ALLELE MINING FOR HEAT STRESS RELATED STARCH SYNTHASE GENE IN CULTIVATED AND WILD SPECIES OF WHEAT

Thesis

Submitted to the Punjab Agricultural University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE
in
BIOTECHNOLOGY
(Minor Subject: Plant Breeding and Genetics)

By

Mehakdeep Singh Dhawan
(L-2014-A-168-M)

School of Agricultural Biotechnology
College of Agriculture

LUDHIANA-141 004

2017

CERTIFICATE I

This is to certify that the thesis entitled. **"ALLELE MINING FOR HEAT STRESS RELATED STARCH SYNTHASE GENE IN CULTIVATED AND WILD SPECIES OF WHEAT"** submitted for the degree of **Master of Science** in **Biotechnology** (Minor subject: **Plant Breeding and Genetics**) of the Punjab Agricultural University, Ludhiana, is a bonafide research work carried out by **Mr. Mehakdeep Singh Dhawan (L-2014-A-168-M)** under my supervision and that no part of this thesis has been submitted for any other degree.

The assistance and help received during the course of investigation have been fully acknowledged.

(Dr. Satinder Kaur)
Major advisor
Asst. Molecular Biologist
School of Agricultural Biotechnology
Punjab Agricultural University
Ludhiana-141004

CERTIFICATE II

This is to certify that the thesis entitled, **"ALLELE MINING FOR HEAT STRESS RELATED STARCH SYNTHASE GENE IN CULTIVATED AND WILD SPECIES OF WHEAT"** submitted by **Mehakdeep Singh Dhawan (L-2014-A-168-M)** to the Punjab Agricultural University, Ludhiana, in partial fulfillment of the requirements for the degree of **Master of Science** in the subject of **Biotechnology** (Minor subject: **Plant Breeding and Genetics**) has been approved by the Student's Advisory Committee along with External Examiner after an oral examination on the same.

(Dr. Satinder Kaur)
Major Advisor

(Dr. Prabhjeet Singh)
External Examiner
Professor
Department of Biotechnology
Guru Nanak Dev University,
Amritsar 143005 (Punjab)

(Dr. Parveen Chhuneja)
Head of the Department

(Dr. Neelam Grewal)
Dean Postgraduate Studies

ACKNOWLEDGEMENTS

Fore mostly, I would like to express my thanks and gratitude to the Almighty lord for bestowing me with the wisdom and knowledge and keeping me steady to make this endeavour possible. This thesis owes its existence to the assistance, guidance and contribution in assorted ways by several people. I would therefore like to offer my sincere thanks to all of them.

At this moment of accomplishment, I would like to express my special appreciation and thanks to my major advisor, Dr. Satinder Kaur, Asst. Molecular Biologist, School of Agricultural Biotechnology, for being a tremendous mentor for me. I would like to thank her for encouraging my research with unflinching and insightful comments when my steps faltered. Her patience and support helped me to overcome many crisis situations. I hope that I am always hovered with her guidance and blessings throughout my life.

I am also indebted to the members of my advisory committee, Dr. Priti Sharma (Assistant Professor, School of Agricultural Biotechnology), Dr. Achla Sharma (Assistant Plant Breeder, Department of Plant Breeding and Genetics) and Dr. Parveen Chhuneja (Nominee Dean PGS and Director, School of Agricultural Biotechnology) for their valuable advice, constructive criticism, extensive discussions around my work and meticulously scrutinizing this manuscript. Particular thanks goes to Dr. Kuldeep Singh (Director NBPGR, New Delhi) and Mr. Inderjit Singh Yadav (Asst. Professor of Bioinformatics) for their generous support and guidance during my research programme.

My special thanks is extended to my Parents Sh. H. S. Dhawan and Smt. Charanjit Kaur, Sisters Tarandeep and Jasmine, Brother-in-laws Arjun and Tanvir and Nephews Nadish and Gurseerat for their support and encouragement. I convey special acknowledgement to Shivani and Jaskaran for their advice and willingness to share their bright thoughts with me, which were very fruitful for shaping up my ideas and research. I expand my thanks to all the field workers and lab attendants Komal and Basant for their valuable contribution in completion of my work.

The warmest feelings are exclusively reserved for my dearest friends, Bhupinder, Hari, Arshpreet, Sumedha, Harsimer and Harman for their moral support, valuable suggestions and encouragement.

All may not be mentioned, but none is forgotten.

Dated:

Place: **Mehakdeep Singh Dhawan**

Title of the Thesis	**:**	Allele mining for heat stress related starch synthase gene in cultivated and wild species of wheat
Name of the student and Admission No.	**:**	Mehakdeep Singh Dhawan L-2014-A-168-M
Major Subject	**:**	Biotechnology
Minor Subject	**:**	Plant Breeding and Genetics
Name of Major Advisor and Designation	**:**	Dr. Satinder Kaur Asst. Molecular Biologist
Degree to be Awarded	**:**	M.Sc. (Biotechnology)
Year of Award of Degree	**:**	2017
Total Pages in Thesis	**:**	72 + VITA
Name of University	**:**	Punjab Agricultural University, Ludhiana-141 004, Punjab (India)

ABSTRACT

The present study was carried out to identify allelic variants in soluble starch synthase I (*SSI*) gene in 20 wheat genotypes including nine wild wheats, one durum wheat and ten hexaploid wheats. Candidate *SSI* gene located on wheat chromosome 7 (7A, 7B, 7D) was selected. After primer designing, *SSI* gene was amplified, cloned and sequenced in these 20 wheat genotypes though sequences of three wheat genotypes could not be obtained. After manual curation of raw data of these sequences, contigs were assembled and 9 exons and 8 introns were predicted like the reference *SSI* gene. There were 286 intronic SNPs and 45 exonic SNPs contributing to 221 transitions and 110 transversions. First exon has 18 SNPs while 3-6 SNPs were detected in rest of the exons. No exonic SNP was detected in wild genotype *Triticum dicoccoides acc. pau14801* and four cultivated genotypes Giza, Arbon, C306 and PBW343. Out of 45 exonic SNPs, 12 contributed towards non-synonymous amino acid substitutions and none of these substitutions lies in active sites of *SSI* protein. Also important catalytic domain of *SSI* gene glycosyltransferase 5 (GT-5) lies outside predicted eight active sites of protein. GT-5 domain lies from exons 1- 6 and six SNPs corresponding to non-synonymous substitutions falls in this region. These six SNP based alleles from four wild progenitor species of *Aegilops tauschii acc. pau14102, Aegilops tauschii acc. pau3747, Aegilops speltoides acc. pau15081* and *Triticum dicoccoides acc. pau7107* and two cultivated genotypes Impala and C591 could be candidate SNPs for improved heat stress tolerance.

Keywords: Starch, Wheat, Starch synthase, PCR, Cloning, Sequencing, SNP, Allelic variation

Signature of Major Advisor **Signature of the Student**

ਖੋਜ ਗ੍ਰੰਥ ਦਾ ਸਿਰਲੇਖ	:	ਕਣਕ ਦੀ ਕਾਸ਼ਤ ਅਤੇ ਜੰਗਲੀ ਸਪੀਸੀਜ਼ ਵਿੱਚ ਗਰਮੀ ਤਣਾਅ ਸਬੰਧਤ ਸਟਾਰਚ ਸਿੰਥੇਜ਼ ਜੀਨ ਦੀ ਐਲੀਲ ਮਾਈਨਿੰਗ
ਵਿਦਿਆਰਥੀ ਦਾ ਨਾਮ ਅਤੇ	:	ਮਹਿਕਦੀਪ ਸਿੰਘ ਧਵਨ
ਦਾਖਲਾ ਕ੍ਰਮਾਂਕ		ਐੱਲ-2014-ਏ-168-ਐਮ
ਪ੍ਰਮੁੱਖ ਵਿਸ਼ਾ	:	ਬਾਇਓਟੈਕਨੋਲੋਜੀ
ਸਹਿਯੋਗੀ ਵਿਸ਼ਾ	:	ਪਲਾਂਟ ਬਰੀਡਿੰਗ ਅਤੇ ਜੈਨੇਟਿਕਸ
ਮੁੱਖ ਸਲਾਹਕਾਰ ਦਾ ਨਾਮ ਅਤੇ ਅਹੁਦਾ	:	ਡਾ. ਸਤਿੰਦਰ ਕੌਰ ਅਸਸਿਸਟੈਂਟ ਮੋਲੇਕੁਲਰ ਬਿਓਲੋਜਿਸਟ
ਡਿਗਰੀ	:	ਐਮ. ਐਸ. ਸੀ
ਡਿਗਰੀ ਨਾਲ ਸਨਮਾਨਿਤ ਕਰਨ ਦਾ ਸਾਲ	:	2017
ਖੋਜ ਖੇਤਰ ਵਿਚ ਕੁਲ ਪੰਨੇ	:	72 + ਵੀਟਾ
ਯੂਨੀਵਰਸਿਟੀ ਦਾ ਨਾਮ	:	ਪੰਜਾਬ ਖੇਤੀਬਾੜੀ ਯੂਨੀਵਰਸਿਟੀ, ਲੁਧਿਆਣਾ -141004 ਪੰਜਾਬ, ਭਾਰਤ

ਨਿਚੋੜ

ਮੌਜੂਦਾ ਅਧਿਐਨ ਵਿਚ 20 ਕਣਕ ਜੀਨੋਟਾਈਪਸ ਜਿਸ ਵਿਚ ਨੈ ਜੰਗਲੀ ਕਣਕ, ਇੱਕ ਦੁਰੁਮ ਕਣਕ ਅਤੇ ਦਸ ਹੇਕਸਾਪਲੋਈਡ ਕਣਕ ਸ਼ਾਮਲ ਸਨ ਜਿਸ ਵਿੱਚ ਘੁਲ ਸਟਾਰਚ ਸਿੰਥੇਜ਼ । (*SSI*) ਜੀਨ ਦੀ ਐਲੈਲਿਕ ਰੂਪ ਦੀ ਪਛਾਣ ਕਰਨ ਕੀਤਾ ਗਿਆ । ਉਮੀਦਵਾਰ *SSI* ਕਣਕ ਗੁਣਸੂਤਰ 7 'ਤੇ ਸਥਿਤ ਜੀਨ (7 ਏ, 7 ਬੀ, 7 ਡੀ) ਚੁਣਿਆ ਗਿਆ ਹੈ । ਪਰਾਈਮਰ ਡਿਜ਼ਾਈਨਿੰਗ ਤੋਂ ਬਾਅਦ, *SSI* ਜੀਨ ਦੀ ਅਮਪਲੀਫਿਕੇਸ਼ਨ, ਕਲੋਨਿੰਗ ਅਤੇ ਸੇਕੁਐਨਸਿੰਗ 20 ਕਣਕ ਜੀਨੋਟਾਈਪਸ ਵਿਚ ਕੀਤੀ ਗਈ ਪਰ ਤਿੰਨ ਕਣਕ ਜੀਨੋਟਾਈਪਸ ਵਿਚ ਪ੍ਰਾਪਤ ਨਹੀਂ ਕੀਤਾ ਗਿਆ । ਇਹ ਕ੍ਰਮ ਦੇ ਕੱਚੇ ਡਾਟਾ ਦੇ ਦਸਤਾਵੇਜ਼ ਕਿਉਰੇਸ਼ਨ ਬਾਅਦ, ਕੈਂਟੀਗਸ ਬਣਾਏ ਗਏ ਅਤੇ 9 ਅਘਜ਼ੋਨਜ਼ ਅਤੇ 8 ਇੰਟਰੰਨਜ਼ ਦੀ ਭਵਿੱਖਬਾਣੀ ਕੀਤੀ ਗਈ ਜੋ ਕਿ ਰੈਫਰੈਂਸ *SSI* ਜੀਨ ਨਾਲ ਮਿਲਦੀ ਸੀ । ਕੁਲ 286 ਇੰਟਰੰਨਿਕ SNPs ਅਤੇ 45 ਅਘਜ਼ੋਨਿੱਕ SNPs ਪ੍ਰਾਪਤ ਹੋਏ ਜਿਸ ਵਿੱਚ 221 ਟ੍ਰਾਂਜ਼ਿਸ਼ਨਜ਼ ਅਤੇ 110 ਟ੍ਰਾਂਸਵਰਸਿਨਜ਼ ਸਨ । ਪਹਿਲੇ ਅਘਜ਼ੋਨ ਵਿੱਚ 18 SNPs ਜਦਕਿ 3-6 SNPs ਬਾਕੀ ਦੇ ਅਘਜ਼ੋਨਜ਼ ਵਿੱਚ ਖੋਜੇ ਗਏ ਸੀ । ਕੋਈ ਵੀ ਅਘਜ਼ੋਨਿੱਕ SNP ਜੰਗਲੀ ਜੀਨੋਟਾਈਪ *T. dicoccoides acc. pau14801* ਅਤੇ ਚਾਰ ਕਾਸ਼ਤ ਜੀਨੋਟਾਈਪਸ ਗਿਜ਼ਾ, ਅਰਬੋਨ, C306 ਅਤੇ PBW343 ਵਿੱਚ ਨਹੀਂ ਖੋਜਿਆ ਗਿਆ ਸੀ । 45 ਅਘਜ਼ੋਨਿੱਕ SNPs ਦੇ ਵਿੱਚੋ, 12 ਨੈਨ-ਸਿਨੋਨੀਮਸ ਅਮੀਨੋ ਐਸਿਡ ਪ੍ਰਤੀਸਥਾਪਨ ਮਿਲੇ ਅਤੇ ਇਹ ਪ੍ਰਤੀਸਥਾਪਨ ਕੋਈ ਵੀ *SSI* ਪ੍ਰੋਟੀਨ ਦੇ ਸਰਗਰਮ ਸਾਈਟ ਵਿੱਚ ਨਹੀਂ ਪਿਆ ਗਿਆ । ਇਸ ਦੇ ਨਾਲ *SSI* ਜੀਨ ਵਿਚ ਗਲਾਈਕੋਟਰਾਂਸਫੇਰਸ 5 (ਜੀ.ਟੀ.-5) ਮਹੱਤਵਪੂਰਨ ਕੇਟਾਲੀਟਿਕ ਡੋਮੇਨ ਪਾਇਆ ਗਿਆ ਜੋ ਕਿ ਪ੍ਰੋਟੀਨ ਦੇ ਅੱਠ ਸਰਗਰਮ ਸਾਈਟ ਦੀ ਭਵਿੱਖਬਾਣੀ ਤੋਂ ਬਾਹਰ ਸੀ । ਜੀ.ਟੀ.-5 ਡੋਮੇਨ ਵਿਚ 1- 6 ਅਘਜ਼ੋਨਜ਼ ਮੌਜੂਦ ਸਨ ਅਤੇ ਛੇ ਨੈਨ-ਸਿਨੋਨੀਮਸ ਪ੍ਰਤੀਸਥਾਪਨ ਇਸ ਖੇਤਰ ਵਿੱਚ ਡਿੱਗ ਕਰਨ ਲਈ ਅਨੁਸਾਰੀ ਤੱਕ ਹੈ । ਚਾਰ ਜੰਗਲੀ ਸਪੀਸੀਜ਼ *Ae. tauschii acc. pau14102*, *Ae. tauschii acc. pau3747*, *Ae. speltoides acc. pau15081* ਅਤੇ *T. dicoccoides acc. pau7107* ਅਤੇ ਦੇ ਕਾਸ਼ਤ ਜੀਨੋਟਾਈਪਸ ਇਮਪਾਲਾ ਅਤੇ C591 ਦੇ ਛੇ SNPs ਅਧਾਰਿਤ ਐਲੀਲਜ਼ ਗਰਮੀ ਤਣਾਅ ਸਹਿਣਸ਼ੀਲਤਾ ਸੁਧਾਰ ਲਈ SNPs ਉਮੀਦਵਾਰ ਹੋ ਸਕਦਾ ਹੈ ।

ਮੁੱਖ ਸ਼ਬਦ: ਸਟਾਰਚ, ਕਣਕ, ਸਟਾਰਚ ਸਿੰਥੇਜ਼, PCR, ਕਲੋਨਿੰਗ, ਸੇਕੁਐਨਸਿੰਗ, SNP, ਅਲੈਲਿਕ ਪਰਿਵਰਤਨ

_____ _____

ਮੁੱਖ ਸਲਾਹਕਾਰ ਦੇ ਹਸਤਾਖਰ **ਵਿਦਿਆਰਥੀ ਦੇ ਹਸਤਾਖਰ**

CONTENTS

CHAPTER I

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) ranks second in primary cereal crops of the temperate region. For 40% of the world's population, it serves as a staple food (FAO) with three sub genomes A, B and D with genome size of 17,000 Mb (Devos and Gale 2000). It constitutes one of the largest genome of cereals. From two successive interspecific hybridizations and polyploidizations, the genome architecture of wheat was originated. Around 0.3–0.5 million years ago, the first hybridization occurred between progenitor *Triticum urartu* (A genome AuAu, $2n = 2x = 14$) and *Aegilops speltoides* Tausch. (SS, $2n = 2x = 14$) which is considered similar to B genome progenitor species, that corresponds to evolution of wild emmer wheat *T. turgidum ssp. dicoccoides* (AuAuBB, $2n = 4x = 28$) (Maestra and Naranjo 1998). Human selection led to the production of cultivated emmer wheat *T. turgidum ssp. Dicoccon*, that hybridized spontaneously around 9,000 years ago with *Ae. tauschii* (DD, $2n = 2x = 14$) to produce allohexaploid wheat, *T. aestivum* L. ($2n = 6x = 42$, AuAuBBDD) [Dvorak *et al* 1998]. It is estimated before 2020, due to increasing world population the global demand for wheat will increase further by 40% therefore for wheat breeding programs, higher yield becomes predominating objective.

India ranks second in the world for production of wheat (Limbore and Khillare 2015). Starch is major determinant of wheat grain yield, which consists of 65–75% grain dry weight and 80% of the endosperm dry weight (Rahman *et al* 2000). The synthesis of starch is particularly sensitive to high temperature. Processing, cooking, organoleptic qualities and the digestibility of starch-based food products, all food qualities are affected by starch (Lafiandra e*t al* 2010). Primary product of photosynthesis in leaves is starch, which is transiently stored in the chloroplast during day for its subsequent use during the night. At various levels photosynthetic partitioning into starch is regulated to control the flux of carbon from the Calvin cycle into starch biosynthesis.

Amylose and amylopectin are two major components of wheat starch. Amylopectin constitutes 75% of the dry weight of starch, while amylose constituting 25% and both the molecules are linked by α-(1,4) glycosidic linkages. Amylopectin accounts for molecular weight of 10^8 and is highly branched with α-(1,6) linkages, while amylose is shorter unbranched molecule with a molecular weight of 10^6 (Ball *et al* 1996). Formation of ADPglucose, a glucose donor, catalyzed by ADP-glucose pyrophosphorylase (AGPase) is the first step for the starch biosynthesis in green plants (James *et al* 2003). For formation of starch molecules ADP-glucose is used by starch synthase enzyme which add glucose units to the growing polymer chain (Buléon *et al* 1998).

The starch biosynthesis depends upon several enzymes involving different isoforms

of soluble starch synthase (*SS*), starch branching enzyme (*SBE*), starch debranching enzyme (*SDE*) - isoamylases, pullulanase, starch phosphorylase (*SP*) and disproportionating enzyme (*DPE*) [Keeling and Myers 2010]. Among the enzymes involved in the starch biosynthetic pathway in wheat endosperms, soluble starch synthase (*SS*) is the most sensitive to high temperature (Keeling *et al* 1993), and has low optimum temperature (20-25°C) for maximum activity (Denyer *et al* 1994). Starch synthases exhibited two distinct isoforms i.e. granule bound starch synthase (*GBSS*) and soluble starch synthases *(SSI, SSII, SSIII* and *SSIV*), which converts ADP-glucose to the starch polymers. Bound to the starch granule, the *GBSS* isoform is responsible for amylose production. *GBSS* in combination with ADP-Glc synthesizes linear amylose fraction of starch that diffuses into the granule matrix (Tatge *et al* 1999) and lead to the synthesis of amylopectin chains (Fulton *et al* 2002). The generation of thousand unbranched chains of glucose residues in length was made through *GBSS* processive mode of action (Denyer *et al* 1999). For formation of amylopectin, *SS* and *SBE* genes are involved. Amylopectin chains elongation are initiated by *SS* isoforms *SSI, SSII* and *SSIII* that are localized in the plastid stroma and a fraction of the protein is bound to the granule.

Synthesis of the shortest chains of amylopectin about 10 glucosyl units or less degree of polymerization (DP 7-11) is mainly due to *SSI*. Synthesis of intermediate chains is predominately due to *SSII* and further extensions to produce longer chains are catalysed by *SSIIa* and/or *SSIII* (Commuri and Keeling 2001). *SSIIa* has major affect on starch quality as it is predominantly expressed in cereal endosperm that affects amylopectin structure (Morell *et al* 2003). *SSIIb* expression is low in leaves at initial stage of grain filling (Hirose and Terao 2004). The expression of *SSIIIa* and *SSIIIb* genes are dominant in endosperm, but its transient expression in leaves were also reported (Dian *et al* 2005; Hirose *et al* 2006). Therefore, *SSI* preferentially elongates the shortest, *SSII* and *SSIII* preferentially elongates medium-length and longer-length chains respectively (Tomlinson and Denyer 2003). *SSI, SSIIa*, and *SSIII* in maize (Gao *et al* 1998), *SSI, SSII*, and *SSIII* in wheat (Li *et al* 1999a) and *SSI* and *SSIIa* in barley (Morell *et al* 2003) are early published starch synthase genes in cereals which are homologous to genes encoding *SSI, SSIIa*, and *SSIIIa* in rice (Hirose and Terao 2004).

Increasing temperature associated with climate change could negatively affect global wheat yields resulting in potential increase in food insecurity and poverty (Tubiello *et al* 2000). As a result, the frequency of extreme events such as the number of hot days is also anticipated to increase (Pittock 2003). As a consequence of climate change in the North Western Plains Zone (NWPZ), 51% of its area might be reclassified as the heat stress prone to mega-environment 1 (ME 1) by 2050 instead of favourable and high yield potential ME 1, which represents 15% of global wheat production (Ortiz *et al* 2008). Over the last 30 years, damages caused by heat to wheat crops decreased worldwide yield by 5.5%, translating to total loss of 35 M tonnes (Lobell *et al* 2011). Heat stress is having an increasingly negative

impact on spike fertility and the onset of senescence (Lobell *et al* 2012). Yield penalty is associated with long periods of high temperatures as well as short heat waves, when temperature rises above 32°C during heading and grain filling stages (Wardlaw *et al* 1994).

Numerous studies involving maize (Keeling *et al* 1994), wheat (Bhullar and Jenner 1985) and barley (MacLeod and Duffus 1988) have reported down regulation of starch deposition in the kernel due to heat stress. This may be due to impairment of conversion of sucrose to starch at high temperature and limits starch synthesis (Bhullar and Jenner 1985). Heat stress also results in down regulation of starch synthesis genes in the grain, and *SS* genes were found to be the most affected (Hurkman *et al* 2003). Exposure to high temperatures can cause considerable morpho-physiological damage, which hastens leaf senescence, reduces photosynthesis, pollen viability and starch biosynthesis. Starch accumulation and transcript profiles for the starch biosynthetic enzymes are greatly affected by high temperatures. With decline in grain-filling, the time to maximum kernel water content, kernel dry weight and harvest maturity also gets affected demonstrating that high temperature significantly affects the overall development of grain.

Peng *et al* (2001) reported that the *SSI* gene encodes 75 kDa protein which is located on chromosome 7S of wheat, its distribution is between the starch granule and the soluble phase. Full length starch synthase-I cDNA of 2591 bp (Li *et al* 1999a) and 2575 bp (Peng *et al* 2001) have been cloned, both of which encode 647 amino-acid polypeptide residues identical in sequence except for single amino acid difference at position 321, where leucine is replaced by valine. Li *et al* (2013) investigated partial sequences of *SSI* from *Aegilops tauschii*, *Triticum turgidum*, *Triticum aestivum* and synthetic hexaploid wheat for evaluating the relationship among starch synthase genes of different species.

Allele mining is a technique for identification of natural allelic variants at candidate loci that has utility in improvement of crops for controlling key agronomic traits. This approach focuses on finding mutation - single nucleotide polymorphisms (SNPs) or insertions:deletions (InDels) which could explain the phenotypic differences arising between two genotypes with different alleles for the desired trait. Specified positions of a chromosomal site at which the DNA sequence of two genotypes differ by a single base is defined as SNP and set of SNPs that pertaining to a particular trait is referred as haplotype. SNPs are genetic markers that are identified from sequence data and have numerous applications such as characterizing allelic variation, genome-wide mapping and marker-assisted selection. Through sequencing SNP discovery uncover variation in agriculturally-relevant populations (Robbins *et al* 2011, Shirasawa *et al* 2010). SNP detection is accommodated through high-throughput genotyping which provide tool for the identification of sequence variation.

The present work was undertaken to identify sequence variation in starch synthase

gene in diploid and tetraploid wild progenitors and in tetraploid and hexaploid cultivated wheat to identify the useful alleles for heat stress tolerance with following objectives:

1. Identification of candidates for starch synthase gene using bioinformatics tools.
2. Primer designing and cloning of starch synthase gene.
3. Identification of novel haplotypes for starch synthase gene in selected wheat germplasm.

CHAPTER II

REVIEW OF LITERATURE

Hexaploid or common wheat (*Triticum aestivum* L. 2n=42, AABBDD) is a globally important food crop. Common wheat has evolved from two spontaneous hybridization events. Tetraploid *T. turgidum* L. wheat (2n=28, AABB) hybridized with diploid *Aegilops tauschii* Coss. (2n=14, DD) followed by spontaneous chromosome doubling is the widely accepted pathway for the origin of common wheat (Feldman 2001). Starch ranks second biopolymer of the world, due to which it substitute major dietary form for the world population. In the non-growing season, cereal grains possess storage properties that help them to maintain constant calories and nutrients. According to FAO, annually major food worldwide supply came from 2,500 million tons harvested starch containing cereals. In plants, starch serves as an important carbohydrate storage molecule. The breakdown of starch ensures retaining essential life processes in plants during periods of non-photosynthetic activity, especially in darkness or during seed germination. Starch granules consist of two major components namely, amylopectin (branched) and amylose (linear). The amylopectin is responsible for the semicrystalline form of starch granules. The literature is divided into the following subheadings:

2.1 Architecture of the Starch Granule

2.2 Starch structure

2.3 Biosynthesis of Reserve Starch

2.4 Soluble Starch Synthase (*SS*)

2.5 Cloning and characterization of soluble starch synthase genes

2.6 Effect of heat stress on wheat

2.7 Effect of high temperature stress on soluble starch synthase activity

2.8 Effect of high temperature stress on soluble starch synthase gene expression

2.9 Allele mining

2.1 Architecture of the Starch Granule

Through different techniques, starch granule structure can be observed and described as follows:

Granular Rings

Granular rings appeared as contrasting light and dark rings that have thickness in the interior portion and thinner toward the exterior (Wikman *et al* 2014). Amylopectin and amylose are main components of crystalline and amorphorous rings, respectively (Koroteeva *et al* 2007) [Fig. 1].

Blocklet Structure

Hanson and Katz (1934) observed blocklets as projection from surface of granules

and Ohtani *et al* (2000) reported blockets of similar size in many crop plants such as sweet potato, maize, rice, and wheat. In wheat these are reported at various stages of the kernel development. Gallant *et al* (1997) reported that larger and smaller blocklets made semicrystalline and amorphous rings respectively.

Ridout *et al* (2003, 2006) also assumed that the blocklets largely consist of amylopectin and that they are embedded in an amorphous matrix that mainly consists of amylose. This matrix swells and gives rise to the amorphous granular rings, which appear bright and soft. The dark, semicrystalline rings arise because they contain defects that do not swell, which results in a discontinuation of the swelled areas.

Lamellar Structure

Sterling (1962) reported lamellar distance (9-10 nm) of starch granules through X-ray diffraction. In granules of starch, the amorphous lamellae appeared thicker whereas crystalline lamellae tend to be thinner, but contrasting situation is present in granules with high amylose. The major component of crystalline lamella is devoted by amylopectin, which makes 11–15 glucose residues of double helices (Bertoft *et al* 2008) which are further characterised into allomorph of A-type (present in cereal starches) or allomorph B-type (present in root and tuber starches) [Imberty *et al* 1991]. Six double helices having monoclinic unit cell exhibits A-type crystal with less water (Popov *et al* 2009), while B-type crystal exhibites hexagonal having six double helices, but with cavity accumulated with water (Imberty *et al* 1988). Stacks in semicrystalline rings were formed by amorphous lamella in combination with crystalline lamella (Cameron and Donald 1992).

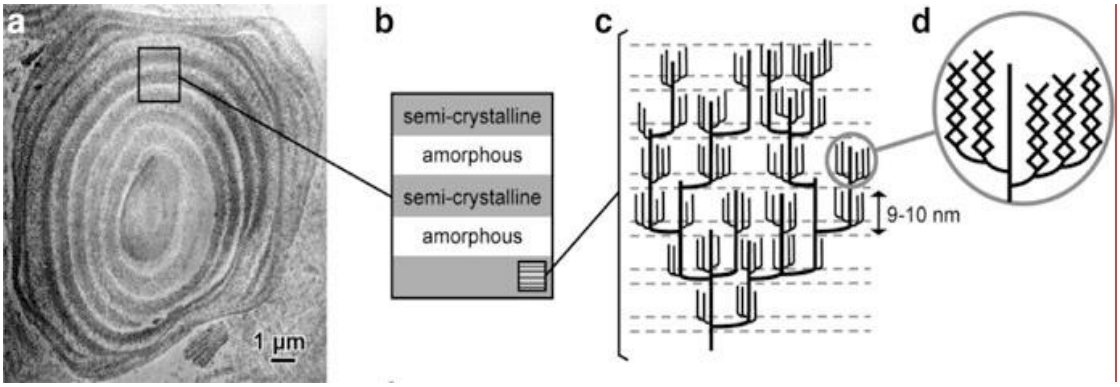

Fig. 1 Diagrammatic representation of the starch granule: (a) waxy maize ultrathin section (b) contrast nature of amorphous and semicrystalline growth rings (c) amylopectin clustered model (d) detailed structure of double helices from the branches of amylopectin (Buleon *et al* 2007)

The amorphous lamella constitutes internal part in form of branches consisting of amylopectin (French 1972). The amylose interferes with crystalline double helice arrangement of amylopectin (Koroteeva *et al* 2007).

2.2 Starch structure

Amylopectin and amylose are two major components of wheat starch. Amylopectin constitutes 75% of the dry weight of starch, while amylose constituting 25% and both the molecules are linked by α-(1,4) glycosidic linkages. Amylopectin accounts for molecular weight of 10^8 and is highly branched with α-(1,6) linkages, while amylose is shorter unbranched molecule with a molecular weight of 10^6 (Ball *et al* 1996) [Fig. 2].

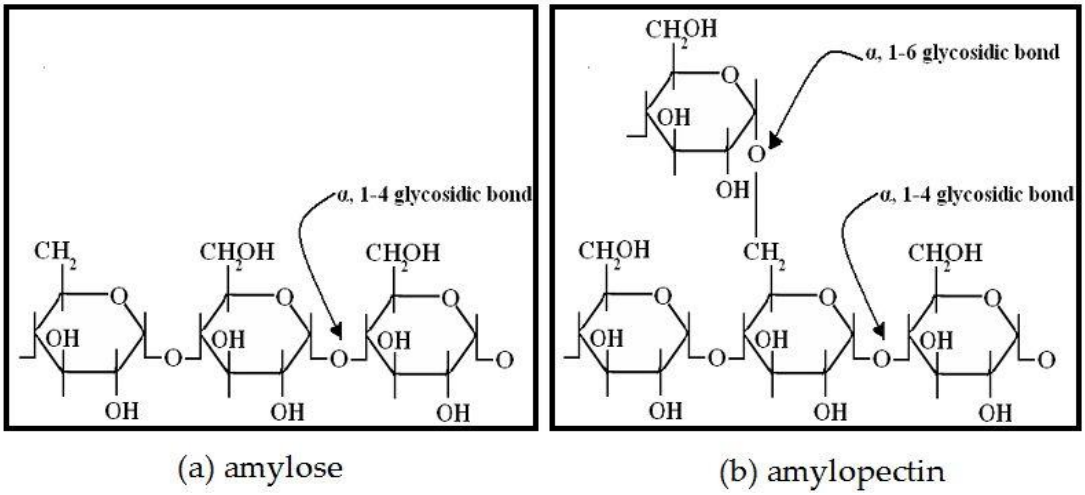

Fig. 2 Schematic structure of (a) amylose showing linear of α-(1,4)-linked D-glucan chains and (b) amylopectin showing branched α-(1,6)-linked D-glucan chains (Vandeputte and Delcour 2004)

Starch Major Component: Amylopectin

Amylopectin consists of linear chains linked by α-(1,4) D-glycosidic bonds interlinked through reducing end side by α-(1,6)-glycosidic bonds forming macromolecule (Buleon *et al* 1998).

