amino acid residue physiochemical properties) functional divergence (θ_I and θ_{II}, respectively) of duplicated genes between two groups. If the value of θ_I and θ_{II} are significantly greater than zero, it shows the significant shifts of amino acid physiochemical properties of amino acids. The critical amino acid site (CAAS) residues responsible for functional divergence are estimated using a site specific profile based on the posterior probability (Qk).

3.2.12. GC content and CpG island analyses

The sequences of genes (including upstream and downstream sequences) encoding epsilon, beta, mu and sigma subunits of *A. thaliana*, *Z. mays* and *G. max* were retrieved from the Phytozome database and GC content and CpG islands were determined by using EMBOSS CpGplot online server (http://www.ebi.ac.uk/tools/seqstats/emboss_cpgplot/).

3.2.13. Promoter analysis

The upstream sequences (~2kb) and downstream sequences (~1kb) of genes encoding four subunits of AP-4 complex of *A. thaliana*, *Z. mays* and *G. max* were retrieved from Phytozome 10, The Arabidopsis Information Resource (TAIR) and Soybean knowledge base (SoyKB) databases. The *cis*-regulatory elements and transcription factor binding sites in these sequences were identified with the help of Plant Promoter Analysis Navigator (http://PlantPAN2.itps.ncku.edu.tw) tool and PLACE-plant cis-acting regulatory DNA elements (http://www.dna.affrc.go.jp/htdocs/PLACE/).

3.2.14. Identification and characterization of single nucleotide polymorphisms (SNPs)

The SNPs in the genes coding for four subunits of AP-4 complex of *Z. mays* and soybean were obtained from Maize GDB, and Soybean Knowledge Base (SoyKB), respectively. Each SNP was characterized with respect to position on CDS, position in protein, polymorphism, site (5'UTR, 3'UTR, exon and intron), mutation type (transition or transversion) etc. The SNPs of the coding regions were further categorized into synonymous (showing no change in amino

acid) and non-synonymous (change in amino acid) types. The restriction enzyme mining for SNPs in the genes of AP-4 complex was done with NEBcutter tool.

The deleterious effect of each nonsynonymous SNP was detected by PANTHER-PSEP (Protein ANalysis THrough Evolutionary Relationships) software [252]. The PANTHER software estimates the length of time (in a million years) amino acid of the reference protein remain conserved in the lineage. This software categorized SNPs into four classes such as probably damaging, damaging, bengin and probably bengin.

The functional binding sites in the protein sequence of epsilon, beta, mu and sigma subunits of AP-4 complexes of *Z. mays* and *G. max* were predicted by DoGSiteScorer server [270]. The DoGSiteScorer is a grid-based method which is based on a difference of Gaussian filter for the detection of potential binding pockets. The refined PDB file generated by I-TASSER software was uploaded with default parameters and potential pockets were predicted on the surface of protein based on the heavy metals coordinates. Global properties of each pocket/ subpocket(s) including size, shape and chemical features were calculated. The nonsynonymous and synonymous SNPs were identified in these functional binding pockets.

3.2.15. Expression profiling

The RNA-Seq analysis of genes coding for AP4E, AP4B, AP4M and AP4S were studied in *A. thaliana*, *Z. mays* and *G. max* while genes encoding ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARFB1C, ARF3, Epsin1, Epsin 2 and VSR1 were studied in *A. thaliana* and *Z. mays*. The RNA-Seq data of different tissues of *A. thaliana* (root, anther, flower, siliqua and seed), *Z. Mays* (root, stem, bract, anther, seed and endosperm) and *G.max* (root, stem, leaf, flower, seed and endosperm), available as sequence read archives (SRAs), were downloaded from the NCBI database. The downloaded SRA files in Fastq format were uploaded to Bowtie 2/2.0.0-beta7 software [145], aligned with the genes of interest of the *A. thaliana* TAIR, MaizeB73 and *G. max* Wm82.av1 reference genomes. Total number of reads and reads exactly aligned one time with the reference gene were obtained and further used for calculation of reads per kilobase of transcript per million mapped reads (RPKM) value based on the formula given below:

RPKM= Number of reads mapped to a gene sequence/ (Gene length/1000

* Total number of reads/1,000,000)

The coefficient of variance (CV) and standard deviation (SD) were calculated for each gene. Heatmaps were constructed for the genes involved in vesicular trafficking in different tissues of *A. thaliana* and *Z. mays* by using Pearson correlation coefficient in Heatmapper software

(http://www.heatmapper.ca). The fold change in expression of a gene in a sample with respect to the expression of this gene in seed was calculated using the following formula:

Log 2 Fold change= Log2(B)-Log2(A)

The log2 fold change values were used to plot graphs. The RNA-Seq data was normalized and then Z-score value of each sample of a given tissue was calculated. A positive value of Z-score indicates high expression. In a tissue, the gene pair was considered co-expressed only if Z-score value of each member of the pair is positive.

Z-score value = $x - \mu / \sigma$

Where x= test score

μ is the sample mean

σ is the sample standard deviation

3.2.16. Primer designing and amplification of genes coding for ARFs, Epsins and VSR1 of maize

The genomic sequences of these genes were retrieved from Phytozome 10v. The primers (Table 3.2) for genes coding for ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARFB1C, Epsin 1 and VSR1 proteins of maize were designed manually from the genomic sequences of maize spanning the exon-intron regions and analyzed by OligodtAnalyzer tool.

The genes coding for ARFs, Epsin and VSR1 involved in AP-4 mediated vesicular transport were amplified from the cDNA of maize using the Applied Biosystems® thermal cycler. The different range of temperature (°C) used for optimization of annealing temperature (Ta) for genes were: ARFA1A (55-59), ARFA1B (54-61), ARFA1C (55-60), ARFA1E (55-60), ARFA1F (54-62), ARFB1B (54-63), ARFA1C (51-59), Epsin 1 (55-60) and VSR1 (53-69).

Table 3.2. Primer sequences designed from genes coding for ARFs, epsin and VSR1 of maize.

Sr. No.	Primer Name	Primer Sequence Forward	Primer Sequence Reverse
1	Actin 1	AAAAGCCATGAGCTTCCTGA	CTTGCTCATACGGTCTGCAA
2	ARFA1A	GGGGTCAGGACAAGATCAGA	ACAAGCAGCACAGCATCAC
3	ARFA1B	CGATGCAGCTGGTAAGACCAC	CTGACCTCCAACATCCCAGAC
4	ARFA1C	TGCTGAATGAGGATGAGCTG	AGCCCCTCATACAACCCTTC
5	ARFA1E	GCTCGGCGAGATCGTCAC	CTGATCTTGTCCTGGCCACC
6	ARFA1F	GGAGGCTGTCAGTGTCTGTCTC	GTGGCATTCACTGTGTGGGATG
7	ARFB1B	GTGGCATTCACTGTGTGGGATG	GTCTGGAACTCTTGCCTCGC
8	ARFB1C	CTCGGCACCAGCGAGATGA	CCTCCACAAGGGCCTCAC
9	Epsin 1	GCTCAGCCTCATTTGGTAGC	CTCCTATCAGGCTTCGCATC
10	VSR 1	GCAGTGGAGGCTTGCTTTAC	GTTGTCCAGGGGCATGTACT

3.2.17. Validation of gene expression profiling obtained by RNA-Seq analysis

Quantitative real-time analysis was performed using Applied Biosystems, 7900HT Fast Real Time PCR System in a 96-well plate. Power SYBR® Green dye (Applied Biosystems) was used to measure the cycle threshold (Ct) value. A 10 µl reaction mixture containing 5 µl SYBR Green dye, 0.5 µl forward primer, 0.5 µl reverse primer, 1 µl cDNA and 3 µl nuclease free water was loaded in a well. A reaction in a qRT-PCR program was carried out using the following program: 2 min at 50 °C (1 cycle), 10 min at 95 °C (1 cycle), 15 s at 95 °C and 30 s at 57-60 °C (40 cycles). The dissociation curve for the primers of *Z. mays*, *G. max* and guar was added in the above program. The actin 1 gene of *Z. mays*, actin 1 gene of *G. max* and actin of guar were used as internal control in real time experiments of *Z. mays*, *G. max* and guar, respectively. Relative quantification was calculated for each gene. Each qRT-PCR reaction was performed with three biological and three technical replicates.

The gene sequences of Actin 1 of *Z. mays* and *G. max*

Zm Actin 1

Forward primer: 5’ CCTTCGAATGCCCAGCAATG 3’
Reverse primer: 5’ACCGTGTGGCTCACACCAT 3’

Length of amplicon 109bp

Gm Actin 1

Forward primer: 5’ATCTTGACTGAGCGTGGTTATTCC3’
Reverse primer: 5’GCTGGTCCTGGC TGTCTCC 3’

Length of amplicon 130bp

Guar Actin

Forward Primer: 5’ AGTTGGTGGAGCGATTTG
Reverse Primer: 5’ AAGGGCATCACAGACCTG

Length of amplicon 140bp

Delta Ct=Ct(gene of interest)-Ct (reference gene)

3.2.18. Protein protein interaction studies

The gene pairs of *Z. mays* with positive value of Z-score were selected for study of protein protein interactions (PPIs). The direct physical and indirect functional interactions were identified with the help of STRING v10.0 database (http://string-db.org). The full length amino acid sequences of AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARFB1C, ARF3, Epsin1, Epsin 2 and VSR1 were uploaded to database with default parameters. The data generated from STRING database were used for visualization of PPI and biological network by Cytoscape v3.6.1 (http://www.cytoscape.org/) software.

3.2.19. *In-silico* studies of AP-4 complex

3.2.19.1. Physico-chemical properties and structural analysis of the protein

ExPASy ProtParam tool (http://www.expasy.ch/tools/protparam.html) was used for the comparative analysis of various physico-chemical parameters and calculations of molecular weight (Mw), theoretical isoelectric point (pI), positive and negative residues, extinction coefficient [87], instability index [102], grand average hydropathy (GRAVY) and aliphatic index.

3.2.19.2. Secondary structure analysis

The secondary structure prediction of the amino acid sequences of AP4E, AP4B, AP4S and AP4M of *C. tetragonoloba* was performed by the PSIPRED v3.3 (Psi-BLAST based secondary structure prediction) server. Hydropathy plots were generated by Kyte-Doolittle hydropathy plot server [143]. The Pfam (http://pfam.sanger.ac.uk) and InterPro (http://www.ebi.ac.uk/interpro/) softwares were used to identify known motifs in epsilon, beta, mu and sigma polypeptides of *C. tetragonoloba* and *Z. mays*.

3.2.19.3. Homology modelling of AP-4 complex

The crystal structures of AP-4 complex subunits and ARFA1A of guar and *Z. mays* are not available in the Protein Data Bank (PDB), (http://www.rcsb.org/pdb/home/home.do). Hence, homology modelling method was used for prediction of the structures of proteins. BLAST search against the PDB was performed to find out homologous protein structures which can be used as templates for homology modeling [5]. The pairwise sequence alignment of the query sequence with the template sequences was done with Clustal omega program. The I-TASSER (Iterative Threading ASSEmbly Refinement) software [289] was used for homology modelling. The templates for structure were identified by LOMETS (Local MEta-Threading-Server) from PDB library. Top five models for AP-4 complex subunits (AP4E, AP4B, AP4M and AP4S) and ARFA1A proteins satisfying the spatial restraints based on target-template alignment were built by SPICKER program based on monte carlo approach. The model with the lowest free enegy has highest confidence score, high TM score and better quality. The best model was chosen based on confidence score (c-score) and TM score. The generated model was further refined by two step procedure of ModRefiner software described in flow chart (Fig. 3.3) [287]. The energy minimized model was further validated by PROCHECK [149] and VERIFY3D. A detailed analysis of the stereochemical quality of the protein structures in the form of Ramachandran plot was performed by PROCHECK. The compatibility of a 3D structure of of each model with its amino acid sequence was tested by VERIFY3D. The quality of the model structures was also checked with ProSA server [279]. To visualize the energetically favored, allowed and disallowed regions. Ramachandran plots were generated by the Rampage server (http://mordred.bioc.cam.ac.uk).

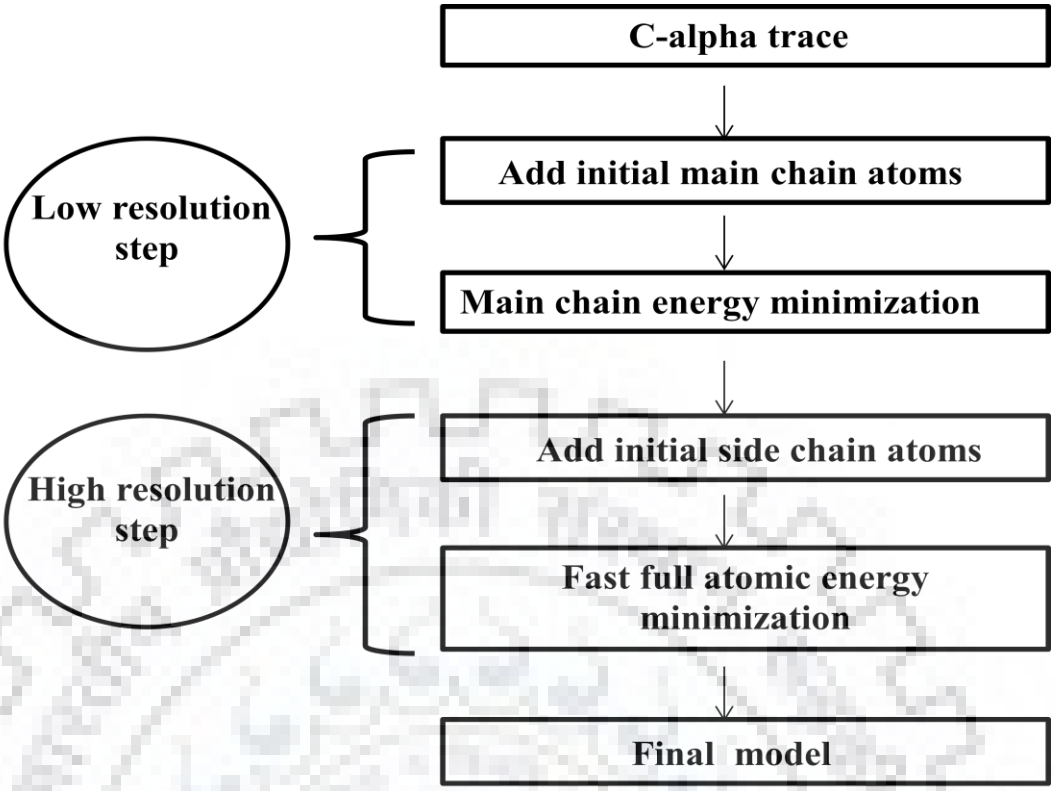


Fig. 3.3. Flowchart showing the two step procedure of ModRefiner software to construct refined model. Modified from Xu and Zhang, 2011.

3.2.19.4. Docking of AP-4 complex subunits with ARFA1A of *Z. mays*

The AP-4 complex subunits and ARFs accessory proteins of *Z. mays* showed protein protein interactions so the docking was carried out for Z. mays proteins. The amino acid sequences of ARF1 family members are well conserved hence docking of epsilon, beta, mu and sigma subunits of AP-4 complex was performed with ARFA1A protein. The docking of each subunit of AP- complex with ARFA1A was performed with the help of ZDOCK server [209] based on Fast Fourier Transform (FTT) algorithm and top 1000 predictions were obtained. Each docking experiment generated 10 models and the best docked structure was selected based on its energy profiles. PyMol v 4.2.1 software was used for visualizing the interactions between the amino acids of AP-4 complex subunits and ARFA1A accessory protein of *Z. mays* [59].

4. RESULTS

4.1. Amplification of AP-4 complex genes of guar

Primers were designed for genes encoding epsilon, beta, mu and sigma subunits of AP-4 complex from the soybean. Amplification of these genes was done on both genomic and cDNA regions of guar at the optimized annealing temperatures. The Figure 4.1 showed the amplification of full length CDS of epsilon (2500bp), beta (2300bp), mu (1332bp) and sigma (423) genes. The sequencing results were confirmed further by the BLASTN tool. The nucleotide sequence of each gene was translated into amino acid sequence using Expasy translate tool. The nucleotide and protein sequences of AP-4 complex genes of guar were deposited to NCBI. The NCBI accession numbers of the submitted complete CDS (cDNA) sequences are MG009494 (AP4E), MG009493 (AP4B), MG009492 (AP4M), and MG009491 (AP4S) (Appendix 1).

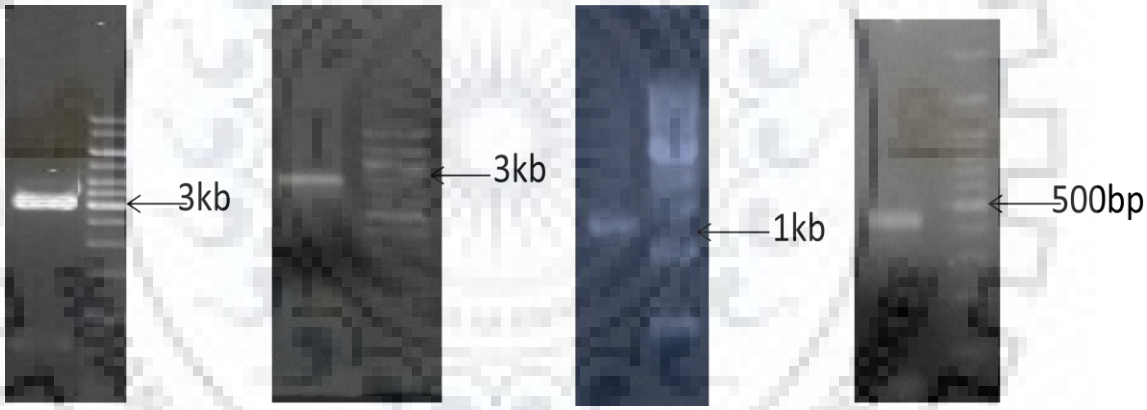


Fig. 4.1. Agarose gel (1%) image showing the amplification of genes coding for (A) epsilon, (B) beta (C) mu and (D) sigma subunits of AP-4 complex on leaf tissue of guar.

4.2. Study of evolutionarily conserved genes and proteins

4.2.1. Identification of gene and protein sequences of AP-4 complex

The orthologous/homologous gene/protein sequences for the epsilon, beta, mu and sigma subunits of AP-4 complex were found in all the studied species of monocots and dicots, and a moss. Nucleotide and protein accession numbers of the selected sequences have been given in Appendix 2-5.

4.2.2. Analysis of exons and introns in AP-4 complex genes

The predicted numbers and sizes of exons and introns in the genes coding for epsilon, beta, mu and sigma subunits have been shown in Tables 4.1-4.4. In the epsilon subunit genes, the numbers of exons were found to vary from 5 in *Brassica rapa* to 13 in *P. patens* (moss). Two monocots (*Z. mays* and *O. sativa*) and two dicots (*Eucalyptus grandis* and *Citrus sinensis*) contained 11 exons and two legumes (*G. max and Medicago truncatula*) and *A. thaliana* contained 10 exons in the epsilon subunit gene. All studied plants contained 11 exons except 10 in *B. rapa* in the beta subunit gene. In mu subunit 13 exons were found in all plants except *E. grandis* which had 14 exons. Four exons of equal length were observed in sigma subunit gene of all plants.

The minimum and maximum exon lengths were 50 bp and 1100 bp, respectively, in the genes coding for epsilon subunits of the studied plant species. The exon length ranged from 50 bp to 450 bp in the beta subunit genes of various plants. A variation was also observed in the exon lengths of genes for mu (50bp to 200 bp) and sigma (70bp to 150 bp) subunits of the studied plant species. The intron length in genes of epsilon, beta and mu subunits of various plant species varied from about 70 bp to 2500 bp. Minimum and maximum intron lengths in the sigma subunit coding genes were found out to be 80 bp and 1500 bp, respectively.

4.2.3. Phylogenetic relationships

Phylogenetic trees constructed from the protein sequences of epsilon, beta, mu and sigma subunits of AP-4 complexes of studied plant species and the protein sequences of outgroups are presented in Figs. 4.2-4.5. In the phylogenetic tree of epsilon subunit, monocots and dicots formed separate clades (Fig. 4.2). The dicots clade was further divided into four subclades where as the monocots remained in a single clade. The patterns of clade formation for beta and mu subunits were similar to that of epsilon subunit (Figs 4.3 and 4.4). In the phylogenetic tree of sigma subunit the dicots were divided into five subclades where as the monocots formed a single clade (Fig. 4.5). *P. patens* (moss) showed close relationship with monocot species.

Table. 4.1. Exon and intron sizes of the epsilon subunit of AP-4 complexes from studied plant species

Exon Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*	Intron Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*
1	783	788	566	566	788	788	602	602	46	1	143	1191	905	361	95	75	571	524	404
2	0	0	0	0	0	0	0	0	615	2	0	0	0	0	0	0	0	0	205
3	0	0	185	185	0	0	191	185	185	3	0	0	1040	402	0	0	124	92	363
4	215	215	215	215	200	585	215	215	230	4	97	111	111	93	81	68	86	94	452
5	110	110	107	110	110	110	110	110	110	5	73	1516	100	777	97	100	268	440	375
6	197	197	197	197	197	0	197	197	197	6	93	93	101	107	91	0	88	101	181
7	131	131	131	131	131	0	131	131	131	7	98	96	895	219	234	0	453	509	151
8	56	56	56	56	56	43	56	56	56	8	104	96	97	83	82	0	74	85	209
9	116	116	116	120	116	0	116	113	116	9	1324	2515	1872	609	368	0	109	446	192
10	77	77	77	77	77	0	77	77	77	10	1096	1514	761	536	81	0	1147	790	381
11	143	143	143	143	143	0	143	143	143	11	0	0	0	0	0	0	0	0	534
12	0	0	0	0	0	0	0	0	474	12	1369	1143	1184	880	70	1040	1556	234	283
13	1095	1094	1091	1101	974	1656	1067	1010	769										

Table 4.2. Exon and intron sizes of the beta subunit of AP-4 complexes from studied plant species

Exon Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*	Intron Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*
1	55	55	52	52	52	52	55	55	533	1	721	401	824	307	397	200	88	104	407
2	516	516	516	516	516	516	516	516	127	2	1026	1579	105	548	112	82	1321	2414	326
3	124	124	124	124	124	124	124	125	198	3	651	812	690	590	84	107	82	469	129
4	198	198	198	198	198	198	200	198	228	4	609	877	521	389	249	267	720	295	293
5	228	228	228	228	228	228	228	228	166	5	140	157	136	118	96	97	99	120	266
6	166	166	116	116	116	166	495	116	209	6	162	156	75	172	103	91	110	99	262
7	209	209	209	209	209	320	209	209	110	7	1507	1521	883	666	113	0	834	508	477
8	110	110	110	110	110	0	110	110	176	8	101	122	76	321	92	87	94	79	301
9	176	176	176	176	176	176	176	176	257	9	499	1088	802	186	109	96	2039	212	305
10	455	452	458	446	446	443	443	458	176	10	377	342	373	534	292	83	398	390	388
11	290	290	290	290	290	290	290	290	290										

Table 4.3. Exon and intron sizes of the mu subunit of AP-4 complexes from studied plant species

Exon Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*	Intron Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*
1	57	57	57	57	60	60	57	57	57	1	925	503	328	204	334	349	97	142	790
2	94	94	100	94	94	94	94	94	94	2	523	634	288	236	69	167	720	684	379
3	203	203	203	203	203	203	203	203	203	3	419	713	421	464	123	98	202	212	288
4	116	116	116	134	116	116	116	116	116	4	153	97	94	102	66	68	121	123	106
5	128	128	128	128	128	128	128	128	128	5	1149	782	765	477	170	132	577	377	190
6	136	126	126	126	126	126	126	126	132	6	454	1315	284	386	311	282	2488	455	473
7	73	79	103	103	100	100	103	103	109	7	113	83	264	108	79	80	259	90	157
8	104	104	104	104	104	104	104	104	104	8	252	361	128	171	94	97	86	92	257
9	94	94	112	94	94	94	94	94	94	9	0	0	446	0	0	0	0	0	0
10	87	87	65	76	91	91	91	85	87	10	1007	1308	252	537	155	112	107	111	228
11	0	0	27	0	0	0	0	0	0	11	145	90	107	121	82	74	318	106	246
12	60	60	60	75	60	60	60	62	60	12	101	84	131	153	82	86	102	86	161
13	73	73	73	73	73	73	66	73	73	13	609	736	606	246	76	91	311	677	191
14	96	92	98	98	98	98	98	98	98										

Table 4.4. Exon and intron sizes of the sigma subunit of AP-4 complexes from studied plant species

Exon Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*	Intron Number	*Glycine*	*Medicago*	*Eucalyptus*	*Citrus*	*Arabidopsis*	*Brassica*	*Zea*	*Oryza*	*Physcomitrella*
1	140	140	140	140	140	140	140	140	140	1	98	363	1461	104	378	241	224	215	969
2	86	86	86	86	86	86	86	86	86	2	1043	325	231	154	180	88	7706	766	309
3	68	68	68	68	68	68	68	68	68	3	99	176	85	127	101	81	111	98	246
4	134	134	134	134	134	134	134	134	134										

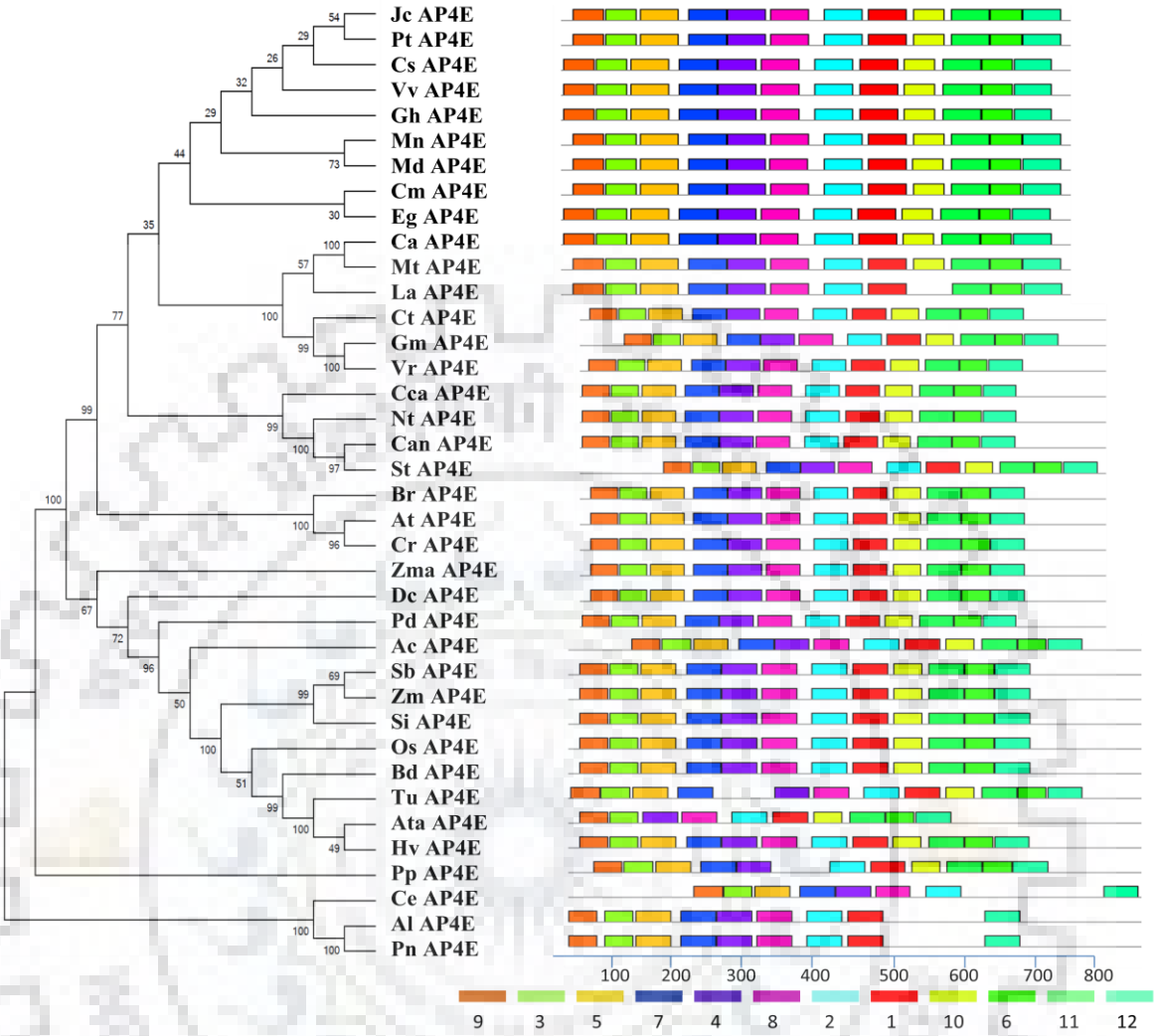


Fig. 4.2. Phylogenetic tree and motifs of protein sequences of epsilon subunit of AP-4 complex from 39 plant species. The unrooted tree was built with MEGA6 software after multiple sequence alignment of protein sequences of epsilon subunits. The branch support bootstrap values were obtained by 1000 replicates using the NJ (neighbor-joining) method. The conserved motifs of protein sequences were obtained by MEME software and shown by different colored boxes.

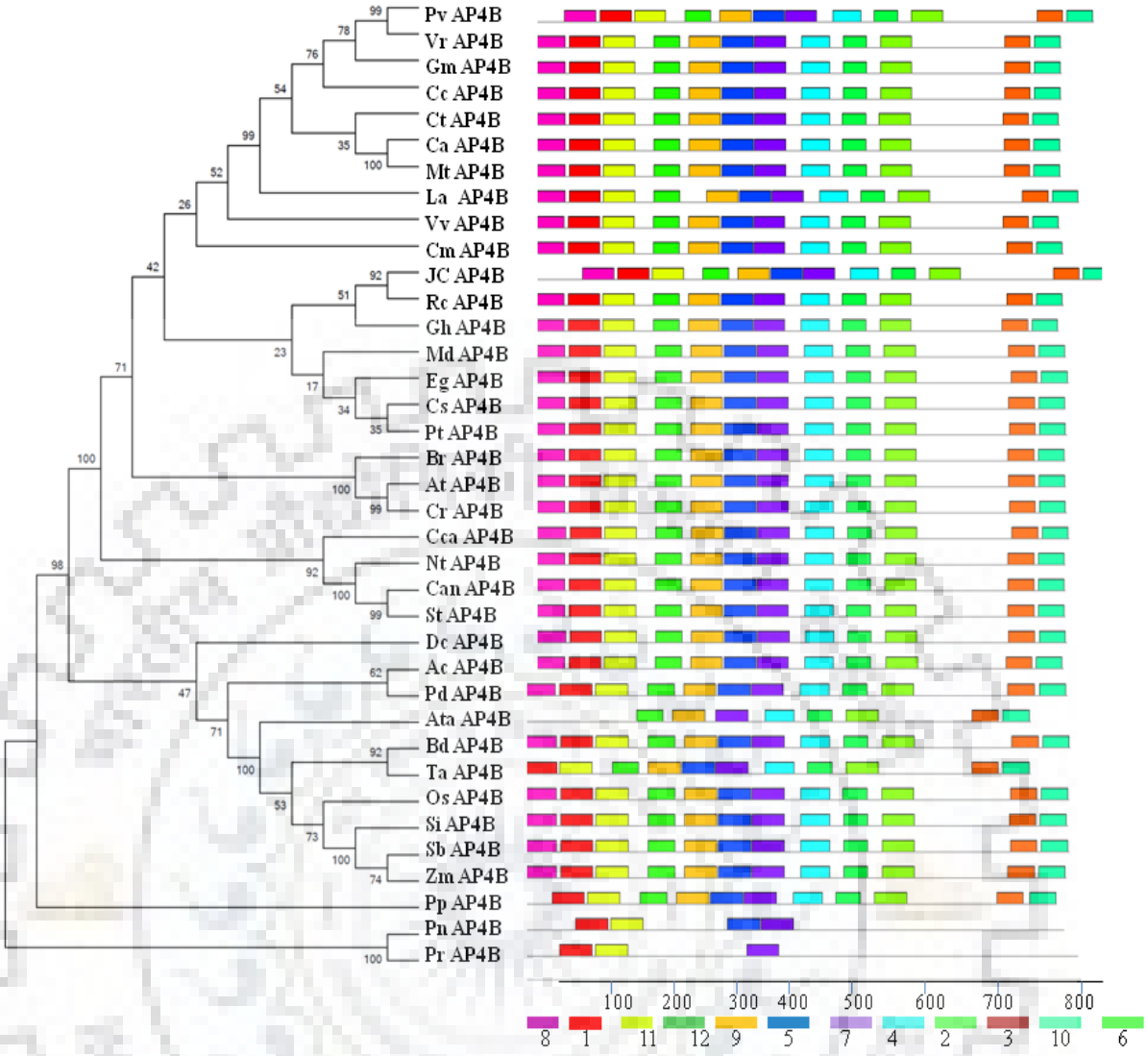


Fig. 4.3. Phylogenetic tree and motifs of protein sequences of beta subunits of AP-4 complexes from 36 plant species. The unrooted tree was built with MEGA6 software after the multiple sequence alignment of protein sequences of beta subunits. The branch support bootstrap values were obtained by 1000 replicates using the NJ (neighbor-joining) method. The conserved motifs of protein sequences were obtained by MEME software and shown by different colored boxes.

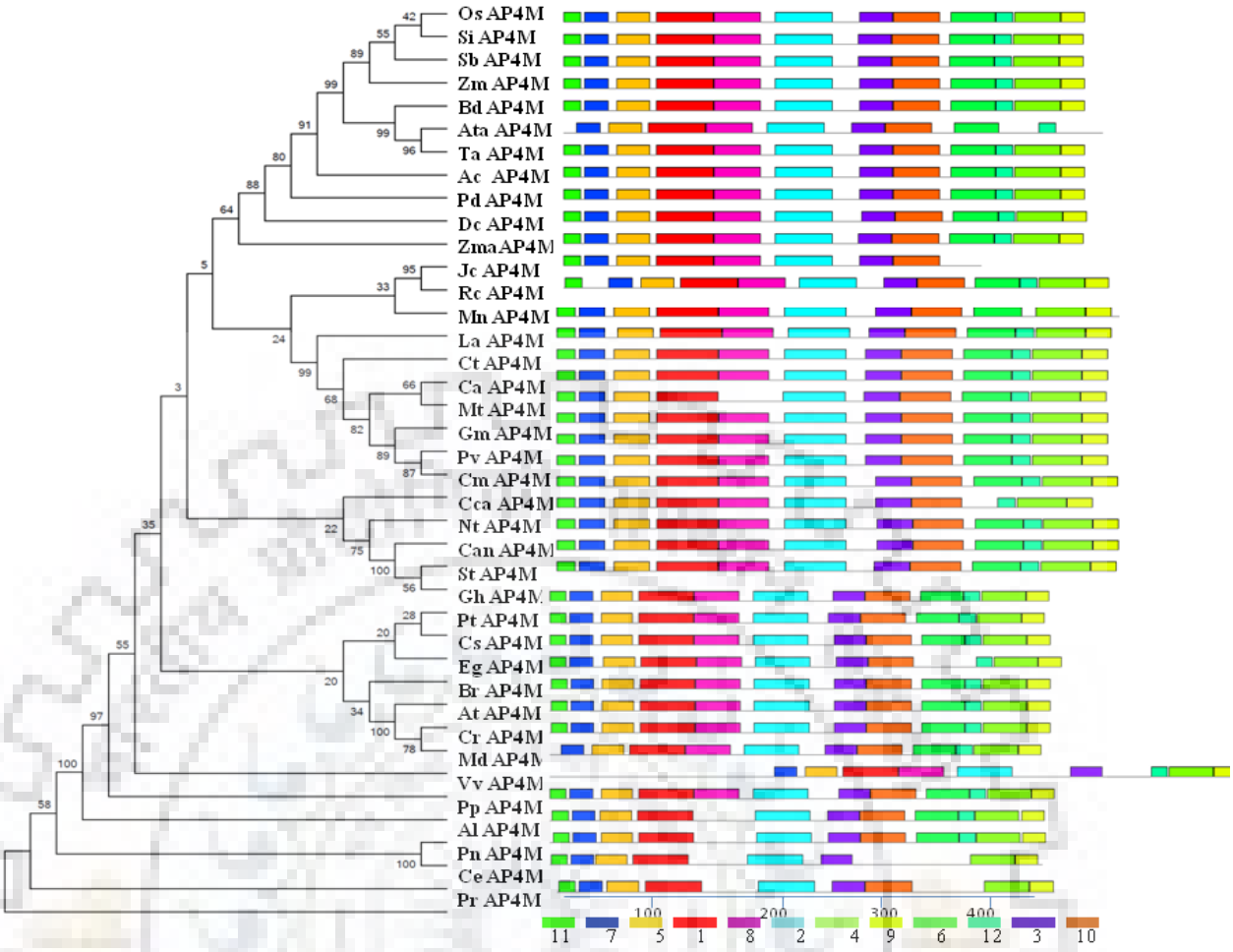


Fig. 4.4. Phylogenetic tree and motifs of protein sequences of mu subunits of AP-4 complexes from 39 plant species. The unrooted tree was built with MEGA6 software after multiple sequence alignment of protein sequence alignments of mu subunits. The branch support bootstrap values were obtained by 1000 replicates using the NJ (neighbor-joining) method. The conserved motifs of protein sequences were obtained by MEME software and shown by different colored boxes.

