ASSOCIATION ANALYSIS OF NOVEL SINGLE NUCLEOTIDE POLYMORPHISM AND ITS EFFECTS ON STRUCTURE AND FUNCTION OF SS1 INVOLVED IN STARCH BIOSYNTHESIS IN RICE LANDRACES OF HIMACHAL PRADESH

Enrollment Number – 111807

Name of Student – Dipti Sharma

Name of Supervisor – Dr. Harvinder Singh

MAY - 2015

Submitted in partial fulfillment of the degree of

Bachelor of Technology

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT

TABLE OF CONTENTS

Chapter No.	Topics	Page No.
	Certificate from the Supervisor	II
	Acknowledgement	III
	Summary	VI
	List of Figures	V
	List of Tables	VI
	List of Symbols and acronyms	VII
Chapter-1	INTRODUCTION	1-5
1.1	Background	1-4
1.2	Different physicochemical properties of rice	4-5
1.3	Research gap	5
1.4	Objective	5
Chapter-2	REVIEW OF LITERATURE	6-16
Chapter-3	MATERIALS AND METHODS	17-24
3.1	Primer Designing	18
3.2	DNA isolation	18-20
3.3	Polymerase chain reaction	20-21
3.4	Gene sequencing of target region	21-22
3.5	Putative SNP identification	22
3.6	Amylose estimation	23-24
Chapter-4	RESULTS AND DISCUSSION	25-34
4.1	In-Silico Analysis	25-26
4.2	Auto Dimer Results	26-27
4.3	Seed germination	27-28
4.4	DNA Isolation Results	29
4.5	Amylose Estimation Results	30-32
4.6	Java Code Devised for Nucleotide Base Splitting	33-34
Chapter-5	CONCLUSIONS	35
Chapter-7	REFERENCES	36-38

CERTIFICATE

This is to certify that project report entitled "**Association Analysis of novel single nucleotide polymorphism and its effect on structure and function of SS1 involved in starch biosynthesis in rice landraces of Himachal Pradesh**", submitted by **Dipti Sharma** in partial fulfillment for the award of degree of Bachelor of Technology in Biotechnology to Jaypee University of Information Technology, Waknaghat, Solan has been carried out under my supervision.
This work has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Signature of Supervisor: On leave
Name of Supervisor: Dr. Harvinder Singh
Designation: Associate Professor
Date: 25th May, 2015

ACKNOWLEDGMENTS

Any assignment puts to litmus test an individual's knowledge; credibility and experience and thus, sole efforts of an individual are not sufficient to accomplish the desired work. Successful completion of a project involves interest and efforts of many people and so this becomes obligatory on my part to record thanks to them. Therefore, first of all I would like to thank my guide and mentor Dr. Harvinder Singh for his guidance, help and constant encouragement throughout this project. Working under him was an enriching, learning and motivating experience.

I express my heartfelt thanks to the Head of Department, Dr. R.S Chauhan for providing me the opportunity of doing this final year project.

I would like to thank my parents, Mr. Omkar Sharma and Mrs. Kanta Sharma for always supporting me and encouraging me to do my work diligently.

Last but not the least I would like to thank Ms. Bushra Saeed for helping me in completing this report, without her help it would not have been possible.

Signature of the Student:

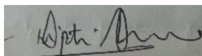
Name of Student: Dipti Sharma

Date: 25th May, 2015

SUMMARY

SSI (starch synthase I) is a major gene involved in many grain quality properties such as amylose content and gel consistency. This gene influences human health through its contribution to the glycemic index and levels of resistant starch. The incomplete digestion-absorption of resistant starch in the small intestine leads to non-digestible starch fractions with physiological functions similar to dietary fiber with significant beneficial impacts. The characteristics of starch, such as gelatinization temperature (GT), apparent amylose content (AAC), pasting temperature (PT) and other physicochemical properties, determine the quality of various products of rice such as eating, cooking and processing qualities e.g., low GT is preferred in manufacturing rice breads and beer.

In the present study, eleven varieties of *Indica* rice were grown at 28 °C. Eleven sets of primers were designed to amplify the SS1 gene sequence. Amplification and sequencing was carried out for SS1 gene amplicons. Point mutation analysis was carried out using Bioinformatics software, TASSEL 5.

Signature of Student: 
Name: Dipti Sharma
Date: 25th May, 2015

Signature of Supervisor: On leave
Name: Dr. Harvinder Singh
Date: 25th May, 2015

LIST OF FIGURES

FIGURE NO	TITLE OF FIGURE	PAGE NUMBER
Figure 1	Amylose structure	2
Figure 2	Amylopectin structure	3
Figure 3	Starch metabolism pathway	3
Figure 4	Procedure followed to work on TASSEL 5.0	13
Figure 5	Comparison of the minimum Bayes factor (BF)	16
Figure 6	Screenshot of Primer 3	18
Figure 7	Polymerase chain reaction	21
Figure 8	Screenshot of Auto Dimer (Primer Dimer Check)	26
Figure 9	Screenshot of Auto Dimer (Hair pin Check)	27
Figure 10	Seeds germination results	28
Figure 11	DNA Isolation results	29-30
Figure 12	Screenshot of TASSEL (in working mode)	34

LIST OF TABLES

TABLE NO	TITLE OF TABLE	PAGE NUMBER
Table No. 1	List of genotypes used in the study	17
Table No. 2	PCR Conditions	21
Table No. 3	List of primers with their GC content	26
Table No. 4	Quantification of Amylose	32
Table No. 5	Amylose content Quantification on 20 *Indica* Varieties	33
Table No. 6	Co-relation between amylose content and Glycemic index of rice	33

LIST OF SYMBOLS AND ACRONYMS

SSI	Starch synthase I
AAC	Apparent amylose content
GT	Gelatinization temperature
SNP	Single nucleotide polymorphism
CTAB	Hexadecyl trimethyl ammonium bromide
C:I	Chloroform isoamyl alcohol
EDTA	Ethylene diamine tetra acetic acid
Nacl	Sodium chloride
PVP	Polyvinyl pyrolidone
TE	Tris EDTA
TAE	Tris acetate EDTA
B-M	Beta-mercaptoethanol
EtBr	Ethidium bromide
dNTPs	Deoxyribonucleotide triphosphate
KI	Potassium iodide
NaOH	Sodium hydroxide
µl	Microliter
Ml	Milli litre
PCR	Polymerase chain reaction
PCA	Principal Component Analysis
GLM	General Linear Model
MLM	Mixed Linear Model
BF	Bayes Factor

SD	Standard Deviation
TASSEL	Trait Analysis by Association, Evolution and Linkage

INTRODUCTION

Rice is the seed of the monocot plants *Oryza sativa* (Asian rice) or *Oryza glaberrima* (African rice). As a cereal grain, it is the most widely consumed staple food for a large part of the world's human population, especially in Asia and the West Indies. It is the grain with the second-highest worldwide production, after maize (corn). There are many varieties of rice and culinary preferences tend to vary regionally. In the far east, there is a preference for softer and stickier varieties.

The many varieties of rice, for many purposes, are distinguished as long-, medium-, and short-grain rice. The grains of fragrant long-grain rice (high amylose) tend to remain intact after cooking; medium-grain rice (high amylopectin) becomes more sticky.

Oryza sativa (rice) contains two major subspecies: the sticky, short-grained japonica or sinica variety, and the nonsticky, long-grained indica variety. Japonica varieties are usually cultivated in dry fields, in temperate East Asia, upland areas of Southeast Asia and high elevations in South Asia. Indica variety is mainly lowland rice, grown mostly submerged, throughout tropical Asia. Rice is known to come in a variety of colors, including: white, brown, black, purple, and red.

The awareness of the general public related to health foods has been on the rise recently and people are looking for the right variety of rice. Rice is a major human food composed largely of starch (Kharabian-Masouleh et al 2012).

1.1 BACKGROUND

Starch or amylum is a carbohydrate consisting of a large number of glucose units joined by glycosidic bonds. It consists of two types of molecules: the linear and helical amylose and the branched amylopectin. Depending on the plant, starch generally contains 20 to 25% amylose and 75 to 80 % amylopectin by weight. Glycogen, the glucose store of animals, is a more branched version of amylopectin.

Rice contains two types of starch: amylose and amylopectin. Amylose is a long straight starch molecule that does not gelatinize during microwave cooking and hence rice with more amylose content tends to cook fluffy, with separate grains. Besides, amylose also hardens and forms crystals during cooking and melts when the

rice is re-heated. Rice that is high in amylose has a lower Glycemic Index number. This is because amylose is harder to break down than simple sugars like glucose etc. and ensures a sustained release of sugar into blood without spiking immediately after a meal. Amylose is important in plant energy storage. It is less readily digested than amylopectin; however, because it is more linear than amylopectin, it takes up less space. As a result, it is the preferred starch for storage in plants.

AMYLOSE

Amylose is a linear polymer made up of D-glucose units. Because of its tightly packed structure, amylose is more resistant to digestion than other starch molecules and is therefore an important form of resistant starch, which has been found to be an effective prebiotic.

Amylose is made up of α (1→4) bound glucose molecules. The carbon atoms on glucose are numbered, starting at the aldehyde (C=O) carbon, so, in amylose, the 1- carbon on one glucose molecule is linked to the 4-carbon on the next glucose molecule (α (1→4) bonds). (Jain et al, 2012).

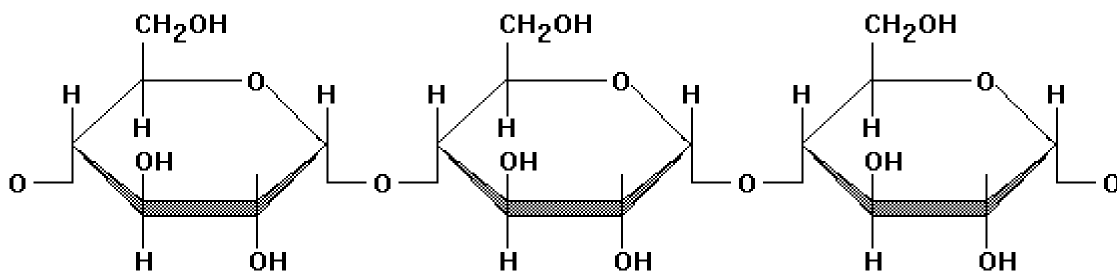


Figure 1. Amylose structure

AMYLOPECTIN

Amylopectin is a soluble polysaccharide and highly branched polymer of glucose found in plants. Glucose units are linked in a linear way with α (1→4) glycosidic bonds.