Unit Chains of Amylopectin

The amylopectin is differentiated into short and long chains, where short chains forms the major group having distributions of size ranging from degree of polymerization (DP) 6-36 and peak pertaining to DP 11–15 whereas long chains are much smaller and in the different plants, ratio attained to S:L-chains is quite different for starches. Abundance of L-chains led to high ratio 10:1 to 22:1 in type-A cereal starches whereas type-B crystalline samples exhibited ratios between 6:1 and 8:1 (Bertoft *et al* 2008). Hizukuri (1986) studied that the S-chains consist of A-chains (DP 6–15) and short B-chains (DP 15–36). The L-chains were all further subdivided into B2-chains (DP peak 38–45), B3-chains (DP peak 62–74) and B4-chains (DP peak >80). Higher S:L chains ratio of A-type allomorph crystalline starches formed larger clusters than in B-type allomorphs.

Bertoft *et al* (2008) divided the starch samples into four different structural types based on the internal unit chain profile of their amylopectin component. In type 1, long chains are not differentiated from short BS chains. A-type allomorph crystallinity has been found in

type 1 amylopectin of several crops e.g. barley, oat, and rye. A and C-type allomorph exhibits type 2 amylopectin and groove in the chromatograms of DP 23 distinguished BL-chains from BS-chains (Buleon *et al* 1998) and includes maize, rice, and finger millet. Type 3 amylopectin reported more BL-chains and includes arrowroot and tapioca starch. All B-type allomorph starches form type 4 amylopectin which includes potato and canna starch. The wheat amylopectin structure was intermediate between type 1 and 2 amylopectin (Kalinga *et al* 2013).

Models of the Semicrystalline Structure of Amylopectin

The Cluster Model

The cluster model demonstrated that clusters are formed by short chains and number of chains in each cluster are affected by S:L chain ratio for amylopectin. Radial nature of the granule is due to penetration of stacks constituting amorphous and crystalline lamellae in the starch granules (Hizukuri 1986) [Fig. 3].

The Building Block Backbone Concept

The concept depicted that major structural building blocks are situated along a backbone that includes major part i.e. long chains. Semicrystalline lamellar stacks responsible for internal architecture of starch granules are produced by orienting tangentially amylopectin molecules in layers on each other. (Bertoft *et al* 2010).

Fine Structure of Amylose

Amylose constitutes second important component in starch composition forming 20–30 % of its weight. The linear nature helps to form complexes with iodine which allows the colorimetric (Juliano *et al* 2012), potentiometric (Bates *et al* 1943) and amperometric methods (Takeda *et al* 1987) for determination of the amylose content. Amylose content within same sample differs by starch granule size (Takeda *et al* 1999) e.g. cereal amyloses are smaller than root and tuber starches while branched amylose molecules are larger than linear molecules (Takeda *et al* 1989, Hizukuri *et al* 1989).

Unit Chains of Amylose

Chains of amylose molecules are classified into A, B and C chains; A and B-chains are defined as side chains whereas C-chain is known as main chain. On weight basis it shows mono-modal size distribution in which DPs < 1,000 has been reported in maize, wheat, barley and rice (Hizukuri and Takagi 1984) whereas on molar basis, it showed mono- or multi-modal size distribution where DPs >1,000 has been reported in potato and sweet potato (Hanashiro and Takeda 1998).

Evidence for the Presence of Branches in Amylose

The atomic force microscopy (AFM) imaging of amylose from pea starch at high resolution showed the presence of a long chain with multiple-branched short chains (Gunning *et al* 2003). Structurely, amylose exhibit variation in their branched architecture where as

amylopectin maintain same branched structure (Takeda *et al* 2003).

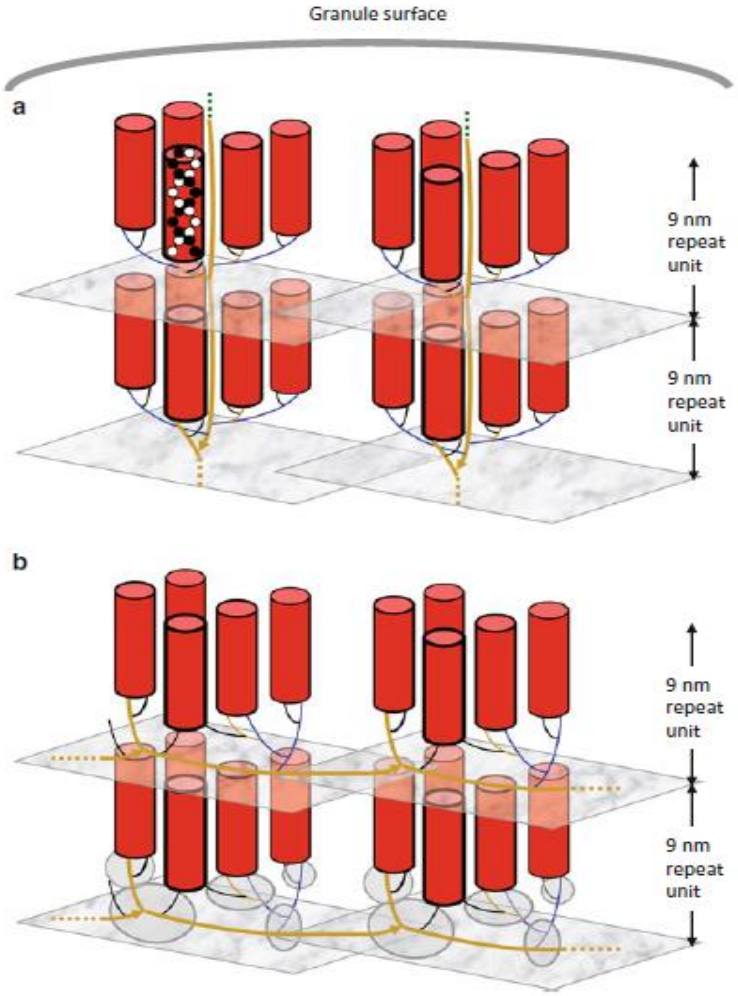

Fig. 3 Schematic drawings of amylopectin molecules forming the alternating crystalline and amorphous lamellae (9 nm repeat distance) in a starch granule according to (a) the cluster model based on Robin *et al* (1974) and Hizukuri (1986) and (b) the building block backbone model based on Bertoft *et al* (2010)

2.3 Starch Biosynthesis

Plants have two different pathways for starch biosynthesis - photosynthetic and non photosynthetic cells. Starch granules are accumulated inside the amyloplast of the higher plants. During photosynthesis plants accumulate starch and sucrose as storage compounds from CO_2 and water. CO_2 first enters into the Calvin- Benson (C_3) cycle, where in combination with Rubisco forms 3-phosphoglycerate (PGA) and fructose 6-P (F6P) through which starch and sucrose are synthesized in the chloroplast and cytosol, respectively. The photophosphorylation and the Z scheme (ETS) system are responsible for production of ATP and NADPH required for starch and sucrose formation, respectively.

Formation of ADPglucose, a glucose donor, catalyzed by ADP-glucose pyrophosphorylase (AGPase) is the first step for the starch biosynthesis in green plants (James *et al* 2003). The chain elongation is supported by shifting glucose residues of

ADPglucose to the non-reducing end of glucan chain. *SS* and *GBSS* are responsible for elongation of amylopectin and amylose chains respectively. *SBE* induces the α-1,6-glucosidic linkage and the final amylopectin structure is trimmed by *SDE*. The synchronized actions of *SS, SBE,* and *SDE* enzymes lead to synthesis of amylopectin whereas *GBSS* is responsible for amylose formation (Fig. 4).

2.4 Soluble Starch Synthase (*SS*)

Amylopectin existing α-1,4-glucosidic chains were elongated by *SS* by shifting glucose residues of ADPglucose to the non-reducing end of glucan chain. The four distinct *SS* isoforms are *SSI, SSII, SSIII,* and *SSIV*. The four *SS* isoforms and one *GBSS* isoform, exist that belongs to the CAZy glucosyltransferase family (GT) having common core sequence (60 kDa) and two conserved regions separated by a linker region namely GT5 and GT1 domains. The GT5 domain contains the KXGGL conserved catalytic motif, while the N-terminal extension regulates catalytic activity, thus altering binding capacity of *SS* for carbohydrates (Li *et al* 2000, Leterrier *et al* 2008).

SSI

The cereal endosperm constitutes 60–70 % of the *SS* activity for *SSI* isoform. Commuri and Keeling (2001) suggested that shortest glucan chains (DP 5-10) are synthesized by *SSI* in amylopectin molecule. Its affinity for glucan polymers increases with the increasing of glucan chain length. Fujita *et al*
(2006) reported that *SSI* activity of rice mutant lines (*ss1*) in endosperm had less chains (DP 8-12) and enhanced chains (DP 6-7 and 16-19) whereas low long chains (DP-21) were observed. The similar *SSI* suppression was also reported in wheat endosperm (McMaugh *et al* 2014).

SSII

Two isoforms of *SSII* are found in monocots, referred to as *SSIIa* and *SSIIb. SSII* gene structure and primary sequences are highly conserved in both monocots and dicots (Li *et al* 2003). *SSII* is responsible for synthesis of chains having DP 12-25 (Umemoto *et al* 2002). Compared to *SSIIb, SSIIa* activity has been defined in multiple species, including potato (Edwards *et al* 1999), wheat (Yamamori *et al* 2000), rice (Umemoto *et al* 2002) and barley (Morell *et al* 2003). Craig *et al* (1998) in pea reported that *SSII* gene mutations declined the proportions of intermediate chains and enhanced the short chains proportion. A report by Umemoto *et al* (2002) and Nakamura *et al* (2005) revealed that the *japonica* rice variety had increased chains (DP 6–10) and decreased chains (DP 12–24) showing inactive mutated *SSIIa* gene, in contrast with *indica* variety having an active *SSIIa* gene. Similar analysis were shown by Zhang *et al* (2004) in maize *sugary-2* mutant endosperm, Kitahara *et al* (2005) in sweet potato and Zhang *et al* (2008) in *Arabidopsis*.

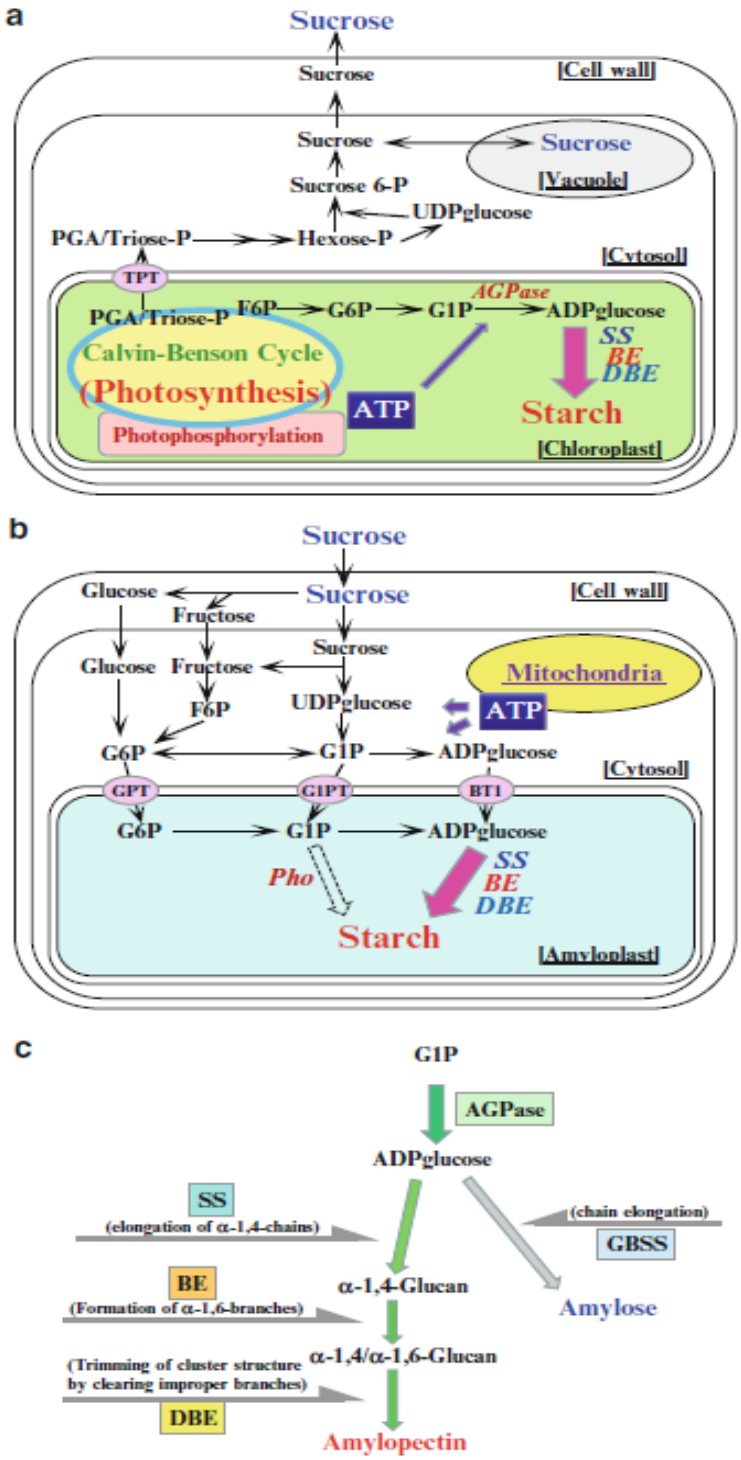

Fig. 4 Starch biosynthetic pathways in photosynthetic and non-photosynthetic tissues (a) The production of starch in the leaf (b) The reserve starch production in storage organs (c) The starch biosynthetic pathway from G1P (Fettke *et al* 2010)

Common characteristics of starch produced in the absence of *SSII* activity across all these mutant species, include higher content of amylose, reduced total starch content, modified amylopectin chain length distribution, altered starch granule morphology, and reduced crystallinity (Ball and Morell 2003). Thus, concluded that *SSII* is mainly responsible

for amylopectin intermediate chains.

SSIII

The *SSIII* isoform has been identified in potato, maize, and wheat (Abel *et al* 1996, Gao *et al* 1998, Li *et al* 2000). Only one copy was identified in dicots, such as in Arabidopsis. Two copies of *SSIII* were identified in some monocots, such as in rice. Two members of *SSIII* were cloned in rice, designated *OSSIII*-1/-2 (Dian *et al* 2005). Expression profile analysis indicates that *OSSIII*-1 is mainly expressed in the leaves, whereas *OSSIII*-2 is mainly expressed in endosperm. Gao *et al* (1998) concluded mutation (*dull1*) in maize induced by disrupted *SSIIIa* gene had altered granular morphology and physicochemical properties. Fontaine *et al* (1993) and Ral *et al* (2006) stated *SSIII* role in *Chlamydomonas reinhardtii ss3* mutant *sta-3* in the synthesis of intermediate and long chains of amylopectin. The *ss3* mutant of barley endosperm showed decreased shorter chains (DP 9–10) and increased intermediate chains with DP 15–24 whereas no variation was detected in long chains (Li *et al* 2011). Antisense suppression of *SSIII* in potato leads to the reduced synthesis of amylopectin of *SSIII* in potato lead to the reduced synthesis of amylopectin with a modified chain length distribution, and to starch granules with severely distorted morphology (Abel *et al* 1996).

SSIV

SSIV consists of least proportion of the total *SS* activity. Loss of *SSIV* cause the reduction in the starch granules of *Arabidopsis*, concluded it as important component for the initiation of starch granule (Roldan *et al* 2007). Toyosawa *et al* (2015) reported that in rice, a single *SSIV* gene mutation produced least effects on phenotype of starch, while the double mutants (*ss4b/ss3a*) resulted into the alteration of the polygonal into the spherical granules.

Biochemical analysis of the individual *SS* isoforms in different plant species showed that the contribution of each isoform to the total *SS* activity is different between species and also between different organs within a species. For example, in potato tuber, *SSII* and *SSIII* are the primary enzymes involved in the tuber starch synthesis (Edwards *et al* 1995). *SSII* accounted for only about 15% of the total activity, whereas *SSIII* contributed approximately 80% of the activity in the soluble fraction (Marshall *et al* 1996). Contrastly, *SSII* isoform accounts for 60-70% of the soluble *SS* activity in pea embryo (Denyer *et al* 1993). Cao *et al* (1999) reported that *SSI* have the major soluble activity (70- 80%) and *SSIII* contributes 20-30% of the total soluble activity in maize endosperm although *SSIIa* is also present in the endosperm.

2.5 Cloning and characterization of starch synthase genes

Several different starch synthases have been isolated and characterized in higher plants such as rice (Baba *et al* 1993), maize (Knight *et al* 1998), potato (Kossmann *et al* 1999), wheat (Li *et al* 1999a), barley (Gubler *et al* 2000) and Arabidopsis (Delvalle *et al* 2005).

In wheat several starch synthase isoforms (*GBSSI, GBSSII, SSI, SSIIa, SSIIb, SSIIc, SSIIIa, SSIIIb* and *SSIVb*) have been identified and cloned. Vrinten and Nakamura (2000) extracted *GBSSII* cDNA and revealed that the obtained *GBSSI* and *GBSSII* amino acid sequences of wheat have 66% identity. It was observed that *GBSSII* was expressed in leaf tissue but not in endosperm whereas *GBSSI* expression was present in endosperm tissue.

Li *et al* (1999b) isolated three classes of cDNA from a wheat endosperm cDNA library by nucleic acid hybridization and demonstrated that sequences were *SSIIa* of wheat and had the highest homology with maize *SSIIa* gene. Gao and Chibbar (2000) characterized three near-full-length putative homoeologous cDNA (*SSIIa-1, SSIIa-2*, and *SSIIa-3*) in wheat endosperm which were almost similar to the maize *SSIIa* gene. Li *et al* (2000) isolated and characterized cDNA sequences encoding wheat *SSIII* gene which was 5,346 bp long and contained an open reading frame that encoded a 1,628 amino acid polypeptide.

Leterrier *et al* (2008) cloned *SSIVb* gene in wheat and observed that *SSIV* cDNA of wheat and rice were similar to rice *SSIVb* and reported that *SSIVb* gene expression was present in leaf. Phylogenetic analysis showed that *SSIII* and *SSIV* formed one cluster, while *SSI, SSII* and *GBSSI* falls in second cluster. Yan *et al* (2009) cloned two *SSII* cDNAs i.e. *ZmSSIIc* and *TaSSIIc*, each containing an open reading frame of 2328 and 2229 bp in length from maize and wheat, shared 49.8–88.9% identity with those of other *SSII*s. Pan *et al* (2011) cloned and characterized *GBSS, SSI, SSII, SSIII* and *SSIV* in sorghum and *SSIIb* and *SSIIIb* genes in wheat.

Peng *et al* (2001) revealed that the *SSI* gene encoding protein (75 kDa) is located on wheat 7S chromosome and is distributed between the starch granule and the soluble phase. *SSI* is located on chromosomes 7A, 7B and 7D similar to the rice *SSI* gene structure. Full length *SSI* cDNA of 2591 bp (Li *et al* 1999a) and 2575 bp (Peng *et al* 2001) have been cloned, both of which encode polypeptide of 647 amino-acid residues identical in sequence except for single amino acid difference at position 321, where leucine is replaced by valine. Li *et al* (2013) investigated partial sequences of *SSI* from *Aegilops tauschii*, *Triticum turgidum, Triticum aestivum* and synthetic hexaploid wheat for evaluating the relationship among starch synthase genes of different species. Knight *et al* (1998) isolated the full length 2907 bp maize endosperm cDNA starch synthase (*SS*) clone and contained an ORF of 1866 bp corresponding to a polypeptide of 622 amino acid residues. The obtained maize and rice *SSI* amino acid sequence have 75.7% sequence identity.

2.6 Effect of heat stress on wheat

Wheat is a winter cereal and relatively narrow temperature range from 12-20°C is considered optimum for reproductive development. Therefore temperature higher from optimum can induce heat stress, which affects both crop growth and productivity. Wheat plants are affected through heat stress at growth developmental stages. But the period

immediately around flowering is especially sensitive to heat stress as this can influence floral fertility, grain number and weight and grain yield. There are yield losses in combination with both high temperature as well as heat shock at mid or late reproductive stage of wheat, and grain filling where temperature is above 32°C. Increasing temperature associated with climatic change could negatively affect global wheat yields resulting in potential increases in food insecurity and poverty (Tubiello *et al* 2000).

Exposure to high temperature can cause considerable morpho-physiological damage, which hastens leaf senescence, reduces photosynthesis and pollen viability and also reduces starch biosynthesis and water soluble carbohydrates (WSC) reserves and shorten the grain filling period. Elevated temperature between anthesis to maturity, affects grain yield of wheat due to shortening of grain filling period to accumulate current or reserve photo-assimilates as compared to normal conditions.

2.7 Effect of heat stress on *SS* activity

Starch is a main constituent of the wheat grain, having upto 70% of dry mass and reduction in starch deposition is the main reason for decline in grain mass. The biosynthesis of starch depends on three enzymes, *AGPase*, *SS* and *GBSS* out of which *SS* regulates starch synthesis is sensitive to heat stress. There are three different classes of wheat starch granules namely A, B and C-type starch granules. During grain filling stage, high temperature stress may reduce starch content and decrease starch granular size in wheat grain. This physiological process may limit the grain size and reduce grain yield of wheat (Fig. 5).

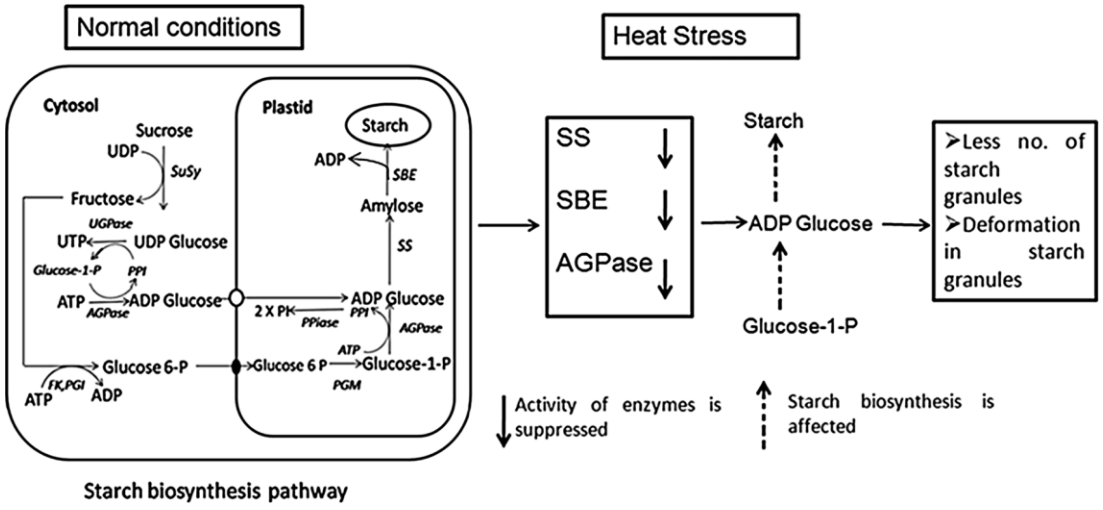

Fig. 5 Flow chart showing the significance of three important enzymes during heat stress (Gupta and Saripalli 2015)

Peng *et al* (2004) revealed that thermal stress (temperatures above 37 °C) during the critical grain filling period can affect the biochemical processes of starch deposition causing yield loss and starch defects. Zakaria *et al* (2002) demonstrated that failure in starch deposition results in lightly packed granules and grain chalkiness. The intrinsic property of

starch synthase gets affected at temperatures well within the physiological range. Denyer and Smith (1992) reported that pea *SS* activity was disrupted at high temperature (> 35 °C) leading to loss of tertiary structure of the enzyme.

Rijven (1986) studied that the wheat grains grown at 21°C reduced by >50% of the *SS* activity by transferring to 37°C. Jenner *et al* (1993) revealed that at 20°C, there was no loss in *SS* activity; when heated at 30°C resulted in 15% loss and further at 40°C the loss increased to 67%. High temperature treatment of wheat ears at 35°C, showed that *SS* activity was reduced to one- half of the value in unheated grains in one day (Hawker and Jenner 1993). Keeling *et al* (1993) observed that when wheat ears heated at 40°C, 97% of *SS* activity was lost.

Prakash *et al* (2009) revealed the synchronized effect of heat stress, on growth of wheat grain and *SS* activity, indicating that soluble starch synthase is the key component providing sensitivity to elevated heat stress on wheat grain. Labuschagne *et al* (2009) reported that the lower ratio of starch in the endosperm in response to heat stress is because of the heat inactivation of *SS*. Zhao *et al* (2008) investigated the effects of high temperature regime on cultivars of winter wheat for the enzymes activity that are responsible for starch and protein and revealed that under elevated temperature (28°C), *SS* synthesis declined from 21-28 days in contrast to lower temperatures (20°C).

Kumar *et al* (2013) studied the activity of *SS* enzyme in seeds of wheat cultivars C-306 (thermotolerant) and PBW343 (thermosusceptible) and observed increase in *SS* activities during afternoon (35°C), compared to morning (29°C) and evening (32°C) in both the cultivars. The plants were further exposed to heat stress of 42°C for 2 h and reduction in the *SS* synthesis activity was recorded in both cultivars suggesting that *SS* was quite thermostable and could sustain temperature upto 35°C but decrease in the activity occurred due to thermal denaturation of the enzyme.

2.8 Effect of high temperature stress on soluble starch synthase gene expression

Transcription of *SS* gene has been studied in wheat over the grain development period by semi quantitative RT-PCR, qRT-PCR and microarray analysis, but only a very few reports are available on effect of high temperature stress on *SS* gene transcription. Stamova *et al* (2009) analyzed the expression of genes encoding *SS* present in wheat using DNA microarrays and showed that *SSI* gene had bell-shaped expression pattern with a sharp increase at about 14 DAA, when the dry matter and storage reserves are rapidly accumulating, and declined as the caryopsis matures. In a similar study, Ganeshan *et al* (2010) studied wheat carbohydrate metabolism genes using real-time PCR analysis and observed that level of *SSI* transcripts peaked at 10 DAA when starch synthesis was actively on followed with reduction in expression to 30 DAA.

Hurkman *et al* (2003) examined the expression of starch biosynthesis enzymes during wheat developmental stages under effect of heat stress. He observed that during 12-16 days

post-anthesis period (DPA) at 24/17°C regimen, steady state transcript levels for AGPase, *SSI*, *SSII*, *SSIII*, *GBSS*, *SBEI* and *SBEII* were higher, in contrast to 37/17°C regimen, transcript levels were reduced. He deduced that among all starch biosynthetic enzymes, *SS* transcript levels decreased greatly under heat stress. Altenbach and Kothari (2004) studied accumulation timing of genes transcript profile through RT-PCR in wheat grain endosperm and observed that under 24/17°C regimen *SSI* transcripts were most predominant at 7, 14 and 20 DPA and expression fell abruptly after 20 DPA while under 37/28°C regimen, transcript expression was high upto 8 DAA only. Zhong *et al* (2012) studied *SS* expression after anthesis at temperature 22 and 32°C in rice and reported that *SSIIa* and *SSI* were more affected by heat stress as compared to *GBSS*.

2.9 Allele Mining

A large amount of sequence information has been deposited into GenBank in the last 20 years (Madris 2008) so as to identify valuable alleles imparting tolerance to biotic and abiotic stresses, enhanced nutrient use efficiency, increased yield and improved quality. Allele mining is a technique for identification of natural allelic variants at candidate loci that has utility in improvement of crops for controlling key agronomic traits. This approach focuses on finding mutations- single nucleotide polymorphisms (SNPs) or insertions:deletions (InDels) which could explain the phenotypic differences arising between two genotypes with different alleles for the trait of interest, identifying the evolution of alleles, new haplotypes identification and development of allele-specific markers for use in MAS.

Specified positions of a chromosomal site at which the DNA sequence of two genotypes differ by a single base is defined as SNP. There are two types of nucleotide base substitutions resulting in SNPs:

1. **Transition:** Substitution between purines or pyramidines.
2. **Transversion:** Substitution between purine and pyramidine.

A SNP present in the coding region directly influence protein, an intronic SNP can affect splicing, SNP in a promoter can alter gene expression etc. Haplotype refers to allelic combination of different loci on same chromosome. A haplotype is a set of DNA variations that tend to be inherited together. For finding out haplotypes, SNP detection is necessary. DNA sequencing is used for SNP discovery. Sequencing data often suffers with errors, thus sequence accuracy and manual curation is a must. It is critical to take into consideration the quality of bases called at candidate SNP position. The key to robust SNP detection is high quality sequencing data.

Shimbata *et al* (2005) revealed the mutations in wheat *SSII* gene by using PCR-based DNA markers. In the *SSII-A* allele, a 289 bp deletion was identified near the initiation codon; in *SSII-B* allele, 175 bp insertion was detected in exon 8 and in *SSII-D* allele, 63 bp deletion was found in fifth exon-intron junction. Thus showed the availability of the suitable DNA

markers in marker assisted selection (MAS) breeding program. Rose *et al* (2015) sequenced three AGP-L genes encoded the large subunit of AGPase in 47 spring and winter wheat genotypes. AGPL-A1 and AGPL-D1 alleles contained one and two haplotypes, respectively. In AGP-L-B allele, 67 SNPs and 13 indels were identified forming five haplotypes. All 13 indels and 58 SNPs were located in the introns, while the nine SNPs were detected in the exonic region.

Monari *et al* (2005) characterized mutated waxy loci in four tetraploid and hexaploid wheat genotypes wheat by PCR, Southern blotting and nucleotide sequences analysis. Three primer pairs were developed and revealed that gene deletion resulted in null allele at the *Wx-B1* locus in one hexaploid wheat, whereas sequencing analysis reported deletion of 724 bp in the *Wx-D1* locus of one hexaploid wheat genotype and 89 bp insertion was identified in the *Wx-A1* locus in one tetraploid wheat, respectively. Saito and Nakamura (2005) reported a single nucleotide insertion in the *T. dicoccoides Wx-A1* allele and a single nucleotide deletion in the *T. dicoccum Wx-A1* allele that caused frameshift mutations resulted in premature termination. The transcript profiles of the *T. dicoccoides* and *T. dicoccum* alleles were decreased by 6.5 and 1.5% of wild-type. Yamaguchi *et al* (2005) studied membrane-localised protein (*TaALMT1*) imparting aluminium resistance to wheat genotypes. Raman *et al* (2006) reported that variation in third intron of *TaALMT1* is due to SSRs and indels. The *TaALMT1* promoter region contained SNPs and VNTRs (Sasaki *et al* 2006), responsible for the different levels of expression in wheat cultivars.

Two major techniques for the sequence polymorphism identification in gene in the naturally occurring populations:

(i) EcoTilling based allele mining

(ii) Sequence based allele mining

EcoTilling based allele mining

TILLING (Targeting Induced Local Lesions IN Genomes) uses traditional mutagenesis and nucleotide polymorphism discovery methods in reverse genetics (McCallum *et al* 2000). TILLING can be applied to chemically mutagenized populations or a collection of genotypes (e.g. cultivars, ecotypes, landraces and wild accessions), this latter approach also known as EcoTILLING (Comai 2004). Many important EcoTILLed genes have been screened in naturally occurring populations (Kadaru *et al* 2006, Wang *et al* 2007, Wang *et al* 2008, Wang *et al* 2010).

Sequencing based allele mining

Sequencing-based allele mining involves PCR-based amplification of alleles of a gene in varied genotypes and then DNA sequencing to recognise nucleotide variance in the alleles. It helps to analyze individuals for haplotype structure and diversity to infer genetic association studies in plants. Sequencing-based allele mining does not require much

sophisticated equipments or involve tedious steps like Eco-TILLING, but the high sequencing cost of the targeted genes for this technique was a major issue which is being reduced with the advanced sequencing methods and techniques (Hutchison 2007). In wheat many new haplotypes have been identified for various candidate genes such as *Pm3* for powdery mildew (Kaur *et al* 2008), *Wx-A1* for waxy gene and amylose biosynthesis (Saito and Nakamura 2005), *Ap1* and *PhyC* for vernalisation response (Beales *et al* 2005), *SSII* for endosperm starch biosynthesis (Shimbata *et al* 2005), *Wx-B1* for waxy protein (Monari *et al* 2005), *PSY-1* and *PSY-2* for grain yellow pigment content (Zhang and Dudcovsky 2008), *Viviparous-1* for pre-harvest sprouting tolerance (Xia *et al* 2008) and *TaALMT1* for aluminium resistance (Raman *et al* 2008).