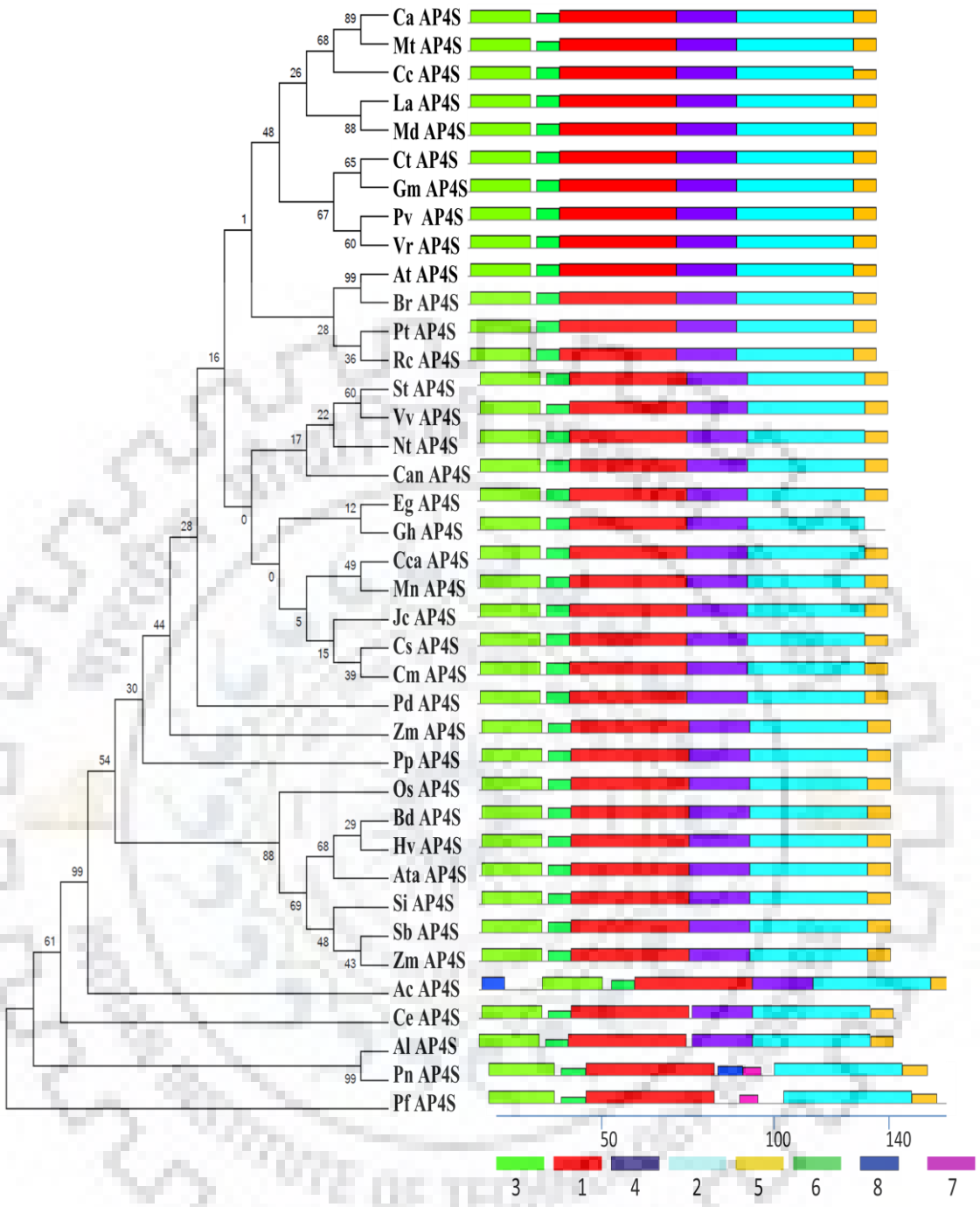


Fig. 4.5. Phylogenetic tree and motifs of protein sequences of sigma subunits of AP-4 complexes from 38 plant species. The unrooted tree was built with MEGA6 software after the multiple sequence alignment by using protein sequence alignments of sigma subunits. The branch support bootstrap values were obtained by 1000 replicates using the NJ (neighbor-joining) method. The conserved motifs of protein sequences were obtained by MEME software and shown by different colored boxes.

4.2.4. Conserved motifs in epsilon, beta, mu and sigma subunits

The conserved motifs found in the protein sequences of epsilon, beta, mu and sigma subunits have been given in Figs. 4.2 – 4.5. In the epsilon subunit, 12 conserved motifs were obtained in all the studied plant species (Fig. 4.2). The beta subunit also contained 12 conserved motifs in all plants except *Triticum urartu* and *Aegilops tauschii* in which 8th motif was absent (Fig. 4.3). The 10th motif was found to contain a region which is homologous to the *H. sapiens* and *M. musculus* AP-4 beta subunit polypeptide region which recognizes and binds to hydrophobic residues [LFxG(M/L)x(L/V)] of accessory protein tepsin. In *H. sapiens and M. musculus*, the binding role is mainly played by three amino acids, namely, tryptophan, tyrosine and glutamine, present at positions, 635, 682 and 696, respectively, in the conserved region (Frazier et al. 2016). These three amino acids were also found in the conserved region in the studied plant species with the difference that tyrosine was replaced by phenylalanine, another aromatic amino acid (Fig 4.6A).

The numbers of conserved motifs in the mu subunit varied from 8 to 12 in various plants; *Jatropha curcas* and *Vitis vinifera* contained 8 and 9 motifs, respectively, whereas 11 motifs were found in *M. domestica, Cicer arietinum, E. grandis* and *Coffea canephora* (Fig. 4.4). The 3rd motif was found to contain a region (of amino acids) which is homologous to the AP-4 mu subunit polypeptide region (reported in *H. sapiens* and *M. musculus*) which recognizes and binds to the sorting signal of a cargo protein. Seven amino acids, namely, phenylalanine, histidine, phenylalanine, lysine, glycine, glutamine and tyrosine (positions 255, 256, 264, 270, 277, 278 and 284, respectively, in *H. sapiens* and *M. musculus*), which play an important role in recognition of sorting signal, were also present in all studied plant species (Fig. 4.6B). Six motifs were present in sigma subunit of each studied plant except *Ananas comosus* in which 7 motifs were identified (Fig. 4.5).

4.2.5. Estimation of divergence time

The divergence time estimations for the genes coding for epsilon, beta, mu and sigma subunits of AP-4 complex have been presented in Figures 4.7- 4.10. The results showed that about 151 million years ago (MYA), during the Jurassic period, the ancestral gene for epsilon subunit had undergone divergence. The epsilon coding gene did not show any further divergence until 94 MYA, when speciation of legumes occurred (Fig. 4.7). The ancestral genes for beta, mu and sigma subunits appear to have diverged during the Jurassic period about 160 MYA, 147 MYA and 152 MYA, respectively (Figs 4.8-4.10).

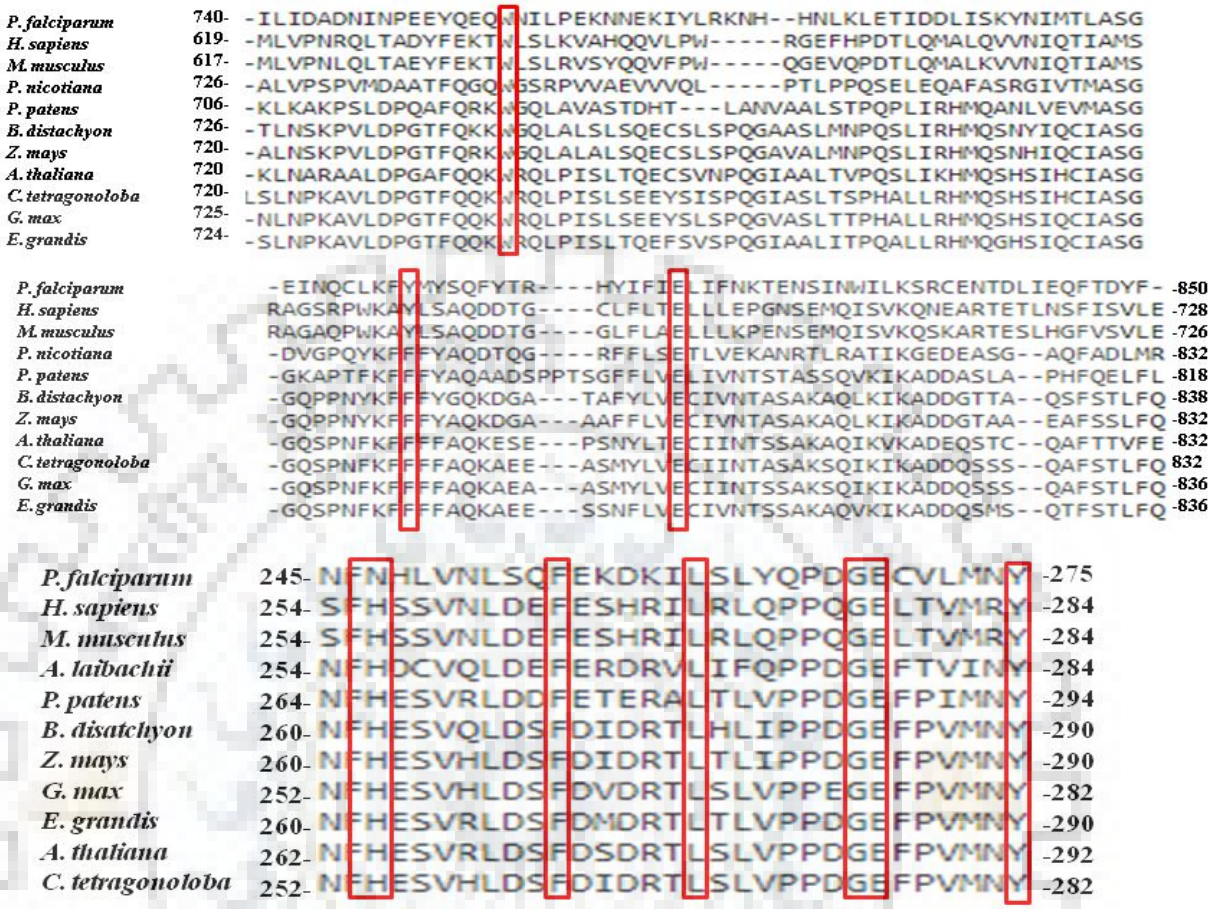


Fig. 4.6. Multiple sequence alignments of the amino acid sequences of the beta and mu subunits of AP-4 complex of *Plasmodium falciparum*, *Homo sapiens*, *Mus musculus*, *Phytophthora nicotiana*, *Physcomitrella patens*, *Brachypodium disatchyon*, *Zea mays*, *Arabidopsis thaliana*, *Glycine max*, *Cyamopsis tetragonoloba* and *Eucalyptus grandis* using CLUSTAL O. (A) Multiple sequence alignment of the amino acid sequences of the beta subunit. The red boxes show critical amino acids (Trp^{635}, Phe^{682} and Glu^{696}) of *H. sapiens* and the corresponding amino acids of other species (B) Multiple sequence alignment of the amino acid sequences of the mu subunit. The red boxes show critical amino acids (Phe^{255}, His^{256}, Phe^{264}, Lys^{270}, Gly^{277}, Glu^{278} and Tyr^{284}) of *H. sapiens* and the corresponding amino acid residues of other species.

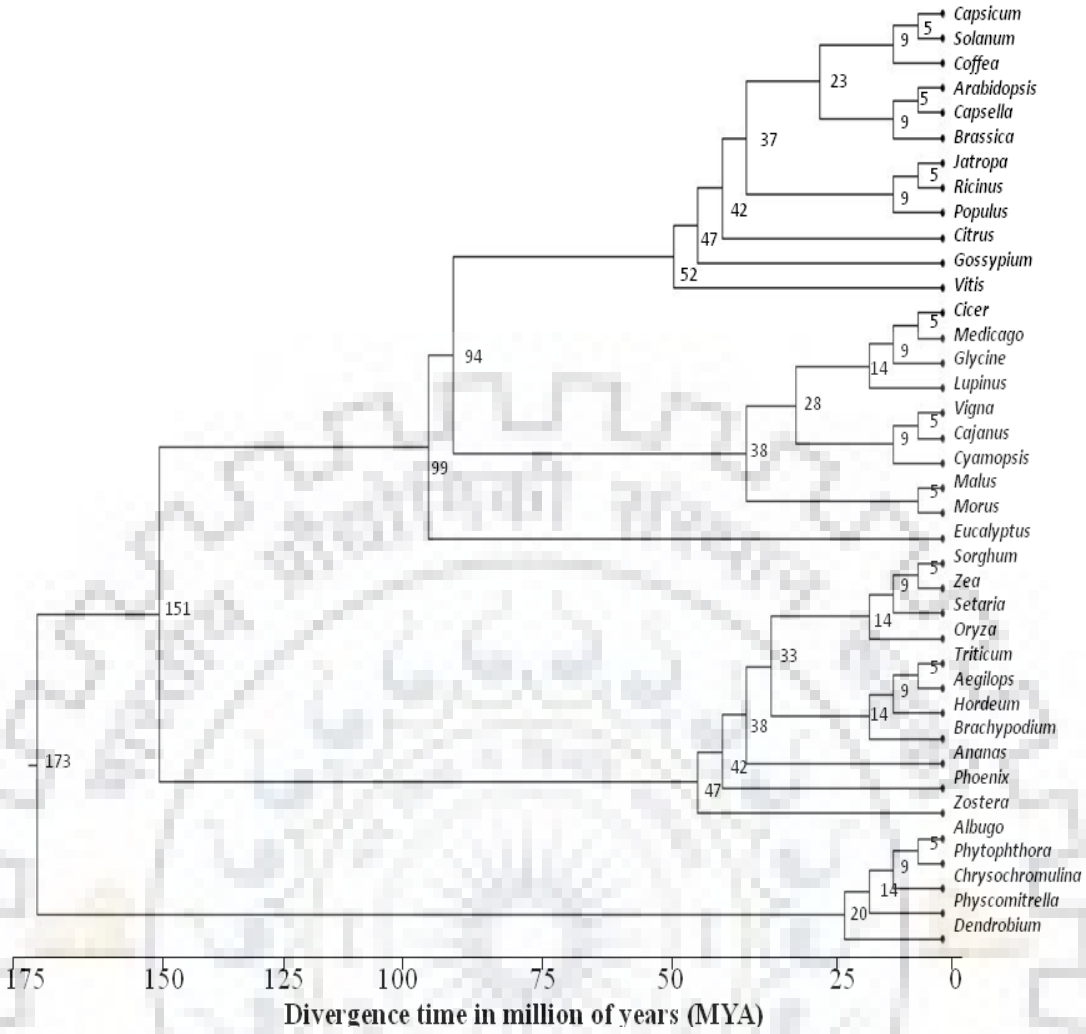


Fig. 4.7. Estimated divergence time values for various plants on the basis of sequence analysis of gene coding for epsilon subunit of AP-4 complex. The tree was built using Bayesian MCMC analysis with a relaxed molecular clock approach with 20 million MCMC steps.

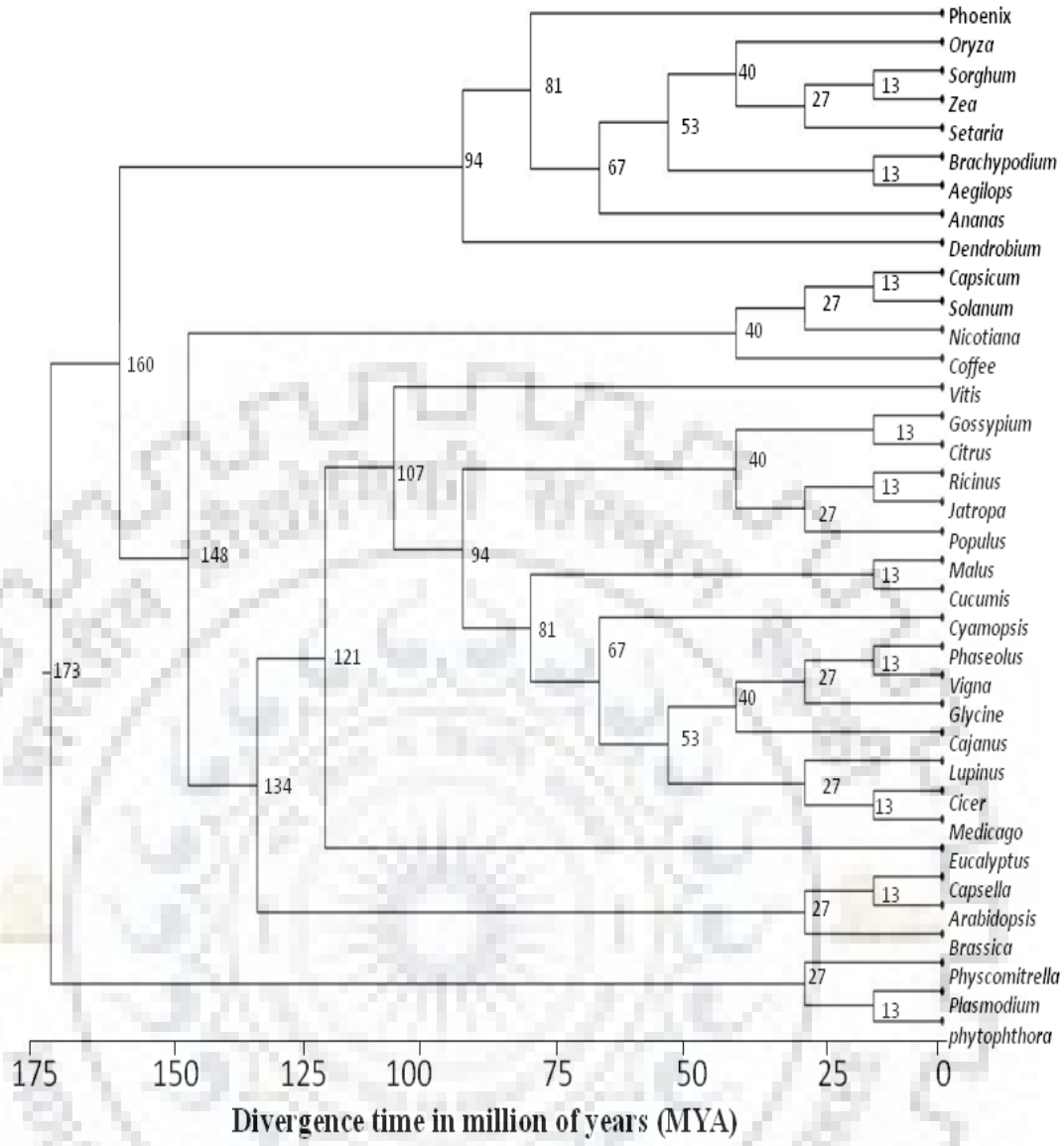


Fig. 4.8. Estimated divergence time values for various plants on the basis of sequence analysis of gene coding for beta subunit of AP-4 complex. The tree was built using Bayesian MCMC analysis with a relaxed molecular clock approach with 20 million MCMC steps.

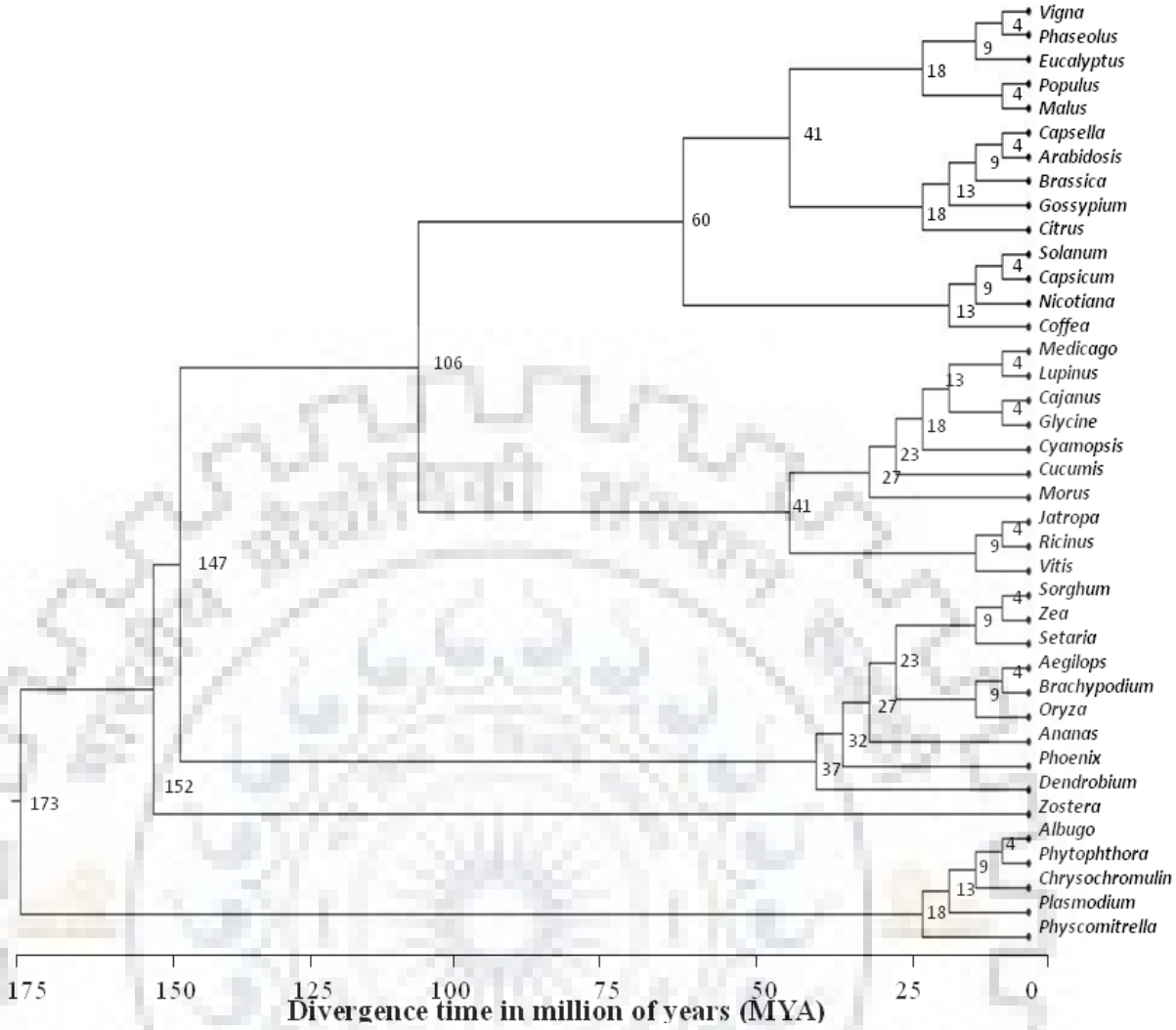


Fig. 4.9. Estimated divergence time values for various plants on the basis of sequence analysis of gene coding for mu subunit of AP-4 complex. The tree was built using Bayesian MCMC analysis with a relaxed molecular clock approach with 20 million MCMC steps.

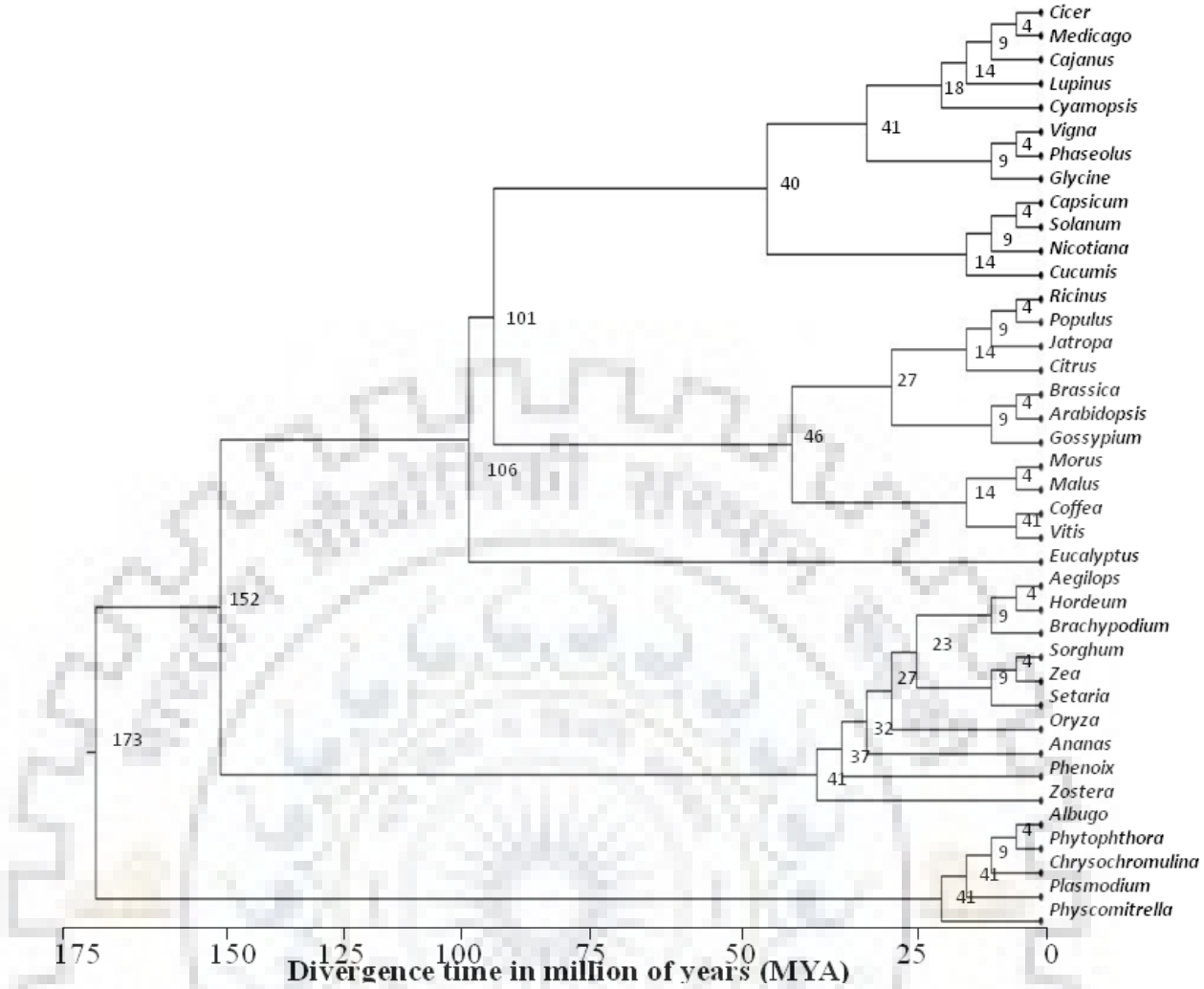


Fig. 4.10. Estimated divergence time values for various plants on the basis of sequence analysis of gene coding for sigma subunit of AP-4 complex. The tree was built using Bayesian MCMC analysis with a relaxed molecular clock approach with 20 million MCMC steps.

4.2.6. Functional divergence analysis of AP-4 complex subunits

The results of functional divergence analysis of AP-4 subunits have been depicted in the Tables 4.5-4.7. The Type I functional divergence in protein sequences of epsilon subunits was observed between I/III and II/III pairs (Table 4.5). In beta and sigma subunits, Type I functional divergence was found between I/III, I/IV, II/III, II/IV and III/IV pairs, and between I/II and I/IV pairs, respectively (Tables 4.6 and 4.7). The Type II functional divergence was noticed for epsilon (between I/III pair), beta (between I/III, I/IV, II/III and II/IV pairs) and sigma (between I/II and I/III pairs) subunits. The mu subunit did not show any functional divergence. The numbers of critical amino acid sites (CAAS) for Type I divergence were one for epsilon subunit, 15 for beta subunit and 145 for sigma subunit. For Type II functional divergence, the CAAS were 7, 8 and 12 for epsilon, beta and sigma subunits, respectively (Tables 4.5-4.7).

Table 4.5. Functional divergence of epsilon subunit of AP-4 complex in studied plant species

Comparison	Type-I				Type-II		
	θI	θ_{SE}	LRT	Qk > 0.85	θII	θ_{SE}	Qk > 0.85
Group I vs. Group II	0.1688	0.0867	3.7905	0	-0.0194	0.0374	0
Group I vs. Group III	0.3160	0.0767	16.9340**	1	0.0306	0.0386	7
Group I vs. Group IV	0.1532	0.0943	2.6381	0	0.0049	0.0335	0
Group II vs. Group III	0.2209	0.0863	6.5441**	0	-0.0118	0.0376	0
Group II vs. Group IV	0.0010	0.0223	0.000	0	-0.0330	0.0314	0
Group III vs. Group IV	0.1969	0.1088	3.2715	0	-0.0060	0.0332	0

θI and θII are coefficients of type I and type II functional divergence between two gene clusters, θ_{SE} Standard error of the maximum likelihood estimate of θI

LRT is likelihood ratio statistic and Qk is posteror probability

LRT> 6.147

θI > 0

** P < 0.01

Table 4.6. Functional divergence of beta subunit of AP-4 complex in studied plant species

Comparison	Type-I				Type-II		
	θI	θ_{SE}	LRT	Qk>0.85	θII	θ_{SE}	Qk>0.85
Group I vs. Group II	0.2208	0.0979	5.0780	0	-0.0255	0.0383	0
Group I vs. Group III	0.2800	0.0635	19.387**	3	0.0103	0.0384	2
Group I vs. Group IV	0.2192	0.0629	12.1246**	0	0.0137	0.0393	2
Group II vs. Group III	0.4232	0.0749	31.8768**	7	0.0166	0.0252	2
Group II vs. Group IV	0.3848	0.0721	28.4544**	4	0.0205	0.0269	2
Group III vs. Group IV	0.2024	0.0563	12.908**	1	0.0029	0.0286	0

θI and θII are coefficients of type I and type II functional divergence between two gene clusters, θ_{SE} Standard error of the maximum likelihood estimate of θI

LRT is likelihood ratio statistic and Qk is posteror probability

LRT> 6.147

θI > 0

** P < 0.01

Table 4.7. Functional divergence of sigma subunit of AP-4 complex in studied plant species

Comparison	Type-I				Type-II		
	θI	θ_{SE}	LRT value	Qk>0.85	θII	θ_{SE}	Qk>0.85
Group I vs. Group II	0.78246	0.21854	12.8186**	10	0.0765	0.0354	0
Group I vs. Group III	0.48320	0.25857	3.4921	1	0.0499	0.0438	5
Group I vs. Group IV	1.09314	0.25146	18.897**	133	0.0690	0.0357	7
Group II vs. Group III	0.00100	0.02236	0.000	0	-0.0150	0.0389	0
Group II vs. Group IV	0.50628	0.20590	6.0459	1	0.0092	0.0285	0
Group III vs. GroupIV	0.31680	0.26045	1.4794	0	-0.0013	0.0394	0

θI and θII are coefficients of type I and type II functional divergence between two gene clusters, θ_{SE} Standard error of the maximum likelihood estimate of θI

LRT is likelihood ratio statistic and Qk is posteror probability

LRT> 6.147

θI > 0

** P < 0.01

4.3. GC contents and CpG islands

Each of the four genes coding for AP-4 complexes in monocots and *P. patens* (moss) contained high GC content as compared to that in the respective gene for AP-4 complexes of dicots (Figs. 4.11-4.14). A CpG island of 627 bp (49 nt to 675 nt) was found in the gene of epsilon subunit of *Z. mays* (Fig. 4.11). The epsilon subunit gene of *G. max* was found to contain a CpG island of 567 bp (284 nt to 850 nt) (Fig.4.11). A relatively smaller CpG island of 209 bp (3801 nt to 4009) was observed in the epsilon subunit gene of *A. thaliana* (Fig. 4.11). The beta subunit gene of *Z. mays* contained a CpG island of 671 bp (48 nt to 718 nt) (Fig. 4.12). In *Z. mays*, three CpG islands of different lengths 906 bp, 523 bp and 1793 bp were found in the gene for mu subunit while 4 CpG islands of 570bp, 629bp, 208bp and 203bp lengths were present in sigma subunit gene (Figs. 4.13 and 4.14).

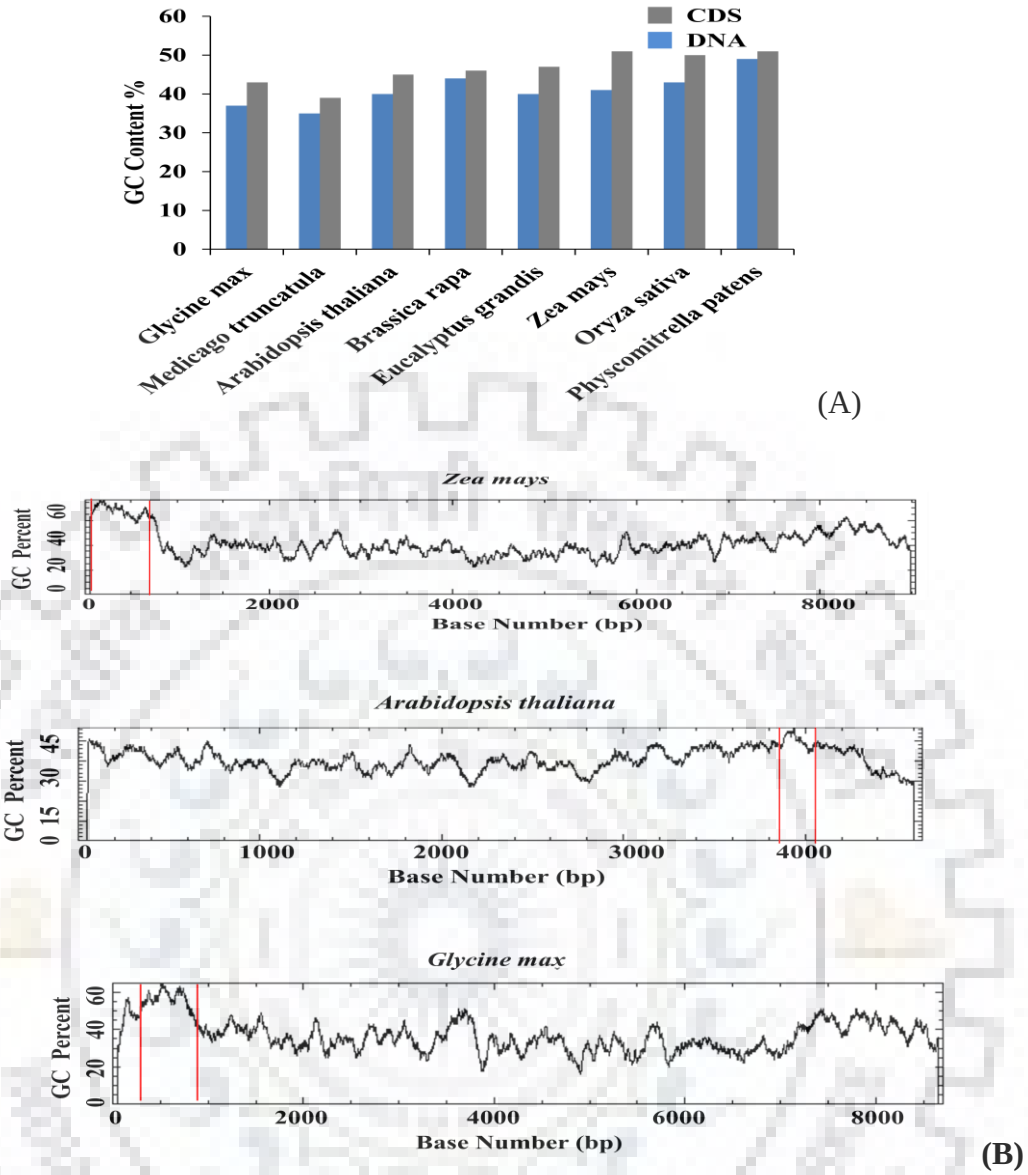


Fig. 4.11 (A) GC contents for genes encoding epsilon subunits of AP-4 complexes of moss, monocots and dicots. (B) CpG islands in genes of epsilon subunits were identified with CpGplot of EMBOSS (http://www.ebi. ac.uk/Tools/emboss/) server with the following parameters: a gap of 100 bp, Observed/Expected ratio >0.60, Percent C+ Percent G >50.00 and Length >200. One CpG island was found in *Zea mays* 627 bp (49 nt to 675 nt), *Arabidopsis thaliana* 209 bp (3801 nt-4009 nt) and *Glycine max* 567 bp (284 nt to 850 nt).