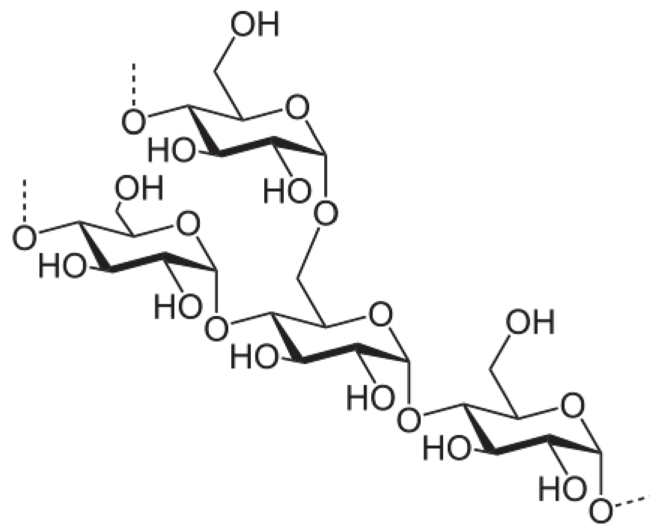


Figure 2. Amylopectin structure

STARCH METABOLISM PATHWAY

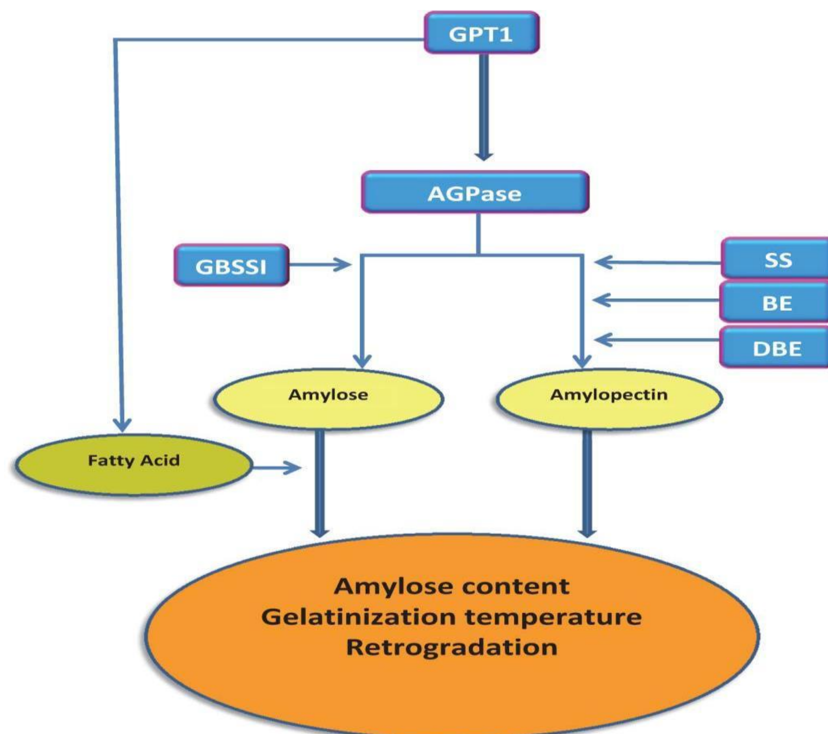


Figure 3. Starch metabolism pathway

There are 18 genes involved in the Starch biosynthetic pathway which are responsible for the control of physicochemical properties, out of which I am working on SSI (starch synthase I) gene. This gene appear to be the major loci controlling traits under human selection in rice, SSI influences amylose content and retro gradation, it explains much of the variation in cooking characteristic.

Amylose content showed significant positive association with elongation ratio and volume expansion. It correlates positively with volume expansion of cooked rice. Rice having high amylose content shows high volume expansion. (Lang and Buu, 2004).

SSI gene, this gene mainly associates with apparent amylose content (AAC), gel consistency and grain elongation.

1.2 DIFFERENT PHYSICOCHEMICAL PROPERTIES OF RICE

1. Pasting temperature: The temperature at which irreversible swelling of the starch granules occur.

2. Chalkiness: is determined by the clarity of the endosperm. If part of the milled rice kernel is not transparent (or opaque), it is often characterized as chalky. While chalkiness disappears upon cooking and has no effect on taste or aroma, it downgrades the quality of milled rice.

3. Gelatinization temperature: Gelatinization is the irreversible melting process of starch. It is an endothermic process, and the peak temperature at which starch absorbs heat is called gelatinization temperature (GT).There are several ways to measure GT. The most common method is alkali spreading value (ASV), for which the degree of disintegration of milled rice is graded after immersion in potassium hydroxide for 24 hours. A more modern assay is by differential scanning calorimetry (DSC). This detects the peak temperature absorbed by the flour-water mixture during gelatinization and is thus more precise than ASV.

4. Peak viscosity: The peak viscosity is defined as the maximum viscosity that occurs prior to the initiation of sample cooling. The highest viscosity reached during

gelatinization of starch usually corresponding to the point where all the granules are swollen to occupy a high proportion of the available volume with each in contact with its immediate neighbors. Note that on cooling the starch paste, the viscosity may rise above this level, but only the initial shoulder or true peak values determined on the hot paste are termed peak viscosity (Belitz, et al 2004).

5. Retrogradation: Retrogradation is a reaction that takes place in gelatinized starch when the amylose and amylopectin chains realign themselves, causing the liquid to gel.

In this study we investigated the role of one starch-related gene i.e. SSI gene and its SNPs by assessing its contribution to variation in starch properties in a rice breeding population.

1.3 RESEARCH GAP

With the advancement in time people are getting more aware about healthy food. However, in *Indica* rice consuming countries like India long grain rice with intermediate amylose and intermediate gelatinization temperature is preferred since it is soft, fluffy and non-sticky after cooking. Improvement in rice starch quality, especially increasing the amylose level in high yielding rice with Amylose content (AC), is an important determinant. High degree of Variability for Amylose and Amylopectin levels have been reported in Indica subspecies in comparison to Japonica variety but no significant research has been carried out in Indica Subspecies with respect to amylose and amylopectin content. This will be the first study where association analysis will be carried out on Indica rice genotypes.

1.4 OBJECTIVES

a. Detection of novel putative Single Nucleotide Polymorphism in sequences SS1.
b. Association analysis with amylose content.
c. Modelling of their protein structures and deducing the effect of putative SNP in starch biosynthesis.

REVIEW OF LITERATURE

Oryza sativa, commonly known as Asian rice, is the plant species most commonly referred in English as rice. Rice is an important crop in the world. *Oryza sativa* is the cereal with the smallest genome, consisting of just 430Mb across 12 chromosomes. It is renowned for being easy to genetically transform, and is a model organism for cereal biology. (PhilRice and IIRR. 2000).

Oryza sativa was the cereal selected to be sequenced as a priority and has gained the status "model organism". It has the smallest genome of all the cereals: 430 million nucleotides and it can serve as a model genome for one of the two main groups of flowering plants, the monocotyledons. Because it has been the subject of studies on yield, hybrid vigor, genetic resistance to disease and adaptive responses, scientists have taken advantage of the existence of a multitude of varieties that have adapted to a very wide range of environmental conditions.

There are about 120,000 varieties known to exist. Two of the types sequenced are Indica and japonica. Rice is well known for its divergence between the Indica and japonica subspecies.

1. Japonica varieties are usually cultivated in dry fields, in temperate East Asia, upland areas of Southeast Asia and high elevations in South Asia.
2. Indica varieties are mainly low land rice, grown mostly submerged, throughout tropical Asia. Rice is known to come in a variety of colors, including: white, brown, black, purple, and red.

Rice is the most important staple food for a large part of the world's human population, especially in East Asia, Southeast Asia, South Asia, the Middle East, and the West Indies. The awareness of the general public related to health foods has been on the rise recently and people are looking for the right variety of rice and more efficient methods of cooking it for diabetics. Due to the low insulin concentration in the body, the release of sugar into blood is less controllable than normal, leading to spikes in blood sugar after meals for them. High blood sugar, if left untreated, can cause dehydration, electrolyte imbalance etc. over short term and retinopathy, nephropathy over long term. Hence, this is an acute problem faced by all diabetics.

The deep genetic structure in rice may be influenced by the natural history of ancestral populations prior to domestication, as well as by the autogamous breeding

system and complexity of the breeding practices exercised by humans (Garris et al. 2005). Even in well-designed studies, modest amounts of population stratification can still exist (Freedman et al. 2004), indicating the difficulties in detecting and controlling population stratification as a source of false positive associations. To elucidate the mechanism of starch biosynthesis in amyloplast and chloroplast, plant starch-synthesizing enzymes have been identified, characterized, and cloned (reviewed by Preiss and Sivak 1995).

Starch properties determine the key functional properties of rice such as cooking temperature and influence human health through its contribution to the glycemic index and levels of resistant starch. The incomplete digestion-absorption of resistant starch in the small intestine leads to non-digestible starch fractions with physiological functions similar to dietary fiber with significant beneficial impacts.

Starch molecules accumulate to form starch grains, which are visible in many plant cells, notably in storage organs such as the potato tuber, and in seeds of cereals and legumes. In rice amylose is found inside the microscopic granules, which make up a single rice grain. Amylopectin forms the crystalline or ice-like "skeleton" of these granules. Scientists think the amylose forms into long chain-like arrangements to fill up the spaces inside of the granule. The protein in rice grains is found in pockets between granules of starch. The cooking and eating characteristics of rice are influenced by the amount of amylose found in the grains. This is because the starch granules in the grain expand during cooking, forcing out the chains of amylose in a process scientists call leaching. As the cooked rice cools, the leached amylose chains line up, lock together and form a gel. When rice cools to room temperature or beyond, the chains of amylose crystallise. Generally, the higher the amylose content of rice, the firmer the cooked grain of rice will be. Some types of rice are between 25% and

30% amylose. These high amylose levels tend to make the rice cook firm and dry. Rice with a medium amylose content of between 16% and 22% usually cooks softer and the grains stick together more readily.

EFFECT OF COOKING ON AMYLOSE CONTENT OF RICE

In diabetes type 2, there is a deficiency of insulin which results in improper/ slow breakdown of food. This results in sugar level spikes immediately after a meal, which can be harmful. Hence, diabetics must have food stuff which have a slow release rate

so as to not cause any spikes. In order to find out the most suitable method of cooking rice for diabetic patients, we found out the amylose content of rice cooked by different methods and co-related it with Glycemic Index (GI). Glycemic Index is release of glucose in the blood by the breakdown of carbohydrates. Higher the glycemic index, faster breakdown of food and thereby more release of glucose in the blood, so immediate requirement of insulin, which can be toxic for diabetic patients. Glycemic index and amylose content are inversely proportional to each other. There is a wide variation in the amylose content of rice depending on the way it is cooked. The effect of cooking on amylose content of rice is described using various experimental approaches.
Various method of cooking involves traditional method, microwave and steam cooked method.
The amylose content of the rice is then co-related with its glycemic index.

