Identification of Hypervariable SSR Loci in the Rice Genome

THESIS

Submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

HARVINDER SINGH

Under the supervision of

Dr. N.K.Singh

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

PILANI (RAJASTHAN) 333 031

INDIA

2009

Identification of Hypervariable SSR Loci in the Rice Genome

THESIS

Submitted in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

By

HARVINDER SINGH

Under the supervision of

Dr. N.K.Singh

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

PILANI (RAJASTHAN) 333 031

INDIA

2009

ACKNOWLEDGEMENT

Indeed the words at my command are not adequate, either in form of spirit, to express the depth of my humility and humbleness before Almighty one without whose endless benevolence and blessings this tedious task could not have been accomplished.

I express my sincere gratitude to Vice-Chancellor Prof. L.K. Maheswari, and Dean Research and Consultancy Division Prof. Ravi Prakash, BITS, Pilani, for their kind support extended to me for the successful completion of this research work.

I place on record unbounded gratitude to my teacher and supervisor, Dr. N. K. Singh, Principal Scientist, NRC on Plant Biotechnology, IARI, New Delhi for showing me the real path of sincerity and dedication. In him I found that the best are not only efficient, effective and result driven, but at the "core they are persons with the best qualities as human beings". Humanity, scientific potentials, kindness, coolness, simplicity, novelty of ideas is the arms of his great personality. I am grateful and indebted to Dr. N. K. Singh for allowing me to work under his supervision, and for his invaluable painstaking efforts taken towards my study.

I express my deepest sense of gratitude and sincere thanks to Dr T. Mohapatra, Principal Scientist, Dr. T.R. Sharma, Principal Scientist, Dr. Kishor Gaikwad, Scientist (Senior scale), NRC on Plant Biotechnology, IARI, New Delhi have always been a great source of inspiration with their invaluable guidance and expert evaluation. I wish to place my heartfelt thanks to Dr Rakesh Singh, Senior Scientist, NBPGR, New Delhi for his support, and fruitful advice. I would like to express my profound gratitude to Dr.P.Anand Kumar, Project Director, NRCPB, IARI, New Delhi for giving me steadfast inspiration during the period of my research.

Lexicon fails to provide me enough words in pen totter scanning for Prof. S.K. Verma,Dr Utpal Roy Prof.A.K Das, Group Leader, Biological Sciences Group, BITS, Pilani, Mr Prateek Jain and Dr S.R.Thakur for all the suggestions and encouragement. For their concrete guidance and ever-encouraging attitude has always been a great source of inspiration. Their faith in my abilities has helped me to set higher goals for myself and achieve all that I deserve.

I am deeply indebted to my friends and colleagues Subodh, Dr.P.K.Singh ,Awadhesh, Pankaj Deepak sharma and my Bitsian friends and seniors Dr. S. Ramachandran, K.Laxminarayanan, Ashok, Dr. Ravi Vannela, Dr. Vishal Saxena, Dr Joy for their help and encouragement during my tough times. A mere 'thank-you' will not be enough for all the support they have given me.

I wish to express my appreciation and heartiest thanks to my friends and colleagues Rupesh, Devanshi, Sapna, Sambit,Irfan, Suhas, Chitra, Dr. Archana, Dr. Mahavir, Dr. Rekha, Dr. Batra, Vivek, Swaroop, Anand, and Shashi Ojha for their co-operation, moral support and help during my stay in the department. I would like to express my appreciation to our laboratory, field and office staff of NRCPB, for their kind help especially Sunil, Neeraj, Manmohan, Ravi, Jitendra and Birju.

It will never be possible for me to pay for the price of sacrifices which my parents have made for my success, the hardships to which they came across during my study, the afflictions with which they suffered, the tears which they shed for me and an agonizing long spell which they crossed for my ultimate settlement. These are the devout prayers of my mother Mrs Parkash Kaur, dreams of my father S.Swinder Singh and fervent feelings of my lovely younger brother Manvinder singh which brought me to this zenith. I also express my heartfelt reverence to my father and mother in laws sardarni and S.Manmohan singh, Harmeet, Chintu and Gunmeet, my aunty Baji , uncle and sister shalu as their love, affection and blessings are unreturnable.

Diction is not enough to express my gratitude to my beloved wife Rimpu, whose selfless love, constant encouragement, obstinate sacrifices, sincere prayers and blessings have always been the most vital source of inspiration and motivation in my life. She has always been my source of constant strength and love and has stood by me in every step of my life, through thick and thin.

I will be failing in my duties if I do not acknowledge the financial assistance and research facilities provided by Department of Biotechnology, Govt. of India, during Rice Genome Project.

New Delhi

Date: , 2009 ***(Harvinder Singh)***

ABSTRACT

Simple sequence repeats (SSR) are the DNA markers of choice for plant genetic analysis due to their abundance, high polymorphism and reproducible assays. In an attempt to find most variable SSR loci in rice, we studied the abundance, density and evaluated relationship between SSR length and level of polymorphism in a set of eight diverse rice genotypes. A total of 70,274 SSR loci were identified in the complete sequence of the rice genome. Di-nucleotide SSR were the most abundant making up 52.82% of all SSR loci. The abundance and density of SSR was comparable across the genome, for genic regions: 5' and 3'untranslated region (UTR), exons, introns and intergenic regions showed distinct patterns of distribution. At whole genome level the twelve rice chromosomes were found to be rich in (A+T%) whereas coding region was rich in (G+C%). The density of each class of SSRs in the genome was analyzed and it was observed that the overall SSR density was similar across the 12 rice chromosomes with an average of 190 SSR per Mb. The highest density of 197 SSR/ Mbp was observed for chromosome 6 and the lowest density of 167/ Mbp was for chromosome 4. For the evaluation of hypervariable SSR the rice SSR were grouped into six main classes based on their repeat length and primers were designed for 201 random SSR loci of different repeat motifs and lengths, representing genic as well as intergenic regions of the twelve rice chromosomes. There was a positive correlation between SSR length and average number of alleles per locus but linearity of this relationship was limited to the SSR length range of 10-70 bp. The highest level of polymorphism was in the SSR length range of 51-70 bp, beyond which there was a decline and stabilization of polymorphism in SSRs longer than 70 bp. Based on this trend, polymorphism level of 45 additional SSR loci with repeat lengths of 51-70 bp from the long arm of rice chromosome 11 was validated with an average of 2.1 alleles per locus. Here we describe a genome wide set of 832 hypervariable SSR (HvSSR) markers with repeat lengths of 51-70 bp representing all the twelve rice chromosomes that will be useful in creation of framework linkage maps for QTL mapping and fingerprinting in rice. About 436 of these highly variable SSR (HvSSR) markers were validated for their consistent amplification and high polymorphism. In the parental lines of three different mapping populations, the HvSSR loci showed more than twice the level of polymorphism than random SSR markers with average repeat length of 34 bp, and therefore are suitable for QTL mapping and fingerprinting studies in rice.

TABLE OF CONTENTS

ANNEXURES

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BAC	:	Bacterial Artificial Chromosome
BLAST	:	Basic Local Alignment Search Tool
bp	:	Base Pair
cDNA	:	Complementary DNA
Chr.	:	Chromosome
cM	:	Centi Morgan
CTAB	:	Cetyl Trimethyl Ammonium Bromide
cv	:	Cultivar
EDTA	:	Ethylene Diamine Tetraacetic Acid
EMBL	:	European Molecular Biology Laboratory
EST	:	Expressed Sequence Tag
e-value	:	Expect Value
HTGS	:	High Throughput Genome Sequencing
InDel	:	Insertion Deletion
IRGSP	:	International Rice Genome Sequencing Project
kbp	:	Kilo Base pair
MAS	:	Marker Assisted Selection
Mbp	:	Megabase pair
NCBI	:	National Centre for Biotechnology Information
PAGE	:	Polyacrylamide Gel Electrophoresis
		Pusa Basmati-1 (PB-1)
		Basmati-370 (B-370)
PIC	:	Polymorphism information content

PCR	:	Polymerase Chain Reaction
QTL	:	Quantitative Trait Loci
RGP	:	Rice Genome Research Program
SDS	:	Sodium Dodecyle Sulphate
SNP	:	Single Nucleotide Polymorphism
SSR	:	Simple Sequence Repeats
TIGR	:	The Institute of Genomic Research
Tm	:	Melting Temperature
CDS	:	Coding sequence

Rice (*Oryza sativa* L.) is cultivated worldwide and is the staple food for about a half of the world's population. It is agronomically and nutritionally most important cereal essential for food security, poverty alleviation and improved livelihoods (White, 1994). During the year 2005-06, rice production in India had increased manifold to 87.0 million tons from 42.4 million hectares of land as compared to 20.58 million tons from 30.81 million hectares in 1950-1951.

The genus *Oryza* includes 22 species: of which two are cultivated and 20 are wild. *O.sativa* is cultivated worldwide, and the word "rice" generally indicates a plant and a crop of this species. The basic chromosome number of the genus *Oryza* is 12. The progenitors of *O. sativa* are considered to be the Asian AA genome diploid species *O. nivara* and *O.rufipogon* and those of *O.glaberrima* are the African AA genome wild diploid species *O.barthii* and *O. longistaminata* (Fig. 1.1).

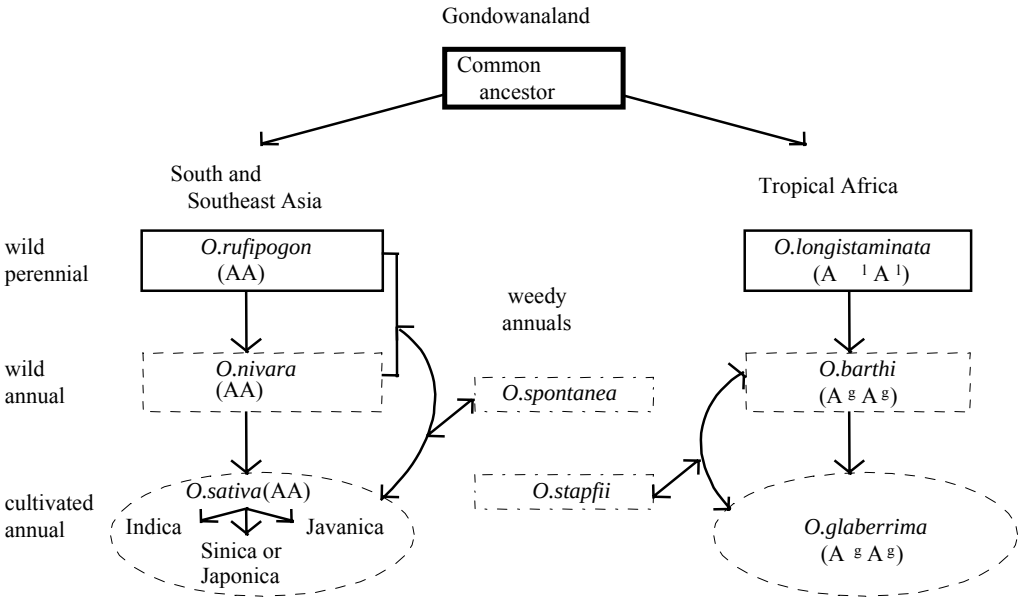


Fig.1.1. Evolutionary pathway of the cultivated species of rice

Rice has one of the largest *ex situ* germplasm collection in the world, and this accessible collection of diverse cultivated and wild rice germplasm has made great contribution to rice breeding.

The new revolution in rice production will be driven by the knowledge of rice genome sequence information and high throughput biology. Rice has one of the smallest genome among the crop plants with an estimated size of only 389 Mbp, and became the first economically important plant genome to be fully decoded, compare that to 16,000 Mbp genome size of bread wheat, the second most important food crop. The rice genome sequence provides a basic platform to study the distribution and function of the genes for yield, quality and resistance against biotic and abiotic stresses. Information obtained from sequencing the rice genome will impact agricultural and biological understanding in rice as well as other cereal crops. The sequencing was done using *O. sativa* L. ssp. japonica cv. Nipponbare, adopting a clone-by-clone approach, based on a comprehensive physical map containing minimum tiles of BAC/PAC clones and 10X coverage. The tremendous milestone of decoding complete rice genome sequence was achieved by the international Rice Genome Sequencing Project a consortium of research institutions from ten countries (IRGSP 2005). The finished quality sequence revealed a total of 37,544 non transposable-element-related protein coding genes as compared to 28,000-29,000 genes in *Arabidopsis* with a lower gene density of one gene per 9.9 kb in rice. A total of 2,859 gene were found to be unique to rice and other cereals, some of which might differentiate the monocot and dicot lineages.

Information on genetic relationships among individuals is of tremendous importance to plant breeders for variety and hybrid development. An estimate on the genetic similarity of breeding materials is best obtained using DNA markers. The development of DNA-based genetic markers has had a revolutionary impact on crop genetics. With DNA markers, it is possible to observe and exploit genetic variation in the entire genome. Popular genetic markers in the plant community include restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSR), single nucleotide polymorphism (SNP) and expressed sequence tags (EST) markers. The application of DNA markers has allowed rapid progress in the investigations of genetic variability and inbreeding, species and strain identification, and the construction of high-resolution genetic

linkage maps for plant species. Well-designed studies using these genetic markers will undoubtedly accelerate identification of genes involved in quantitative trait loci (QTL) for marker-assisted selection.

Among all the markers, SSR and SNP are of huge importance as these markers are the most abundant and are inherited in a co-dominant manner and can, therefore, distinguish between heterozygotes and homozygotes. All these factors together have made SSR an ideal marker for plant breeding, genetic linkage analysis, gene mapping, resolution of IPR disputes, conservation biology and population genetics. Simple sequence repeats (SSR) are stretches of DNA consisting of tandemly repeated short units of 1-6 base pairs in length. They are ubiquitous in eukaryotic genomes and can be analyzed through PCR technology. The rice SSR have been categorized into two groups based on length of SSR tracts and their potential as informative genetic markers: Class I SSRs contain perfect SSRs ≥20 nucleotides in length and Class II contain perfect SSRs >12 nucleotides and <20 nucleotides in length. The rationale for these two categories is that longer perfect repeats (Class I) are highly polymorphic, as evidenced by the experimental data originally reported for human (Weber 1990). The sequences flanking specific SSR loci in a genome are believed to be conserved within a particular species, across species within a genus and some times even across related genera. These flanking sequences, therefore, have been used to design primers for individual SSR loci and the SSR markers reveal polymorphisms due to variation in the lengths of the SSR at specific loci; they are multi-allelic and co-dominant in nature, reproducible, distributed throughout the genome therefore are very informative.

The present investigation is an effort to identify relationship between SSR structure and polymorphism level in rice that will in turn help to target the polymorphic SSR loci and such a panel of highly polymorphic SSRs will be helpful in finding genome wide polymorphism between any two parents or cultivars for generating high density linkage maps.The present investigation was undertaken using eight rice varieties with the following objectives:

1. In silico analysis of the SSR loci in the rice genome for their distribution in the genic and intergenic regions.

2. Identification and experimental validation of hypervariable SSRs in the EST database of rice.

3. Identification and experimental validation of hypervariable SSR on the whole rice genome.

Chapter 2 Review of Literature

Natural or artificially created genetic variability is a *sine qua non* for making selection for varietal improvement. In conventional plant breeding, selection is based on phenotypic expression and the individual giving highest phenotypic measurement is supposed to be the one having best genotypic constitution based on *a priori*. Such selection has given desirable results in enhancing the crop yield but it is certainly not a fool proof method of selecting plants with the most suitable genetic constitution. With the advent of new generation DNA markers for marker assisted breeding and fast accumulating genome sequences of various organisms in the public domain, it is now possible to map and tag a gene of interest for molecular breeding.

2.1 Decoding of the rice genome

For many years, rice has been the subject of numerous breeding studies aimed at developing cultivars with higher yield potential and better cooking and eating quality. It has also become a useful plant for studying biology, as a model plant of monocotyledons (monocots). The Rice Genome Research Program (RGP) in Japan has established the genomic resources such as EST sequences and YAC (yeast artificial chromosome) framework physical maps as well as the high density genetic maps and markers (Saji *et al.*, 2001). These maps and sequences served as the framework for establishing the map-based sequence of the rice genome. All biological processes involved in the life cycle of rice such as fertilization, germination, growth, development, photosynthesis, metabolism, and response to the environment are encoded in the genome. Obtaining the master plan (genome sequence) of the rice plant was therefore indispensable in understanding rice biology. For these reasons, rice has long been a target plant for a comprehensive genome analysis. However, although many researchers recognized the impact of this undertaking, it is such an enormous task for one research institution or even a single country. The International Rice Genome Sequencing Project (IRGSP), involving institutions from 10 countries, was therefore established in 1997 to pool resources and manpower, accelerate the

sequencing of the genome and ensure public access to the sequence data. After seven years of global collaboration, it successfully attained its goal and released an accurate Nipponbare genome sequence of 389 Mbp. The IRGSP set the accurate rice genome sequence as the ultimate goal (http://rgp.dna.affrc.go.jp/IRGSP/bnl/Guidelines.html) of the project. To achieve this goal, the clone-by-clone methodology coupled with the hierarchical shotgun sequencing was chosen. This strategy required a sequence-ready physical map that covered the whole genome with genomic BAC/PAC clones. The IRGSP constructed several genomic libraries with PAC and BAC clones with more than 68 times genome coverage (IRGSP, 2005).

2.2 Use of Molecular marker in Rice

DNA markers have distinct advantages over other markers as they are ubiquitous and unlimited in number, discrete and non-deleterious, inherited in simple Mendelian fashion, cover the whole genome and have none of the problems experienced with morphological markers and are free from epistatic interactions and pleiotropic effects (Swathi *et al.*, 1999). The DNA markers are more widely employed for molecular tagging of genes and molecular marker assisted selection (Gupta *et al.*, 1999). In several crops, molecular markers closely linked to traits of economic importance have been developed (Caetano-Anolles and Gresshoff, 1996) that allows indirect selection for desirable traits in early segregating generations at the seedling stage by side stepping the confounding effect of the environment. This situation changed dramatically since the first RFLP map of rice was constructed in 1980 at the Cornell University followed by the development of high-density maps (McCouch *et al.*, 1988).Today various types of DNA markers are employed for identification and manipulation of genes governing agronomically important traits in many crop plants. Ideal DNA markers should have some desirable properties such as: high polymorphism, co dominant inheritance (for determination of homozygous and heterozygous states of diploid organisms), frequent occurrence in the genome, selective neutral behavior (the DNA sequences of any organism are neutral to environmental conditions or management practices), easy access (availability),

easy and fast assay, high reproducibility and easy exchange of data between laboratories.