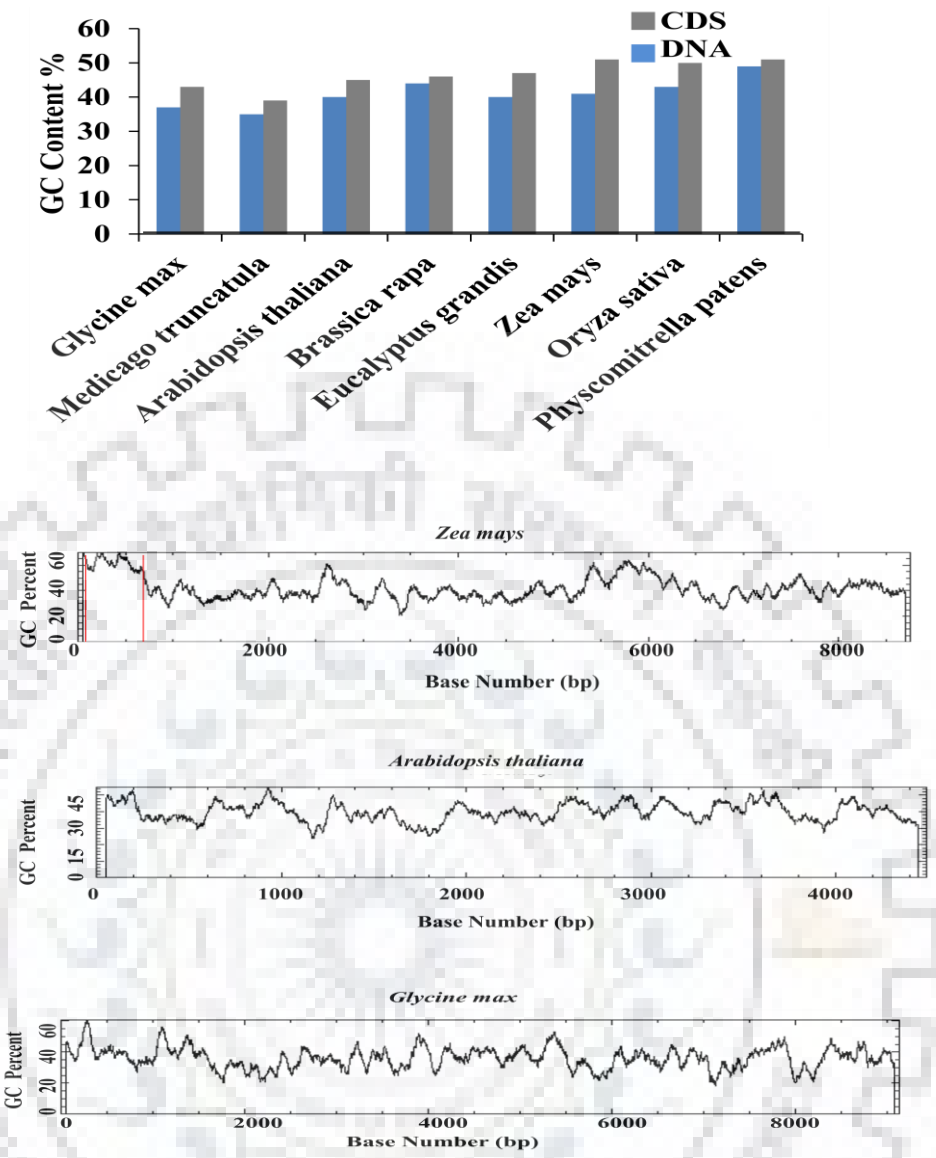


Fig. 4.12. GC content and CpG islands of beta subunit of AP-4 complex: (A) GC content for genes encoding beta subunits of AP-4 complexes of moss, monocots and dicots. (B) CpG islands in genes of beta subunits were identified with CpGplot of EMBOSS (http://www.ebi.ac.uk/Tools/emboss/) server with the following parameters: gap of 100 bp, Observed/Expected ratio >0.60, Percent C+ Percent G >50.00 and Length >200. One CpG island was found in *Zea mays* 671bp (48 nt to 718 nt) while no CpG island was observed in *Arabidopsis thaliana* and *Glycine max.*

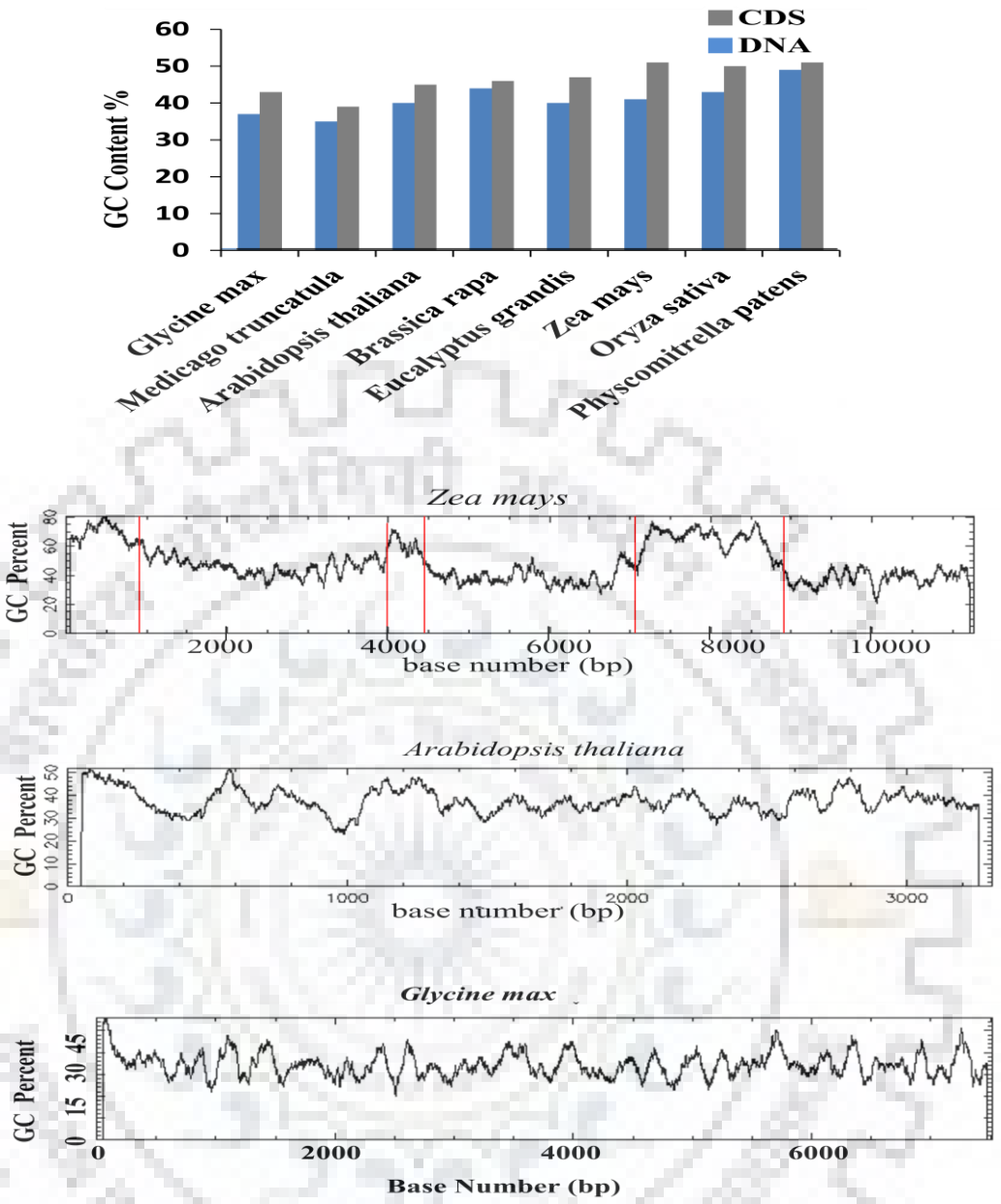


Fig. 4.13. (A) GC contents for genes encoding mu subunits of AP-4 complexes of moss, monocots and dicots. (B) CpG islands in genes of mu subunits were identified with CpGplot of EMBOSS (http://www.ebi. ac.uk/Tools/emboss/) server with the following parameters: gap of 100 bp, Observed/Expected ratio >0.60, Percent C+ Percent G >50.00 and Length >200. Three CpG islands were found in *Zea mays* ranging from 906 bp (48 nt to 953 nt), 523 bp (3940 nt to 4463nt) and 1793 bp (708 nt to 8890 nt) while no CpG island was observed in *Arabidopsis thaliana* and *Glycine max.*

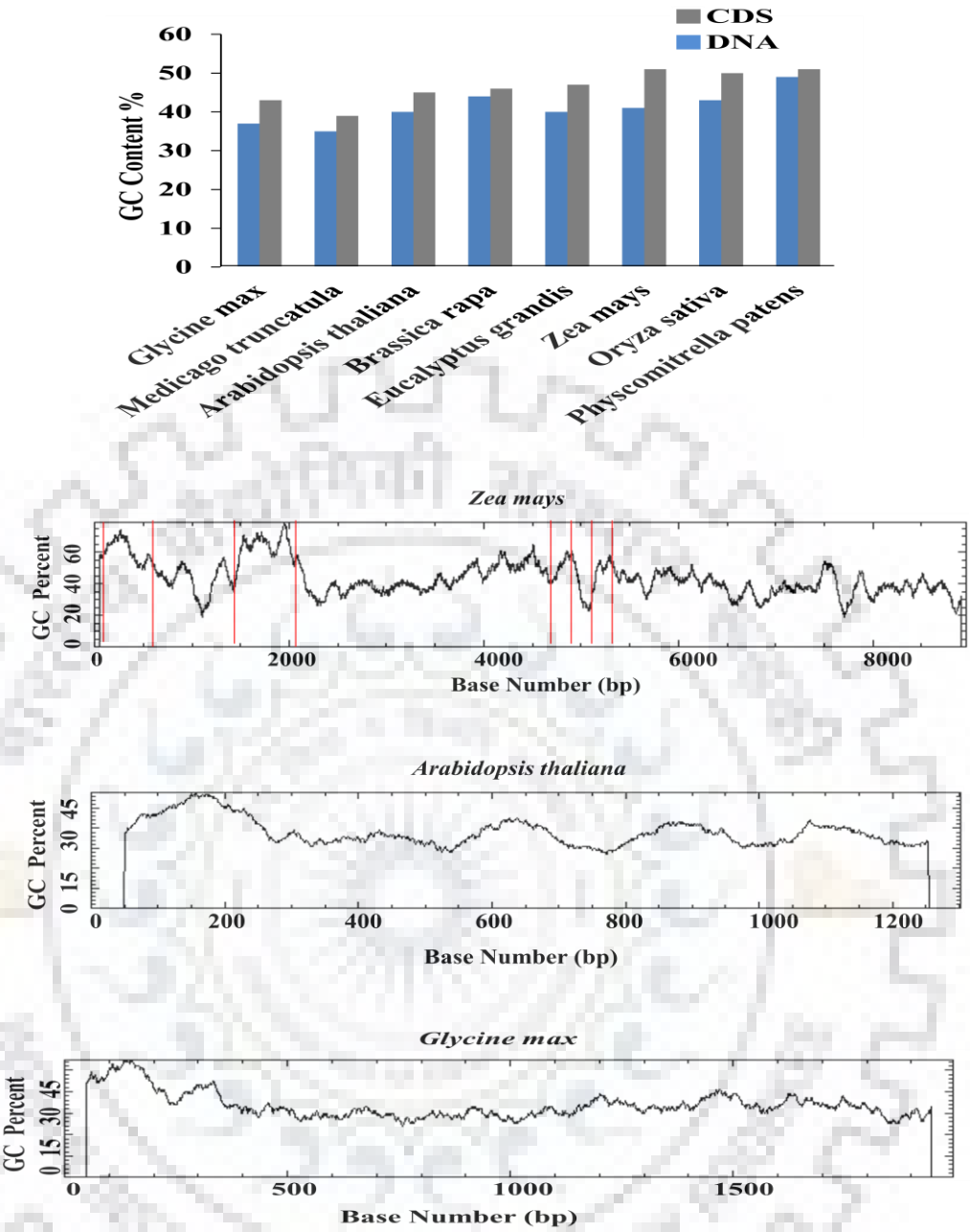


Fig. 4.14. (A) GC contents for genes encoding sigma subunits of AP-4 complexes of moss, monocots and dicots. (B) CpG islands of genes of sigma subunits were identified with CpGplot of EMBOSS (http://www.ebi. ac.uk/Tools/emboss/) server using the following parameters: a gap of 100 bp, Observed/Expected ratio >0.60, Percent C+ Percent G >50.00 and Length >200. Four CpG islands ranging from 570bp (49 - 618), 692bp (1455-2146), 208bp (4749-4956) and 203bp (5149-5351) were found in *Zea mays*, while no CpG island was observed in *Arabidopsis thaliana* and *Glycine max.*

4.4. *Cis* regulatory elements (CREs) and transcription factors

The promoters of AP-4 complex genes of *A. thaliana*, *Z. mays* and *G. max* contained 116, 122 and 136 CREs, respectively. The common CREs identified in the promoter regions of genes coding for epsilon, beta, mu and sigma subunits of AP-4 complexes of *A. thaliana*, *Z. mays* and *G. max* have been shown in Table 4.8. These CREs were categorized into three groups, namely, stress response, cellular development and hormonal regulation, based on their biological functions (Fig. 4.15). The more frequent CREs were GATABOX, IBOXCORE, GT1CONSENSUS, MYCOCCONESUSAT, BIHD1OS, GT1GMSCAM4, CAATBOX, CACTFTPPCA1, POLNEN1LELAT52, DOFCOREZM, SEF4MOTIFGM7S, EBOXBNNAPA, ROOTMOTIFTAPOX1, ARR1AT and WRKY71OS. The transcription factors binding with these CREs were found to be myb, E2F, HD-ZIPIII, homeodomain, bZIP, ARF, AP2-EREBP, ARR, MYC and WRKY (Table 4.8).

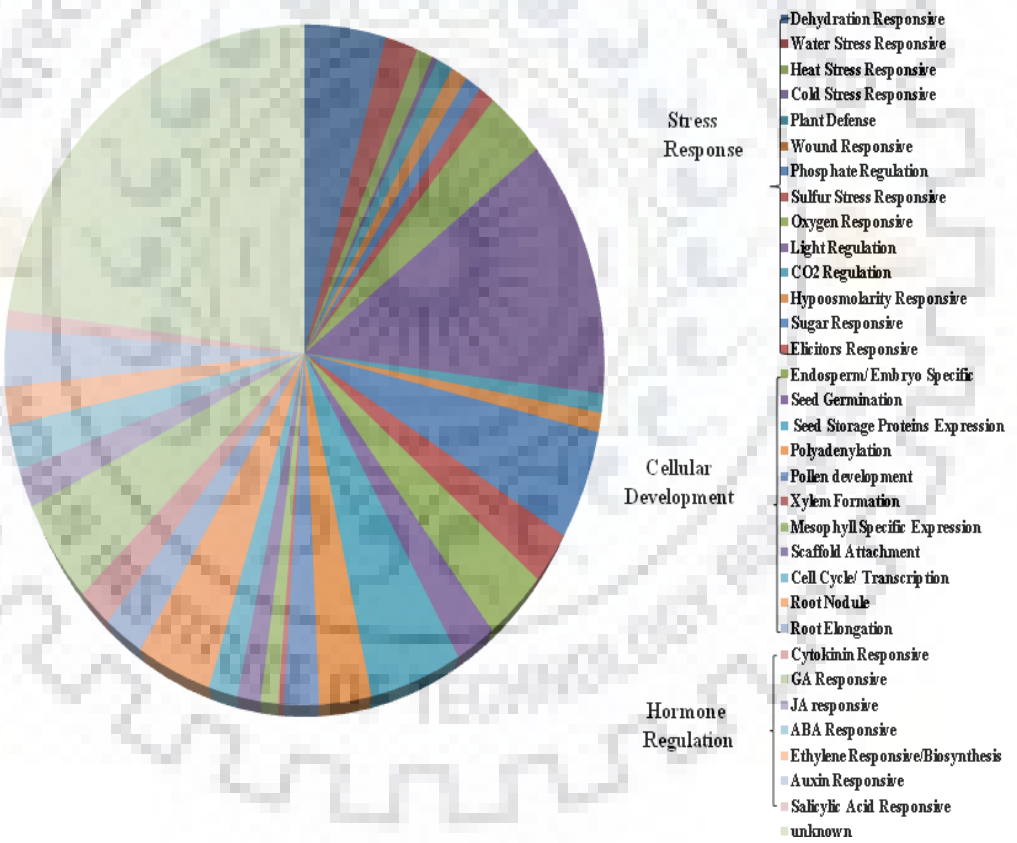


Fig. 4.15. Pie distribution of identified *cis* regulating elements in promoters of genes coding for epsilon, beta, mu and sigma subunits of AP-4 complexes from *Arabidopsis thaliana*, *Zea mays* and *Glycine max* with the PLACE and PlantPAN2.0 softwares based on their biological functions.

Table 4.8. Characterstics of *cis* regulating elements in promoter region of genes for AP-4 complex of *Arabidopsis thaliana, Zea mays* and *Glycine max*

Sr. No.	*Cis* Regulating Element	Sequence	Function	Interacting Transcription Factor
1	ACGTATERD1	5'-ACGT-3'	Dehydration responsive	MYB
2	ARR1AT	5'-NGATT-3'	Cytokinin	ARR1
3	BOXIINTPATPB	5'-ATAGAA-3'	In gene of chloroplast	-
4	BIHD1OS	5'-TGTCA-3'	Plant defence system	Homeodomain
5	CAATBOX1	5'-CAAT-3'	Expression of legumin gene	-
6	CACTFTPPCA1	5'-YACT-3'	Key component of mesophyll expression module	-
7	CCAATBOX1	5'-CCAAT-3'	Heat shock protein	-
8	DOFCOREZM	5'-AAAG-3'	Endosperm-specific element	C2C2-Dof
9	EBOXBNNAPA	5'-CANNTG-3'	Storage protein	bHLH/MYB
10	GATABOX	5'-GATA-3'	Seed specific elemnet	GATA
11	GTGANTG10	5'-GTGA-3'	Pollen expression element	-
12	GT1CONSENSUS	5'-GRWAAW-3'	Light regulation	trihelix
13	GT1GMSCAM4	5'-GAAAAA-3'	Light responsive elemnet	GT-1-like
14	IBOXCORE	5'-GATAA-3'	Light regulated gene	-
15	INRNTPSADB	5'-YTCANTYY-3'	Light responsive	-
16	MARTBOX	5'-TTWTWTTWTT-3'	Scaffold attachment	-
17	MYCCONSENSUSAT	5'-CANNTG-3'	Cold responsive gene	MYC
18	NODCON2GM	5'-CTCTT-3'	Nodule expression	-
19	OSE2ROOTNODULE	5'-CTCTT-3'	Root nodule	-
20	POLASIG1	5'-AATAAA-3'	Polyadenylation	-
21	POLASIG3	5'-AATAAT-3'	Polyadenylation	-
22	POLLEN1LELAT52	5'-AGAAA-3'	Pollen specific element	-
23	RAV1AAT	5'-CAACA-3'	Rosette leaves and root specific	RAV1
24	ROOTMOTIFTAPOX1	5'-ATATT-3'	Root nodule	-
25	SEF4MOTIFGM7S	5'-RTTTTTR-3'	beta-conglycinin Interaction	SEF4
26	TAAAGSTKST1	5'-TAAAG-3'	Guard cell expression	StDof
27	TATABOX5	5'-TTATTT-3'	Regulation of glutamine synthetase	TFIID
28	WRKY71OS	5'-TGAC-3'	Gibberellin signaling response	WRKY

4.5. SNP analysis in *Z. mays* and soybean

In AP-4 complex genes (AP4E, AP4B, AP4M and AP4S) of *Z. mays*, total 1316 SNPs were found. The number of transitions (Ts) and transversions (Tv) 854 and 462 were, respectively. The ration of Ts/Tv was 1.85. Total 38 non- synonymous SNPs were found among 1316 SNPs (Appendix 7). Out of 38 non- synonymous SNPs, 22 were observed in functional binding pockets of epsilon (6), beta (14), mu (1) and sigma (1). The analysis of non-synonymous SNPs by PANTHER software revealed that out of these 22 SNPs, 2 SNPs in epsilon (A942 and S476) (Fig. 4.16) and 1 in sigma (K57R) (Fig. 4.17) subunits were showing probably damaging effects.

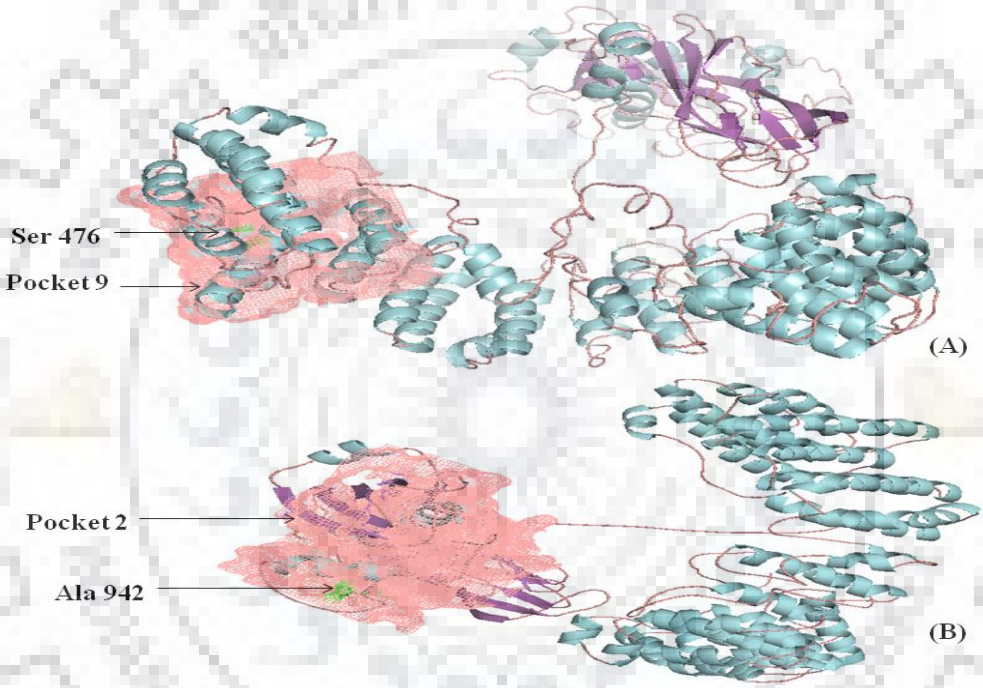


Fig.4.16. 3-D structures of epsilon subunit of *Zea mays* showing the amino acid residues substituted by SNPs, (A) The amino acid residues (Ser 476) of epsilon subunit substituted by SNPs located on pocket 9 and (B) The amino acid residue (Ala 942) of epsilon subunit substituted by SNP located on pocket 2. The alpha helices, B-sheets, and turns are shown in cyan, magenta and orange colors, respectively. The pockets in the 3-D structures are displayed with a mesh model.

In soybean, 156 SNPs were observed in the genes encoding various subunits of AP-4 complex. The number of transitions (Ts) and transversions (Tv) were 80 and 76, respectively (Fig 4.19). The ratio of Ts/Tv was 1.05. Seventy percent SNPs were found in intron region, 8 % in 3' UTR, 1% in 5'UTR and 20 % in the coding region. In soybean, 22 SNPs resulted in non-synonymous and 9 synonymous substitutions (Appendix 6). The ratio of non-synonymous to synonymous substitutions was 2.5. The PANTHER software analysis of the SNPs revealed that out of 22 non-synonymous SNPs, 8 were showing probably damaging effects. Out of the 22 non-synonymous SNPs, 9 were found to be in the functional binding pockets of epsilon (7), beta (1) and mu (1). In epsilon subunit of these 9 SNPs, 3 SNPs (P78S in mu and G153S, A156T in epsilon subunits) of *G. max* may be causing damaging effects (Fig. 4.18A and B).

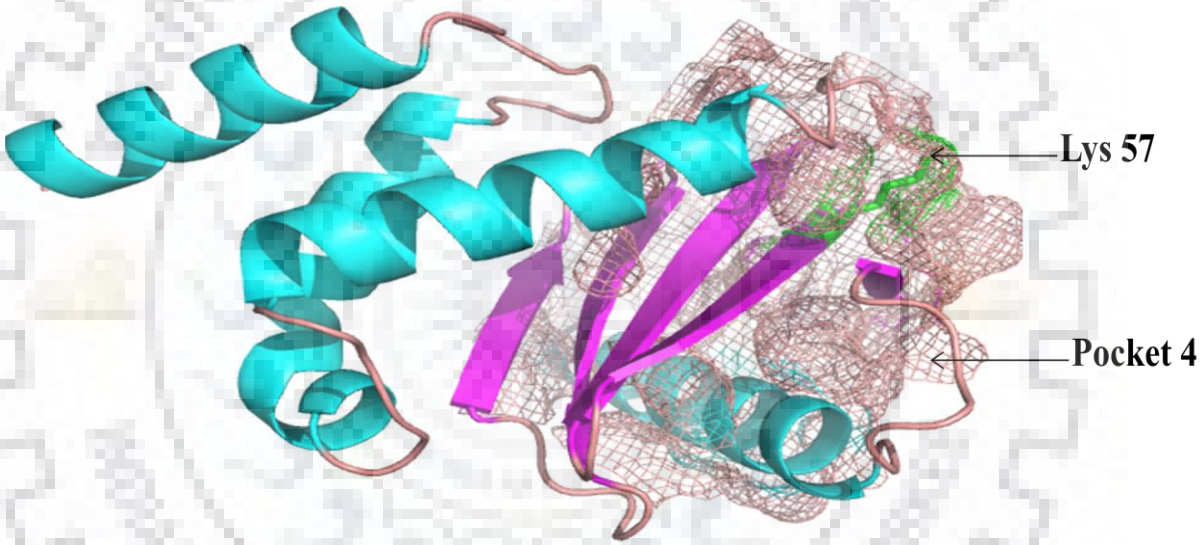


Fig. 4.17. 3-D structures of beta and sigma subunits of *Zea mays* showing the amino acid residues (Lys 57) of sigma subunit substituted by SNP located on pocket 4. The alpha helices, B-sheets, and turns are shown in cyan, magenta and orange colors, respectively. The pockets in the 3-D structures are displayed with a mesh model.

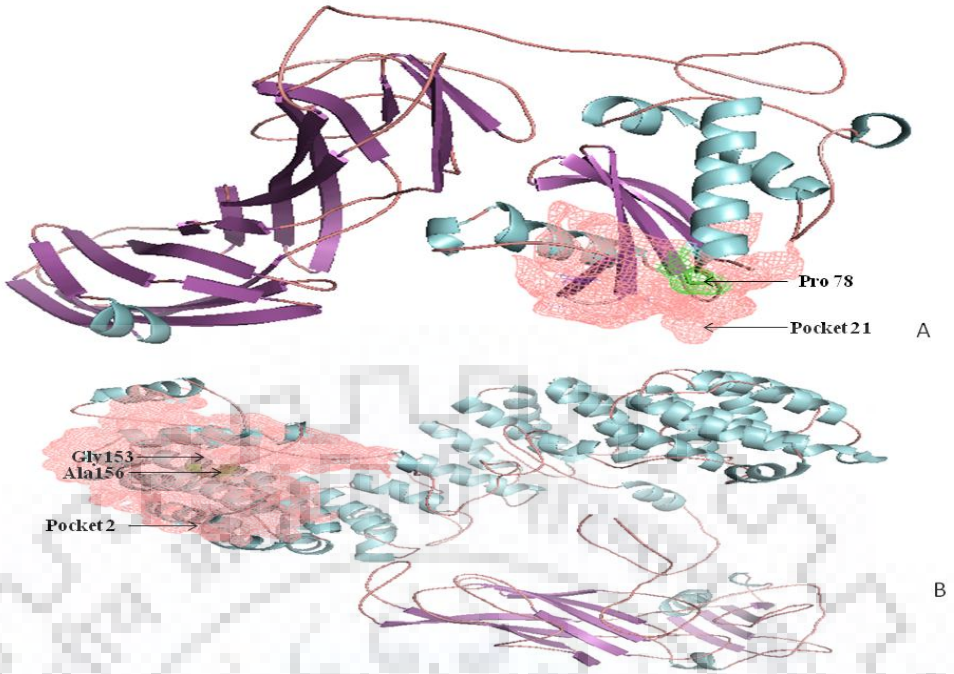


Fig. 4.18. 3-D structures of epsilon and mu subunits of *Glycine max* showing the amino acid residues substituted by SNPs, (A) The amino acid residues (Gly 153 and Ala 156) of epsilon subunit substituted by SNPs located on pocket 2 and (B) The amino acid residue (Pro78) of mu subunit substituted by SNP located on pocket 21. The alpha helices, B-sheets, and turns are shown in cyan, magenta and orange colors, respectively. The pockets in the 3-D structures are displayed with a mesh model.

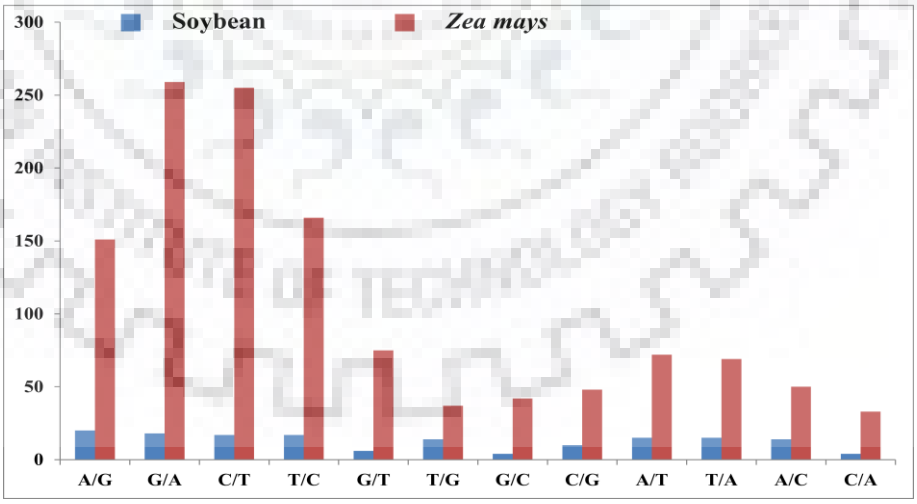


Fig. 4.19. The frequency distribution of different nucleotide substitution types (transitions and transversions) in *Zea mays* and soybean. In *Z. mays*, all substitutions were more in number in comparison to soybean

4.5.1. SNP mining by restriction enzymes

Total three restriction sites were identified in coding regions of AP-4 complex genes of *Z. mays* and soybean. In gene coding for sigma subunit of *Z. mays*, 2 restriction sites for Csp6I and BPMI restriction enzymes were predicted at positions G57A and T343C, respectively. In gene encoding epsilon subunit of soybean, MfeI restriction enzyme acts on T7C position.

4.6. High expression of AP-4 complex genes in seeds

The fold change expression of the *A. thaliana*, *Z. mays* and *G. max* genes encoding epsilon, beta, mu and sigma subunits of AP-4 complexes in different tissues compared to seed have been shown in the Fig. 4.20. In *A. thaliana*, the genes of beta and mu subunits had more expression in seed as compared to all other studied tissues whereas the epsilon subunit gene showed high expression in seed as well as flower. High expression of the sigma subunit gene was observed in seed, siliqua and flower of *A. thaliana*; the expression in flower and siliqua was more as compared to that in seed (Fig. 4.20A). In *Z. mays*, high expression of genes coding for epsilon, beta and sigma subunits was observed in seed. The mu subunit gene of *Z. mays* had high level of expression in seed as well as anther; the expression in anther was more than that in seed (Fig. 4.20B). The genes of epsilon, beta and sigma subunits of *G. max* AP-4 complex had high expression in seed. The *G. max* mu subunit gene showed high expression in seed as well as flower, the expression in flower was more than that in seed (Fig. 4.20C).

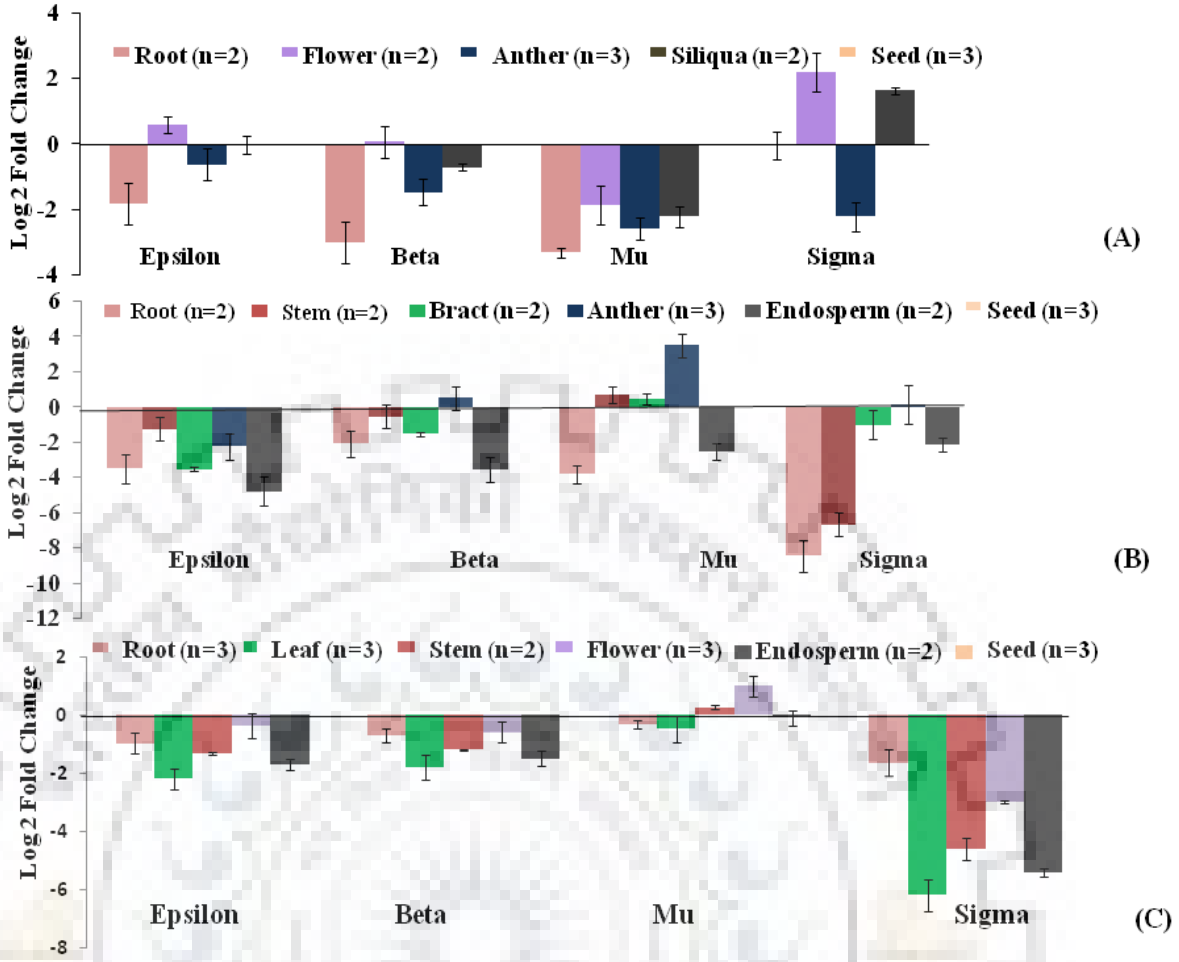


Fig. 4.20. Fold change expression of the genes encoding epsilon, beta mu and sigma subunits of AP-4 complex in different tissues of (A) *Arabidopsis thaliana* compared to seeds, (B) *Zea mays* compared to seed and (C) *Glycine max* compared to seeds. A bar represents standard error of the mean.

4.7. Real time PCR analysis of AP-4 complex genes

Total RNA was isolated from the different tissues of guar (root, leaf, flower and seed), *Z. mays* (root, bract, anther and seed) and *G. max* (root, leaf, flower and seed). The cDNA was synthesized from each isolated RNA sample. The confirmation of the cDNA synthesis in each sample of was carried out by beta-actin gene amplification by PCR. The results of semi-quantitative PCR showed the expression of genes coding for epsilon, beta, mu and sigma subunits of AP-4 complex in guar, *Z. mays* and *G. max* (Fig. 4.21). Tissue specific expression of AP-4 complex genes of guar, *Z. mays* and *G. max* in comparison to seed tissue was also studied by quantitative real time PCR. The results of qRT-PCR experiments revealed that in *Z. mays* expression levels of genes coding for beta and mu subunits were higher in anther than in other tissues, whereas the genes coding for epsilon and sigma subunits showed high expression

in seed (Fig. 4.22A). In roots of *Z. mays*, the expression levels of the AP-4 complex genes were lowest as compared to other tissues. In *G. max*, the levels of expression of epsilon and mu genes of AP-4 complex were found to be higher in flower tissue in comparison to other tissues (Fig. 4.22B). The high expression of beta and sigma genes was observed in seed. The expression of AP-4 complex genes of *G. max* was lowest in leaf as compared to other tissues. The qRT-PCR studies in guar showed that the levels of expression of genes encoding epsilon and beta subunits were higher in seed tissues. The mu subunit had high levels of expression in flower and seed tissues while sigma subunit showed higher expression in leaf as compared to other tissues (Fig. 4.22C) The expression levels of AP-4 complex genes in root tissue of guar was lowest in comparision to other tissues.

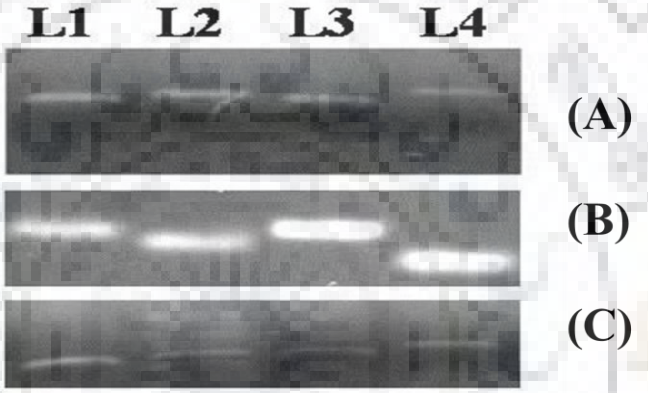
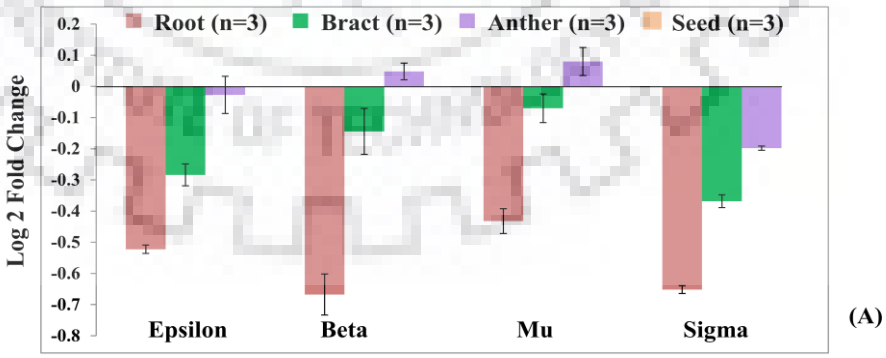
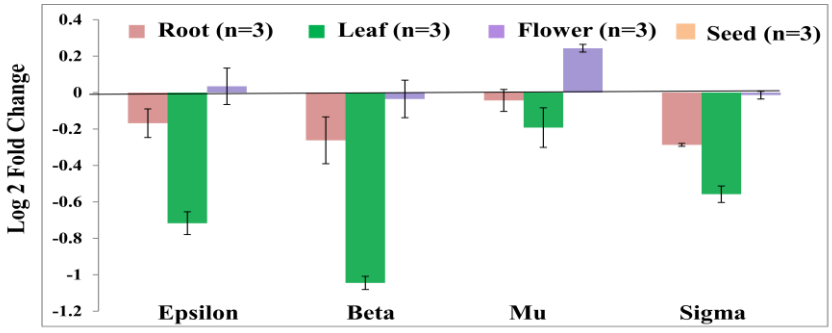
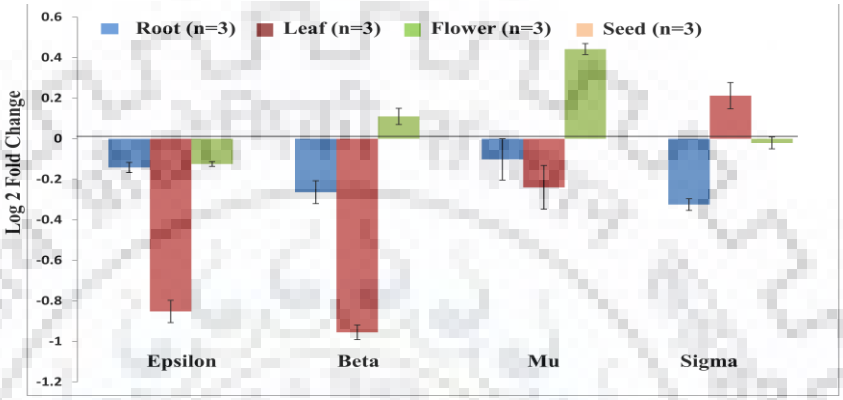


Fig. 4.21. Agarose gel (1%) showing amplification of genes coding for epsilon, beta, mu and sigma subunits of AP-4 complex on cDNA of (A) guar, (B) *Z. mays* and (C) *G.max*. M is molecular marker, L1 AP4E, L2 AP4B, L3 AP4M and L4 AP4S.