Retrogradation describes the hardening of cooked starch after cooling due to re- crystallization of gelatinized starch components during storage. It is believed there is a significant correlation between the tendency of any one starch sample to retrograde and its levels of resistant starch. High-amylose rice cultivars usually have more resistant starch (RS) and lower estimated glycemic index (EGS), suggesting highly-retrograded cooked rice cultivars tend to a reduction of hydrolysis index (HI) and glycemic index (GI). Conversely, starch of low-amylose rice which have higher HI are more quickly hydrolyzed than intermediate and high-amylose rice (high HI). Characteristics of high amylose rice cultivars are normally determined by RVA (Rapid Visco Analysis) which are described by parameters such as peak viscosity (PKV), hot paste viscosity (HPV) and cool paste viscosity (CPV).
There are seven starch synthesis enzyme classes including ADP-glucose pyrophosphorylase (AGPase), granule bound starch synthase (GBSS), starch synthase (SS), branching enzyme (BE), debranching enzyme (DBE), starch phosphorylase (PHO) and glucose 6-phosphate translocator (GPT). These genes/enzymes contribute directly or indirectly to the production of starch granules. GBSSI and SSIIa genes are most correlated with starch properties, whereas SSIIb and SPHOL genes do not explain variation in starch properties while other genes have low to medium effects on rice starch quality. Seven genes (AGPS2b,

SPHOL, SSIIb, SSIVb and ISA1,BEIIa and BEIIb) of 18 do not contribute to starch physiochemical properties.

Gene

ADP glucose pyrophosphorylasre (AGPase)

Granule bound starch synthases (GBSS I and GBSS II)

Starch synthases (SSI, SSIIa, SSIIb, SSIIIa, SSIIIb, SSIVa, SSIVb)

Branching enzymes (BEI, BEIIa, BEIIb)

Debarnching enzymes (ISA1, ISA2, Pullulanase)

Starch phosphorylase (SPHOL)

Glucose 6-phosphate translocator (GPT1)

Starch synthase, which catalyzes the elongation of α-1,4-glucosidic bonds on amylose and amylopectin molecules by transfer of ADP-Glc, is present in two forms: SSS and GBSS enzymes (reviewed by Preiss, 1991). Two genes appear to be the major loci controlling traits under human selection in rice, GBSSI (waxy gene) and SSIIa. GBSSI influenced amylose content and retrogradation. Other genes contributing to retrogradation were GPT1, SSI, BEI and SSIIIa. SSIIa explained much of the variation in cooking characteristics. Other genes had relatively small effects. The genes encoding SSl and GBSS are closely located on rice chromosome 6.

It is generally thought that GBSS is responsible for the synthesis of amylose in starch granules (Sprague et al., 1943; Nelson and Rines, 1962; Echt and Schwarz, 1981; Shure et al., 1983). The role of SSS in starch synthesis still remains unclear, although this enzyme appears to play an important role(s) in the amylopectin synthesis by the cooperative reaction with branching enzyme and/or to participate in the amylose synthesis together with GBSS (Hagenblad et al. 2004).

Three apparent isoforms of SSS with molecular sizes of 55 and 57 kD in developing rice (*Oryza sativa* L.) seeds were identified (Baba et al., 1993). Analysis of the amino terminal amino acid sequence reveals that the three isoforms of SSS are identical except that the 55-kD isoform lacks the first eight amino acids at the amino terminus. Thus, it appears that these three isoforms are products of the same gene, although additional evidence is necessary. The cDNA cloning of the rice SSS form

(Baba et al., 1993), termed SSS1, demonstrates that this enzyme shares only a limited degree of sequence identity with rice GBSS (Wang et al., 1990) and *E.coli* glycogen synthase (Kumar et al., 1986). The gene expression pattern is also different between SSSl and GBSS; the SSSl gene is expressed at the same level in both leaves and developing seeds (Baba et al., 1993), whereas the transcript of the GBSS gene is present only in the developing seeds (Hirano and Sano, 1991). These facts may imply the distinct role(s) of SSSl and GBSS in starch synthesis. At any rate, because of the extreme differences of the exon/intron organizations between two genes encoding rice SSSl and GBSS and the significant differences between their primary structures
(Baba et al., 1993), it is possible that the two genes have evolved from an ancestral gene in a different way or that these genes are products of different ancestral genes that have converged during evolution (Tanaka et al. Plant Physiol. Vol. 108, 1995).

The gene encoding starch synthase IIa (SSIIa), alk, is exclusively expressed in the rice endosperm and has been extensively studied in the context of its effect on cooking quality and starch texture. Two SNPs within exon 8, [A/G] and [GC/TT] are significantly associated with rice alkali disintegration and eating quality and starch gelatinization temperature.

Single nucleotide polymorphisms (SNP) are the most abundant type of genetic variation found within all species and many important plant traits and human diseases are attributed to these sequence variations. Identifying SNP and associating them with grain starch quality advances our understanding of the starch biosynthesis pathway and highlights ways to improve crops that are higher yielding and of better quality, directly impacting food security and human nutrition and health. Massively parallel sequencing (MPS) technology is a high throughput platform for genetic analysis based on ultra deep DNA sequencing. Kharabian-Masouleh et al. (2011) discovered more than 501 SNPs and 113 In/dels in 17 starch synthesis genes in an Australian rice breeding population using a combination of a target pooled long range PCR and MPS. By combining MPS with high throughput genotyping technologies such as multiplexed-MALDI-TOF (Sequenom), rapid polymorphism discovery followed by association analysis is now possible.

Only one ´T/C´ nsSNP at position 5153 of this gene showed minor associations with FV, SB and Martin test (MT), with R2 values of 0.16, 0.11, 0.16, respectively SSIIa. Highly significant associations were found between SNP of SSIIa

and PT, peak time (PKT), GT and breakdown viscosity. The highest F-test value of 199.65 was observed for the [GC at position 4827–4828 of SSIIa and PT. This SNP is associated with PT, PKT and BDV with R2 values of 0.642, 0.323 and 0.168, respectively. This SNP has one of the strongest associations among the physiochemical properties studied in this rice population. GBSSI and SSIIa are major genes involved in many grain quality properties such as amylose content and gelatinization temperature SSIIa had a high association with pasting temperature, gelatinization temperature and peak time.Gelatinization is the irreversible melting process of starch. It is an endothermic process, and the peak temperature at which starch absorbs heat is called gelatinization temperature (GT). GT is one of the key traits measured in breeding programs in rice. It is important because it affects the texture of cooked rice and it is said to be related to the cooking time of rice. Given the number of times that rice is cooked in the world every day, selecting rice varieties with low GT represents huge potential savings in fuel consumption. The effect of this gene on cooking quality and starch texture has been extensively studied by many authors. Highly significant associations were found between SSIIa SNPs and important physiochemical properties such as PT, PKT, GT and BDV. Melting of starch crystalline regions is measured by pasting and gelatinization temperature and peak time (Bao, et al.2006).

The characteristics of starch, such as gelatinization temperature (GT), apparent amylose content (AAC), pasting temperature (PT), Gel Consistency (GC) and other physicochemical properties, determine the quality of various products of rice, e.g., eating, cooking and processing qualities. The GT of rice flour is controlled by the alk locus, which has been co-mapped to the starch synthase IIa (SSIIa) locus. Differences in starch properties of rice grains determine their eating, cooking and processing qualities.

Apparent amylose content (AAC) and gelatinization temperature (GT) are two of the well-established parameters used to evaluate these qualities (Juliano 1998; Bergman et al. 2004). For example, low-GT rice have a softer texture than high-GT rice among freshly cooked waxy and low-AAC rice; and among intermediate-and high-AAC rice, intermediate-GT rice are softer than low-GT rice when freshly cooked, but they have similar hardness values on accelerated staling (Perez et al. 1993). Different rice products require different types of rice with different starch

properties, e.g., low GT is preferred in manufacturing rice breads and beer (Juliano 1998). Knowledge of the relationship between physicochemical properties and end- use qualities can direct breeding activities in selecting desired rice with unique qualities. However, lack of information on the genetic basis and related techniques for selection retards the breeding progress. The GT of starch can be measured indirectly by alkali spreading value or directly by differential scanning calorimetry (DSC) as peak temperature (Tp) from the endotherm. The pasting temperature (PT) measured by visco amylography such as using Rapid Visco (McKenzie and Rutger 1983).

Umemoto et al. (2002) reported that the starch synthase IIa (SSIIa) gene is located at the alk locus on chromosome 6 in the rice genome. Association analyses have also been conducted for the SSIIa gene in relation to GT, but only two or three SNPs were used in such studies (Chen et al. 2003; Umemoto et al. 2004), resulting in the same haplotypes consisting of rices with both high and low GT. Furthermore, previous studies began with alignment of SSIIa sequences to find SNPs among a few rice (Chen et al. 2003; Umemoto et al.2004) e.g., the two genome-sequenced rice cultivars, 93-11 and Nipponbare (Chen et al. 2003). The few sequences studied unlikely represent all naturally occurring variations in this gene. Therefore, it is necessary to search for more SNPs from a different set of germplasm in order to find those that can differentiate high- and low-GT rice varieties. Furthermore, nucleotide diversity in SSIIa has not been investigated. The SSIIa plays an important role in the elongation of short chains of DP < 10 that leads to the formation of intermediate chains of amylopectin, which is especially responsible for gelatinization of starch (Umemoto et al. 2002; Gao et al. 2003; Jiang et al. 2004; Nakamura 2002). The present association analysis also indicated that the SSIIa GC/TT explained more than

60% of the total variation of the phenotypes. The present association study indicates that both Wx and SSIIa SNP loci are important in controlling these traits, each explaining > 50% of the total variation for both traits (Theor Appl Genet .2006).