CHAPTER III

MATERIAL AND METHODS

The experiments conducted in present study were carried out in the Fields and Molecular Biology Laboratory of School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India during the year 2014-2016. This chapter describes the materials used and methods followed for various experiments.

3.1 Plant Material

20 different wheat genotypes which include wild diploid and tetraploid genotypes and cultivated tetraploid and hexaploid wheats were raised in the fields of the School of Agricultural Biotechnology, PAU using standard agronomic practices (Table 1).

Table 1: List of 20 different wheat germplasm lines used in present study

	Wild wheat
1.	*Triticum monococcum acc. pau14087*
2.	*Triticum boeoticum acc. pau5088*
3.	*Ae. speltoides acc. pau3809*
4.	*Ae. speltoides acc. pau15081*
5.	*Ae. speltoides acc. pauTA1784*
6.	*Ae. tauschii acc. pau3747*
7.	*Ae. tauschii acc. pau14102*
8.	*Triticum dicoccoides acc. pau14801*
9.	*Triticum dicoccoides acc. pau7107*
	Durum wheat
10.	MACS9
	Winter bread wheat
11.	Arbon
12.	Impala
	Spring bread wheat
13.	Giza
14.	C306
15.	C591
16.	WH542
17.	PBW343
18.	Halna
19.	Raj3765
20.	K9644

3.2 Identification of candidates for starch synthase gene using bioinformatics tools

3.2.1 Downloading *Triticum aestivum SS* nucleotide sequences from Public Database

Entries for nucleotide sequences were searched for the *SS* gene using the NCBI and

EMBL databases. The resulted sequences were then divided in four different files, each containing one of four different isoform of starch synthase (*SSI*, *SSII*, *SSIII* and *SSIV*). Common entries of gene accession from NCBI (https://www.ncbi.nlm.nih.gov/) and EMBL (https://www.embl.org) were unified and files containing representative sequences were made.

3.2.2 Clustering of the selected sequences

These selected representative sequences in each file were clustered using CD-HIT suite tool (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi) with 95% identity to remove redundancy.

3.2.3 Online BLAST against IWGSC (Ensembl Plants) *Triticum aestivum* database

Clustered sequences were used as query for online BLAST against IWGSC (Ensembl Plants) *Triticum aestivum* (http://plants.ensembl.org/*Triticum_aestivum*/Info/Index) database as subject. It contains genome assembly of *T. aestivum* cv. Chinese Spring. Accession number for each selected sequence had been recorded. Selection was further narrow down on the basis of low e-value, high bit score, multiple hits of single entry and gene functional annotation.

3.2.4 Standalone BLAST of selected sequences against cDNA *Triticum aestivum* database

The full length coding sequences (CDS) were downloaded from TriFLDB database (http://trifldb.psc.riken.jp/) and clustered into previously clustered file of IWGSC database using 99% identity. Standalone (offline) BLAST of common selected sequences as query was done against cDNA *Triticum aestivum* database as subject downloaded from IWGSC (Ensembl Plants) tool. The following command was typed in terminal in linux system:

makeblastdb -in <database_name> -dbtype nucl

blastn-query <query_name> -db <database_name> -out <output_file name> -outfmt 6-evalue

3.2.5 Standalone BLAST of selected full length CDS sequences against genomic database

To reconfirm the position of full length gene/CDS, standalone BLAST of these sequences was done as queries against IWGSC (Ensembl Plants) *Triticum aestivum* genomic database as subject.

3.2.6 Comparison of results of gDNA and cDNA BLAST

Results of both genomic and complementary DNA sequences databases were compared and common accession IDs in gDNA and cDNA databases were selected. The chromosomal location of candidate genes was then noted. The available literature was also referred to confirm the chromosomal position of most likely candidate genes.

3.2.7 Estimation of diversity among selected full length CDS sequences

The multiple sequence alignment of selected full length CDS sequences was carried out using ClustalX tool and tree was constructed using Phylogeny.fr (www.phlogeny.fr) tool to assess the phylogenetic relationship/divergence analysis among the different selected sequences.

3.2.8 Confirmation of selected sequence IDs using gene expression analysis studies in grain development

For functional characterization of *SS* gene, the accessions of selected gene sequences were compared with gene expression profiles study using WheatExp (http://wheat.pw.usda.gov/WheatExp/) database that showed expression transcript profiles of homoeologous of *SS* gene across a wide range of tissues (grain, leaf, root, spike and stem) from different growth developmental stages and under heat and drought stress conditions. The accessions showing maximum expression in grain and under heat stress condition were selected.

The detailed outline of candidate gene identification of *SS* using bioinformatics tools is shown in Fig. 6.

3.3 Primer designing and cloning of starch synthase gene

3.3.1 Designing primers for homoeologous *SSI* gene

The nucleotide sequence of the target *SSI*-7B gene (Fig. 7) was used for the primer designing using the software PerlPrimer (http://perlprimer.sourceforge.net/) tool. Three sets of primers were designed ranging from full length of gene to primers amplifying 800-900 bp of gene to ensure uniform amplification. All primers were tested on different wheat genotypes for amplification. In first set of primers, one pair of primer can amplify the whole gene. In second set of primers, the target gene sequence, were divided into relatively two equal parts (Fig. 8) with an approximately 150-200 bp overlap in the middle. SSL1 and SSL2 primers for each part of the gene were designed in cultivated genotypes for present study (Table 2). In third set of primers, the seven overlapping primers covering the full gene were designed for amplification and sequencing in wild genotypes (Fig. 9) with the amplicon length ranging from 830 to 900 bases (Table 3). These same set of seven overlapping primers were used for sequencing the cloned fragments in cultivated genotypes except SSOL4 primer. This primer was redesigned into two primers (Fig. 10) as discussed in section 4.2.1 (Table 4).

All the primers designed were synthesized through facility of Eurofins Genomics, India. 1X TE buffer was used to dissolve each primer to make master stock of 100pM. They were further diluted using nuclease free water to make working primer solutions of 5pM.

Table 2: Sequence of two primers designed to amplify *SSI*-7B gene in two parts

Primer ID	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	Amplicon size
SSL1	TCAGCAGACATATCACTCGACTGACC	CAGATCCACAAACAACGTACACTGGCAC	2289 bases
SSL2	TCATAAGTTGTACCACCTCGCTGACC	TACTGGATCTGTTGGCTACAAACTGCT	2677 bases

Table 3: Sequence of seven primers designed to amplify *SSI*-7B gene in seven parts

Primer ID	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	Amplicon size
SSOL1	TCGTGAGATGATTGATTGGG	CCACAATTATGCTGTCAATGG	830 bases
SSOL2	ATCTAGCCAAGAACAGAATGCAG	CCTTATTCACATGAAGTGCTTGG	855 bases
SSOL3	TGTACTCTTGCGCATGAAGG	TCTCCATATAAACTTCCTGGTCTG	844 bases
SSOL4	TTTGCTCCTTACAAGAGTGCC	TGACACAATCAAGGCTGAGG	894 bases
SSOL5	GAGTTTATGTACAGTCCCAAAGAG	CCATGTACAGATAATGAGAAGTGG	872 bases
SSOL6	CCAGTGTTGCATGTTAAATTGG	TAGTTAGTGAGAACATCATGGG	880 bases
SSOL7	AGTGTCTCCCTCATCATTATTCTG	ATATCAACACCTCTCCTCACC	870 bases

Table 4: Sequence of two primers redesigned for sequencing in cultivated genotypes

Primer ID	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	Amplicon size
SSOL4A	TTTGCTCCTTACAAGAGTGCC	CAGATCCACAAACAACGTACAC	449 bases
SSOL4B	CATAAGTTGTACCACCTCGCTG	TGACACAATCAAGGCTGAGG	634 bases

3.3.2 Genomic DNA isolation

Genomic DNA was isolated from the leaf tissues of field grown plants using miniprep CTAB extraction method (Murray and Thompson 1980):

1. Young leaves were collected from field grown plants.
2. The leaves were crushed using liquid nitrogen to form a fine powder with help of mortar and pestle and collected in 2 ml centrifuge tubes.
3. To the powder 800 µl of CTAB buffer (2% CTAB, 1.4 M NaCl, 100 mM TrisHCl, 20 mM EDTA, 0.5% sodium bisulphite) containing 1 % mercaptoethanol was added and the tubes were incubated in water bath at 65°C for 1 h.
4. After incubation, 800 µl of chloroform: isoamyl alcohol (24:1) was added to the mixture and shaked for 30-45 min.

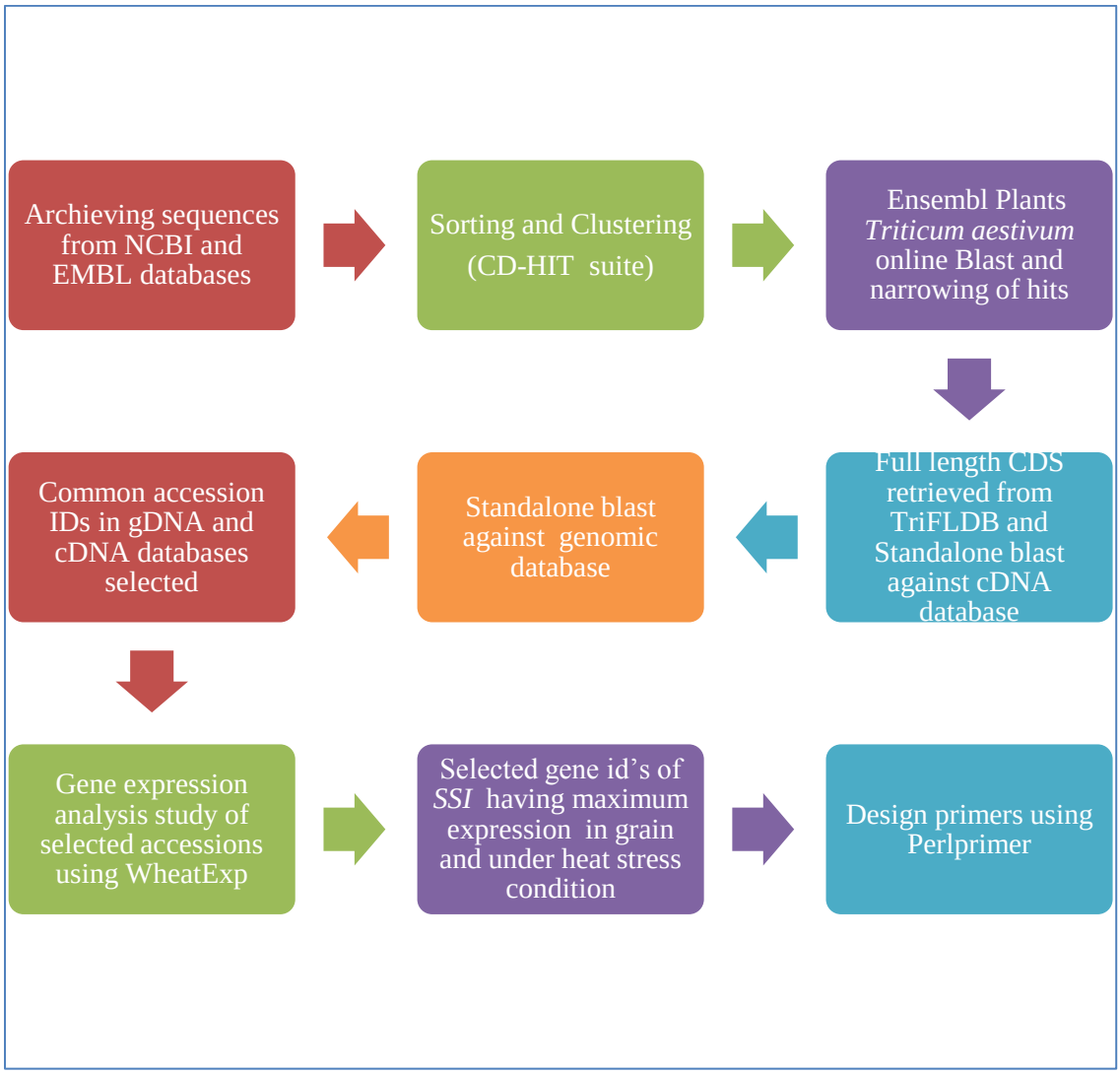

Fig. 6 **Outline of identification of candidates for starch synthase (*SS*) gene using bioinformatics tools**

>TGACv1_scaffold_592658_7BS dna:scaffold scaffold:TGACv1:TGACv1_scaffold_592658_7BS:30992:43775:-1

CGAGACCGACGCGGGCAGATCAATAGATTTTTCTAGACGTTTTATTGGATCGTGAGATGATTGATTGGGGTGGCGTGTCGATACGATAGCGGTGCACCGCC
GATGTATCGGGGCCTGTGCACGTGGTTGGGTCTCAGCAGACATATCACTCGACTGACCGGAGGGAGTACTACTAGAAAGAAGACTGAAAAGGCTAGGCCA
AGTGCACGCATGTTGGGAACGTAGTTAAATTGAGGAGTTTGTCTTTTGCTTGGCCTGGTATTATTACCAAAAATGGTGGAGTCCCTGTACTTAATGGGAAAA
TCTTAACATGACACTGGGGTTTATGAGTCTCCAATTGTATATTCTCAGCACTCAACTGATTTTACTGATAGACATACACATTTTTTTTTTGAGGGGTAGTGGA
CATGCACATGAGCCCCCTTCTTCAAGGAATGCAATGCTTCTTTCTGTTTTATATGACAGGAACTAGAAGAGCTTCCACCTTTGAGTACAGAAGTAGTATCGA
AATTGAAGACCCTTGTATTACTGTCTTGTTTTTCAATGAAATGGGAGGCCCATGCAGTAAGTCACATGGGCGCCTGGGAGGCTGGGATCATGTGTGCTTTGC
AGAGTACTAGACCCAGCTCACTCTCTGTTAGATTCTTTGTTGGGTTGCTACTTTGTGTTTGCTGTGTAGTATATTAGACATCCTGAATTTGGCATCTAGCCAA
GAACAGAATGCAGGTTGCGCCAATTTTTAAAGAATGTGATCTTCTAAGTATTAATTAATCATGTTCTGCTAATGCCTGTCCTAACTCTGGTCTTGACAAATA
TACCATATGAATATTTTCCATTTTGCAACCAGGGATTGCTGAGGATTCCATTGACAGCATAATTGTGGCTGCAAGTGAGCAGGATTCTGAGATCATGGATGC
CAAGGATCAACCTCAAGCTAAAGTTACTCGCAGCATCGTGTTTGTGACTGGGGAAGCTGCTCCTTATGCAAAGTCAGGGGGGTTGGGAGATGTTTGTGGTT
CGTTACCAATTGCTCTTGCTGCTCGTGGTCACCGAGTGATGGTTGTAATGCCAAGATACTTAAATGGGTCCTCTGATAAAAACTATGCAAAGGCATTATACA
CTGCGAAGCACATTAAGATTCCATGCTTTGGGGGATCACATGAAGTGACCTTTTTTCATGAGTATAGAGACAACGTCGATTGGGTGGGTACACAATCACCT
TCTTATTCTCTGTTAAATTCTAGCAACTGTTTATCCTTGTTTACACTTTCTTTTAGCCCTGCAAAGACATATGTGATCTCCATACTTTCTTGTTATTTCCCTTGT
ACTCTTGCGCATGAAGGTCAAAATATGATATACCCATGGAAGTCATGCATGTGCCTAGTATTTTTGGTGTCGGTGCCTTTAACTTTCAGGGATTAATACGTG
GAATTTGATAACTAAAGTTTATTTTATTGAAAAAAATTGTAGGTTGGCTGAGCCCACAGCCATGCAGTGCCACCACTGCTTGCACATGATTTTGCATTTCTG
TTTGCACCAAGCACTTCATGTGAATAAGGTGTAAAATCATAAAGTACCAATTTTATTCTGCCAATTGCTCTTAAGAGTATATACATTTATCTTGGCCTCAAT
CATGGGAGTACTGTGCATTCAGTGCACCATCATTGTTCTAAGGAGAAAATGTGGGTGCAAGGAAGACACTTTTGTCCCTTAATAAAAGGCAGGCACTCTGT
TGTCATATAGATAGAAAGCAACAAACTTATTTCAAAGAGCTAACAATGGCAAAAGAACCAAAAAAAGCATGCTAAGGCGGCGACACCAAAAGGTGAGGG
GGGCCTTGTGACTGACAGCACCCCAAACTATTGCCATTGTTTTACTAAATGAAGATCATTTTAGAAGCTCTCAGGAACTTCGGAAACAGTGGCTTTCCGCCC
ACAGATCGTCTGTTGATATTTTTGTCCACTGATACTTTTTTTGCTCCTTACAAGAGTGCCTATGTTGACATAGATGTTAAGTTGTTCATAAGTTTACTTCTTAT
TCTAAACAGCAAGTGCCTAATGCTTGCATTTATTTTGGCTATTTATTTTTATTCTCATTTCAATCAACACTTTTGTTCAGGTGTTTGTCGATCATCCGTCATAT
CACAGACCAGGAAGTTTATATGGAGATAATTTTGGTGCTTTTGGTGATAATCAGGTACATTACACTATACTAAGCTCCTAGTTGACTAAGTCATAAGTTGTA
CCACCTCGCTGACCGGCTGCTCTATGTCGTGCAGTTCAGATACACACTCCTTTGCTATGCTGCATGCGAGGCCCCACTAATCCTTGAATTGGGAGGATATAT
TTATGGACAGAATTGCATGTTTGTTGTGAACGATTGGCATGCCAGCCTTGTGCCAGTGTACGTTGTTTGTGGATCTGAAAGTCCAATCCTTTATTCATTCTCT
GCTTTGCAGTGTGCCCATGTCTACATTTCTTTTATGCTTTTTTCATGTCTGTTCTTATATTGCATATATGCTTATGGAGTCTAAAAGTTACCGGAGGGAATAA
CTCTTAAGGATTTCCTCAATCAATTATCTTTAGCTTTAGTTAACATTTATTGTGGCAAACATAATGTGTTTTGAGATTTACAAGTTCAGAGATTGCACTTCAC
CAGTTCGTAGCTAATCTGATGTTTTCCCCGAGAAAATGCCTAAAGCTTTGTGTCTTGATGCATTGATAGAAAAAGAGTTTATGTACAGTCCCAAAGAGGGG
ACACAAAATTACAACACCGCACCCCTGAGAACTAGGCGCTGCCGGAAGAAGCGATGCAAGCCCCACTGCCCCTGCCTTCGCTCACAGCCGGGCCTCAGCC
TTGATTGTGTCAAGTAAGCTAGCAGTGCTAGACTGCGCAAGGTCGATTCGTCGAAGATGACAGTGTTGCGCTGCTTCCAAATCCACCAAACTATGAGCATG
ATCACTGGAGATGTACCTTTTCTCGCGGCTGAGGGGGTGGACTGGTGGTCTGCTGCTGCCAATTTTCAGATAATCTGAAAAATGCATGTTTTGATGATTTTA
GTATCTTGCGGACCCTGGGTACCACCTAAGCTTTCACACAGTAATTTGCAGTTACACCTATAAAAGTAACGGTCATGATATGCATGTGTTTTGGGTAGATCA
TGGTGCATGCATTTTAGGAATTAGGACATGTCAGAACCACGTGAGGCTTATGGGGCAATTCATTTGTTCCATTATACGAGTCACGAATATGGTTCAGCATGT
TTGGACACTACTTGTTTGGGGCAATTTGAGATGGTGAATTGTAGCTGCTTGATGTTGGCTAGGTGGATTATTTTGTACAAGTATCGATGTTAGATGCATATT
TCCTTTTGCCTTGTGCTGTTTGCCATGTCGTGTTCCCCTTTTTTGTCGCCAGTGTTGCATGTTAAATTGGTTTTCATTACATAATCAACTTTGTTGCTGACATC
AGTCATTTTTATTCAGCCTTCTTGCTGCAAAATATAGACCATACGGTGTTTACAGAGATTCCCGCAGCACCCTTGTTATACATAATTTAGCACATCAGGTTTG
GGTCTACCACTTCTCATTATCTGTACATGGCTTTGTAAGTCAGTTCACACGTATCGTCATACTGTTCAATGTCATTAGGGTGTGGAGCCTGCAAGTACATATC
CTGATCTGGGATTGCCTCCTGAATGGTATGGAGCTTTAGAATGGGTATTTCCAGAATGGGCAAGGAGGCATGCCCTTGACAAGGGTGAGGCAGTTAACTTT
TTGAAAGGAGCAGTTGTGACAGCAGATCGGATTGTGACCGTCAGTCAGGTGAAATACTCATACTTCTCTTTTTTCTTTGCGGGATGTTCTTCAGTTCAATTG
CCCTGTCTTTCACGCAATTAAGAAATGATTTAATCTTTTGTTTCTAGGGTTATTCATGGGAGGTCACAACTGCTGAAGGTGGACAGGGCCTCAATGAGCTCT
TAAGCTCCCGAAAAAGTGTATTGAATGGTAACTATATTTGAATCCACTTATCTTCTTCTGAAACATATTTACAGAAATAGATGGATGGGTTGCAAGTATTAA
TTCAGTTTGCTCTTTCGGTATGCAGGAATTGTAAATGGAATTGACATTAATGATTGGAACCCCACCACAGACAAGTGTCTCCCTCATCATTATTCTGTCGAT
GACCTCTCTGGAAAGGTGTGTGGATAGTACCCTATATAATAACATGTATATGTGATCTAGCACTTTCTTTTTCTTTGCTAGTTTGCCTCCCATGATGTTCTCA
CTAACTAATCCTATGTGGTTTGGCATACTTGTCAGGCCAAATGTAAAGCTGAATTGCAGAAGGAGTTGGGTTTACCTGTAAGGGAGGATGTTCCTCTGGTTA
GATACCAACCCCTAAGATATATATTTTTTAAATCCCTAAAAAAAACTTGCCGATCATCTCATTAGCTTGATTCACAGATTGGCTTTATTGGAAGACTGGATT
ACCAGAAAGGCATTGATCTCATTAAAATGGCCATTCCAGAGCTCATGAGGGAGGACGTGCAATTTGTAAGTTCATATTCTTTTTCTTGAGACTAGAGTATTA
AATCAAACATCAGTAATATGCTACTGCAAGTTTTGAAACAAAGTTGGAGGCAATAAATATATGTGTAGCACTGACCATGCAGTGCCACTATAGCTGGAGTG
TCCTGTAGTCTATGTGATCTAACACACTCAACAACATGTTTTCGCATACAAACACATGCGTGCGCGCAACAAACATACTCTACAATAAAATTGGCTTGGTG
AACTGCAGACATGCTCTTATATCCATTCCAACATTTCTTGTTTCATCATTGGCTGAAGACTAAGAGAAAGGGGACCCAGGGTGATTTTAGCAGTTTGTAGCC
AACAGATCCAGTAAGGAAGCTAGCCGAGCCTAGGAGAATTCGCTTAGGTAGCTGGAATGTAGGGTCCCTGACAGGGAAGCTTCGGGAGCTAGTCGATGCA
GCGGTGAGGAGAGGTGTTGATATCCTTTGCGTCCAAGAAACCAAATGGAGGGGACAGAAGGCGAAGGAGGTGGAGGATACCGGCTTCAAGCTGTGGTAC
ACGGGGACGGCTGCAAAATGCAAACAGAAAGGGCGTAGGCATTTTGATCAACAAGAG

Fig. 7 Sequence of the gene *SSI*-7B (5156bp) adopted from Ensembl Plants IWGSC database

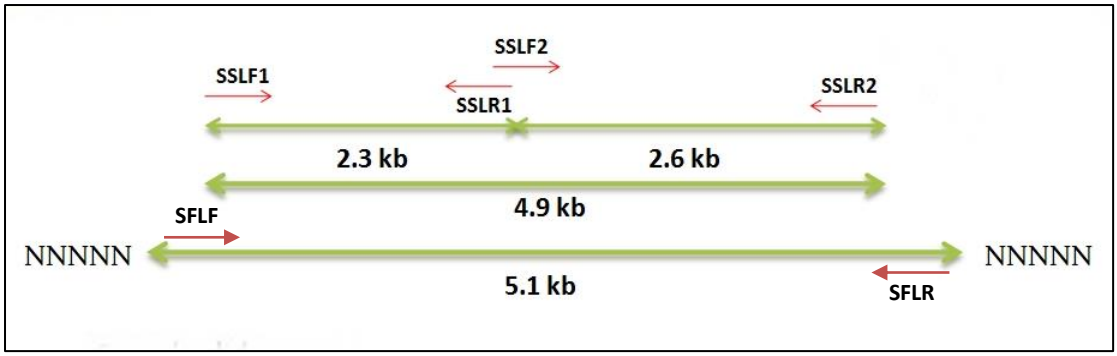

Fig. 8 **Diagrammatic representation of *SSI*-7B gene and two sets of primers designed; The total length of the gene was 5.1 kb but due to presence of NNN nucleotides upstream and downstream flanking the *SSI* gene, first set of primer SFL, with forward primer SFLF from position 130 bp and reverse primer SFLR from position 4906 bp were designed; For 4.9 kb gene second set of two overlapping primers SSL1, with forward primer SSL1F from position 130 bp and reverse primer SSL1R from position 2419 bp; SSL2 with forward primer SSL2F from position 2229 bp and reverse primer SSL2R from position 4906 bp were designed**

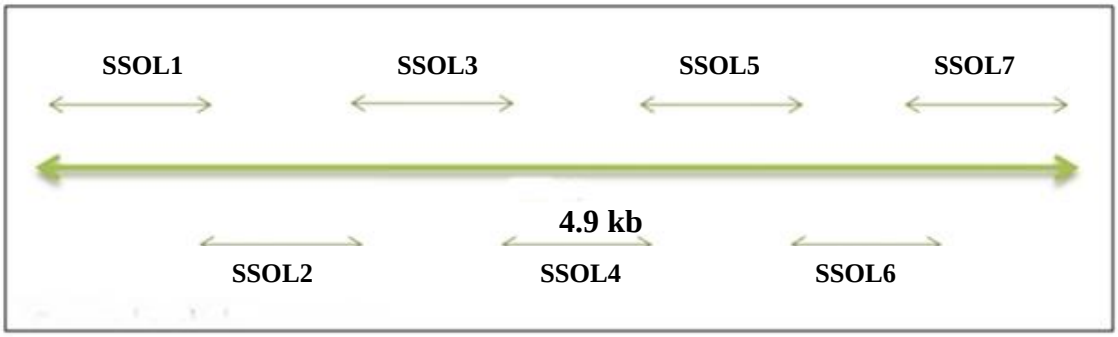

Fig. 9 **Diagrammatic representation of *SSI* gene (4.9 kb) and third set of seven primers SSOL1, SSOL2, SSOL3, SSOL4, SSOL5, SSOL6 and SSOL7 represents corresponding overlapping primers used in amplification and sequencing reaction for wild genotypes**

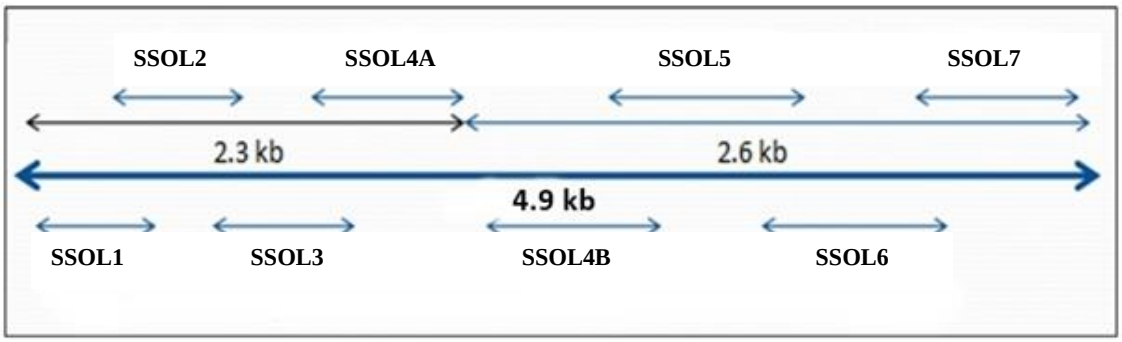

Fig. 10 **Diagrammatic representation of *SSI* gene (4.9 kb) and primers SSOL1, SSOL2, SSOL3, SSOL4A, SSOL4B, SSOL5, SSOL6 and SSOL7 represents corresponding overlapping primers used in sequencing reaction for cultivated genotypes**

5. After shaking, centrifugation at 10,000 rpm for 15 min in microcentrifuge machine was done.
6. The supernatant was separated from the tissue debris in fresh 1.5 ml centrifuge tube using pipette and RNase treatment was given by adding 10 µl of RNase to each tube, and incubated at 37°C for 30 min.
7. To this solution, 800 µl of chilled isopropyl alcohol was added, inverted gently and incubated at 4°C for 15 min.
8. The tubes were again centrifuged at 10,000 rpm for 10 min.
9. The supernatant was removed and 70% ethanol was used to wash the pallet.
10. Then the pellet was air dried and suspended in 50 µl of TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA).
11. The DNA samples were stored at 4°C overnight.

3.3.2.1 Assessment of quantity and quality of DNA

Agarose gel electrophoresis method was used to estimate the quantity and quality of isolated DNA samples. The gel was prepared by dissolving 0.8 g agarose in 100 ml of 0.5 X TBE buffer. The 10X TBE buffer (45mM Tris base, 45mM Boric acid, 1mM EDTA, pH 8.0) was prepared in autoclaved distilled water which was subsequently diluted to 0.5X TBE buffer for the preparation of the gel. The gel mixture was heated to dissolve agarose completely and cooled to 60°C with constant stirring. To this solution, 5 µl of ethidium bromide was added. The agarose solution was poured into gel casting tray for solidification at room temperature (RT). After solidification, 2 µl of each DNA sample was mixed with 8 µl of 1X loading dye and loaded in individual wells of agarose gel. The DNA samples were resolved by running the gel in 0.5X TBE buffer at 5 V/cm. The visualization of gel was done under UV light and captured using UVP Geldoc-It® imaging system.

3.3.3 *In vitro* DNA amplification using Polymerase Chain Reaction (PCR)

In vitro amplification using PCR (Saiki *et al* 1988) was performed in an 8 well PCR strip or 96 well PCR plate in thermocycler (Eppendorf) in a 25 µl reaction. As the amplified PCR product was further used for sequencing purpose hence, hi fidelity highly processive Taq polymerase *Ex Taq* (Takara) was used for PCR amplification. A PCR reaction of 25 µl was made containing 100 ng template DNA (4 µl), 5X PCR buffer plus Mg^{2+} (2.5 µl), 10 mM dNTPs (2.0 µl), 5 pM forward and reverse primers (1.25 µl), 5 units/ µl Ex Taq DNA Polymerase (0.25 µl), and sterile water (13.75 µl). Different temperature profiles were employed for generation of different size amplicons (Table 5 and 6). After completion of the mentioned steps of PCR reaction, mixture was held at 4°C until it was taken out of the thermocycler. The amplified product so obtained was visualized on 0.8% agarose gel.

Table 5: Temperature profile of primers SSL1 and SSL2 used to amplify *SSI*-7B gene in two parts

Steps		Temperature (°C)	Time
I	Initial Denaturation	94°C	5 min.
II	Denaturation	94°C	1 min.
III	Primer annealing	60°C	1 min.
IV	Extension	68°C	3 min. (Repeat II to IV steps for 14 cycles)
V	Denaturation	94°C	1 min.
VI	Primer annealing	60°C	1 min.
VII	Extension	68°C	3 min. (Repeat V to VII steps for 16 cycles with 20 seconds increments per cycle)
VIII	Final extension	72°C	10 min.
IX	Hold	4°C	-

Table 6: Temperature profile of overlapping primers used to amplify *SSI*-7B gene in seven parts

Steps		Temperature (°C)	Time
I	Initial Denaturation	95°C	5 min.
II	Denaturation	95°C	1 min.
III	Primer annealing	55°C	1 min.
IV	Extension	72°C	1 min. (Repeat II to IV steps for 35 cycles)
V	Final extension	72°C	7 min.
VI	Hold	4°C	-

3.3.4 Fractionation of amplification products

After amplification, 3 µl of 6X loading dye was mixed thoroughly with each of the amplified products in an eppendorf centrifuge. 25 µl of each sample was loaded on 0.8 percent agarose gel using 1.0X TAE buffer. PCR products were resolved by running gel at constant voltage depending upon the size of band. The visualization of gel was done under UV light and captured using UVP Geldoc-It® imaging system. The composition of 50X TAE buffer is shown in Table 7.