(B)



(C)

Fig. 4.22. Fold change expression of the genes encoding epsilon, beta mu and sigma subunits of AP-4 complex in different tissues of (A) *Zea mays* compared to seeds, (B) *Glycine max* compared to seed and (C) guar compared to seeds. A bar represents standard error of the mean.

4.8. Functional annotation of genes

The genes coding for AP-4 complex, ARFs, epsins and VSR1of *A. thaliana* and *Z. mays* were annotated based on their biological process, cellular component and molecular function characteristics (Table 4.9). The common GO terms found in annotated genes were GO: 0097708 (intracellular vesicle) and GO: 0006886 (intracellular protein transport). The results of the GO categories analysis were found to be similar in both monocot and dicot plants. The reactome pathway database revealed that the studied plant genes are involved in *trans*-Golgi network, vesicle budding, ER to Golgi anterograde, Golgi to ER retrograde, intra Golgi and metabolism of proteins.

Table 4.9. GO and KEGG enrichment analyses of the genes involved in AP-4 complex–mediated vesicular trafficking in *Arabidopsis thaliana* and *Zea mays*.

Sr. No.	Gene	Gene Name	KEGG	Biological Function	Molecular Function	Cellular Component (S)
1	AP4E	Epsilon subunit	Membrane trafficking	Intracellular protein transport	Protein binding	AP-4 adaptor complex
2	AP4B	Beta subunit	Membrane trafficking	Intracellular protein transport	Clathrin binding	AP-4 adaptor complex
3	AP4M	Mu subunit	Membrane trafficking	Intracellular protein transport	Protein binding	AP-4 adaptor complex
4	AP4S	Sigma subunit	Membrane trafficking	Protein transport	Protein binding	AP-4 adaptor complex
5	ARFA1A	ADP-ribosylation factor A1A	Membrane trafficking	Intracellular transport	GTP- binding	Intracellular vesicle
6	ARFA1B	ADP-ribosylation factor A1B	Membrane trafficking	Intracellular transport	GTP- binding	Intracellular vesicle
7	ARFA1C	ADP-ribosylation factor A1C	Membrane trafficking	Intracellular transport	GTP- binding	Intracellular vesicle
8	ARFA1E	ADP-ribosylation factor A1E	Membrane trafficking	Intracellular transport	GTP- binding	Intracellular vesicle
9	ARFA1F	ADP-ribosylation factor A1F	Membrane trafficking	Intracellular transport	GTP- binding	Intracellular vesicle
10	ARFB1B	ADP-ribosylation factor B1B	GTP-binding proteins	Intracellular transport	GTP- binding	Intracellular vesicle
11	ARFB1C	ADP-ribosylation factor B1C		Intracellular transport	GTP- binding	Intracellular vesicle
13	Epsin 1	Epsin 1	Membrane trafficking	Protein targeting to vacuole	Clathrin binding	Cytoplasmic vesicle, Golgi apparatus
14	Epsin 2	Epsin 2	Membrane trafficking	Protein transport	Clathrin binding	Cytoplasmic vesicle and Golgi apparatus
15	VSR 1	Vacuolar sorting receptor 1		Golgi to vacuole transport	Amino-terminal vacuolar sorting propeptide binding	Golgi apparatus and Membrane

4.9. Co-expression of genes

The results of RNA-Seq analysis of the *A. thaliana* and *Z. mays* genes coding for epsilon, beta, mu and sigma subunits of AP-4 complexes, ARF family members, epsin 1, epsin 2 and VSR1 in different tissues have been shown in the Figs 4.23 and 4.24. All these genes were found to express in all the studied tissues. In *A. thaliana*, the expression levels of genes encoding AP4E, AP4S, ARFA1A, ARFA1E, ARFA1C, epsin 1, epsin 2 and VSR1 were high in flower. The genes coding for AP4B, AP4M, ARFA1B, ARFA1F, ARFB1B, ARFB1C and ARF3 showed high expression in seed. In anther tissue of *A. thaliana*, the expression levels of all studied genes were found to be low (Fig. 4.24A). The RNA-Seq analysis of *Z. mays* revealed that the levels of expression of genes encoding AP4E, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARF3 and VSR1 were high in seed tissue; all the studied genes showed low expression in stem tissue (Fig. 4.24B).

The data obtained by RNA-Seq was normalized and Z-score values were calculated. The Z- score values of the studied genes have been listed in Tables 4.10 and 4.11. The ARFA1A gene had a positive Z- score value in all studied tissues (except root) of *A. thaliana* and those of *Z. mays*. The ARF3 and ARFA1E genes showed a negative value of Z-score in most of the tissue of *A. thaliana* and *Z. mays*. In anther and seed tissues of *A. thaliana*, positive Z-score values were recorded for the genes encoding AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1E, ARFA1F and epsin 2. The genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1F, epsin 1 and VSR were found to have positive values in reproductive tissues, namely, anther, seed and endosperm of *Z. mays*.

The genes with positive Z-score values in a tissue are supposed to co-express in that tissue. On this basis, in *A. thaliana*, the genes encoding AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1E, ARFA1F, epsin2 and VSR1 proteins were the co-expressing genes in anther tissue while the genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1E, ARFA1F, ARFB1B, epsin 1 and epsin 2 proteins were found to co-express in the seed tissue.

In *Z. Mays*, the genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARF3, epsin 1 and VSR1 proteins were observed to co-express in the anther tissue, whereas the genes encoding AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARF3, epsin 1, epsin 2 and VSR1 proteins were the co-expressing genes in seed. In stem tissue of *Z. mays* and root tissue of *A. thaliana*, co-expression associations of the AP-4/ARFs were least in number as compared to other tissues.

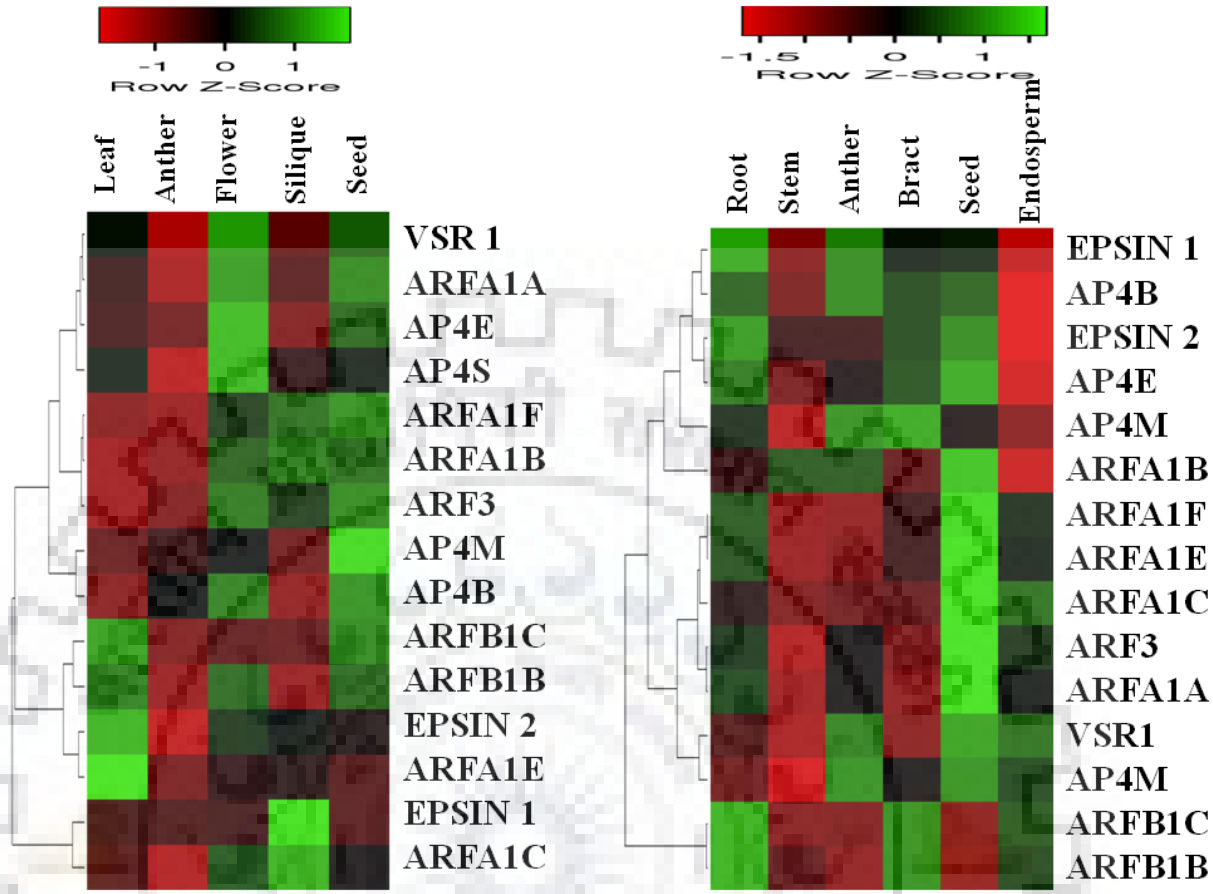


Fig. 4.23. Heatmap showing expression patterns in the form of reads per kilobase of transcript per million mapped reads (RPKM) of genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARFB1C, ARF3, Epsin1, Epsin2 and VSR1 proteins in different tissues of (A) *Arabidopsis thaliana* and (B) *Zea mays*.

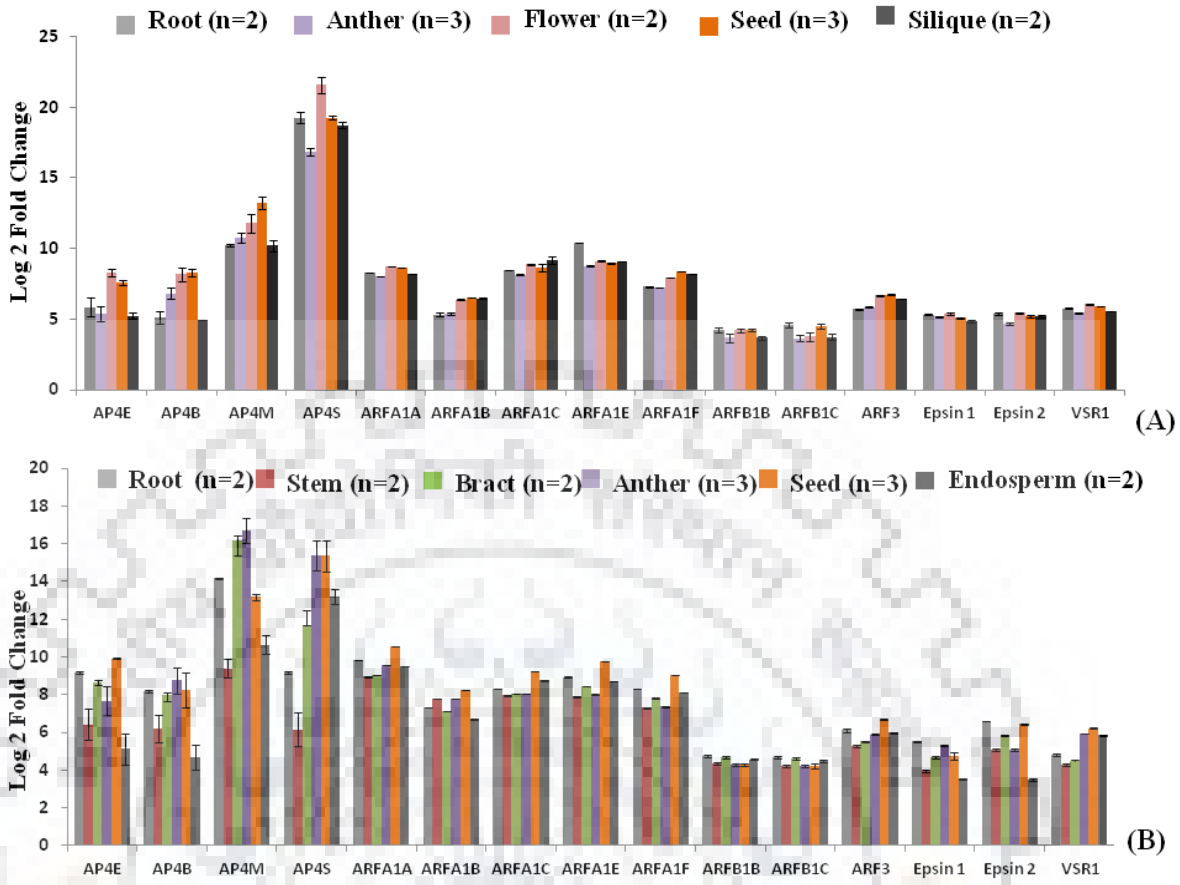


Fig. 4.24. Fold change expression of the genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARFB1C, ARF3, Epsin1, Epsin2 and VSR1 proteins in different tissues of (A) *Arabidopsis thaliana* and (B) *Zea mays*. A bar represents standard error of the mean

Table 4.10. The Z- score values of genes coding for AP-4 complex subunits, ARFs, epsins and VSR1 of *Arabidopsis thaliana*.

Gene Name	Root	Silique	Flower	Anther	Seed
AP4E	0.01	0.01	0.01	0.03	0.01
AP4B	0.01	0.01	0.01	0.01	0.07
AP4M	0.07	0.01	0.16	0.03	0.08
AP4S	0	0.02	0.05	0.01	0.01
ARFA1A	0	0.22	0.17	0.01	0.06
ARFA1B	0	0.02	0	0.01	0.06
ARFA1C	0	0	0.01	0	0
ARFA1E	-0.01	0	0	0.01	0.02
ARFA1F	-0.01	0	0	0.01	0.07
ARFB1B	0.01	0	0.01	0	0.01
ARFB1C	0.01	0	0	0	0
ARF3	0	-0.01	0	0	0
Epsin1	-0.09	0	0.01	0	0.01
Epsin 2	0	0.76	0	0.04	0.01

VSR1	0	0.01	0.01	0.02	0

Table 4.11. The Z- score values of genes coding for AP-4 complex subunits, ARFs, epsins and VSR1 of *Zea mays*.

Gene Name	Root	Stem	Bract	Anther	Seed	Endosperm
AP4E	0.01	0.01	0.01	0.01	0.02	0.03
AP4B	0.01	0	0	0.01	0.02	0.01
AP4M	0.01	0	-0.01	0.02	0.03	0.01
AP4S	0.02	0.01	0	0.01	0.01	0.04
ARFA1A	0.12	0.01	0.01	0.26	0.06	0.4
ARFA1B	0.18	0	-0.08	0.15	0.11	0.02
ARFA1C	0.01	0	0.04	0.08	0.01	0.02
ARFA1E	-0.03	0	0	0.08	0.19	0
ARFA1F	0.02	0	0	0.06	0.05	0.01
ARFB1B	0.01	0.01	0.04	0.01	0.03	0
ARFB1C	-0.01	0	0.02	-0.01	0	0
ARF3	0	0	0	0.01	0.11	-0.01
Epsin1	0.01	0.29	0.03	0.06	0.07	0.11
Epsin 2	0.04	-0.21	0.03	0	0.17	0
VSR1	-0.01	0.01	0	0.05	0.01	0.01

4.10. Real time PCR studies revealed tissue specific expression of vesicular trafficking genes in *Z. mays*

Semi-quantitaive studies of genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFB1B, ARFB1C, epsin 1 and VSR1 was performed along with actin 1 of Z. mays in root, bract, anther and seed tissues (Fig. 4.25). All these genes showed amplification in 4 studied tissues of *Z. mays*. The relative expression profiles of these genes in root, bract, anther, and seed tissues of *Z. mays* have been presented in Figure 4.26. The expression levels of ARFA1A gene were found higher in all tissues as compared to the expression levels of other ARF genes. The expression levels of ARFA1A gene was the highest among all the studied genes and ranged from 3.4 to 3.8 fold. The genes coding for AP4B, AP4S, ARFA1A, ARFA1B and ARFA1E proteins showed high expression levels in seed. In anther, high expression was recorded for the genes coding for proteins, namely, AP4E, AP4M, ARFA1F, ARFB1C and VSR1. The genes for ARFA1C, ARFB1B and epsin 1 proteins showed high expression levels in bract tissue. In all tissues, the expression levels of genes for ARFB1B and ARFB1C proteins were lower than those of all other studied genes.

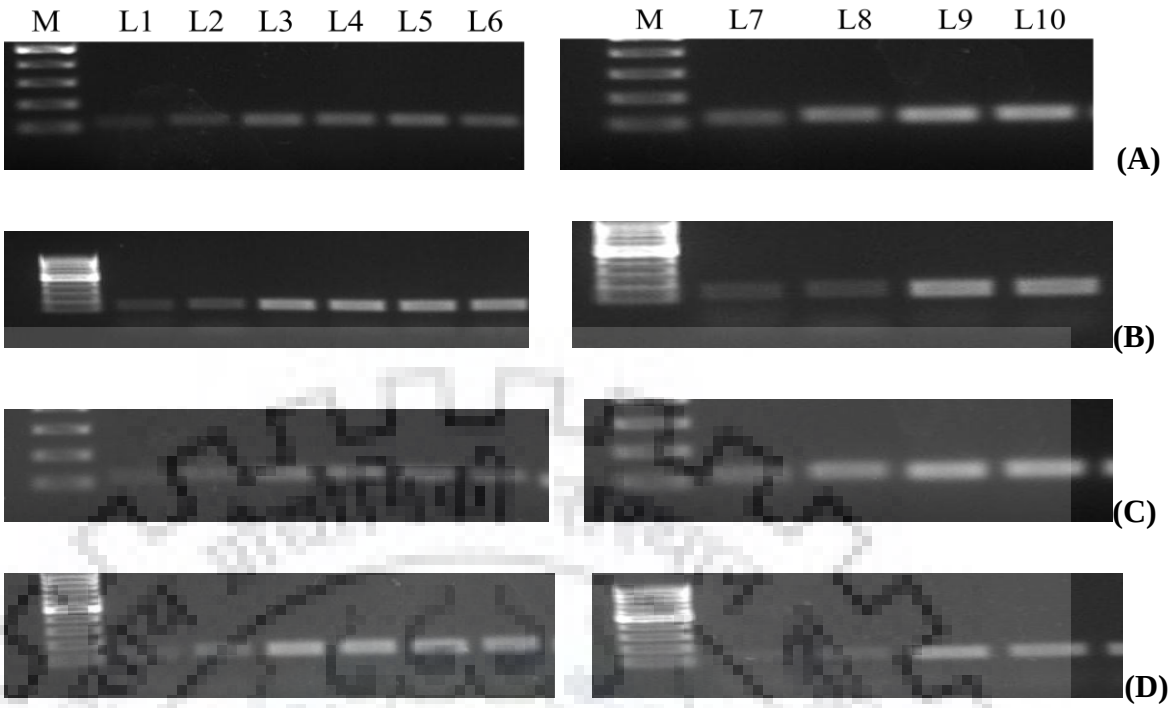


Fig. 4.25. Agarose gel (1%) showing amplification of genes coding for actin, ARFs, Epsin 1 and VSR1 on cDNA of different tissues (A) root, (B) bract, (C) anther and (D) of *Z . mays*. M is molecular marker, L1 actin, L2 ARFA1A, L3 ARFA1B, L4 ARFA1C, L5 ARFA1E, L6 ARFA1F, L7 ARFB1B, L8 ARFB1C, L9 Epsin1 and L10 VSR1.

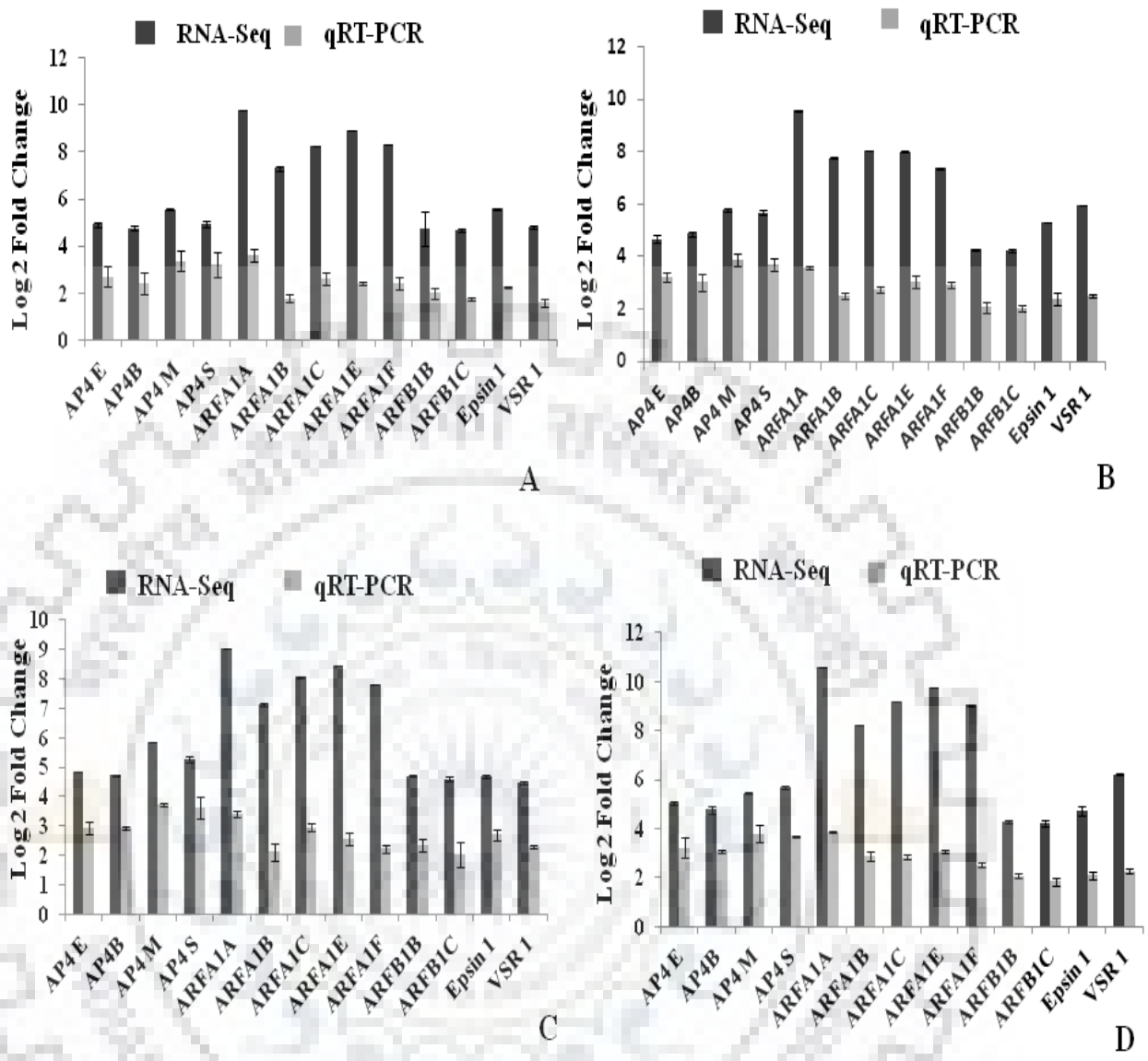


Fig. 4.26. Fold change expression of the genes coding for AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARFB1C, Epsin1 and VSR1 proteins of *Zea mays* by qRT-PCR of (A) root, (B) anther, (C) bract and (D) seed tissues. A bar represents standard error of the mean

4.11. PPIs of co-expressing genes

The protein - protein interactions network of co-expressing genes of *Z. mays* has been shown in Fig. 4.27. A total of 13 nodes, 4 for AP-4 complex (one for each subunit) and 9 for accessory proteins (one for each accessory protein), and 39 edges/interacting combinations (presented in solid lines) were observed in the network. The accessory proteins, namely, ARFA1A, ARFA1B, ARFA1C, ARFA1F, ARFB1B and ARFB1C, were found to interact with AP4E, AP4B, AP4M and AP4S subunits of AP-4 complex. The epsin 1 and epsin 2 proteins showed

interaction with AP4E, AP4B and AP4M subunits. The AP4E, AP4B, AP4M and AP4S proteins had interaction with most of the accessory proteins, thus forming the central hub of PPIs network. No interaction was detected among the ARFA1E, ARF3 and AP-4 complex subunits. The VSR1 protein interacted with epsin 1 protein but not with four subunits of AP-4 complex.

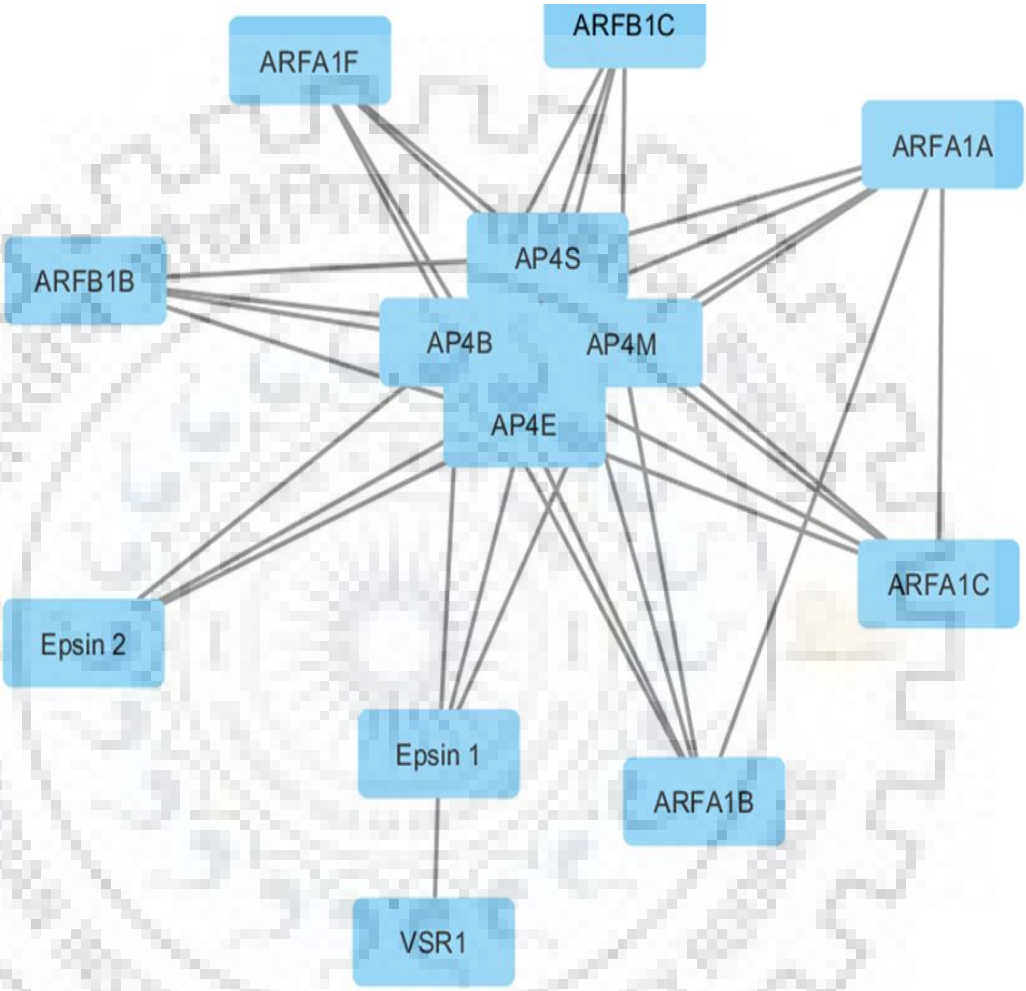


Fig. 4.27. The protein - protein interaction network of AP-4 complex subunits (AP4E, AP4B, AP4M and AP4S), ARFs (ARFA1A, ARFA1B, ARFA1C, ARFA1F, ARFB1B and ARFB1C), epsins (Epsin1 and Epsin 2) and VSR1 protein of *Z. mays*. Blue solid boxes show nodes and black lines represent interacting combinations of co-expressing genes.

4.12. *In silico* analysis of AP-4 complex proteins

4.12.1. Primary structure prediction

The AP-4 complex gene and protein sequences were used for comparing the physico-chemical properties in *C. tetragonoloba*, *Z. mays*, *G. max*, *A. thaliana*, *H. sapiens* and *M. musculus*. The AP-4 complex genes and proteins identified in C. tetragonoloba were written as CtAP4E, CtAP4B, CtAP4M and CtAP4S. The CDS length of the large subunit epsilon ranged from 2520 bp to 3105 bp whereas the CDS length for the other large subunit beta varied from 2481bp to 2856 bp. The minimum and maximum CDS lengths for the medium subunit mu were found to be 1251 bp and 1860 bp, respectively. The small subunit sigma had a narrow CDS length range of 423 bp to 438 bp. The CDS lengths of epsilon subunits of all studied plant species were smaller than those of *H. sapiens* and *M. musculus* whereas the CDS lengths of beta subunits were larger than those of above mammalian species. The CDS lengths of mu and sigma subunits of the investigated plant species were almost similar to those of the *H. sapiens* and *M. musculus*. The amino acid sequences were subjected to ExPASy ProtParam tool (http://www.expasy.ch/tools/protparam.html) for the analysis of various physico-chemical parameters. A brief comparative analysis of various parameters like length of protein, molecular weight, theoretical pI, stability, GRAVY index, etc. has been given in Table 4.12. The aliphatic index shows a direct relationship with the mole fraction of Ala, Ile, Leu, and Val in the protein and act as a criteria to measure the stability of a protein.The instability index analysis showed that AP4E and AP4B proteins are unstable in studied plants and mammals. The AP4M and AP4S proteins are predicted to be stable in plants except AP4M in *A. thaliana* while unstable in mammals. The each subunit of AP-4 complex in all studied organism number contains more number of negatively charged amino acids as compared to positively charged amino acids. The AP4E, AP4B, AP4M and AP4S subunits of AP-4 complex in plants have negative value of GRAVY score indicating that these proteins are non- polar in nature while in mammals, except AP4S, remaining subunits have a negative value of GRAVY score.

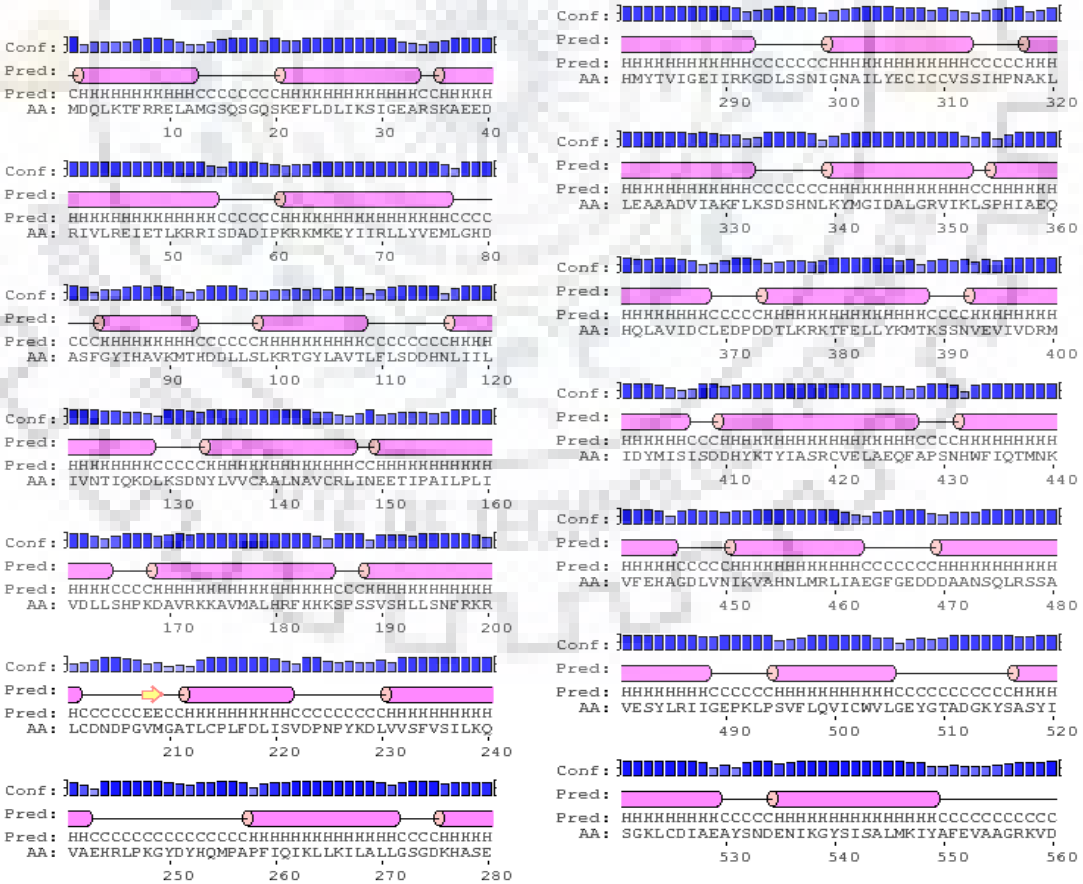
Table 4.12. Comparison of physicochemical properties of AP4E, AP4B, AP4M and AP4S using the protparam tool of EXPASY

Gene	Nucleotide	Amino Acid number	Molecular Weight	Theoretical pI	Negatively Charged Residue	Positively Charged Residues	Extinction coefficient E (M^{-1} cm^{-1})		Instability Index	Aliphatic Index	GRAVY
							Oxidized (Cystine)	Reduced (Cysteine)			
Ct AP4E	2904	967	106604.30	6.08	122	107	73410	72660	42.53	92.81	-0.280
Ct AP4B	2535	844	93236.32	5.18	100	74	81555	80680	41.29	92.81	-0.109
Ct AP4M	1332	443	49965.35	6.64	51	50	38850	38850	33.46	91.44	-0.128
Ct AP4S	432	143	16917.6	5.37	22	16	9190	8940	29.12	97.41	-0.120
Zm AP4E	2910	969	105537.86	5.50	123	100	64470	63720	50.96	88.21	-0.248
Zm AP4B	2526	841	92620.35	5.20	99	73	78700	77700	42.55	92.14	-0.120
Zm AP4M	1356	451	50537.78	6.16	53	50	43320	43320	36.72	87.07	-0.162
Zm AP4S	432	143	16907.61	5.67	21	16	7700	7450	30.04	98.74	-0.078
Gm AP4E	2946	981	107758.68	5.75	127	109	66420	65670	46.89	90.67	-0.274
Gm AP4B	2538	845	93326.31	5.21	99	74	81430	80680	42.34	93.27	-0.124
GmAP4M	1332	443	49704.99	6.69	52	51	37485	37360	35.17	89.93	-0.152
Gm AP4S	432	143	16917.2	5.37	22	16	9190	8940	29.12	97.41	-0.120
At AP4E	2817	938	103734.85	5.33	130	102	64930	64180	48.62	89.85	-0.315
At AP4B	2526	841	93266.27	5.26	100	75	83170	82170	44.68	91.52	-0.134
At AP4M	1356	451	50946.20	7.71	52	53	44810	44810	40.89	87.67	-0.204

At AP4S	432	143	16690.69	5.49	22	16	13200	12950	30	96.78	-0.106
Hs AP4E	3414	1137	127287.41	5.66	138	110	105880	104630	44.70	98.00	-0.114
Hs AP4B	2220	739	83259	5.59	91	75	79770	78270	50.49	102.21	-0.036
Hs AP4M	1362	453	49977.03	6.71	47	45	33140	32890	45.96	95.89	-0.131
Hs AP4S	435	144	17004.76	5.08	22	16	10555	10430	42.56	108.89	0.042
Mm AP4E	3369	1122	124845.35	5.49	140	108	96940	95690	48.59	98.61	-0.097
Mm AP4B	2214	737	82838.56	5.81	89	75	79520	78270	47.40	101.55	-0.067
Mm AP4M	1347	448	49410.32	6.17	46	40	34755	34380	46.65	97.63	-0.115
Mm AP4S	435	144	16817.51	5.21	21	16	10555	10430	39.92	108.89	0.074

4.12.2 Secondary structure and hydropathy plot of AP-4 complex protein of guar

Secondary structures of the CtAP4E, CtAP4B and CtAP4S proteins revealed the presence of more helices than beta strands (Figs. 4.28, 4.29 and 4.31). The CtAP4M contained more number of strands as compared to helices (Fig. 4.30). The hydropathy plot analysis of AP-4 complex protein detected with PDBsum server revealed the average score of hydophobicity score of each subunit. The average score of hydrophobicity was more towards the negative side for epsilon, beta and mu subunits showing that these proteins are hydrophilic in nature (Fig. 4.32 A-C). For sigma subunit the average score of hydrophobicity was more towards positive side indicating that this protein is hydrophobic in nature (Fig. 4.32 D). The motif analysis by Pfam and InterPro softwares showed that the armadillo type fold was present at the N-terminal end in both epsilon and beta subunits while beta adaptin appendage C- terminal domain was present only in the beta subunit. Longin domain was present in both mu and sigma subunits. The mu homology domain (MHD) was found at the C-terminal end of mu subunit.