FLOW CHART FOR RUNNING TASSEL v5.0

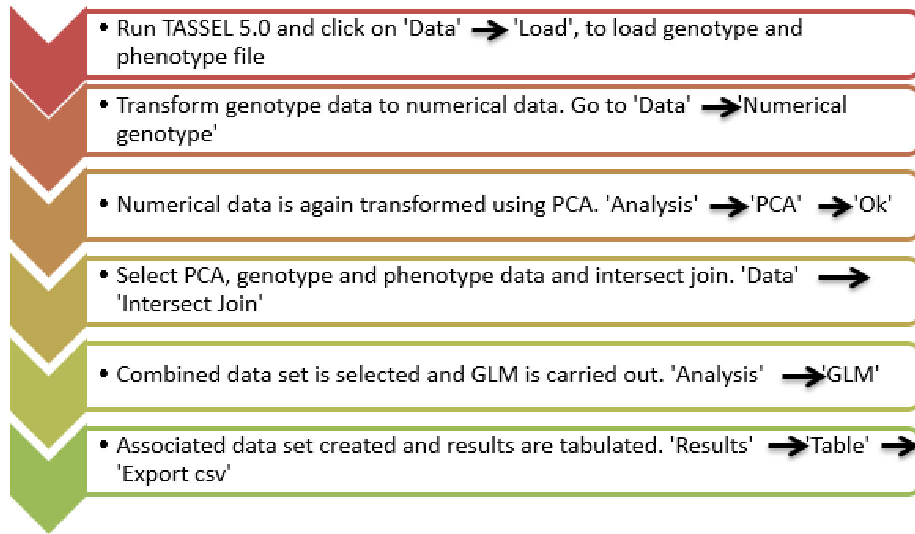


Figure 4. Procedure followed to work on TASSEL 5.0

WHY PRINCIPAL COMPONENT ANALYSIS (PCA) IS DONE?

In this particular context, PCA is mainly used to account for population-specific variations in alleles distribution on the SNPs under investigation. Such "population substructure" mainly arises as a consequence of varying frequencies of minor alleles in genetically distant ancestries (e.g. Japonica and Indica). The construction of principal axes follows from the classical approach to PCA, which is applied to the scaled matrix (individuals by SNPs) of observed genotypes (AA, AB, BB; say B is the minor allele in all cases), to the exception that an additional normalization to account for population drift might be applied. It all assumes that the frequency of the minor allele (taking value in {0,1,2}) can be considered as numeric, that is we work under an additive model (also called allelic dosage) or any equivalent one that would make sense. As the successive orthogonal PCs will account for the maximum variance, this provides a way to highlight groups of individuals differing at the level of minor allele frequency. Apart from clustering subpopulations, this approach can also be used for detecting outliers which might arise in two cases (AFAIK): (a) genotyping errors, and (b) when working with a homogeneous population (or assumed so, given self-reported ethnicity), individuals exhibiting unexpected genotype. What is usually done in this case is to apply PCA in an iterative manner, and remove individuals whose scores are below ±6 SD on at least one of the first 20 principal axes; this amount to "whiten" the sample, in some sense. Note that any such measure of genotype distance (this also holds when using Multidimensional Scaling in place of PCA) will allow to spot relatives or siblings.

WHY INTERSECT JOIN IS DONE?

This option joins multiple data sets by the intersection of their taxa. The common data sets between the phenotype and genotype data set are to be collaborated hence intersect join option is used rather than using union join. Taxa must be present in both data sets to be included.

MLM (Mixed Linear Model)

A mixed model is one which includes both fixed and random effects. Including random effects gives MLM the ability to incorporate information about relationships among individuals. When a genetic marker based kinship matrix (K) is used jointly with population structure (Q), the "Q+K" approach improves statistical power compared to "Q" only. MLM can be described in Henderson's matrix notation as
follows:

$$y = X\beta + Zu + e$$

where y is the vector of observations; β is an unknown vector containing fixed effects, including genetic marker and population structure (Q); u is an unknown vector of random additive genetic effects from multiple background QTL for individuals/lines; X and Z are the known design matrices; and e is the unobserved vector of random residual.

TASSEL implements several methods to improve statistical power and reduce computing time. The Restricted Maximum Likelihood (REML) estimates of and are obtained through the Efficient Mixed-Model Association (EMMA) algorithm which is much faster than the expectation and maximization (EM) algorithm. TASSEL also implements a method called compression which reduces the dimensionality of the kinship matrix to reduce computational time and improve model fitting.

GLM (General Linear Model)

This function performs association analysis using a least squares fixed effects linear model. TASSEL utilizes a fixed effects linear model to test for association between segregating sites and phenotypes. The analysis optionally accounts for population structure using covariates that indicate degree of membership in underlying populations. A main effects only model is automatically built using all variables in the input data. A separate model is built and solved for each trait and marker combination. Any factors, covariates, reps or locations are included in every model as main effects. How the data is used must be defined either in the input data files or using the Trait Filter after the data has been imported but before it has been joined with a genotype. General Linear Model (GLM) can be run using a numeric data set only or using numeric data joined to genotype data. If only numeric data is selected, best linear unbiased estimates (BLUEs or least square means) will be generated for the taxa for each trait. [Note: only factors and covariates intended to control field variation should be included at this stage. Population structure covariates which are intended to control for marker effects should only be included when markers are also in the analysis.] If numeric data with genotypes are analyzed, each trait by marker combination will be tested and two reports will be produced, one containing trait by marker F-tests and the other containing allele estimates.

GLM vs MLM

According to research paper "Association Mapping for Important Agronomic Traits in Core Collection of Rice (*Oryza sativa* L.) with SSR Markers", association analysis was performed using the software TASSEL (www.maizegenetics.net/tassel). For the mixed linear model (MLM) method, both K and Q matrices were incorporated, whereas for the GLM method, only population structure information (Q-matrix) was used as a covariate. Significance of associations between loci and traits were determined by their P values (P,0.05). Since MLM method performs better in controlling spurious associations than GLM method, we first ranked the significant (P,0.05) association from MLM and then compared the significance of these markers (P,0.05) in the permutation based on GLM association tests. For the comparison, we calculated and used other two significant thresholds (i.e. Minimum Bayes factor (BF) and Bonferroni threshold) besides the P value. BF was calculated using the following

formula: BF =2e*P*ln(P). Duncan multiple comparisons was implemented in SPSS for comparisons of performance of agronomic traits relevant to different alleles of the significant trait-marker associations.

BF value

P values that are not miniscule are difficult to interpret. When the p value is moderate, supplementing the p value with a Bayes factor is especially important for demonstrating the level of evidence against realistic alternative hypotheses.

p value	*Z*	Simple BF	False discovery rate BF	Normal-prior BF	Symmetric-decreasing-prior BF
0.1	1.64	0.26	0.1	0.7	0.63
0.05	1.96	0.15	0.05	0.47	0.41
0.01	2.17	0.04	0.01	0.15	0.13
0.001	2.58	0.004	0.001	0.02	0.02
0.0001	3.28	0.0005	0.0001	0.003	0.003

* For the simple BF and the false discovery rate BF, the minimum occurs when the observed effect size is at the mode of the alternative distribution. The minimum simple BF for each *p* value and *Z* is $\exp(-Z^2/2)/\exp(0)$, since the critical point under the null hypothesis is $\hat{\theta}^2/(\sigma^2/m) = Z^2$ and the critical point under the alternative distribution is $(\hat{\theta} - \theta)^2/(\sigma^2/m) = 0$. The minimum false discovery rate BF for each *p* value is $p/1 = p$, because the maximum "power" for the alternative hypothesis is 1. The minimum normal-prior BF is found by minimizing Ioannidis's equation 4 (7) and is $Z\exp(0.5 - Z^2/2)$ for $Z > 1$. The minimum symmetric-decreasing-prior BF is the minimum BF over all priors symmetric and decreasing around the mean of the null, and is equal to $-ep\ln(p)$ (20).

Figure 5. Comparison of the minimum Bayes factor (BF) possible for each p value, minimized over the possible alternative hypotheses allowed for each BF*

BONFERRONI THRESHOLD

To calculate the Bonferroni corrected p-value you multiply your p-values by the number of SNPs, any corrected p-values under 0.05 are significant. Bonferroni is considered to be very conservative so as an alternative Benjamini-Hochberg method can also be used. In statistics, the Bonferroni correction is a method used to counteract the problem of multiple comparisons. It is considered the simplest and most conservative method to control the familywise error rate

MATERIALS AND METHODS

Strategy flowchart:

1. Primer Designing
2. Checking the stability of primers via Auto Dimer
3. Isolation of DNA
4. Amplification of target gene
5. Sequencing of the target region
6. Association Analysis using TASSEL 5.0
7. Protein structure modelling using PYMOL.

Seeds of twenty varieties of rice were procured (Table 1). For good germination dormancy breakdown treatment was given to the genotypes by soaking them in 500 ppm gibberellic acid overnight at temperature 37 °C.

S/No	Genotypes
1	S5
2	S2
3	S3
4	S4
5	7
6	1
7	2
8	3
9	4
10	5
11	11
12	12
13	13
14	14
15	15
16	6
17	18

Table No. 1 List of genotypes used in the study.

3.1. PRIMER DESIGNING:

A primer is a strand of nucleic acid that serves as a starting point for DNA synthesis. They are required for DNA replication because the enzymes that catalyze this process, DNA polymerases, can only add new nucleotides to an existing strand of DNA. The polymerase starts replication at the 3'-end of the primer, and copies the opposite strand.