Table 7: Composition of 50X TAE buffer (pH 8.0)

Components	Composition
Tris base	242 g
Glacial Acetic acid	57.1 ml
Disodium EDTA	37.2 g
Double distilled water	To make 1 litre volume

3.3.5 Purification of PCR products

The agarose containing the specific DNA band was removed from the gel with a clean, sharp blade and placed into 1.5 ml microcentrifuge tube. As per manufacturer's protocol, QIAquick Gel Extraction Kit (Qiagen®) was used to purify the DNA fragments. The agarose block was weighed (g) and placed into 1.5 ml microcentrifuge tube. To the one volume of gel slice (100 mg ~ 100 µl), three volumes QG Buffer was added and incubated at 50°C for 10-15 min, after that one volume of isopropanol was added and mixed. To bind DNA, the sample was poured into the QIAquick spin column provided with 2 ml collection tube and centrifuge for 1 min at 17,900 x g. After discarding the supernatant, it was placed back into the 2 ml collection tube and 0.5 ml QG Buffer was added to centrifuge at 17,900 x g for 1 min. For washing, 0.75 ml PE Buffer was added and centrifuged for 1 min, further the column was allowed to stand for 2–5 min. For removal of residual buffer it was centrifuged once again for 1 min at 17,900 x g. It was then placed into fresh 1.5 ml microcentrifuge tube and 12-15 µl nuclease free water was added to the center of the QIAquick membrane and incubated at RT for 1 min and again centrifuged for 1 min at 17,900 x g. The concentration of the eluted DNA was estimated by spectrophotometer readings.

3.3.6 Cloning of purified DNA fragments in PCR cloning vector pGEM®-Teasy

pGEM®-Teasy vector (Promega) are prepared by cutting with EcoR V and adding a 3´ terminal thymidine to both ends which are used for cloning PCR products. The efficiency of ligation for PCR product into plasmids was improved due to single 3´T overhangs at the insertion site that prevents recircularization of the vector, thus providing a compatible overhang for PCR products produced by thermostable polymerases. The single deoxyadenosine was added by these polymerases to the 3´ends of the amplified fragments. MCS sites of the pGEM®-T easy vector ar flanked by recognition sites for the various restriction enzymes such as *EcoR* I, *BstZ* I and *Not* I, which provides three single-enzyme digestions for the insert release. The purified DNA fragments were ligated using this vector (Fig. 11) and the ligation product was transformed into *Escherichia coli* DH-5α host strain.

3.3.7 Ligation of PCR DNA fragments into pGEM®-Teasy vector

The ligation of purified PCR amplified fragments into cloning vector was performed in a 1.5 ml microfuge tube with 10 µl of ligation reaction and was incubated overnight at 4°C as given in Table 8.

Molar ratio of PCR product: vector and concentration of insert DNA to be used required optimization depending on insert size. For calculation of the approximate amount of PCR product (insert) for the ligation reaction to occur, following equation was carried out:

$$\frac{ng\ of\ vector \times kb\ size\ of\ insert}{kb\ size\ of\ vector} \times insert{:}\ vector\ molar\ ratio = ng\ of\ insert$$

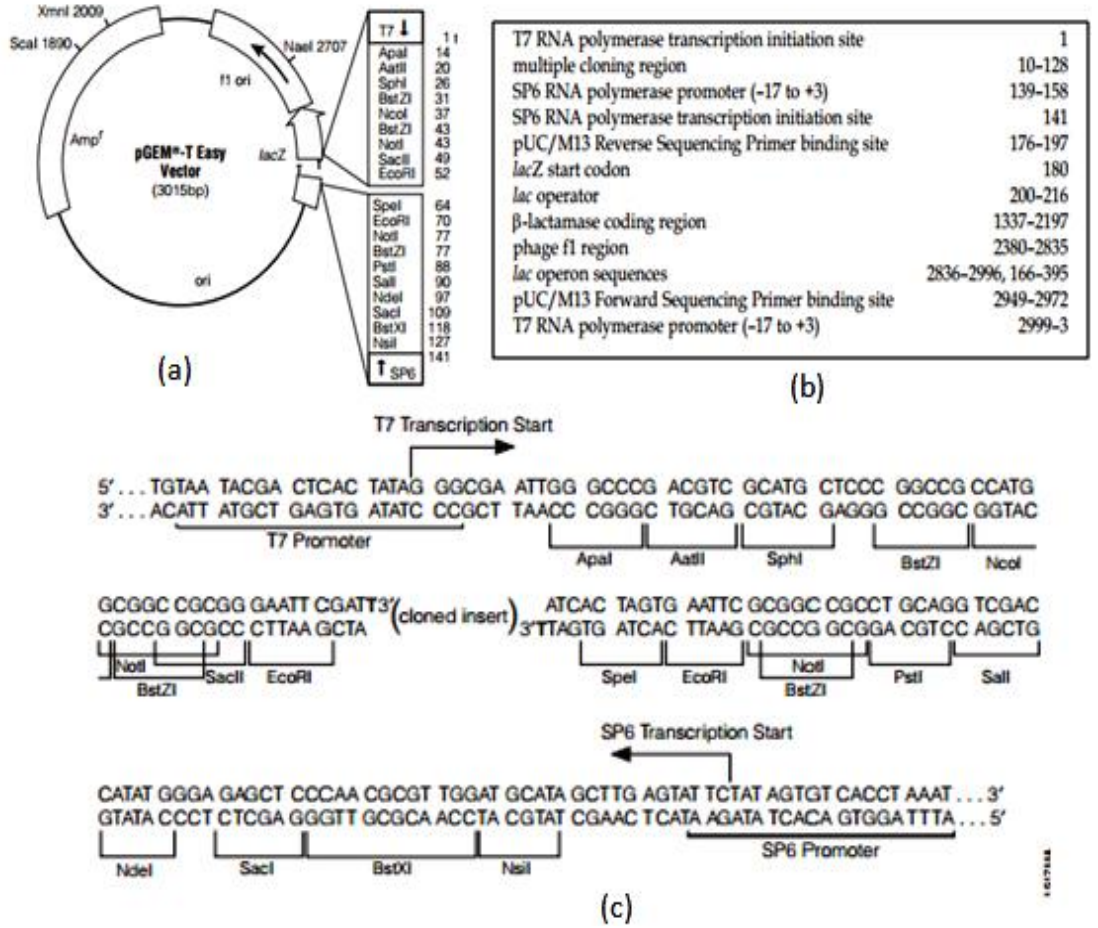

Fig. 11 (a) pGEM®-Teasy vector circle map (b) pGEM®-Teasy vector sequence reference points (c) The promoter and multiple cloning sequence

Table 8: Composition of ligation reaction

Component	Volume (µl)
pGEM®-T Easy Vector (50ng)	1.0
Purified PCR fragment (Insert)	3.0
2X Rapid Ligation Buffer	5.0
T4 DNA ligase (3 Weiss units/µl)	1.0
Total	10

3.3.8 Preparation of medium for transformation of ligation reaction into host *E. coli*, for isolation of recombinant clones

To make the Luria Bertani medium all components (Table 9) were dissolved in water (800 ml), pH adjusted to 7.0 and final volume was made to 1000 ml. For LB-agar medium, LB-broth was supplemented with bacteriological agar at 1.6 percent level. Both the LB-broth and LB-agar were autoclaved at 15 psi of steam for 15 min, before use.

Table 9: Composition of LB (Luria Bertani) medium (1000 ml)

Component	Amount (g/l)
Bacto Tryptone	10
Bacto Yeast extract	5
NaCl	10
Agar	16
Distilled water to make	1000 ml
pH adjusted to	7.0

2M Mg^{2+} stock was prepared by dissolving 20.33g $MgCl_2.6H_2O$ and 24.65g $MgSO_4.7H_2O$ in double distilled water, making final volume to 100 ml and the stock solution obtained was filter sterilized.

For preparing SOC media, Bacto®-tryptone, Bacto®-yeast extract, NaCl and KCl was added to distilled water (97 ml) and dissolved with constant stirring (Table 10). 2M Mg^{2+} stock and 2M glucose was added to make final concentration of 20mM after autoclaving and cooling to room temperature. Final volume was made to 100 ml (pH 7.0) and the media was filtered by 0.2 µm unit.

Table 10: Composition of SOC medium (100 ml)

Component	Amount
Bacto Tryptone	2g
Bacto Yeast extract	0.5g
NaCl (1M)	1ml
KCl (1M)	0.25ml
Mg2+ stock, filter sterilized (2M)	1ml
Glucose, filter sterilized (2M)	1ml
Distilled water to make	100 ml

3.3.8.1 Use of Ampicillin in LB broth and preparation LB-Agar-Ampicillin plates

The cloning plasmid vector pGEM®-Teasy contained Ampicillin-resistance gene (Amp^R) as a positive selectable marker gene. Therefore, for selective growth of *E. coli* cells, which harboured the cloning vector (recombinant or non-recombinant), ampicillin was consistently included in LB-broth and LB-agar at a level of 100 µg/ml. For use in liquid media, 100 ml of LB-broth was supplemented with 100 µl of ampicillin stock solution (100 mg ampicillin-sodium salt/ml sterilized Milli-Q water). For making solidified media, 100 ml of LB-agar medium was melted and cooled to ~55°C. The media (30-35 ml) mixed with 100 µl of ampicillin stock solution was then poured into 85mm petri dishes, and solidified at RT.

3.3.8.2 Preparation of LB-ampicillin agar plates for Blue/White selection of recombinant clones

The MCS in the plasmid vector pGEM®-Teasy was located in the β- galactosidase gene under the transcriptional control of *LacZ* promoter, to serve as screenable marker gene for recombinant plasmids in the presence of substrate X-gal (5- bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and inducer IPTG (isopropyl-β-D-thiogalactopyranoside) in the medium. Therefore, the solidified LB-Ampicillin agar medium in plates was uniformly smeared upon with 20 µl of X-gal solution (50mg/ml in N,N- dimethylformamide). The X-gal solution was prepared by dissolving 100 mg of 5-bromo-4-chloro-3-indolyl-b-D-galactoside in 2 ml N,N´-dimethyl-formamide, covered with aluminum foil and stored at -20°C.

3.3.9 Preparation of competent cells of *E. coli* strain DH-5α

E. coli DH-5α strain was used to prepare competent cells with a transformation efficiency of 10^5 to 10^6, using calcium chloride method, as described by Cohen *et al* (1972). A single colony was isolated from overnight freshly grown colonies on Luria-Bertani (LB) agar plates and transferred into 100 ml LB medium in a 1 litre flask by incubating at 37°C until the OD_{600} reached 0.3-0.4 (2 to 3 hours). Cells were transferred to polypropylene tubes (50 ml) cooled to 0°C and collected by centrifugation. Supernatant was removed and tubes were inverted for 1 to 2 min to drain away traces of media. Cell pellets were then suspended in ice-cold 0.1M calcium chloride (10 ml per 50 ml original culture) and cooled to 0°C. After centrifugation and draining, cells were suspended in ice-cold solution of 0.1M calcium chloride and 15 percent glycerol (1 ml/50 ml original culture) and resuspended in 1.5 ml microtubes. The competent cells prepared were stored at -80°C.

3.3.10 Transformation and selection of recombinant clones

For transformation, 80 µl of competent cells mixed with 5 µl of ligation reaction product was added in 1.5 ml polypropylene tube and was allowed to stand in ice for 30 min. The tube was transferred to a thermostat maintained at 42°C and incubated for 50-55 seconds. The tube was then quickly transferred to ice and allowed to stay for 20 min. After vector's

insertion in bacteria, 200 µl SOC media was added and incubated at 37°C for 90-120 min at 150 rpm and spread on a LB plate containing ampicillin and X-gal. The plates were incubated overnight at 37°C (upside down) and observed for growth of blue (non-recombinant) and white (recombinant) coloured colonies. Colony PCR was performed using the desired primers and the product was resolved on 1.2 percent gel to identify the positive clones which were then used for the plasmid DNA isolation followed by sequencing.

3.3.11 Isolation of recombinant plasmid DNA from recombinant clones of *E. coli*

Individual recombinant (white) clones from individual plates were picked up with a sterile tip, inoculated into a well of 96-deepwell culture plate containing 1.2 ml of LB-Ampicillin broth and the plate was incubated (covered with breathable seal) for 20-24h at 37°C at 220 rpm. The plasmid isolation was done using R.E.A.L® prep 96 (QIAGEN). As per the protocol, after culturing of the selected white colonies, bacterial cells were harvested by centrifugation for 5 min at 1500 x g. Media was removed by inverting the plate after removing the seal. Each bacterial pellet was resuspended (leaving no cell clumps) in 300µl of R1 Buffer (RNase added), tapping of block was done and vortexed. In each well 300µl of R2 Buffer was added and the block was taped and mixed by inverting 10 times and incubated at RT for 5 min. At the end of incubation the lysate appeared viscous and free from cell clumps. In each well 300µl of R3 Buffer was added and the block was taped and mixed by inverting 10 times. QIAfilter 96 plate was placed on the new 96-deepwell plate. Eight channel pipette was used to transfer the lysate to the QIAfilter 96 plate placed over the new 96-deepwell plate. This assembly was joined with paper tape to avoid separation during rotation in a centrifugation. Centrifugation was done at 1500 x g for 10 min. After centrifugation the clear lysate get collected into a new 96-deep well plate placed under the QIAfilter to which 700 µl of isopropanol was added. The block was taped and inverted for 3 times. The pellet of plasmid DNA was obtained by centrifugation at 2500 x g for 15 mins at RT. The supernatants were removed by inverting the block over upside down onto a tissue paper. The 500 µl of 70% ethanol was used to wash DNA pellet. For reconcentration of the pellet, it was centrifuged at 2500 x g for 2 min. Pellets were air dried for 15 mins to evaporate the residual ethanol. To these plasmid DNA pellets, 25 µl of nuclease free water was added. Isolated plasmid after pellet get dissolved, was quantified over 1.2 percent agarose gel and the quantity was estimated for sequencing reaction.

3.3.12 Preparation of glycerol stocks

Selected white clones were picked using a sterile pipette tip or toothpick and inoculated in well of 96-deepwell plate containing 800 µl of LB-Ampicillin broth (per well). The inoculated plates were cultured for 12-18h at 37°C under shaking conditions (150 rpm) so that broth is moderately turbid from bacterial growth. To this culture plates 200 µl of 80 percent glycerol was added and vortexed gently for few seconds to mix the glycerol and

culture. These plates were stored at -80°C as the future source of the desired transformants.

3.4 Identification of novel haplotypes for starch synthase gene in selected wheat germplasm.

3.4.1 Sequencing of candidate gene region

The sequence data on nucleotide sequence of the *SSI*-7B DNA regions was produced using separate sequencing reactions for forward and reverse primers with Big Dye Terminator v3.1 (ABI 3730xl Sequencer) were outsourced through facility of Eurofins Genomics, India.

3.4.2 Analysis of nucleotide sequences

Chromatograms of sequences were analyzed manually. The sequences were extracted from chromatogram files using CHROMAS Lite 2.1.1 (http://technelysium.com.au/). Contigs were generated using the DNA Baser v4.23.0 (http://www.dnabaser.com/) which also helps in automatic detection and trimming of low quality regions of the sequences. Multiple sequence alignment, using the gene sequence retrieved from database earlier, as reference, was carried out using ClustalX 2.1.1. Based on the alignment candidate SNPs and InDels were predicted. The candidate SNPs were then manually curated by analyzing and comparing chromatogram files with ClustalX alignment files. Only those parts of reads where Q-value equaled or exceeded 20 were considered for finding SNPs. The bases which had Q-value below 20 were not considered to be SNPs.

3.4.3 Prediction of gene structure and amino acid sequences

Primary structure and amino acid sequences were predicted in different wild and cultivated genotypes of wheat using Artemis tool (http://www.sanger.ac.uk/science/tools/artemis). The amino acid substitution between wheat genotypes were identified visually with respect to reference (Fig. 12), using the multiple alignment viewer JalView 2.0 (www.jalview.org).

MAATGVGAGCLAPSVRLRADPATAARASACVVRARLRRVARGRYVAELSREGPAARPAQQQ
QLAPPLVPGFLAPPPPAPAQSPAPTQPPLPDAGVGELAPDLLLEGIAEDSIDSIIVAASEQDSEIMD
AKDQPQAKVTRSIVFVTGEAAPYKSGGLGDVCGSLPIALAARGHRVMVVMPRYLNGSSDKNY
AKALYTAKHIKIPCFGGSHEVTFFHEYRDNVDWVFVDHPSYHRPGSLYGDNFGAFGDNQFRYT
LLCYAACEAPLILELGGYIYGQNCMFVVNDWHASLVPVLLAAKYRPYGVYRDSRSTLVIHNLA
HQGVEPASTYPDLGLPPEWYGALEWVFPEWARRHALDKGEAVNFLKGAVVTADRIVTVQGYS
WEVTTAEGGQGLNELLSSRKSVLNGIVNGIDINDWNPTTDKCLPHHYSVDDLSGKAKCKAELQ
KELGLPVREDVPLIGFIGRLDYQKGIDLIKMAIPELMREDVQFVMLGSGDPIFEGWMRSTESSYK
DKFRGWVGFSVPVSHRITAGCDILLMPSRFEPCGLNQLYAMQYGTVPVVHGTGGLRDTVETFN

Fig. 12 Protein sequence of *SSI*-7B gene reference

3.4.4 Protein structure prediction

To develop the tertiary structures of *SSI*-7B protein in *Triticum aestivum* wild and cultivated genotypes, a template protein of *SSI* was detected using the PDB database on the

basis of sequence homology. Homology model was developed using the MODELLER 9.17 program (https://salilab.org/modeller/9.17), starting from the pairwise query-template alignment. Molecular graphics and analysis were carried out with the UCSF Chimera tool (Pettersen *et al* 2004). It helps in viewing the protein structure and comparison of protein from two different sources. Catalytic sites in the protein were determined with the help of PDBsum (www.ebi.ac.uk/pdbsum/) online program. Structural domain of *SSI* protein was analyzed using Pfam (http://pfam.xfam.org/).

CHAPTER IV

RESULTS AND DISCUSSION

The present investigation was carried to identify allelic variants for *SSI* gene in 20 selected wild and cultivated genotypes of wheat (Table 1). Wild wheat includes one accession each of *Triticum monococcum* and *T. boeoticum* (A genome), three accessions of *Aegilops speltoides* (B genome), two accessions of *Aegilops tauschii* (D genome) and two accessions of *T. dicoccoides* (AB genome). Cultivated wheat includes one durum wheat (AABB) and ten hexaploid wheat cultivars including two winter wheat and eight spring wheat genotypes (AABBDD). The results obtained have been presented in this chapter under the following headings:

4.1 Identification of candidates for starch synthase gene using bioinformatics tools

4.1.1 Identification of *SS* gene from NCBI/EMBL database

NCBI and EMBL databases were searched for nucleotide sequences of full length *SS* gene. A total 82 nucleotide sequences of gene were retrieved from these databases. These sequences consisted of four different isoforms of gene (*SSI, SSII, SSIII* and *SSIV*). Presence of four different isoforms of *SS* gene have already been reported as *SSI* (Li *et al* 1999a, Peng *et al* 2001, Li *et al* 2013, McMaugh *et al* 2014), *SSII* (Li *et al* 1999b, Gao and Chibbar 2000, Shimbata *et al* 2005, Huang and Babel 2012), *SSIII* (Li *et al* 2000, Pan *et al* 2011) and *SSIV* (Leterrier *et al* 2008). Different sequences in each of isoform were clustered to remove common sequences using CD-HIT Suite tool at 95% identity (Table 11). As a result, only 24 sequences were selected which includes ten sequences of *SSI* isoform with gene length ranging from 500-3006 bp, eight sequences of *SSII* isoform with gene length ranging from 2025-7010 bp, four sequences of *SSIII* isoform with gene length ranging from 735-5346 bp and two sequences of *SSIV* isoform with gene length ranging from 3386-7137 bp. Maximum number of sequences were obtained for *SSI* isoform followed by *SSII*, *SSIII* and *SSIV* isoforms. Sequences which do not fall into selected clustering criteria were discarded.

Table 11: Number of candidate *SS* genes as obtained from NCBI and EMBL databases and selected sequences after clustering

Gene	Number of *SS* sequences selected from NCBI and EMBL databases	Number of *SS* sequences after clustering using CD-HIT suite at 95% identity
SSI	25	10
SSII	22	8
SSIII	30	4
SSIV	5	2
Total	**82**	**24**

The selected 24 sequences were used as query for online BLAST against IWGSC (Ensembl Plants) *Triticum aestivum* database as subject. 44336 sequence hits were obtained of which 619 sequences were selected for having lowest e-value ($5.7E^{-39}$- 0), highest bit score (523-2946), multiple hits for single entry and sequences having functional annotation. Functional annotation was done for genes having biological process (starch biosynthetic process), cellular component (amyloplast and chloroplast) and/or molecular function (starch synthase activity) and gene sequences showing these related functions were selected. Selected sequences includes maximum number of 221 sequences for *SSII* isoform, 183 sequences for *SSI* isoform, 123 sequences for *SSIII* and 92 sequences for *SSIV* isoform of gene (Table 12).

Table 12: Compiled list of selected 24 representative *SS* gene sequences, their hits on IWGSC database and selected sequences

Gene	Number of *SS* representative sequence selected from NCBI/EMBL database	Total hits against IWGSC Ensembl Plants database	Number of Selected sequences
SSI	1	22	3
	2	114	7
	3	30	15
	4	11579	3
	5	121	41
	6	60	21
	7	859	45
	8	33	6
	9	158	22
	10	100	20
SSII	11	75	21
	12	74	19
	13	15370	36
	14	12422	34
	15	432	41
	16	65	20
	17	61	24
	18	84	26
SSIII	19	212	55
	20	53	10
	21	50	10
	22	176	48
SSIV	23	116	45
	24	2070	47
	Total	**44336**	**619**

4.1.2 Identification of *SS* gene from CDS database

CDS database is also searched for representative sequences of *SS* gene and 16 full length coding sequences (CDS) were retrieved from TriFLDB database. These sequences were clustered into already selected 24 nucleotides sequences (selected from NCBI/EMBL database) of *SS* gene using 99% identity and 24 common sequences were selected.

These common 24 sequences between NCBI/EMBL and CDS databases were selected and used as query and standalone (offline) blast was done against the cDNA IWGSC (Ensembl Plants) *Triticum aestivum* database as subject. A total of 214 sequence hits were obtained. Here also maximum number of 120 hits was obtained for *SSII* isoform of gene while *SSI*, *SSIV* and *SSIII* isoforms of gene have 40, 31 and 23 hits respectively (Table 13).

Table 13: List of different isoforms of *SS* gene selected from CDS database and number of sequences selected after standalone blast against IWGSC cDNA database

Isoform selected from CDS database	Hits obtained against cDNA database
SSI	40
SSII	120
SSIII	23
SSIV	31
Total	**214**

Table 14: List of accession number of 16 selected sequences of *SS* gene common between genomic and CDS databases and their chromosomal location as obtained after blast against IWGSC database

Gene	Selected sequences acc no.	Chromosome Location
SSI	AF091803	7A,7BS, 7D
SSI	AJ292521	7A, 7BS, 7D
SSI	AJ292522	7A, 7BS, 7D
SSII	AB201445	7A,7B,7DS
SSII	AB201446	7A,7B, 7DS
SSII	AB201447	7A,7B,7DS
SSII	AF155217	7A, 7B,7DS
SSII	AJ269502	7A, 7B,7DS
SSII	AJ269503	7A, 7B,7DS
SSII	AJ269504	7A, 7B,7DS
SSII	EU307274	1A, 1B,1D
SSII	EU333947	6A, 6B, 6D
SSIII	AF258608	1A, 1B, 1D
SSIII	EU333946	2A, 2B, 2D
SSIV	AY044844	1A, 1B,1D
SSIV	DQ400416	1A, 1B,1D

These combined 24 sequences were again blasted against genomic (gDNA) *Triticum aestivum* database as subject and total of 266 sequences were obtained. Finally 16 sequences common in both genomic and CDS databases were selected (Table 14) for having maximum query coverage and score. Sequences with single hit were preferred but mostly single hits were partial genes. Thus full gene sequences having location on three homoeologous chromosomes rather than on altogether different chromosomes were selected. Available literature validating full gene was also taken into consideration while selecting gene sequences. Location of *SS* gene has been reported by many workers in wheat as *SSI* (Li *et al* 1999a, Peng *et al* 2001) and *SSII* (Li *et al* 1999b) have been reported to be located on homoeologous chromosome 7 (7A, 7B and 7D) while *SSIII* (Li *et al* 2000) and *SSIV* (Leterrier *et al* 2008) isoforms found to be present on homoeologous group 1 (1A, 1B and 1D).

4.1.3 Phylogenetic and molecular evolutionary analysis

Phylogenetic and molecular evolutionary analysis was conducted using Phylogeny.fr tool using multiple sequence alignment files of 16 selected full length CDS sequences. Tree was generated using neighbour joining method (Fig. 13) and *SS* sequences falls into two main groups. Group I contain *SSI, SSII* and *SSIV* isoforms while group II has only two sequences of *SSIII* isoform indicating that *SSIII* is divergent of all other isoforms. Group I was a large group containing three *SSI*, nine *SSII* and two *SSIV* isoforms sequences respectively and is further divided into two subgroups, subgroup I contain sequences of *SSIV* isoform only while subgroup II contain *SSI* and *SSII* isoforms. Thus *SSIV* is divergent while *SSI* is closely related to *SSII*. Subgroup II was further divided into two clusters one containing *SSI* and other *SSII*. Close relationship between sequences of *SSI* and *SSII* isoforms of wheat have earlier reported by Leterrier *et al* (2008).

Different molecular evolutionary and sequence analysis showed that though each *SS* isoform have evolved from common ancestor, but they are distinct from each other (Leterrier *et al* 2008; Nougue *et al* 2014). A phylogenetic analysis of *SS* sequences conducted by Liu *et al* (2015) in maize indicated that *GBSSI, SSI* and *SSII* isoforms are closely related and formed one cluster while *SSIII, SSIV* and *SSV* isoforms formed second cluster. In wheat also, a study by Leterrier *et al* (2008) supported that close relationship of *SSI, SSII* and *GBSS* was due to identical length of glycosyl transferase group 1 (GT-1) domain of gene which is different in *SSIII* and *SSIV*. Phylogenetic analysis in present study indicated that approach used for selecting different sequences from databases was accurate and yield results in accordance with database available.

4.1.4 Selection of gene accessions of each *SS* isoform common in both genomic and cDNA databases

Three gene sequence IDs for each isoform of *SSI, SSII, SSIII* and *SSIV* (total 12 sequences) were finally selected on basis of maximum length coverage of gene and having

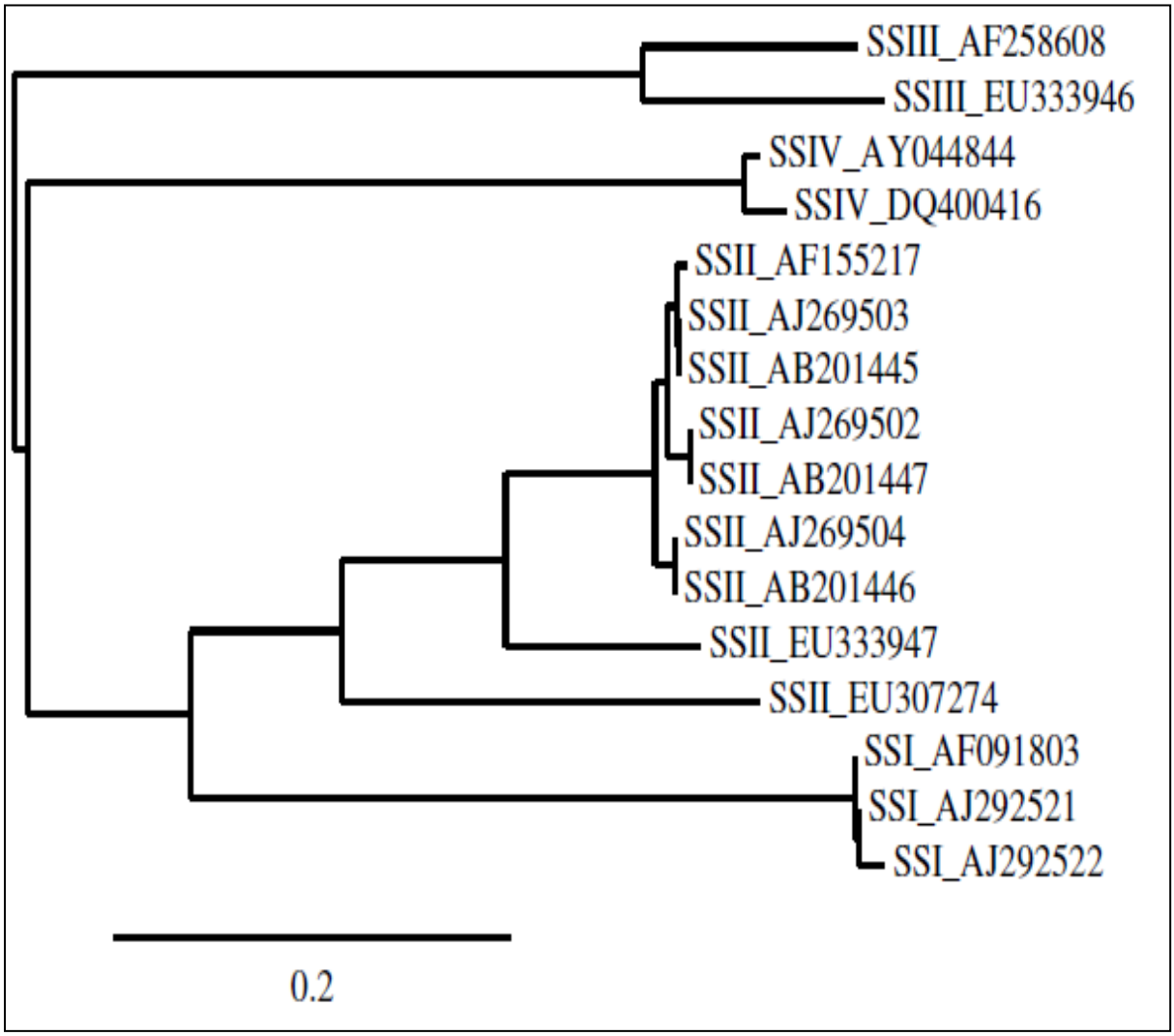

Fig. 13 Phylogenetic tree representing divergence analysis of full length cDNA sequences using neighbour joining method of *SS* gene (Phylogeny.fr tool)

99% or more identity with query sequence for further *in-silico* gene expression profile analysis (Table 15). Each of selected sequences of *SSI* and *SSII* have alleles on homoeologous chromosome 7A, 7B and 7D while selected sequences of *SSIII* and *SSIV* have alleles on each of homoeologous chromosome 1A, 1B and 1D.

Table 15: List of selected gene sequence and their IDs

Gene	Selected gene accession number	Selected gene sequence IDs
SSI	AF091803	Traes_7DS_5159E3934.1
		Traes_7BS_6135B1D85.1
		Traes_7AS_70ED86B15.1
SSII	AB201445	Traes_7DS_E6C8AF743.2
		Traes_7AS_53CAFB43A.2
		Traes_7BS_7BEAF5EC0.1
SSIII	AF258608	Traes_1AS_83A34BFC8.1
		Traes_1BS_AF95964AE.1
		Traes_1DS_1B53199CF.1
SSIV	AY044844	Traes_1AL_5E9E6239D.1
		Traes_1BL_709B74768.2
		Traes_1DL_B0409DEFD.2

4.1.5 Confirmation of selected sequence using gene expression profile study

For functional characterization of selected 12 sequences of different isoform of *SS* gene (three from each *SS* isoform), gene expression profiles of selected sequences were verified using WheatExp database (Pearce *et al* 2015). Two different expression databases were selected; one has collection of transcript from different tissues (grain, leaf, root, spike and stem) of wheat plant and second database has collection of transcripts from seeds at different interval of heat and drought stress conditions. For *SSI* and *SSII*, the expression of three homoeologous genes (7A, 7B and 7D genomes) were more in grain and spike as compared to leaf, root and stem (Fig. 14). The expression profiles of *SSIII* isoform on chromosome 1A, 1B and 1D were more in grain, spike and stem than in leaf and root tissues while in *SSIV* isoform, the expression was more in grain, leaf, spike and stem than in root in all the three homoeologous genes (1A, 1B and 1D). So *SSI* and *SSII* isoforms could be good candidates to study the allelic variation in *SS* gene. But *SSI* isoform was selected for present study because it is major gene in wheat starch synthesis and contribute 60-70% of the *SS* activity (Cao *et al* 1999, Fujita *et al* 2006, Jeon *et al* 2010 and Zeeman *et al* 2010).