2.3 Simple sequence repeats (SSRs)

Although these highly variable loci are certainly a boon for individual assignment, pedigree, or parentage analysis, as well as for mapping studies, their use in classical analysis of population genetic structure has shortcomings. (Kim *et al.*, 1999) showed that estimates of population differentiation may be erroneously low because SSR loci have high within-population heterozygosity or may be inflated because of recent reductions in population size. In addition, a statistically significant result regarding differences in allele frequency between populations may not reflect biological significance because statistical power is often high for hypervariable loci (Kim *et al.*, 1999; Queney *et al.*, 2001). It is reported that population level inference based on microsatellite genetic variation is also affected by time and space.

SSRs are thought to mutate predominantly by slippage of DNA polymerase during replication, which generally results in gains or losses of single repeat units, depending on the DNA strand in which the slippage occurs (Levinson and Gutman 1987*b*; Schlötterer and Tautz 1992 ; Weber and Wong 1993 ; Primmer *et al.* 1996 ; Wierdl, Dominska, and Petes 1997). The mutation mechanism for microsatellites appears consistent with the theoretical stepwise mutation model (SMM; Kimura and Ohta 1978) in which mutations are additions or subtractions of repeat units in the case of microsatellites. Convergent or recurrent types of mutations, although a fundamental part of SMM, are not consistent with the standard infinite alleles model (IAM, Kimura and Crow 1964), which assumes that every mutation that occurs within a population creates a unique allele. A slippage mechanism of mutation, which may occur commonly only in microsatellites because of their molecular structure, clearly has implications for inferences based on population phenotypic diversity because alleles can return to previous allele sizes or states, retarding the separation of allele frequency profiles between populations.

If SSRs evolve in a stepwise fashion, convergence of unrelated alleles to a common size, size homoplasy, should be common at these loci, yet homoplasy has rarely been observed within populations (Estoup *et al.* 1995 ; Garza and Freimer 1996 ; Grimaldi and Crouau-Roy 1997 ; Culver, Menotti-Raymond, and O'Brien 2001). Homoplasy should also obscure the actual genetic distance between populations (Goldstein *et al.* 1995). This creates a paradox: microsatellites are useful because they are polymorphic, yet their mechanism of mutation obscures population differentiation by increasing homoplasy. Nevertheless, microsatellites do produce information generally concordant with other marker types, which suggests that homoplasy does not obscure differentiation of allelic frequencies. For example, microsatellites reliably discriminate the five major stocks of chinook salmon (*Oncorhynchus tschawytscha*) in California's Central Valley (Banks *et al.* 2000).

2.4 Abundance and distribution of SSR in the rice genome

Microsatellite DNA sequences were first studied in humans, where they were found to be abundant and dispersed throughout the genome (Hamada *et al.* 1982). Since that time, they have been found in a wide array of other eukaryotes including several monocot and dicot plant species including rice. Microsatellites consist of tandemly arrayed di-, tri- and tetra-nucleotide repeats, and are hypervariable and ubiquitously distributed throughout eukaryotic genomes. Among different crops species, the frequencies and occurrence of the most common dinucleotide repeats [$(AC)_n$ and $(GA)_n$] have been worked out in relatively greater detail. $(AAG)_n$ and $(AAT)_n$ are the most frequent trinucleotide repeats in plants (Gupta *et al.*, 1996). Human genome is estimated to contain on an average 10-fold more microsatellites than plant genome (Powell *et al.*, 1996). Microsatellites are not clustered in specific regions but rather uniformly distributed in different regions. Tomato is an exception in this regard as it shows clustering of microsatellite (Gupta and Varshney, 2000). Primers are designed using unique sequences on both sides of the flanking regions of the repeat motif region to avoid the complex multiple pattern, since these repeats are present up to thousand times throughout the genome, this ensures SSR to provide a simple banding pattern with a high level of polymorphism detected (Litt and Lutty,

1989; Tautz, 1989; Weber and May, 1989) The high level of polymorphism detected is basically due to the difference of number of repeats in each individual. In several crop plants, including soybean (Akkaya *et al.*, 1992), rice (Wu and Tanksley, 1993; Akagi *et al.*, 1997), barley (Becker and Heun, 1995) and wheat (Roder *et al.*, 1995), specific amplification of microsatellite loci has indicated that microsatellite DNA markers are more variable than RFLP markers. This new type of molecular markers, known also as simple sequence repeats (SSR) or simple sequence length polymorphism (SSLP) has been developed based on DNA sequence variation, is based on a 1-6 nucleotide core element that is tandemly repeated from two to many thousands of times. (Hamada *et al.* 1982, Litt *et al.* 1989; Tautz *et al.* 1986; Akkaya *et al.*, 1992) A different "allele" occurs at an SSR locus as a result of changes in the number of times a core element is repeated, altering the repeat region. This marker system gained a lot of importance in the last few years, and now, more than 8000 microsatellite markers are placed on the rice map (McCouch *et al.* 2002).They are co-dominant in nature, so very ideal for segregating populations. Differences in length at an SSR locus are detected with DNA amplification by the PCR using two oligonucleotide primers that complement unique sequences flanking the SSR locus. Sizes of the amplified products are then precisely determined by electrophoresis in either agarose or Polyacrylamide gels with detection by EtBr staining, autoradiography (using a single ^{32}p- labeled primer) or fluorescence (using a fluorescent labeled primer). Current research has suggested that the length variations between alleles at an SSR locus are created by slippage of DNA polymerase during the replication of the tandem repeat followed by a failure of DNA mismatch repair enzymes to restore the original sequence (Strand *et al.,* 1993)

SSR can provide more information more easily than previous DNA-based genetic marker technologies, such as restriction fragment length polymorphism (RFLP) and randomly amplified DNA (RAPD). Markers are co dominant, many alleles are found among closely related individuals, technically simple, inexpensive PCR technology can easily and rapidly be utilized to generate these markers without the use of radioactivity. It is also sensitive, only a small quantity of DNA is required, analytically simple, data are unambiguously scored, and

highly reproducible. Highly abundant markers are uniformly dispersed throughout genome as frequently as every 0.2 cM (McCouch *et al.* 2002), broadly applicable, loci are frequently conserved between related species and sometimes across genera (Moore et al, 1991), readily transferable, information can be communicated as simple sequences of primer pairs, and does not require the physical transfer of probes among laboratories; flexible, these markers can be used as sequence tagged sites to provide anchors between genetic linkage maps and physical chromosome locations. The current level of genome coverage provided by these simple sequence length polymorphisms (SSLPs) in rice is sufficient to be useful for genotype identification, gene and quantitative trait locus (QTL) analysis, screening of large insert libraries, and marker-assisted selection in breeding (McCouch *et al.*, 1997). The only significant limitation of SSR analysis may be the initial investment and the technical expertise required to clone and sequence the loci, which is not the case in rice since the whole genome sequence is now available. *In silico* identification of SSR loci becomes quite easy. Once primer sequences are designed and published, analysis of SSR loci will be practical for any laboratory capable of PCR and electrophoresis. The throughput and cost effectiveness of screening loci also could be greatly improved by multiplex PCR, which allows the simultaneous amplification and scoring of multiple SSR loci in a DNA sample in a single PCR reaction and a single lane of gel electrophoresis.

Through surveys of sequences in EMBL and Gene Bank databases, it was discovered that all possible di-and trinucleotide repeat motifs were present at 5- to 10- fold greater frequencies than expected by a random distribution. The only exception being (CG)n which occurred at significantly less than a random frequency. Other forms of imperfect and scrambled arrangements of repeat units ("cryptic repeat") also were extremely common in eukaryotic genomes. Slippage during DNA replication was proposed as a mechanism for the creation and hypervariability of these repeat elements. Regions of divergence between conserved genes in closely related species often contained these simple repeats. It was hypothesized (Tautz, 1989) that because SSR were so prevalent in the genome and so highly variable, mutations at these sites may be a more important

source of evolutionary variation than classical point mutations and chromosomal translocations.

The information content of SSR loci as a genetic marker is directly proportional to the number and frequency of alleles present in a population. It is known as the percentage of polymorphism or polymorphism information content (PIC) value. In a study of 100 $(AC)_n$ loci, (Weber and May, 1989), reported that 64% were perfect tandem repeats, 25% were imperfect, and 11% compound. The information content of the loci was directly correlated with the number of tandemly repeated units. PIC values ranged from close to 0 at n=10 up to 0.8 at n=24. As such, it was estimated that the human genome potentially contains ~7000 (AC)n loci with PIC >0.7.

The application of SSR marker analysis to plant genetics is only just beginning, but is being adopted very rapidly throughout the research community. A number of loci have been characterized in agricultural species including rice, soybean, maize, barley, rapeseed, and grape. In each case, preliminary work involved identifying SSR loci in existing sequence database, creating primer pairs for these loci, and surveying a small set of diverse germplasm for polymorphism. Subsequent efforts have involved cloning and sequencing SSR loci by screening size-fractionated genomic libraries, and using primer pairs flanking SSR loci to survey sets of individuals representing both agronomically important and distantly related species (Table 2.1). Electronic PCR was used for *in silico* identifying 2240 SSR loci in the rice genome (McCouch *et al.,* 2002).

SSR loci occur throughout plant genomes, but specific repeat motifs may occur at strikingly different levels of abundance than those found in animals. (Wang *et al.,* 1994), surveyed the relative frequencies of all SSRs with repeat units of 1-4 nucleotides (total length of repeat >20) in the Gene Bank and EMBL plant databases. The most abundant SSR was $(AT)_n$ followed in decreasing order by (A)n, (AG)n, (AAT)n, (AAC)n, (AGC)n, (AAG)n, (AATT)n and (AC)n. Eukaryotic genome is densely interspersed with tandem repeats termed microsatellites or simple sequence repeats (SSR) (Tautz and Renz, 1984; Weber and May, 1989). On average, 1 SSR was found in dicotyledons and

monocotyledons at every 21 and 65 Kb, respectively. Fewer SSRs containing (CG) nucleotides were found in dicots than in monocots. There was no correlation between abundance of SSRs and nuclear DNA content. Also, an extremely low frequency of SSRs was found in organelle DNA (1 per 317 Kb). The abundance of all tri-and tetranucleotide SSR combinations jointly were equivalent to that of the total dinucleotide combinations. Mono-, di-, and tetranucleotide repeats all were located in noncoding regions of DNA, while 57% of those trinucleotide SSRs (containing CG) were located within gene coding regions. All trinucleotide SSRs composed entirely of (A-T) were found in noncoding regions. A total of 18,828 Class 1 di, tri and tetra-nucleotide SSRs, representing 47 distinctive motif families, were identified and annotated on the rice genome provides information about the physical positions of all Class 1 SSRs in relation to widely used published SSRs (McCouch et al,. 2002) There was an average of 51 hypervariable SSRs per Mb, with the highest density of markers occurring on chromosome 3 (55.8 SSR Mb^{-1}) and the lowest occurring on chromosome 4 (41.0 SSR Mb^{-1}) (IRGSP, 2005). SSRs in different genic regions - 5'untranslated region (UTR), 3'UTR, exon, and intron - show distinct patterns of distribution in rice genomes. The much higher density of SSRs in 5'UTRs compared to the other regions and a strong affinity towards trinucleotide repeats in these regions was observed. On a genomic level, trinucleotide repeats are the most prevalent type in rice (Lawson and Zhang, 2006). Taken together, this information suggested that it may be possible to create high density maps of plant chromosomes using a mixture of tri- and tetranucleotide SSR motifs.

The analysis of SSR alleles in plant DNA is being greatly advanced by the capabilities for automated sizing of PCR products by GeneScan™ fluorescence-based detection (Applied Biosystems Division of Perkin-Elmer). Differences in allele sizes of only two nucleotides can be resolved reproducibly. A multiplex of different primer pairs tagged with fluorescent dyes of different colors allows multiple loci to be analyzed in a single PCR reaction and a single gel lane. This technology ultimately may be complemented with automated DNA extraction and PCR set-up, permitting very high rates of sample throughput and low unit cost for large-scale operations, such as genetic resources profiling or scoring segregation in marker assisted breeding programs. SSRs will be useful in the study of genetic

organization and variation in a myriad of ways. Their ease of use and high information content naturally will lead to the complementation and replacement of other types of genetic markers in many situations, but possibly also to novel applications not previously considered for DNA-based markers (Brown *et al.*, 1996). The application of SSR markers to varietal identification and for plant breeders and seed producers' right protection is already underway (Smith, 1994).

Identification of markers linked to useful traits has been based on construction of complete linkage maps and the study of co-segregation, or bulk segregation analysis (BSA) in case of simple traits. However, alternative methods such as the construction of partial maps and combination of pedigree and marker information have also proved useful in identifying marker/trait association. Molecular markers have been looked upon as tools for a large number of applications ranging from localization of a gene to improvement of plant varieties by marker-assisted selection. They have also become extremely popular markers for phylogenetic analysis adding new dimensions to the evolutionary theories. If we look at the history of the development of these markers, it is evident that they have been improved over the last two decades to provide easy, fast and automated assistance to scientists and breeders. Genome analysis based on molecular markers has generated a vast amount of information and a number of databases are being generated to preserve and popularize it. Microsatellite markers, especially SSR markers, have been found to be extremely useful in this regard. Owing to their quality of following clear Mendelian inheritance, they can be easily used in the construction of linkage maps, which can provide an anchor or reference point for specific regions of the genome.

2.5 Polymorphism level of the genomic and EST SSRs

The SSR markers reveal polymorphisms due to variation in the lengths of SSRs at specific individual loci; they are therefore, polyallelic and co-dominant in nature, thus proving to be very informative. Consequently, they have been used extensively not only for mapping SSR loci in many cereal species, but also for tagging a number of genes and for diversity studies in these crops (Li et al, 2000). However, development of SSR markers is expensive, labour intensive and time consuming, in particular, if they are being developed from genomic libraries.

Even though, due to importance of SSRs, they have been developed in a large number of plants including major cereal species such as barley (Temnykh et al, 2000) maize (Temnykh et al, 2001), oats (Saal et al, 1999), rice (Bhattramakki et al, 2000; Roder et al, 1998), rye (Varshney et al, 2000) sorghum (Thiel *et al*, 2001) and wheat (Scott *et al,* 2001; Scott *et al,* 2000). Due to current emphasis on functional genomics, ESTs (expressed sequence tags) are fast accumulating in EST databases of a large number of crop species. These EST databases can be mined for SSR containing ESTs that would serve for designing locus specific primers. Following this procedure, SSR markers can be obtained at significantly reduced costs, as EST derived SSRs are free by-product of the currently expanding EST databases. While EST-derived SSRs have been shown to be less polymorphic than those derived from genomic sequences (Eujayl *et al.*, 2001) they have some intrinsic advantages: they are quickly obtained by electronic sorting, are unbiased in their repeat type, are present in gene rich regions of the genome, and are still abundant . Since they represent the transcribed part of the genome, EST-based SSR markers lead to the direct mapping of genes. Further, compared to SSR markers derived from genomic DNA sequences those based on ESTs have a higher level of transferability among related species as they are located in more conserved regions of the genome. This higher transferability has been demonstrated sugarcane (Cordeiro *et al.*, 2001) at different taxonomic levels. For instance, when the ESTs/ genomic DNA derived SSR markers from sugarcane (Saccharum spp.) were used to related genera such as erianthus (Erianthus spp.) and sorghum (Sorghum spp.), EST-derived SSRs were found to be more superior in terms of transferability (Cordeiro *et al.*, 2001). Due to this attractive feature of the EST derived SSR marker and the availability of complete genome sequences for rice has made it possible to carry out the genome wide analyses.

The SSR loci have been categorized into two groups based on the length of SSR and their potential as informative genetic markers: Class I SSRs contain perfect repeat length of 20 nucleotides or higher and Class II SSRs contain perfect repeat length of 12-20 nucleotides. The rationale for these two categories is that longer repeats are highly polymorphic, as evidenced by the experimental data originally reported in human (Weber 1990). More mutations and contractions were

observed for longer tetranucleotide repeats (Xu *et al.* 2000). A size-dependent mutation bias (in which long alleles are biased toward contraction, whereas short alleles are biased toward expansion) is observed. (Huang *et al.* 2002). Length-dependent mutation patterns of SSR have been observed in different organisms, such as flies (Harr and Schlötterer 2000) and rice (Cho *et al.* 2000; Temnykh *et al.* 2000). The Class II SSRs tended to be less variable, representing sites where SSR expansion may occasionally occur but its probability is limited due to a smaller chance of slipped-strand mispairing over the shorter SSR template. SSRs shorter than 12 bp have a mutation rate that is no different than that of most unique sequences, and therefore demonstrate stochastic variation as has been shown in yeast (Pupko 1999).

2.6 Application of SSRs

Molecular marker technology has wide and diverse applications, they can broadly be classified into two main categories. The first category is genome analysis applications, which include mapping, tagging, map based cloning, gene pyramiding and marker-assisted selection (MAS) for genes of interest in both simple as well as quantitative traits (Wu and Tanksley, 1993). The second category is the fingerprinting applications including varietal identification, ensuring seed purity, phylogeny and evolution studies, diversity analysis, elimination of germplasm duplicates .The DNA markers are more widely employed for molecular tagging of genes and molecular marker assisted selection (Gupta *et al.*, 1999). Microsatellite markers have been used to distinguish the inter-varietal chromosome substitution lines of wheat (Korzun *et al.*, 1997). Microsatellite marker linked to fragrance (Cordeiro *et al.*, 2002) and assessment of purity of rice hybrid using microsatellite marker (Yashitola *et al.*, 2002). Microsatellite markers are powerful in screening large breeding populations (Lopez *et al.*, 2003).

The availability of increasing numbers of mapped SSLP markers can be expected to complement existing RFLP and AFLP maps, increasing the power and resolution of genome analysis in rice. (McCouch *et al.*, 1997). Large-scale cDNA analysis provides several great advantages for genome investigations in rice. Isolated and partially characterized cDNA clones have contributed not only to the

construction of an RFLP linkage map and physical maps of the chromosomes but also to investigations of the mechanisms of expression of various isozymes and family genes. The ultimate aim of large-scale cDNA analysis is to catalogue all the expressed genes of this important cereal, including tissue-specific, developmental stage-specific and stress-specific genes. The sequence data were translated into amino acid sequences for the 3 possible reading frames, and the similarity of these amino acid sequences to known proteins registered in PIR were examined. About 25% of the clones had significant similarities to known proteins. Some of the hit clones showed library-specific distributions, indicating that the composition of the clones in each library reflects, to some extent, the regulation of gene expression specific to differentiation, growth condition, or environmental stress. To further characterize the cDNA clones, including unknown clones, nucleotide sequence similarities of 24728 clones were analyzed and the clones were classified into around 10,000 independent groups, suggesting that around a half or one third of expressed genes in rice have already been captured. These results obtained from our large-scale cDNA analysis provide useful information related to gene expression and regulation in rice (Yamamoto and Sasaki, 1997).