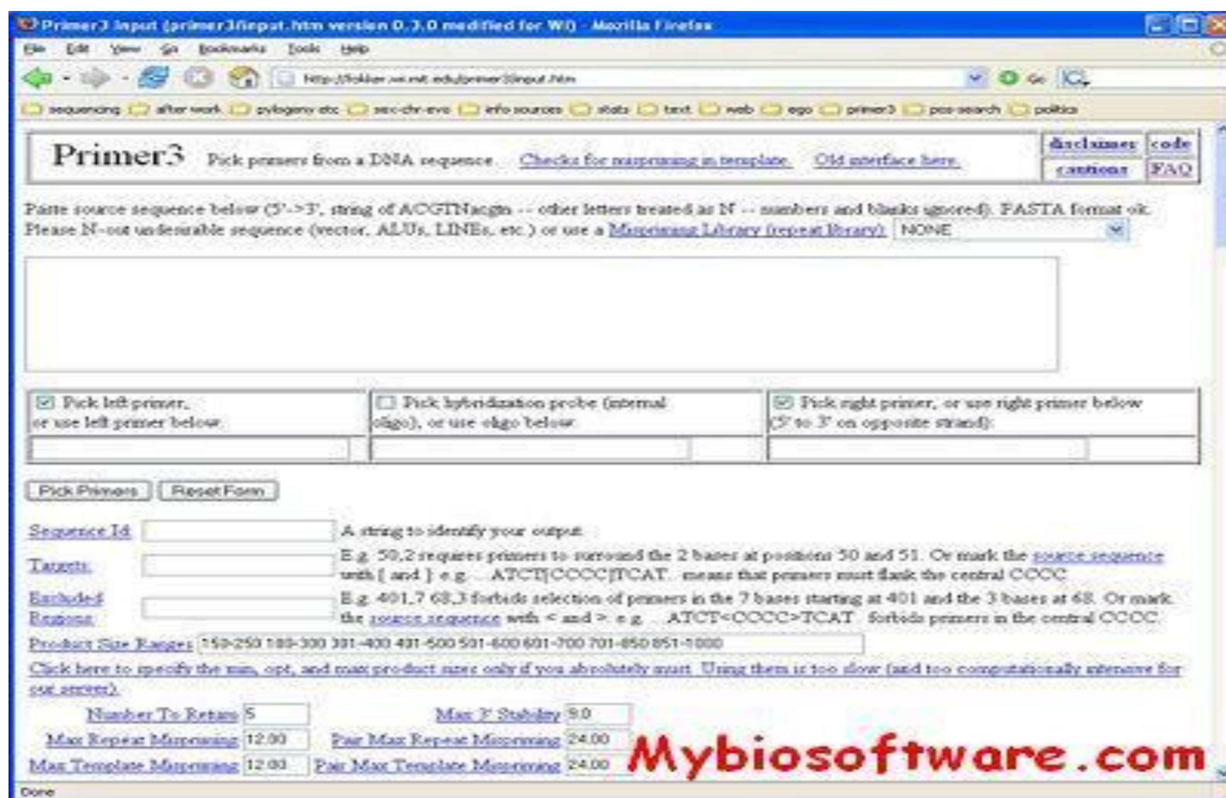


Figure 6. Screenshot of primer 3: a primer designing tool

3.2 DNA ISOLATION CHEMICALS PREPARATION:

1 M Tris (pH 8.0)

Dissolve 12.11 g of Tris base in 80 ml of H2O. Adjust pH to 8.0 by adding concentrated HCL. Allow the solution to cool to room temperature before making the final adjustments to the pH. Adjust the volume to 100 ml with H2O. Sterilize using an autoclave.

0.5 M EDTA (pH 8.0)

Dissolve 18.612 g of EDTA in 80 ml of dH2O. Adjust pH to 8.0 by adding concentrated HCL. Allow the solution to cool to room temperature before making the final adjustments to the pH. Adjust the volume to 100 ml with H2O. Sterilize using an autoclave.

NaCl

Dissolve 29.22 g of NaCl in 100 ml of H2O. Sterilize using an autoclave.

TE buffer

Dissolve 2 ml of EDTA in 10 ml of Tris buffer. Sterilize using an autoclave.

50x TAE buffer (pH 8.5)

242 g Tris Cl

57.1 g acetic acid

100 ml of 0.5M EDTA (pH 8.0) Make up to 1L with water.

To make a 1x working solution, dissolve 2ml of 50x TAE buffer in 98 ml of dH2O.

0.8% Agarose gel

0.8g Agarose dissolved in 100 ml TAE

Protocol (CTAB method)

1. Grind 200 mg of plant tissue to a fine powder in liquid nitrogen.
2. Take the sample in microfuge tubes (2ml).
3. Add 700µl prewarmed cTAB buffer/extraction buffer in microfuge tubes.
4. Incubate the cTAB/plant extract mixture for about 1-2 hr at 60°C in a recirculating water bath.
5. After incubation add 700µl of Chloroform: Isoamyl Alcohol (24:1) and mix the solution by inversion.
6. Spin/centrifuge the CTAB/plant extract mixture at 10000 rpm for 10-15 minutes

7. Transfer the supernatant to clean microfuge tubes (1.5ml).
8. Add 500 µl of chilled isopropanol (-20°C). Mix well.
9. Keep it at -20°C for 1 hr.
10. Centrifuge @10000 rpm for 10 minutes.
11. Discard the supernatant.
12. Wash the pellet with 200µl of 70% ethanol (chilled).
13. Centrifuge at 10000 for 10 minutes.
14. Discard the supernatant and repeat 70% ethanol-washing step for 1-2 minutes.
15. Dry the pallet. There should not be any smell of ethanol left.
16. Dissolve the pellet in 50µl of TE buffer.
17. Keep it at 4°C overnight.

DNA quantification: Agarose gel electrophoresis

1. Prepare 0.8% solution of agarose by melting 0.8 g of agarose in 100ml of 1X TAE buffer/ running buffer in a microwave for approximately 2 min.
2. Allow to cool for a couple of minutes then add 5µl ethidium bromide, stir to mix.
3. Cast a gel using a supplied tray and comb. Allow the gel to set for a minimum of 20 min at room temperature on a flat surface.
4. After proper solidification of gel remove the comb and keep gel in gel running tank. Pour running buffer gently.
5. Load samples with DNA loading dye in separate wells.
6. Run the gel for 45-1hr at 100 V.
7. Expose the gel to UV light and photograph.
8. Confirm DNA quality, presence of a highly resolved high molecular weight band indicates good quality DNA, presence of a smeared band indicates DNA degradation.

3.3 PCR AMPLIFICATION

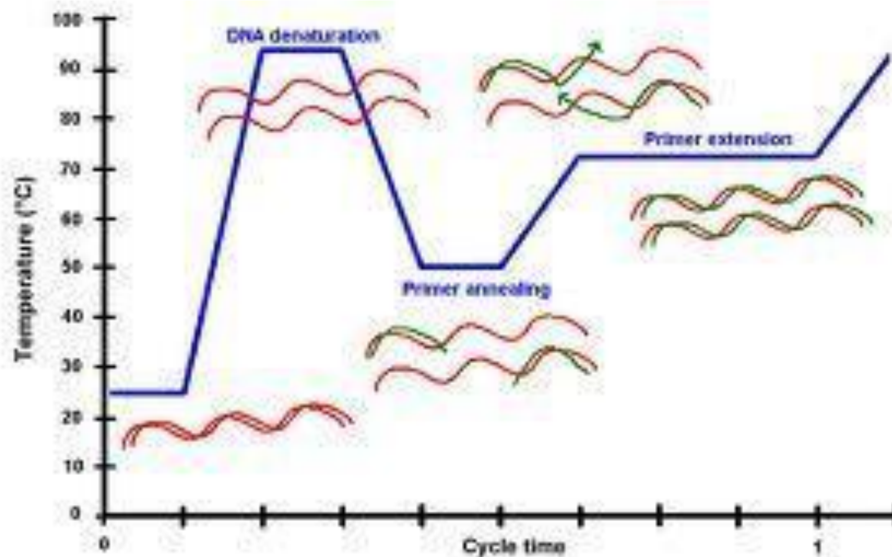


Figure 7. Polymerase chain reaction

Stage	Process	Temperature (°C)	Time(min.)
Stage 1	Initial denaturation	94	4
Stage 2	Denaturation Annealing extension	94 Ta 72	1 1 2
Stage 3	Final extension	72	7

Table No. 2. PCR conditions

Number of Cycles - 35 cycles

3.4 GENE SEQUENCING OF TARGET REGION

DNA sequencing is the process of determining the precise order of nucleotides within a DNA molecule. It includes any method or technology that is used to determine the order of the four bases adenine, guanine, cytosine, and thymine in a strand of DNA.

Sanger sequencing is a method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in

vitro DNA replication. The DNA sample is divided into four separate sequencing reactions, containing all four of the standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and the DNA polymerase. To each reaction is added only one of the four dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP), while three other nucleotides are ordinary ones. Putting it in a more sensible order, four separate reactions are needed in this process to test all four ddNTPs. Following rounds of template DNA extension from the bound primer, the resulting DNA fragments are heat denatured and separated by size using gel electrophoresis. The DNA bands may then be visualized byautoradiography or UV light and the DNA sequence can be directly read off the X- ray film or gel image.

3.5 PUTATIVE SNP IDENTIFICATION

Single nucleotide polymorphisms (SNPs) are single base pair positions in genomic DNA at which different sequence alternatives exist in normal individuals in some populations. In simple words SNP is the polymorphism occurring between DNA samples with respect to single base. SNPs have the potential to provide basis of superior and highly informative genotyping assay. SNPs in coding region have functional significance if the resulting amino acid change causes the altered phenotype. SNP markers associated with phenotypic changes pinpoint functional polymorphism.

At a particular site in a DNA molecule theoretically four possible nucleotides are involved but in reality only two of these four possibilities have been observed at the specific sites in population. Thus SNPs are largly biallelic in nature. SNPs are less mutable as compared to other markers. The low rates of recurrent mutation make them evolutionary stable. They are excellent markers for studying complex genetic traits and for understanding the genomic evolution.

Direct sequencing: Sequence analysis is the most direct way of identifying SNPs. DNA could be sequenced using various methods available like Sanger sequencing and other nucleotide sequencing method.

3.6 AMYLOSE ESTIMATION (Potassium Iodide Method):

The following reagents were used for estimation of amylose content and for preparation of standard:

95% Ethanol- Prepared from 100 % Ethanol
1N NaOH, Iodine - Potassium iodide solution
Standard amylose - Obtained from HIMEDIA
1N Acetic acid - From NICE Chemicals
Glassware - Borosil
Water bath - High Precision water bath from Acmas was used Spectrophotometer- From Amersham Biosciences to measure OD Cuvette - Quartz Cuvette to measure OD Software- HandyGraph Software to draw graph
1N NaOH solution - Dissolve 40g of NaOH in 1000ml distilled water
1N Acetic acid solution - Dilute 57.5 ml glacial acetic acid to 1000ml using distilled water.
Potassium iodide solution - Dissolve 0.26 g of Iodine in 10 ml of Potassium iodide solution containing 2.6 g of KI.
Standard Amylose Solution- Take 40mg of pure potato starch (amylose) in a 100 ml volumetric flask and add 1 ml of 95% ethanol and 9.0 ml of 1N NaOH. Shake well and boil over water bath for 10 minutes and make up the solution to 100 ml using distilled water.

Protocol for amylose estimation

1. Weigh 100 mg well powdered milled rice into 100 ml volumetric flask.
2. Add 1 ml 95% ethanol and 9 ml 1 N NaOH.
3. Heat the sample for 10 minutes in boiling water bath, cool it and make up the volume to 100 ml.
4. Pipette 5 ml from the 100 ml into another 100 ml volumetric flask.
5. Add 1 ml I N acetic acid and then 2 ml iodide solution and make up the volume to 100 ml.
6. Shake, stand for 20 minutes and determine the per cent Transmittance at 620 nm using a colorimeter.
7. Prepare a series of standard starch solution containing 0, 20, 40, 60, 80 and 100% amylose as in the steps 1 to 5

8. Read the transmittance of the standards at 620nm and plot a standard graph.

9. Amylose content of the sample is determined in reference to the standard curve and expressed on percent basis.

10. Making of amylose standards:

a. Pipette out 1, 2, 3, 4 and 5 ml of the standard amylose into 100 ml volumetric flasks in three replications.

b. Keep one flask as blank without adding anything.

c. Add 1.0 ml 1N acetic acid and 2.0 ml I-KI solution to all the flasks including blank. d. Make up all the flasks to 100ml using distilled water and cover all the flasks with a black cloth or aluminum foil to prevent direct light exposure. I-KI disintegrates on exposure to light.

e. Keep for 20 minutes and take reading at 620nm in a spectrophotometer.

f. The standards including blank, correspond to 0%, 4%, 8%, 12%, 16% and 20% of amylose.

g. Draw a standard curve using the absorbance reading.