The B-genome homoeologue of *SSI* gene (accession Traes_7BS_6135B1D85.1) contributes the majority of transcripts to the overall expression as compared to A and D

homoeologous under heat and drought conditions (Fig. 15). Similarly, this accession gives more expression in grain as compared to other tissue. As the 7B homoeologue-specific transcript profile of *SSI* gene have more tissue specific (in grain) and stress specific expression (under heat and drought stress), so *SSI*-7B isoform was selected as candidate gene for present study.

SS isoforms distribution varies between species, tissue, and developmental stages present within the plastid (Commuri and Keeling 2001). Wang *et al* (2014) reported higher *SSI* transcript level than *SSII* and *SSIII* during early seed development stage (6 DAP), but remain steady throughout endosperm development (15 DAP), concluding that *SSI* is constitutively expressed at all stages of seed development (Park *et al* 2012) making it the major isoform among the starch synthase isoforms in cereal crops (Cao *et al* 1999). *SSI* activity exhibited higher thermostability (35 °C) than *SSII* according to Keeling *et al* (1994). Li *et al* (2000) reported that *SSI, SSII* and *SSIII* isoforms have expression in the grain tissue but expression of *SSI* isoform was highly specific in the endosperm.

4.2 Primer designing and cloning of starch synthase gene

4.2.1 Primers designing to capture target homoeologous *SSI* gene

The nucleotide sequence of the target *SSI*-7B gene was used to design gene specific primers using the software PerlPrimer (Marshall O 2004) tool. Size of selected gene is 5156 bp, but primers were designed to amplify gene starting from nucleotide position 130 and ending on position 4906. Primers were not able to design for complete gene as few initial and last nucleotides were not available (NNNN….) in selected candidate gene. Three sets of primers were designed. First set of primer pair (SFLF and SFLR) was designed to amplify full gene from nucleotide position 130 to position 4906. This primer could not amplify gene in any of the wheat genotypes under study. Thus, second set of primers were designed, which include two overlapping primer pairs SSL1 and SSL2. One primer pair SSL1F - SSL1R amplify region of size 2289 base pairs starting from nucleotide positions 130 and ending on nucleotide position 2419 while second primer pair SSL2F - SSL2R amplify region of size 2677 base pairs between nucleotide positions 2229 and 4906. There is an overlap of 190 nucleotides in the center of gene. SSL1 and SSL2 showed amplification at an annealing temperature of 60ºC (Table 2). These primers could amplify the expected fragments in cultivated hexaploid and tetraploid wheats only but not in wild diploid and tetraploid wheat. Thus third set of seven overlapping primer pairs were designed for amplification of gene in seven small parts as SSOL1 (130 – 879 nucleotides), SSOL2 (700 – 1555 nucleotides), SSOL3 (1320 – 2164 nucleotides), SSOL4 (1970 – 2864 nucleotides), SSOL5 (2726 – 3598 nucleotides), SSOL6 (3410 – 4290 nucleotides) and SSOL7 (4152 – 4906 nucleotides) with the amplicon length ranging from 830 to 900 bases. Annealing temperature for these primers was 55ºC (Table 3). These same set of seven overlapping primers were also used for

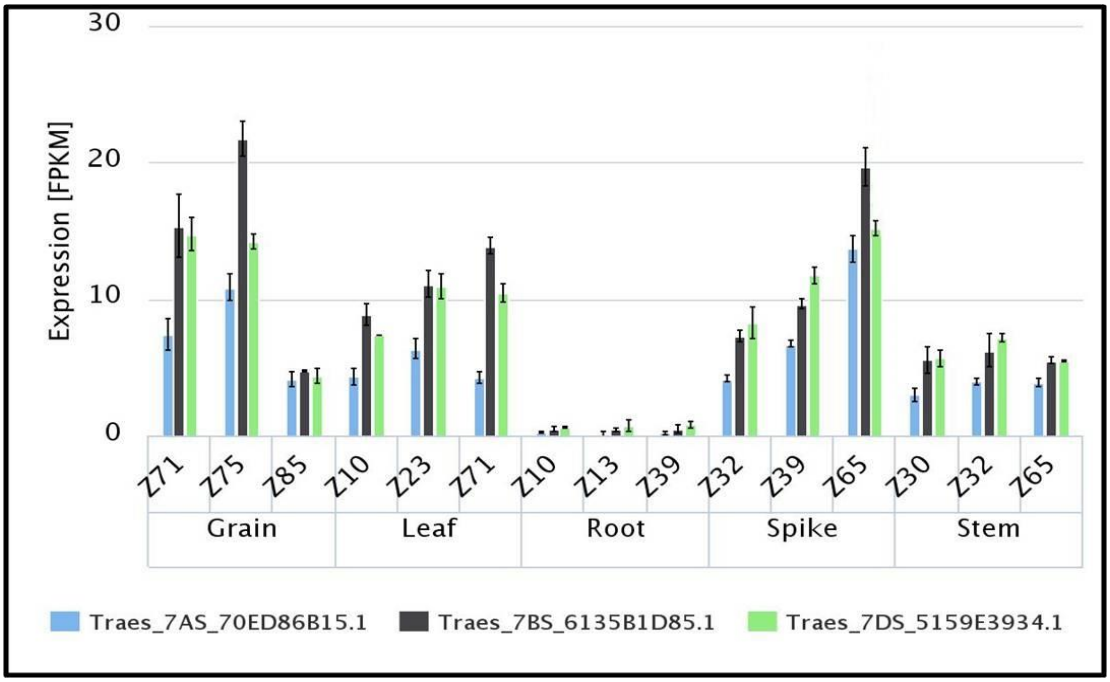

Fig. 14 Dynamic expression profiles of *SSI* homoeologous (7A, 7B and 7D) gene in different wheat tissues

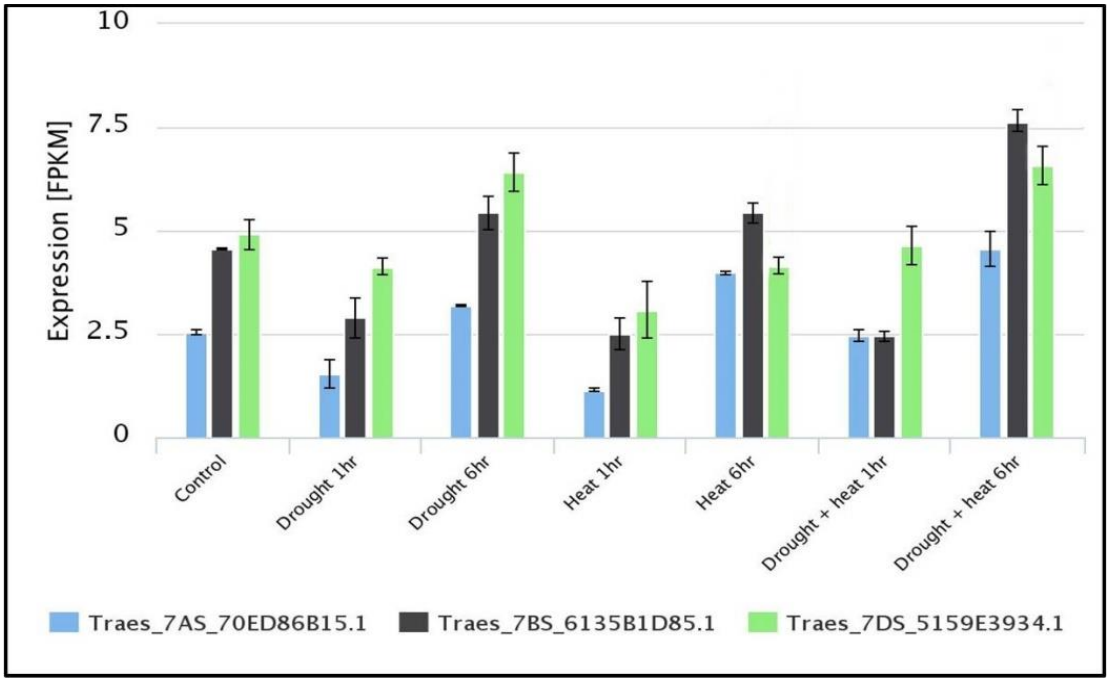

Fig. 15 Dynamic expression profiles of *SSI* homoeologous (7A, 7B and 7D) gene under heat and drought stress conditions

sequencing small PCR amplified fragments generated by primer set SSOL1-SSOL7 and also larger PCR fragments (2289 and 2677 base pairs) generated by primers SSL1 and SSL2. Only primer SSOL4 could not be used in sequencing of larger amplicons of 2289 and 2677 base pairs in cultivated genotypes as this primer starts from nucleotide position 1970 and ended on nucleotide position 2864. These positions included the middle overlap region between SSL1 and SSL2 primers. Therefore, another set of two overlapping primers were designed namely SSOL4A (1970 – 2419 nucleotide) and SSOL4B (2230 – 2864 nucleotide) such that it divide SSOL4 primer into two parts as shown in Table 4. Generation of gene specific primers and their use in sequencing has been reported in many studies. Huang and Babel (2010) synthesized gene specific primers for *Wx* and *SSII* genes with amplicon length ranged from 750-1657 bp from three homoeologous of hexaploid wheat which have applications in detecting sequence diversity and association mapping.

4.2.2 *In vitro* DNA amplification using Polymerase chain reaction (PCR)

The genomic DNA was isolated from young leaves and quantified at concentration of 200-600 ng/µl on agarose gel. No wheat lines could be amplified by first set of primer SFL. Second set of primer pairs SSL1 and SSL2 could amplify fragments of 2289 and 2677 bp in 11 hexaploid wheat genotypes only (Fig. 16 and 17). Remaining two wild tetraploid lines (*T. dicoccoides acc. pau14801* and *T. dicoccoides acc. pau7107*) and seven diploid wild wheat lines could not be amplified by SSL1 and SSL2 primers. So these genotypes were amplified using the third set of seven primer pairs-SSOL1, SSOL2, SSOL3, SSOL4, SSOL5, SSOL6 and SSOL7 that showed the expected amplicon size of 830, 855, 844, 894, 872, 880 and 870 bp respectively (Fig. 18). PCR is simple and routinely used method for characterization of starch biosynthesis genes e.g. PCR amplification of *SSI* cDNA sequence with 2591 bp amplicon (Li *et al* 1999a), *SSIII* cDNA with 5346 bp amplicon length (Li *et al* 2000) and *SSIV* cDNA having amplicon length of 3386 bp (Leterrier *et al* 2008) have been reported.

4.2.3 Cloning of the amplified candidate gene regions

The amplified candidate gene fragments were purified from agarose gels and its quantity was confirmed by running a subsample of 3µl purified sample in 1 percent agarose gel. The gel showed the presence of 500-1200 ng/µl DNA concentration in different samples from cultivated and wild genotypes. The purified fragments were individually cloned and white recombinant clones were isolated on LB-Amp-Xgal agar plates (Fig. 19).

Cloning of candidate gene fragments was done in 13 wheat genotypes including cultivated hexaploid and tetraploid wheats and wild tetraploid wheat genotypes only. In diploid wild wheat genotypes namely *T. monococcum, T. boeoticum, Ae. speltoides* and *Ae. tauschii,* PCR product was sequenced directly without cloning. This is because the hexaploid genotypes expected to carry three copies of gene on homoeologous chromosome and tetraploid carrying two copies while diploid species are expected to amplify single copy of

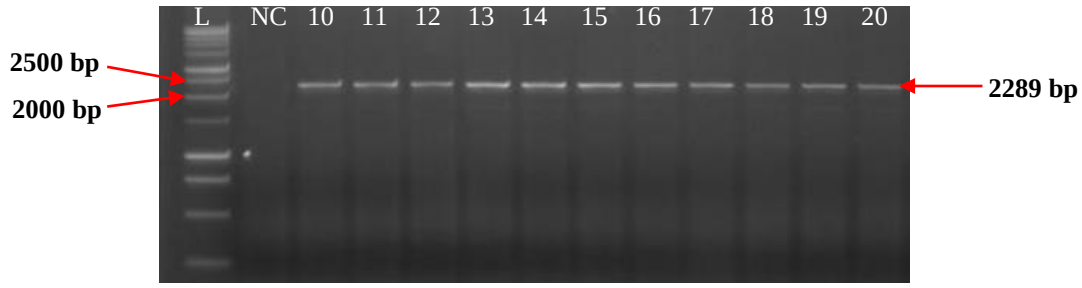

Fig. 16 **PCR amplification profile of tetraploid and hexaploid wheat genotypes with SSL1 primer amplifying the *SSI* gene from nucleotides position 130 to 2419**

L refers to 1 kb ladder (Promega, Cat. No. G5711); NC - negative control; 10 - Giza, 11 - MACS9, 12 - Arbon, 13 - Impala, 14 - C306, 15 - C591, 16 - WH542, 17 - PBW343, 18 - Halna, 19 - Raj 3765, 20 - K9644

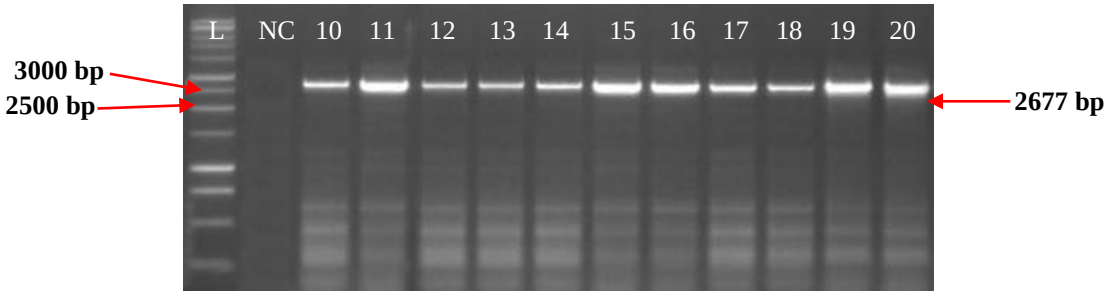

Fig. 17 **PCR amplification of tetraploid and hexaploid wheat genotype using SSL2 primer amplifying the *SSI* gene from nucleotides position 2229 to 4906**

L refers to 1 kb ladder (Promega, Cat. No. G5711); NC- negative control; 10 - Giza, 11 - MACS9, 12 - Arbon, 13 - Impala, 14 - C306, 15 - C591, 16 - WH542, 17 - PBW343, 18 - Halna, 19 - Raj 3765, 20 - K9644

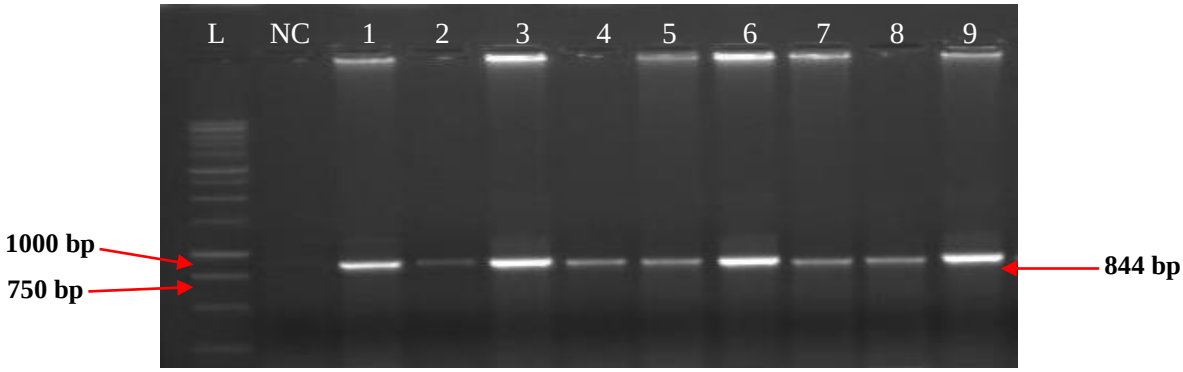

Fig. 18 **PCR amplification profile of wild wheat genotypes using SSOL3 primer pair amplifying the *SSI* gene from nucleotides position 1320 to 2164**

L refers to 1 kb ladder (Promega, Cat. No. G5711); NC - negative control; 1 - *Triticum monococcum acc. pau14087,* 2 - *Triticum boeoticum acc. pau5088,* 3 - *Ae. speltoides acc. pau3809,* 4 - *Ae. speltoides acc. pau15081,* 5 - *Ae. speltoides acc. pauTA1784,* 6 - *Ae. tauschii acc. pau3747,* 7 - *Ae. tauschii acc. pau14102,* 8 - *Triticum dicoccoides acc. pau14801,* 9 - *Triticum dicoccoides acc. pau7107*

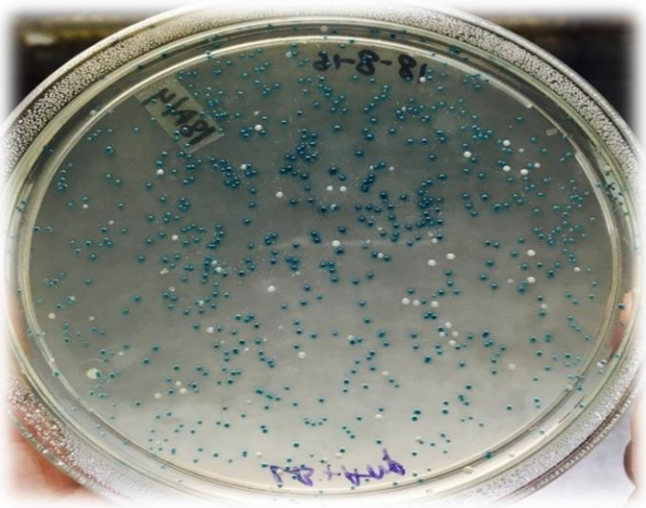

Fig. 19 **Transformed cells selected containing blue and white colonies. White colonies representing transformed clones while blue colonies represent untransformed clones**

gene. Southern hybridization experiment by Jia *et al* (2008) revealed that Calreticulin *(CRT)* gene involved in drought-stressed responses have three copies in hexaploid wheat genome and one copy in its diploid genome. Golovnina *et al* (2010) characterized the *VRN1* gene in wheat responsible for determination of spring-winter growth habit and PCR product in diploid genotypes were directly sequenced without cloning. Cloning of different starch synthase genes in wheat have also been reported by different workers e.g., *SSI* (Li *et al* 1999a and Peng *et al* 2001), *SSII* (Gao and Chibbar 2000), *SSIII* (Li *et al* 2000), *SSIV* (Leterrier *et al* 2008) and *GBSSII* (Vrinten and Nakamura 2000).

4.2.4 Detection of recombinant clones

For critical analysis on the true identity of cloned fragments, the recombinant plasmid DNA was isolated from 10-15 individual recombinant clones (miniprep/midiprep plasmid DNA) for each of the cloned fragments from the selected wild tetraploid and cultivated genotypes and the identity of the true recombinants were established by colony PCR.

4.2.4.1 Colony PCR amplification of Insert DNA from recombinant plasmid DNA

Confirmation of recombinant plasmids carrying *SSI* gene was done by amplifying gene with two out of seven primer pairs (SSOL3 and SSOL5). PCR amplification of the recombinant plasmid DNA with SSOL3 gene-specific primer pair resulted in amplification of fragments of expected size of 844 bp and with SSOL5 primer pair resulted in amplification of fragment of size 872 bp (Fig. 20). The recombinant clones with desired insert size true to the identity of correspondingly cloned amplicon DNA for each of the amplified candidate gene fragments were selected and those which did not amplify any product or amplify a product of undesired length were discarded. Out of the total 621 white colonies, only 180 recombinant clones positive for the *SSI* gene were selected (five for each of 13 wheat genotypes cloned) and rest of the white colonies might have resulted from cells carrying the vector–vector ligation instead of vector–insert ligation. The plasmid DNA of these 180 recombinant clones was isolated using QIAGEN R.E.A.L 96 plasmid kit and its authenticity was confirmed on 1.2 percent agarose gel (Fig. 21).

4.3 Identification of novel haplotypes for starch synthase gene in selected wheat germplasm.

4.3.1 Sequencing of cloned candidate gene fragments

The sequencing grade recombinant plasmid DNA from 180 selected clones of 13 wheat genotypes was purified and custom sequenced through ABI Sequencer 3730xl for both the complementary strands using overlapping gene-specific primers. On the other hand, 49 PCR products of seven diploid wild genotypes are used directly for sequencing. In total, sequencing of 180 clones and 49 PCR products was done.

4.3.2 Generation of contigs

Sequence data of approximately 700-800 bp long reads was obtained using ABI

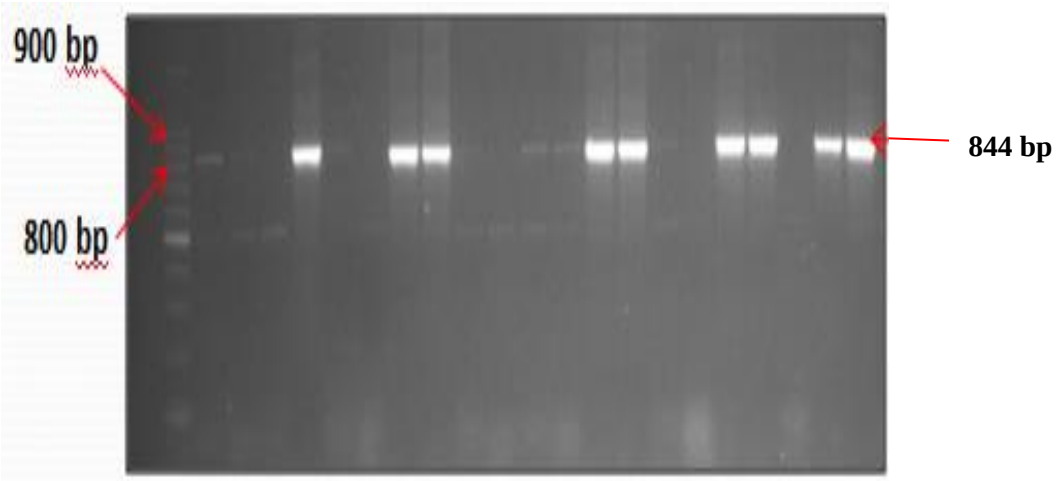

Fig. 20 Colony PCR amplification of positive white colonies containing *SSI*-7B gene using SSOL3 specific primer

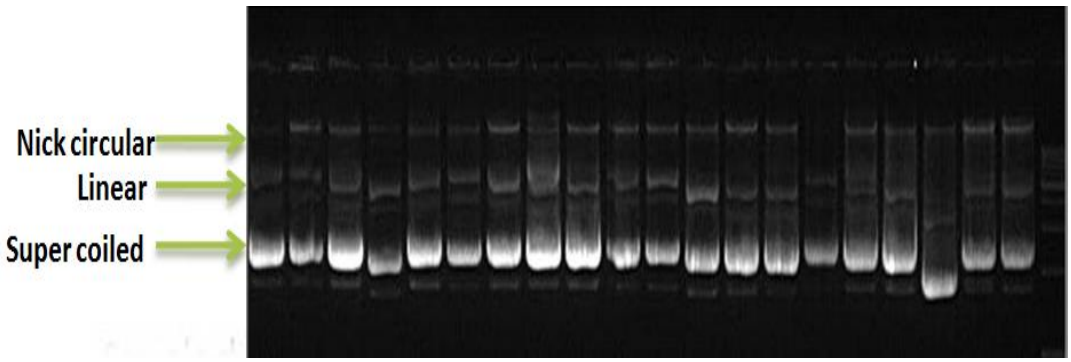

Fig. 21 Confirmation of integrity of recombinant plasmid DNA on 1.2 % agarose gel

3730xl for all genotypes. The sequences were extracted from Chromatograms using Chromas Lite software. In the total of 1118 sequences obtained (each clone was sequenced in both forward and reverse direction) the raw data on the individual clones and for both the complimentary DNA strands were proof-read for selection of high quality bases (Fig. 22). The sequence of the each individual strand was completed by manual aligning of the sequence of one strand with that of the reverse complimentary sequence obtained with reverse primer and misread bases in each sequence were removed (Fig. 23). After combining the overlapping amplicon sequences of the forward and reverse sequences, contigs were generated using DNA Baser v4.23.0 (Fig. 24). DNA Baser also helps in trimming of low quality reads from the sequences and keeps the high quality reads. Out of the 1118 sequences, 273 sequences were not taken into consideration. These 273 sequences correspond to three genotypes *Ae. speltoides acc. pauTA1784*, Raj3765 and K9644. Reason for not obtaining the sequences in these genotypes was the failure of amplification which may be due to poor quality of DNA and thus low quality sequences or no sequence data (NNN nucleotides) obtained. Thus sequence data of 17 wheat genotypes is presented here. After the formation of contigs, these contigs were taken as queries and blasted against sequences of *SSI* homoeologous genes (7A, 7B and 7D) using NCBIblast2seq tool for the authenticity that they belong to 7B chromosome only.

After contigs assembly, the exonic and intronic boundaries of *SSI* gene were predicted using Artemis tool by facilitating the intuitive manual addition of annotation to sequence data (Berriman and Rutherford 2003). Homology base gene identification using homologous protein and sequences analysis using this tool identified 9 exons and 8 introns in these sequences (Fig. 25). The reference *SSI* gene accession Traes_7BS_6135B1D85.1 identified in Ensembl Plants tool also contained 9 exons and 8 introns. Similar number of exons and introns in the genotypes under study showed authenticity of sequence data. This tool provides integrated visualization and computational analysis of different types of high throughout sequencing (HTS) datasets in the context of a reference genome and its corresponding annotation.

4.3.3 Detection of nucleotide variations

The generated gene consensus sequences were aligned with *SSI* reference gene sequence using ClustalX2.0.11 to detect of Single Nucleotide Polymorphisms (SNPs) and Insertion-Deletions (InDels) (Fig. 26). Sequence of *SSI* gene obtained from *Triticum monococcum* and *Triticum boeoticum* were aligned with *SSI*-7A as reference (Traes_7AS_70ED86B15.1), sequences of two accessions of *Ae. tauschii* aligned with *SSI*-7D as reference (Traes_7DS_5159E3934.1) while sequences from all other genotypes *Ae.*

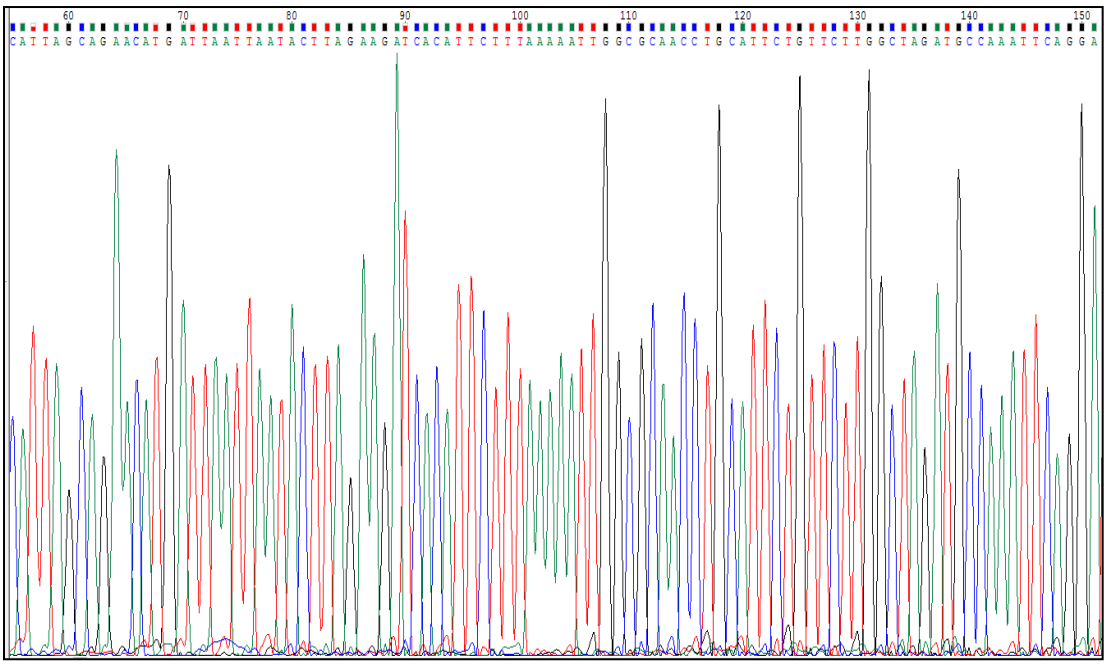

Fig. 22 View of sequencing Chromatogram (retrieved from capillary sequencer) in Chromas Lite software

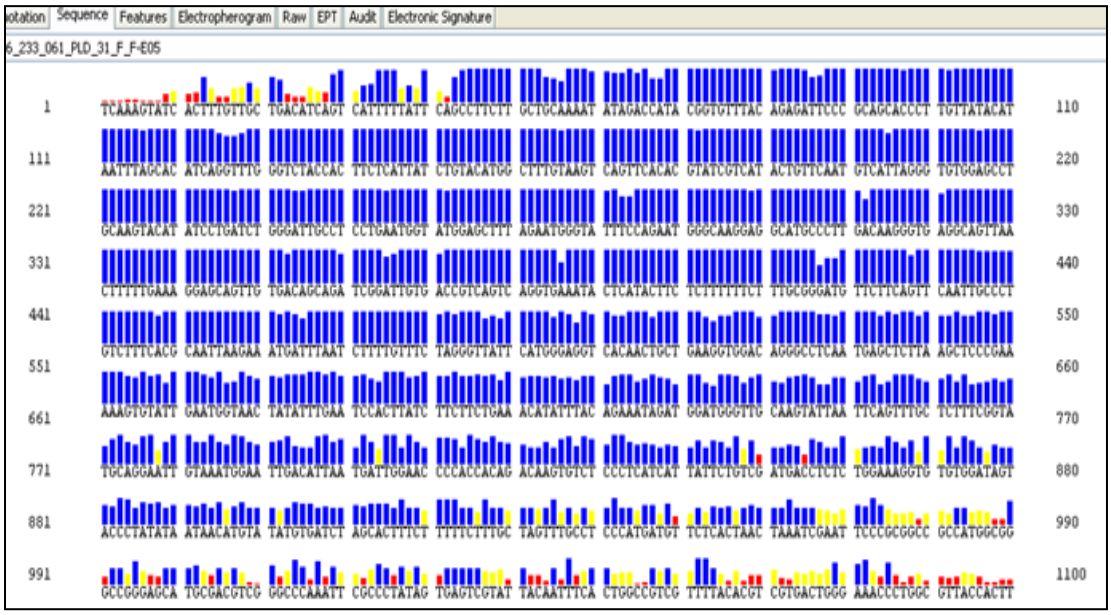

Fig. 23 Sequence obtained using ABI 3730XL sequencer. Blue, yellow and red bars indicate Q-value of particular base

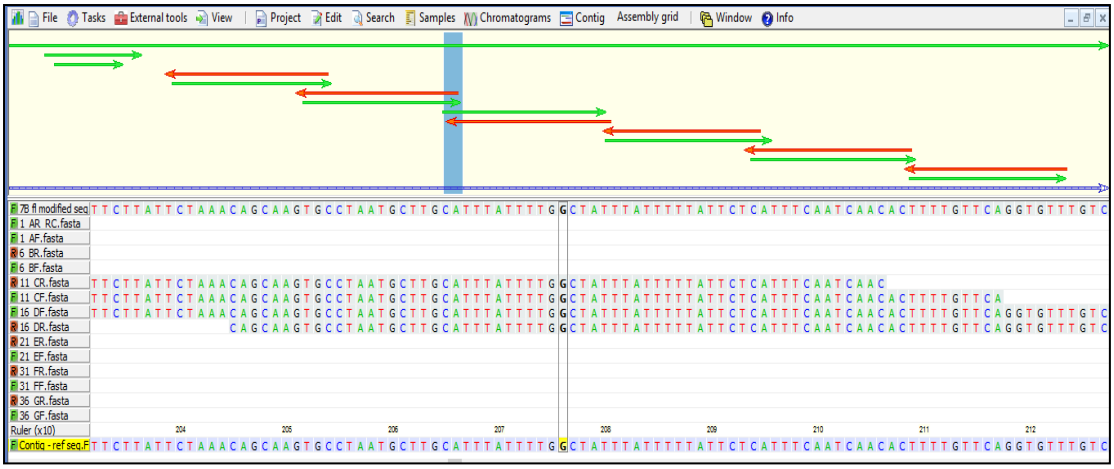

Fig. 24 Forward and Reverse sequences were assembled using DNA BASER to form the contig of complete *SSI*-7B gene