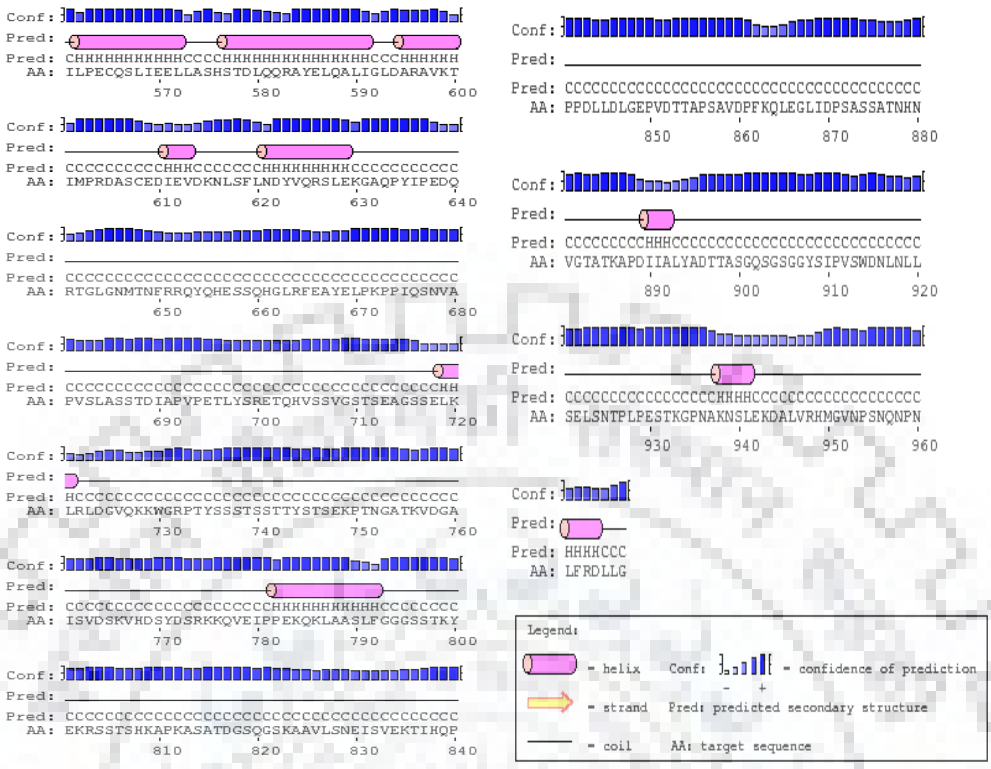


Fig. 4.28. PSIPRED results showed that 38 helices and one strand were found at the various positions in the AP4E protein. The result revealed that random coils dominated among the secondary strcture followed by helices.

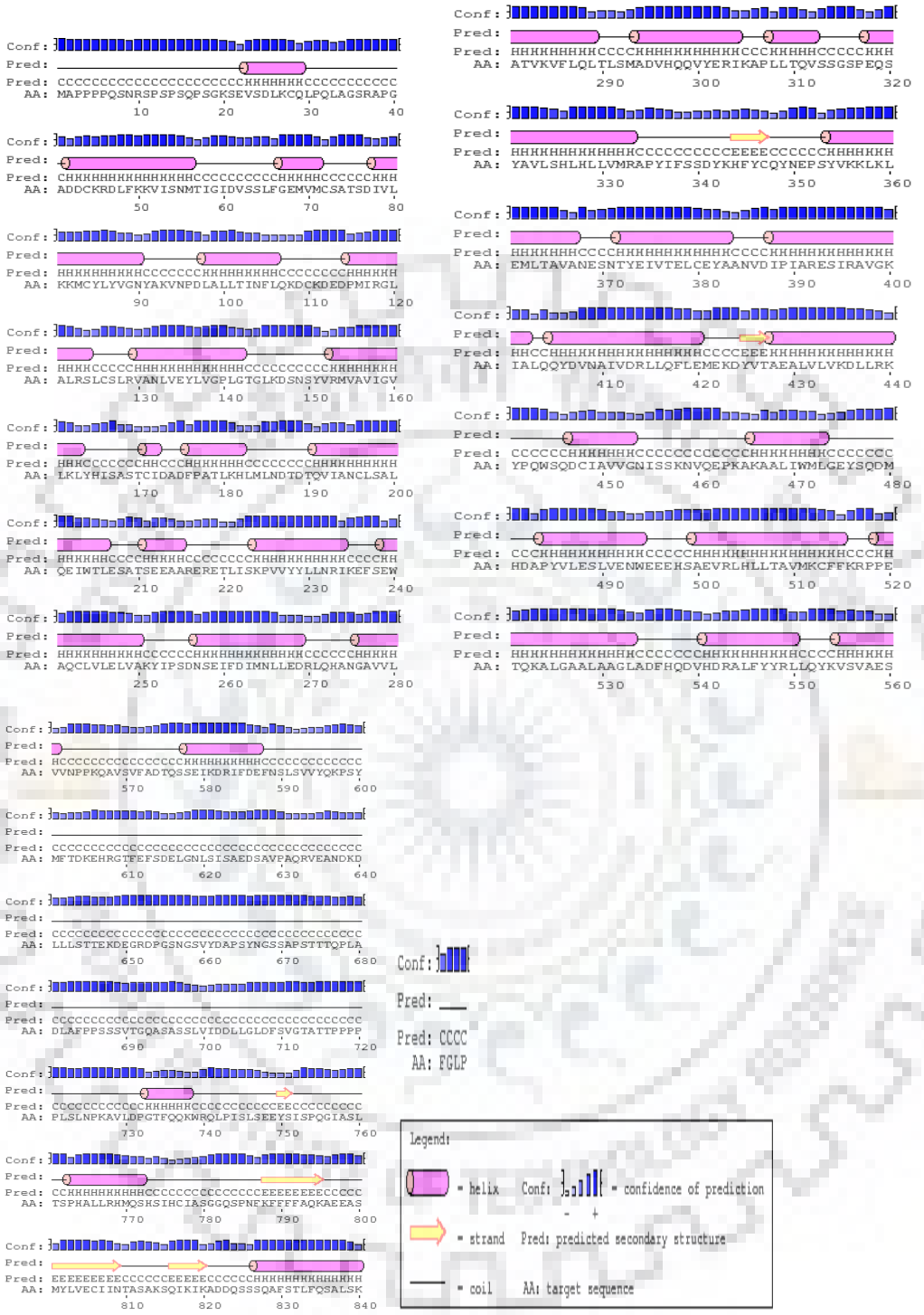


Fig. 4.29. PSIPRED results showed that 34 helices and six strands were found at the various positions in the AP4B protein. The result revealed that random coils dominated among the secondary structure followed by helices.

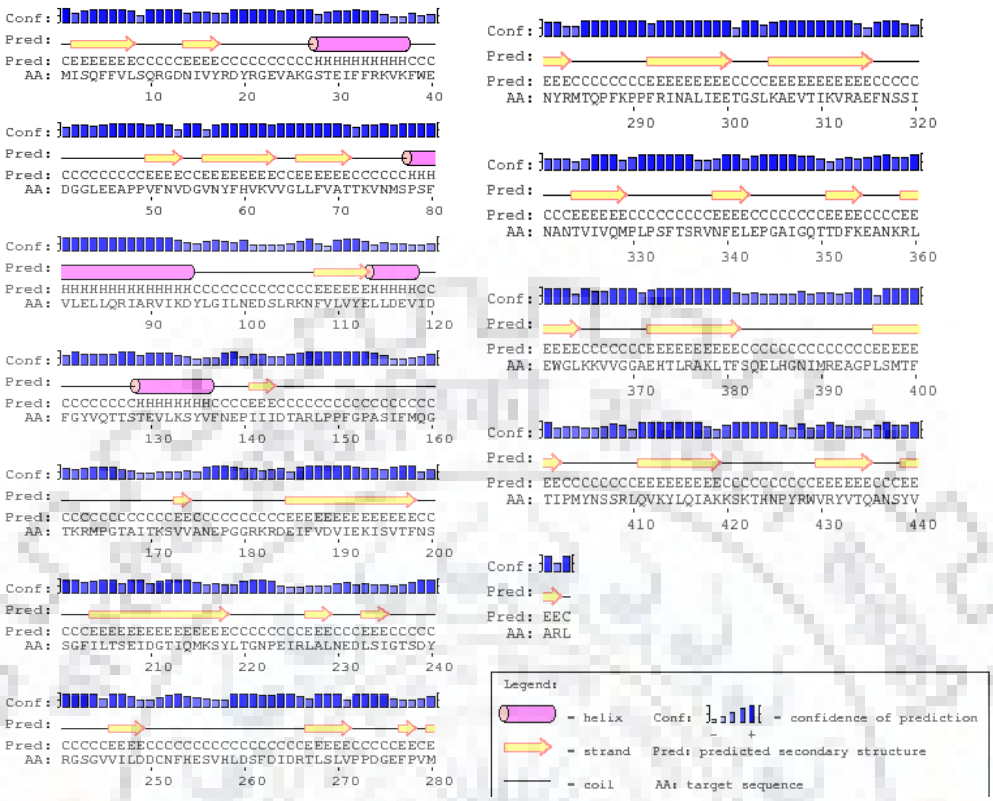


Fig. 4.30. PSIPRED results showed that 4 helices and 29 strands were found at the various positions in the AP4M protein. The result revealed that random coils dominated among the secondary structure followed by strands.

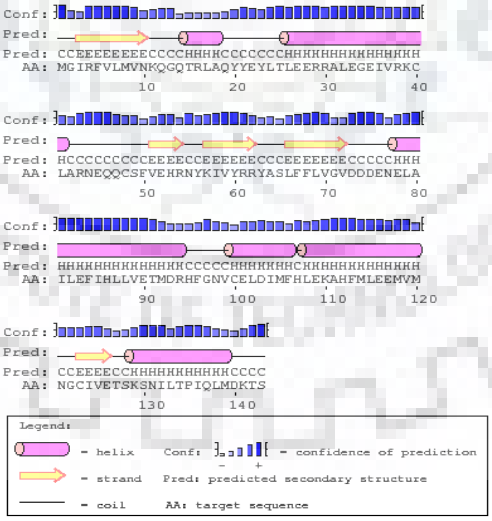


Fig. 4.31. PSIPRED results showed that 6 helices and 5 strand were found at the various positions in the AP4S protein. The result revealed that helices dominated among the secondary structure followed by random coils.

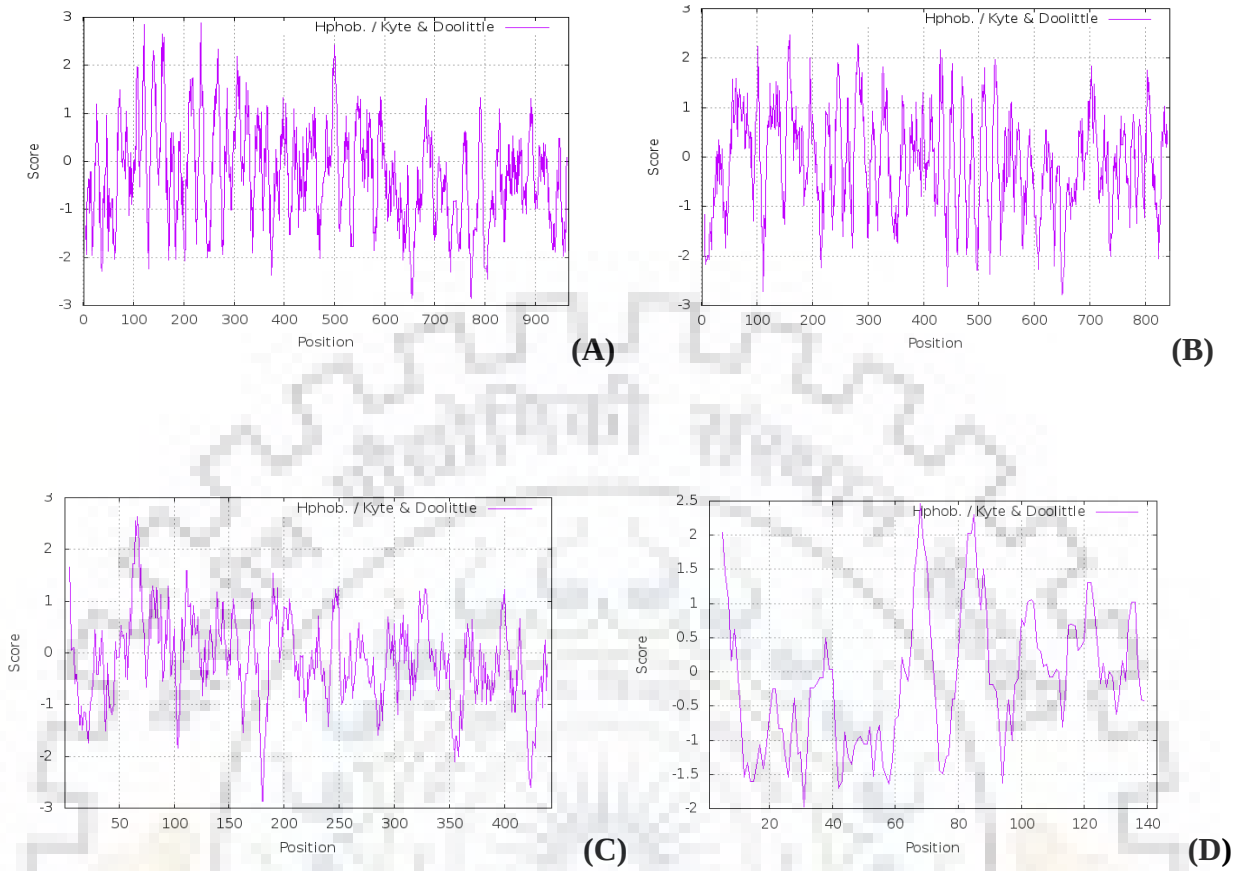


Fig. 4.32. Kyte and Doolittle hydropathy plot of AP-4 complex protein of guar; (A) Epsilon subunit, (B) Beta subunit, (C) Mu subunit and (D) Sigma subunit.

4.12.3. Docked amino acid residues of AP-4 complex subunits with ARFA1A accessory protein

The Figures show the 3D models of ARFA1A accessory protein (Figs. 4.33A-4.36A) and epsilon, beta, mu and sigma subunits of the AP-4 complex (Figs. 4.33B-4.36B), and molecular docking of ARFA1A accessory protein with epsilon (Fig. 4.33C), beta (Fig. 4.34C), mu (Fig. 4.35C), and sigma (Fig. 4.36C) subunits of AP-4 complex of maize. The molecular docking of ARFA1A protein with epsilon subunit indicated that the interacting amino acid residues of epsilon subunit in the binding pocket were Phe 315, Phe 317, Asn 319, Ile 353, Lys 354, Ser 641, Glu 642, Ser 664, Phe 667, Glu 808, Ile 839, Val 855, Ser 857, Ser 858 and Lys 926 (Fig. 4.34C and D, Table 4.13). The Glu 808 was found to interact with Ile 49 of switch I region of ARFA1A (Table 4.13). Out of the interacting amino acids, Phe 315, Phe 317, Asn 319, Ile 353 and Lys 354

are the ones which are known to form hydrophobic interactions. The Phe 317 of epsilon formed ring-ring interaction with Trp 172 of ARFA1A. These interactions predicted the formation of a hydrophobic pocket during the interaction of epsilon subunit of AP-4 complex with ARFA1A accessory protein.

The interacting amino acids in the binding pocket of beta subunit were found to be Ser 348, Tyr 350, Ser 395, Ser 396, Thr 397, Ser 398, Thr 399, Tyr 400, Asn 401, Ala 402, Pro 403, Ser 404, Asp 405, Ile 407, Pro 409, Ser 410, Leu 411 and Leu 456 (Fig. 4.34C and D, Table 4.14). The Tyr 350 of beta subunit formed a hydrophobic bond with Trp 153 of ARFA1A (Fig. 4.34D).

The molecular docking of ARFA1A accessory protein with the mu subunit of AP-4 complex revealed that Glu 113, Glu 117, Asp 120 Phe 121, Gln 125, Thr 126, Glu 130, Val 131, Ser 134, Tyr 135 and Phe 137 were the amino acid residues of mu subunit interacting with ARFA1A protein (Fig. 4.35C and D, Table 4.15).

The Fig. 4.36D shows the docked residues of sigma and ARFA1A proteins. The amino acid residues Gln 12, Gln 14, Arg 16, Gly 97, Asn 98, Cys 100, Glu 101, Leu 102, Asp 103, Met 105, Phe 106, Leu 108 and Glu 109 of the sigma subunit were found to interact with ARFA1A accessory protein (Table 4.16).

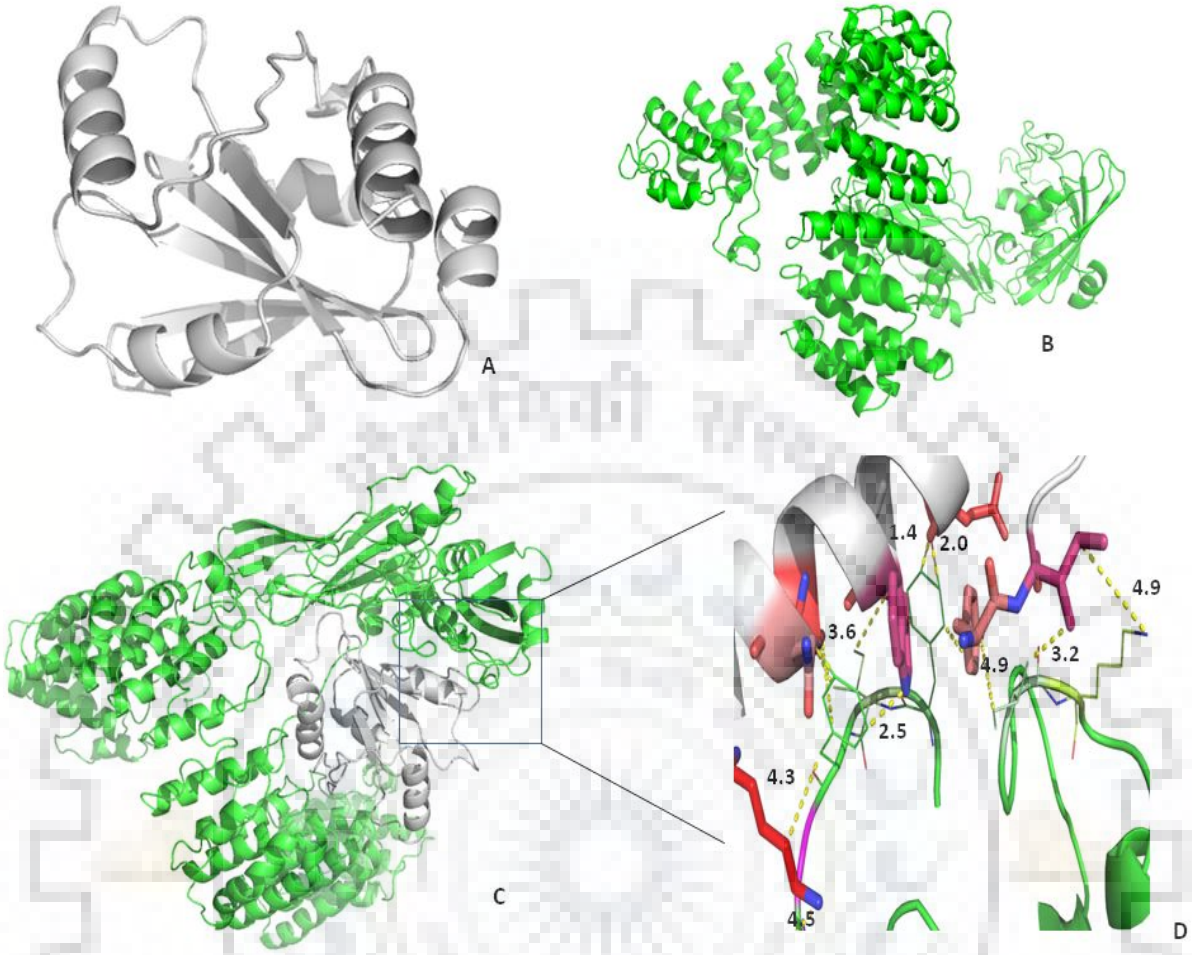


Fig. 4.33. Molecular docking of accessory protein ARFA1A with the epsilon subunit of AP-4 complex of *Z. mays;* (A) Three dimensional structure of ARFA1A, (B) Three dimensional structure of epsilon subunit of AP-4 complex, (C) Structural outline of ARFA1A protein docked with epsilon subunit and (D) Enlarged view of interacting amino acid residues of ARFA1A protein and epsilon subunit. The backbones of both proteins are shown as cartoons and the interacting amino acid residues of ARFA1A protein and epsilon subunit are represented as sticks and lines, respectively.

Table 4.13. List of amino acids of accessory protein ARFA1A interacting with the epsilon subunit of AP-4 complex of *Zea mays.*

Sr.No.	Amino Acid of ARFA1F	Amino Acid(s) of Epsilon subunit
1	Lys 180	Phe 317, Asn 319
2	Asn 176	Phe 317
3	Ser 195	Phe 317, Ile 316
4	Trp 172	Phe 317
5	Asp 171	Ile 316, Phe 315
6	Glu 168	Phe 315
7	Ile 155	Ile 353, Lys 354
8	Tyr 154	Ile 353, Phe 315
9	Gln 83	Lys 926, Ile 839
10	Leu 77	Ser 857
11	Pro 76	Ser 857, Ser 858
12	Ile 49	Gln 808
13	Ile 46	Val 855
14	Thr 44	Phe 667
15	Glu 40	Ser 644
16	Leu 38	Glu 642
17	Lys 37	Ser 641, Glu 642

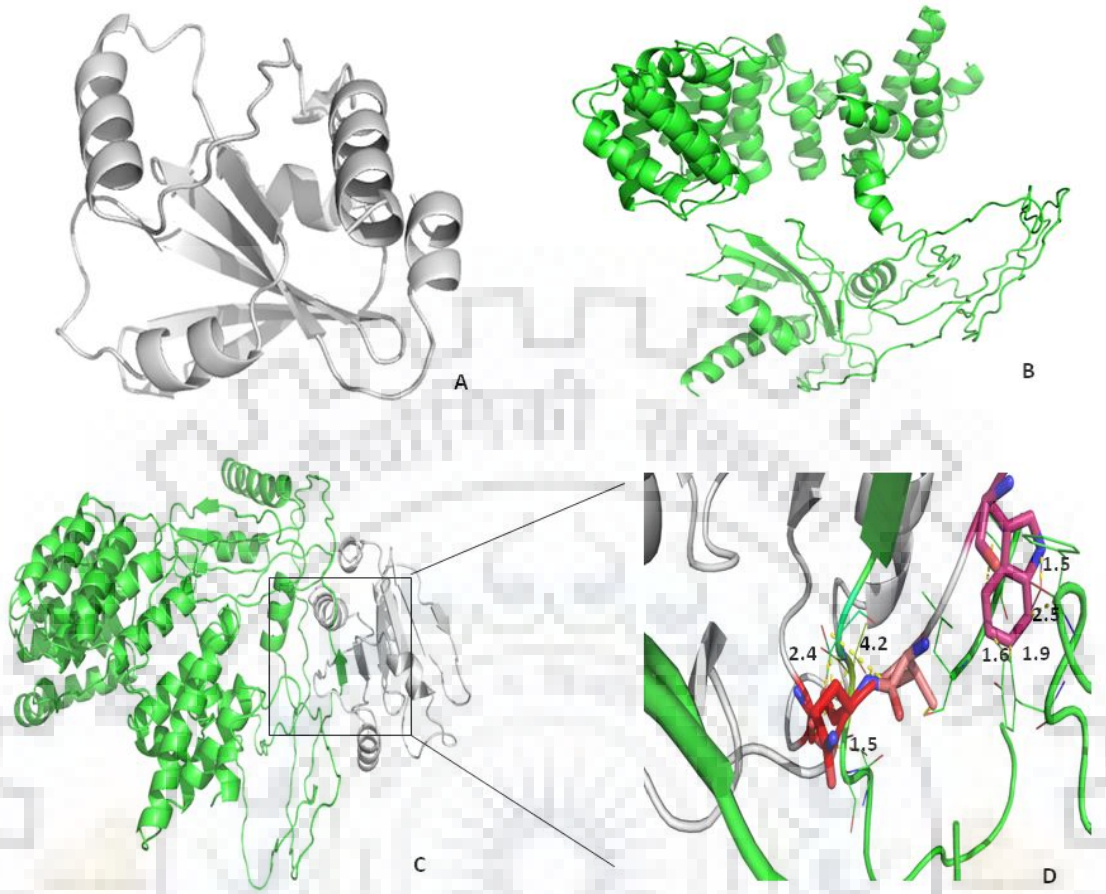


Fig. 4.34. Molecular docking of accessory protein ARFA1A with the beta subunit of AP-4 complex of *Z. mays;* (A) Three dimensional structure of ARFA1A, (B) Three dimensional structure of beta subunit of AP-4 complex, (C) Structural outline of ARFA1A protein docked with beta subunit and (D) Enlarged view of interacting amino acid residues of ARFA1A protein and beta subunit. The backbones of both proteins are shown as cartoons and the interacting amino acid residues of ARFA1A protein and beta subunit are represented as sticks and lines, respectively.

Table 4.14. List of amino acids of accessory protein ARFA1A interacting with the beta subunit of AP-4 complex of *Zea mays.*

Sr. No.	Amino Acid of ARFA1A	Amino Acid(s) of beta subunit
1	Lys 180	Ser404, Ile407
2	Ser 179	Ile407
3	Asn 176	Ser404
4	Leu 172	Tyr400, Ala402
5	Trp 171	Phe348
6	Gln 156	Ser398
7	Ile 155	Leu456, Thr397, Ser 398
8	Trp 153	Try350
9	Ser 94	Ser 395, Ser 396
10	Val 92	Thr 397
11	Phe 90	Thr 399
12	Leu 88	Asn 401
13	Gln 86	Asp 405, Pro 403
14	Ser 10	Leu 411
15	Arg 11	Pro 409, Leu 411, Ser 410

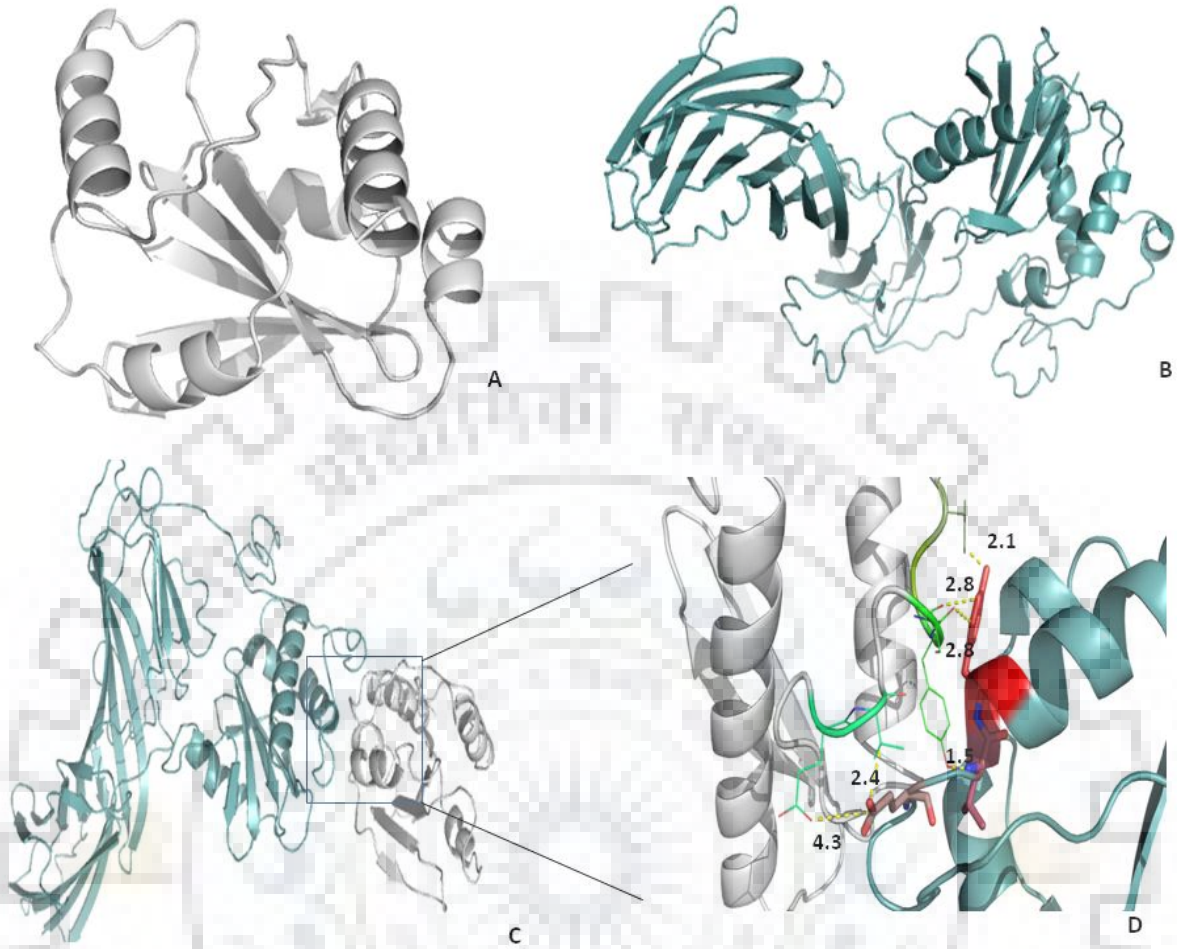


Fig. 4.35. Molecular docking of accessory protein ARFA1A with the mu subunit of AP-4 complex of *Z. mays; (A)* Three dimensional structure of *ARFA1A, (B)* Three dimensional structure of mu subunit of AP-4 complex, (C) Structural outline of ARFA1A protein docked with mu subunit and (D) Enlarged view of interacting amino acid residues of ARFA1A protein mu subunit. The backbones of both proteins are shown as cartoons and the interacting amino acid residues of ARFA1A protein and mu subunit are represented as sticks and lines, respectively.

Table 4.15. List of amino acids of accessory protein ARFA1A interacting with the mu subunit of AP-4 complex of *Zea mays.*

Sr. No.	Amino Acid of ARFA1F	Amino Acid(s) of mu subunit
1	Glu 168	Ser 134, Tyr 135
2	Tyr 167	Phe 137, Ser 134
3	Leu 166	Tyr 135
4	Glu 164	Glu 130, Val 131
5	Gly 163	Glu 117, Gln 125
7	Thr 161	Gln 125
8	Gly 40	Phe 121
9	Lys 38	Asp 120
10	Lys 36	Asp 120
11	Tyr 35	Phe 121

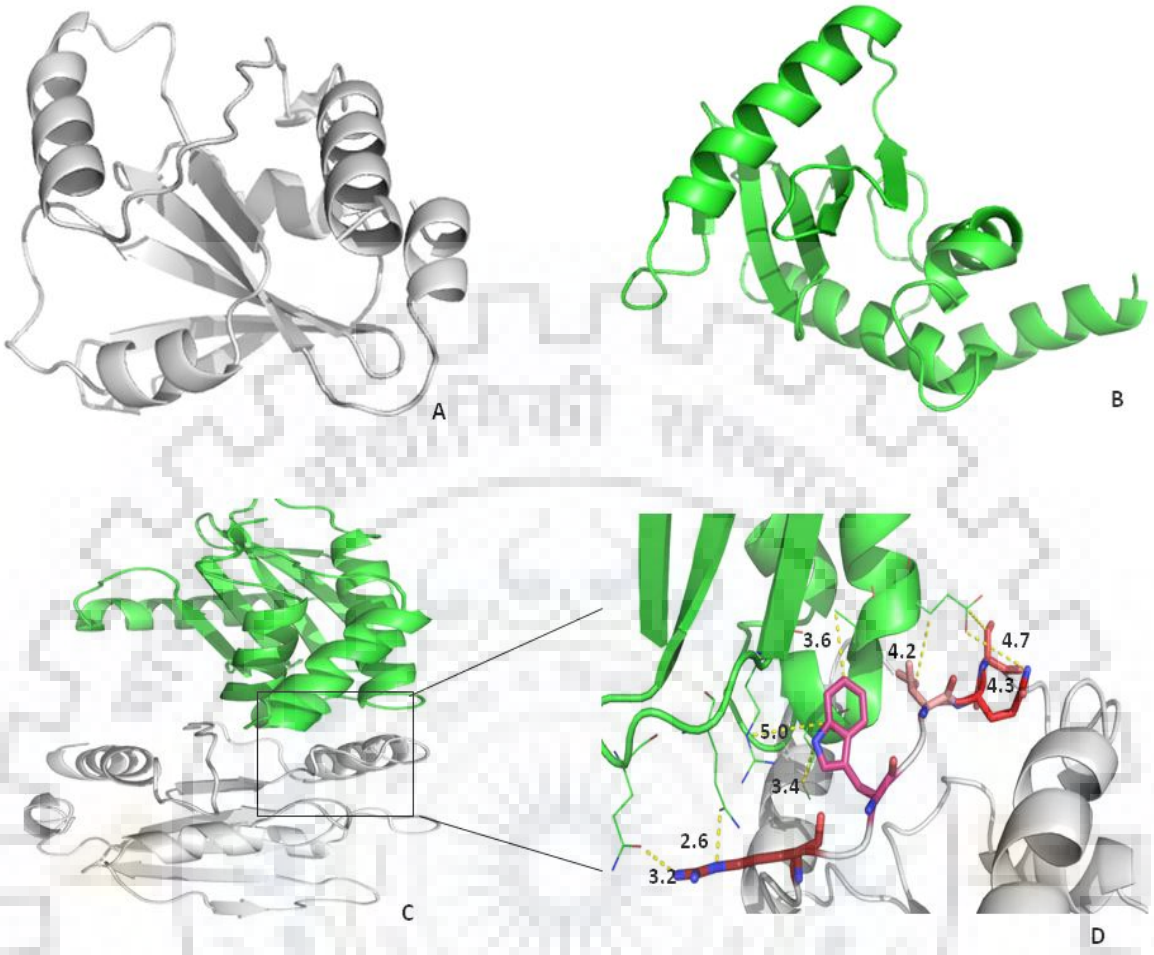


Fig. 4.36. Molecular docking of accessory protein ARFA1A with the sigma subunit of AP-4 complex of *Z. mays;* (A) Three dimensional structure of ARFA1A, (B) Three dimensional structure of sigma subunit of AP-4 complex, (C) Structural outline of ARFA1A protein docked with sigma subunit and (D) Enlarged view of interacting amino acid residues of ARFA1A protein sigma subunit. The backbones of both proteins are shown as cartoons and the interacting amino acid residues of ARFA1A protein and sigma subunit are represented as sticks and lines, respectively.

Table 4.16. List of amino acids of accessory protein ARFA1A interacting with the sigma subunit of AP-4 complex of *Zea mays.*

Sr. No.	Residues of ARFA1F	Residues of sigma subunit
1	Ser 157	Glu 109
2	Gln 156	Glu 109
3	Ile 155	Glu 109
4	Trp 153	Arg 16, Met 105, leu 108
5	Arg 151	Gln 12, Gln 14
6	Hys 146	Glu 101, Leu 102
7	Leu 107	Leu 102, Phe 106
8	Arg 104	Asp 103, Cys100
9	Val 101	Asn 98, Gly 97

5. DISCUSSION

For normal functioning of a living cell, each protein after synthesis must be transported to its specific destination. This intracellular protein transport is accomplished by the specific and complex interactions of several proteins. Any defect in this transport process leads to various physiological abnormalities resulting in several kinds of defects in the organism characteristics. In the last four decades, several details of the intracellular protein transport process have been worked out. These details include the identification of many participating proteins (COPI, COPII, AP-1, AP-2, AP-3, AP-4, AP-5 and their accessory proteins) and characteriztion in various organisms like *Saccharomyces cerevisiae*, *M. musculus*, *H. sapiens* and *A. thaliana*. However, a comprehensive study of the genes coding for AP-4 complex subunits has not been done in plants so far. Here, the detailed analysis of the AP-4 complex genes and their protein products has been presented in guar, maize and soybean like comprehensive study of several genes of *Arabidopsis* [202] and *O. sativa* [295, 128] have been carried out in past. The details regarding the interaction of proteins participating in AP-4 mediated vesicular trafficking have not been worked out in plants. As co-expression analysis of genes is an important tool to work out biological pathways at molecular level, the studies on the co-expression of plant genes coding for AP-4 complex subunits, ARF1s, epsins and VSR1 proteins and their interactions at protein level was carried out which have been reported to participate in vesicular trafficking mechanism in *H. sapiens* and *M. musculus*.

The orthologs of the genes coding for epsilon, beta, mu and sigma subunits of AP-4 complex of guar were found in all the studied plants which included representatives of cereal and legume crops. These orthologous genes, like other eukaryotic genes, contained introns. The number of introns generally varied according to the size of the gene, however, the *B. rapa* epsilon gene, which is the largest gene among the four genes coding for AP-4 complex, contained only 4 introns. With an increase in the number of introns, a decrease in the sizes of exons was observed in the genes coding for epsilon, beta and mu subunits of AP-4 complex. A similar observation has been earlier reported for the genes of *A. thaliana* and *O. sativa* [301]. The CDS lengths of the mu and sigma subunit genes of the studied plant species were almost similar to those of *H. sapiens* and *M. musculus*, however, the CDS lengths of epsilon and beta subunit genes of plants varied considerably from those of the above mammalian species [24].

These genetic changes may have been selected in past for specific adaptations in plants. The phylogenetic analysis of the AP-4 complex genes of studied plants showed that divergence in these genes from their respective ancestral genes had occurred during a time period which roughly coincided with the time period of divergence of monocots and dicots. In the studied plants, functional divergence was observed in the protein sequences of epsilon, beta and sigma subunits, however, no functional divergence was found in the mu subunit protein sequences. This result indicated that amino acid sequences of mu subunits are structurally and functionally more important than the those of the other three subunits.

An adapter protein (AP) complex plays the role of an adapter by recognizing the protein to be transported and the accessory protein for making the membrane of the vesicle. In *H. sapiens* and *M. musculus*, the AP-4 complex recognizes a tyrosine-based signal on the cargo protein with the help of a specific conserved motif of amino acids present on its mu subunit. In this motif, the recognizing role is mainly played by seven amino acids, namely, phenylalanine, histidine, phenylalanine, lysine, glycine, glutamine and tyrosine present at positions, 255, 256, 264, 270, 277, 278 and 284, respectively [38]. The study revealed the presence of a similar conserved motif on the mu subunit of AP-4 complex of each studied plant. All the seven amino acids which play the important recognition role in the above mammalian species were also found to be conserved in all studied plant species. It seems that the protein sorting signal recognizing motif on the mu subunit of AP-4 complex is well-conserved in diverse groups of organisms.

In *H. sapiens* and *M. musculus*, the AP-4 complex recognizes and binds to a hydrophobic stretch of amino acids on the accessory protein tepsin, which makes the outer coat of the vesicle, with the help of a specific conserved motif on its beta subunit. In this motif, three amino acids, namely, tryptophan, tyrosine and glutamine, at positions 635, 682 and 696, respectively, play the main role in binding [79]. The results have shown that a similar conserved motif is present on the beta subunit of AP-4 complex of each studied plant. The above mentioned three amino acids playing the main role of binding in *H. sapiens* and *M. musculus* have been found conserved in the studied plants with the difference that tyrosine is replaced by another aromatic amino acid phenylalanine. This change may have become necessary because of the possibility of presence of a different accessory protein in plants as the searches for tepsin protein sequence from the NCBI and Uniprot databases did not yield any positive result.