RESULTS AND DISCUSSION

4.1 IN SILICO ANALYSIS:

SSI GENE was downloaded from gramene database with following detail: Gramene result:
Identification no. LOC_os06g06560
Sequence- SSI GENE
Location-MSU6:6:3077460:308640
Chromosome: 6

PRIMER LIST	SEQUENCES OF PRIMERS	GC%	TM	LENGTH
SS-I_1F	GTGAAGCAACGGAGAGAAGG	45.00	59.69	20
SS-I_1R	TCACCATTCCTCTTCCCACT	50.00	59.72	20
SS-I_2F	AGGAATGGTGAGGTTTGGTG	55.00	59.97	20
SS-I_2R	CCATCGACCACAATTCAGTG	50.00	60.10	20
SS-I_3F	TATTTTTCTTGGGCCTCTCG	40.00	57.15	20
SS-I_3R	GGAATCAATCTTGCAACCAAA	50.00	59.71	21
SS-I_4F	GCCAGTCTTGTGCCAGTGTA	36.36	59.76	20
SS-I_4R	AGTTCTCCTGCGCTTTTGAG	50.00	60.43	20
SS-I_5F	GCCAGTCTTGTGCCAGTGTA	45.00	59.96	20
SS-I_5R	GGCATCCCTGTAAACACCAT	50.00	58.88	20
SS-I_6F	CCTTTGTTTTGGCATGTTCC	55.00	59.81	20
SS-I_6R	AGGCTCAAACCTGATGTGCT	39.13	59.35	23

SS-I_7F	GCTCCCGGAAGAGTGTATTG	50.00	59.76	20
SS-I_7R	AAGCAGGTGCTCCAATTGAT	42.86	58.23	20
SS-I_8F	GGCAGCAAATGTCTTGATGA	42.86	58.61	20
SS-I_8R	TCACCAGATCCAAGCATGAC	50.00	60.11	20
SS-I_9F	GGCAGCAAATGTCTTGATGA	50.00	60.69	20
SS-I_9R	ATGGTTAGTGGCGAGAATGC	50.00	60.14	20
SS-I_10F	CCGTTTGCTGAGAAAGGAGA	55.01	59.71	20
SS-I_10R	ATGGTTAGTGGCGAGAATGC	64.76	60.54	20
SS-1_11F	GGTTGCTAAATTGCTCGACTG	45.82	55.07	21
SS-I_11R	TGATGGGAGCTAAATCCTATTGA	66.67	60.77	23

Table No. 3. List of primers with guanine/cytosine content (GC %) and annealing temperature (TM)

4.2 AUTO DIMER RESULTS:

22 primer sequences designed above were checked for dimer formations and hairpin structures. The results were found positive for both and no dimers /hairpins were found. Below are the screen shots after running auto dimer software. Minimum SCORE requirement was at 20.

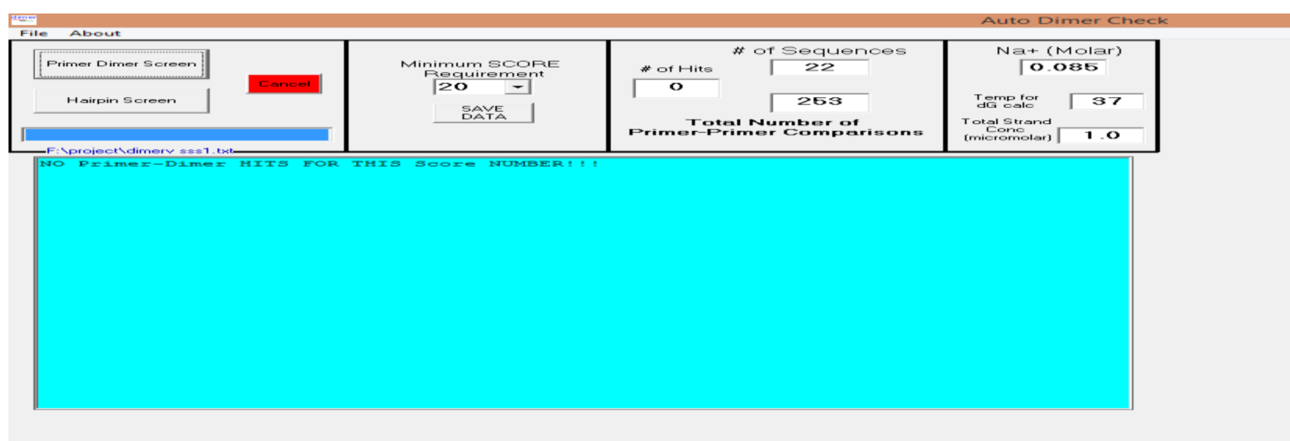


Figure 8. Result screen of Auto Dimer for Primer-Dimer check.

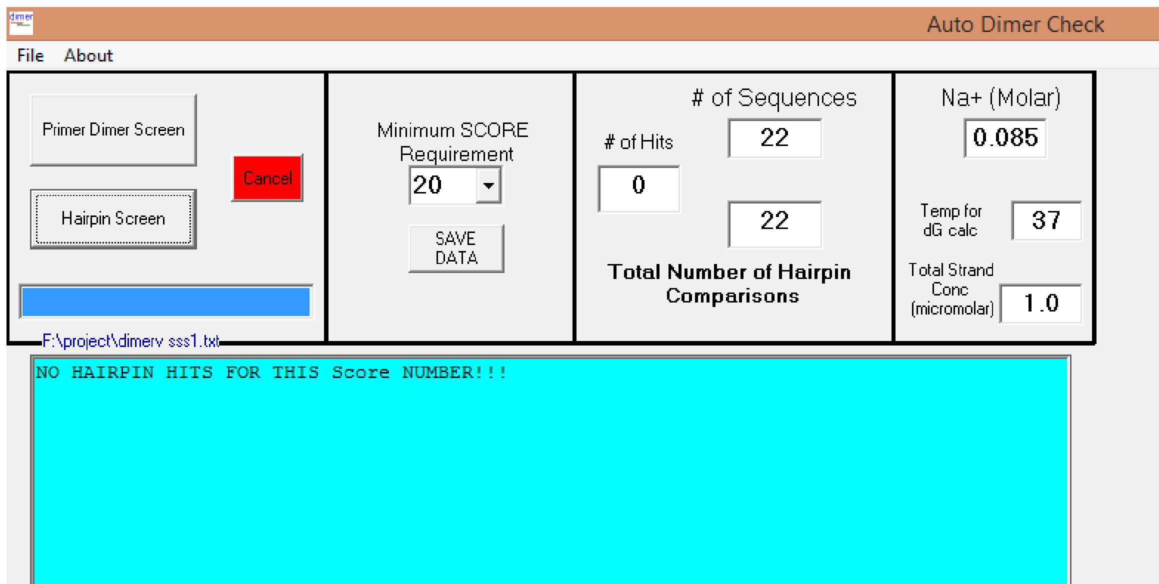


Figure 9. Result screen of Auto Dimer for Hair pin check.

4.3 SEED GERMINATION:

Seeds of twenty varieties of rice were procured from NBPGR Shimla and NBPGR Delhi (Table 1). For good germination dormancy breakdown treatment was given to the genotypes by soaking them in 500 ppm Gibberellic acid overnight at temperature 37 °C.

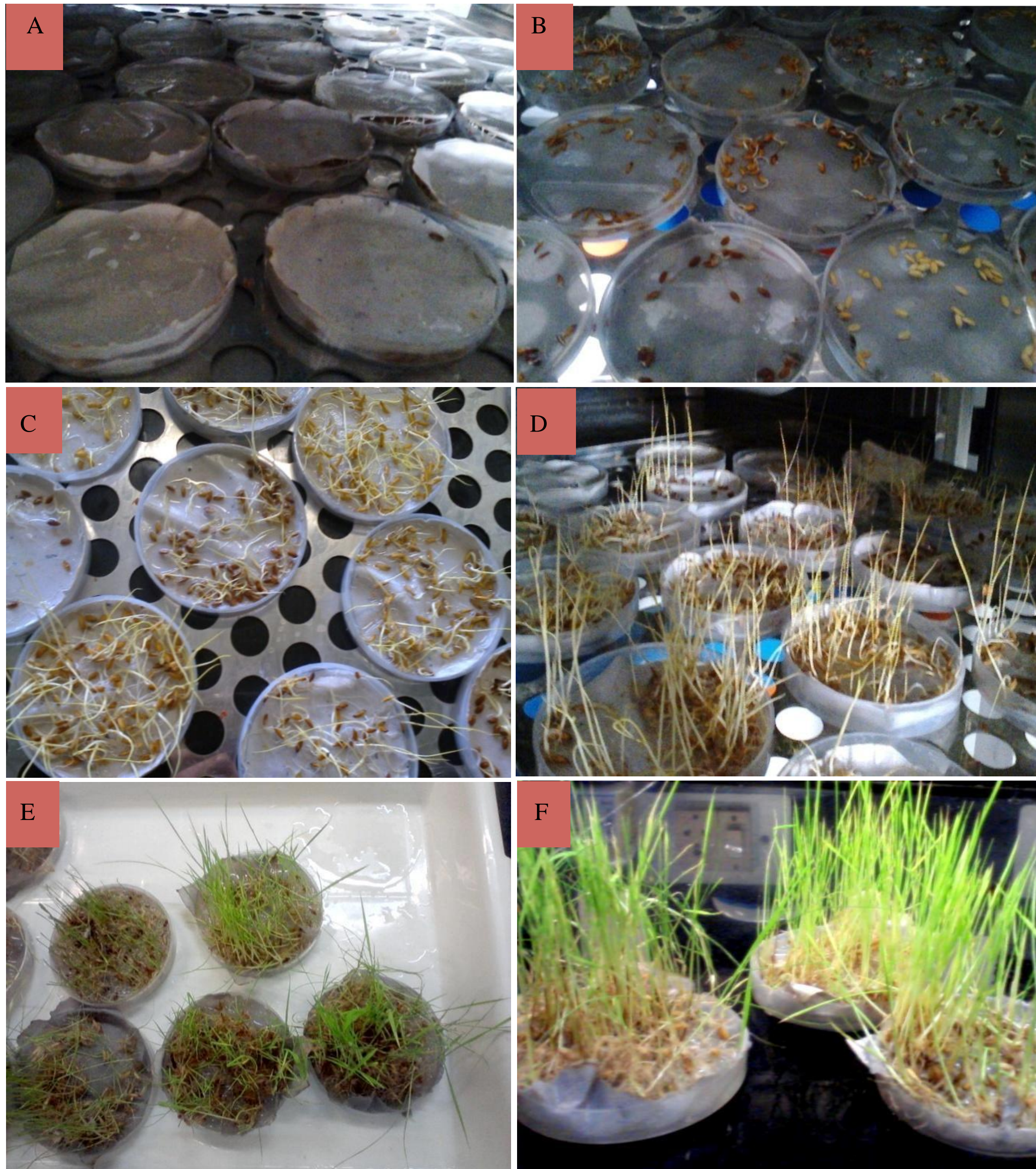


Figure 10. Different rice varieties growing in Petri plates containing blotting paper.