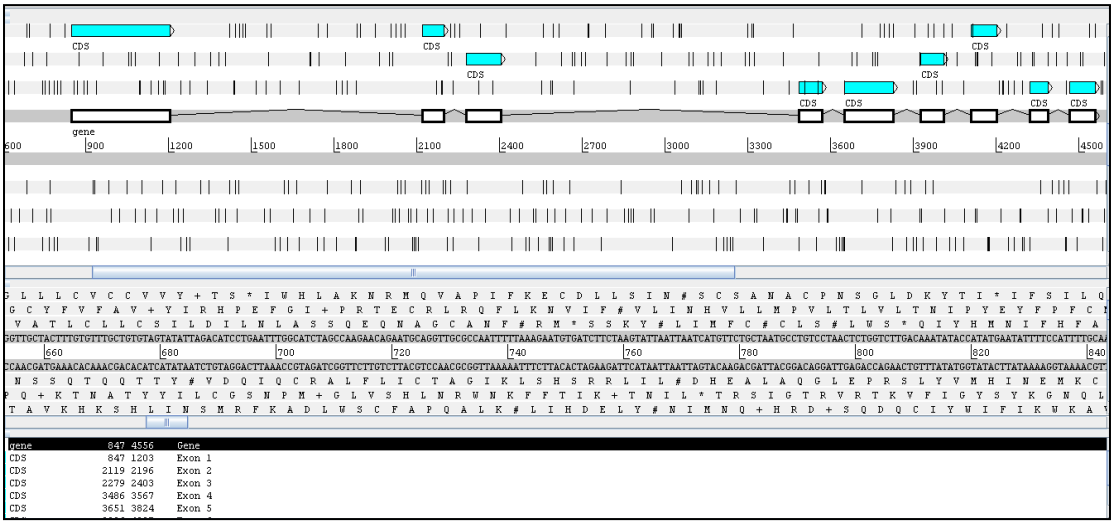

Fig. 25 Prediction of Intronic and Exonic boundaries of *SSI*-7B gene using Artemis tool

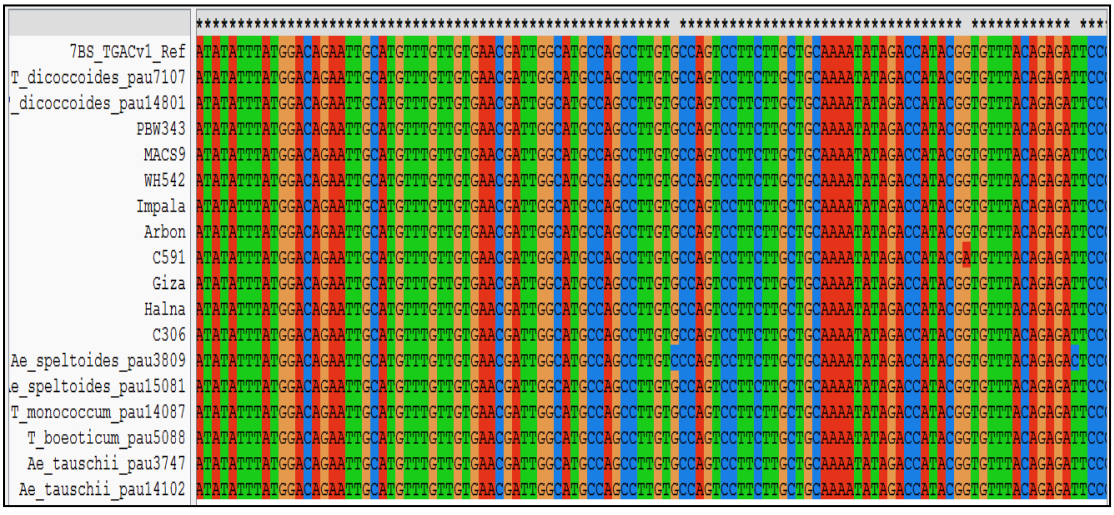

Fig. 26 ClustalX alignment of *SSI*-7B gene nucleotide sequences of different wild and cultivated genotypes with reference representing SNPs

Table 16: SNPs identified in Exon and Intron of *SSI* gene in different wheat genotypes

Genotype	Exon									Total SNPs	Intron								Total SNPs
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	
Triticum monococcum acc. pau14087				1						1	9		28	3	1	1	2		44
Triticum boeoticum acc. pau5088	4			1						5	14		31	3	1	1	2		52
Ae. speltoides acc. pau3809			1	2	2			1	2	8	20		36	3	1		3		63
Ae. speltoides acc. pau15081	7			1	2	1	2			13	34		29		2				65
Ae. tauschii acc. pau3747	2									2	3		5			1			9
Ae. tauschii acc. pau14102	1									1	2		11						13
Triticum dicoccoides acc. pau14801										0	1		2						3
Triticum dicoccoides acc. pau7107	2						1	1		4	2		5	5					12
MACS9								1		1	1								1
Arbon										0	1								1
Impala	1	1	1				1			4	2		2						4
Giza										0	2		1		1		1		5
C306										0	1						1		2
C591	1			1						2	1		1						2
WH 542						1			1	2	2		1						3
PBW 343										0	4			1	1				6
Halna		1	1							2			1						1
Total	**18**	**2**	**3**	**6**	**4**	**2**	**4**	**3**	**3**		**99**	**0**	**153**	**15**	**7**	**3**	**9**	**0**	

Table 17: Total no of SNPs in Exon and Intron of *SSI*-7B gene in different wheat genotypes

Genotype	Exonic SNP	Intronic SNP
Triticum monococcum acc. pau14087	1	44
Triticum boeoticum acc. pau5088	5	52
Ae. speltoides acc. pau3809	8	63
Ae. speltoides acc. pau15081	13	65
Ae. tauschii acc. pau3747	2	9
Ae. tauschii acc. pau14102	1	13
Triticum dicoccoides acc. pau14801	0	3
Triticum dicoccoides acc. pau7107	4	12
MACS9	1	1
Arbon	0	1
Impala	4	4
Giza	0	5
C306	0	2
C591	2	2
WH 542	2	3
PBW 343	0	6
Halna	2	1
Total	**45**	**286**

speltoides, T.dicoccoides, durum wheat and hexaploid wheats were aligned using *SSI*-7B as reference (Traes_7BS_6135B1D85.1). Variations/SNPs detected in the wheat genotypes under study were then manually curated by checking the quality of the corresponding reads in the chromatogram file. Total number of variations has been detected for each exon and intron of the gene with respect to the reference sequence (Table 16). In total, 45 SNPs in exonic region and 286 SNPs in intronic region were identified in gene in selected wheat genotypes (Table 17).

4.3.3.1 Exonic SNPs

In exon, maximum SNPs were detected in wild genotypes of *Ae. speltoides acc. pau15801* (13) *and Ae. speltoides acc. pau3809* (8). No SNP was detected in wild genotype *T. dicoccoides acc. pau14081* and cultivated genotypes Giza, Arbon, C306, PBW 343. First exon showed maximum SNPs (18) while second and sixth exons had minimum SNPs (2). Maximum SNPs were in wild genotypes as *Triticum boeoticum acc. pau5088, Ae. speltoides acc. pau15081, Ae. tauschii acc. pau3747, Ae. tauschii acc. pau14102* and *Triticum dicoccoides acc. pau7107* had four, seven, two, one and two SNPs respectively in first exon

while *Triticum monococcum acc. pau14087, Ae. speltoides acc. pau3809* and *Triticum dicoccoides acc. pau14801* do not have any SNP in first exon. In cultivated genotypes first exon has SNPs only in Impala and C591. In second exon SNPs were found only in hexaploid cultivated wheat Impala and Halna and no SNP was found in any of wild genotypes. Third exon has three SNPs one each in *Ae. speltoides* acc. *pau3809,* Impala and Halna. Fourth exon was second most variable exon with 6 SNPs in five genotypes of *Ae. speltoides acc. pau3809, Ae. speltoides acc. pau15801, Triticum monococcum acc. pau14087, Triticum boeoticum acc. pau5088* and C591. Fifth exon is *Ae. speltoides* specific as it contain 4 SNPs, two each in *Ae. speltoides acc. pau3809* and *acc. pau15081*. Sixth exon again has minimum number of 2 SNPs, one each in *Ae. speltoides* acc. *pau15081* and WH542. Four SNPs of seventh exon were found in *Ae. speltoides acc. pau15081* (2), *T. dicoccoides acc. pau7107* (1) and WH542 (1). Eigth exon had SNPs in wild and durum wheat as *Ae. speltoides acc. pau 3809* (1), *T. dicoccoides acc. pau7107* (1) and MACS9 (1). Ninth exonic SNPs were found in *Ae. speltoides acc. pau3809* and WH542.

In A genome genotypes of *Triticum boeoticum acc. pau5088* five SNPs in first and fourth exons have been detected at positions 459, 531, 567, 603 and 933 while only SNP at position 933 (T→C) was found to be common in both *Triticum boeoticum acc. pau5088* and *Triticum monococcum acc. pau14087* genotypes. SNPs T→C and A→G were found to be most common in A genome. *Ae. speltoides acc. pau3809* and *pau15081* contained eight exonic SNPs (870, 918, 930, 999, 1110, 1336, 1380 and 1480 positions) and 13 exonic SNPs (336, 371, 414, 438, 468, 469, 604, 930, 999, 1113, 1191, 1233 and 1257 positions) respectively of which two SNP at positions 930 (C→T) and 999 (T→A) were common in both B genome genotypes. T→C, C→T and T→A were most common type of SNPs detected in *Ae. speltoides*. In *Ae. tauschii acc. pau3747* and *Ae. tauschii acc. pau14102*, D genome genotypes, common SNP at 593 position (G→C) have been identified but *acc. pau14102* contained one additional SNP at 348 exonic position also. In wild tetraploid *Triticum dicoccoides acc. pau7107*, exonic SNPs were detected at positions 423, 602 and 1330 with respect to reference.

In cultivated genotypes, SNPs positions identified were MACS9 (1396), Impala (456, 740 and 1222), C591 (388 and 905), WH542 (1164 and 1373) and Halna (732 and 819) and no common SNP position was found among these genotypes. But T→C and A→G were common SNPs among cultivated hexaploid genotypes. T→C was the most commonly occurring exonic SNP detected over wild and cultivated genotypes.

4.3.3.2 Intronic SNPs

In introns, maximum SNPs have been detected in the third intron (153) followed by first intron (99) and minimum SNPs in the sixth intron (3). No SNP was detected in second and eighth introns in any of genotypes indicated these are conserved in nature. Introns

exhibited more SNP variation than exons in wild genotypes as *Triticum monococcum acc. pau14087, Triticum boeoticum acc. pau5088, Ae. speltoides acc. pau3809 and Ae. speltoides acc. pau15081* contained 44, 52, 63 and 65 SNPs respectively. The cultivated genotypes MACS9, Arbon, Impala, Giza, C306, C591, WH 542, PBW 343 and Halna showed less number of SNPs (1-6) as compared to wild genotypes.

In wild wheat, A genome species *Triticum monococcum acc. pau14087* and *Triticum boeoticum acc. pau5088* shared common intronic SNPs at 37 positions 1739, 1893, 1960, 2005, 2081, 2174, 3098, 3105, 3106, 3116, 3185, 3215, 3225, 3246, 3281, 3325, 3345, 3384, 3475, 3477, 3481, 3497, 3531, 3533, 3597, 3625, 3639, 3642, 3646, 3663, 3677, 3690, 3715, 3902, 3909, 3919 and 4165. In the B genome genotypes, *Ae.speltoides acc. pau3809 and Ae. speltoides acc. pau15081* common SNPs were detected at 20 positions 1437, 1491, 1599, 1605, 1606, 1715, 1812, 1902, 2881, 2918, 2946, 3014, 3099, 3114, 3155, 3172, 3326, 3330, 3335 and 3446 in intronic region while D genome species, *Ae. tauschii acc. pau3747* and *Ae. tauschii acc. pau14102* showed common SNPs at two positions at 1819 and 2156 with respect to reference. The wild tetraploid wheat *Triticum dicoccoides acc. pau14801* and *Triticum dicoccoides acc. pau7107* detected common intronic SNPs at one position 3446. In cultivated genotypes, C306 and C591 identified one common intronic SNP at 1637 position only while rest of the genotypes do not contained any common SNPs in introns (Table 18).

Intronic variation was more in *Triticum monococcum acc. pau14087, Triticum boeoticum acc. pau5088, Ae. speltoides acc. pau3809 and Ae. speltoides acc. pau15081* but exonic variation was maximum in *Ae. speltoides acc. pau3809 and Ae. speltoides acc. pau15081*. Maximum exonic and intronic variations were found in *Aegilops speltoides* species and therefore could be essence of variation. Waines (1994), Pradhan *et al* (2012) and Awlachew *et al* (2016) reported that *Ae. speltoides* could serve as a source of genetic variability for improved thermotolerance in wheat.

In present study introns found to be more polymorphic than exons. Further wild genotypes exhibited more allelic variation than cultivated genotypes as expected. Ramakrishna *et al* (2002) and Haga *et al* (2002) reported that introns contained more genetic variations than in exons in grass genomes. Introns were more diversified in terms of size and nucleotide sequences (Breathnach and Chambon 1981). In closely related species, introns exhibited a large number of insertion/deletion differences than exons (Clark *et al* 1996). Huang and Babel (2012) also reported more allelic variation in intronic regions of *SSIIa* gene in wheat but detected no association between these variation and yield related triats. In wheat, Rose *et al* (2015) sequenced three *AGP-L* genes which encoded the large subunit of AGPase and showed that *AGP-L-B* allele identified 13 indels and 58 SNPs in the introns, while only nine SNPs were found in the exon.

Table 18: List of identified intronic SNPs along with the position against reference in selected wild and cultivated genotypes

S. No	Genotype	Position	Nucleotide conversion	S. No	Genotype	Position	Nucleotide conversion
1	*Triticum monococcum acc. pau14087*	1739	G→A	26	*Triticum monococcum acc. pau14087*	3531	T→G
2		1877	A→G	27		3533	A→G
3		1893	C→T	28		3597	A→G
4		1960	T→C	29		3625	G→T
5		2005	T→C	30		3639	C→T
6		2015	A→G	31		3642	C→G
7		2081	C→A	32		3646	C→T
8		2149	A→C	33		3663	T→C
9		2174	G→A	34		3677	G→T
10		3098	C→G	35		3690	T→C
11		3105	T→C	36		3715	A→G
12		3106	G→A	37		3902	C→T
13		3116	G→A	38		3909	T→A
14		3185	T→C	39		3919	C→T
15		3215	G→A	40		4165	T→G
16		3225	G→T	41	*Triticum boeoticum acc. pau5088*	1520	G→A
17		3246	G→C	42		1579	G→A
18		3281	A→T	43		1584	G→C
19		3325	G→A	44		1602	C→T
20		3345	T→A	45		1656	A→C
21		3384	G→A	46		1670	T→G
22		3475	T→C	47		1739	G→A
23		3477	T→C	48		1783	C→T
24		3481	T→C	49		1893	C→T
25		3497	T→G	50		1960	T→C

S. No	Genotype	Position	Nucleotide conversion	S. No	Genotype	Position	Nucleotide conversion
51	*Triticum boeoticum acc. pau5088*	2005	T→C	76	*Triticum boeoticum acc. pau5088*	3597	A→G
52		2081	C→A	77		3625	G→T
53		2174	G→A	78		3632	C→T
54		2250	T→A	79		3639	C→T
55		3098	C→G	80		3642	C→G
56		3105	T→C	81		3646	C→T
57		3106	G→A	82		3663	T→C
58		3116	G→A	83		3677	G→T
59		3142	A→G	84		3690	T→C
60		3185	T→C	85		3715	A→G
61		3215	G→A	86		3902	C→T
62		3225	G→T	87		3909	T→A
63		3230	G→A	88		3919	C→T
64		3246	G→C	89		4165	T→G
65		3281	A→T	90	*Ae. speltoides acc. pau3809*	1437	T→A
66		3325	G→A	91		1491	T→C
67		3345	T→A	92		1500	A→G
68		3384	G→C	93		1599	T→A
69		3475	T→C	94		1605	G→A
70		3477	T→C	95		1606	A→C
71		3481	T→C	96		1633	C→T
72		3497	T→G	97		1683	G→A
73		3531	T→G	98		1715	T→C
74		3533	A→G	99		1725	G→A
75		3579	C→G	100		1730	C→T

S. No	Genotype	Position	Nucleotide conversion	S. No	Genotype	Position	Nucleotide conversion
101	*Ae. speltoides acc. pau3809*	1793	A→T	126	*Ae. speltoides acc. pau3809*	3116	C→T
102		1812	G→A	127		3155	T→G
103		1837	C→G	128		3172	A→G
104		1902	C→A	129		3194	A→G
105		1942	G→A	130		3201	C→T
106		2066	A→G	131		3214	G→C
107		2074	T→C	132		3239	G→T
108		2075	G→A	133		3244	A→G
109		2086	T→C	134		3264	T→A
110		2542	G→A	135		3270	A→G
111		2577	T→C	136		3291	G→C
112		2636	G→T	137		3326	G→C
113		2846	G→A	138		3330	A→C
114		2856	C→T	139		3335	T→C
115		2863	G→A	140		3356	G→C
116		2881	G→A	141		3373	G→A
117		2893	G→A	142		3409	T→C
118		2918	A→C	143		3446	A→G
119		2946	A→G	144		3579	C→T
120		3014	T→C	145		3585	C→T
121		3051	A→G	146		3614	A→G
122		3070	G→T	147		3892	G→C
123		3099	A→C	148		4245	C→T
124		3108	C→T	149		4270	C→T
125		3114	C→A	150		4271	C→T

S. No	Genotype	Position	Nucleotide conversion	S. No	Genotype	Position	Nucleotide conversion
151	*Ae. speltoides acc. pau15081*	1211	C→T	176	*Ae. speltoides acc. pau15081*	1875	T→G
152		1215	A→G	177		1902	C→A
153		1223	C→A	178		1915	G→A
154		1240	T→G	179		1932	C→T
155		1248	C→T	180		1947	T→G
156		1270	T→C	181		1958	G→C
157		1299	C→T	182		1963	C→G
158		1331	T→C	183		1977	G→C
159		1341	A→C	184		1997	T→C
160		1359	C→T	185		2881	G→A
161		1437	T→A	186		2901	G→A
162		1491	T→C	187		2918	A→C
163		1495	G→T	188		2922	T→C
164		1574	G→A	189		2925	C→T
165		1599	T→A	190		2927	C→T
166		1604	A→G	191		2946	A→G
167		1605	G→A	192		2963	T→C
168		1606	A→G	193		2983	C→T
169		1706	G→A	194		3014	T→C
170		1715	T→C	195		3015	G→A
171		1724	G→A	196		3099	A→C
172		1756	A→G	197		3114	C→G
173		1812	G→A	198		3128	A→G
174		1855	C→T	199		3144	T→G
175		1866	C→G	200		3155	T→G

S. No	Genotype	Position	Nucleotide conversion	S. No	Genotype	Position	Nucleotide conversion
201	*Ae. speltoides acc. pau15081*	3172	A→G	226	*Ae. tauschii acc. pau3747*	2881	A→C
202		3202	G→A	227		3550	G→A
203		3207	G→A	228		4218	C→A
204		3270	A→G	229	*Ae. tauschii acc. pau14102*	1819	G→A
205		3291	G→C	230		2156	T→G
206		3321	G→A	231		2992	T→C
207		3326	G→C	232		3072	A→T
208		3330	A→C	233		3122	G→A
209		3335	T→C	234		3348	T→C
210		3340	A→G	235		3373	G→A
211		3346	A→G	236		3394	C→G
212		3395	G→A	237		3429	C→G
213		3446	A→G	238		3433	C→A
214		3850	T→A	241		3477	T→C
215		3871	T→A	242		3498	T→C
216		4077	T→A	243		3501	A→G
219		4098	C→G	244	*Triticum dicoccoides acc. pau14801*	1327	A→G
220	*Ae. tauschii acc. pau3747*	1819	G→A	245		2607	G→C
221		1914	C→T	246		3446	A→G
222		2156	T→G	247	*Triticum dicoccoides acc. pau7107*	1491	T→C
223		2705	T→C	248		1816	A→C
224		2846	A→C	249		2726	A→C
225		2878	G→A	250		3102	A→G

S. No	Genotype	Position	Nucleotide conversion	S. No	Genotype	Position	Nucleotide conversion
251	*Triticum dicoccoides acc. pau7107*	3309	T→C	270	Giza	2753	G→A
252		3321	G→A	271		3633	A→G
253		3446	A→G	272		4267	T→C
254		3579	C→T	273	C306	1637	G→T
255		3585	C→T	274		4276	A→T
256		3610	A→T	275	C591	1637	G→T
257		3614	A→G	276		2757	A→G
258		3621	A→T	277	WH 542	1576	A→G
259	MACS9	1771	A→G	278		2050	A→C
260	Arbon	1723	A→G	279		2739	T→C
261	Impala	1302	C→T	280	PBW 343	1207	G→A
263		1740	A→G	281		1527	T→C
264		3070	G→A	282		1660	C→T
265		3144	T→A	283		1703	T→C
266	Giza	1252	T→C	284		3614	A→G
267		1825	G→A	285		3890	A→G
268		2695	C→T	286	Halna	3348	C→T
269		2717	A→G				

4.3.3.3 Transition and Transversion

SNPs identified in present study contribute to both transition (A↔ G/ C↔ T) and transversion (A↔ T/C ↔ G). There were 221 base transitions and 110 base transversions. In total, base transitions were found to be more common then transversions (Table 19). In wild genotypes maximum transition and transversion were found in *Ae. speltoides* with 47 transition and 22 transversion in *acc. pau3809* and 48 transition and 32 transversion in *acc. pau15081* followed by *Triticum boeoticum acc. pau5088* with 36 transition and 20 transversion and *Triticum monococcum acc. pau14087* with 31 transition and 14 transversion. *Triticum dicoccoides acc. pau14801* and *Ae.tauschii acc. pau3747 and Ae. tauschii acc. pau14102* showed comparatively less number of transition (3-8) and transversion (1-6).

In cultivated genotypes, also transition was more than transversion. Impala and Giza have maximum transition followed by PBW 343. Other cultivated genotypes have only 1-3 transition. Transversion in cultivated genotypes were found in Impala, C306, C591 and WH 542 only. Transitions are usually favoured over transversions (Sankoff *et al* 1976) because transition mutations do not alter 3D structures of protein as they resulted into synonymous substitution because selection pressure acts to conserve chemical properties of amino acids (Vogul and Kopun 1977). Luo *et al* (2016) provided evidence that universal bias occurs in favour of transitions over transversions as the process of transversion including size conformation is more complicated than transition.

Table 19: Total no of Transitions and Transversions in different wheat genotypes

Genotype	Transition	Transversion
Triticum monococcum acc. pau14087	31	14
Triticum boeoticum acc. pau5088	36	20
Ae. speltoides acc. pau3809	47	22
Ae. speltoides acc. pau15081	48	32
Ae. tauschii acc. pau3747	6	5
Ae. tauschii acc. pau14102	8	6
Triticum dicoccoides acc. pau14801	2	1
Triticum dicoccoides acc. pau7107	11	4
Giza	7	0
MACS9	2	0
Arbon	1	0
Impala	7	1
C306	0	2
C591	3	1
WH 542	3	2
PBW 343	6	0
Halna	3	0
Total	**221**	**110**

Thus four wild wheat genotypes, *Triticum monococcum acc. pau14087, Triticum boeoticum acc. pau5088, Ae. speltoides acc. pau 3809* and *Ae. speltoides acc. pau15081* showed maximum variation in form of transition and transversion, indicating that could be source for some useful allelic variants in *SSI* gene.

4.3.3.4 InDels

No insertions were detected in any of 17 selected wheat genotypes. However, deletions were identified only in the wild genotypes namely *Triticum monococcum acc. pau 14087, Triticum boeoticum acc. pau 5088, Ae. speltoides acc. pau 3809, Ae. speltoides acc. pau15081*, *Ae. tauschii acc. pau 14102* and *Triticum dicoccoides acc. pau14801*. These deletions were found only in intronic regions while no deletion was detected in exonic regions in present study (Table 20). Presence of more number of Indels in introns have been reported earlier by Clark *et al* (1996) in *Mlc1* gene in *Drosophila* and Rose *et al* (2015) in *AGPase* gene in wheat where introns contained more of both insertion and deletion variations than exons.

Table 20: Deletions detected in wild genotypes along with their respective position with reference

Genotype	Position	Deletion
Triticum monococcum acc. pau14087	1885-1891	TATTCTG
	2274	T
	3388	C
	4889-4890	TT
Triticum boeoticum acc. pau5088	1885-1891	TATTCTG
	2274	T
	3388	C
	4889-4890	TT
Ae. speltoides acc. pau3809	1241	C
	1607-1609	GTA
	3262	G
Ae. speltoides acc. pau15081	1241	C
	1607-1609	GTA
	3016-3017	CC
	3262	G
	3409	T
Ae. tauschii acc. pau14102	4807-4817	AAACAAAGTTG
Triticum dicoccoides acc. pau14801	2850	C

4.3.4 Detection of variations in protein sequences

Protein sequences of the predicted gene in selected wild and cultivated wheats were identified using the Artemis tool. The predicted sequences were aligned with the the reference

sequence using Jalview 2.0 tool to detect variations in amino acids (Fig. 27). Out of total 45 exonic SNPs, 12 non-synonymous substitution and 33 synonymous substitution were identified in 12 wild and cultivated genotypes (Table 21). The wild genotypes *Triticum monococcum acc. pau14087, Triticum boeoticum acc. pau5088* and *Ae. speltoides acc. pau3809* contained one, five and eight synonymous substitutions respectively while *Ae. speltoides acc. pau15081* detected two non-synonymous and eleven synonymous substitutions. Each of *Ae.tauschii acc. pau3747* and *Ae.tauschii acc. pau14102* genotypes contained one non-synonymous substitution each while former have one synonymous substitution also. The wild tetraploid *Triticum dicoccoides acc. pau7107* identified one synonymous and three non-synonymous substitutions only. Out of seven wild genotypes, four had non-synonymous substitution. In case of cultivated genotypes, MACS9 contained one while Impala and C591 contained two non-synonymous substitutions each. Other cultivated genotypes Impala, WH542 and Halna identified two synonymous substitutions each.

Of all the synonymous substitutions, amino acid G-Glycine substitution occurred in maximum number of genotypes (6) followed by I-Isoleucine substitution (5), D-Aspartic acid substitution (4), T-Threonine substitution (4), P-Proline substitution (4) while minimum substitution have been detected for H-Histidine (2), A-Alanine (2), L-Leucine (2), V-Valine (1), Q-Glutamine (1), R-Arginine (1) and S-Serine (1). The maximum non-synonymous substitutions were identified for D-Aspartic acid (3) followed by A-Alanine (2) and E-Glutamic acid (2) while L-Leucine, Q-Glutamine, T-Threonine, N-Asparagine and P-Proline occurred only once. *Triticum dicoccoides acc. pau7107* genotype contributed maximum non-synonymous substitutions (3) followed by *Ae. speltoides acc. pau15081* (2), Impala (2) and C591 (2) while one non-synonymous substitutions detected in each of *Ae. tauschii acc. pau3747, Ae. tauschii acc. pau14102* and MACS9. A synonymous substitution constitutes evolutionary based substitution of one base for another present in exon of gene so that the amino acid sequence produced is not modified. Modification of protein is denied because of degeneracy of amino acid. Non-synonymous substitution on the other hand, lead to alteration of amino acid which is further distincted into conservative (change to an amino acid with same physiochemical properties), semi-conservative (e.g. negative changed to positively charged amino acid), or radical (vastly different amino acid).

Non-synonymous substitutions were further divided into conservative, semi-conservative and radical substitution based on physiochemical properties like aliphatic, aromatic, sulphur containing, hydroxyl group, positively charged (basic), negatively charged (acidic) and amid containing group. In the present study, three conservative (two G→A and one L→P conversions); two semi-conservative (two K→E conversions) and seven radical (two G→D, one M→L, K→Q, I→T, D→N and N→D conversions) substitutions were identified (Table 21).

Fig. 27 Multiple sequence alignment and analysis of amino acid sequences of *SSI*-7B gene using Jalview tool

Table 21: Synonymous and Non-synonymous Substitution in Exonic regions of *SSI* gene in 12 wheat genotypes

	SNP Position in reference			Amino acid Conversion	
Genotype	**CDS**	**Protein**	**Nucleotide Conversion**	**Synonymous Substitution**	**Non-Synonymous Substitution**
Triticum monococcum acc. pau14087	933	311	T→C	H→H	
Triticum boeoticum acc. pau5088	459	153	G→A	G→G	
	531	177	A→G	P→P	
	567	189	A→G	A→A	
	603	201	A→T	P→P	
	933	311	T→C	H→H	
Ae. speltoides acc. pau3809	870	290	G→C	V→V	
	918	306	T→C	D→D	
	930	310	C→T	T→T	
	999	333	T→A	P→P	
	1110	370	T→C	D→D	
	1336	446	T→C	L→L	
	1380	460	T→C	I→I	
	1455	485	A→G	Q→Q	
Ae. speltoides acc. pau15081	336	112	T→A	I→I	
	414	138	T→A	T→T	
	379	127	A→T		M→L*
	438	146	G→T	G→G	
	468	156	G→C	G→G	
	469	157	T→C	L→L	
	604	202	A→C		K→Q*
	930	310	C→T	T→T	

Genotype	SNP Position in reference		Nucleotide Conversion	Amino acid Conversion	
	CDS	Protein		Synonymous Substitution	Non-Synonymous Substitution
	999	333	T→A	P→P	
	1113	371	G→A	R→R	
	1191	397	C→T	S→S	
	1233	411	C→T	D→D	
	1257	419	A→G	T→T	
Ae. tauschii acc. pau3747	348	116	T→C	I→I	
	593	198	G→C		G→A#
Ae. tauschii acc. pau14102	593	198	G→C		G→A#
Triticum dicoccoides acc. pau7107	423	141	C→T	I→I	
	602	201	T→C		I→T*
	1231	411	G→A		D→N*
	1330	444	A→G		K→E+
MACS9	1367	456	T→C		L→P#
Impala	456	152	A→G	A→A	
	740	247	G→A		G→D*
	825	275	A→G	G→G	
	1222	498	A→G		N→D*
C591	388	130	A→G		K→E+
	905	302	G→A		G→D*
WH 542	1164	388	T→C	G→G	
	1373	459	A→C	D→D	
Halna	732	244	T→C	G→G	
	819	273	T→C	I→I	

*** radical non-synonymous substitution, # conservative non-synonymous substitution and + semi-conservative non-synonymous substitution**

4.3.5 Structure prediction of *SSI* protein

Homology modeling approach was employed to determine the structure of proteins based on the known structure of template proteins. The models were constructed using Modeller 9.17 tool. No known *SSI* protein structure in wheat was available in PDB. So *SSI* protein in barley (PDB id 4hln) showing 98% sequence identity with wheat reference sequence was used as template. By using this template, three-dimensional structures were predicted for *SSI* protein in different wild and cultivated genotypes (Fig. 28). RMSD (root mean square deviation) values were calculated using chimera tool and were found to be less than 2 Å indicating the accuracy of generated structures. Then the predicted structures were superimposed and analyzed for structural differences (Fig. 29) mainly in active sites of the protein.

The active sites are clefts in the protein structure where ligand binds resulting in catalysis. The residues at the active site formed hydrophobic interactions or hydrogen bonds required for catalysis of the protein conformation and functioning. The eight active site residues namely F (535), P (537), F (538), E (543), W (548), T (564), H (572) and S (575) were identified in the template *SSI* barley protein using PDBsum (www.ebi.ac.uk/pdbsum/) [Fig. 30]. *SSI* barley protein consists of two catalytic domains GT-1 and GT-5 as predicted by Pfam and it was observed that all eight active site residues fall outside of both GT domains in barley template. The structural differences detected in the wild and cultivated genotypes were shown in Table 22.