The grasses are distinguished from other angiosperms by the characteristic feature of having a high genomic GC content ranging from 43% to 50% [16, 243]. A high GC content in the AP-4 complex genes of monocots in comparison to dicots was observed. High GC content in the genes may be an adaptation to survive under changing environmental conditions [242]. The development and physiology of an organism is controlled by regulation of gene expression with the help of *cis*-regulatory sequences like promoters and enhancers [282]. The cis-regulatory sequences contain distinct DNA sequences called CREs which bind to specific transcription factors and/or other regulatory molecules required for starting and continuation of transcription [193]. The presence of a CRE in the promoter of a gene may indicate the biological function of this gene in the organism. Among the total number of CREs observed by us, 15 were found to be more frequent in the promoters of the AP-4 complex genes of the three studied plants, *A. thaliana*, *Z. mays* and *G. max*. Out of these 15 CREs, six CREs, viz., GATABOX, IBOXCORE, GT1CONSENSUS, MYCOCCONESUSAT, BIHD1OS and GT1GMSCAM4, have been reported to involve in the regulation of abiotic and biotic stress responses [144, 255, 162]. Seven CREs, viz., CAATBOX, CACTFTPPCA1, POLNEN1LELAT52, DOFCOREZM, SEF4MOTIFGM7S, EBOXBNNAPA and ROOTMOTIFTAPOX1 are known to regulate cellular developmental activities like flower and seed development, and seed germination [4, 237, 178,75]. Two CREs, viz., ARR1AT and WRKY71OS have been earlier shown to be involved in the regulation of cytokinin and gibberllin signaling responses [227, 296]. These findings indicate that the AP-4 complex may have a significant role in plants in stress response, cellular development and hormonal regulation.

SNPs are widely occurring single nucleotide changes in the genomes of various organisms. These mutations generally cause deleterious alterations in the expression of the genes. An SNP may result in a major change in the level of expression/protein products of a gene resulting in drastic changes in the phenotype determined by that gene. Total 1316 and 156 SNPs in the genes coding for epsilon, beta, mu and sigma subunits of AP-4 complexes of *Z. mays* and soybean were found, respectively; a little more than half of the SNPs were transitions. The presence of non-synonymous SNPs in the functional binding pockets of predicted 3-D structures of AP-4 complex proteins showed that the functions of these proteins could be altered. The genes in which SNPs were observed in the candidate functional binding sites could be applied as functional markers. The identification of the SNPs that directly affect gene and/or protein

function(s) is quite important for molecular plant breeding. For breeding of crop plants information regarding genetic resources is required [22]. The identification of SNPs increase the information on genetic variation of AP-4 complex genes in crop plants.

Though none of these SNPs altered the amino acids essential for the binding of AP-4 complex with the sorting signals of cargo and the accessory proteins, the potential negative effect of these SNPs, especially transversions, on the normal functioning of the AP-4 complex could not be ruled out. The defective functioning of AP-4 complex may cause reduction in the efficiency of cargo protein transport activity and/or mislocalization of the cargo protein, resulting in drastic loss of stress tolerance ability of plant and/or quality and quantity of plant product. There is a need to identify more number of SNPs and study their impact on the structure, function and regulation of AP-4 complex genes. Hereditary spastic paraplegia, which is a group of neurodegenerative disorders, has been found to be caused by specific mutation in AP-4 complex. Therefore, a significant conclusion from this study is that the germplasm to be used in a plant breeding programme to develop an improved variety must be screened for the SNPs and other mutations in the genes coding for epsilon, beta, mu and sigma subunits of AP-4 complex to ensure that the parental lines to be used in breeding have normal intracellular protein transport. As the plant breeders in past have never paid any attention to the AP-4 complex genes and other genes involved in intracellular protein transport, it is likely that some of our current crop varieties have a defect in some of these genes and as a result of it, these are not able to perform to their maximum potentials.

Subsequently, two reports have been published on the structural and functional characterization of AP-4 complex genes tobacco and *A. thaliana*. In the first report, the mu subunit of AP-4 complex was shown to bind with the tail of vacuolar sorting receptor (VSR) in tobacco [86]. In the second one, the mu subunit of AP-4 complex was found to recognize the vacuolar sorting receptor 1 (VSR1) in *A. thaliana* [82]. The mutations in the genes coding for each of the four subunits led to defective sorting of major seed storage protein 12S globulins in the protein storage vacuoles. The findings reported in the above two research papers suggest a possible role of AP-4 complex in transport of proteins to protein storage vacuoles in seeds. In the above two reports, comparative analysis of AP-4 complex genes have not been made in different tissues of the plant. The RNA-Seq studies revealed that the genes coding for epsilon, beta, mu and sigma subunits of AP-4 complexes expressed in all the studied tissues of *A. thaliana*, *Z.*

mays and G. max; however, the expression of these genes was significantly more in the seed and other reproductive tissues. These results of RNA-Seq analysis were validated by the results of our qRT-PCR studies in guar, *Z. mays* and *G. max*. This result appears to indicate that these proteins are involved in vesicular trafficking of proteins in plants. Hence, similar proteins seem to participate in the intracellular protein transport in animals and plants.

The genes coding for proteins involved in a pathway in an organism are expected to express simultaneously or almost simultaneously. RNA-Seq analysis of *A. thaliana* and *Z. mays* revealed that genes coding for ARF family members, epsin 1, epsin 2 and VSR1 expressed in all the studied tissues and more expression was observed in reproductive tissues, similar to AP-4 complex genes. These results of RNA-Seq analysis were validated by the results of our qRT-PCR studies in *Z. mays*. Similar validations for different genes has already been reported in in maize [288], Rice [118, 85, 28], *Lycoris aurea* [163], tomato [212] and apple [300], *Pleurotus eryngii* [286], white clover [294], *Arabidopsis* [256]. The findings from RNA-Seq and qRT-PCR studies suggested more involvement of AP-4 complex subunits, ARFs, epsins, and VSR1 proteins in flower and seed development.

The results of RNA-Seq co-expression analysis indicate that AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1E, ARFA1F, epsin2 and VSR1 proteins are necessary in the development/ functioning of anther tissue of *A. thaliana* whereas the proteins AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1E, ARFA1F, ARFB1B, epsin 1 and epsin 2 appear to be important in the seed development of *A. thaliana*. Almost similar conclusions were obtained for *Z. mays* in which AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARF3, epsin 1 and VSR1 proteins appear to be necessary for the anther development/ functioning, and AP4E, AP4B, AP4M, AP4S, ARFA1A, ARFA1B, ARFA1C, ARFA1E, ARFA1F, ARFB1B, ARF3, epsin 1, epsin 2 and VSR1 proteins seem to be required for seed development. The AP-4 and ARFs proteins do not appear to make significance contribution in the development of stem and root tissues of *A. thaliana* and *Z. mays*. The ARF protein has been reported to play a role in seed size development in *A. thaliana* [273] and embryo development in rice [299]. Our findings give support to these conclusions and also provide a possible mechanism for its role.

The results of protein- protein interactions network analysis showed that AP4E, AP4B, AP4M and AP4S proteins had interaction with most of the accessory proteins and formed the

central hub. This is expected as these subunits together form AP-4 complex which plays the dual roles of recognizing the cargo protein and recruiting the accessory proteins for vesicular coat formation. The PPI results revealed that the VSR1 protein interacts with epsin 1 which further interacts with AP4E, AP4B and AP4M subunits of AP-4 complex. This result is similar to that of Song et al. [245] who reported that in *A. thaliana* epsin 1 plays a role in vacuolar trafficking of proteins by interacting with VSR1. These findings seem to suggest that the mechanism of action of VSR1 in vesicular trafficking is similar in *A. thaliana* and *Z. mays*.

The docking results of epsilon subunit of AP-4 complex with ARFA1A accessory protein revealed that the epsilon subunit interacts with the ARFA1A. The Glu 808 residue of epsilon was found to interacts with Ile 49 of switch I region of ARFA1A (Table 4.13). Bohem et al. [26] reported that the substitution of Ile at 49th position in ARF1 with Thr disrupts the interaction of ARF1 with epsilon and in two previous reports the same substitution has been shown to abolish the interaction of ARF1 with its effector molecules in human cell lines [136, 213]. The above information suggests that Ile 49 residue of ARFA1A is critical for its interaction with epsilon subunit of AP-4 complex in *Z. mays*. The mu subunit of AP-4 complex was also observed to interact with the ARFA1A protein, however the interacting amino acid residues of ARFA1A were different from those which interacted with epsilon subunit of AP-4 complex. Similar mode of dual interaction of ARFA1A protein with epsilon and mu subunits of AP-4 complex has been reported in mammals [26]. The interaction mechanism of ARFA1A protein with AP-4 complex appears similar in animals and plants. A hydrophobic pocket found in the interaction of epsilon subunit with ARFA1A protein in *Z. mays* may provide a possibility of binding of a ligand for regulating the activity of AP-4 complex.

These conclusions are likely to provide a boost to the research work aiming to understand the molecular mechanisms of intracellular protein transport in plants. These research findings make useful additions to the understanding of molecular mechanism of AP-4 complex - mediated vesicular trafficking in plants. This knowledge is expected to contribute significantly to the molecular studies of development and functioning of different plant tissues and organs, and is likely to be helpful in the plant breeding programs aiming to produce improved varieties having desirable characteristics.

6. REFERENCES

1. Aguilar RC, Boehm M, Gorshkova I, Crouch RJ, Tomita K, Saito T, Ohnu H, Bonifacino JS. Signal binding specificity of the mu4 subunit of the adaptor protein complex AP-4. J. Biol. Chem. 276: 13145-13152, 2001.
2. Ahle S, Ungewickell E. Purification and properties of a new clathrin assembly protein. EMBO J. 5: 3134-3149, 1986.
3. Ahmad S, Singh M, Lamb-Palmer ND, Lefsrud M, Singh, J. Assessment of genetic diversity in 35 Pisum sativum accessions using microsatellite markers. Can. J. Plant Sci. 92: 1075-1081, 2012.
4. Allen RD, Bernier F, Lessard PA, Beachy RN. Nuclear factors interact with a soybean beta-conglycinin enhancer. Plant Cell. 1: 623-631, 1989.
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular biology. 215(3): 403-10, 1990.
6. Andag U, Schmitt HD. Dsl1p, an essential component of the Golgi-endoplasmic reticulum retrieval system in yeast, uses the same sequence motif to interact with different subunits of the COPI vesicle coat. J. Biol. Chem. 278: 51722-51734, 2003.
7. Aoki K, Ogata Y, Shibata D. Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol. 48(3): 381-90, 2007.
8. Appiah-Kubi D, Asibuo JY, Quain MD, Oppong A, Akromah R. Diversity studies on soybean accessions from three countries. Bio catalysis and Agricultural Biotechnology. 3: 198-206, 2014.
9. Arora RN, Pahuja SK. Mutagenesis in guar [Cyamopsis tetragonoloba (L.) taub.]. Plant Mutation Reports. 2(1): 7-9, 2008.
10. Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9(3): 208-218, 1991.
11. Asare AT, Gowda BS, Galyuon IKA, Aboagye LL, Takrama JF, Timko MP. Assessment of the genetic diversity in cowpea [Vigna unguiculata (L.) Walp.] germplasm from Ghana using simple sequence repeat markers. Plant Genet. Resour. 8: 142-150, 2010.

12. Atias O, Chor B, Chamovitz D. Large-scale analysis of Arabidopsis transcription reveals a basal coregulation network. BMC Syst. Biol. 3: 86, 2009.
13. Ayyanagar GNP, Krishnaswami N. A note on the chromosome number in cluster bean (Cyamopsis psoralioides DC). Indian J. Agri. Sci. 3: 934-935, 1933.
14. Bar-Hen A, Charcosset A, Bourgoin M, Cuiard J. Relationships between genetic markers and morphological traits in a maize inbred lines collection. Euphytica 84: 145-154, 1995.
15. Barlowe C, Orci L, Yeung T, Hosobuchi M, Hamamoto S, Salama N, Rexach, MF, Ravazzola M, Amherdt M, Schekman R. COPII: A membrane coat formed by sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell. 77: 895-907, 1994.
16. Barow M, Meister A. Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. Cytometry. 47: 1-7, 2002.
17. Bauer P, Leshinsky-Silver E, Blumkin L, Schlipf N, Schröder C, Schicks J, Lev D, Riess O, Lerman-Sagie T, Schöls L. Mutation in the AP4B1 gene cause hereditary spastic paraplegia type 47 (SPG47). Neurogenetics. 13: 73-6, 2012.
18. Beadle GW. Teosinte and the origin of maize. J. Hered. 30: 245-47, 1939.
19. Bergmann S, Ihmels J, Barkai N. Similarities and differences in genome-wide expression data of six organisms. PLoS Biology. 2: E9, 2004.
20. Bernardo R. Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids. Crop Sci. 34(1): 20-25, 1994.
21. Bethune J, Wieland F, Moelleken J. COPI-mediated transport. J. Membr. Biol. 211: 65-79, 2006.
22. Bhullar NK, Michael Mackay M, Beat Keller B. Genetic Diversity of the Pm3 Powdery Mildew Resistance Alleles in Wheat Gene Bank Accessions as Assessed by Molecular Markers. Diversity. 2: 768-786, 2010.
23. Blumstein J, Faundez V, Nakatsu F, Saito T, Ohno H, Kelly RB. The neuronal form of adaptor protein-3 is required for synaptic vesicle formation from endosomes. J. Neurosci. 21: 8034-8042, 2001.

24. Boehm M, Bonifacino JS. Adaptins: the final recount. Mol. Biol. Cell. 12: 2907-20, 2001.
25. Boehm M, Bonifacino JS. Genetic analyses of adaptin function from yeast to mammals. Gene. 286: 175-186, 2002.
26. Boehm M, Aguilar RC, Bonifacino JS. Functional and physical interactions of the adaptor protein complex AP-4 with ADP-ribosylation factors (ARFs). EMBO J. 20: 6265-6276, 2001.
27. Bonetta L. Protein–protein interactions: interactome under construction. Nature. 468: 851-854, 2010.
28. Borah P, Sharma E, Kaur A, Chandel G, Mohapatra T, Kapoor S, Khurana JP. Analysis of drought-responsive signalling network in two contrasting rice cultivars using transcriptome-based approach. Sci. Rep. 7: 42131, 2017.
29. Borner GHH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, Sahlender DA, Robinson MS. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. J. Cell Biol. 197(1): 141-160, 2012.
30. Boyer CD, Hannah LC. Kernel mutants of corns A.R. Hallaner. In Speciality Corns, CRC Press. Second Edition: 13-44, 1994.
31. Breyan H, Ross, Yimo L, Esteban A, Corales, Patricia V, Burgos, Gonzalo A, Mardones. Structural and Functional Characterization of Cargo-Binding Sites on the m4-Subunit of Adaptor Protein Complex 4. PLoS One. 9: e88147, 2014.
32. Broich SL, Palmer RG. A cluster analysis of wild and domesticated soybean phenotypes. Euphytica. 29: 23-32, 1980.
33. Brown HA, Gutowski S, Moomaw CR, Slaughter C, Sternweis PC. ADP-ribosylation factor, a small GTP-dependent regulatory protein, stimulates phospholipase D activity. Cell. 75: 1137-1144, 1993.
34. Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci. 40: 815-823, 2000.
35. Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A. Evolution of DNA Sequence Nonhomologies among Maize Inbreds. The Plant Cell. 17: 343-360, 2005.

36. Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z,Wang B, Xu D,Yang B, Xie C, Fan L, Gao S, Xu X, Zhang G, Li Y, Jiao Y, John FD, Jeffrey RI, Anne L, Vince B, M Cinta R, Edward SB, Doreen W, Jinsheng L, Qi S, Yunbi X. Construction of the third generation Zea mays haplotype map. BioRxiv. 026963, 2015.
37. Burgess J, Jauregui M, Tan J, Rollins J, Lallet S, Leventis PA, Boulianne GL, Chang HC, Le Borgne R, Kramer H, Brill JA. AP-1 and clothing are essential for the secretory granule biogenesis in *Drosophila*. Mol. Biol. Cell. 22: 2094-2105, 2011.
38. Burgos PV, Mardones GA, Rojas AL, daSilva LL, Prabhu Y, Hurley JH, Bonifacino JS. Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. Dev. Cell. 18: 425-36, 2010.
39. Burle ML, Fonseca JR, Kami JA, Gepts P. Microsatellite diversity and genetic structure among common bean (Phaseolus vulgaris L.) landraces in Brazil, a secondary center of diversity. Theor. Appl. Genet. 121: 801-813, 2010.
40. Cai H, Yu S, Menon S, Cai Y, Lazarova D, Fu C, Reinisch K, Hay JC. Ferro-Novick, S. TRAPPI tethers COPII vesicles by binding the coat subunit Sec23. Nature. 445: 941-944, 2007.
41. Carvalho VP, Ruas CF, Ferreira JM, Moreria RMP, Ruas PM. Genetic diversity among maize (Zea mays L.) landraces assessed by RAPD markers. Genet. Mol. Biol. 27(2): 228-236, 2004.
42. Cavenagh MM, Whitney JA, Carroll K, Zhang CJ, Boman AL, Rosenwald AG, Mellman I, Kahn RA. Intracellular distribution of Arf proteins in mammalian cells. Arf6 is uniquely localized to the plasma membrane, J. Biol. Chem. 271: 21767-21774, 1996.
43. Chaudhary BS, Lodhi GP. Studies on the inheritance of five qualitative characteristics in clusterbean [cyamopsis tetragonoloba (L.) taub]. Euphytica 30(1): 161-165, 1981.
44. Chaudhary BS, Paroda RS, Solanki KR. New crossing technique in clusterbean [Cyamopsis tetragonoloba (L.) taub.]. Current science 1974.
45. Chauhan DK, Bhat JA, Thakur AK, Kumari S, Hussain Z, Satyawathi CT. Molecular characterization and genetic diversity assessment in soybean [Glycine max (L.) Merr.] varieties using SSR markers. Ind. J. Biotechnol. 14: 504-510, 2015.

46. Chen H, Fre S, Slepnev VI, Capua MR, Takei K, Butler MH, Di Fiore PP, De Camilli P. Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis. Nature. 394: 793-797, 1998.
47. Cheng J, Khan MA, Qiu WM, Li J, Zhou H, Zhang Q, Guo W, Zhu T, Peng J, Sun F, Li S. Diversification of genes encoding granule bound starch synthase in monocots and dicots is marked by multiple genome-wide duplication events. PLOS One. 7: e30088, 2012.
48. Chia JM, Song C, Bradbury PJ, Costich D, Leon ND, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE. Holland J, Hufford BM, Jinsheng L, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D. Maize hapmap2 identifies extant variation from a genome in flux. Nat. Genet. 44: 803-807, 2012.
49. Clark J, Moore L, Krasinskas A, Way J, Battey J, Tamkun J, Kahn RA. Selective amplification of novel members of the ADP-ribosylation factor (ARF) family: cloning of new human and Drosophila ARL genes, Proc. Natl. Acad. Sci. U. S. A. 90: 8952-8956, 1993.
50. Cockcroft S, Thomas GM, Fensome A, Geny B, Cunningham E, Gout I, Hiles I, Totty NF, Truong O, Hsuan JJ. Phospholipase D: a downstream effector of ARF in granulocytes. Science. 263: 523-526, 1994.
51. Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol. 158: 824-834, 2012.
52. Cosson P, Letourneurtr F. Coatomsr (COPI)-coated vesicles: role in intracellular transport and protein sorting. Curr. Opin. Cell Biol. 9(4): 484-487, 1997.
53. Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. Genome Biology. 9: R130, 2008.

54. Cui J, Li P, Li G, Xu F, Zhao C, Li YH, Yang ZN, Wang G, Yu QB, Li YX, Shi T L. AtPID: Arabidopsis thaliana protein interactome database-an integrative platform for plant systems biology. Nucleic Acids Res. 36: D999-D1008, 2008.
55. Cui Z, Carter Jr, TE, Burton JW. Genetic base of 651 Chinese soybean cultivars released during 1923 to 1995. Crop Sci. 40: 1470-1481, 2000.
56. De Camilli P, Chen H, Hyman J, Panepucci E, Bateman A, Brunger AT. The ENTH domain. FEBS Lett. 513: 11-18, 2002.
57. De Marcos Lousa C, Gerschlick DC, Denecke J. Mechanisms and concepts paving the way towards a complete transport cycle of plant vacuolar sorting receptors. Plant Cell. 24: 1714-1732, 2012.
58. Delannay X, Rogers DM, Palmer RG. Relative genetic contributions among ancestral lines to North America soybean cultivars. Crop Sci. 23: 944-949, 1983.
59. DeLano WL. The PyMOL Molecular Graphics System. Delano Scientific. 2002; (http://www.pymol.org).
60. Dell'Angelica EC, Mullins C, Bonifacino JS. AP-4, a novel protein complex related to clathrin adaptors. J. Biol. Che. 274: 7278-7285, 1999.
61. Dhugga KS, Barreiro R, Whitten B, Stecca K, Hazebroek J, Randhawa GS, Dolan M, Kinney AJ, Tomes D, Nichols S. Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. Science. 303(5656): 363-366, 2004.
62. Ding XD, Richter T Chen, M, Fujii, H, Seo, YS, Xie, MT, Zheng, XW, Kanrar S, Stevenson RA, Dardick C, Li Y, Jiang H, Zhang Y, Yu F, Bartley EL, Chern M, Bart R, Chen X, Zhu L, Farmerie WG, Gribskov M, Zhu JK, Fromm ME, Ronald PC, Song WY. A Rice Kinase-Protein Interaction Map. Plant Physiol. 149: 1478-1492, 2009.
63. Doldi ML, Vollmann J, Lelly T. Genetic diversity in soybean as determined by RAPD and micro satellite analyses. Pl. Br. 116 (4): 331-335, 1997.
64. Donaldson JG, Jackson, CL. Regulators and effectors of the ARF GTPases. Curr. Opin. Cell Biol. 12: 475-482, 2000.
65. Doray B, Lee I, Knisely J, Bu G, Kornfeld S. The gamma/sigma1 and alpha/sigma2 hemicomplexes of clathrin adaptors AP-1 and AP-2 harbor the dileucine recognition site. Mol. Biol. Cell. 18: 1887-1896, 2007.

66. Dortay H, Gruhn N, Pfeifer A, Schwerdtner M, Schmulling T, Heyl A. Toward an interaction map of the two-component signaling pathway of Arabidopsis thaliana. J. Prot. Res. 7: 3649-3660, 2008.
67. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. Focus. 12: 13-15, 1990.
68. Drummond AJ, Rambaut A. BEAST: bayesian evolutionary analysis Sampling trees. BMC Evol. Biol. 7: 214, 2007.
69. Dudhate A, Shinde H, Tsugama D, Liu S, Takano T. Transcriptomic analysis reveals the differentially expressed genes and pathways involved in drought tolerance in pearl millet [Pennisetum glaucum (L.) R. Br]. Plos one. 3(4): e0195908, 2018.
70. Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK, Datta S. Singh MN, Bashasab F, Kulwal P, Wanjari KB, Varshney RK, Cook DR, Singh NK. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus Cajan (L.) Millspaugh]. BMC Plant Biol. 11: 17, 2011.
71. Elhaik E, Pellegrini M, Tatarinova TV. Gene expression and nucleotide composition are associated with genic methylation level in *Oryza sativa*. BMC Bioinformatics. 15: 23, 2014.
72. Faundez VV, Kelly RB. The AP-3 complex required for endosomal synaptic vesicle biogenesis is associated with a casein kinase Ialpha-like isoform. Mol. Biol. Cell. 11: 2591-2604, 2000.
73. Ficklin SP, Feltus FA. Gene coexpression network alignment and conservation of gene modules between two grass species: maize and rice. Plant Physiol. 156(3): 1244-56, 2011.
74. Fields IC, King SM, Shteyn E, Kang RS, Folsch H. Phosphatidylinositol 3,4,5-trisphosphate localization in recycling endosomes is necessary for AP-1B-dependent sorting in polarized epithelial cells. Mol. Biol. Cell. 21: 95-105, 2010.
75. Filichkin SA, Leonard JM, Monteros A, Liu PP, Nonogaki H. A novel endo-beta-mannanase gene in tomato LeMAN5 is associated with anther and pollen development. Plant Physiol. 134: 1080-1087, 2004.

76. Flint-Garcia SA, Bodnar AL, Scott MP. Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. Theor. Appl. Genet. 119: 1129-1142, 2009.
77. Folsch H, Ohno H, Bonifacino JS, Mellman I. A novel clathrin adaptor complex mediates basolateral targeting in polarized epithelial cells. Cell. 99: 189-198, 1999.
78. Fotin A, Cheng Y, Sliz P, Grigorieff N, Harrison SC, Kirchhausen T, Walz T. Molecular model for a complete clathrin lattice from electron cryomicroscopy. Nature. 432: 573-579, 2004.
79. Frazier MN, Davies AK, Voehler M, Kendall AK, Borner GHH, Chazin WJ, Robinson MS, Jackson LP. Molecular basis for the interaction between Adaptor Protein Complex 4 (AP4) β4 and its accessory protein, tepsin. Traffic. 17: 400-415, 2016.
80. Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z,Wang B. Zhai L, Dai C, Xu J, Wang W, Li X, Zheng J, Chen L, Luo L, Liu J, Qian X, Yan J, Wang J & Wang G. RNA sequencing reveals the complex regulatory network in the maize kernel. Nat. Commun. 4: 2832, 2013.
81. Fuji K, Shimada T, Takahashi H, Tamura K, Koumoto Y, Utsumi S, Nishizawa K, Maruyama N, Hara-Nishimura I. Arabidopsis vacuolar sorting mutants (green fluorescent seed) can be identified efficiently by secretion of vacuole-targeted green fluorescent protein in their seeds. Plant Cell. 19: 597-609, 2007.
82. Fuji K, Shirakawa M, Shimono Y, Kunieda T, Fukao Y, Koumoto Y, Takahashi H, Hara-Nishimura I, Shimada T. The adaptor complex AP-4 regulates vacuolar protein sorting at the trans-Golgi network by interacting with VACUOLAR SORTING RECEPTOR1. Plant Physiol. 170: 211-9, 2016.
83. Gaidarov I, Keen JH. Phosphoinositide-AP-2 interactions required for targeting to plasma membrane clathrin-coated pits. J. Cell Biol. 146: 755-764, 1999.
84. Galinat WC. The origin of maize. Annu. Rev. Genet. 5: 447-478, 1971.
85. Gaur VS, Singh US, Kumar A. Transcriptional profiling and in silico analysis of Dof transcription factor gene family for understanding their regulation during seed development of rice Oryza sativa L. Mol. Biol. Rep. 38: 2827-2848, 2011.

86. Gershlick DC, de Marcos Lousa C, Foresti O, Lee AJ, Pereira EA, daSilva LLP, Bottanelli F, Denecke J. Golgi-dependent transport of vacuolar sorting receptors is regulated by COPII, AP1, and AP4 protein complexes in Tobacco. The Plant Cell. 26: 1308-29, 2014.
87. Gill SC, Von Hippel, PH. Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry. 182(2): 319-326, 1989.
88. Gill, S. L. Evaluation of reciprocal hybrid crosses in guar: Texas Tech University: 2009.
89. Gillet JB. Indigofera (Microcharis) in tropical Africa with related genera Cyamopsis and Rhynchotropis. Kew Bulletin, Additional Series. 1: 1-16, 1958.
90. Ginwal HS, Gera M. Genetic variation in seed germination and growth performance of 12 Acacia nilotica provenances in India. J. Trop. Forest Sci. 12(2): 286-297, 2000.
91. Ginwal HS, Maurya SS, Chauhan P. Genetic diversity and relationship between cultivated clones of Dalbergia sissoo of wide geographical origin using RAPD markers. J. Forestry Res. 22(4): 507-517, 2011.
92. Ginwal HS, Mittal N, Barthwal S. Development and characterization of polymorphic chloroplast microsatellite markers in sweet flag (Acorus calamus L.). I. J. Gene Plant Breed. 69(3): 256-259, 2009.
93. Goldberg, J. Structural basis for activation of ARF-GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. Cell. 95: 237-248, 1998.
94. Grabner CP, Price SD, Lysakowski A, Cahill AL, Fox AP. Regulation of large dense-core vesicle volume and neurotransmitter content mediated by adaptor protein 3. Proc.Natl. Acad. Sci. USA. 103: 10035-10040, 2006.
95. Gravotta D, Carvajal-Gonzalez JM, Mattera R, Deborde S, Banfelder JR, Bonifacino JS, Rodriguez-Boulan E. The clathrin adaptor AP-1A mediates basolateral polarity. Dev. Cell 22: 811-823, 2012.
96. Gresta F, Avola G, Cannavò S, Santonoceto C. Morphological, biological, productive and qualitative characterization of 68 guar [Cyamopsis tetragonoloba (L.) Taub.] genotypes. Ind. Crops Prod. 114: 98-107, 2018.
97. Gresta F, Sortino O, Santonoceto C, Issi L, Formantici C, Galante, YM. Effects of sowing times on seed yield, protein and galactomannans content of four varieties of

guar (cyamopsis tetragonoloba l.) in a mediterranean environment. Ind. Crops Prod. 41: 46-52, 2013.

98. Gu X, Velden KV. DIVERGE. Phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics. 18: 500-501, 2002.
99. Gu X. Statistical methods for testing functional divergence after gene duplication. Mol. Biol. Evol. 16: 1664-1674, 1999.
100. Gupta R, Gray JC. Characterization of a single-copy Arabidopsis gene encoding a protein showing limited similarity to the N-terminus of the mammalian clathrin-assembly protein AP180. DNA Res. 6: 51-5, 1999.
101. Gupta SK, Bansal R, Vaidya UJ, Gopalakrishna T. Assessment of genetic diversity at molecular level in mungbean (Vigna radiata (L.) Wilczek). J. Food Legumes. 26(3 & 4): 19-24, 2013.
102. Guruprasad, K., Reddy, B. V. B. and Pandit, M. W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Engineering 4(2): 155-161, 1990.
103. Hake S, Walbot V. The genome of Zea mays, its organization and homology to related grasses. Chromosoma. 79(3): 251-270, 1980.
104. Hirst J, Irving C, Borner GH. Adaptor protein complexes AP-4 and AP-5: new players in endosomal trafficking and progressive spastic paraplegia. Traffic. 14(2): 153-64, 2013a.
105. Hirst J, Borner GH, Edgar J, Hein MY, Mann M, Buchholz F, Antrobus R, Robinson MS. Interaction between AP-5 and the hereditary spastic paraplegia proteins SPG11 and SPG15. Mol. Biol. Cell. 24: 2558-2569, 2013b.
106. Hirst J, Barlow LD, Francisco GC, Sahlender DA, Seaman MN, Dacks JB, Robinson MS. The fifth adaptor protein complex. PLOS Biol. 9: e1001170, 2011.
107. Hirst J, Bright NA, Rous B, Robinson MS. Characterization of a fourth adaptor-related protein complex. Mol. Biol. Cell. 10: 2787-2802, 1999.
108. Hoffman GR, Rahl PB, Collins RN, Cerione RA. Conserved structural motifs in intracellular trafficking pathways: structure of the gamma COP appendage domain. *Mol. Cell.* 12(3): 615-25, 2003.

109. Holloway ZG, Velayos-Baeza A, Howell GJ Levecque C, Ponnambalam S, Sztul E, Monaco AP. Trafficking of the Menkes copper transporter ATP7A is regulated by clathrin, AP-2, AP-1 and Rab22 dependent steps. Mol. Biol. Cell. 24: 1735-1748, 2013.

110. Honda A, Nogami M, Yokozeki T, Yamazaki M, Nakamura H, Watanabe H, Kawamoto K, Nakayama K, Morris AJ, Frohman MA, Kanaho Y. Phosphatidylinositol 4-phosphate 5-kinase α is a downstream effector of the small G protein ARF6 in membrane ruffle formation. Cell. 99: 521-532, 1999.

111. Huang HC, Chen CC, Chang WC, Tao MH, Huang C. Entry of hepatitis B virus into immortalized human primary hepatocytes by clathrin-dependent endocytosis. J. Virol. 86: 9443-9453, 2012.

112. Hunt AG, Xu RQ, Addepalli B, Rao S, Forbes KP, Meeks LR, Xing DH, Mo M, Zhao HW, Bandyopadhyay A, Dampanaboina L, Marion A, Von Lanken C, Li QSQ. Arabidopsis mRNA polyadenylation machinery: comprehensive analysis of protein-protein interactions and gene expression profiling. BMC Genom. 9: 220, 2008.

113. Husain SM, Singh RK, Rani A. Analysis of genetic diversity of soybean germplasm using randomly amplified polymorphic DNA. World Soybean Research Conf., August 10-15, 2009, Beijing, China. pp.52, 2009.

114. Hymowitz T, Matlock RS. Guar in the United States. Oklahoma Agricultural Experiment Station Technical Bulletin. 611: 1-34, 1963.

115. Hymowitz, T. The trans-domestication concept as applied to guar. Economic Botany. 26(1): 49-60, 1972.

116. Jackson CL, Casanova JE. Turning on ARF: the Sec 7 family of guanine nucleotide exchange factors. Trend Cell Biol. 10: 60-67, 2000.

117. Jackson LP, Kelly BT, McCoy AJ, Gaffry T, James LC, Collins BM, Honing S, Evans PR, Owen DJ. A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. Cell. 141: 1220-1229, 2010.

118. Jain M, Nijhawan A, Arora, R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, Khurana JP. F-Box Proteins in Rice. Genome-Wide Analysis, Classification,

Temporal and Spatial Gene Expression during Panicle and Seed Development, and Regulation by Light and Abiotic Stress. Plant Physiol. 143: 1467-1483, 2007.

119. Jamra AR, Philippe O, Raas-Rothschild A, Eck SH, Graf E, Buchert R, Borck G, Ekici A, Brockschmidt FF, Nothen MM. Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. Am. J. Hum. Genet. 88: 788-795, 2011.
120. Janvier K, Bonifacino JS. Role of the endocytic machinery in the sorting of lysosome-associated membrane proteins. Mol. Biol. Cell. 16: 4231-4242, 2005.
121. Janvier K, Kato Y, Boehm M, Rose JR, Martina JA, Kim BY, Venkatesan S, Bonifacino J. Recognition of dileucine-based sorting signals from HIV-1 Nef and LIMP-II by the AP-1 gamma-sigma1 and AP-3 delta-sigma3 hemicomplexes. J. Cell Biol. 163: 1281-1290, 2003.
122. Johal GS, Balint-Kurti P, Weil CF. Mining and harnessing natural variation: A little magic. Crop Sci. 49: 2066-2073, 2008.
123. Jongeneel V, Estreicher A, Baxevanis AD, Ouellette BFF, Wolfsberg TG, Landsman D. et al. EXPRESSED SEQUENCE TAGS (ESTs). Bioinforma A Pract Guid to Anal Genes Proteins. 10(1): 57-63, 2001.
124. Kalthoff C, Alves J, Urbanke C, Knorr R, Ungewickell EJ. Unusual structural organization of the endocytic proteins AP180 and EPSIN 1. J. Biol. Chem. 277: 8209-8216, 2002.
125. Kamiguchi H, Long KE, Pendergast M, Schaefer AW, Rapoport I, Kirchhausen T, Lemmon V. The neural cell adhesion molecule L1 interacts with the AP-2 adaptor and is endocytosed via the clathrin-mediated pathway. J. Neurosci. 18: 5311-5321, 1998.
126. Kanoh H, Williger BT, Exton JH. Arfaptin 1, a putative cytosolic target protein of ADP-ribosylation factor, is recruited to Golgi membranes. J. Biol. Chem. 272: 5421-5429, 1997.
127. Kastning K, Kukhtina V, Kittler JT, Chen G, Pechstein A, Enders S, Lee SH, Sheng M, Yan Z, Haucke V. Molecular determinants for the interaction between AMPA receptors and the clathrin adaptor complex AP-2. Proc. Natl. Acad. Sci. USA. 104: 2991-2996, 2007.

128. Kaur A, Pati PK, Pati AM, Nagpal AK. *In-silico* analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa.* PLOS ONE. 12(9): e0184523, 2017.
129. Keen JH. Clathrin assembly proteins: affinity purification and a model for coat assembly. J. Cell Biol. 10: 1989-98, 1987.
130. Keim P., Diers B.W., Olson T.C., Shoemaker R.C. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics. 126: 735-742, 1990.
131. Kelly BT, McCoy AJ, Spate K, Miller SE, Evans PR, et al. A structural explanation for the binding of endocytic dileucine motifs by the AP2 complex. Nature. 456: 976-979, 2008.
132. Khalil HB,Brunetti SC, Pham UM, Maret D, Laroche A, Gulick PJ. Characterization of the caleosin gene family in the Triticeae. BMC Genomics. 15: 239, 2014.
133. Kirchhausen T. Adaptors for clathrin-mediated traffic. Annu. Rev. Cell Dev. Biol. 15: 705-732, 1999.
134. Kirchhausen T, Owen D, Harrison SC. Molecular Structure, Function, and Dynamics of Clathrin-Mediated Membrane Traffic. Cold Spring Harb Perspect Biol. 6(5): a016725, 2014.
135. Krauss M, Kinuta M, Wenk MR, De Camilli P, Takei K, Haucke V. ARF6 stimulates clathrin/AP-2 recruitment to synaptic membranes by activating phosphatidylinositol phosphate kinase type Igamma. J. Cell Biol. 162: 113-124, 2003.
136. Kuai J, Boman AL, Arnold RS, Zhu X, Kahn RA. Effects of activated ADP-ribosylation factors on Golgi morphology require neither activation of phospholipase D1 nor recruitment of coatomer. J. Biol. Chem. 275: 4022-4032, 2000.
137. Kumar S, Parekh M J, Patel CB, Zala HN, Sharma R, Kulkarni KS, Fougat R S, Bhatt RK, Sakure AA. Development and validation of est-derived ssr markers and diversity analysis in cluster bean (Cyamopsis tetragonoloba). Journal of Plant Biochemistry and Biotechnology. 25(3): 263-269, 2016.
138. Kumar D, Singh NB. Guar in India. Jodhpur: Scientific Publishers. 2002.
139. Kumar S, Hash CT, Nepolean T, Mahendrakar MD, Satyavathi CT, Singh G, Rathore A, Yadav RS, Gupta R, Srivastava RK. Mapping Grain Iron and Zinc

Content Quantitative Trait Loci in an Iniadi-Derived Immortal Population of Pearl Millet. Genes. 9: 248, 2018.