Gramene (http://www.gramene.org) is a comparative genome mapping database for grasses and a community resource for rice. Gramene replaces the existing AceDB database 'RiceGenes' with a relational database based on Oracle. Gramene provides curated and integrative information about maps, sequences, genes, genetic markers, mutants, QTLs, controlled vocabularies and publications. It aims to use the rice genetic, physical and sequence maps as fundamental organizing units, to provide a common denominator for moving from one crop grass to another and is to serve as a portal for interconnecting with other web-based crop grass resources. This synteny among the crop grasses suggests that the rice genomic sequence will be more than a tool for understanding the biology of a single species because it can function as a window into the structure and function of genomes in the other crop grasses as well. Gramene provides researchers with access to the most widely used rice genetic maps. Associated with the maps is information on the underlying restriction fragment length polymorphisms (RFLPs), simple sequence repeats (SSRs), amplified fragment length polymorphism (AFLPs) and mapping populations. For sequence-based markers

such as SSRs, AFLPs and cDNA_RFLPs, Gramene provides researchers with information on experimental conditions, such as the primer, amplicon-sequence information, as well as the PCR amplification conditions (Ware *et al.*, 2002).

Table 2.1 A summary of major milestones in the development of genetic linkage maps for rice

Sr. no	Year	Reference	Finding/ Contribution
1.	1927	Chao	Early version of linkage map of rice chromosomes
2.	1963	Nagao and Takahashi	First rice map consisting of 12 linkage groups, corresponding to the haploid number of chromosomes
3.	1984	Kinoshita	Updated rice linkage map annually
4.	1984	Khush *et al.*	Primary trisomic stocks were used to assign the 12 linkage groups to their respective chromosomes
5.	1988	McCouch *et al.*	First RFLP map of rice was constructed from F_2 population from a cross between two major subspecies (*indica* and *japonica*) of cultivated rice covering 1,389 cM of the rice genome. Primary trisomics were used to assign linkage groups to each of the 12 chromosomes.
6.	1991	Saito *et al.*	Second RFLP map of rice based on different *indica*/*japonica* cross
7.	1994	Causse *et al.*	Saturated molecular map with 726 markers (mainly RFLP) was developed based on the interspecific backcross population covering 1491cM with an interval size of 4.0cM on the framework map and 2.0cM overall.
8.	1991	Knapp, S.J.	Described methods for estimating means of QTL genotypes and recombination frequencies between marker and quantitative trait loci using multilocus backcross, double haploid, recombinant inbred and test cross progeny models.
9.	1996	Singh *et al.*	>170 RFLP markers were assigned to specific chromosome arms through gene dosage analysis using the secondary and telotrisomics and the centromeres positions were mapped on all 12 linkage groups
10.	1996	Akagi *et al.*	Out of new 369 complete microsatellites, of which $(CGG/GCC)_n$ was the most frequent, in 11, 798 rice sequences in the database. Of these microsatellites, 35 out of the 45 could be

			successfully converted into microsatellite DNA markers using sequence information in their flanking regions. Integration of these markers with the published microsatellite DNA markers showed that about 35% of the rice chromosomes were covered by the 56 microsatllite DNA markers (35 newly published and 23 already published). They suggested that at least 90 microsatellite markers should be developed to cover for the construction of map and even more for construction of a microsatellite map
11.	1996	Tsunematsu *et al.*	Used 71 recombinant inbred lines (RILs) at F_6 and F_7 generations for developing RFLP framework map of rice, that covered a distance of 1275cM containing 375 markers. The RI lines showed a distorted segregation in some regions of chromosome 1, 3, 6, 11 and 12.
12.	1997	Chen *et al.*	94 microsatellite markers were integrated into existing RFLP framework maps of four rice populations and an interspecific backcross population. The SSRs used were predominantly poly (GA) motifs and were abundant in rice. The 94 SSRs with 27 previously described microsatellites provided genome wide coverage of the 12 chromosomes, with an average distance of 1 SSLP (simple sequence length polymorphism) per 16-20cM.
13.	1998	Temynkh *et al.*	Framework map solely based on microsatllie markers consisting of 300 SSLP markers
14.	1998	Harushima *et al.*	A high density rice genetic linkage map was constructed with 2275 markers (1455 out of total were ESTs) using 186 F_2 individuals covering 1521.6cM in the Kosambi function. 615 of the total ESTs used showed significant similarities to known genes including single copy, family and isozyme genes.
15.	2000	Temynkh *et al.*	188 new microsatellite markers were developed and evaluated for allelic diversity and finally integrated into existing map with124 SSR loci. The total 312 SSR markers provided whole genome coverage with an average density of one SSLP per 6cM
16.	2002	McCouch *et al.*	A total of 2414 new di-, tri- and tetra-nucleotide non-redundant SSR primer pairs, representing 2240 unique marker loci were developed and experimentally validated for rice

Summary

Among the different kinds of DNA markers, simple sequence repeats (SSR) of 2-6 bp motifs are of tremendous importance because of their relative abundance (1 SSR locus every 10 kbp), multiple alleles, co-dominant inheritance, uniform genome coverage and simple reproducible assays. The SSR markers have been used widely for rice germplasm evaluation and genetic dissection of quantitative trait loci controlling important agronomic traits. Use of SSR to interpret population structure provides much greater resolution than other types of markers because of the high level of polymorphism at individual SSR loci. It has been reported in rice that intervarietal polymorphism is positively associated with the SSR length but there is no systematic genome wide study to validate these assertions.

The aim of present study was to: (a) analyze the distribution of the SSR loci in the rice genome in the genic and intergenic regions. (b) To Identify and experimental validation of hypervariable SSRs in the EST database of rice. (c) Identification and experimental validation of hypervariable SSR on the whole rice genome. Such a relationship between SSR length and allelic polymorphism in rice will in turn help to identify hypervariable SSR loci with maximum polymorphism. Such panels of SSR markers will help find genome wide polymorphism between parental lines of QTL mapping populations and for fingerprinting of rice germplasm.

Chapter 3 Materials and Methods

3.1 Plant material

Sets of eight rice genotypes based on their diverse agro morphological and quality traits were used for the study of hypervariability (Table3.1).

Table 3.1. Rice Genotypes used for the analysis of polymorphism of SSR loci

Genotypes	Characters	Maturity
Basmati370	Aromatic	Medium
CSR27	Salt tolerant	Medium
Pusa1121	Aromatic	Medium
Jaya	High yielding	Medium
Swarna	High yielding	Late
CB11	NPT	Medium
Pusa Basmati1	Aromatic	Medium
Pusa1342	Long grain	Medium

3.2 In silico analysis tools:

The complete Rice genome sequences for twelve chromosomes were downloaded from TIGR rice pseudomolecules build 4.0 and the complete sequences for the gene were downloaded from our local database (www.nrcpb.org) for further analysis.

3.2.1 BLAST

The Basic Local Alignment Search Tool (BLAST)(http://www.ncbi.nlm.nih.gov/BLAST) was performed for the cDNA sequences against rice ESTs present in local database using optimized search parameter values of G=5, E=0.05, q=-1, r=1 at a word size of 11, Expect value of 10 and without complexity filter. Hits with bit score values of 100 or more were treated as significant and selected for further analysis

3.2.2 Multiple sequence alignment

For the identification of the hypervariable SSRs among the selected EST's for a particular cDNA, rice EST's were multiple aligned using T-COFFEE and BOXSHADE software's

3.2.2.1 T-Coffee (Tree-based Consistency Objective Function for alignment Evaluation)

The main feature of T-Coffee is the optimization method, which is used to find the multiple alignment that best fits the pair-wise alignments in the input library, it uses the progressive method. The valid input sequence format for this tool is fasta. (Notredame et.al., 2000)

3.2.2.2 BOXSHADE

This is a release of the PASCAL program BOXSHADE (http://www.ch.embnet.org) is a program for creating good looking printouts from multiple-aligned protein or DNA sequences.This server takes a multiple-alignment file in either GCG's MSF-format or Clustals ALN-format.

3.2.3 MISA (MIcroSAtellite) for mining of SSR

The software was used to scan the individual rice chromosomal sequences, downloaded from pseudomolecule sequence (build 3.0). Perfect dinucleotide to hexanucleotide simple sequence repeats were identified using the MISA (Thiel *et al.*, 2003) Perl scripts, specifying a minimum of five dinucleotide to hexanucleotide repeats and a maximum of 100-bp interruption for compound repeats. It takes input sequence file in fasta format and has the capacity to process multiple fasta files. Results of the microsatellite search are stored in two files:

1. In "<FASTfile>.misa" the localization and type of identified microsatellite(s) are stored in a tablewise manner.
2. The file "<FASTfile>.statistics" summarizes different statistics as the frequency of a specific microsatellite type according to the unit size or individual motifs.

3.2.4 SSR primer discovery tool for primer design

It is an automated process to identify and design PCR primers for the amplification of SSR loci this tool integrates SPUTNIK, an SSR repeat finder, with Primer 3, a PCR primer design program, into one pipeline tool, SSR Primer. On submission of multiple FASTA formatted sequences, the script screens each sequence for SSRs using SPUTNIK and the results are parsed to Primer3 for locus specific primer design.

The complete genomic sequences of the twelve rice chromosome pseudomolecules build 4.0 were downloaded from the TIGR site (www.tigr.org) and SSR loci were identified in these sequences at every 20 cM intervals in 5 cM windows. The SSR loci were grouped into six classes based on their repeat lengths, i.e. 10-30 bp, 31-50 bp, 51-70 bp, 71-90 bp, 91-110 bp and ≥111bp, and primer pairs were designed from the unique sequences flanking of these SSR sites (Annexure VI). Total 201 SSR loci containing di-, tri-, tetra- and penta-nucleotide repeat motifs were randomly selected from 68 different genomic intervals representing the whole rice genome. The pseudomolecule sequences of all the 12 rice chromosomes were split in to files using UNIX commands, feeding each of the SSR source sequences and specifying the target regions for PCR primer design using SSR primer discovery tool (Robinson *et al.* 2004).

For development of hypervariable SSR (HvSSR) markers the localization of SSRs was done by subjecting the twelve rice chromosome pseudomolecules build 5.0 to MISA tool (www.pgrc.ipk-gatersleben.de). All the SSR loci in the repeat size range of 51-70 bp were extracted and based on their position in the pseudomolecules sub dataset of sequence files were extracted using TIGR rice genome data extractor (www.tigr.org). The sub dataset sequence files were then subjected to SSR primer discovery tool for primer design. Total 833 HvSSR primer pairs were designed in this way, specifying an optimal Tm of 55°C with a minimum and maximum of 50°C and 60°C, respectively, GC content of 30-70% with low chance of di-nucleotide or hairpin loop formation and PCR product size of 100-400 bp.

3.3 Plant DNA Extraction

DNA was isolated by CTAB (Cetyl- Tetra Methyl Ammonium Bromide) method of Murray and Thompson, 1980.

3.3.1 Reagents:

1. 1M Tris-Cl (pH 8.0):

121g Tris-Cl base was dissolved in 800 ml double distilled H_2O, and pH was adjusted using concentrated HCl to 8.0, volume was made up to 1000ml and then solution was autoclaved.

2. 0.5 M EDTA (pH 8.0):

186 g di-sodium salt of EDTA was dissolved in 800 ml double distilled H_2O, and pH was adjusted to 8.0 using NaOH pellets, volume was made up to 1000ml and then autoclaved.

3. 5M NaCl:

292.2 g of NaCl was dissolved in 800 ml double distilled H_2O, volume was adjusted to 1000ml and then autoclaved.

4. DNA extraction buffer (100 ml working solution):

1M Tris-Cl (pH 8.0)	10.0 ml	(100mM)
0.5 M EDTA (pH 8.0)	4.0 ml	(20mM)
5M NaCl	28 ml	(1.4 M)
CTAB	2g	(2%)
β-Mercaptoethanol	200 µl	(0.2%)

5. 3M Sodium Acetate pH 5.2:

6. RNase A (10mg/ml):

10 mg RNase A was dissolved in 1 ml Autoclaved double distilled H_2O, vortexed well and then kept in water bath at 100^oC for 10-15 minutes, cool down and store at -20^oC.

7. Phenol Chloroform Isoamyl alcohol (PCI) mixture:

Ready-made Phenol-Chloroform Isoamyl alcohol (25:24:1) v/v was obtained from Amresco company.

8. Chloroform: Isoamyl alcohol (24:1)

Freshly prepared by mixing 96ml Chloroform and 4ml Isoamyl alcohol

9. Tris-EDTA buffer10:1(TE):

Tris (10mM): EDTA (1mM) buffer was prepared in 100 ml as follows:

1mM Tris-Cl, pH 8.0	1 ml
0.5M EDTA pH 8.0	0.2 ml
Autoclaved double distilled H_2O	98.8 ml

10. EtBr Stock Solution (10mg/ml)

10 mg ethidium bromide was dissolved in 1 ml double distilled H_2O.

3.3.2 DNA isolation protocol using Tissue Lyser (Qiagen):

25 mg of freshly harvested leaves from the eight rice varieties was placed into each tube of two collection microtube racks containing one tungsten carbide bead, the tubes were sealed with the caps supplied. Cool the collection microtubes in liquid nitrogen for 20 min. Now sandwich each rack of collection microtubes between adapter plates and fix into the tissue lyser clamps. Shake the sample for 2min. at 30 Hz. twice so as to ground the leaves to fine powder. Add 500µl of pre-warmed (65^{o}C) DNA extraction buffer. Suspension was incubated at 65^{o}C for one hour with intermittent mixing. The tubes were cooled to room temperature and equal volume of chloroform: isoamyl alcohol (24:1) was added and mixed by gentle inversion for 5 minutes. The contents were then centrifuged at 6000 g for 20 min. at room temperature. The aqueous phase was transferred to a fresh centrifuge tube with a wide bore tips to avoid DNA shearing and 0.6 volume chilled isopropanol was added. Contents were mixed by gentle inversion and incubated at –20^{0}C for 30 minutes for DNA precipitation. The precipitated DNA so obtained was pooled out by centrifugation at 10000 rpm for 5 minutes. The DNA pellet was washed twice with 70% ethanol, and dried at 37^{o}C for one hour, and dissolved in 500 µl TE buffer (pH 8.0).The crude DNA extracted was purified

by adding RNase (10mg/ml) to the sample at the rate of 5 µl/1g of leaf tissue used for the extraction, and was mixed gently by inversion. Mixture was incubated in water-bath for 1hour at 37^{o}C with intermittent mixing. Equal volume of Phenol Chlorpform Isoamyl alcohol solution was added to the mixture, mixed gently by inverting for about 5 minutes. Mixture was centrifuged at 10000 rpm for 10 minutes and the aqueous phase was taken out in a new 1.5 ml eppendorf tube. Equal volume of was added to the aqueous phase, mixed gently by inversion for 5 minutes and centrifuged at 10000 rpm. Chloroform: isoamyl alcohol (24:1) extraction was performed twice to remove all impurities. To the aqueous phase, 1/10th volume of 3M sodium acetate (pH 5.2) was added, a mixed gently, and then 2 volume of chilled ethanol was added to the mixture. Contents were mixed gently and incubated at –20^{o}Cfor 2 hr, centrifuged for 5 minutes at 10000rpm. The supernatant was discarded and pellet was washed twice with 70% ethanol. DNA Pellet was dried properly and dissolved in 100 µl TE buffer pH 8.0.

DNA was quantified using gel quantification method, in which the samples along with known concentration of λ genomic DNA as standard were loaded on 0.8% agarose gel prepared in 0.5 x TAE buffer. The intensity of individual samples was compared with a range of known amount of λ DNA (25, 50, 75 and 100 ng). Accordingly, samples were diluted to a concentration of 25ng /µl, and again rechecked loaded on 0.8% gel.

3.4 PCR Amplification of the SSR loci

PCR reactions were set up in 10µl volume containing 50 ng of rice template DNA, 5 pmole (13 ng) each of forward and reverse primers, 0.1 mM dNTPs, 1x PCR buffer (10mM Tris, pH 8.0, 50 mM KCl and 50 mM ammonium sulphate), 1.8 mM $MgCl_2$, and 0.2 unit of Taq DNA polymerase. The volume was made up to 10µl by autoclaved double distilled water. The cocktail was subjected to PCR amplification in a thermal cycler (Biometra or Applied Biosystems, USA). The PCR cycling conditions involved initial denaturation at 94^{o}C for 5 min followed by 35 cycles comprising of denaturation at 94^{o}C for 1 min primer annealing was in the range of 50.0^{o}C – 60.0^{o}C for 1 min and primer extension at 72^{o}C for 2min. This was followed by a final extension step at 72^{o}C for 7, min followed by storage at 4.0^{o}C before electrophoresis.

3.5 Electrophoretic separation of the PCR products

3.5.1 Agarose Gel Electrophoresis*:*

Three percent metaphor agarose (BMA, USA) was used to resolve the PCR amplified DNA. The gel was prepared by dissolving of metaphor agarose in small pinches to avoid clotting in 0.5x TBE buffer and heated for dissolution. Gel was cooled to ~65^{o}C and Gel star stain (BMA, USA) was added at the rate of 2.5µl /100ml. The gel was poured in to the cassette with combs and allowed to polymerize at room temperature. The gel along with tray and combs were shifted to the electrophoresis tank and combs were removed carefully. PCR samples were prepared for loading by mixing with 1x loading dye and loaded in the preformed wells. Electrophoresis was carried out in 0.5x TBE buffer at 80 volts for 3.5 hours. Gel photos were taken using gel documentation system (AlphaImager ^{TM} image acquisition with CCD camera, San Leandro California).

3.5.2 Poly Acrylamide Gel Electrophoresis (PAGE)

Stock solution of 30% Acrylamide: Bis-acrylamide (29:1) was prepared by dissolving 145g Acrylamide and 5 g Bis-acrylamide in 500 ml double distilled H_2O and the solution was filtered and degassed using Millipore filtering assembly (0.2µm size filter). The stock solution was stored in Dark bottle 4^{o}C.A 10% Ammonium persulphate (APS) solution was prepared afresh for each runby dissolving 100 mg in 1 ml distilled water.

For 100ml of 10% PAGE working solution, 67ml of 0.5x TBE buffer, and 33ml of 30% Acrylamide: Bis-acrylamide stock solution, 600µl of 10% APS and 110µl TEMED were mixed well and poured immediately into the gel cassette (Hoefer vertical-gel apparatus SE600, Amersham Biosciences) and left for polymerization for 30 min. After polymerization the samples were loaded and run at a constant current of 20mA per gel for 4.0 hrs. The gels were stained with EtBr solution (100µl of EtBr stock solution/liter double distilled H_2O) for 10 min and destained with double distilled H_2O for 15 min. Destained gels were exposed to UV light for visualization and images were captured with gel documentation system.