Table 22: Active site residues identified in template barley 4hln protein and structural differences detected in four wild and three cultivated genotypes of wheat having non-synonymous substitutions

	Genotype						
Active site residues with position (4hln)	*Ae. tauschii acc. pau 14102*	*Ae. tauschii acc. pau 3747*	*Ae. speltoides acc. pau 15081*	*T. dicoccoides acc. pau7107*	MACS 9	Impala	C591
F (535)	E (340)	E (340)	E (340)	E (340)	E (340)	E (340)	E (340)
P (537)	G (342)	G (342)	G (342)	G (342)	G (342)	G (342)	G (342)
F (538)	L (343)	L (343)	L (343)	L (343)	**R (346)**	L (343)	L (343)
E (543)	D (348)	D (348)	D (348)	D (348)	D (348)	D (348)	D (348)
W (548)	G (353)	G (353)	G (353)	G (353)	**P (350)**	**L (351)**	G (353)
T (564)	**M (369)**	**M (369)**	**F (381)**	**K (368)**	**E (373)**	**E (373)**	**F (381)**
H (572)	R (376)	R (376)	**E (377)**	R (376)	R(376)	R (376)	R (376)
S (575)	Q (380)	Q (380)	Q (380)	Q (380)	Q (380)	Q (380)	Q (380)

Letter in bold refer to variable active site in the predicted protein structure

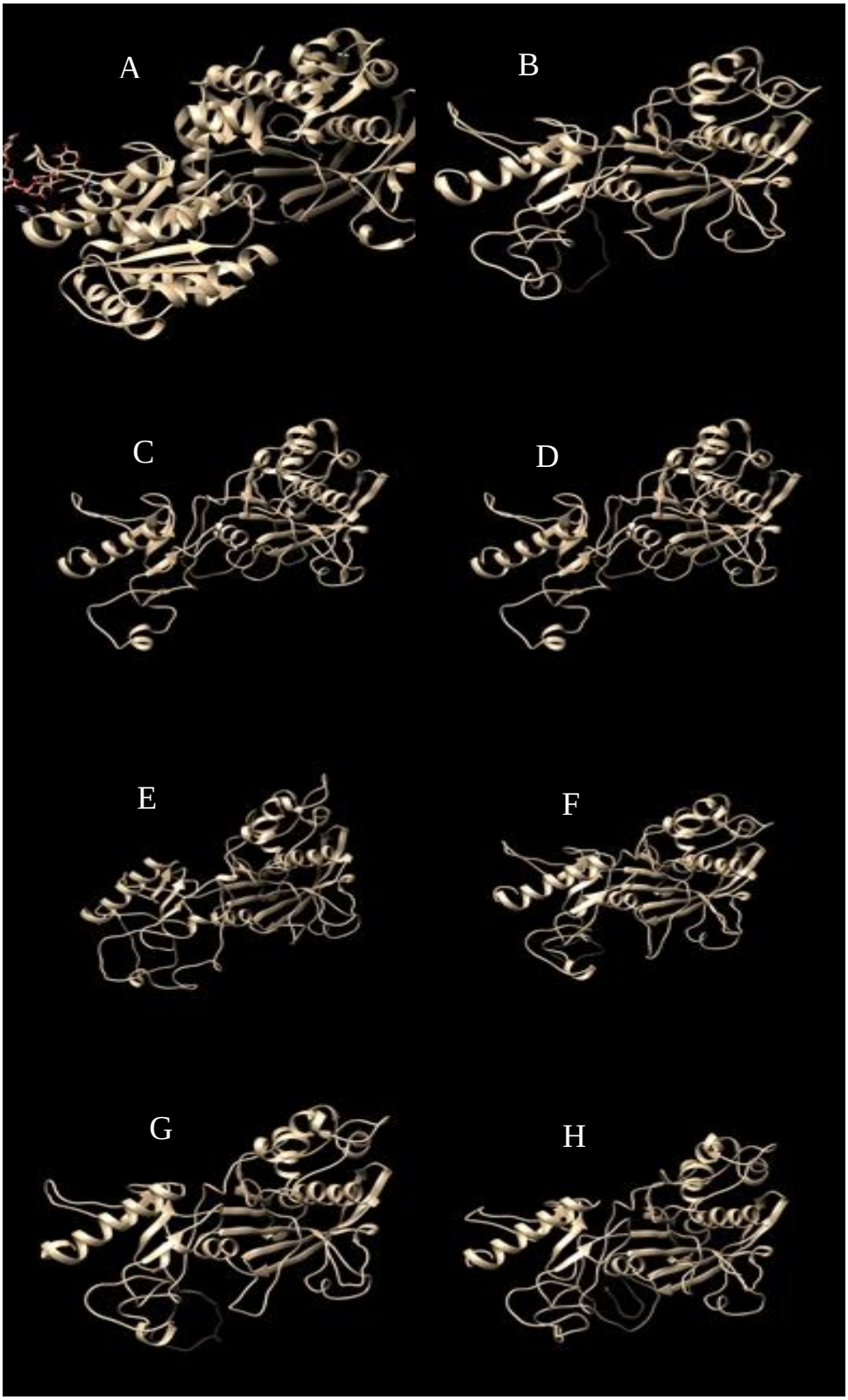

Fig. 28 3D structures predicted for *SSI* protein using MODELLER 9.17

A - Reference Barley template (4hln), B - *Aegilops speltoides acc. pau15081*, C - *Aegilops tauschii acc. pau3747*, D - *Aegilops tauschii acc. pau14102*, E - *Triticum dicoccoides acc. pau7107*, F - C591, G - Impala, H - MACS9

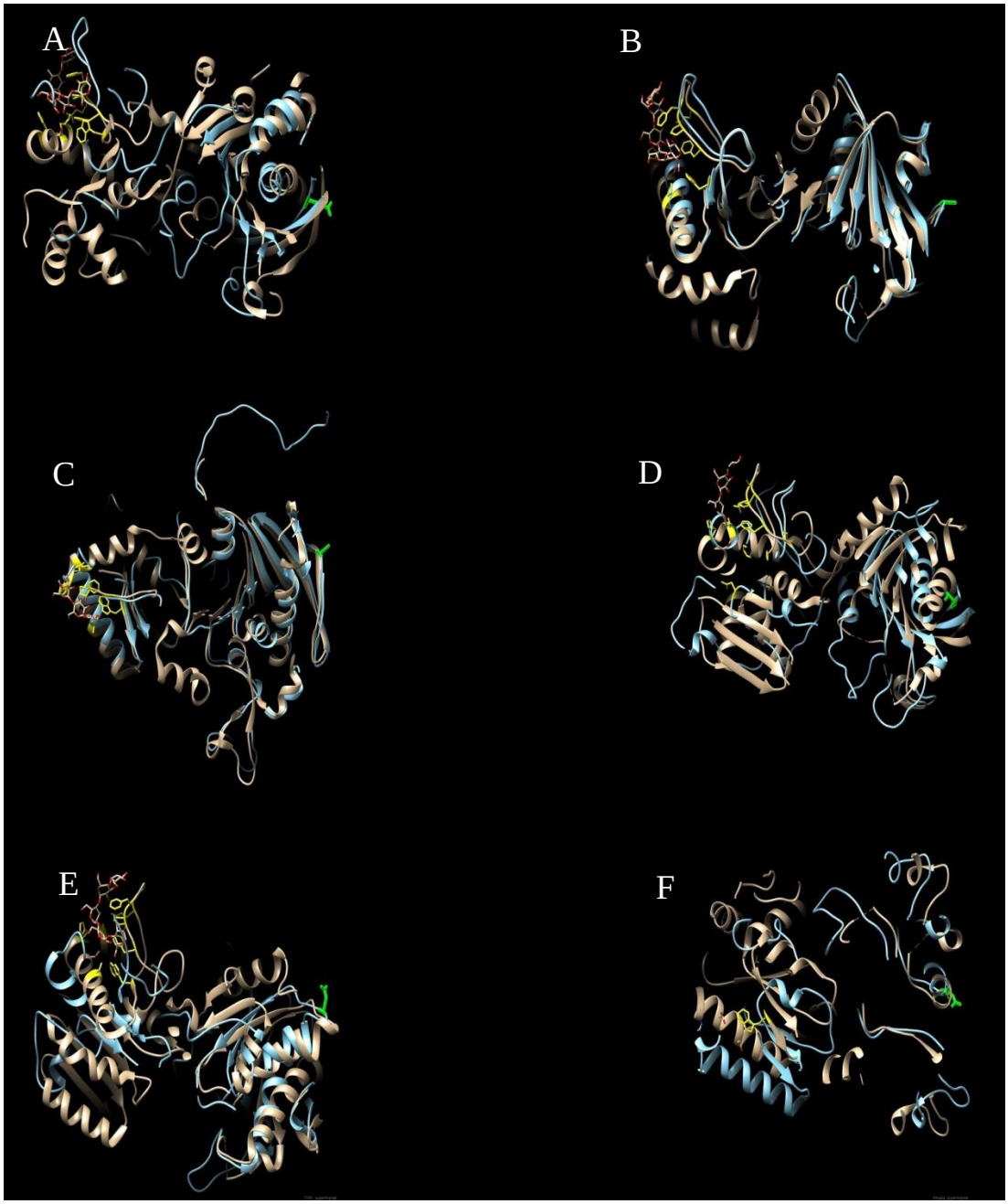

Fig. 29 The predicted structures superimposed and analyzed for structural differences

A - *Aegilops speltoides acc. pau15081*, B - *Aegilops tauschii acc. pau3747*, C - *Aegilops tauschii acc. pau14102*, D - *Triticum dicoccoides acc. pau7107*, E - C591, F - Impala. Yellow colour represents active site residues identified in template barley 4hln protein

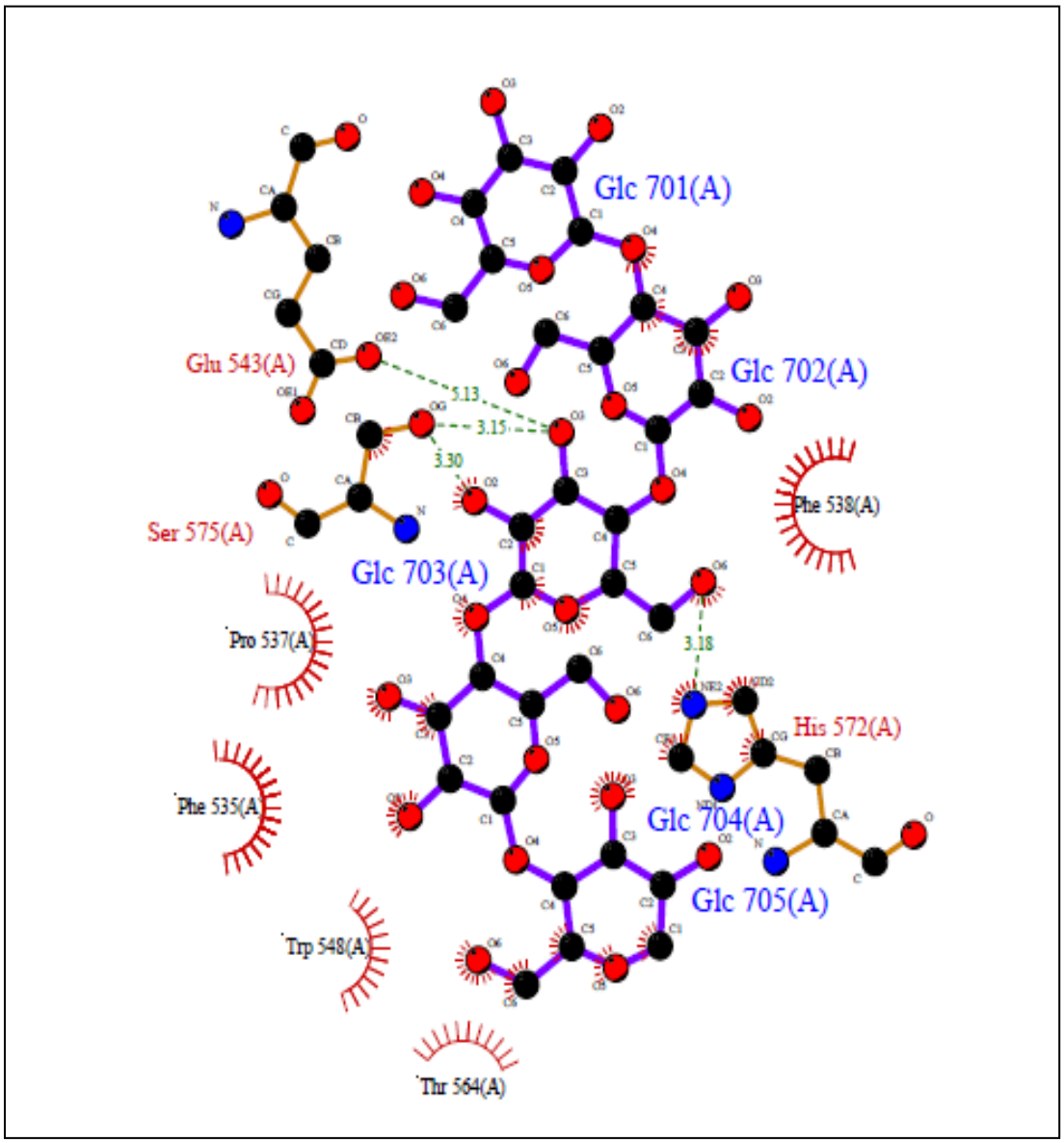

Fig. 30 Active site residues in template barley (4hln) protein along with their position identified using PDBsum

The active site residues that take part in catalysis and substrate binding i.e. F (535), P (537), E (543) and S (575) showed common structural differences in the selected wild and cultivated genotypes when compared with barley protein. The amino acids residues within build models are conserved among themselves except few active site residues as R (346) in MACS9; P (350) and L (351) in MACS9 and Impala respectively and E (377) in *Ae. speltoides acc. pau15081*. T (564) active site residue of barley protein showed maximum structural differences as it was different in all the genotypes - M (369) in *Ae. tauschii acc. pau14102* and *Ae. tauschii acc. pau3747;* F (381) in *Ae. speltoides acc. pau15081* and C591; E (373) in MACS9 and Impala and K (368) in *T. dicoccoides acc. pau7107* as shown in Table 21. Thus active site corresponding to barley active site T (564) was the most variable active site in present study. But none of predicted active sites in the genotypes under study corresponds to non-synonymous substitutions. Literrier *et al* (2008) studied homology modeling for protein structure prediction of *SSIV* gene using TITO (Tool for Incremental Threading Optimization) and suggested that valine residue in the K-**X**-G-G-L conserved motif in *SSIII* and *SSIV* isoforms could be important factor for protein specificity as compared to *SSI* and *SSII* isoforms.

4.3.6 Prediction of domains in the *SSI* protein

Domains and motifs in the *SSI* protein were determined with the help of Pfam (http://pfam.xfam.org/) online program. Results revealed the presence of important starch synthase catalytic domain belonging to glycosyltransferase 5 (GT-5) family in sequences of *SSI* gene from cultivated and wild genotypes under study (Table 23). In most of the known *SS* isoforms, GT-1 domain occupies C-terminal half and GT-5 domain occupies N-terminal half of catalytic active motifs, and are responsible for binding with glucosyl donor i.e. ADPglucose (Keeling and Myers 2010). Leterrier *et al* (2008) studied starch synthase glycosyl transferase (GT-5) domain homology which was compared to prokaryotic *SS* and revealed that GT-5 domain (Pfam PF08323) was similar among all *SS* (*SSI-SSIV*) isoforms.

Table 23: Domain of *SSI* protein in selected wild and cultivated species of wheat

Family	Description	Entry type	Clan	Alignment	
				Start	End
glyco_transf_5	Starch synthase catalytic domain	Domain	CL0113	35	293

As shown in Table 23, the alignment of the starch synthase catalytic domain GT-5 starts from amino acid on 35 position and end at amino acid 293. Predicted GT-5 domain found in present study found to be of similar length in all the selected wild and cultivated wheat genotypes (Fig. 31). Thus no variations for catalytic domain were identified. Of the total nine exons in *SSI* gene, six falls in GT-5 domain region and within these exons, there were 35 SNPs with first exon having maximum 18 SNPs (Table 24). The six allelic variants

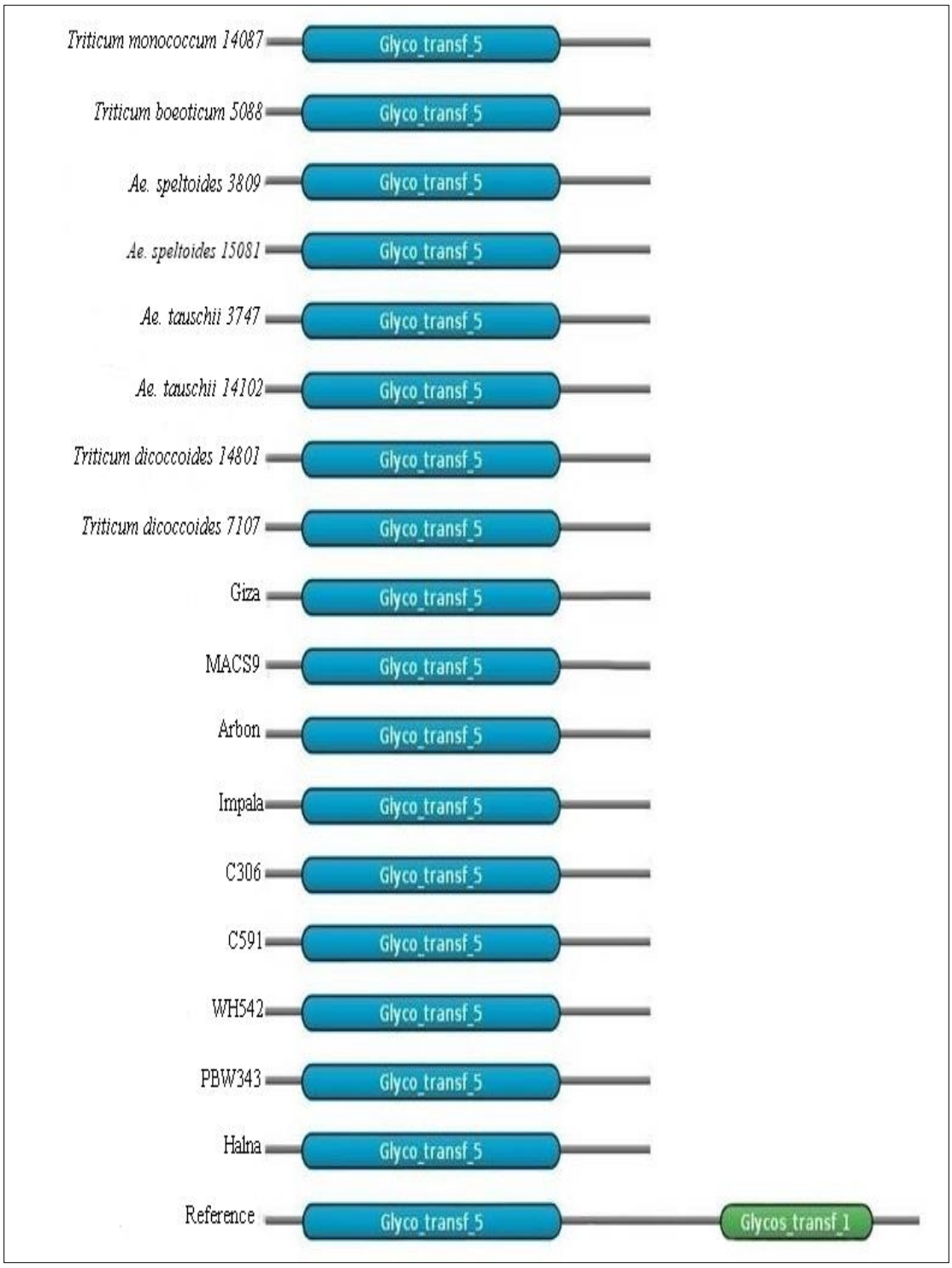

Fig. 31 Domain structure prediction for *SSI* protein sequences in Pfam

with non-synonymous substitution in four wild and two cultivated genotypes (Fig. 32) could be candidate SNPs used for designing markers to identify alleles for improved heat tolerance.

Table 24: Non-synonymous substitutions (with position) lying in the GT-5 domain region, and corresponding exon

Genotype	Non-synonymous substitutions lying in the GT-5 domain region	Exon Number
Ae. tauschii acc.pau 14102	A (93)	1
Ae. tauschii acc. pau 3747	A (93)	1
Ae. speltoides acc. pau 15081	Q (97)	1
T. dicoccoides acc. pau 7107	T (96)	1
Impala	D (142)	2
C591	D (197)	4

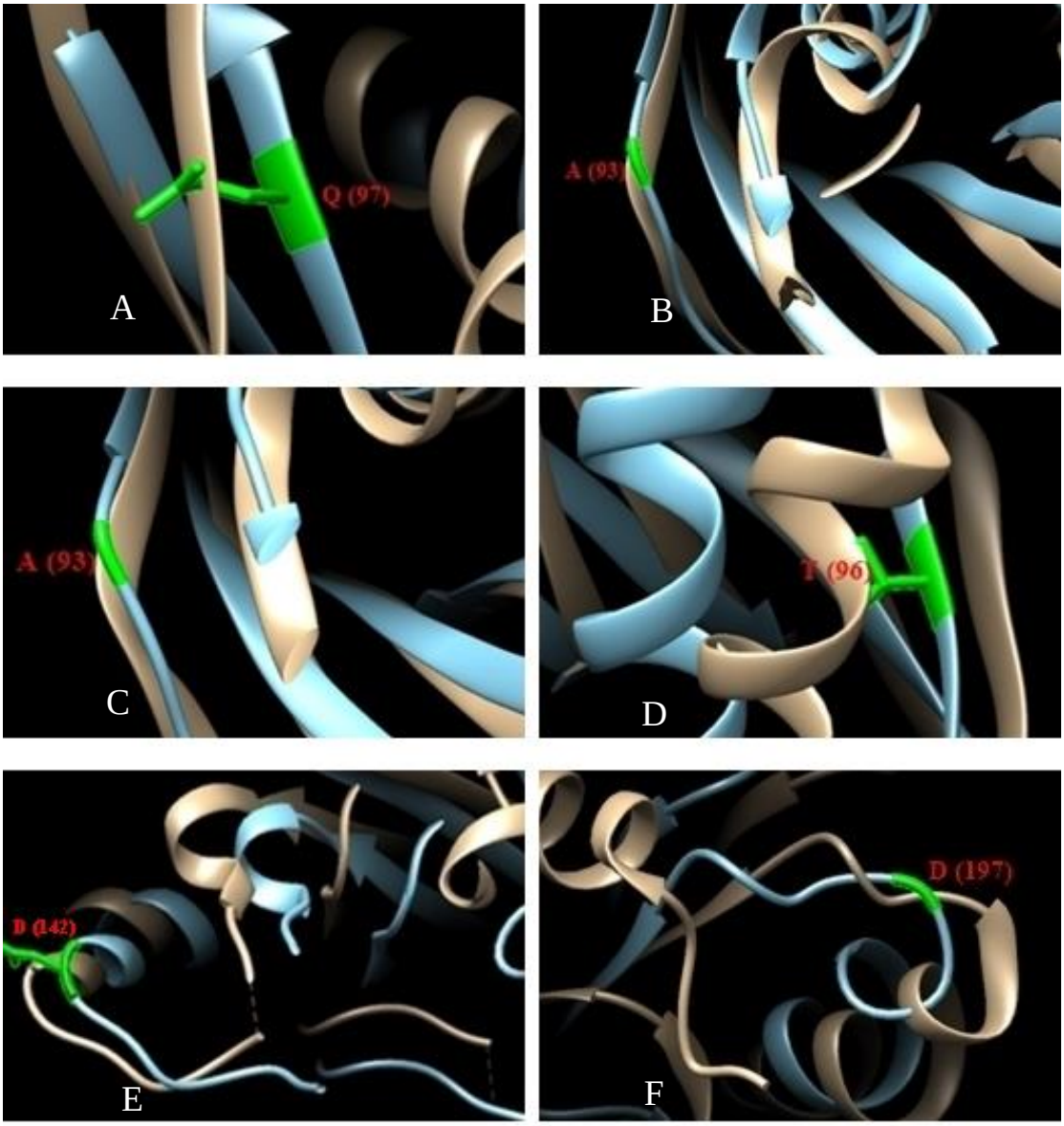

Fig. 32 Non-synonymous substitution along with the position (green colour) detected in GT-5 domain

(A) *Ae. speltoides acc. pau 15081* - Q (97) (B) *Ae. tauschii acc. pau 14102* - A (93) (C) *Ae. tauschii acc. pau 3747* - A (93) (D) *T. dicoccoides acc. pau 7107* - T (96) (E) Impala - D (142) (F) C591 - D (197)

CHAPTER V

SUMMARY

Bread wheat (*Triticum aestivum* L.) ranks second in primary cereal crops of the temperate region. For 40% of the world's population, it serves as a staple food (FAO) with three sub genomes A, B and D with genome size of 17,000 Mb (Devos and Gale 2000). It is estimated before 2020, due to increasing world population the global demand for wheat will increase further by 40% therefore for wheat breeding programs, higher yield becomes predominating objective. Increasing temperature associated with climate change could negatively affect global wheat yields resulting in potential increase in food insecurity and poverty. The grain-filling period in wheat is most sensitive to high temperature, which in turn reduce starch accumulation and transcript profiles for the starch biosynthetic enzymes, demonstrating that high temperature significantly affects the overall development of grain.

In the present study, allelic variation for soluble starch synthase I (*SSI*) gene was identified in different wild and cultivated genotypes of wheat. Nine wild genotypes includes two A genome species - *T. monococcum* and *T. boeoticum*, two D genome species - *Ae. tauschii*, three S genome species - *Ae. speltoides* and two AB genome species - *T. dicoccoides*. 11 cultivated genotypes including one durum wheat - MACS9, two hexaploid winter wheats - Arbon and Impala and eight hexaploid spring wheat genotypes C306, C591, Giza, PBW343, WH542, Halna, Raj3765 and K9644 (Table 1).

Candidate gene for soluble starch synthase (*SS*) gene was identified by *in-silico* approach using NCBI/EMBL and CDS database. NCBI and EMBL databases were searched for nucleotide sequences of full length *SS* gene. 24 sequences including four different isoforms of *SS* gene - *SSI, SSII, SSSIII* and *SSIV* were selected after clustering to remove common sequences using CD-HIT Suite tool at 95% identity. The selected 24 sequences were used as query for online BLAST against IWGSC (Ensembl Plants) *Triticum aestivum* database as subject and 619 sequences were selected on basis of lowest e-value, highest bit score and gene functional annotation.

Full length coding sequences (CDS) of *SS* gene were also retrieved from TriFLDB database. Common 24 sequences between NCBI/EMBL and CDS databases were selected and used as query and standalone (offline) blast was done against the cDNA IWGSC (Ensembl Plants) *Triticum aestivum* database as subject. These combined 24 sequences were also blasted against genomic (gDNA) *Triticum aestivum* database as subject and finally 16 common sequences in genomic and CDS databases were selected for having maximum query coverage. Phylogenetic and molecular evolutionary analysis of these 16 selected full length CDS sequences were conducted using Phylogeny.fr tool using multiple sequence alignment files indicated that *SSI* and *SSII* isoforms of gene were diverse from *SSIII* and *SSIV* isoforms.

Out of four *SS* isoforms, *SSI* showed higher expression in grain as compared to other tissues of plant and was also giving higher expression under heat stress condition. Within *SSI*, 7B homoeologue contributes the majority of transcripts to the overall expression as compared to A and D homoeologue. Therefore, *SSI*-7B homoeologue was selected as candidate gene for present study.

Using *SSI*-7B homeologue gene (4.9 kb), three different sets of primers were designed using Perlprimer software. Size of selected gene is 5156 bp, but primers were designed to amplify gene starting from nucleotide position 130 and ending on position 4906. First set of primer pair (SFLF and SFLR) was designed to amplify full gene from nucleotide position 130 to position 4906. This primer could not amplify gene in any of the wheat genotypes under study. Second set of primers include two overlapping primer pairs SSL1 and SSL2 amplifying gene in two fragments of size 2289 and 2419 base pairs. These primers could amplify the expected fragments in cultivated hexaploid and tetraploid wheats only but not in wild diploid and tetraploid wheat. Thus third set of seven overlapping primer pairs (SSOL1-SSOL7) were designed for amplification and sequencing of gene in seven small parts with the amplicon length ranging from 830 to 900 bases. Two additional primers SSOL4A and SSOL4B were also designed for sequencing two fragments of size 2289 and 2419. Long range (LR) PCR amplification was carried out using Takara Ex Taq Polymerase. The PCR amplified products were resolved by 0.8 % agarose gel and the desired fragments were purified and cloned into pGEM®-Teasy vector and transformed into *E. coli*, DH-5α host strain. White colonies carrying the recombinant plasmid were selected through blue/white screening and desired clones carrying inserts were confirmed by colony PCR amplification using the SSOL3 and SSOL5 overlapping primers. Sequencing was performed with Big Dye Terminator v3.1 (ABI Sequencer 3730xl).

Sequence data of approximately 700-800bp long reads was obtained using ABI 3730xl for all genotypes. The sequences were extracted from Chromatograms using Chromas Lite software. In the total of 1118 sequences obtained (each clone was sequenced in both forward and reverse direction), the sequence of the each individual strand was completed by manual aligning of the sequence of one strand with that of the reverse complimentary sequence obtained with reverse primer and misread bases in each sequence were removed. After combining the overlapping amplicon sequences of the forward and reverse sequences, contigs were generated using DNA Baser v4.23.0. 273 sequences corresponding to three genotypes *Ae. speltoides acc. pauTA1784,* Raj3765 and K9644 could not be obtained.

The contigs were taken as queries and blasted against sequences of *SSI* homoeologous genes (7A, 7B and 7D) using NCBIblast2seq tool for the authenticity that they belong to 7B chromosome only. Exonic and intronic boundaries of *SSI* gene were predicted using Artemis tool and 9 exons and 8 introns identified in these sequences were similar to reference *SSI* gene

accession Traes_7BS_6135B1D85.1. In the selected 17 genotypes intronic regions found to be more polymorphic than exonic regions as introns contained 286 SNPs while exons contained 45 SNPs. First intron (99 SNPs) and third intron (153 SNPs) showed maximum variations while second and eighth introns were fully conserved. Fourth intron has 15 SNPs, while intron five (7 SNPs), six (3 SNPs) and seven (9 SNPs) have comparatively few SNPs. In case of exons maximum SNPs (18) were detected in first exon while in rest of the exons 3-6 SNPs were detected. The detected SNPs found to be comprised of 221 base transitions and 110 base transversions.

No insertions were detected in any of 17 selected wheat genotypes. However, deletions were identified in intronic regions of the wild genotypes namely *Triticum monococcum acc. pau14087*, *Triticum boeoticum acc. pau5088*, *Ae. speltoides acc. pau3809*, *Ae. speltoides acc. pau15081*, *Ae. tauschii acc. pau14102* and *Triticum dicoccoides acc. pau14801*. No deletion was detected in exonic regions.

Homology modeling approach was employed to determine the structure of proteins based on the known structure of template proteins using Modeller 9.17 tool. *SSI* protein in barley (PDB id 4hln) showing 98% sequence identity with wheat reference *SSI* sequence was used as template. The eight active site residues namely F (535), P (537), F (538), E (543), W (548), T (564), H (572) and S (575) fall outside of both GT domains (GT-1 and GT-5) in barley template. The active site residues showed common differences in the selected wild and cultivated genotypes when compared with barley protein. Active site corresponding to barley T (564) was the most variable as it has different variant in all seven genotypes with non-synonymous substitution. But none of predicted active sites in the genotypes under study corresponds to non-synonymous substitutions.

Out of 45 exonic SNPs, 33 synonymous substitutions were identified in seven wild and five cultivated genotypes. 12 non-synonymous substitutions were identified in four wild (*Aegilops speltoides acc. pau15081, Ae. tauschii acc. pau3747, Ae. tauschii acc. pau14102, T. dicoccoides acc. pau7107*) and three cultivated genotypes (MACS9, Impala, C591). Six of these non-synonymous substitutions fall in important catalytic domain glycosyltransferase 5 (GT-5). Four of these non-synonymous substitutions were from exon 1 while other two are from exon 2 and exon 4 respectively. Thus alleles identified in four wild genotypes of *Aegilops speltoides acc. pau15081, Ae. tauschii acc. pau3747, Ae. tauschii acc. pau14102, T. dicoccoides acc. pau7107* and two cultivated genotypes Impala, C591 were important candidates for variants in *SS* gene.

REFERENCES

Abel G J W, Springer F and Willmitzer L (1996) Cloning and functional analysis of a cDNA encoding a novel 139 kDa starch synthase from potato (*Solanum tuberosum* L.). *Plant J* **10**: 981-91.

Altenbach S B and Kothari K M (2004) Transcript profiling of genes expressed in endosperm tissue are altered by high temperature during wheat grain development. *J Cereal Sci* **40**: 115-26.

Awlachew Z T, Singh R, Kaur S, Bains N S and Chhuneja P (2016) Transfer and mapping of the heat tolerance component traits of *Aegilops speltoides* in tetraploid wheat *Triticum durum. Mol Breed* **36**: 78.

Baba T, Nishihara M and Mizuno K (1993) Identification, cDNA cloning, and gene expression of soluble starch synthase in rice (*Oryza sativa* L.) immature seeds. *Plant Physiol* **103**: 565-73.

Ball S G and Morell M K (2003) From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* **54**: 207-33.

Ball S, Guan H, James M, Myers A, Keeling P, Mouille G, Buleon A, Colonna P and Preiss J (1996) From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. *Cell* **86**: 349-52.