140. Kuravadi NA, Verma S, Pareek S, Gahlot P, Kumari S, Tanwar UK, Bhatele P, Choudhary M, Gill KS, Pruthi V, Tripathi SK, Dhugga KS, Randhawa GS. Guar: an industrial crop from marginal farms, In: Agricultural sustainability: Progress and prospects in crop research, edited by G S Bhullar and N K Bhullar Elsevier, Amsterdam. 47, 2013.
141. Kuravadi NA, Tiwari PB, Choudhary M, Randhawa GS. Genetic diversity study of cluster bean [Cyamopsis tetragonoloba (L.) taub] landraces using rapd and issr markers. International Journal of Advance Biotechnology and Research 4(4): 460-471, 2013.
142. Kuravadi NA, Tiwari PB, Tanwar UK, Tripathi SK, Dhugga KS, Gill KS, Randhawa GS. Identification and characterization of est-ssr markers in cluster bean (spp.). Crop Science 54(3): 1097-1102, 2014.
143. Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology. 157(1): 105-132, 1982.
144. Lam E, Chua NH. ASF-2: A factor that binds to the cauliflower mosaic virus 35S promoter and a conserved GATA motif in cab promoters. Plant Cell. 1: 1147-1156, 1989.
145. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9: 357-9, 2012.
146. Lankey KR, Hallauer AR, Kahler AL. Allelic difference at enzyme *loci* and hybrids performance in maize. J. Hered. 78: 231-234, 1997.
147. Lanza LLB, De Souza Jr CL, Ottoboni LMM, Vieira LMC, De Souza AP. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor. Appl. Genet. 94: 1023-1030, 1997.
148. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and clustal X version 2.0. Bioinformatics. 23: 2947-2948, 2007.
149. Laskowski, R. A. Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. Bioinformatics. 23: 1824-1827, 2007.

150. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P. Coexpression analysis of human genes across many microarray data sets. Genome Res. 14(6): 1085-94, 2004.
151. Lee M, Goldhalk FB, Lamkey KR, Wodmar WW. Association of restriction fragment length polymorphism among maize inbreeds with agronomic performance of their crosses. Crop Sci. 29: 1067-1071, 1989.
152. Lee M, Orci L, Hamamoto S, Futai E, Ravazzola M. et al. Sar1p Nterminal helix initiates membrane curvature and completes the fission of a COPII vesicle. Cell 122: 605-617, 2005.
153. Lee MCS, Miller EA, Goldberg J, Orci L, Schekman R. Bi-directional protein transport between the ER and Golgi. Ann. Rev. Cell Dev. Bio. 20: 87-123, 2004.
154. Lee TH, Kim YK, Pham TTM, Song SI, Kim JK, Kang KY, An G, Jung KH, Galbraith DW, Kim M, Yoon UH, Nahm BH. RiceArrayNet: a database for correlating gene expression from transcriptome profiling, and its application to the analysis of coexpressed genes in rice. Plant Physiol. 151(1): 16-33, 2009.
155. Letourneur F, Gaynor EC, Hennecke S, Demolliere C, Duden R, Emr SD, Riezman H, Cosson P. Coatomer is essential for retrieval of dilysine tagged proteins to the endoplasmic reticulum. Cell. 79(7): 1199-1207, 1994.
156. Lewin DA, Mellman I. Sorting out adaptors. Biochim. Biophys. Acta. 1401: 129-145, 1998.
157. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai YC, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y,Wang G, Li J, Yan J. Genomewide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. 45: 43-50, 2013.
158. Liu R, Vitte C, Ma J, Mahama AA, Dhliwayo T, Lee M, Bennetzen JL. A GeneTrek analysis of the maize genome Proc. Natl. Acad. Sci. USA. 104(28): 11844-9, 2007.
159. Lopes AD, Scapim CA, da Silva Machado MFP, Mangolin CA, Silva TA, Cantagali LB, Teixeira FF, Mora F. Genetic diversity assessed by microsatellite markers in sweet corn cultivars. Scientia Agricola Sci. Agric. 72(6): 513-519, 2015.
160. Lorenz A, Hoegemeyer T. The phylogenetic relationships of US maize germplasm. Nat. Genet. 45: 844-845, 2013.

161. Lu Y, Huggins P, Bar-Joseph Z. Cross species analysis of microarray expression data. Bioinformatics. 25: 1476-1483, 2009.
162. Luo H, Song F, Goodman RM, Zheng Z. Up-regulation of OsBIHD1, a rice gene encoding BELL homeodomain transcriptional factor, in disease resistance responses. Plant Biol. (Stuttg). 7: 459-468, 2005.
163. Ma R, Xu S, Zhao Y, Xia B, Wang R. Selection and validation of appropriate reference genes for quantitative real-time PCR analysis of gene expression in *Lycoris aurea*. Front. Plant Sci. 7: 536, 2016.
164. MacLennan NK, Dong J, Aten JE, Horvath S, Rahib L, Ornelas L. et al. Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. Mol. Genet. Metab. 98(1-2): 203-14, 2009.
165. Mahammad S, Comfort DA, Kelly RM, Khan SA. Rheological Properties of Guar Galactomannan Solutions during Hydrolysis with Galactomannanase and r-Galactosidase Enzyme Mixtures. Biomacromolecules. 8: 949-956, 2007.
166. Malik MFA, Qureshi AS, Ashraf M, Khan MR. Genetic diversity of soybean germplasm from five different origins using RAPD markers. World Soybean res. Conf, August 10-15, 2009, Beijing, China. 2009.
167. Manjaya JG, Bapat VA. Studies on genetic divergence in soybean [Glycine max (L.) Merrill]. J. Oilseeds Res. 25: 178- 180, 2008.
168. Mao L, Van Hemert JL, Dash S, Dickerson JA. Arabidopsis gene co-expression network and its functionalmodules. BMC Bioinformatics. 10: 346, 2009.
169. Mardones GA, Burgos PV, Lin Y, Kloer DP, Magadan JG, Hurley JH, Bonifacino JS. Structural basis for the recognition of Tyrosine-based sorting signals by the Mu3A subunit of the AP-3 adaptor complex. J. Biol. Chem. 288(13): 9563-9571, 2013.
170. Marsan PA, Egidy G, Monfredini G, Disilvestro S, Motto M. RAPD markers in maize genetic-analysis. Maydica. 38: 259-264, 1993.
171. Marti L, Fornaciari S, Renna L, Stefano G, Brandizzi F. COPII-mediated traffic in plants. Trends Plant Sci. 15: 522-528, 2010.
172. Masalia RR, Bewick AJ, Burke JM. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. PLoS ONE 12(7): e0182289, 2017.

173. Matheson LA, Hanton SL, Rossi M, Latijnhouwers M, Stefano G, Renna L, Brandizzi F. Multiple roles of ADP-ribosylation factor 1 in plant cells include spatially regulated recruitment of coatomer and elements of the Golgi matrix. Plant Physiol. 143(4): 1615-27, 2007.
174. Matsuda S, Miura E, Matsuda K, Kakegawa W, Kohda K, Watanabe M, Yuzaki M. Accumulation of AMPA receptors in autophagosomes in neuronal axons lacking adaptor protein AP-4. Neuron. 57: 730-45, 2008.
175. Matsui W, Kirchhausen T. Stabilization of clathrin coats by the core of the clathrin-associated protein complex AP-2. Biochemistry. 29 (48): 10791-10798, 1990.
176. Mattera R, Guardia CM, Sidhu S, Bonifacino JS. Bivalent Motif-Ear Interactions Mediate the Association of the Accessory Protein Tepsin with the AP-4 Adaptor Complex. J Biol Chem. M115.683409, 2015.
177. McMahon HT, Mills IG. COP and clathrin-coated vesicle budding: different pathways, common approaches. Curr. Opin. in Cell Bio. 16: 379-391, 2004.
178. Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P. An endosperm specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm. Plant J. 16: 53-62, 1998.
179. Menetrey J, Macia E, Pasqualato S, Franco M, Cherfils J. Structure of Arf6-GDP suggests a basis for guanine nucleotide exchange factors specificity. Nature Struct. Biol. 7: 466-469, 2000.
180. Mills IG, Praefcke GJK, Vallis Y, Peter BJ, Olesen LE, Gallop JL, Butler PJG, Evans PR, McMahon HT. EpsinR: an AP1/clathrin interacting protein involved in vesicle trafficking. J. Cell Biol. 160: 213-222, 2003.
181. Miranda GV, Sediyama CS, Reis MS, Cruz CD. Genetic diversity among elite Brazilian soybean cultivars with narrow genetic base. Crop Breed. Appl. Biotechnol. 1: 115-123, 2001.
182. Moeller DA, Schaal BA. Genetics relationships among native American maize accessions of the Great Plains assessed by RAPDs. Theor. Appl. Genet. 99: 1061-1067, 1999.

183. Moreno-De-Luca A, Helmers SL, Mao H, Burns TG, Melton AM, Schmidt KR, Fernhoff PM, Ledbetter DH, Martin CL. Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. J. Med. Genet. 48: 141-144, 2011.
184. Morse WJ, Carter JL, Williams LF. Soybeans: Culture and Varieties, U.S. Department of Agriculture Farmers' Bulletin No. 1520, 1-38, 1949.
185. Morse WJ. Soy beans: Culture and Varieties, U.S. Department of Agriculture Farmers' Bulletin No. 1520, 1-33, 1927.
186. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by rna-seq. Nature methods. 5(7): 621-628, 2008.
187. Nakasugi K, Crowhurst RN, Bally J, Wood CC, Hellens RP, Waterhouse PM. De novo transcriptome sequence assembly and analysis of rna silencing genes of Nicotiana benthamiana. PLoS One 8(3): 59534, 2013.
188. Nandha PS, Singh J. Comparative assessment of genetic diversity between wild and cultivated barley using gSSR and EST-SSR markers. Plant Breeding. 133: 28-35, 2014.
189. Ohno H, Fournier MC, Poy G, Bonifacino JS. Structural determinants of interaction of tyrosine-based sorting signals with the adaptor medium chains. J. Biol. Chem. 271: 29009-29015, 1996.
190. Ohno H, Stewart J, Fournier MC, Bosshart H, Rhee I, Miyatake S, Saito T, Gallusser A, Kirchhausen T, Bonifacino JS. Interaction of tyrosine-based sorting signals with clathrin-associated proteins. Science. 269: 1872-1875, 1995.
191. Oldham CE, Mohney RP, Miller SLH, Hanes RN, O'Bryan JP. The ubiquitin-interacting motifs target the endocytic adaptor protein epsin for ubiquitination. Curr. Biol. 12: 1112-1116, 2002.
192. Olukolu BA, Wang GF, Vontimitta V, Venkata BP, Marla S, Ji J, Gachomo E, Chu K, Negeri A, Benson J, Nelson R, Bradbury P, Nielsen D, Holland JB, Balint-Kurti PJ, Johal G. A genome-wide association study of the maize hypersensitive defense response identifies genes that cluster in related pathways. PLOS Genet. 10(8): e1004562, 2014.

193. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. 12: 283-293, 2011.
194. Ooi CE, Dell'Angelica EC, Bonifacino JS. ADP-ribosylation factor 1 (ARF1) regulates recruitment of the AP-3 adaptor complex to membranes. J. Cell Biol. 142: 391-402, 1998.
195. Oshlack A, Robinson MD, Young MD. From RNA-seq reads to differential expression results. Genome biology. 11(12): 220, 2010.
196. Overstreet E, Chen X, Wendland B, Fischer JA. Either part of a Drosophila epsin protein, divided after the ENTH domain, functions in endocytosis of delta in the developing eye. Curr. Biol. 13: 854-860, 2003.
197. Owen DJ, Evans PR. A structurfal explanation for the recognition of tyrosine-based endocytotic signals. Science. 282: 1327-1332, 1998.
198. Palade G. Intracellular aspects of the process of protein synthesis. Science. 189(4200): 347-58, 1975.
199. Paleotti O, Macia E, Luton F, Klein S, Partisani M, Chardin P, Kirchhausen T, Franco M. The small G-protein Arf6GTP recruits the AP-2 adaptor complex to membranes. J. Biol. Chem. 280: 21661-21666, 2005.
200. Panwar P, Saini RK, Sharma N, Yadav D, Kumar A. Efficiency of RAPD, SSR and Cytochrome P450 gene based markers in accessing genetic variability amongst finger millet (Eleusine coracana) accessions. Mol. Biol. Rep. 37: 4075-4082, 2010.
201. Parrish JR, Gulyas KD, Finley RL. Yeast two-hybrid contributions to interactome mapping. Curr. Opin. Biotechnol. 17: 387-393, 2006.
202. Pasrija R, Thakur JK. Analysis of differential expression of Mediator subunit genes in Arabidopsis. Plant Signal. Behav. 7(11): 1676-1686, 2012.
203. Pathak R, Singh M, Henry A. Genetic diversity and interrelationship among clusterbean (Cyamopsis tetragonoloba) genotypes for qualitative traits. Indian J. Agr. Sci. 81(5): 402-406, 2011.
204. Pathak, R., Singh, S. K., Singh, M. and Henry, A. Molecular assessment of genetic diversity in cluster bean (Cyamopsis tetragonoloba) genotypes. Journal of Genetics 89(2): 243-246, 2010.

205. Pearse BM. Coated vesicles from pig brain: purification and biochemical characterization. J Mol Biol. 97:93-98, 1975.
206. Peden AA, Oorschot V, Hesser BA, Austin CD, Scheller RH, Klumperman, J. Localization of the AP-3 adaptor complex defines a novel endosomal exit site for lysosomal membrane proteins. J. Cell Biol. 164: 1065-1076, 2004.
207. Pena-Rosas, JP. Garcia-Casal MN. Pachon H. Mclean MS. Arabi M. Technical considerations for maize flour and corn meal fortification in public health: consultation rationale and summary. ANNALS OF THE NEW YORK ACADEMY OF SCIENCES Issue: Technical Considerations for Maize Flour and Corn Meal Fortification in Public Health. 2014.
208. Persson S, Wei H, Milne J, Page GP, Somerville CR. Identification of genes required for cellulose synthesisby regression analysis of public microarray data sets. Proc. Natl. Acad. Sci. USA. 102(24): 8633-8, 2005.
209. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatic. 30: 1771-1773, 2014.
210. Piperno DR, Flannery KV. The earliest archaeological maize (*Zea mays*L.) from highland Mexico: New accelerator mass spectrometry dates and their implications. Proc. Natl. Acad. Sci. USA. 98(4): 2101-2103, 2001.
211. Poats FJ. Guar, a summer row crop for the southwest. Economic Botany. 14(3): 241-246, 1960.
212. Pombo MA, Zheng Y, Fei Z, Martin GB, Rosli HG. Use of RNA-seq data to identify and validate RT-qPCR reference genes for studying the tomato-Pseudomonas pathosystem. Sci. Rep. 7: 44905, 2017.
213. Puertollano R, Randazzo PA, Presley JF, Hartnell LM, Bonifacino JS. The GGAs promote ARF-dependent recruitment of clathrin to the TGN. Cell. 105: 93-102, 2001.
214. Punia A, Yadav R, Arora P, Chaudhury A. Molecular and morphophysiological characterization of superior cluster bean (Cymopsis tetragonoloba) varieties. Journal of Crop Science and Biotechnology. 12(3): 143-148, 2009.

215. Radhakrishna H, Donaldson JG. ADP-ribosylation factor 6 regulates a novel plasma membrane recycling pathway. J. Cell Biol. 139: 49-61, 1997.
216. Rafalski JA. Association genetics in crop improvement. Curr. Opin. Plant Biol. 13: 174-180, 2010.
217. Randolph LF. The origin of maize. Indian J. Genet. Plant Breed. 19: 1-12, 1959.
218. Rappoport JZ. Simon SM. Endocytic trafficking of activated EGFR is AP-2 dependent and occurs through preformed clathrin spots. J. Cell Sci. 122: 1301-1305, 2009.
219. Ren X, Farias GG, Canagarajah BJ, Bonifacino JS, Hurley JH. Structural basis for recruitment and activation of the AP-1 clathrin adaptor complex by Arf1. Cell. 152, 755-767, 2013.
220. Robinson MS, Bonifacino JS. Adaptor-related proteins. Curr. Opin. Cell Biol. 13: 444-453, 2001.
221. Robinson MS. Adaptable adaptors for coated vesicles. Trends Cell Biol. 14(4): 167-74, 2004.
222. Robinson DG, Hinz G, Holstein SEH. The molecular characterization of transport vesicles. Plant Mol. Biol. 38 (1-2): 49-76, 1998.
223. Robinson DG, Pimpl P. Clathrin and post-Golgi trafficking: a very complicated issue. Trends Plant Sci. 19: 134-139, 2014.
224. Roth MG. Snapshots of ARF1: implications for mechanisms of activation and inactivation. Cell. 97: 149-152, 1999.
225. Rupp JV, Mangolin CA, Scapim CA, Machado MFPS. Genetic structure and diversity among sweet corn (su1- germplasm) progenies using SSR markers. Maydica. 54: 125-132, 2009.
226. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-25, 1987.
227. Sakai H, Aoyama T, Oka A. Arabidopsis ARR1 and ARR2 response regulators operate as transcriptional activators. Plant J. 24: 703-711, 2001.
228. Salado-Navarro LR, Sinclair TR, Hinson K. Changes in yield and seed growth traits in soybean cultivars released in the Southern USA from 1945 to 1983. Crop Sci. 33: 1204-1209, 1993.

229. Sambrook J, Russell DW. Molecular cloning: a laboratory manual (CSHL press). 2001.
230. Satyavathi CT, Bhat KV, Bharadwaj C, Tiwari SP, Chaudhury VK. AFLP analysis of genetic diversity in Indian soybean [Glycine max (L.) Merr.] varieties. Genet. Resour. Crop Evol. 53: 1069-1079, 2006.
231. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E , Peto M, Grant D, Shu S, Goodstein D, Barry K, Griggs MF, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman1 A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA. Genome sequence of the palaeopolyploid soybean. Nature. 463: 178-183, 2010.
232. Schoonheim PJ, Veiga H, da Costa PD, Friso G, van Wijk KJ, de Boer AH. A comprehensive analysis of the 14-3-3 interactome in barley leaves using a complementary proteomics and two-hybrid approach. Plant Physiol. 143: 670-683, 2007.
233. Schröder S, Ungewickell E. Subunit interaction and function of clathrin-coated vesicle adaptors from the Golgi and the plasma membrane. J. Biol. Chem. 266: 7910-7918, 1991.
234. Seo YS, Chern M, Bartley LE, Han M, Jung KH, Lee I. Walia H, Richter T, Xu X, Cao P, Bai W, Ramanan R, Amonpant F, Arul L, Canlas PE, Ruan R, Park CJ, Chen X, Hwang S, Jeon JS, Ronaldet PC. Towards establishment of a rice stress response interactome. PLoS Genet. 7: e1002020, 2011.
235. Serventi IM, Cavenaugh E, Moss J, Vaughan M. Characterization of the gene for ADP-ribosylation factor (ARF) 2, a developmentally regulated, selectively expressed member of the ARF family of approximately 20-kDa guanine nucleotide-binding proteins. J. Biol. Chem. 268: 4863-4872, 1993.
236. Shimada T, Fuji K, Tamura K, Kondo M, Nishimura M, Hara-Nishimura I. Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA. 100: 16095-16100, 2003.

237. Shirsat A, Wilford N, Croy R, Boulter D. Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Mol. Gen. Genet. 215: 326-31, 1989.
238. Shteyn E, Pigati L, Folsch H. Arf6 regulates AP-1B-dependent sorting in polarized epithelial cells. J. Cell Biol. 194, 873-887, 2011.
239. Simmen T, Honing S, Icking A, Tikkanen R, Hunziker W. AP-4 binds basolateral signals and participates in basolateral sorting in epithelial MDCK cells. Nat. Cell Biol. 4: 154-159, 2002.
240. Singh RJ, Hymowitz T. Soybean genetic resources and crop improvement. Genome. 42: 605-616, 1999.
241. Sleper DA, Poehlman JM. Breeding Field Crops. 5th Edition. Iowa State Press. Ames, USA. 2006.
242. Smarda P, Bures P. The variation of base composition in plant genomes. In Plant genome diversity (Springer, Vienna). 1: 209-235, 2012.
243. Smarda P, Bures P, Horová L, Leitch IJ, Mucina L, Pacini E, Tichý L, Grulich, V. Ecological and evolutionary significance of genomic GC content diversity in monocots. Proc. Natl. Acad. Sci. USA. 111: 4096-4102, 2014.
244. Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JF, Cregan PB. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109: 122-128, 2004.
245. Song J, Lee MH, Lee GJ, Yoo CM, Hwang I. Arabidopsis EPSIN1 plays an important role in vacuolar trafficking of soluble cargo proteins in plant cells via interactions with clathrin, AP-1, VTI11, and VSR1. Plant Cell. 18: 2258-2274, 2006.
246. Stamnes MA, Rothman JE. The binding of AP-1 clathrin adaptor particles to Golgi membranes requires ADP-ribosylation factor, a small GTP-binding protein. Cell. 73: 999-1005, 1993.
247. Stephens DJ, Banting G. Specificity of interaction between adaptorcomplex medium chains and the tyrosine-based sorting motifs of TGN38 and lgp120. Biochem J. 335 (3): 567-572, 1998.
248. Stuart JM, Segal E, Koller D, Kim SK. A gene coexpression network for global discovery of conserved genetic modules. Science. 302: 249-255, 2008.

249. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 321(5891): 956-960, 2008.
250. Tai YS. Interactome of signaling networks in wheat: the protein-protein interaction between TaRAR1 and TaSGT1. Mol. Biol. Rep. 35, 337-343, 2008.
251. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30: 2725-2729, 2013.
252. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics. 32: 2230-2232, 2016.
253. Tanwar UK, Pruthi V, Randhawa GS. RNA-Seq of guar (*Cyamopsis tetragonoloba*, L. Taub.) Leaves: *De novo* Transcriptome Assembly, Functional Annotation and Development of Genomic Resources. Front. Plant Sci. 8: 91, 2017.
254. Tardif G, Kane NA, Adam H, Labrie L, Major G, Gulick P. et al. Interaction network of proteins associated with abiotic stress response and development in wheat. Plant Mol. Biol. 63: 703-718, 2007.
255. Terzaghi WB, Cashmore AR. light-regulated transcription. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46: 445-474, 1995.
256. Thakur JK, Agarwal P, Parida S, Bajaj D, Pasrija R. Sequence and expression analyses of KIX domain proteins suggest their importance in seed development and determination of seed size in rice, and genome stability in Arabidopsis. Mol. Genet. Genomics. 288(7-8): 329-346, 2013.
257. Theos AC, Tenza D, Martina JA, Hurbain I, Peden AA, Sviderskaya EV, Stewart A. Robinson MS, Bennett DC, Cutler DF, Bonifacino JS. Functions of adaptor protein (AP)-3 and AP-1 in tyrosinase sorting from endosomes to melanosomes. Mol. Biol. Cell. 16: 5356-5372, 2005.
258. Tian H,Wang F, Zhao J, Yi H,Wang L,Wang R, Yang Y, Song W. Development of maizeSNP3072, a high-throughput compatible SNP array, for DNA fingerprinting identification of Chinese maize varieties. Mol Breed. 35: 1-11, 2015.

259. Tito CM, Poggio L, Naranjo CA. Cytogenetic studies in the genus *Zea*: DNA content and heterochromatin in species and hybrids. Theor. Appl. Genet. 83: 58-64, 1991.
260. Traub LM, Downs MA, Westrich JL, Fremont DH. Crystal structure of the α appendage of AP2 reveals a recruitment platform for clathrin-coat assembly. Proc. Natl. Acad. Sci. USA. 96: 8907-8912, 1999.
261. Traub LM. Tickets to ride: selecting cargo for clathrin-regulated internalization. Nat. Rev. Mol. Cell Biol. 10: 583-596, 2009.
262. Tsuchiya M, Price SR, Tsai SC, Moss J, Vaughan M. Molecular identification of ADP-ribosylation factor mRNAs and their expression in mammalian cells. J. Biol. Chem. 266: 2772-2777, 1991.
263. Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H, Bertani C, Davassi A, Mayer KF, Schön CC. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genomics. 15: 232-233, 2014.
264. Van Verk MC, Hickman R, Pieterse CMJ. Van Wees SCM. RNA-Seq: Revelation of the messengers. Trends Plant Sci. 18(4): 175-179, 2013.
265. Vandamme EJ, De Baets S, Steinbuchel A. Polysaccharides II, polysaccharides from Eukaryotes: Biopolymers. 6: 245-274, 2002.
266. Verkerk AJ, Schot R, Dumee B, Schellekens K, Swagemakers S, Bertoli-Avella AM, Lequin MH, Dudink J, Govaert P, van Zwol AL. Mutation in the AP4M1 gene provides a model for neuroaxonal injury in cerebral palsy. Am. J. Hum. Genet. 85: 40-52, 2009.
267. Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, Legrand J. et al. The auxin signalling network translates dynamic input into robust patterning at the shoot apex. Mol. Syst. Biol. 7: 508, 2011.
268. Vitale A, Chrispeels MJ. Sorting of proteins to the vacuoles of plant cells BioEssays. 14: 151-160, 1992.
269. Vitale A, Galili G. The endomembrane system and the problem of protein sorting. Plant Physiol. 125: 115-118, 2001.

270. Volkamer A, Griewel A, Grombacher T, Rarey M. Analyzing the Topology of Active Sites: On the Prediction of Pockets and Sub-pockets. J. Che. Inf. Model. 50: 2041-2052, 2011.
271. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. Comparative assessment of large-scale data sets of protein-protein interactions. Nature. 417: 399-403, 2002.
272. Wang J, Li Y, Lo SW, Hillmer S, Sun SSM, Robinson DG, Jiang L. Protein mobilization in germinating mung bean seeds involves vacuolar sorting receptors and multivesicular bodies. Plant Physiol. 143: 1628-1639, 2007.
273. Wang Q, Xue X, Li Y, Dong Y, Zhang L, Zhou Q, Deng F, Ma Z, Qiao D, Hu C, Ren Y. A maize ADP-ribosylation factor ZmArf2 increases organ and seed size by promoting cell expansion in *Arabidopsis*. Physiol. Plant. 156: 97-107, 2016.
274. Wang Y, Hu Z, Yang Y, Chen X, Chen G. Function annotation of an SBP-box gene in arabidopsis based on analysis of co-expression networks and promoters. Int. J. Mol. Sci. 10(1): 116-32, 2009.
275. Wang YJ, Wang J, Sun HQ, Martinez M, Sun YX, Macia E, Kirchhausen T, Albanesi JP, Roth MG, Yin HL. Phosphatidylinositol 4 phosphate regulates targeting of clathrin adaptor AP-1 complexes to the Golgi. Cell. 114: 299-310, 2003.
276. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Rev. Genet. 10(1): 57-63, 2009.
277. Wass MN, Fuentes G, Pons C, Pazos F, Valencia A. Towards the prediction of protein interaction partners using physical docking. Mol. Syst. Biol. 7: 469. 2011.
278. Watson PJ, Frigerio G, Collins BM, Duden R, Owen DJ. Gamma-COP appendage domain - structure and function. Traffic. 5(2): 79-88, 2004.
279. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35(2): 407-410, 2007.
280. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett C J, Rogers J, BÃhler JR. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. 453(7199): 1239-1243, 2008.

281. Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic acids research 18(22): 6531-6535, 1990.
282. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. 13: 59-69, 2011.
283. Wolf JBW. Principles of transcriptome analysis and gene expression quantification: An rna-seq tutorial. Mol. Ecol. Resour. 13(4): 559-572, 2013.
284. Wu G, Zhang L, Yin Y, Wu J, Yu L, Zhou Y, Li M. Sequencing, *de novo* assembly and comparative analysis of *raphanus sativus* transcriptome. Front. plant sci. 6: 198, 2015.
285. Xiang L, Ed Etxeberria E, den Ende WV. Vacuolar protein sorting mechanisms in plants. FEBS J. 280: 979-993, 2013.
286. Xie C, Gong W, Zhu Z, Yan L, Hu Z, Peng Y. Comparative transcriptomics of Pleurotus eryngii reveals blue-light regulation of carbohydrate-active enzymes (CAZymes) expression at primordium differentiated into fruiting body stage. Genomics. 110: 201-209, 2018.
287. Xu D, Zhang Y. Improving physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J. 101: 2525-2534, 2011.
288. Xu XH, Chen H, Sang YL, Wang F, Ma JP, Gao XQ, Zhang XS. Identification of genes specifically or preferentially expressed in maize silk reveals similarity and diversity in transcript abundance of different dry stigmas. BMC Genomics. 13: 294, 2012.
289. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. Nature Methods. 12: 7-8, 2015.
290. Yap CC, Murate M, Kishigami S, Muto Y, Kishida H, Hashikawa T, Yano R. Adaptor protein complex-4 (AP-4) is expressed in the central nervous system neurons and interacts with glutamate receptor delta2. Mol. Cell Neurosci. 24: 283-295, 2003.
291. Yonekura-Sakakibara K, Tohge T, Fumio Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K. Comprehensive Flavonol

Profiling and Transcriptome Coexpression Analysis Leading to Decoding Gene–Metabolite Correlations in Arabidopsis. The Plant Cell. 20: 2160-2176, 2008.

292. Yoshihisa T, Barlowe C, Schekman R. Requirement for a GTPase activating protein in vesicle budding from the endoplasmic reticulum. Science. 259: 1466-1468, 1993.
293. You Q, Zhang L. Yi X, Zhang k, Yao D, Zhang X, Wang Q,Zhao X, Ling Y, Xu W, Li F, Su Z Co-expression network analyses identify functional modules associated with development and stress response in Gossypium arboretum. Sci. Rep. 6: 38436, 2016.
294. Zhang H, Tian H, Chen M, Xiong J, Cai H, Liu Y. Transcriptome analysis reveals potential genes involved in flower pigmentation in a red-flowered mutant of white clover (Trifolium repens L.). Genomics. 110: 191-200, 2018.
295. Zhang S, Chen C, Li L, Meng L, Singh J, Jiang N, Deng X-W, He Z-H, Lemaux PG. Evolutionary Expansion, Gene Structure, and Expression of the Rice Wall-Associated Kinase Gene Family. Plant Physiol. 139: 1107-1124, 2005.
296. Zhang ZL, Xie Z, Zou X, Casaretto J, Ho TH, Shen QJ. A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells. Plant Physiol. 134: 1500-1513, 2004.
297. Zhong SL, Lin ZF, Grierson D. Tomato ethylene receptor-CTR interactions: visualization of NEVER-RIPE interactions with multiple CTRs at the endoplasmic reticulum. J. Exp. Bot. 59: 965-972, 2008.
298. Zhou X, Carter Jr. TE, Cui Z, Miyazaki S, Burton JW. Genetic base of Japanese soybean cultivars released during 1950 to 1988. Crop Sci. 40: 1794-1802, 2000.
299. Zhou X, Li J, Cheng W, Liu H, Li M, Zhang Y, Li W, Han S, Yingdian Wang Y. Gene structure analysis of rice ADP-ribosylation Factors (OsARFs) and their mRNA expression in developing rice plants. Plant Mol. Bio. Rep. 28: 692-703, 2010.
300. Zhou Z, Cong P, Tian Y, Zhu Y. Using RNA-seq data to select reference genes for normalizing gene expression in apple roots. PLOS One. 12(9): e0185288, 2017.
301. Zhu L, Zhang Y, Zhang W, Yang S, Chen JQ, Tian D. Patterns of exon-intron architecture variation of genes in eukaryotic genomes. BMC Genomics. 10: 47, 2009.
302. Zhu Y, Drake MT, Kornfeld S. ADP-ribosylation factor 1 dependent clathrin-coat assembly on synthetic liposomes. Proc. Natl. Acad. Sci. USA. 96: 5013-5018, 1999.

303. Zouhar J, Munoz A, Rojo E. Functional specialization within the vacuolar sorting receptor family: VSR1, VSR3 and VSR4 sort vacuolar storage cargo in seeds and vegetative tissues. Plant J. 64: 577-588, 2010.
304. Zwiewka M, Feraru E, Möller B, Hwang I, Feraru MI, Kleine-Vehn J, Weijers D, Jiří Friml J. The AP-3 adaptor complex is required for vacuolar function in Arabidopsis. Cell Res. 21: 1711-1722, 2011.

LIST OF PUBLICATIONS

- **Nishu Mittal,** Dharmendra Singh and Gursharn S. Randhawa.*Structural, evolutionary, SNP genotyping and expression analyses of AP-4 complex genes of plants. (Communicated in Plos ONE)

- **Nishu Mittal**, Poonam S. Jaiswal, Dharmendra Singh and Gursharn S. Randhawa.* Co-expression Analysis and Protein-Protein Interaction Studies in AP-4 Complex - Mediated Vesicular Trafficking in *Arabidopsis thaliana* and *Zea mays.* (Communicated in Genomics)

- Poonam Jaiswal, **Nishu Mittal** and Gursharn S. Randhawa.* Cyamopsis tetragonoloba type 1 metallothionein (CtMT1) gene is upregulated under drought stress and its protein product has an additional C-X-C motif and unique metal binding pattern. (Communicated in International journal of Biological Macromolecules).