3.6 Estimation of polymorphism information content (PIC) values

The basic information about molecular markers that determines their application in genetic mapping is calculated for each marker using Polymorphism Information Content (PIC). The term PIC was originally introduced in the human genetics by Botstein *et al.* 1980. It refers to the value of a marker for detecting polymorphism within a population or set of genotypes by taking into account not only the number of alleles that are expressed but also the relative frequencies of alleles per locus. PIC = $1- \sum pi^2$ Where p_i is the frequency of the *i*th (presence of band) allele.

Chapter 4 Results

Genetic diversity is the primary requirement for the development of molecular markers and successful application of marker-assisted selection (MAS) for any trait in a crop. The excellent attributes and abundance of simple sequence repeat (SSR) markers in the rice genome have contributed significantly to the genomic studies. The SSRs are present at comparatively higher frequency in the non coding regions than coding region of the genome (Subramanian *et al.*, 2003). Very little is known about the biological significance of non-coding regions, therefore, a comprehensive analysis of SSRs is likely to help in understanding their importance. Information of SSRs for their abundance and distribution patterns in the coding as well as non-coding regions of the genome may give clue to the function of SSRs in gene regulation. With the public availability of rice genome sequence information, there is growing interest in identifying and characterizing genes associated with both qualitative and quantitative forms of phenotypic variation. Of particular interest to rice breeders is the possibility of using existing germplasm resources for gene and allele discovery on the basis of association mapping strategies. The present study was therefore, conducted with the aim to analyse the abundance, distribution and variability of SSRs not only at the genomic level but at specific genomic regions as well such as coding and non-coding sequence regions. The significant results of the study are described as under.

4.1 Distribution and abundance of SSRs in the rice genome

The MISA (MIcroSAtellite) software was used to scan the individual rice chromosomal sequences, downloaded from pseudomolecule sequence (build 3.0). The SSRs generated through MISA in the rice genome were traced for their distribution and abundance.

The total number of loci for all the SSRs on individual rice chromosome was determined. The number of repeat units at each locus was calculated for Di-, Tri-, Tetra-, Penta- and Hexa-nucleotide SSRs (Table 4.1; Fig. 4.1). The mono-nucleotide SSRs were not considered for the study because of their low

Table 4.1. Distribution and density (count/Mbp) of different classes of SSR in rice genome with more than 5X repeats identified using MISA tool

	Number of SSR loci in different repeat type						
Chromosome	**Di-**	**Tri-**	**Tetra-**	**Penta-**	**Hexa-**	**Total**	**Density(bp)**
Chr.1	4475	3556	303	77	43	**8454**	195
Chr.2	3516	3191	279	59	25	**7070**	196
Chr.3	3569	3230	242	65	24	**7130**	197
Chr.4	3177	2489	177	51	26	**5920**	167
Chr.5	2893	2651	207	65	28	**5844**	196
Chr.6	3363	2593	205	49	23	**6233**	203
Chr.7	2729	2289	198	44	12	**5272**	178
Chr.8	2900	2282	191	45	22	**5440**	191
Chr.9	2345	1769	186	30	15	**4345**	191
Chr.10	2313	1782	140	28	29	**4292**	189
Chr.11	2928	1826	239	41	10	**5044**	178
Chr.12	2914	2038	208	48	22	**5230**	189
Total	**37122**	**29696**	**2575**	**602**	**279**	**70274**	**189**
Total(bp)	74244	89088	10300	3010	1674	178316	

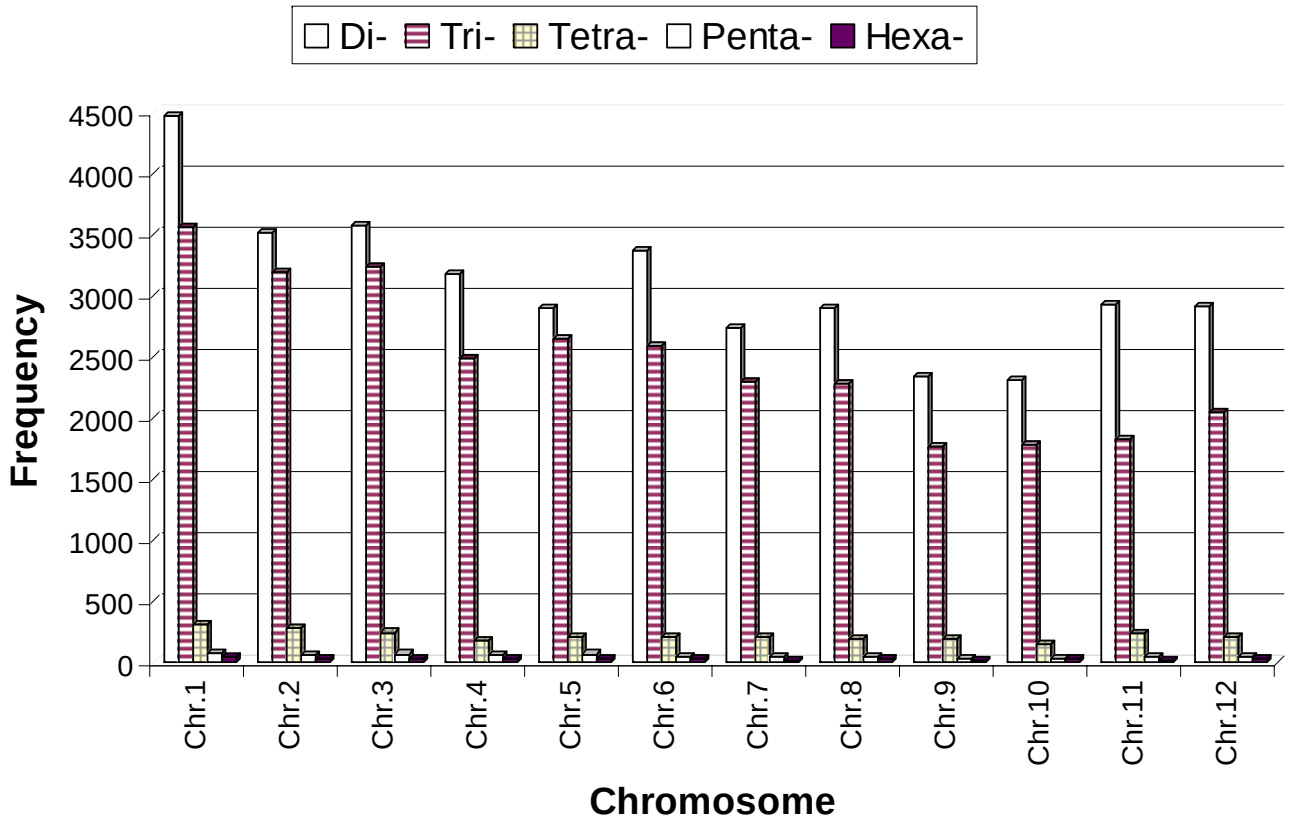


Fig. 4.1: Distribution of Simple Sequence Repeats (SSR) in rice genome

polymorphism levels. A total of 70,274 SSRs were identified. Di-nucleotide SSRs were the highest (37,122) with 52.82% coverage of total SSRs and covered a total length of 74,244bp in rice genome. Di-nucleotide SSRs were followed by tri-nucleotide SSRs (29,696) with 42.25% coverage of the total SSRs and covered 89,088bp of the rice genome. A total of 2,575 tetra-nucleotide SSRs were observed with 3.66% coverage and covered 10,300bp of the rice genome. Distribution of Penta- and Hexa-nucleotide SSRs were comparatively lower being 602 and 279, respectively and both covered less than 1% in the rice genome individually. Chromosome 1 (8454) has the highest and chromosome 10 (4292) the lowest SSR repeats occurrence (Fig. 4.1).

Most of the SSRs were found in the repeat range of 5X-15X (X=repeat no; Annexure III). The maximum dinucleotide repeats were within the range of 10-30bp (31,200) with the highest count on chromosome 1 (3,823) and the lowest on chromosome 10 (1,961), followed by 32-50bp (3,126) with the highest count on chromosome 11 (412) and the lowest on chromosome 9 (173). In, the repeat range 52-70bp (1676 SSRs) the highest count was observed on chromosome 11 (206) and the lowest on chromosome 10 (94). In repeat length of 72-90bp (856 SSRs) the highest count was observed on chromosome 11 (101) and the lowest on chromosome 4 (6). For the repeat length range of 92-110bp (193) the highest frequency was observed on chromosome 11(30) and the lowest on chromosome 4 (6). For the higher repeat number ranges the total SSR frequency was less than 50 ranging between 5and 20.

Similarly, for trinucleotide repeats 15-45bp (29,214) the highest count was observed on chromosome 1 (3,497) and the lowest on chromosome 9 (1,743). In the repeat length range of 48-75bp (295) the highest was on chromosome 1 and chromosome 11 (38) and the lowest on chromosome 4 (12). In repeat length range of 78-105bp the highest frequency was observed in chromosome 1 and chromosome 11 (20) and the lowest on chromosome 8 and 10 (4). For the higher repeat length ranges the total frequency was less than 50 ranging between 5 and 35.

The tetranucleotide followed the same pattern with highest total frequency in 5X-15X range i.e.20-60bp (2,477) with the highest count on chromosome 1

(296) and the lowest on chromosome 10 (128). For the repeat length of 64-100bp (65) the highest frequency was observed on chromosome 1 and chromosome 11(16) and the lowest on chromosome 4 (2). For the higher repeat ranges the total frequency was less than 50 ranging between 0 and 35.

For pentanucleotide repeat the highest frequency was observed for 25-75bp (600) with the highest count on chromosome 1 (77) and the lowest on chromosome 10 (28). For the repeat length range of more than 80bp the total frequency was less than 2 or no SSR was observed. The hexanucleotide repeat length between 30- 90bp (258) showed the highest frequency with the highest count on chromosome 1 (42) and lowest on chromosome 10 and 11 (10). For the repeat range of more than 95bp in hexa nucleotide repeats one or no SSR was observed.

Following whole genome pattern the order of SSR frequency for non-coding regions, intergenic (20,546) > introns (5,162)> UTRs (2,568) at 5X-15X repeat a range was observed to be highest (Annexure IV). Repeats for intergenic and introns were highest on the chromosome 1 (2,633 and 723, respectively) and the lowest on chromosome 9 (1,234) and chromosome 10 (284), respectively. However, in the UTR region the highest frequency for this range was observed on chromosome 3 (396) and the lowest on chromosome 11 (95). In 16X-25X (2,271) repeat range, the intergenic regions were observed at the highest frequency on chromosome 1 (297) and the lowest on chromosome 9 (126). Similarly, for this range in intron regions, chromosome 1 (76) showed the highest SSR frequency and the lowest on chromosome 9 and 10 (38). In UTR (166) region chromosome 3 (27) showed the highest frequency, whereas, the lowest frequency was observed on chromosome 10 (4). In 26X-35X repeat range, the highest frequency of SSR was observed on chromosome 1 (149) and the lowest on chromosome 10 (80). In introns (330) for 26X-35X repeat range the highest frequency was observed on chromosome 2 (43) and the lowest on chromosome 10 (17). In UTR (22), no SSR was found on chromosome 9 and 12. For intergenic regions in 36X-45X range (710) the highest frequency was observed on chromosome 2 (78) and the lowest on chromosome 4(31). In introns (163) the highest frequency for this range was on chromosome 9 (396) and the lowest on chromosome 3 and 4 (5), whereas, in

UTRs, 2 SSRs were observed one each on Chromosome 6 and 9. No SSR was observed in further classes >46X for UTR, whereas, in intergenic regions 144 and 33 total SSRs were observed in the repeat range of 46X-55X and ≥56X, respectively, however, on individual chromosome the SSR density was less than 20. In introns, 35 and 7 total SSRs were observed in the repeat range of 46X-55X and ≥56X, respectively, however, on individual chromosome the frequency was between 1-5 for 46X-55X and frequency for >=56X repeat class only 0-2 SSRs were found.

For trinucleotide repeats, in 5X-15X repeat class the frequency for total number of SSRs was comparable between introns and UTRs i.e. intergenic (8941)> UTRs (4887)> introns (3363). For intergenic and intron regions the highest frequency of 1059 and 420, respectively was observed on chromosome 1 and lowest on chromosome 10 (569 and 199, respectively), whereas, in UTRs the highest frequency was observed on chromosome 3 (722) and the lowest on chromosome 11 (212). In 16X-25X repeat range for intergenic regions (194), the highest frequency was observed on chromosome 6 (25) and the lowest on chromosome 4 (7). Similarly, for this range in introns (67) the frequency of SSRs on individual chromosomes was between 5 and 11. In UTRs the total number of SSRs was 16 with 0-5 SSRs on individual chromosome.

Similarly, for the occurrence of total number of SSRs in tetranucleotide repeats all the six classes of repeat ranges followed the similar pattern of Intergenic > introns > UTRs. In 5X-15X repeat range, both intergenic SSRs (1,765) and UTR-SSRs (223) were at the highest frequency on chromosome 1 (212 and 35) and the lowest on chromosome 10 (63) and 6(1) respectively. For introns (656) the highest frequency was observed on chromosome 2 (91) and the lowest on chromosome 10 (30). In 16X-25X repeat range, intergenic SSRs (51) were at the highest frequency on chromosome 7 (9) and the lowest on chromosome 4 (2). In introns for the 16X-25X class SSRs were observed in the range of 0-3. No SSR was observed for the higher repeat motif class. For UTRs in 16X-25X only one SSR was observed on chromosome 6 and for the rest of the motif class no SSR was observed.

Penta-nucleotide and Hexa-nucleotide showed the similar results for the Intergenic (431 and 129) > introns (105 and 37) > UTRs (82 and 23) in 5X-15X repeat range but the frequency was comparable across the chromosomes as in penta-nucleotide. The highest frequency SSRs in the intergenic regions was observed on chromosome 1 (55) and the lowest on chromosome 10 (20). For the introns, the highest frequency was observed on chromosome 5 (14) and the lowest on chromosome 7 (3). The UTRs showed the highest frequency on chromosome 3 (14) and the lowest on chromosome 11 (1). For the higher repeat classes no SSR was observed. Hexanucleotide repeats for the 5X-15X repeat range showed the highest frequency on chromosome 1 (23) and the lowest on chromosome 7 (3) in intergenic region. In introns, the highest frequency was observed on chromosome 2 (10) and no SSR was observed on chromosome 11. Similarly, in UTRs no SSR was observed on chromosome 9 and 10. SSRs on rest of the chromosomes ranged between 1 and 5. For the 16X-25X repeat range class only one SSR was observed in intergenic region on chromosome 12, whereas, all the higher classes for both penta-nucleotide and hexa-nucleotide repeats no SSR was observed on the 12 rice chromosomes for the three genomic regions.

For coding regions (cDNA) the SSR frequency showed the same pattern for di-, tri-, tetra-, penta- and hexa-nucleotide with the maximum frequency occurring in the range of 5X-15X (Annexure V). For di-nucleotide repeats in this range the highest frequency was observed on chromosome 3 (314) and lowest on chromosome 11 (103). For 16X-25X range the highest frequency was observed on 3 (20) and lowest on chromosome 11 (2). For 26X-35X repeat range the frequency was observed to be between 0-3, whereas only two SSRs were observed in 36X-45X range on chromosome 6 and 9 (1).

Trinucleotide repeats showed the highest frequency of SSRs (12626) as compared to other repeat types. The highest frequency was observed in the range of 5X-15X with the highest on chromosome 3 (1864) and lowest on chromosome 9 (528). For 16X-25X repeat range the highest frequency was observed on chromosome 1 (7) and no SSR was observed on chromosome 8 and 12. For 26X-35X repeat range the frequency was observed to be between 0-3, whereas no SSRs were observed in 36X-45X range.

In 5X-15X range for tetranucleotide the highest frequency was observed on chromosome 1 (24) and lowest on chromosome 11(9). For this range in pentanucleotide the highest frequency was observed on chromosome 3 (17) and lowest on chromosome 11(1). For hexanucleotide repeats in 5X-15X range the highest frequency was observed on chromosome 1 (16) and lowest on chromosome 9 (3). No SSR was observed for the rest of the higher ranges for tetra, penta and hexanucleotide repeat.

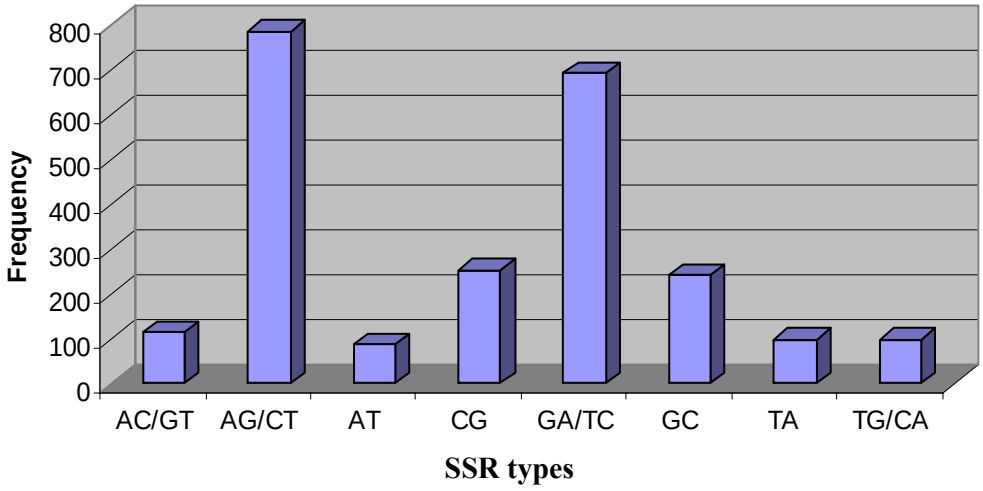
The frequency of dinucleotide repeats considering sequence complementarity AT/TA, AC/GT, TG/CA was lower in coding regions as compared to non coding regions, whereas, the dinucleotide repeats GC, CT/AG, CG and GA/TC were at higher frequency in the coding regions. In the UTR region CT/AG and GA /TC repeats were abundant, whereas, at the whole genome level AT, TA, CT/AG and GA/TC repeats were higher (Fig 4.2 a and b).