Bates F L, French D and Rundle R E (1943) Amylose and amylopectin content of starches determined by their iodine complex formation. *J Am Chem Soc* **65**: 142-48.

Beales J, Laurie D A and Devos K M (2005) Allelic variation at the linked *AP1* and *PhyC* loci in hexaploid wheat is associated but not perfectly correlated with vernalization response. *Theor Appl Genet* **110**: 1099-107.

Berriman M and Rutherford K (2003) Viewing and annotating sequence data with Artemis. *Brief bioinform* **4**: 124-32.

Bertoft E, Laohaphatanaleart K and Piyachomkwan K (2010) The fine structure of cassava amylopectin. Part 2. Building block structure of clusters. *Int J Biol Macromol* **47**: 325-35.

Bertoft E, Piyachomkwan K and Chatakanonda P (2008) Internal unit chain composition in amylopectins. *Carbohydr Polym* **74**: 527-43.

Bhullar S S and Jenner C F (1985) Differential responses to high temperature of starch and nitrogen accumulation in the grain of four cultivars of wheat. *Aust J Plant Physiol* **12**: 363-75.

Breathnach R and Chambon P (1981) Organization and expression of eukaryotic split genes coding for proteins. *Annu Rev Biochem* **50**: 349-83.

Buléon A, Colonna P, Planchot V and Ball S (1998) Starch granules: structure and biosynthesis. *Int J Biol Macromol* **23**: 85-112.

Buleon A, Veronese G and Putaux J L (2007) Self-association and crystallization of amylose. *Aust J Chem* **60**: 706-18.

Cameron R E and Donald A M (1992) A small-angle X-ray scattering study of the annealing and gelatinization of starch. *Polymer* **33**: 2628-635.

Cao H, Inparl-Radosevich J, Guan H, Keeling P L, James M G and Myers A M (1999) Identification of the soluble starch synthase activities of maize endosperm. *Plant Physiol* **120**: 205-15.

Clark A G, Leicht B G, and Muse S V (1996) Length Variation and Secondary Structure of Introns in the *Mlcl* Gene in Six Species of *Drosophila*. *Mol Biol Evol* **13**: 471-82.

Cohen S N, Chang A C Y and Hsu L (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci USA* **69**: 2110-114.

Comai L, Young K, Till B J, Reynolds S H, Greene E A and Codomo C A (2004) Efficient discovery of DNA polymorphisms in natural populations by EcoTilling. *Plant J* **37**: 778-86.

Commuri P D and Keeling P L (2001) Chain-length specificities of maize starch synthase I enzyme: studies of glucan affinity and catalytic properties. *Plant J* **25**: 475-86.

Craig J, Lloyd J R and Tomlinson K (1998) Mutations in the gene encoding starch synthase II profoundly alter amylopectin structure in pea embryos. *Plant Cell* **10**: 413-26.

Delvallé D, Dumez S, Wattebled F, Roldán I, Planchot V, Berbezy P, Colonna P, Vyas D, Chatterjee M, Ball S, Mérida A and D'Hulst C (2005) Soluble starch synthase I: a major determinant for the synthesis of amylopectin in *Arabidopsis thaliana* leaves. *Plant J* **43**: 398-412.

Denyer K, Waite D, Motawia S, Moller B L and Smith A M (1999) Granule-bound starch synthase I in isolated starch granules elongates malto-oligosaccharides processively. *Biochem J* **340**: 183-91.

Denyer K and Smith A M (1992) The purification and characterisation of two forms of soluble starch synthase from developing pea embryos. *Planta* **186**: 609-17.

Denyer K, Hylton C M and Smith A M (1994) The effect of high temperature on starch synthesis and the activity of starch synthase. *Aust J Plant Physiol* **21**: 783-89.

Denyer K. Sidebottom C, Hylton C M and Smith A M (1993) Soluble isoforms of starch synthase and starch branching enzyme also occur within starch granulesin developing pea embryos. *Plant J* **4**: 191-98.

Devos K M and Gale M D (2000) Genome relationships: the grass model in current research. *Plant Cell* **12**: 637-46.

Dian W, Jiang H and Wu P (2005) Evolution and expression analysis of starch synthase III and IV in rice. *J Exp Bot* **56**: 623-32.

Dvorak J, Luo M C and Yang Z L (1998) Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross- fertilizing *Aegilops* species. *Genetics* **148**: 423-34.

Edwards A, Borthakur A and Bornemann S (1999) Specificity of starch synthase isoforms from potato. *Eur J Biochem* **266**: 724-36.

Edwards A, Marshall J, Sidebottom C, Visser R G F, Smith A M and Martin C (1995) Biochemical and molecular characterization of a novel starch synthase from potato tubers. *Plant J* **8**: 283-94.

Feldman M (2001) origin of cultivated wheat. In: Bonjean A P, Angus WJ (ed). *The world wheat book: A history of wheat breeding.* Pp 3-56. Paris: Lavoisier Tech and Doc.

Fettke J, Albrecht T and Hejazi M (2010) Glucose 1-phosphate is efficiently taken up by potato (*Solanum tuberosum*) tuber parenchyma cells and converted to reserve starch granules. *New Phytol* **185**: 663-75.

Fontaine T, D'Hulst C and Maddelein M L (1993) Toward an understanding of the biogenesis of the starch granule. Evidence that *Chlamydomonas* soluble starch synthase II controls the synthesis of intermediate size glucans of amylopectin. *J Biol Chem* **268**: 16223-30.

French D (1972) Fine structure of starch and its relationship to the organization of starch granules. *J Jpn Soc Starch Sci* **19**: 8-25.

Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H and Nakamura Y (2006) Function and characterization of starch synthase I using mutants in rice. *Plant Physiol* **140**: 1070-84.

Fulton D C, Edwards A, Pilling E, Robinson H L, Fahy B, Seale R, Kato L, Donald A M, Geigenberger P, Martin C and Smith A M (2002) Role of granule-bound starch synthase in determination of amylopectin structure and starch granule morphology in potato. *J Biol Chem* **277**: 10834-41.

Gallant D J, Bouchet B and Baldwin P M (1997) Microscopy of starch: evidence of a new level of granule organization. *Carbohydr Polym* **32**: 177-91.

Ganeshan S, Drinkwater J M, Repellin A and Chibbar R N (2010) Selected carbohydrate metabolism genes show coincident expression peaks in grains of *in vitro* cultured immature spikes of wheat (*Triticum aestivum* L.). *J Agr Food Chem* **58**: 4193-201.

Gao M and Chibbar R N (2000) Isolation, characterization and expression analysis of starch synthase IIa cDNA from wheat (*Triticum aestivum* L.). *Genome* **43**: 768-75.

Gao M, Wanat J, Stinard P S, James M G and Myers A M (1998) Characterization of *dull1*, a maize gene coding for a novel starch synthase. *Plant Cell* **10**: 399-412.

Golovnina K A, Kondratenko E Y, Blinov A G and Goncharov N P (2010) Molecular characterization of vernalization loci *VRN1* in wild and cultivated wheats. *BMC Plant Biol* **10**: 168.

Gubler F, Li Z, Fieg S, Jacobsen J V and Morell M K (2000) Cloning and characterization of a starch synthase I gene (Accession No. AF234163) from barley (PGR00-047). *Plant Physiol* **122**: 1459.

Gunning A P, Giardina T P and Faulds C B (2003) Surfactant-mediated solubilisation of amylase and visualisation by atomic force microscopy. *Carbohydr Polym* **51**: 177-82.

Gupta P K and Saripalli G (2015) AGPase: its role in crop productivity with emphasis on heat tolerance in cereals. *Theor Appl Genet* **128**: 1893-916.

Haga H, Yamada R, Ohnishi Y, Nakamura Y and Tanaka T (2002) Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. *J Hum Genet* **47**: 605-10.

Hanashiro I and Takeda Y (1998) Examination of number-average degree of polymerization and molar- based distribution of amylose by fluorescent labeling with 2-aminopyridine. *Carbohydr Res* **306**: 421-26.

Hanson E A and Katz J R (1934) Abhandlungen zur physikalischen Chemie der Starke und der Brotbereitung. XVII. Uber Versuche die gewachsene Struktur des Starkekorns mikroskopisch sichtbar zu machen besonders an lintnerisierter Starke. *Z physikal Chem abt* **168**: 339-52.

Hawker J S and Jenner C F (1993) High temperature affects the activity of enzymes in the committed pathway of starch synthesis in developing wheat endosperm. *Aust J Plant Physiol* 20: 197-209.

Hirose T and Terao T (2004) A comprehensive expression analysis of the starch synthase gene family in rice (*Oryza sativa* L.). *Planta* **220**: 9-16.

Hirose T, Ohdan T, Nakamura Y and Terao T (2006) Expression profiling of genes related to starch synthesis in rice leaf sheaths during the heading period. *Physiol Plant* **128**: 425-35.

Hizukuri S (1986) Polymodal distribution of the chain lengths of amylopectins, and its significance. *Carbohydr Res* **147**: 342-47.

Hizukuri S and Takagi T (1984) Estimation of the distribution of molecular weight for amylose by the low-angle laser-light-scattering technique combined with high-performance gel chromatography. *Carbohydr Res* **134**: 1-10.

Hizukuri S, Takeda Y and Maruta N (1989) Molecular structures of rice starch. *Carbohydr Res* **189**: 227- 35.

Huang X Q and Babel A B (2010) Development of genome-specific primers for homoeologous genes in allopolyploid species: the *waxy* and *starch synthase II* genes in allohexaploid wheat (*Triticum aestivum* L.) as examples. *BMC Res Not* **3**: 140-51.

Huang X Q and Babel A B (2012) Sequence diversity, haplotype analysis, association mapping and functional marker development in the waxy and starch synthase IIa genes for grain-yield-related traits in hexaploid wheat (*Triticum aestivum* L.). *Mol Breed* **30**: 627-45.

Hurkman W J, McCue K F, Altenbach S B, Korn A, Tanaka C K, Kothari K M, Johnson E L, Bechtel D B, Wilson J D, Anderson O D and DuPont F M (2003) Effect of temperature on expression of genes encoding enzymes for starch biosynthesis in developing wheat endosperm. *Plant Sci* **164**: 873-81.

Hutchison C A (2007) DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res* **35**: 6227-237.

Imberty A, Buleon A and Tran V (1991) Recent advances in knowledge of starch structure. *Starch Starke* **43**: 375-84.

Imberty A, Chanzy H and Perez S (1988) The double-helical nature of the crystalline part of A-starch. *J Mol Biol* **201**: 365-78.

James M G, Denyer K and Myers A M (2003) Starch synthesis in the cereal endosperm. *Curr Opin Plant Biol* **6**: 215-22.

Jenner C F, Siwek K and Hawker J S (1993) The synthesis of (14C) starch from (14C) sucrose in isolated wheat grains is dependent upon the activity of soluble starch synthase. *Aust J Plant Physiol* **20**: 329-35.

Jeon J S, Ryoo N, Hahn T R, Walia H and Nakamura Y (2010) Starch biosynthesis in cereal endosperm. *Plant Physiol Biochem* **48**: 383-92.

Jia X Y, Xu C Y, Jing R L, Li R Z, Mao X G, Wang J P and Chang X P (2008) Molecular cloning and characterization of wheat calreticulin (CRT) gene involved in drought-stressed responses. *J Exp Bot* **59**: 739-51.

Juliano B O, Tuano A P P and Monteroso D N (2012) Replacement of acetate with ammonium buffer to determine apparent amylose content of milled rice. *Cereal Food World* **57**: 14-19.

Kadaru S B, Yadav A S, Fjellstrom R G and Oard J H (2006) Alternative Ecotilling protocol for rapid, cost-effective single-nucleotide polymorphism discovery and genotyping in rice (*Oryza sativa* L.). *Plant Mol Biol Rep* **24**: 3-22.

Kalinga D, Waduge R and Bertoft E (2013) On the differences in granular architecture and starch structure between pericarp and endosperm wheat starches. *Starch Starke* **65**: 791-800.

Kaur N, Street K, Mackay M, Yahiaoui N and Keller B (2008) Allele mining and sequence diversity at the wheat powdery mildew resistance locus *Pm3*. *Mol Breed* **78**: 1-3.

Keeling P L and Myers A M (2010) Biochemistry and Genetics of Starch Synthesis. *Annu Rev Food Sci* **1**: 271-303.

Keeling P L, Bacon P J and Holt D C (1993) Elevated temperature educes starch deposition in wheat endosperm by reducing the activity of soluble starch synthase. *Planta* **191**: 342-48.

Keeling P L, Banisadr R, Barone L, Wasserman B P and Singletary G W (1994) Effect of temperature on enzymes in the pathway of starch biosynthesis in developing wheat and maize grain. *Aust J Plant Physiol* **21**: 8072-77.

Kitahara K, Fukunaga S and Katayama K (2005) Physicochemical properties of sweet potato starches with different gelatinization temperatures. *Starch* **57**: 473-79.

Knight M E, Harn C, Lilley C E R, Guan H P, Singletary G, Mu-Forster C, Wasserman B P and Keeling P L (1998) Molecular cloning of starch synthase I from maize (W64A) endosperm and expression in *E. coli*. *Plant J* **14**: 613-22.

Koroteeva D A, Kiseleva V I and Krivandin A V (2007) Structural and thermodynamic properties of rice starches with different genetic background. *Int J Biol Macromol* **41**: 534-47.

Kossmann J, Abel G J W, Springer F, Lloyd J R and Willmitzer L (1999) Cloning and functional analysis of a cDNA encoding a starch synthase from potato (*Solanum tuberosum* L.) that is predominantly expressed in leaf tissue. *Planta* **208**: 503-11.

Kumar R R, Sharma S K, Goswami S, Singh G P, Singh R and Singh K (2013) Characterization of differentially expressed stress-associated proteins in starch granule development under heat stress in wheat (*Triticum aestivum* L.). *Indian J Biochem Biophys* **50**: 126-38.

Labuschagne M T, Elago O and Koen E (2009) The influence of temperature extremes on some quality and starch characteristics in bread, biscuit and durum wheat. *J Cereal Sci* **49**: 184-89.

Lafiandra D, Sestili F, D'Ovidio R, Janni M, Botticella E, Ferrazzano G, Silvestri M, Ranieri R and DeAmbrogio E (2010) Approaches for the modification of starch composition durum wheat. *Cereal Chem* **87**: 28-34.

Leterrier M, Holappa L D, Broglie K E and Beckles D M (2008) Cloning, characterisation and comparative analysis of a starch synthase IV gene in wheat: functional and evolutionary implications. *BMC Plant Biol* **8**: 98.

Li J, Ezquer I and Bahaji A (2011) Microbial volatile-induced accumulation of exceptionally high levels of starch in *Arabidopsis* leaves is a process involving NTRC and starch synthase classes III and IV. *Mol Plant-Microbe Interact* **24**: 1165-178.

Li W, Liu A, Sheng Y Z, Pu Z, Liu Y and Cheng G Y (2013) Molecular characterization of *SSI* gene in *Triticum* L. and *Aegilops tauschii*. *J Biol Sci* **13**: 528-33.

Li Z, Chu X, Mouille G, Yan L, Hashemi B K, Hey S, Napier J, Shewry P, Clarke B, Appels R, Morell M K and Rahman S (1999b) The localization and expression of the class II starch synthases of wheat. *Plant Physiol* **120**: 1147-55.

Li Z, Mouille G, Hashemi B K, Rahman S, Clarke B, Gale K R, Appels R and Morel M K (2000) The structure and expression of the wheat starch synthase III gene. Motifs in the expressed gene define the lineage of the starch synthase III gene family. *Plant Physiol* **123**: 613-24.

Li Z, Rahman S, Kosar-Hashemi B, Mouille G, Appels R and Morell M K (1999a) Cloning and characterization of a gene encoding wheat starch synthase I. *Theor Appl Genet* **98**: 1208-216.

Li Z, Sun F, Xu S, Chu X, Mukai Y, Yamamoto M, Ali S, Rampling L, Kosar-Hashemi B, Rahman S and Morell M K (2003) The structural organisation of the gene encoding class II starch synthase of wheat and barley and the evolution of the genes encoding starch synthases in plants. *Funct Int Genom* **3**: 76-85.

Limbore V N and Khillare S K (2015) An analytical study of Indian agriculture crop production and export with reference to wheat. *Rev Res* **4**: 1-8.

Liu H, Yu G, Wei B, Wang Y, Zhang J, Hu Y, Liu Y, Yu G, Zhang H and Huang Y (2015) Identification and phylogenetic analysis of a novel starch synthase in maize. *Front Plant Sci* **6**: 1-13.

Lobell D B, Schlenker W and Costa-Roberts J (2011) Climate trends and global crop production since 1980. *Science* **333**: 616-20.

Lobell D B, Sibley A and Ortiz-Monasterio J I (2012) Extreme heat effects on wheat senescence in India. *Nat Clim Change* **2**: 186-89.

Luo G H, Li X H, Han Z J, Zhang Z C, Yang Q, Guo H F and Fang J C (2016) Transition and Transversion Mutations Are Biased towards GC in Transposons of *Chilo suppressalis* (Lepidoptera: Pyralidae). *Genes* **7**: 72.

MacLeod L C and Duffus C M (1988) Reduced starch content and sucrose synthase activity in developing endosperm of barley plants grown at elevated temperature. *Aust J Plant Physiol* **15**: 36.

Maestra B and Naranjo T (1998) Homoeologous relationships of *Aegilops speltoides* chromosomes to bread wheat. *Theor Appl Genet* **97**: 181-86.

Mardis E R (2008) Next-generation DNA sequencing methods *Annu Rev Genom Hum Genet* **9**: 387- 402.

Marshall J, Sidebottom C and Debet M (1996) Identification of the major starch synthase in the soluble fraction of potato tuber. *Plant Cell* **8**: 1121-35.

Marshall O (2004) PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real- time PCR. *Bioinformatics* **20**: 2471-72.

McCallum C M, Comai L, Greene E A and Henikoff S (2000) Targeted screening for induced mutations. *Nat Biotechnol* **18:** 455-57.

McMaugh S J, Thistleton J L, Anschaw E, Luo J, Konik-Rose C, Wang H, Huang M, Larroque O, Regina A, Jobling S A, Morell M K and Li Z (2014) Suppression of starch synthase I expression affects the granule morphology and granule size and fine structure of starch in wheat endosperm. *J Exp Bot* **65**: 2189-220.

Monari A M, Simeone M C, Urbano M, Margiotta B and Lafiandra D (2005) Molecular characterization of new waxy mutants identified in bread and durum wheat. *Theor Appl Genet* **110**: 1481-89.

Morrell P L, Lundy K E and Clegg M T (2003) Distinct geographic patterns of genetic diversity are maintained in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite migration. *P Natl Acad Sci USA* **100**: 10812-17.

Murray M G and Thompson W F (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* **8**: 4321-25.

Nakamura Y, Francisco P B and Hosaka Y (2005) Essential amino acids of starch synthase IIa differentiate amylopectin structure and starch quality between *japonica* and *indica* rice varieties. *Plant Mol Biol* **58**: 213-27.

Nougue O, Corbi J, Ball S G, Manicacci D and Tenaillon M I (2014) Molecular evolution accompanying functional divergence of duplicated genes along the plant starch biosynthesis pathway. *BMC Evol Biol* **14**: 103.

Ohtani T, Yoshino T and Hagiwara S (2000) High-resolution imaging of starch granule structure using atomic force microscopy. *Starch Starke* **52**: 150-53.

Ortiz R, Sayre K D, Govaerts B, Gupta R, Subbarao G V, Ban T, Hodson D, Dixon J M, Ortiz- Monasterio J I and Reynolds M (2008) Climate change:Can wheat beat the heat? *Agric Ecosys Environ* **126**: 46-58.

Pan X, Yan H, Li M, Wu G and Jiang H (2011) Evolution of the Genes Encoding Starch Synthase in Sorghum and Come Wheat. *Mol Plant Breed* **2**: 60-67.

Park Y J and Nishikawa T (2012) Characterization and expression analysis of the starch synthase gene family in grain amaranth (*Amaranthus cruentus* L.). *Genes Genet Syst* **87**: 281-89.

Pearce S, Gross H V, Herin S Y, Hane D, Wang Y, Gu Y Q and Dubcovsky J (2015) WheatExp: an RNA-seq expression database for polyploid wheat. *BMC Plant Biol* **15**: 299.

Peng M, Hucl P and Chibbar R N (2001) Isolation, characterization and expression analysis of starch synthase I from wheat (*Triticum aestivum* L.). *Plant Sci* **161**: 1055-62.

Peng S, Huang J, Sheehy J E, Laza R C, Visperas R M, Zhong X, Centeno G S, Khush G S and Cassman K G (2004) Rice yields decline with higher night temperature from global warming. *Proc Natl Acad Sci USA* **101**: 9971.

Pettersen E F, Goddard T D, Huang C C, Couch G S, Greenblatt D M, Meng E C and Ferrin T E (2004) UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-12.

Pittock B (2003) *Climate Change: An Australian Guide to the Science and Potential of Impacts*. Department for the Environment and Heritage, Australian Greenhouse Office, Canberra, ACT.

Popov D, Buleon A and Burghammer M (2009) Crystal structure of A-amylose: a revisit from synchrotron microdiffraction analysis of single crystals. *Macromolecules* **42**: 1167-74.

Pradhan G P, Prasad P V V, Fritz A K, Kirkham M B and Gill B S (2012) High temperature tolerance in *Aegilops* species and its potential transfer to wheat. *Crop Sci* **52**: 292-304.

Prakash P, Kumari A, Singh D V, Pandey R, Sharma N P and Ghildiyal M C (2009) Starch synthase activity and grain growth in wheat cultivars under elevated temperature: a comparison of responses. *Indian J Plant Physiol* **14**: 364-69.

Rahman S, Li Z, Batey I, Cochrane M P, Appels R and Morell M (2000) Genetic alteration of starch functionality in wheat. *J Cereal Sci* **31**: 91-110.

Ral J P, Colleoni C and Wattebled F (2006) Circadian clock regulation of starch metabolism establishes *GBSSI* as a major contributor to amylopectin synthesis in *Chlamydomonas reinhardtii*. *Plant Physiol* **142**: 305-17.

Ramakrishna W, Dubcovsky J, Park Y J, Busso C, Embereton J, SanMiguel P and Bennetzen J L (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. *Genetics* **162**: 1389-1400.

Raman H, Raman R, Wood R and Martin P (2006) Repetitive indel markers within the *ALMT1* gene controlling aluminium tolerance in wheat (*Triticum aestivum* L). *Mol Breed* **18**: 171-83.

Raman H, Ryan P R, Raman R, Stodart B J, Zhang K, Martin P, Wood R, Sasaki T, Yamamoto Y and Mackay M (2008) Analysis of *TaALMT1* traces the transmission of aluminum resistance in cultivated common wheat (*Triticum aestivum* L.). *Theor Appl Genet* **116**: 343-54.

Ridout M J, Parker M L and Hedley C L (2003) Atomic force microscopy of pea starch granules: granule architecture of wild-type parent, *r* and *rb* single mutants, and the *rrb* double mutant. *Carbohydr Res* **338**: 2135-47.

Ridout M J, Parker M L and Hedley C L (2006) Atomic force microscopy of pea starch: granule architecture of the *rug3-a, rug4-b, rug5-a* and *lam-c* mutants. *Carbohydr Polym* **65**: 64-74.

Rijven A H G C (1986) Heat inactivation of starch synthase in wheat endosperm tissue. *Plant Physiol* **81**: 448-53.

Robbins M D, Sim S C, Yang W, Deynze A V, Knaap E, Joobeur T and Francis D M (2011) Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato. *J Exp Bot* **62**: 1831-845.

Robin J P, Mercier C and Charbonniere R (1974) Lintnerized starches. Gel filtration and enzymatic studies of insoluble residues from prolonged acid treatment of potato starch. *Cereal Chem* **51**: 389-406.

Roldan I, Wattebled F and Lucas M M (2007) The phenotype of soluble starch synthase IV defective mutants of *Arabidopsis thaliana* suggests a novel function of elongation enzymes in the control of starch granule formation. *Plant J* **49**: 492-504.

Rose M K, Huang X Q and Babel A B (2015) Molecular characterization and sequence diversity of genes encoding the large subunit of the ADP-glucose pyrophosphorylase in wheat (*Triticum aestivum* L.*). J Appl Genet* **57**: 15-25.

Saiki R K, Gelfand D H, Stoffel S, Scharf S J, Higuchi R, Horn G T, Mullis K B and Erlich H A (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-91.

Saito M and Nakamura T (2005) Two point mutations identified in emmer wheat generate null Wx- A1 alleles. *Theor Appl Genet* **110**: 276-82.

Sankoff D, Cedergren R J and Lapalme G (1976) Frequency of insertion-deletions, transversion and transition in the evolution of 5S ribosomal RNA. *J Mol Evol* **7**: 133-49.

Sasaki T, Ryan P R and Delhaize (2006) ESequence upstream of the wheat (*Triticum aestivum* L.) *ALMT1* gene and its relationship to aluminum resistance. *Plant Cell Physiol* **47**: 1343-54.

Shimbata T, Nakamura T, Vrinten P, Saito M, Yonemaru J and Seto Y (2005) Mutations in wheat starch synthase II genes and PCR-based selection of a SGP-1 null line. *Theor Appl Genet* **111**: 1072-79.

Shirasawa K, Isobe S, Hirakawa H, Asamizu E, Fukuoka H, Just D, Rothan C, Sasamoto S, Fujishiro T, Kishida Y, Kohara M, Tsuruoka H, Wada T, Nakamura Y, Sato S and Tabata S (2010) SNP discovery and linkage map construction in cultivated tomato. *DNA Res* **17**: 381-91.

Stamova B S, Laudencia-Chingcuanco D and Beckles D M (2009) Transcriptional analysis of starch biosynthesis in the developing grain of hexaploid wheat. *Int J Plant Genom* **8**: 23.

Sterling C (1960) Crystallinity of potato starch. *Starch Starke* **12**: 182-85.

Takeda C, Takeda Y and Hizukuri S (1989) Structure of amylomaize amylose. *Cereal Chem* **66**: 22-25.

Takeda Y, Hizukuri S and Takeda C (1987) Structures of branched molecules of amyloses of various origins, and molar fractions of branched and unbranched molecules. *Carbohydr Res* **165**: 139-45.

Takeda Y, Shibahara S and Hanashiro I (2003) Examination of the structure of amylopectin molecules by fluorescent labeling. *Carbohydr Res* **338**: 471-75.

Takeda Y, Takeda C and Mizukami H (1999) Structures of large, medium and small starch granules of barley grain. *Carbohydr Polym* **38**: 109-14.

Tatge H, Marshall J, Martin C, Edwards E A and Smith A M (1999) Evidence that amylose synthesis occurs within the matrix of the starch granule in potato tubers. *Plant Cell Environ* **22**: 543-50.

Tomlinson K and Denyer K (2003) Starch synthesis in cereal grains. *Adv Bot Res* **40**: 1-61.

Toyosawa Y, Kawagoe Y and Matsushima R (2015) Deficiency of starch synthase IIIa and IVb leads to dramatic changes in starch granule morphology in rice endosperm. *Starch* **51**: 58-62.

Tubiello F N, Rosenzweig C, Goldberg R A, Jagtap S and Jones J W (2000) U.S. National assessment technical report effects of climate change on U.S. crop production part I: Wheat, Potato, Corn and Citrus.

Umemoto T, Yano M and Satoh H (2002) Mapping of a gene responsible for the difference in amylopectin structure between *japonica*-type and *indica*-type rice varieties. *Theor Appl Genet* **104**: 1-8.

Vandeputte G E and Delcour J A (2004) From sucrose to starch granule to starch physical behaviour: a focus on rice starch. *Carbohydr Polym* **58**: 245-66.

Vogul F and Kopun M (1977) Higher frequencies of transitions among point mutations. *J Mol Evol* **9**: 159.

Vrinten P L and Nakamura T (2000) Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues. *Plant Physiol* **122**: 255-64.

Waines J G (1994) High temperature in wild wheats and spring wheats. *Aust J Plant Phys* **21**: 705-15.

Wang G X, Tan M K, Rakshit S, Saitoh H, Terauchi R, Imaizumi T, Ohsako T and Tominaga T (2007) Discovery of single-nucleotide mutations in acetolactate synthase genes by Ecotilling. *Pestic Biochem Physiol* **88**: 143-48.

Wang J, Sun J, Liu D, Yang W, Wang D, Tong Y and Zhang A (2008) Analysis of *Pina* and *Pinb* alleles in the micro-core collections of Chinese wheat germplasm by Ecotilling and identification of a novel *Pinb* allele. *J Cereal Sci* **48**: 836-42.

Wang N, Shi L, Tian F, Ning H, Wu X, Long Y and Meng J (2010) Assessment of *FAE1* polymorphisms in three Brassica species using EcoTILLING and their association with differences in seed erucic acid contents. *BMC Plant Biol* **10**: 137.

Wang Z, Li W, Qi J, Shi P and Yin Y (2014) Starch accumulation, activities of key enzyme and gene expression in starch synthesis of wheat endosperm with different starch contents. *J Food Sci Technol* **51**: 419-29.

Wardlaw I F and Wrigley C W (1994) Heat tolerance in temperate cereals: an overview. *Aust J Plant Physiol* **21**: 695-703.

Waterhouse A M, Procter J B, Martin D M, Clamp M and Barton G J (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-91.

Wikman J, Blennow A and Buleon A (2014) Influence of amylopectin structure and degree of phosphorylation on the molecular composition of potato starch lintners. *Biopolymers* **101**: 257- 71.

Xia L Q, Ganal M W, Shewry P R, He Z H, Yang Y and Roder M S (2008) Exploiting the diversity of *Viviparous-1* gene associated with pre-harvest sprouting tolerance in European wheat varieties. *Euphytica* **159**: 411-17.

Yamaguchi M, Sasaki T, Sivaguru M, Yamamoto Y, Osawa H, Ahn S J and Matsumoto H (2005) Evidence for the plasma membrane localization of Al-activated malate transporter (*ALMT1*). *Plant Cell Physiol* **46**: 812-16.

Yamamori M, Fujita S and Hayakawa K (2000) Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose. *Theor Appl Genet* **101**: 21-29.

Yan H B, Jiang H W, Pan XX, Li M R, Chen, Y P and Wu G J (2009) The gene encoding starch synthase IIc exists in maize and wheat. *Plant Sci* **176**: 51-57.

Zakaria S, Matsuda T, Tajima S and Nitta Y (2002) Effect of high temperature at ripening stage on the reserve accumulation in seed in some rice cultivars. *Plant Prod Sci Tokyo* **5**: 160-68.

Zeeman S C, Kossmann J and Smith A M (2010) Starch: Its metabolism, evolution, and biotechnological modification in plants. *Annu Rev Plant Biol* **61**: 209-34.

Zhang W and Dubcovsky J (2008) Association between allelic variation at the Phytoene synthase 1 gene and yellow pigment content in the wheat grain. *Theor Appl Genet* **116**: 635-45.

Zhang X, Colleoni C and Ralushana V (2004) Molecular characterization demonstrates that the *Zea mays* gene *sugary2* codes the starch synthase isoform *SSIIa*. *Plant Mol Biol* **54**: 865-79.

Zhang X, Szydlowski N and Delvalle D (2008) Overlapping functions of the starch synthase *SSII* and *SSIII* in amylopectin biosynthesis in *Arabidopsis*. *BMC Plant Biol* **8**: 96-113.

Zhao H, Dai T, Jiang D and Cao W (2008) Effects of high temperature on key enzymes involved in starch and protein formation in grains of two wheat cultivars. *J Agron Crop Sci* **194**: 47- 54.

Zhong L J, Dong H, Cai X B, Feng Y N, Ren P and Cheng F M (2012) Gene expression of the key enzymes controlling starch synthesis and metabolism in rice grain endosperm under effects of high temperature after anthesis. *Ying Yong Sheng Tai Xue Bao* **23**: 745-50.

VITA

Name of the student	:	MEHAKDEEP SINGH DHAWAN
Father's name	:	Mr. H. S. DHAWAN
Mother's name	:	Mrs. CHARANJIT KAUR DHAWAN
Nationality	:	INDIAN
Date of birth	:	24.01.1992
Permanent home address	:	H.No. 523-R, Model Town Ludhiana-141002

EDUCATIONAL QUALIFICATIONS

Bachelor degree	:	B.Sc. (Hons.) Biotechnology
University and year of award	:	Punjab Agricultural University, Ludhiana, India, 2014
OCPA	:	8.17/10.00
Master's degree	:	M.Sc. (Biotechnology)
University and year of award	:	Punjab Agricultural University, Ludhiana, India, 2017
OCPA	:	8.18/10.00
Title of Master Thesis	:	"Allele mining for heat stress related starch synthase gene in cultivated and wild species of wheat"