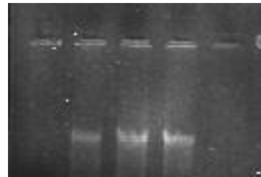
A. Day 1 seeds covered with blotting paper.
B. Day 5 seeds started to germinate and upper layer of blotting paper has been removed.
C. Seedlings of day 8.
D. Seedlings of day 14.
E. Seedlings of day 18.
F. Seedlings of day 24.

4.3 DNA ISOLATION RESULT:

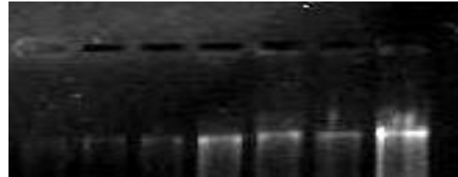
A

B

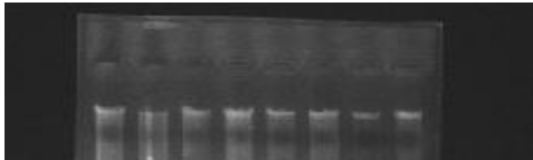
C



D



E



F

G

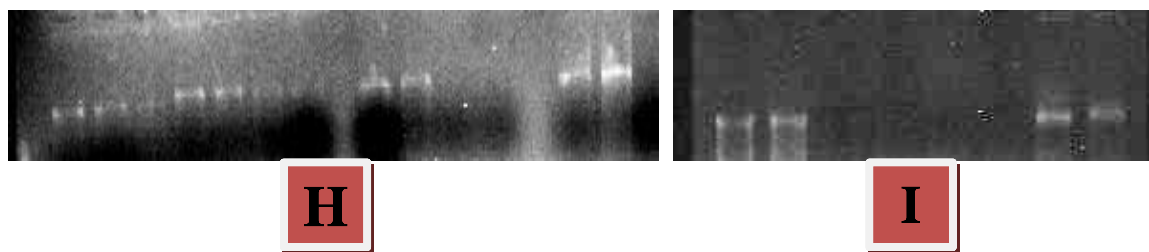


Fig.7 Different bands of DNA were obtained from different varieties of rice seedlings.

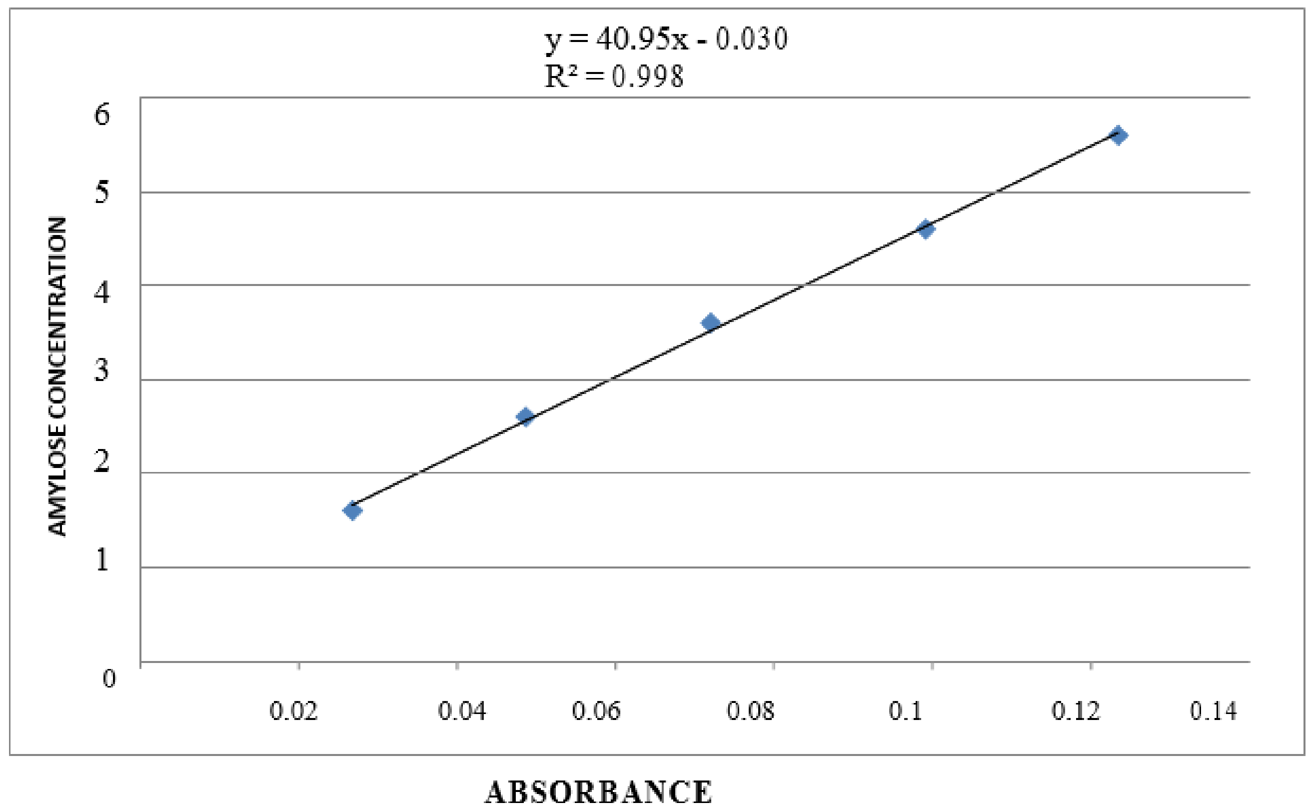
a. DNA bands obtained by running gel containing samples 1, 2, 3, 4, 5, &6.
b. DNA bands obtained by running gel containing samples 3, 4 & 5.
c. DNA bands obtained by running gel containing samples 6 & 7.
d. DNA bands obtained by running gel containing samples 8, 9 & 10.
e. DNA bands obtained by running gel containing samples 10, 11 & 12.
f. DNA bands obtained by running gel containing samples 13, 14, 15 & 16.
g. DNA bands obtained by running gel containing samples S2, S3, S4 & S5.
h. DNA bands obtained by running gel containing samples 10, 6, 7, 2, S3, S4 & 15.
i. DNA bands obtained by running gel containing samples 11, 12, S5 & S2.

4.4 AMYLOSE ESTIMATION RESULTS:

Quantification of Amylose:

Standard	A1 (Abs)(620nm	A2 (Abs 620nm)	A3 (Abs 620nm)	Avg (Absorbance)
Blank	0	0	0	0
1	0.0245	0.0265	0.0293	0.026766667
2	0.05	0.046	0.0499	0.048633333
3	0.0776	0.0691	0.0693	0.072
4	0.0989	0.0989	0.0996	0.099133333
5	0.1239	0.1288	0.1176	0.123433333

Table No. 4: This table is used in making the graph shown below. The procedure followed to obtain these readings is described in the Materials and Methods section. A1, A2, and A3 are the triplicated values.



The above graph is obtained by taking Concentration on the Y-axis and Absorbance at 620nm on the X-axis for each of the set of values.

Sample	Amylose Content (mg/ml)	Characterization
S5	22.406505	Intermediate
s2	8.8152	Very Low
S3	6.812745	Very Low
S4	11.01012	Very Low
7	7.70955	Very Low
1	9.908565	Very Low
2	9.621915	Very Low
3	14.912655	Low
4	9.155085	Very Low
5	14.544105	Low
11	15.88317	Low
12	13.1559	Low
13	14.1387	Low
14	11.833215	Very Low
15	15.35082	Low
6	10.39587	Very Low
8	14.07318	Low
9	9.53592	Very Low
10	9.875805	Very Low
16	13.82748	Low

Table No. 5: Amylose content Quantification on 20 *Indica* Varieties

This table lists the amylose content of 20 varieties of rice. The values are obtained from extrapolating from the standard amylose curve and based on which characterization is done.

Table No. 6: Co-relation between amylose content and Glycemic index of rice. It is seen that amylose content is inversely related to the Glycemic index.

If	Then
Amylose content is high	Low Glycemic Index and the rice grains will show high volume expansion (not necessarily elongation) and a high degree of flakiness. The rice grains cook dry, are Less tender, and become hard upon cooling.
Amylose content is low	High glycemic Index and the rice grains will cook moist and sticky

4.5 JAVA CODE FOR NUCLEOTIDE BASE SPLITTING:

TASSEL 5.0 follows a specific excel format of files to run the various inbuilt function.