The trinucleotide repeats CGC, GCG, GCC, GGC and CGG were common to all the genomic regions. Interestingly, the frequency of trinucleotides in the coding regions (12,626) was higher to non-coding regions (intergenic, introns and UTRs). Within the non-coding region frequency of trinucleotide SSRs was the highest (4,887) in UTRs (Table 4.2) and the lowest in the introns (3363)

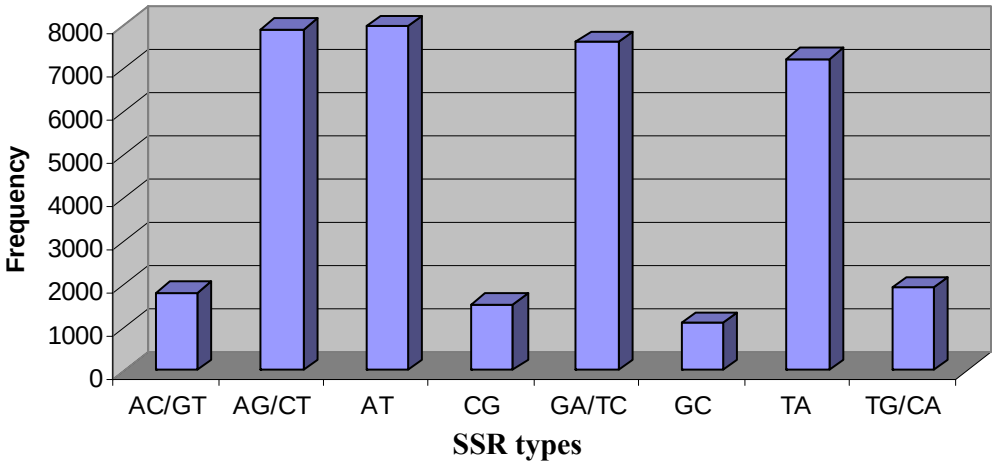
The occurrence of tetranucleotide repeats in whole genome and intergenic regions showed affinity towards ATAG, TATG, TATC and ATCT, whereas, rest of the genomic region did not showed affinity towards any particular repeat type. For pentanucleotide repeats at the whole genome level was found to be rich in AAAAG, GAAAA and TCTTT, however, within different genomic region the frequency of particular repeat type was variable. In case of hexanucleotide repeats the occurrence of different type of SSRs varied on all the chromosomes for all the genomic regions.

4.2 Comparative analysis of SSR density within coding and non-coding region

The density of each class of SSRs (Table 4.1) was analyzed and it was observed that the overall density is comparable across the 12 rice chromosomes



(a)



(b)

Fig 4.2: Frequency of dinucleotide SSR (a) in coding regions and (b) intergenic sequence of the rice genome

Table 4.2. Occurrence of di-, tri-, tetra-, penta- and hexa-nucleotide SSRs in the different regions of the rice genome

Repeat type	Number of SSR loci with different repeat numbers						
	05-15X	16-25 X	26-35 X	36-45 X	46-55 X	>=56 X	Total
Di-nucleotide							
Intergenic	20546	2271	1291	710	144	33	24995
Intron	5162	650	330	163	35	7	6347
UTRs	2568	166	22	2	0	0	2758
Coding (cDNA)	2153	106	18	2	0	0	2279
Tri-nucleotide							
Intergenic	8941	194	94	29	5	4	9267
Intron	3363	67	26	4	5	5	3470
UTRs	4887	16	4	0	0	0	4907
Coding (cDNA)	12626	30	11	0	0	0	12667
Tetra-nucleotide							
Intergenic	1565	51	13	4	2	4	1639
Intron	556	7	0	0	0	0	563
UTRs	223	1	0	0	0	0	224
Coding (cDNA)	189	0	0	0	0	0	189
Penta- nucleotide							
Intergenic	431	1	0	0	0	0	432
Intron	105	1	0	0	0	0	106
UTRs	82	0	0	0	0	0	82
Coding (cDNA)	67	0	0	0	0	0	67
Hexa-nucleotide							
intergenic	129	0	0	0	0	0	129
intron	37	1	1	0	0	0	39
UTRs	23	0	0	0	0	0	23
Coding (cDNA)	91	0	0	0	0	0	91
	63744	3562	1810	914	191	53	70,274

X=number of times a motif is repeated

.

Table 4.3: Average density (count/Mbp) of different classes of SSRs in the rice genome

Chromosome	Di-	Tri-	Tetra-	Penta-	Hexa-
1	103	82	7	2	0.99
2	97	88	8	2	0.69
3	98	89	7	2	0.66
4	89	70	5	1	0.73
5	97	89	7	2	0.94
6	109	84	7	1	0.74
7	92	77	6	1	0.4
8	101	80	6	1	0.77
9	103	77	8	1	0.66
10	101	78	6	1	1.27
11	103	64	8	1	0.35
12	105	73	7	1	0.79
Mean (whole genome)	99	79	7	1.4	0.7

with an average of 189 SSR per Mbp. The highest density was observed on chromosome 6 (203), followed by chromosome 3 (197), chromosome 2 and 5 {196), chromosome 1 (195) chromosome 9 and 8 (191) chromosome 12 and 10 (189) chromosome 7 and 11 (178), whereas, the lowest density (167) was observed on the chromosome 4.

Comparison of different repeat type of SSRs (Table 4.3) at different chromosomes showed that for the average density of Dinucleotide repeats, chromosome 6 (109) was densely populated, followed by chromosome 12 (105), chromosome 1, 9 and 11 (103), chromosome 8 and 10 (101), chromosome 3 (98), chromosome 2, 5 and 7 (97) and chromosome 4 (89). For trinucleotide repeats, highest density was observed on chromosome 3 and 5 (89), followed by chromosome2 (88), chromosome 6 (84), chromosome 1 (82), chromosome 8 (80), chromosome 10 (78), chromosome 7 and 9 (77), chromosome 12 (73), chromosome 4 (70) and on chromosome 11 (64). For tetranucleotide repeats, the highest density was observed on chromosome 2, 9 and 11 (8), followed by chromosome 1, 3, 5, 6 and 12 (7), chromosome 7, 8 and 10 (6) and on chromosome 4 (5). The penta-nucleotide repeats were at the highest density on chromosome 1,2,3,5 (2) and the lowest on chromosome 4, 6, 7, 8, 9, 10, 11 and 12 (1) The hexa-nucleotide repeats were at the highest density on chromosome 10 (1.27), followed chromosome 1 (0.99), chromosome 5 (0.94), chromosome 12 (0.79), chromosome 8 (0.77), chromosome 6 (0.74), chromosome 4 (0.73), chromosome 2 (0.69), chromosome 3 and 9 (0.66), chromosome 7 (0.4) and on chromosome 11 (0.35),

Average density (SSRs per Mbp region) for intergenic, intronic UTR and coding regions (Table 4.4) showed that for intergenic regions chromosome 6 showed the highest density (109.30 and the lowest density was observed on chromosome 4 (88.44). For intronic regions, chromosome 1 showed the highest density (32.17) and the lowest density was observed on chromosome 7 (24.52). The average density for UTRs (5' and 3') was the highest on chromosome 3 (32.71) and the lowest density was observed on chromosome 11 (11.49). For coding regions the highest average density was observed on chromosome 3 (62.11) and the lowest on chromosome 11 (25)

Table 4.4: Average SSR density (count/Mbp.) in the intergenic, introns, UTRs and coding regions of the twelve chromosomes of the rice genome

Chromosome	Intergenic	Intron	UTR	Coding (cDNA)
1	105	32	24	48
2	96	32	26	52
3	94	28	33	62
4	88	26	18	33
5	100	29	24	49
6	109	28	21	43
7	89	25	20	38
8	98	30	20	37
9	100	29	19	30
10	98	26	17	32
11	99	27	11	25
12	101	30	18	31
Mean	**98**	**28**	**21**	40

Fig.4.3. Multiple sequence alignment of coding sequence (CDS) of the gene 01-3726 with rice ESTs showing SSR length variability (nksrssr01_24287*)

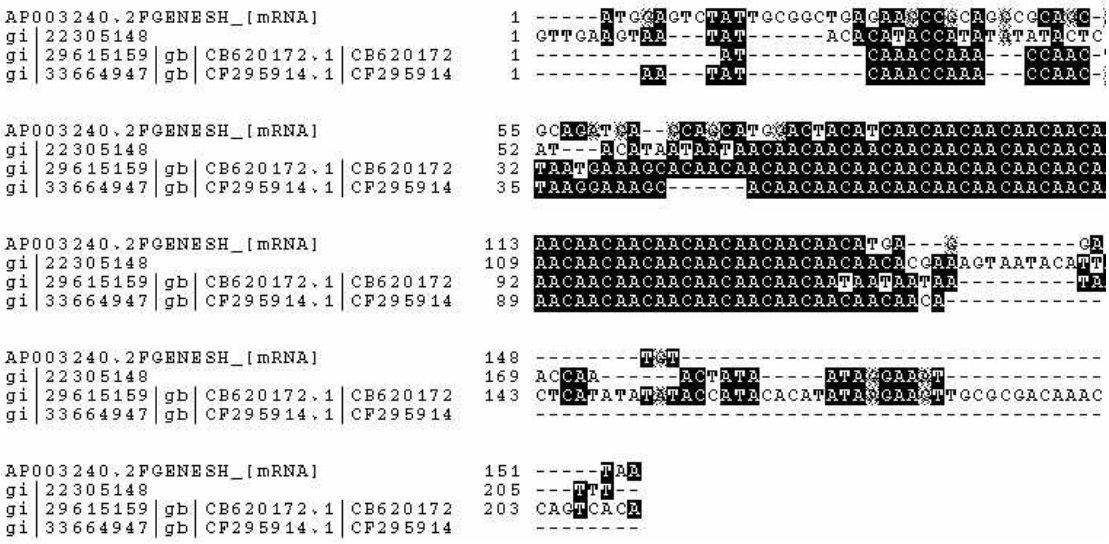


Fig.4.4. Multiple sequence alignment of coding sequence (CDS) of the gene 01-5610 with rice ESTs showing SSR length variability (nksrssr01_36481*)

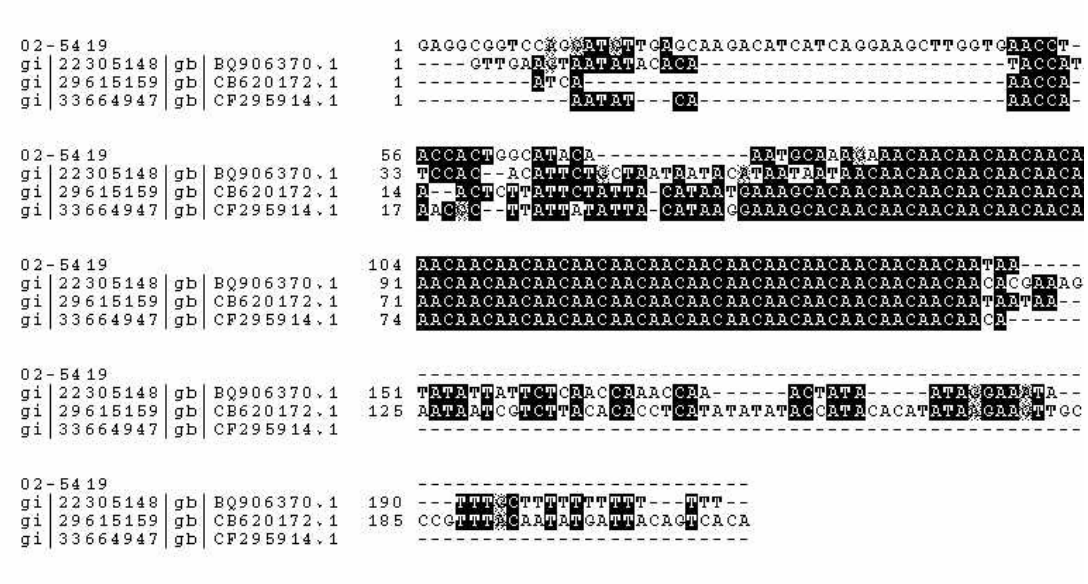
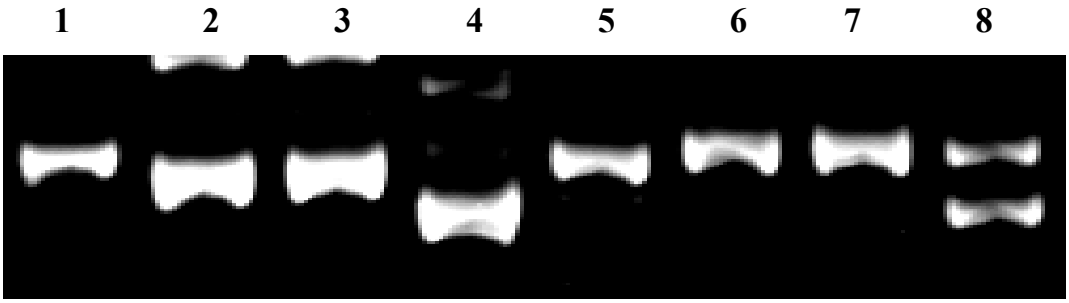
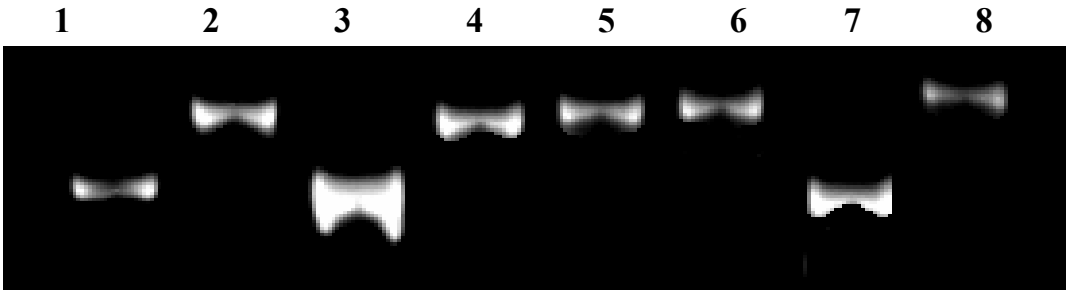


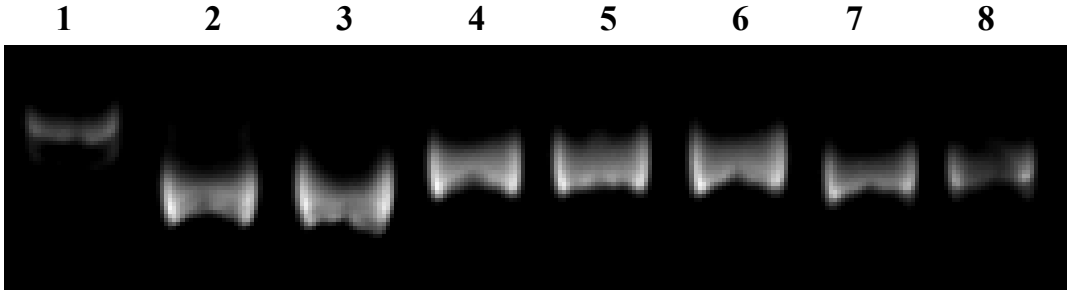
Fig.4.5 **Multiple sequence alignment of coding sequence (CDS) of the gene 02-5419 with rice ESTs showing SSR length variability (nksrssr02_35496*)**



(a)



(b)



(c)

Fig. 4.6 **PAGE gels showing different levels of polymorphism for coding regions in eight diverse rice genotypes with representative SSR loci. a (nksrssr01_24287*), b (nksrssr01_36481*), c (nksrssr02_35496*), 1 Basmati 370, 2 CSR 27, 3 Pusa 1121, 4 Jaya, 5 Swarna, 6 Pusa 1266, 7 Pusa Basmati 1, 8 Pusa 1342**

Table 4.5. Identification of hypervariable SSRs in genic regions (cDNA) through multiple sequence alignment of rice ESTs.

Sl No.	Gene ID	SSR	Size
1	nksrssr01_240*	(CAG)	30
2	nksrssr01_24287*	(ACA)	51
3	nksrssr01_36481*	(CAA)	57
4	nksrssr02_19743*	(GAA)	60
5	nksrssr02_35496*	(AAC)	66
6	nksrssr02_3423*	(AAG)	87
7	nksrssr05_21398*	(GAA)	75
8	nksrssr05_23557*	(CTT)	75
9	nksrssr05_5010*	(AGA)	45
10	nksrssr08_12289*	(AAG)	30
11	nksrssr09_22356*	(AAG)	66
12	nksrssr10_2214*	(GAA)	87
13	nksrssr11_16750*	(AAG)	66

The comparative analysis of average densities between intergenic, introns, UTR and coding regions depicts similar distribution pattern for average density between coding and UTR regions (Table 4.4) for all the chromosomes except for chromosome 10 where the average density increases for coding region and decreases for UTRs.

4.3 Identification and validation of the hypervariable SSRs in the coding regions of rice genome

The 56,298 cDNA sequences were downloaded from the local database (www.nrcpb.org) for the identification and localization of SSRs using MISA software. A total of 488 cDNA, containing both compound and perfect type I SSR (with SSR length above 20 bp) were selected for further study. The BLASTn search was performed for the selected 488 cDNA sequences against 2,60,282 Rice ESTs present in local database using optimized search parameter values of G=5, E=0.05, q=-1, r=1 at a word size of 11, Expect value of 10 and without complexity filter (Singh *et al.*, 2004). Hits with bit score values of 100 or more were treated as significant and selected for further analysis. A total of 313 cDNA (64.13%) showed significant similarity with the Rice EST's.

To identify the hypervariable SSRs among the selected EST's for a particular cDNA, rice EST's (bit score ≥100 and ≥20bp length) were multiple aligned using T-COFFEE and BOXSHADE software's. Out of the 313 cDNA with their respective EST analyzed, only 13 cDNAs showed considerable variability for SSRs (Table 4.5), of these three genes each were observed on chromosome 1, 2 and 5, one each on chromosome 8, 9, 11 and 10.

In the coding regions, the validation was performed for the 13 coding regions with variable SSRs e.g. nksrssr01_24287* (ACA) SSR coding Putative flower development regulator LEUNI (Fig. 4.3; Fig 4.6), nksrssr02_19743* with (GAA) SSR and nksrssr10_2214* with (GAA) SSR coding for hypothetical protein showed 37.5% polymorphism. The nksrssr01_240* showing variability for SSR CAG and coding for the OsNAC04-like protein, nksrssr05_21398* and nksrssr05_23557* were found to be monomorphic. Whereas, nksrssr02_35496* with (AAC) (Fig. 4.5; Fig 4.6) SSR coding a hypothetical protein, nksrssr08_12289* with (AAG) SSR coding for Putative far-red impaired response

protein, nksrssr09_22356* (AAG) SSR coding for Hypothetical 83.6kDa protein, nksrssr11_ 19358* (AAG) SSR coding hypothetical protein, gene nksrssr01_36481* (Fig. 4.4; Fig 4.6) with (CAA) SSR and nksrssr05_5010* (AGA) SSR coding for seven transmembrane protein Mlo6 were able to differentiate 25% of genotypes.