CONFERENCES

- Nishu Mittal, Poonam S. Jaiswal, Santosh Kumar, Gursharn S. Randhawa. Identification and characterization of genes encoding mu, beta, epsilon and sigma subunits of AP-4 complex of Cyamopsis tetragonoloba (L.) Taub. pp 29. 3rd Mini Symposium on Cell Biology. 23rd May, 2017, National Centre for Cell Science, Pune.
- Nishu Mittal, Dharmendra Singh,Gursharn Singh Randhawa. Role of AP-4 Complex Genes in Seed Development. Pp 93. International Conference on "Trends in Biochemical and Biomedical Research: Advances and Challenges" February 13-15, 2018, Department of Biochemistry, Institute of Science, Banaras Hindu University , Varanasi. Uttar pradesh

Appendix 1

>Seq 1 [Organism *Cyamopsis tetragonoloba*] epsilon subunit of AP-4 complex (AP4E), full CDS sequence

Accession number: MG009494

```
ATGGATCAGTTGAAAACATTCAGGAGGGAGCTAGCAATGGGCTCCCAGAGCGGCCAATCCAAGGAATTTCTCGATCT
CATCAAGTCCATCGGTGAAGCCCGATCCAAGGCGGAGGAAGACCGCATCGTCCTCCGTGAGATCGAAACCCTCAAAC
GCCGCATCTCCGACGCCGACATCCCCAAACGGAAGATGAAGGAGTACATCATACGCCTCCTCTACGTCGAAATGCTC
GGCCACGACGCTTCCTTCGGTTACATCCACGCCGTTAAGATGACTCACGACGATTTGCTTTCACTCAAACGGACCGG
TTACCTCGCCGTTACTCTCTTCCTCAGCGACGACCACAACCTCATCATCCTCATCGTTAATACAATTCAGAAGGATC
TTAAGTCCGATAACTACCTCGTCGTTTGCGCAGCTCTCAACGCCGTTTGCCGCCTCATCAATGAGGAAACCATCCCT
GCCATTTTGCCTCTTATTGTTGACCTCCTTTCCCACCCGAAAGACGCCGTTCGGAAGAAAGCCGTTATGGCTCTTCA
TCGCTTCCACCACAAATCACCTTCTTCCGTTTCCCATTTGCTCTCCAATTTCCGCAAGCGGTTGTGTGATAATGACC
CTGGTGTTATGGGTGCCACTCTGTGCCCTCTTTTTGACCTCATATCTGTTGATCCTAATCCTTACAAGGACCTCGTT
GTTAGCTTCGTTAGCATTCTTAAACAAGTCGCTGAACACAGGTTGCCCAAGGGCTATGACTATCATCAAATGCCTGC
ACCTTTCATTCAGATTAAGCTGCTTAAAATACTGGCATTGTTGGGGAGCGGGGATAAGCATGCTAGCGAACACATGT
ATACTGTGATTGGTGAAATAATTAGGAAGGGCGATTTGTCTAGCAACATAGGGAATGCTATTCTTTATGAGTGCATA
```

TGCTGTGTCTCTTCCATACACCCGAATGCTAAGTTGTTAGAAGCTGCTGCGGATGTTATTGCCAAGTTTTTGAAGAG
TGATAGTCATAATCTGAAGTACATGGGCATTGACGCCCTTGGTCGGGTGATAAAGTTAAGTCCACATATTGCAGAAC
AACATCAACTGGCTGTCATTGACTGCTTAGAGGACCCTGATGATACTCTGAAGCGGAAAACTTTTGAACTGCTATAC
AAGATGACCAAGTCCTCCAACGTGGAAGTGATTGTTGACCGTATGATTGATTACATGATTAGCATTAGTGATGACCA
TTATAAGACCTATATAGCATCTCGGTGTGTTGAACTTGCCGAGCAATTTGCACCAAGTAATCATTGGTTTATACAGA
CCATGAATAAAGTTTTTGAGCATGCTGGAGATCTTGTGAATATTAAGGTAGCACATAATTTGATGCGGTTGATTGCT
GAAGGATTTGGAGAGGACGATGATGCTGCAAATAGTCAGTTGAGATCATCTGCTGTTGAGTCATATTTGCGCATTAT
TGGAGAGCCAAAGCTGCCATCTGTATTCCTTCAAGTCATCTGTTGGGTTCTGGGGGAATATGGCACGGCTGATGGAA
AGTATTCTGCTTCCTATATCAGTGGGAAGTTGTGTGACATTGCAGAGGCATATTCTAACGATGAAAACATTAAGGGT
TATTCAATTTCAGCATTGATGAAAATTTATGCATTTGAAGTAGCAGCTGGGAGGAAAGTGGATATCCTACCTGAGTG
TCAATCTCTGATTGAAGAATTATTAGCATCCCACTCCACAGATTTGCAACAACGTGCTTATGAATTGCAAGCCCTTA
TTGGTTTGGACGCACGAGCAGTTAAAACAATAATGCCACGTGATGCAAGTTGTGAAGACATTGAGGTTGATAAGAAT
CTTTCTTTTCTCAATGATTATGTCCAGCGGTCTCTAGAAAAGGGTGCCCAGCCTTACATTCCTGAGGATCAACGTAC
TGGACTGGGGAATATGACTAACTTCAGAAGACAATATCAGCATGAATCTTCGCAGCATGGTCTCAGATTTGAGGCGT
ATGAACTTCCAAAGCCTCCAATACAATCAAATGTTGCTCCAGTTTCACTTGCATCCTCAACAGACATTGCTCCAGTA
CCAGAGACATTGTATTCTAGAGAGACGCAGCATGTCTCATCAGTGGGATCAACATCAGAGGCTGGATCGTCTGAGCT
TAAACTACGGCTTGATGGTGTTCAAAAGAAATGGGGTAGGCCAACTTATTCGTCTTCAACATCTTCCACCACATATT
CTACTTCTGAAAAGCCAACCAACGGGGCAACAAAGGTGGATGGTGCAATATCTGTTGACTCAAAAGTTCATGATAGT
TATGATTCAAGAAAGAAACAGGTTGAGATTCCTCCAGAAAAGCAGAAGCTTGCTGCTTCGTTGTTTGGTGGTGGTAG
TTCAACTAAGTATGAGAAAAGATCATCCACTAGTCATAAGGCTCCAAAAGCTAGTGCTACTGATGGATCCCAGGGAT
CAAAGGCTGCAGTTCTTTCCAATGAAATATCTGTGGAGAAAACAATTCATCAACCTCCACCTGATTTGCTCGACTTG
GGTGAACCGGTTGACACTACAGCTCCTTCAGCCGTTGACCCATTCAAGCAATTGGAAGGACTTATTGACCCAAGTGC
TAGCTCTGCAACAAACCACAATGTTGGTACTGCAACAAAAGCACCTGATATTATAGCACTCTATGCAGATACTACTG
CAAGTGGACAGAGTGGAAGTGGTGGCTATTCTATCCCTGTGAGTTGGGATAATCTAAATCTTTTATCTGAGTTGTCA
AATACACCTTTGCCTGAATCTACGAAGGGTCCAAATGCCAAAAATTCGTTGGAGAAGGATGCATTAGTGAGACACAT
GGGTGTGAATCCCTCAAATCAGAATCCCAACTTGTTCAGGGATTTGCTTGGCTAA

>Seq 2 [Organism *Cyamopsis tetragonoloba*] beta subunit of AP-4 complex (AP4B), full CDS sequence

Accession Number: MG009493

ATGGCTCCACCGCCACCGCAGTCTAATCGATCGCCCTCGCCGTCGCAACCGTCGGGGAAGAGTGAAGTCTCCGATCT
GAAATGCCAGCTCCCGCAGCTTGCTGGGAGCCGAGCTCCGGGCGCTGATGATTGTAAGAGGGATCTCTTCAAGAAGG
TCATTTCCAACATGACTATAGGTATTGATGTCTCCTCTCTCTTTGGGGAGATGGTAATGTGCTCAGCAACATCAGAC
ATTGTTCTGAAAAAAATGTGCTACCTTTATGTTGGGAATTATGCAAAGGTCAATCCTGATCTTGCACTTTTGACGAT
TAATTTTCTGCAAAAAGATTGCAAGGATGAGGATCCAATGATTCGGGGATTAGCATTGAGGAGTTTGTGTTCACTCA
GGGTGGCAAACTTGGTGGAGTATTTGGTTGGGCCATTAGGGACTGGATTGAAGGACAGTAACAGTTATGTTAGGATG
GTAGCAGTTATCGGAGTTCTGAAACTGTATCATATATCAGCTTCCACGTGTATTGATGCAGATTTTCCAGCAACACT
GAAGCATTTGATGCTTAATGACACAGATACTCAGGTAATTGCAAATTGTTTGTCTGCTTTACAAGAAATTTGGACCT
TGGAGTCAGCCACATCAGAGGAAGCGGCCAGGGAGAGAGAAACTTTGATTAGCAAGCCGGTTGTTTATTACCTTCTG
AATCGCATTAAGGAATTCAGTGAATGGGCACAATGTCTTGTGCTGGAATTAGTGGCCAAGTACATTCCATCAGATAA
CAGTGAAATATTTGATATAATGAATCTCCTTGAAGATAGACTCCAGCATGCAAATGGTGCTGTTGTCTTGGCAACTG
TTAAAGTATTCCTACAGTTGACTTTATCTATGGCCGATGTTCATCAGCAGGTATATGAGCGCATCAAAGCCCCTCTG
TTAACTCAAGTGAGTTCAGGAAGTCCAGAACAATCTTACGCAGTTTTAAGCCATCTGCATCTGTTGGTCATGCGTGC
ACCTTATATATTTTCCTCAGACTACAAACACTTCTATTGCCAGTATAATGAGCCATCATATGTCAAAAAGTTGAAGC
TTGAAATGCTGACTGCAGTTGCAAATGAAAGTAACACCTATGAGATAGTGACAGAATTGTGTGAATACGCTGCAAAT
GTCGACATCCCAATTGCTAGGGAATCAATTAGGGCTGTTGGGAAAATAGCTTTGCAGCAGTACGATGTCAATGCTAT
TGTTGATCGACTTCTACAATTTCTTGAGATGGAAAAGGACTATGTGACTGCAGAAGCTCTGGTTCTTGTGAAAGATC
TGCTCAGGAAATATCCACAATGGAGTCAGGATTGTATTGCTGTTGTAGGGAATATCAGTAGCAAAAATGTTCAAGAA
CCCAAGGCTAAGGCAGCTCTTATATGGATGCTGGGGGAATATTCTCAGGATATGCATGATGCTCCGTATGTTTTGGA
GAGTTTAGTTGAAAATTGGGAAGAGGAGCATTCTGCTGAGGTCCGCTTACATCTTCTTACTGCAGTCATGAAGTGTT

TCTTTAAGAGACCGCCTGAGACTCAAAAAGCATTAGGAGCTGCATTGGCTGCTGGTCTTGCTGATTTTCACCAGGAT
GTTCATGATAGGGCCTTGTTTTATTACAGGCTTCTGCAATACAAAGTATCAGTGGCAGAGAGTGTGGTGAATCCCCC
TAAGCAAGCAGTTTCAGTATTTGCTGATACTCAGAGCAGTGAAATCAAAGATAGAATATTTGATGAGTTTAACAGTT
TATCTGTTGTATATCAGAAGCCATCTTACATGTTCACTGATAAGGAACACCGAGGAACATTTGAGTTTTCAGATGAA
CTTGGAAATCTATCTATTAGTGCAGAAGATTCTGCTGTTCCTGCTCAGAGGGTGGAGGCAAATGACAAGGATCTGCT
TCTAAGTACCACAGAGAAAGACGAAGGTAGAGATCCCGGTAGCAATGGTTCTGTCTATGATGCGCCTTCCTATAATG
GTTCATCTGCCCCTTCTACTACTACACAACCGCTAGCAGATTTGGCATTTCCTCCAAGTTCCAGTGTAACCGGTCAA
GCTTCAGCATCTAGCTTGGTAATCGATGATCTGCTCGGTTTAGATTTTTCAGTTGGGACTGCAACCACACCTCCACC
TCCTCCATTAAGCCTCAACCCAAAGGCGGTTCTGGATCCTGGCACATTTCAGCAGAAATGGCGTCAGCTGCCAATAT
CTTTATCAGAGGAATATTCAATTAGTCCTCAAGGAATTGCATCATTAACAAGTCCCCATGCACTACTCCGGCACATG
CAAAGTCATTCGATACATTGCATTGCATCCGGTGGTCAGTCTCCCAACTTCAAGTTCTTTTTCTTTGCCCAGAAAGC
GGAAGAAGCTTCAATGTATCTTGTAGAGTGTATAATTAACACAGCGTCAGCCAAGTCACAGATTAAAATAAAAGCTG
ACGACCAAAGTTCATCCCAGGCATTCTCAACATTATTCCAATCAGCCTTGTCCAAGTTTGGTTTGCCTTGA

>Seq 3 [Organism *Cyamopsis tetragonoloba*] mu subunit of AP-4 complex (AP4M), full CDS sequence

Accession Number: MG009492

ATGATCTCACAGTTCTTCGTCCTGTCACAGCGTGGCGATAACATCGTTTATCGTGACTACCGTGGTGAAGTGGCCAA
GGGAAGTACAGAGATATTTTTCCGCAAAGTAAAGTTTTGGGAAGATGGGGGGCTGGAGGAGGCACCACCTGTCTTTA
ATGTGGATGGTGTAAATTACTTTCATGTGAAAGTTGTTGGATTATTGTTTGTGGCAACTACAAAGGTTAATATGTCA
CCTTCTTTTGTCTTGGAGCTTTTACAAAGAATTGCTCGAGTTATTAAAGATTACCTTGGGATTCTCAATGAAGACTC
ATTGCGAAAAAACTTTGTGCTTGTGTACGAGTTACTTGATGAAGTTATTGATTTTGGTTATGTGCAAACAACATCTA
CTGAGGTTTTGAAGTCATATGTTTTCAATGAGCCAATTATCATTGATACTGCACGGCTGCCTCCCTTTGGTCCTGCT
TCCATATTTATGCAAGGGACCAAAAGAATGCCAGGTACAGCTATTACAAAATCTGTTGTTGCTAATGAACCTGGTGG
TAGGAAGAGGGATGAGATCTTTGTTGATGTAATCGAGAAAATAAGTGTTACATTCAATTCCAGTGGATTTATTCTTA
CTAGTGAGATAGATGGCACCATTCAAATGAAGAGTTACCTTACCGGTAACCCGGAGATTCGACTTGCTCTCAATGAG
GACCTGAGTATCGGAACAAGTGATTATAGAGGTTCAGGTGTTGTGATTTTAGATGATTGTAACTTCCACGAGTCTGT
ACACCTTGATAGTTTTGATATTGACCGAACTTTGTCACTGGTACCGCCAGATGGTGAATTCCCTGTCATGAATTATC
GTATGACTCAACCATTTAAGCCACCCTTTCGTATTAATGCATTGATTGAAGAAACAGGATCTCTGAAGGCTGAAGTG
ACCATTAAAGTGCGAGCTGAATTCAACTCAAGCATCAATGCTAACACAGTTATTGTACAGATGCCACTGCCATCATT
TACATCTCGTGTTAATTTTGAGTTAGAACCTGGAGCAATTGGGCAGACAACTGATTTTAAGGAAGCGAACAAAAGAC
TGGAATGGGGCCTGAAAAAGGTTGTTGGTGGAGCTGAACATACGTTACGAGCAAAGTTGACATTTTCACAGGAATTA
CATGGAAATATCATGAGAGAAGCAGGACCTCTTAGCATGACTTTCACTATACCTATGTACAATTCTTCAAGGCTTCA
GGTGAAGTACTTGCAAATTGCAAAGAAGTCAAAAACTCATAATCCATATCGATGGGTGAGATATGTAACCCAAGCAA
ACTCATACGTAGCCCGGTTGTAA

>Seq 4 [Organism *Cyamopsis tetragonoloba*] sigma subunit of AP-4 complex (AP4s), full CDS sequence

Accession Number: MG009491

ATGGGGATCCGATTCGTACTGATGGTGAACAAGCAAGGCCAAACGCGTCTTGCCCAATACTACGAATACCTCACTCT
CGAAGAAAGGCGTGCTCTTGAAGGTGAAATCGTTCGCAAATGCCTCGCTCGTAACGAACAGCAGTGTTCATTTGTTG
AGCATCGCAACTACAAAATTGTGTATAGGCGCTATGCATCATTGTTTTTCCTGGTTGGAGTTGATGATGATGAGAAT

GAGCTGGCTATTCTGGAATTTATACATCTCTTAGTTGAAACTATGGATCGTCATTTTGGCAATGTGTGTGAGCTAGA
TATCATGTTCCATTTAGAAAAGGCACATTTTATGTTAGAGGAAATGGTCATGAATGGTTGCATTGTGGAGACAAGCA
AGTCAAATATTCTGACTCCGATTCAGCTGATGGATAAAACATCTTGA

Appendix 2. Nucleotide/protein accession numbers of epsilon subunit of AP-4 complexes from various plant species.

Sr. No.	Nucleotide Accession Number	Protein Accession Number	Chromosome Number	Species Name	Abbrevation of Species
1	XM_014665277.1	A0A1S3VQU9	11	*Vigna radiate*	Vr
2	XM_004509436.2	A0A1S2YT50	Ca7	*Cicer arietinum*	Ca
3	AET03493	G7L7U1	8	*Medicago truncatula*	Mt
4	EXC31026	W9S8A1	-	*Morus notabilis*	Mn
5	XM_016860865.1	A0A1U8IUW2	-	*Gossypium hirsutum*	Gt
6	XM_016716770.1	A0A1U8GLV5	1	*Capsicum annuum*	Can
7	XM_016596934.1	A0A1S3YJY8	-	*Nicotiana tabacum*	Nt
8	XM_008446014.2	A0A1S3B9W4	-	*Cucumis melo*	Cm
9	NM_102908.3	NP_174454.2	1	*Arabidopsis thaliana*	At
10	KOO35705	A0A0M0KA45	-	*Chrysochromulina* sp	Ce
11	KUF76162	A0A0W8BWE5	-	*Phytophthora nicotianae*	Pn
12	EMT21569	M8BHW5	-	*Aegilops tauschii*	Ata
13	KMZ63002	A0A0K9P441	-	*Zostera marina*	Zma
14	NM_001176239.1	NP_001169710.1	3	*Zea mays*	Zm
15	OAY80461	A0A199VU55	-	*Ananas comosus*	Ac
16	ERP62950	U5GHH1	LGIV	*Populus trichocarpa*	Pt
17	EMS64146	M7ZWW2	-	*Triticum urartu*	Tu
18	XM_010646179.2	XP_010644481.1	2	*Vitis vinifera*	Vv
19	XM_003564868.3	XP_003564916.1	2	*Brachypodium distachyon*	Bd
20	XM_010057819.2	XP_010056121.1	-	*Eucalyptus grandis*	Eg
21	XM_019567790.1	XP_019423335.1	LG18	*Lupinus angustifolius*	La
22	XM_015761161.1	XP_015616647.1	1	*Oryza sativa*	Os
23	XM_015314599.1	XP_015170085.1	-	*Solanum tuberosum*	St
24	XM_002458937.2	XP_002458982.1	3	*Sorghum bicolor*	Sb
25	KDO72339	A0A067FY71	-	*Citrus sinensis*	Cs
26	XM_008377177.2	XP_008375399.1	7	*Malus domestica*	Md
27	XM_020835210.1	XP_020690869.1	-	*Dendrobium catenatum*	Dc
28	XM_008811929.2	XP_008810151.1	-	*Phoenix dactylifera*	Pd
29	XM_004970987.3	XP_004971044.1	V	*Setaria italic*	Si
30	KYP54408	A0A151SI10	11	*Cajanus cajan*	Cc
31	XM_003534292.3	XP_003534340.2	9	*Glycine max*	Gm
32	XM_012227595.2	XP_012082985.1	-	*Jatropha curcas*	Jc
33	XM_002519890.2	XP_002519936.1	-	*Ricinus communis*	Rc
34	EOA39602	R0GUH7	-	*Capsella rubella*	Cr
35	CDP07217	A0A068UG44	-	*Coffea canephora*	Cca
36	CCA22889	F0WNJ8	-	*Albugo laibachii*	Al
37	AK357914.1	BAJ89128.1	-	*Hordeum vulgare*	Hv
38	XM_009117085.2	XP_009115333.1	A9	*Brassica rapa*	Br
39	MG009494	MG009494	-	**Cyamopsis tetragonoloba*	Ct

Appendix 3. Nucleotide/protein accession numbers of beta subunit of AP-4 complexes from various plant species.

Sr. No.	Nucleotide Accession Number	Protein Accession Number	Chromosome Number	Species Name	Abbreviation of Species Name
1	KRH62012	I1JUR7	4	*Glycine max*	Gm
2	ESW09162	V7AU47	9	*Phaseolus vulgaris*	Pv
3	XM_014645932.1	A0A1S3U5V2	5	*Vigna radiate*	Vr
4	XM_020360942.1	A0A151TDQ2	6	*Cajanus cajan*	Cc
5	KEH35972	A0A072V2B0	3	*Medicago truncatula*	Mt
6	XM_004501132.2	A0A1S2Y9Y8	Ca5	*Cicer arietinum*	Ca
7	OIV96551	A0A1J7G852	LG15	*Lupinus angustifolius*	La
8	CBI20897	D7SUE8	-	*Vitis vinifera*	Vv
9	KCW80232	A0A059CPU4	-	Eucalyptus *grandis*	Eg
10	CDP16153	A0A068V8A3	-	*Coffea canephora*	Cc
11	XM_016883343.1	A0A1U8NIE1	6	*Gossypium hirsutum*	Gh
12	XM_008468028.2	A0A1S3CQS8	-	*Cucumis melo*	Cm
13	XM_002516531.2	B9RS58	-	*Ricinus communis*	Rc
14	KDP34217	A0A067KDR3	-	*Jatropha curcas*	Jc
15	XM_016647702.1	A0A1S4CPM3	-	*Nicotiana tabacum*	Nt
16	XM_006356583.1	M1B5K7	-	*Solanum tuberosum*	St
17	EOA19965	R0H5G0	-	*Capsella rubella*	Cr
18	CAB87709	Q9LDK9	5	*Arabidopsis thaliana*	At
19	XM_009132960.2	M4CPG2	A3	*Brassica rapa*	Br
20	KMZ63439	A0A0K9P371	-	*Zostera marina*	Zma
21	EES01104	C5XED0	3	*Sorghum bicolor*	Sb
22	EEC71123	B8ABZ7	1	*Oryza sativa*	Os
23	KQK08816	A0A0Q3GBT1	2	*Brachypodium distachyon*	Bd
24	AQK96797	K7UZT9	8	*Zea mays*	Zm
25	EDQ64303	R7W430	-	*Physcomitrella patens*	Pp
26	XM_020297148.1	A0A0W8BVA3	-	*Aegilops tauschii*	Ata
27	KUF75757	A0A0W8BVA3	-	*Phytophthora nicotianae*	Pn

28	XM_008359196.2	XP_008357418.1	8	*Malus domestica*	Md
29	OAY62891	A0A199UDS2	-	*Ananas comosus*	Ac
30	XM_016701046.1	XP_016556532.1	-	*Capsicum annuum*	Ca
31	EEE93091	B9H9C3	LGVI	*Populus trichopora*	Pt
32	XM_008794122.2	XP_008792344.1	-	*Phoenix dactylifera*	Pd
33	XM_020818517.1	XP_020674176.1	-	*Dendrobium catenatum*	Dc
34	KDO79711	A0A067GWQ1	-	*Citrus sinensis*	Cs
35	MG009493	MG009493	-	**Cyamposis tetragonoloba*	Ct
36	KQL05972	K3XEI3	-	*Setaria italic*	Si

Appendix 4. Nucleotide/ protein accession numbers of mu subunit of AP-4 complexes from various plant species.

Sr. No.	Nucleotide Accession Number	Protein Accession Number	Chromosome Number	Species Name	Abbreviation of Species Name
1	XM_004500930.2	A0A1S2Y8D3	Ca5	*Cicer arietinum*	Ca
2	XM_003523734.3	C6TKW5	4	*Glycine max*	Gm
3	KEH36188	B7FLE7	3	*Medicago truncatula*	Mt
4	XM_014646780.1	A0A1S3U8C5	5	*Vigna radiata*	Vr
5	ESW13255	V7AVL6	8	*Phaseolus vulgaris*	Pv
6	EEE81613	B9GN34	LGII	*Populus trichocarpa*	Pt
7	XM_008468302.2	A0A1S3CRM0	-	*Cucumis melo*	Cm
8	XM_016713554.1	A0A1U8GCJ3	4	*Capsicum annuum*	Ca
9	XM_016593819.1	XP_016449305.1.	-		Nt
10	XM_016893398.1	XP_016748887.1.	-	*Gossypium hirsutum*	Gh
11	NM_001328593.1.	NP_001315522.1.	2		Zm
12	NM_118588.7	Q9SB50	4	*Arabidopsis thaliana*	At
13	KMZ68436	A0A0K9PHJ9	-	*Zostera marina*	Zma
14	XM_010109384.1	W9SC15	-	*Morus notabilis*	Mn
15	EEF52620.1	B9R7Z0	-	*Ricinus communis*	Rc
16	EMT19894	R7WGA3	-	*Aegilops tauschii*	Ata
17	KOO27030	A0A0M0JKN7	-	*Chrysochromulina sp*	C
18	AAN35786	Q8IIH2	11	*Plasmodium falciparum*	Pf
19	ACJ85581	B7FLE7	3	*Medicago truncatula*	Mt
20	KYP64615	A0A151TC62	7	*Cajanus cajan*	Cc
21	KUF65421	A0A0W8B0I0	-	*Phytophthora nicotianae*	Pn
22	CCA26603	F0WYP4	-	*Albugo laibachii*	Al

23	KQK15198	I1GS95	1	*Brachypodium distachyon*	Bd
24	CDP11497	A0A068UT51	-	*Coffea canephora*	Cca
25	XM_020244872.1	XP_020100461.1	LG12	*Ananas comosus*	Ac
26	XM_009139968.2	XP_009138216.1	A1	*Brassica rapa*	Br
27	XM_010065932.2	XP_010064234.1	-	*Eucalyptus grandis*	Eg
28	XM_020679122.1	XP_020534781.1	-	*Jatropha curcas*	Jc
29	XM_019560248.1	XP_019415793.1	LG14	*Lupinus angustifolius*	La
30	XM_006283664.1	XP_006283726.1	-	*Capsella rubella*	Cr
31	XM_008380356.2	XP_008378578.1	8	*Malus domestica*	Md
32	XM_010666421.2	A5BF65	18	*Vitis vinifera*	Vv
33	XM_006338006.2	XP_002461036.1	-	*Solanum tuberosum*	St
34	XM_002460991.2	XP_002461036.1	2	Sorghum bicolor	Sb
35	KDO81151	A0A067H0R9	-	*Citrus sinensis*	Cs
36	KQL26704	K3ZT90	II	*Setaria italica*	Si
37	XM_020819530.1	XP_020675189.1	-	*Dendrobium catenatum*	Dc
38	XM_008796943.2	XP_008795165.1	-	*Phoenix dactylifera*	Pd
39	MG009492	MG009492	-	**Cyamopsis tetragonoloba*	Ct

Appendix 5. Nucleotide/protein accession numbers of sigma subunit of AP-4 complexes from various plant species.

Sr. No.	Nucleotide Accession Number	Protein Accession Number	Chromosome Number	Species name	Abbreviation of Species Name
1	XM_003516265.3	C6SXR3	1	*Glycine max*	Gm
2	KYP61396	A0A151T2W6	8	*Cajanus Cajan*	Cc
3	XM_014650885.1	A0A1S3UK22	7	*Vigna radiate*	Vr
4	XM_016684280.1	A0A1U8E025	8	*Capsicum annuum*	Can
5	EEF48142	B9RKE4	-	*Ricinus communis*	Rc
6	XM_004503453.1	A0A1S2YE76	Ca6	*Cicer arietinum*	Ca
7	XM_010088605.1	W9QC48	-	*Morus notabilis*	Mn
8	XM_016834419.1	A0A1U8JMI4	-	*Gossypium hirsutum*	Gh
9	XM_008452582.1	A0A1S3BQ09	-	*Cucumis melo*	Cm
10	XM_016583199.1	A0A1S3XFM9	-	*Nicotiana tabacum*	Nt
11	AAC62137	O82201	2	*Arabidopsis thaliana*	At
12	EMT28245	M8BTB4	-	*Aegilops tauschii*	Ata
13	LFYR01000614.1	A0A0K9PXE7	-	*Zostera marina*	Zma
14	ACF85502	B4FTQ9	4	*Zea mays*	Zm
15	OAY69518	A0A199UXJ6	-	*Ananas comosus*	Ac
16	KOO27776	A0A0M0JMH0	-	*Chrysochromulina sp.*	Ce
17	KUF80718	A0A0W8C9E3	-	*Phytophthora nicotianae*	Pn
18	XM_008345782.1	XP_008344004.1	2	*Malus domestica*	Md
19	XM_009103888.2	XP_009102136.1	A7	*Brassica rapa*	Br
20	XM_015757331.1	XP_015612817.1	8	*Oryza sativa*	Os

21	XM_002444239.2	XP_002444284.2	7	*Sorghum bicolor*	Sb
22	XM_012221437.1	XP_012076827.1	-	*Eucalyptus grandis*	Eg
23	XM_010049345.2	XP_010047647.1	4	*Phaselous vulgaris*	Pv
24	XM_007152887-1	XP_007152949.1	-	*Medicago truncatula*	Mt
25	*XM_003625467.2*	*XP_003625515.1*	3	*Brachypodium distachyon*	Bd
26	XM_003574410.3	XP_003574458.1	LGVI	*Populus trichocarpa*	Pt
27	XM_002308250.2	XP_002308286.1	-	*Solanum tuberosum*	St
28	XM_006338131.2	XP_006338193.1	-	*Coffea canephora*	Cca
29	CDP03226	A0A068U426	4	*Citrus sinensis*	Cs
30	KDO55890	A0A067EPS3	-	*Phoenix dactylifera*	Pd
31	XM_008801538.1	XP_008799760.1	-	*Hordeum vulgare*	Hv
32	AK361712.1	BAJ92916.1	-	*Dendrobium catenatum*	Dc
33	XM_020833183.1	XP_020688842	VI	*Setaria italica*	Si
34	MG009491	MG009491	-	**Cyamopsis tetragonoloba*	Ct
35	KQL01640	K3YK35	-	*Albugo laibachii*	Al
36	CCA14478	F0W0A7	-	*Capsella rubella*	Cr
37	XM_006298599.11	XP_006298661.	-	*Physcomitrella patens*	Pp
38	XM_001785266.1	XP_001785318	11	*Vitis vinifera*	Vv

*Nucleotide and protein sequences submitted in NCBI database

The sequences of *Albugo laibachii*, *Chrysochromulina* sp, *Plasmodium falciparum* and *Phytophthora nicotianae* have been included for comparative studies.

Appendix 6. Distribution and characterization of SNPs in coding region of AP-4 complex genes of soybean

Gene Name	Chromosome position	Polymorphism	CDS position	Protein Position	Amino Acids	Codon	PSEP	Effect
AP4E	41433462	T/C	7	3	S/P	Tca/ Cca	-	-
AP4E	41433494	C/T	39	13	A/A	gcC/gcT	-	-
AP4E	41433497	C/A	42	14	S/R	agC/agA		
AP4E	41433776	T/C	321	107	A/A	gcT/gcC		
AP4E	41433912	G/A	457	153	G/S	Ggc/Agc	1629	Probably damaging
AP4E	41433921	G/A	466	156	A/T	Gcc/Acc	1629	Probably damaging
AP4E	41434013	C/T	558	186	V/V	gtC/ gtT		
AP4E	41434019	C/T	564	188	C/C	tgC/ tgT		
AP4E	41434069	C/T	614	205	A/V	gCc/gTc	1629	Probably damaging
AP4E	41434088	C/T	633	211	V/V	gtC/ gtT		
AP4E	41436042	A/G	1517	506	N/S	aAt/ aGt	593	Probably damaging
AP4E	41437822	A/T	1768	590	N/Y	Aat/Tat		
AP4E	41440754	T/G	2235	745	D/E	gaT/gaG		
AP4E	41440773	A/G	2254	752	T/A	Acc/Gcc		
AP4E	41440810	T/C	2291	764	V/A	gTt/gCt		
AP4E	41441490	C/G	2971	991	Q/A	Caa/Gaa	148	Probably bengin
AP4E	41441492	A/T	2973	991	Q/H	caA/caT	148	Probably bengin
AP4E	41441493	T/G	2974	992	S/A	Tct/Gct		
AP4E	41441502	G/T	2983	995	G/C	Ggt/Tgt	936	Probably damaging
AP4E	41441503	G/T	2984	995	G/V	gGt/gTt		
AP4B	6649984	T/C	6	2	A/A	gcT/gcC		
AP4B	6650902	G/A	202	68	E/K	Gag/Aag	936	Probably damaging
AP4B	6653988	T/A	999	333	R/R	cgT/cgA		
AP4B	6656412	C/T	1611	537	H/H	caC/caT		
AP4B	6657320	T/G	1917	639	N/K	aaT/aaG	148	Probably bengin
AP4B	6657456	T/G	2053	685	S/A	Tca/Gca	148	Probably bengin
AP4B	6658060	C/G	2279	760	S/W	tCg/tGg		
AP4B	6658196	A/G	2415	805	I/M	atA/atG		

AP4M	45139831	T/C	1148	383	L/P	cTa/cCa		
AP4M	45143810	G/A	393	131	L/L	ttG/ttA		
AP4M	45144390	C/T	232	78	P/S	Cct/Tct	1639	Probably damaging

Appendix 7. Distribution and characterization of SNPs in coding region of AP-4 complex genes of *Z. mays.*

Gene Name	Chromosome Position	Polymorphism	CDS Position	Protein Position	Amino Acids	Codon	PSEP	Effect
AP4E	163507467	G/A	33	11	L/L	ctG/ctA	-	
AP4E	163507530	C/A	96	32	I/I	atC/atA	-	
AP4E	163507554	G/A	120	40	E/E	gaG/gaA	-	
AP4E	163507860	G/A	426	142	A/A	ctG/ctA	-	
AP4E	163507868	C/T	435	145	A/V	gcC/gcT	936	Probably damaging
AP4E	163507899	C/T	465	155	I/I	atC/atT	-	
AP4E	163507917	G/A	483	161	Q/Q	caG/caA	-	
AP4E	163507944	G/A	510	170	K/K	aaG/aaA	-	
AP4E	163508643	G/T	637	213	A/S	Gca/Tca	1629	Probably damaging
AP4E	163508723	C/T	717	239	I/I	atC/atT	-	
AP4E	163508765	T/C	759	253	D/D	gaT/gaC	-	
AP4E	163508768	C/T	762	254	Y/Y	taC/taT	-	
AP4E	163508780	C/T	774	258	P/P	ccC/ccT	-	
AP4E	163508790	A/T	784	262	I/F	Att/Ttt	936	Probably damaging
AP4E	163508953	A/G	816	272	V/V	gtA/gtG	-	
AP4E	163509081	T/C	944	315	F/S	Tca/Cca		
AP4E	163509319	A/G	1095	365	L/L	ctA/ctG	-	
AP4E	163509335	C/T	1111	371	L/L	Cta/Tta	-	
AP4E	163509705	A/G	1212	404	E/E	gaA/gaG	-	
AP4E	163509716	A/G	1222	408	N/S	aAc/aGc	148	Probably begign
AP4E	163509738	G/A	1245	415	K/K	aaG/aaA	-	
AP4E	163509923	C/T	1341	447	A/A	gcC/gcT	-	
AP4E	163509957	C/T	1375	459	L/L	Ctg/Ttg	-	
AP4E	163509986	A/C	1404	468	G/G	ggA/ggC	-	
AP4E	163510001	T/C	1419	473	G/G	ggT/ggC	-	
AP4E	163510008	A/T	1426	476	S/C	Agt/Tgt	593	Probably damaging
AP4E	163510485	G/C	1449	483	V/V	gtG/gtC	-	
AP4E	163510506	C/T	1470	490	V/V	gtC/gtT		

AP4E	163510686	G/A	1575	525	K/K	aaG/aaA		
AP4E	163511832	A/G	1623	541	A/A	gcA/gcG	-	
AP4E	163511838	G/A	1629	543	A/A	gcG/gcA	-	
AP4E	163511865	G/A	1656	552	A/A	gcG/gcA	-	
AP4E	163511882	G/A	1674	558	G/E	gGa/gAa	936	Probably damaging
AP4E	163511904	T/A	1695	565	P/P	ccT/ccA	-	
AP4E	163513094	G/A	1737	579	S/S	tcG/tcA	-	
AP4E	163513097	G/A	1740	580	T/T	acG/acA	-	
AP4E	163513172	C/T	1815	605	P/P	ccC/ccT	-	
AP4E	163513193	C/T	1836	612	D/D	gaC/gaT	-	
AP4E	163514834	G/T	1920	640	E/D	gaG/gaT	593	Probably damaging
AP4E	163514861	C/T/G	1947	649	V/V	gtC/gtT	-	
AP4E	163514990	T/C	2076	692	D/D	gaT/gaC	-	
AP4E	163515146	C/T	2132	711	S/F	tCt/tTt	148	Probably begign
AP4E	163515149	C/T	2235	745	P/P	ccC/ccT		
AP4E	163515177	C/A	2263	755	P/T	Ccc/Acc		
AP4E	163515184	G/A	2270	757	G/E	gGg/gAg	148	Probably begign
AP4E	163515203	G/C	2289	763	G/G	ggG/ggC		
AP4E	163515227	G/A	2313	771	R/R	CgG/cgA		
AP4E	163515266	A/G	2352	784	E/E	gaA/gaG		
AP4E	163515386	G/A	2472	824	V/V	gtG/gtA		
AP4E	163515398	T/C	2484	828	N/N	aaT/aaC		
AP4E	163515426	G/C	2512	838	V/L	Gta/Cta	148	Probably begign
AP4E	163515503	C/T	2589	863	D/D	gaC/gaT		
AP4E	163515516	C/T	2602	868	L/L	Cta/Tta		
AP4E	163515550	C/G	2636	879	P/R	cCg/cGg	148	Probably begign
AP4E	163515553	T/C	2639	880	V/A	gTg/gCg	593	Probably damaging
AP4E	163515600	C/T	2688	896	L/L	cCt/cTt		
AP4E	163515658	C/T	2744	915	T/I	aCt/aTt		
AP4E	163515688	G/A	2774	925	R/H	cGc/cAc		
AP4E	163515715	T/A	2801	934	M/K	aTg/aAg		
AP4E	163515738	G/A	2824	942	A/T	Gcc/Acc	593	Probably damaging
AP4E	163515756	G/A	2842	948	A/T	Gct/Act	148	Probably begign
AP4E	163515787	T/A	2873	958	V/E	gTg/gAg	-	
AP4M	26730633	C/A	146	49	P/P	ccC/ccA	-	
AP4M	26731801	C/T	393	131	V/V	gtC/gtT	-	
AP4M	26731884	T/C	473	158	M/T	aTg/aCg		
AP4B	149794314	C/T	99	33	V/V	gtC/gtT	-	
AP4B	149795204	T/C	268	90	S/P	Tca/Cca	148	probably begign
AP4B	149795384	A/G	348	116	L/L	ctA/ctG	-	
AP4B	149795489	G/A	453	151	R/R	agG/agA	-	
AP4B	149795660	G/T	513	171	V/V	gtG/gtT	-	
AP4B	149795708	A/T	561	187	A/A	gcA/gcT	-	
AP4B	149795714	G/C	567	189	V/V	gtG/gtC	-	

AP4B	149795726	C/T	579	193	S/S	agC/agT	-	
AP4B	149795729	C/T	582	194	S/S	agC/agT	-	
AP4B	149796771	G/A	789	263	G/G	ggG/ggA	-	
AP4B	149796773	A/C	791	264	D/A	gAt/gCt		
AP4B	149796785	C/A	803	268	A/D	gCt/gAt	936	probably damaging
AP4B	149796932	C/T	855	285	Y/Y	taC/taT		
AP4B	149797010	C/T	933	311	D/D	gaC/gaT		
AP4B	149797046	T/C	969	323	D/D	gaT/gaC		
AP4B	149797059	C/A	982	328	L/I	Cta/Ata	1629	probably damaging
AP4B	149797062	T/A	985	329	S/T	Tca/Aca	1628	probably damaging
AP4B	149799125	G/A	1008	336	S/S	tcG/tcA	-	
AP4B	149799137	A/G	1020	340	T/T	acA/acG	-	
AP4B	149799152	C/A	345	135	R/R	cgC/cgA	-	
AP4B	149799239	T/C	1122	374	D/D	gaT/gaC	-	
AP4B	149799242	T/C	1125	375	N/N	aaT/aaC	-	
AP4B	149799323	T/C	1206	204	A/A	gcT/gcC	-	
AP4B	149799329	C/G	1212	404	S/S	tcC/tcG	-	
AP4B	149799341	T/C	1224	408	S/S	agT/agC	-	
AP4B	149799363	G/A	1246	416	A/T	Gca/Aca	-	
AP4B	149799384	G/A	1267	423	V/I	Gta/Ata	-	
AP4B	149799475	A/C	1358	453	Q/P	cAa/cCa	-	
AP4B	149799480	A/G	1363	455	T/A	Aca/Gca	-	
AP4B	149799483	C/T	1366	456	L/L	Ctg/Ttg	-	
AP4B	149799486	G/A	1369	457	A/T	Gcc/Acc	-	
AP4B	149799552	G/A	1435	479	A/T	Gca/Aca	-	
AP4B	149799557	G/A	1440	480	L/L	ctG/ctA	-	
AP4B	149799968	T/C	1452	484	C/C	tgT/tgC	-	
AP4B	149799986	G/A	1470	490	G/G	ggG/ggA	-	
AP4B	149800007	C/A	1491	497	P/P	ccC/ccA	-	
AP4B	149800019	C/T	1503	501	I/I	atC/atT	-	
AP4B	149800052	T/C	1536	512	I/I	atT/atC	-	
AP4B	149800061	C/T	1545	515	G/G	ggC/ggT	-	
AP4B	149800100	A/G	1584	528	Q/Q	caA/caG		
AP4B	149800101	A/C	1585	529	K/Q	Aaa/Caa	148	Probably begign
AP4B	149800113	G/A	1597	533	A/T	Gct/Act	-	
AP4B	149800128	G/A	1612	538	V/I	Gtt/Att	148	Probably begin
AP4B	149800190	A/G	1674	558	G/G	ggA/ggG	-	
AP4B	149800192	C/A	1676	559	T/N	aCt/aAt	148	Probably begin
AP4B	149800216	T/C	1700	567	L/S	tTg/TCg	148	Probably begin
AP4B	149800235	C/G	1719	573	S/S	tcC/tcG	-	

AP4S	76976880	G/A	57	19	Q/Q	caG/caA	-	-
AP4S	76977218	A/G	170	57	K/R	aAg/aGg	1629	Probably damaging
AP4S	76977231	G/A	183	61	R/R	agG/agA	-	-
AP4S	76977234	C/T	186	62	R/R	cgC/cgT	-	-
AP4S	76985173	C/T	306	102	L/L	ctC/ctT	-	-
AP4S	76985210	T/C	343	131	N/N	aaC/aaT	-	-
AP4S	76985287	G/A	420	140	E/E	gaG/gaA	-	-
AP4S	76985296	C/A	429	143	S/S	tcC/tcA	-	-

PSEP is Position- specific evolutinary preservation