Code:

```
The package bio;
import java.io.BufferedReader;
import java.io.IOException;
import java.io.InputStreamReader;
public class dipti {
public static void main(String[] args) throws IOException { String a;
char c[]; double b; int i;
InputStreamReader ir= new InputStreamReader(System.in); BufferedReader br=new
BufferedReader(ir); System.out.println("Enter the String:\n");
a=br.readLine(); b=a.length(); c=a.toCharArray(); for(i=0;i<=b;i++)
{
```

```
}}}            System.out.print(c[i]+"\t");
```

The code is run on **NET BEANS.**

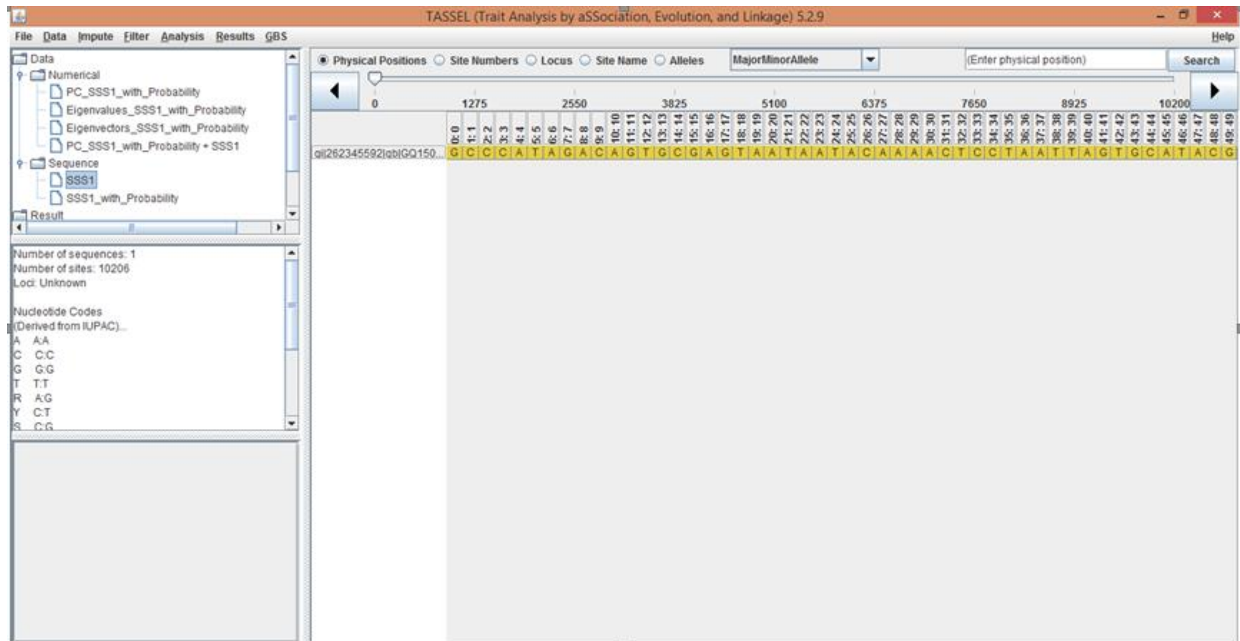


Figure 12. Screenshot of TASSEL 5.0 after loading a file with bases split. The sequence loaded is of SS1 ge

CONCLUSIONS

SSI is a major gene involved in many grain quality properties such as amylose content and grain elongation. SSI had a high association with apparent amylase content. This gene influence human health through its contribution to the glycemic index and levels of resistant starch. The incomplete digestion-absorption of resistant starch in the small intestine leads to non-digestible starch fractions with physiological functions similar to dietary fiber with significant beneficial impacts.

The characteristics of starch, such as, apparent amylose content (AAC), Gel Consistency (GC) and other physicochemical properties, determine the quality of various products of rice such as eating, cooking and processing qualities.

Primers were successfully designed using Primer 3 software. To check the primers, software Auto Dimerv was used. It efficiently listed the TM and GC percentage of the primers as well as was able to detect if the primers formed any hair pins and primer - dimers, so that amplification can be done efficiently.

Amylose content is inversely related to the Glycemic index. If amylose content is high then Glycemic Index will be low and the rice grains will show high volume expansion (not necessarily elongation) and a high degree of flakiness. The rice grains cook dry, are less tender, and become hard upon cooling. If amylose content is low then glycemic Index will be high and the rice grains will cook moist and sticky. The lowest amylose content of 6.812745 mg/ml was observed in S3 variety and intermediate amylose content of 22.406505 mg/ml was observed in S5 variety.

Java code for making TASSEL 5.0 excel sheet ran efficiently splitting the bases properly as to fit the format for the software.

REFERENCES

1. Asp, N. G. & Bjo¨rck, I. Resistant starch. Trends Food Sci Tech 3, 111–114 (1992).

2. Fan, J. &Marks, B. Retrogradation kinetics of rice flours as influenced by cultivar, Cereal Chem 75, 153–155 (1998).

3. Faisant, N. et al. Structural features of resistant starch at the end of the human small intestine, Eur J Clin Nutr 47, 285 (1993).

4. Ring, S.G., Gee, J. M.,Whittam,M., Orford, P.&Johnson, I. T. Resistant starch: its chemical form in foodstuffs and effect on digestibility in vitro. Food Chem 28, 97–109 (1988).

5. Philpot, K., Martin, M., Butardo Jr, V., Willoughby, D. & Fitzgerald, M.Environmental factors that affect the ability of amylose to contribute to retrogradation in gels made from rice flour. J Agric Food Chem 54, 5182–5190 (2006).

6. Hu, P., Zhao, H., Duan, Z., Linlin, Z. & Wu, D. Starch digestibility and the estimated glycemic score of different types of rice differing in amylose contents. J Cereal Sci 40, 231–237 (2004).

7. Chung, H. J., Lim, H. S. & Lim, S. T. Effect of partial gelatinization and retrogradation on the enzymatic digestion of waxy rice starch. J Cereal Sci 43, 353–359 (2006).

8. Chen, M. H., Bergman, C., Pinson, S. & Fjellstrom, R. Waxy gene haplotypes: Associations with apparent amylose content and the effect by the environment in an international rice germplasm collection. J Cereal Sci 47, 536–545 (2008).

9. Cai, X. L.,Wang, Z. Y., Xing, Y. Y., Zhang, J. L. & Hong, M. M. Aberrant splicing of intron 1 leads to the heterogeneous 5 UTR and decreased expression of waxy gene in rice cultivars of intermediate amylose content. Plant J 14, 459–465 (1998).

10. Larkin, P. D. & Park, W. D. Association of waxy gene single nucleotide polymorphisms with starch characteristics in rice (Oryza sativa L.). Mol Breed 12, 335–339 (2003).

11. Umemoto, T. et al. Effects of Variations in Starch Synthase on Starch Properties and Eating Quality of Rice. Plant Prod Sci 11, 472–480 (2008).

12. Umemoto, T. et al.Natural variation in rice starch synthase IIa affects enzyme and starch properties. Funct Plant Biolo 31, 671–684 (2004).

13. Waters, D. L. E., Henry, R. J., Reinke, R. F. & Fitzgerald, M. A. Gelatinization temperature of rice explained by polymorphisms in starch synthase. Plant Biotechnol J 4, 115–122 (2006).

14. Yan, C. J. et al. Genetic analysis of starch paste viscosity parameters in glutinous rice (Oryza sativa L.). Theor Appl Genet 1–14 (2010).

15. Fujita, N. et al. Function and characterization of starch synthase I using mutants in rice. Plant Physiol 140, 1070 (2006).

16. Shastry, B. S. SNP alleles in human disease and evolution. J Hum Genet 47, 561–566 (2002).

17. Imelfort, M., Duran, C., Batley, J. & Edwards, D. Discovering genetic polymorphisms in next generation sequencing data. Plant Biotechnol J 7, 312–317 (2009).

18. Masouleh, A. K.,Waters, D. L. E., Reinke, R. F. & Henry, R. J. A high throughput assay for rapid and simultaneous analysis of perfect markers for important quality and agronomic traits in rice using multiplexed MALDI TOF mass spectrometry. Plant Biotechnology J 7, 355–363 (2009).

19. Ohdan, T. et al. Expression profiling of genes involved in starch synthesis in sink and source organs of rice. J Exp Bot 56, 3229 (2005).

20. Dian, W., Jiang, H., Chen, Q., Liu, F. & Wu, P. Cloning and characterization of the granule-bound starch synthase II gene in rice: gene expression is regulated by the nitrogen level, sugar and circadian rhythm. Planta 218, 261–268 (2003).

21. Kharabian Masouleh,A., Waters, D. L. E., Reinke, R. F. &Henry, R. J.Discovery of polymorphisms in starch related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing{. Plant Biotechnol J 9(9),1074–1085 (2011).

22. Umemoto, T.& Aoki, N. Single-nucleotide polymorphisms in rice starch synthase IIa that alter starch gelatinisation and starch association of the enzyme. Funct Plant Biol 32, 763–768 (2005).

23. Waters, D. L. E.&Henry, R. J. Genetic manipulation of starch properties in plants:Patents 2001–2006. Recent Pat Biotechnol 1, 252–259 (2007).

24. Hirose, T., Ohdan, T., Nakamura, Y. & Terao, T. Expression profiling of genes related to starch synthesis in rice leaf sheaths during the heading period. Physiol Plantarum 128, 425–435 (2006).

25. Cao, H. et al. Identification of the soluble starch synthase activities of maize endosperm. Plant Physiol 120, 205–216 (1999).

26. Kawagoe, Y., Kubo, A., Satoh,H., Takaiwa, F. &Nakamura, Y. Roles of isoamylase and ADP-glucose pyrophosphorylase in starch granule synthesis in rice endosperm. Plant J 42, 164–174 (2005).

27. Tetlow, I. J. et al. Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein–protein interactions. Plant Cell Online 16, 694–708 (2004).

28. Rolletschek, H. et al. Antisense inhibition of the plastidial glucose 6 phosphate/ phosphate translocator in Vicia seeds shifts cellular differentiation and promotes protein storage. Plant J 51, 468–484 (2007).

29. Hirose, T. & Terao, T. A comprehensive expression analysis of the starch synthase gene family in rice (Oryza sativa L.). Planta 220, 9–16 (2004).

30. Sun, C., Sathish, P., Ahlandsberg, S., Deiber, A.& Jansson, C. Identification of four starch-branching enzymes in barley endosperm: partial purification of forms I, IIa and IIb. New Phytol 137, 215–222 (1997).

31. Yamakawa, H., Ebitani, T. & Terao, T. Comparison between locations of QTLs for grain chalkiness and genes responsive to high temperature during grain filling on the rice chromosome map. Breed Sci 58, 337–343 (2008).

32. Nishi, A., Nakamura, Y., Tanaka, N. & Satoh, H. Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. Plant Physiol 127,459 (2001).

33. Fujita, N. et al. Antisense inhibition of isoamylase alters the structure of amylopectin and the physicochemical properties of starch in rice endosperm. Plant Cell Physiol 44, 607 (2003).

34. Bao, et al. , Nucleotide diversity in starch synthase IIa and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (Oryza sativa L.)

35. Tanaka,et al.Structure, Organization, and Chromosomal Location of the Gene Encoding a Form of Rice Soluble Starch Synthase' Plant Physiol. (1995).

36. Ashish Jain et al, "Effect of cooking on amylose content of rice", European Journal of Experimental Biology, 2012, 2 (2):385-388.

37. Bienvenido O Juliano et al, "A gel consistency test for eating quality of rice", Journal of the Science of Food and Agriculture (2012).

38. Jiangbo Zhou, Aiqing You, "Association analysis of important agronomic traits in japonica rice germplasm", African Journal of Biotechnology, 2012, 11(12):2957-2970.