4.4 Relationship of SSR length and SSR motifs with allelic polymorphism

Parental polymorphism survey is often the first requirement for gene/ QTL mapping projects. However the level of polymorphism is not always satisfactory and quite often several hundred SSR markers are screened to find sufficient number of markers to create the framework molecular linkage map. Therefore the aim was to develop a set of hypervariable SSR (HvSSR) markers for the rice genome. The whole rice genome sequence was targeted to identify the HvSSR due to low level of polymorphism in the coding regions. The genomic sequences were downloaded at 20 cM interval with 5cM window and the SSRs identified in these sequences were further partitioned into 7 classes based on the repeat length (bp) to design SSR primers (Table 4.6). Total 201 SSR loci of different repeat lengths spread over the twelve rice chromosomes and representing both genic and intergenic sequences were tested on eight rice genotypes (Basmati 370, CSR 27, Pusa 1121, Jaya, Swarna, CB 11, Pusa Basmati 1 and Pusa 1342) all the 201 primer pairs generated single consistent amplification products. The data scored for these 201 markers was analysed to establish correlation between repeat length (bp) and the level of polymorphism. The 201 loci revealed 352 alleles with a range of 1 to 4 alleles and an average of 1.77 alleles per locus. 106 markers (52.73%) were polymorphic (Table 4.7; Annexure VI) whereas 95 markers generated single monomorphic product. The maximum polymorphism was detected with nine loci generating 4 alleles followed by 28 loci with 3 alleles and 69 loci with 2 alleles per locus (Fig. 4.7). There was a significant positive correlation between SSR length and average number of alleles per locus for both genic and intergenic SSR ($r = 0.234$, $P < 0.01$). However, the polymorphism level increased linearly only in the SSR length range of 11-70 bp, it decreased slightly in the 71-90 bp range and then remained more or less constant for higher repeat lengths (Fig. 4.8). The polymorphism level peaked in the SSR length range of 51-70 bp. The proportion of polymorphic loci also peaked in the 51-70 bp SSR length

group, but it showed a much sharper decline for SSR lengths higher than 111 bp. It was observed that the markers developed from chromosome no. 2 and chromosome no.3 showed the highest PIC values 0.53 and 0.52 respectively whereas the minimum PIC value was shown by the markers developed from the chromosome no. 11 (0.30) for the rest of the chromosomes the PIC values were between the range of 0.41 to 0.49 (Table 4.7).

The analysis of PIC values based on type of repeat motif is shown in (Table 4.8). On the basis of repeat motif the SSR motif AC/GT (0.525), TAA/TTA (0.663), AAC/ GTT (0.560) and AAT/ATT (0.526) showed maximum PIC values for di and tri-nucleotide repeats whereas, the minimum PIC value (0.22) was shown by the markers based on 3 different motifs (di-, tri- and tetra-nucleotide repeat). The markers with intermediate PIC values (0.28 to 0.50) were observed for rest of the 98 different repeat motifs. The proportion of polymorphic SSRs was least for pentanucleotide 3 out of 16. On the other hand, markers developed from dinucleotide repeats were the most polymorphic 42 out of 71 (59.2%), followed by the trinucleotide repeats 42 out of 76 (55.3%) and tetranucleotides 19 out of 38 (50.0%). These results suggest that it is preferable to analyse tri- and di-nucleotide repeat motifs for highest polymorphism among genotypes. The analysis of average PIC values for different repeat types showed highest PIC for the dinucleotide SSRs (0.467) followed by pentanucleotide repeats (0.460) and trinucleotide SSRs (0.445). The minimum PIC was observed for tetranucleotide repeats (0.395).

4.5 Validation of high polymorphism for HvSSR loci from rice chromosome 11

Based on the above results with 201 SSR loci, PCR primers were designed for amplification of additional 45 SSR loci of repeat lengths in the range of 51-70 bp from the genomic sequence of long arm of rice chromosome 11 that was sequenced in our laboratory as part of the International Rice Genome Sequencing Project (IRGSP 2005) for validation purpose (Annexure VII). The chromosome 11 was chosen primarily due to presence of large number of disease resistance genes that are being genetically mapped in our laboratory to facilitate marker-assisted breeding (Rice chromosome 11 and 12 sequencing consortia, 2005). Out

of the forty-five HvSSR loci analysed, forty gave clean amplification generating eighty-two different PCR products with an average polymorphism level of 2.50 alleles per locus. Of the forty amplified SSR loci, five generated three alleles each, thirty-two produced two alleles each and only three were monomorphic, this in contrast to the above results with 201 loci (Annexure VI) where 47.2% of the amplified SSR loci were monomorphic. Overall, the forty polymorphic HvSSR loci on chromosome 11 in the SSR length range of 51-90 bp showed an average PIC value of 0.36, which was higher than the 0.30 PIC value estimated for the ten polymorphic loci on chromosome 11 earlier. These results confirmed that maximum polymorphism was found in the SSR length range of 51-70 bp

4.6 Development and validation of 832 HvSSR markers for the whole rice genome

After establishing the validity of high allelic polymorphism for the SSR length range 51-70 bp, we mined all the 832 potential HvSSR loci containing di-, tri- and tetra-nucleotide repeat motifs with SSR lengths of 51-70 bp from the 12 rice chromosomes (Annexure VIII). The penta-nucleotide repeats were ignored because from the above analysis very small proportion of these loci was polymorphic and moreover there were only 2 such loci in the SSR length range of 51-70 bp. Most of these 832 loci are AT-rich and belong to the intergenic regions of the genome, indicating that the genic sequences normally devoid of large size SSR. Only thirty nine of the 832 HvSSR loci were from genic regions (Annexure VIII). Forward and reverse primer pairs were designed from the unique sequences flanking the 832 HvSSR loci and of these potential HvSSR loci, 436 were validated for consistent amplification at optimized annealing temperatures. The 436 HvSSR markers were evenly distributed in the rice genome, except that some regions of the chromosomes 4, 8 and 11 are poorly represented (Fig. 4.9). Most of these belong to intergenic regions (411 loci); indicating that the genic sequences are largely devoid of long SSR sites. These recombinant inbred lines (RILs) are being utilized for QTL mapping of important agronomic traits e.g. yield, quality and salinity tolerance. Earlier, we have used a large number of random SSR markers of RM series (McCouch *et al.* 2002) for parental polymorphism survey between the parental lines of these RILs with varying

Table 4.6. Frequency distribution of 201 SSR loci with different repeat motifs and lengths analyzed for the pattern of polymorphism in a set of eight diverse rice genotypes

Repeat motif	Repeat length (bp)												Total
	11-30		31-50		51-70		71-90		91-110		111-209		
	G	IG	G	IG	G	IG	G	IG	G	IG	G	IG	
Di	1	8	0	13	0	13	3	12	1	13	1	7	**71**
Tri	7	11	9	12	6	10	7	4	0	4	1	4	**76**
Tetra	0	18	0	6	0	13	0	1	0	0	0	0	**38**
Penta	1	14	0	1	0	0	0	0	0	0	0	0	**16**
Total	**9**	**51**	**9**	**32**	**6**	**36**	**10**	**17**	**1**	**17**	**2**	**11**	**201**

G = genic; IG intergenic

Table 4.7 Polymorphism information content (PIC) values of 106 polymorphic markers from the total 201 SSR markers showing PCR amplification in eight rice genotypes

Primer Id[a]	No. of alleles	PIC	Primer Id[a]	No. of alleles	PIC
nksrssr01_20	4	0.72	nksrssr05_9226	3	0.65
nksrssr01_5940	4	0.72	nksrssr05_9458	2	0.38
nksrssr01_9791	2	0.49	nksrssr05_19770	2	0.38
nksrssr01_10282	4	0.56	nksrssr05_23557*	2	0.50
nksrssr01_19351	3	0.41	nksrssr05_23679	3	0.61
nksrssr01_24287*	3	0.65	nksrssr05_29392	2	0.24
nksrssr01_25171*	2	0.32	nksrssr06_293	2	0.22
nksrssr01_25279	2	0.49	nksrssr06_534	3	0.41
nksrssr01_27499	2	0.22	nksrssr06_535	3	0.45
nksrssr01_28113*	2	0.41	nksrssr06_1086	2	0.47
nksrssr01_36481*	2	0.38	nksrssr06_23125*	3	0.65
nksrssr01_38671*	2	0.50	nksrssr06_30840	3	0.63
nksrssr01_42559*	2	0.50	nksrssr06_30941	2	0.44
nksrssr02_83	2	0.50	nksrssr07_19359	2	0.41
nksrssr02_125	4	0.69	nksrssr07_20547*	2	0.32
nksrssr02_806	4	0.67	nksrssr07_21222*	3	0.57
nksrssr02_1078	2	0.47	nksrssr07_21630	3	0.63
nksrssr02_5324	2	0.49	nksrssr07_22533	2	0.38
nksrssr02_11027	2	0.44	nksrssr07_22775	4	0.72
nksrssr02_12192	2	0.47	nksrssr07_23109	2	0.22
nksrssr02_19743*	3	0.59	nksrssr07_23468	2	0.47
nksrssr02_35496*	2	0.47	nksrssr08_5696	3	0.41
nksrssr03_105	2	0.49	nksrssr08_5791	3	0.50
nksrssr03_1022	2	0.50	nksrssr08_5844	2	0.24
nksrssr03_5219	2	0.41	nksrssr08_5968	2	0.50
nksrssr03_5219	2	0.41	nksrssr08_6088	2	0.32
nksrssr03_5694	2	0.49	nksrssr08_7796	2	0.41
nksrssr03_5942	2	0.22	nksrssr08_9910	3	0.59
nksrssr03_11652	4	0.72	nksrssr08_12289*	2	0.38
nksrssr03_12043	4	0.69	nksrssr08_20398	2	0.47
nksrssr03_12086	2	0.22	nksrssr08_22813	2	0.38
nksrssr03_14678	2	0.22	nksrssr08_23418	2	0.28
nksrssr03_20495	3	0.53	nksrssr09_2291	2	0.50
nksrssr03_26474	2	0.50	nksrssr09_2463	2	0.50
nksrssr03_27842	2	0.22	nksrssr09_2682	3	0.57
nksrssr03_28280	2	0.22	nksrssr09_5365	2	0.44
nksrssr03_28762	2	0.50	nksrssr09_9205	2	0.38
nksrssr04_4030	2	0.50	nksrssr09_9921	3	0.64
nksrssr04_15571*	3	0.59	nksrssr09_10470	2	0.22
nksrssr04_16014	2	0.50	nksrssr09_16935	2	0.38
nksrssr04_16255	4	0.69	nksrssr09_22356*	2	0.49
nksrssr04_19472	3	0.45	nksrssr10_14*	2	0.22
nksrssr04_19845	3	0.41	nksrssr10_22114*	3	0.61
nksrssr04_25032	3	0.61	nksrssr11_7483	2	0.32
nksrssr04_25723	3	0.61	nksrssr11_14014	3	0.59
nksrssr04_30674	2	0.49	nksrssr11_16750*	2	0.49
nksrssr04_34639*	2	0.41	nksrssr11_19358*	3	0.41
nksrssr05_328	3	0.61	nksrssr11_21743	2	0.38
nksrssr05_3420	2	0.24	nksrssr11_23495	2	0.22
nksrssr05_3620	3	0.61	nksrssr11_23831	2	0.28
nksrssr05_5010*	2	0.22	nksrssr11_24183	2	0.24
nksrssr05_7506	2	0.22	nksrssr11_27039	3	0.63
nksrssr05_9030	2	0.41	nksrssr11_27617*	2	0.22

[a]Primer Id, eg. "nksrssr11_19358" includes, lab Id (nks), rice simple sequence repeats (rssr), chromosome number (11), and base pair position in the rice pseudomolecule build 4 (19358 kbp). *genic SSRs

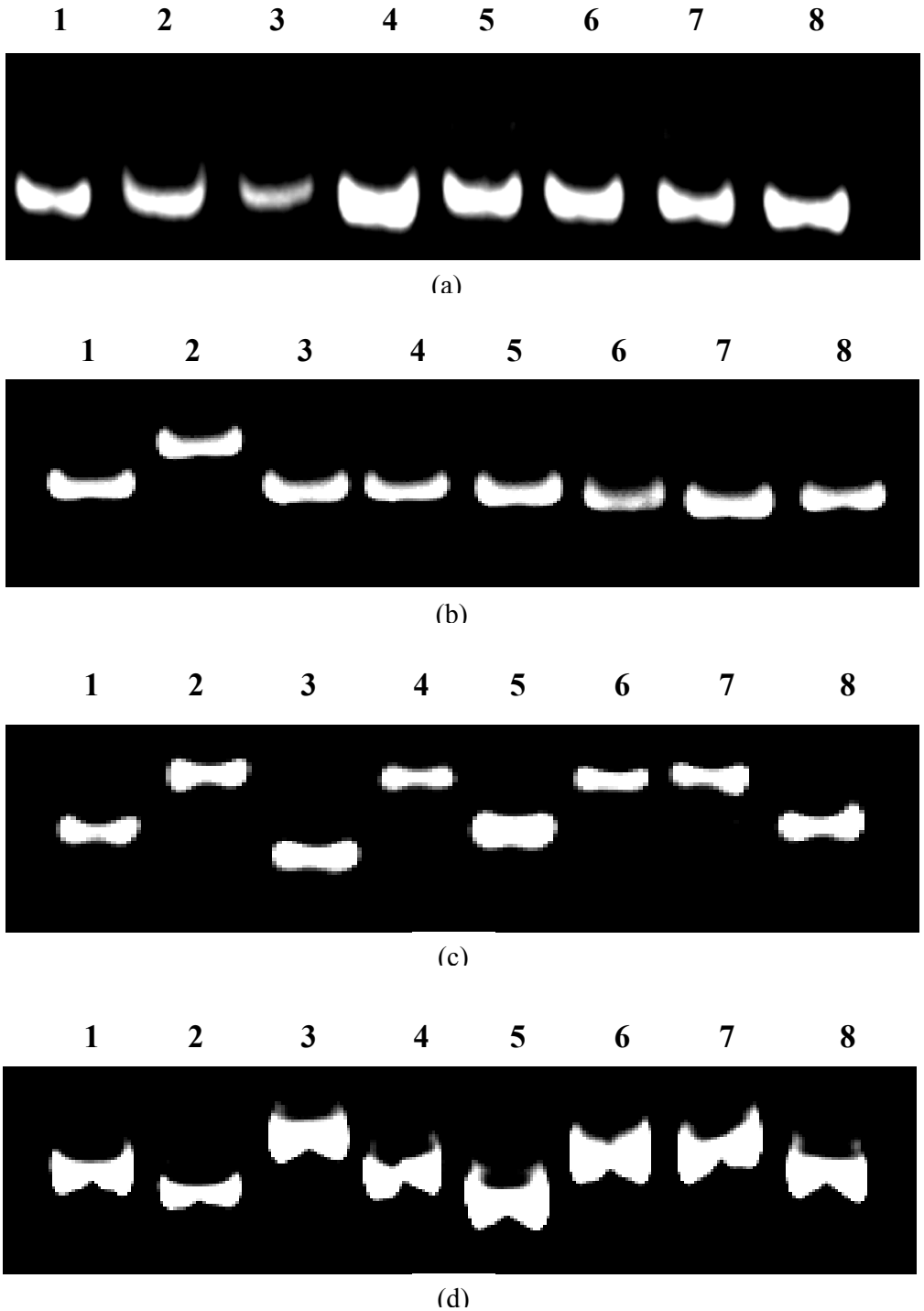


Fig. 4.7. Representative PAGE gels showing polymorphism of the SSR primers in eight rice genotypes 1. Basmati 370, 2. CSR 27, 3. Pusa 1121, 4. Jaya, 5. Swarna, 6. Pusa NPT11, 7. Pusa Basmati 1 and 8. Pusa 1342 (a- nksrssr11_26130 (monomorphic); b- nksrssr08_5844 (2 alleles); c- nksrssr09_2682 (3 alleles); d- nksrssr04_16255 (4 alleles)

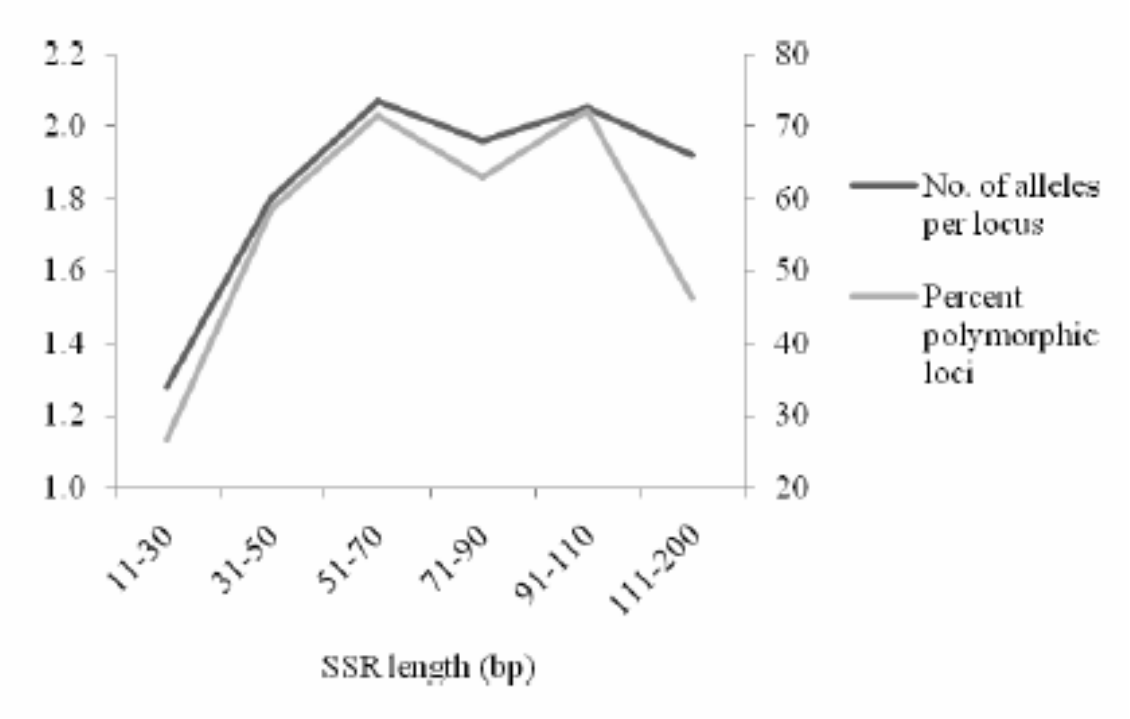


Fig. 4.8. Relationship of average number of alleles per locus and percent of polymorphic loci with the SSR lengths of 201 random SSR loci in eight rice genotypes.

Table 4.8. Average polymorphism information content (PIC) of 106 polymorphic markers with different SSR motifs in eight rice genotypes

Motif	No. of loci	Average PIC
AC/GT	2	0.525
AT/TA	30	0.467
CT/AG	6	0.365
GA/TC	4	0.470
Dinucleotide	42 (71)[a]	0.467
AAC/GTT	2	0.560
AAG/CTT	9	0.394
AAT/ATT	6	0.526
AGA/TCT	5	0.498
ATA/TAT	6	0.503
CAA/TTG	1	0.380
CAG/CTG	1	0.410
CCA/TGG	1	0.220
CCT/AGG	2	0.280
GAA/TTC	2	0.530
GAC/GTC	1	0.500
GAG/CTC	1	0.410
GTA/TAC	1	0.220
TAA/TTA	3	0.663
Trinucleotide	42 (76)	0.445
TATC/GATA	3	0.495
GCAC/GTGC	1	0.220
CTCC/GGAG	1	0.470
AGAC/GTCT	2	0.500
ACAG/CTGT	2	0.495
TCTA/TAGA	1	0.490
ATCT/AGAT	1	0.490
ATAC/GTAT	2	0.300
ATAG/CTAT	2	0.485
TATG/CATA	1	0.470
ATGC/GCAT	1	0.380
TTGT/ACAA	1	0.470
TTAT/ATAA	1	0.380
Tetranucleotide	19 (38)	0.395
CTCTC/GAGAG	1	0.380
GGCTG/CAGCC	1	0.410
TATAT/ATATA	1	0.590
Pentanucleotide	3(16)	0.460
Total	106 (201)	

[a] The figures in parenthesis indicate total number of amplified loci in that category

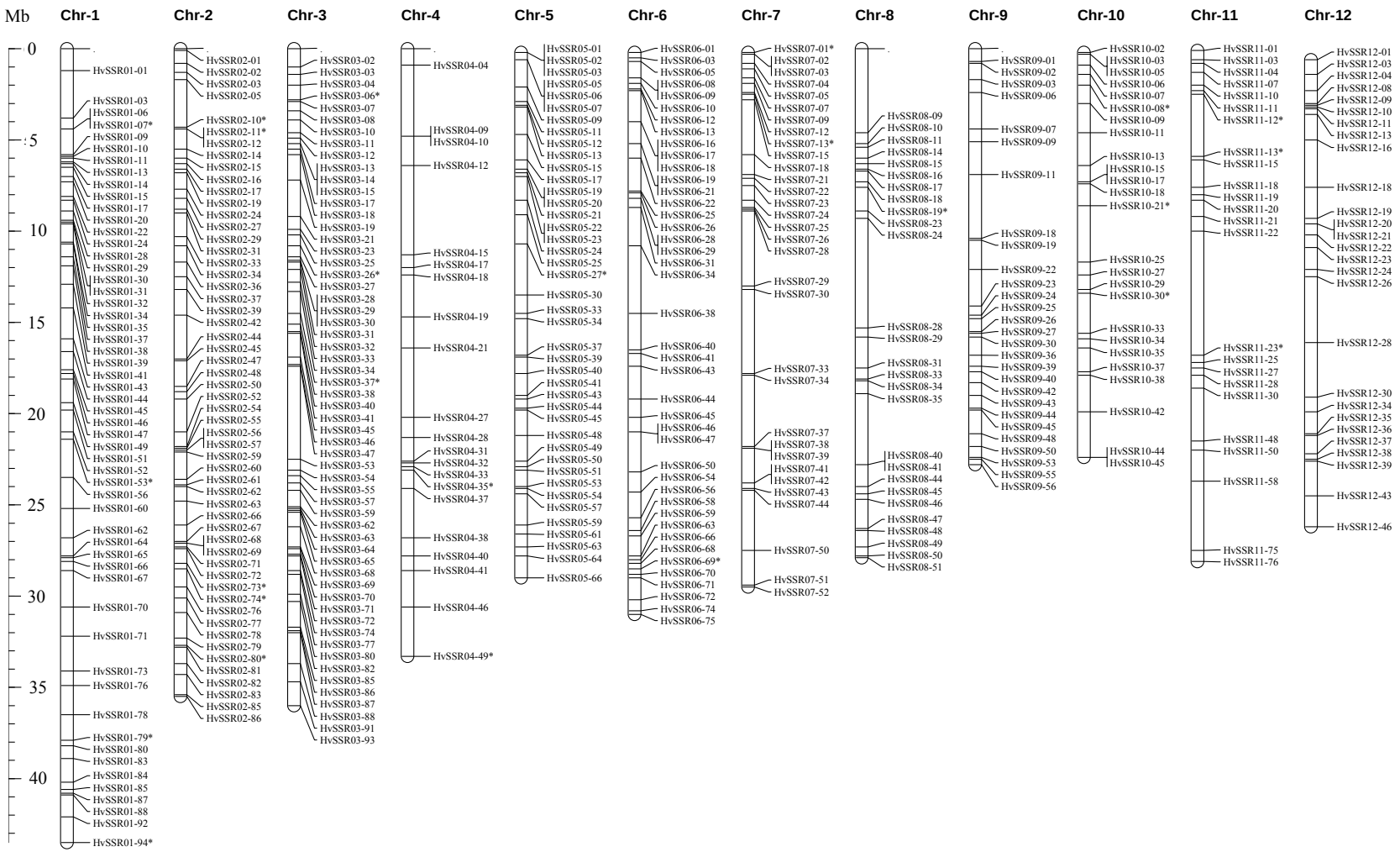


Fig. 4.9. Physical map of the 12 rice chromosomes showing location of 436 validated HvSSR markers in the TIGR rice pseudomolecules release 5

Table 4.9 Validation of higher parental polymorphism success rates with large number of HvSSR markers in four different QTL mapping populations of rice as compared to similar or higher number of random SSR markers of RM series

Sr. no.	Mapping population	RM markers			HvSSR markers		
		No. used	Polymorphic		No. used	Polymorphic	
			no.	%		no.	%
1	Pusa 1121/Pusa 1342	408	77	18.9	397	165	41.6
2	Pusa 1266/Pusa Basmati 1	650	160	24.6	405	218	53.8
3	Pusa 1266/Jaya	812	151	18.6	397	199	50.1
4	CSR 27/MI 48	471	88	18.7	382	118	30.9

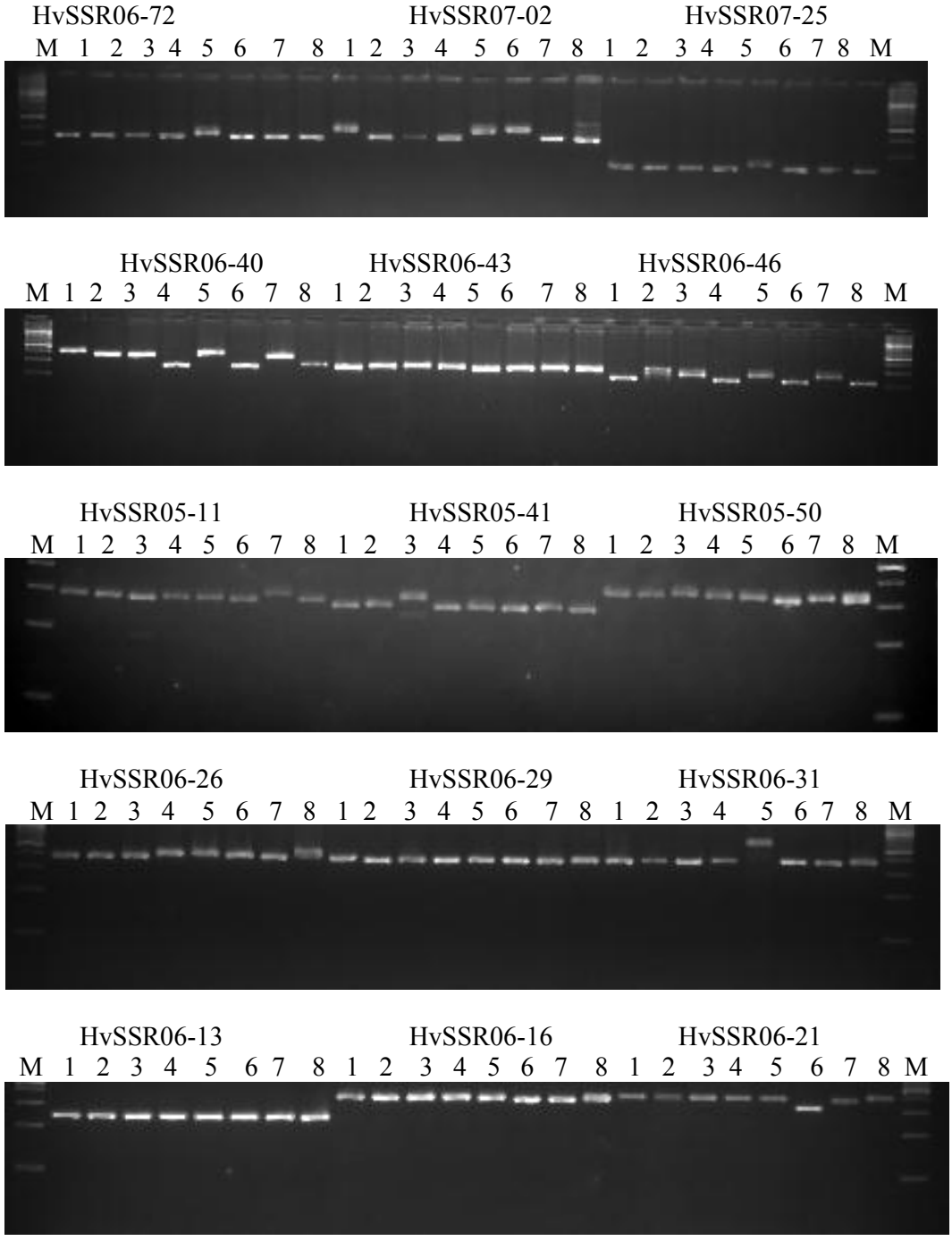


Fig. 4.10 Agarose gels with representative samples showing different levels of polymorphism in 8 rice cultivars with 15 randomly selected HvSSR markers.M: 100 bp. DNA ladder, 1 Jaya, 2 Pusa1266, 3 Pusa Basmati 1, 4 Swarna, 5 Pusa 1121 , 6 Pusa 1342, 7 CSR 27,8 MI 48

degree of success. After analyzing 408-812 RM series SSR markers for the four RIL populations, 77-160 polymorphic loci were identified at success rates of 18.7-24.6% (Table 4.9; Fig. 4.10). However, employing 382-402 HvSSR primers to the same parental lines, the bi-parental polymorphism rate ranged between 30.9-51.6%, which almost double to the rate of polymorphism obtained using random SSR markers (Table 4.9). For Pusa 1121/ Pusa 1342 the percent polymorphism increased from 18.9% to 41.6%, for Pusa 1266/Pusa basmati 1 the percent polymorphism increased from 24.6% to 53.8%, similarly for Pusa1266/Jaya the percent polymorphism just doubled from 18.6% to 50.1% and for genotypes CSR27/MI48 the percent polymorphism increased from 18.7% to 30.9%. In the initial screening 600 of these primers gave good amplification at annealing temperatures of 55-60^{o}C. Out of these, about 400 primers pairs gave good amplification with both the parents of the four QTL mapping population for which data are presented in comparison to our earlier results with large number of random SSR markers (Table 4.9). With the use of HvSSR markers now we have large number of polymorphic loci for the creation of high density framework linkage maps.

Chapter5 Discussion

To best utilize rice as a model system in plant genomics, it is necessary to integrate classical genetic resources, functional analysis and sequence information. Saturating the existing rice genetic map with highly informative, technically efficient, sequence-based markers will accelerate the tagging of gene(s) of interest and their efficient use in marker-assisted breeding. Hypervariable SSRs provide a useful source of polymorphic DNA markers for integrating genetic maps with genomic sequences and ultimately with phenotypic variation. In rice (*Oryza sativa* L.), earlier studies demonstrated that SSR markers are uniformly distributed throughout the genome and detect high level of allelic diversity in cultivated varieties and distantly related species (McCouch *et al.*, 1997). The present study was undertaken with the objective: (a) In silico analysis of the SSR loci in the rice genome for their distribution in the genic and intergenic regions. (b) Identification and experimental validation of hypervariable SSRs in the EST database of rice. (c) Identification and experimental validation of hypervariable SSR on the whole rice genome.

In the present study the abundance and density of SSRs across the rice genome has been analysed. The data on individual rice chromosome was split into coding, intergenic, intronic and UTR regions. The overall abundance and density of SSRs in each chromosome was found to be comparable (Table 4.1; Fig. 4.1). The data have been analyzed for di-, tri-, tetra-, penta- and hexa-nucleotides repeats. Again, the density of each class of repeat is comparable across various genomic regions**.** However, different repeat motifs often show tremendous variation in density in different genomic regions, sometimes even in a chromosome-specific manner. For searching the SSR by the MISA script, SSRs were considered to contain motif that are between 2 to 6 nucleotide of size and more than 10bp in length. A total of 70,274 di-, tri-, tetra-, penta- and hexa-nucleotide SSRs were generated. However, the mononucleotide repeat were excluded for further study as they are not useful for polymorphism survey (J. Serapion *et al.*, 2004). The 70,274 SSRs generated in the study are higher in number to the total number of markers reported earlier (IRGSP, 2005). As

expected the total frequency of SSRs depends on the size of chromosome e.g. chromosome 1 being bigger in size has the highest frequency and chromosome 10 has the lowest repeat occurrence (Fig. 4.1). There was an average of 189 SSRs per Mbp (Table 4.1) and the density of SSRs in the study is higher to reported earlier (IRGSP, 2005). The density of each class of repeat motif was observed to be comparable across the 12 chromosomes with the highest density of markers occurring on chromosome 6 which was in contrast to the report on densely populated chromosome 3 (IRGSP, 2005), whereas lowest density was observed on chromosome 4 that was found to be in agreement with the similar report.

Relative abundance of (AT)n, (TA)n and (CT)n/(GA)n repeats to GC can be explained on the basis of A/T richness and the relative ease of strand separation compared to G/C tracts (R Gur-arieb *et al.*, 2000) The $(GC)_n$ SSRs that are abundantly distributed in genic regions may explain the stability of gene structure (Fig. 4.2a and b). Chromatin remodeling and gene silencing via cytosine methylation has been reported to occur more frequently for GC rich SSRs (Razin, 1998). The GT repeats were found to be less frequent in rice genome as a whole and in intergenic regions, the frequency of this repeat was lower in introns and UTRs though not lowest. The studies on GT tandem repeats in human chromosome 22 revealed that they are recombination hotspots (Majewski *et al.*, 2000)

The abundance of SSRs in tri-nucleotide repeats CGC, GCG, GCC, GGC and CGG in all the genomic regions mainly the coding region in the study is a specific feature of monocot genomes, which may be due to their increased G+C content (Morgante *et al.*, 2002). The higher frequency of trinucleotide SSRs in coding region are in consistency with previous observation about differences in abundance of SSR classes (Cardle et al,.2000). The dominance of trinucleotide SSRs over di-, tetra-, and pentanucleotide SSRs may be explained on the basis of the suppression of non-trinucleotide SSRs in coding region due to the risk of frameshift mutation that may occur when those SSRs alternate in size of one unit (Metzger *et al.*, 2000). This observation is in agreement with the findings in maize and human, where GC rich poly-trinucleotide repeats represented the large proportion of SSRs (Jurka and Petiyagoda 1995; Chen *et al.*, 1996). On the

contrary, in the genomes of *Arabidopsis* and yeast (*Saccharomyces cerevisiae*) the majority of trinucleotide repeats found in coding regions are AT-rich (Cardle *et al.*, 2000; Young *et al.*, 2000). It is intriguing that these trinucleotide repeats were found preferentially in yeast genes involved in the regulation of transcription, signal transduction, and cell growth and division, but only rarely in genes controlling common metabolic functions such as glycolysis and respiration. It follows that these associations between different trinucleotide repeats and certain types of genes may be genome- specific. This knowledge, when used in the context of known frequencies of particular trinucleotide repeats in specific functional categories of genes, might be useful in characterizing novel trinucleotide repeats -containing genes during genome-scale annotation.

The tetranucleotide repeats ATAG, TATG, TATC and ATCT in whole genome and intergenic regions are rich in (A+T %) similar to dinucleotide repeats. The repeats in the various classes like AAT, AAC and AAG among trinucleotides, AAAT, AAAC, and AAAG in the case of tetranucleotides, AAAAG, GAAAA, TCTTT, AAAAT and AAAAC in the case of pentanucleotides, AAAAAG and AAAAAG among hexanucleotides repeats in the whole genome is possible as during SSR evolution the poly (A) stretches present in the genome might have mutated to produce the A-rich repeats. It is also possible that the abundance of repeats is influenced by their secondary structures and the effect on DNA replication. If a repeat sequence is selected during evolution for transcriptional regulation or is a target of a binding protein for one or more nuclear processes (such as chromatin organization, DNA replication, transcription, recombination), its abundance and distribution is expected to be controlled whereas, in the genomic region the frequency of particular repeat type was variable. In case of hexanucleotide repeats the occurrence of different type of SSRs varied on all the chromosomes for all the genomic regions (Subramanian *et al.*, 2003).

Average density for non coding regions showed the highest density range for different type of SSRs in intergenic region followed by coding, intronic and UTRs regions (Table 4.4). The comparative analysis of these regions for average densities depicted almost similar trend for coding and UTR regions across the twelve chromosomes. The intergenic and intronic regions showed a different trend

for the average density of SSRs on the twelve chromosomes. The highest and lowest average density of SSRs for intergenic regions showed the similar results as for whole genome with chromosome 6 being densely populated and chromosome 4 the least but for other chromosomes it was comparable.

The abundance of the SSR in the repeat range of 5X-15X may arise because of new mutations, as they were the most common events which generated new simple sequence repeats consisting of two repeat units or it may be the duplication of adjacent sequences. Similar results were obtained from an analysis of pseudogenes by (Zhu *et al.* 2000).

It was observed that occurrence of SSRs was highest in intergenic regions followed by introns, UTRs and coding regions for di-, tetra- and penta- nucleotide repeats. The results on more frequent occurrence of trinucleotide repeats were in coding regions is in agreement to (Toth *et al.*, 2000 and Metzgar *et al.*, 2000). Similarly, hexa-nucleotide repeats were found to be higher in number in coding region as compared to introns and UTRs. The occurrence of SSRs other than repeats of triplets or of other multiples of three in a coding region will completely impair the protein function by causing frameshift mutations.

To find out the relationship between SSR repeat length with polymorphism, inter varietal 201 primers were developed from 68 random intervals in the 12 rice chromosomes (Table 4.6 and Annexure VI). The primers were able to resolve 50% of genotypes, there was a significant positive correlation between SSR length and average number of alleles per locus for both genic and intergenic SSR. However, the polymorphism level increased linearly only in the SSR length range of 11-70 bp, it decreased slightly in the 71-90 bp range and then remained more or less constant for higher repeat lengths. The polymorphism level peaked in the SSR length range of 51-70 bp and there was no specific association of any repeat motifs with the level of allelic polymorphism. The proportion of polymorphic loci also peaked in the 51-70 bp SSR length group, but it showed a much sharper decline for SSR lengths higher than 111 bp. However, the PIC values remained stable after reaching a peak of 0.49 in 51-70 bp SSR length group. The findings could be explained by the rate of expansion mutations and contraction mutations, earlier reports suggests that the rate of expansion mutations

is constant for all alleles and the rate of contraction mutations increases exponentially over repeat length. And at a certain critical repeat length, the rates of expansion and contraction mutations are equal. Mutations in alleles shorter than the critical length are biased towards expansion, whereas mutations in longer alleles favor contraction, there is a strong bias toward expansion or contraction for a particular allele depending on its length. For young SSRs that evolve from shorter allele length and have not reached equilibrium, however, there will be an overall bias toward expansion. So 51-70bp can be considered as a critical length at which the rate of expansion and contraction mutation is equal and hence the chances of getting polymorphism is high (Xu, 2000).

It was observed that the markers developed from chromosome 3 showed maximum polymorphism information content whereas markers developed from chromosome 11 showed the least PIC values for polymorphism studies among the eight genotypes. It is evident with the markers of high PIC values that distribution of alleles towards the optimum level (p=q=0.5) is the utmost requirement to distinguish between the genotypes at a high level. The present study has resolved up to the level of 2 genotypes (25%) of the genotypes studied. Interestingly the markers developed from chromosome 2 and 4 were found to be highly polymorphic whereas markers from chromosome 11 showed the least polymorphism. To further reconfirm that highest polymorphism can be attained in the class of 51-70bp. (Annexure VII) the primers were developed from chromosome 11 as it has higher number of disease resistant like and defense response genes than any other rice chromosome (Rice chromosome 11 and 12 sequencing consortia, 2005) and chromosome shows sequence synteny with group 4 chromosomes of wheat (Singh *et al.*, 2007). Interestingly in the range of 51-70bp. the overall PIC value for chromosome 11 increased considerably.

The polymorphism level among di- and trinucleotide SSRs was found to be different. Markers based on dinucleotide repeats were more polymorphic (59.2%) than trinucleotide SSR markers (55.3%) in the 8 rice genotypes tested. These results are in agreement with the results reported by (Nicot *et al.* 2004) in wheat, in their experiments dinucleotide EST-SSRs were observed to be more polymorphic than trinucleotide repeats. Similarly the analysis of repeat motif revealed AC/GT, AAC/GTT, TAA/TTA and AAT/ATT to be highly polymorphic

which were found to be rich in (A+T%). The high polymorphism detected by AT-based markers may suggest that non transcribed regions of the genes easily accumulate mutations (Gadaleta *et al.* 2007). Though for motif TATAT/ATATA a penta-nucleotide (Table 4.8) PIC value was observed to be on the higher side but this value may not be considered significant as polymorphism was observed with the least number of polymorphic primers. The results suggest analyzing di- and tri- repeat motifs for polymorphism studies to access the variability among genotypes.

It was observed that coding SSR (8.9 %) of the total SSRs amplified showed hypervariability for length in the transcribed region of the genes might have a role in gene expression or function as reported by (Ayers *et al.* 1997; Bao *et al.* 2002) for variation in the number of GA or CT repeats in the 5' UTR of the waxy gene to be correlated with amylose content. Thus, the mechanisms found in human or animal systems might also have a role in generating phenotypic diversity in plant species however, it remains to be seen whether any unusual phenotypic variation might be associated with the length of SSRs in coding regions as was reported for several diseases in humans (Li *et al.* 2004; Cummings and Zoghbi 2000). SSR markers that are physically associated with coding regions of the genome and which enhance the role of genetic markers in germplasm evaluation by enabling the variation in expressed genes to be assayed (Peng *et al.* 2005). Polymorphic SSR markers directly sample variations in transcribed regions of the genome, and this may enhance their value in rice genetic diversity assessment and marker-assisted selection. The results of the present investigation indicate that some SSR types with repeat length of 51-70bp are able to discriminate among cultivars much more than others, a finding which may help researchers choose the appropriate markers to be used in breeding programs.

Chapter 6 Summary

The investigations were carried out for the identification of hypervariable regions in the rice genome. The study was conducted with respect to density, abundance and distribution of SSR (simple sequence repeats) of different lengths (bp) and different repeat motifs in the coding as well as the non-coding regions of the genome, correlation between repeat length and polymorphism level and validation of a set of potential hypervariable SSR markers. The results obtained hitherto are summarized as follows:

1. A total of 70,274 SSR loci were identified in the complete sequence of the rice genome. Di-nucleotide SSR were the most abundant making up 52.82% of all SSR loci ,followed by tri-nucleotide SSRs with 42.25% coverage and tetra-nucleotide SSRs with 3.66% coverage. The abundance of penta- and hexa-nucleotide SSRs were comparatively lower, both covering less than 1% of total SSRs in the rice genome. The mono-nucleotide SSRs were not considered for the present study because of their low polymorphism levels and unreliable assays. Chromosome 1 has the highest and chromosome 10 the lowest number of SSR loci. Most of the SSR loci contained 5-15X repeats of their respective SSR motifs. As a rule the shorter SSR were more abundant than the longer ones.
2. The patterns of SSR frequencies were also analyzed separately for non coding regions where at 5X-15X repeat range the SSR frequency was observed to be highest for intergenic > intronic > UTRs. For trinucleotide repeats, in 5X-15X repeat class the frequency for total number of SSRs was comparable between introns and UTRs i.e. intergenic> UTR (untranslated genic regions) > introns. For coding regions (exon sequences) the SSR frequency showed the same pattern for di-, tri-, tetra-, penta- and hexa-nucleotide with the maximum frequency occurring in the range of 5X-15X.
3. The frequency of SSR containing di-nucleotide repeat motifs AT, AC/GT, TG/CA, TA was lower in coding regions as compared to non-coding regions (intergenic, introns and UTRs) whereas, the dinucleotide repeats

GC, CT/AG, CG and GA/TC were at higher frequency in the coding regions. In the UTR region, CT/AG and GA/TC repeats were abundant, whereas, at the whole genome level and introns AT, TA, CT/AG and GA/TC repeats were most abundant.

4. The SSR containing trinucleotide repeats CGC, GCG, GCC, GGC and CGG were most abundant in all the genomic regions. Interestingly, the frequency of tri-nucleotide repeats was higher in the coding regions as compared to the non-coding regions (intergenic, introns and UTRs). Within the non-coding region, the frequency of trinucleotide SSR was the highest in UTRs and the lowest in the introns.
5. The occurrence of SSR with tetranucleotide repeats in the whole genome showed bias towards ATAG, TATG, TATC and ATCT repeat motifs. The most abundant SSR with pentanucleotide motif at the whole genome level were AAAAG, GAAAA and TCTTT. In case of hexanucleotide repeats the occurrence of different type of SSRs varied on all the chromosomes for all the genomic regions.
6. The density of each class of SSRs in the genome was analyzed and it was observed that the overall SSR density was similar across the 12 rice chromosomes with an average of 190 SSR per Mbp. The highest density of 197 SSR/ Mbp was observed for chromosome 6 and the lowest density of 167/ Mbp was for chromosome 4.
7. Average density (SSRs per Mbp region) was comparable across the twelve rice chromosomes for intergenic, intronic UTR and coding regions. For intergenic regions chromosome 6 showed the highest density (109.30 and the lowest density was observed on chromosome 4 (88.44). For intronic regions, chromosome 1 showed the highest density (32.17) and the lowest density was observed on chromosome 7 (24.52). The average density for UTRs (5' and 3') was the highest on chromosome 3 (32.71) and the lowest density was observed on chromosome 11 (11.49). For coding regions the highest average density was observed on chromosome 3 (62.11) and the lowest on chromosome 11 (25). The comparative analysis of these regions for average densities depicted almost similar trend for coding and UTR regions across the twelve chromosomes. The intergenic and intronic

regions showed a different trend for the average density of SSRs on the twelve chromosomes.

8. The SSR markers developed in this study and tested on 8 rice genotypes (Basmati 370, CSR27, Pusa1121, Jaya, Swarna, CB11, Pusa Basmati1 and Pusa 1342) amplified 228 alleles, with a range of 1 to 4 allels and an average of 1.77 alleles per locus. The graphic representation of relationship between the SSR length (bp) and average number of alleles per locus showed that there was a positive correlation between SSR length and average number of alleles per locus but linearity of this relationship was limited to the SSR length range of 10-70 bp. The highest level of polymorphism was in the SSR length range of 51-70 bp, beyond which there was a decline and stabilization of polymorphism in SSRs longer than 70 bp.

9. The average polymorphism information content (PIC) values for the SSR loci from different chromosomes was fairly uniform and it ranged from 0.53 (chromosome 2 and 4) to 0.34 (chromosome 11).

10. On the basis of SSR repeat motif, the loci with SSR motifs AC/GT (0.525), TAA/TTA (0.663), AAC/ GTT (0.560) and AAT/ATT (0.526) showed the highest PIC values. The average PIC values calculated for different repeat types showed highest PIC for the dinucleotide SSRs (0.467) followed by pentanucleotide repeats (0.460) and trinucleotide SSRs (0.445). The minimum PIC was observed for tetranucleotide repeats (0.395). However, the SSR length was the overriding factor over the SSR motif that determined the level of polymorphism.

11. To further validate this observation primers were designed for additional SSR loci with repeat lengths in the range of of 51-70 bp from the genomic sequence of the long arm of rice chromosome 11 sequenced by India due to its richness for disease resistance genes. Out of total 45 loci tested, 40 gave good amplification and generated 82 alleles with an average of 2.5 alleles per locus. The additional 40 primers for the chromosome 11 showed an average PIC value of 0.36 which was considerably higher than 0.31 observed earlier with random SSR loci including repeat lengths even higher than 70 bp. This suggests that for maximum polymorphism, SSR repeat length of 51-70 bp is the most appropriate.

12. Based on the above validation results, all the 832 potential hypervariable SSRs with the repeat lengths range of 51-70bp were mined from the genomic sequence of 12 rice chromosomes and of these 432 used for parental polymorphism survey in four different RIL mapping populations derived from biparental crosses. As compared to the random SSR markers of RM series the HvSSR loci gave almost double the level of polymorphism which further validated their usefulness for genetic analysis in rice.

Future scope of the work

Simple sequence repeats (SSR) are the DNA markers of choice for plant genetic analysis due to their abundance, high polymorphism and reproducible assays. It has been reported in rice that intervarietal polymorphism is positively associated with the SSR length but there is no systematic genome wide study to validate these assertions

Polymorphic SSR markers directly sample variations in transcribed regions of the genome, and this may enhance their value in rice genetic diversity assessment and marker-assisted selection. The results of the present investigation indicate that some SSR types with repeat length of 51-70bp are able to discriminate among cultivars much more than others, a finding which may help researchers choose the appropriate markers to be used in breeding programmes. Such panels of SSR markers will help find genome wide polymorphism between parental lines of QTL mapping populations and for fingerprinting of rice germplasm.

References

Akagi H, Yokozeki Y, Inagaki A, Fujimara (1996) Microsatllite DNA markers for rice chromosomes. Theor Appl Genet 93: 1071-1077

Akagi H, Yokozeki Y, Inagaki A, Fujimara (1997) Highly polymorphic microstallites of rice consist of AT repeats, and a classification of closely related cultivars with these microsatellite loci. Theor Appl Genet 94 (1): 61-67

Akkaya MS, Bhagawat AA, Cregan PB (1992) Length polymorphism of simple sequence repeat in soybean. Genetics 132: 1131-1139

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ(1990) Basic local alignment search tool.J Mol Biol 215**:**403-410

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94: 773–781

Banks MA, Rashbrook VK, Calavetta MJ, Dean CA, Hedgecock D (2000) Analysis of microsatellite DNA resolves genetic structure and diversity of chinook salmon (*Oncorhynchus tshawytscha*) in California's Central Valley. Can J Fish Aquat Sci 57:915-927

Bao S, Corke H, Sun M (2002) Microsatellites in starch-synthesizing genes in relation to starch physicochemical properties in waxy rice (*Oryza sativa* L.). Theor Appl Genet 105:898-905

Bartley D, Gall GAE, Bentley B, Brodziak J, Gomulkiewicz R, Mangel M (1992) Geographic variation in population genetic structure of chinook salmon from California and Oregon. Fish Bull 90:77-100

Becker J, Heun M (1995) Mapping of digested and undigested random amplified microsatellite polymorphism in barley. Genome 38: 991-998

Bhattramakki D, Dong JM, Chhabra AK, Hart GE(2000) An integrated SSR and RFLP linkage map of Sorghum bicolor (L.) Moench Genome 43: 988-1002

Botstein D, White LR, Sholnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am J Hum Genet 32 : 314-331

BOXSHADE version 3.21 Hofmann K Baron M http://www.ch.embnet.org

Brown SM, Szewc-McFaddan AK, Kresovich S (1996) Development and applications of simple sequence repeat (SSR) loci for plant genome analysis. *In*: Methods of genome analysis in plants, (ed) Jauhar Publ.; CRC Press Inc, Florida, USA.

Caetano-Anolles G, Gresshoff PM (1996) Generation of sequence signatures from DNA amplification fingerprints with mini-hairpin and microsatellite primers. Biotechniques 20: 1044-1056

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156: 847–854

Causse MA, Fulton TM, Cho YG, Ahn SN, Chunwongse J, Wu K, Xiao J, Yu Z, Ronald PC, Harrington SE, Second G, McCouch SR, Tanksley SD (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138: 1251-1274

Chao F (1927) Linkage studies in rice. Genetics 13: 133-169

Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice. (*Oryza sativa* L.). Theor Appl Genet 95: 553-567

Chin ECL, Senior ML, Shu H, Smith, JSC (1996) Maizesimple repetitive DNA sequences: abundance and allele variation.Genome 39: 866–873

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Cordeiro G, Casu R, McIntyre C, Manners J Henry RJ (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160: 1115-1123

Cordeiro GM, Christopher MJ, Henry RJ, Reinke RF (2002) Identification of microsatellite markers for fragrance in rice by analysis of rice genome sequence. Molecular Breeding 9: 245-250

Culver M, Menotti-Raymond MA, O'Brien SJ (2001) Patterns of size homoplasy at 10 microsatellite loci in pumas (*Puma concolor*) Mol Biol Evol 18:1151-1156

Cummings CJ and Zoghbi HY (2000) Trinucleotide repeats:mechanisms and pathophysiology. Annu Rev Genomics Hum Genet 1: 281–328

Dresselhaus T, Heuer S, Lorz H (1999) Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. Mol Gen Genet 261: 416–427

Estoup A, Tailliez C, Cornuet JM, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). Mol Biol Evol 12:1074-1084

Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica 119:39-43

Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S,Tomita M (2003) A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Lett 554 : 17–22

Gadaleta A, Mangini G, Mule G, Blanco A (2007) Characterization of dinucleotide and trinucleotide EST-derived microsatellites in the wheat genome. Euphytica 153: 73-85

Garza JC, Freimer NB (1996) Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Res 6:211-217

Goldstein DB, Linares AR, Cavallis-forza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463-471

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. J Hered 88: 335–342

Gramene, a tool for grass genomics (2002) http//: www.gramene.org

Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. J Mol Evol 44:336-340

Gupta PK Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113: 163-185

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996). Microsatellites in plants: a new class of molecular markers. *Current Science* 70 (1): 45-54

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113: 163-185

Gupta PK, Varshney RK, Sharma PC Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breeding 118: 369-390

Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y (2000) Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. Genome Res 10:62-71

Hamada H, Petrino MG, Kakunga T (1982) A novel repeated element with Z-DNA-forming potentials is widely found in evolutionary distinct eukaryotic genomes. Proc Natl Acad Sci USA 79: 6465-6469

Harr B, Schlötterer C (2000) Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide under representation. Genetics 155:1213-1220

Harushima Y, Yano M, Shomura A, Sato M, Shimano T, Kuboki Y, Yamamoto T Shao YL, Antonio BA, Parco A, Kajiya H, Huang N, Yamamoto K, Nagamura

Y, Kurata N, Khush GS, Sasaki T (1998) A high-density rice linkage map with 2275 markers using asingle F_2 population. Genetics 148: 479-494

Huang QY, Xu FH, Shen H, Deng HY, Liu YJ, Liu YZ, Li JL, Recker RR, Deng HW (2002) Mutation patterns at dinuleotide microsatellite loci in humans. Am J Hum Genet 70:625-634

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436: 793-799

Jurka J, Pethiyagoda C (1995) Simple repetitive DNA sequences from primates: Compilation and analysis. J Mol Evol 40: 120–126

Khush GS, Singh RJ, Sur SC, Librojo AL (1984) Primary trisomic of rice: origin, morphology cytology and use in linkage mapping. Genetics 107: 141-163

Kim TJ, Parker KM, Hedrick PW (1999) Major histocompatibility complex differentiation in Sacramento River chinook salmon. Genetics 151:1115-1122

Kimura M, Crow JF (1964) The numbers of alleles that can be maintained in a finite population. Genetics 49:725-738

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci USA 75:2868-2872

Kinoshita T (1984-1994) Report of the committee on gene symbolization and linkage map. Rice Gnet Newsl 1-1

Knapp SJ (1991) Using molecular markers to map multiple quantitative triat loci: models for backcross, recombinant inbred and doubled haploid progeny. Theor Appl Genet 81: 332-338

Korzun V, Borner A, Worland AJ, Law CN, Roder MS (1997) Application of microsatellite markers to distinguish inter-varietal chromosome substitution lines of wheat. Euphytica 95: 149-155

Kota R, Varshney RK, Thiel T, Dehmer KJ Graner A (2001) Generation and comparison of EST-derived SSR and SNP markers in barley *(Hordeum vulgare* L.). Hereditas 135: 141-151

Lawson MJ, Zhang L (2006) Distinct pattern of SSR distribution in the Arabdopsis thaliana and rice genomes. Genome Biology doi:10.1186/gb-2006-7-2-r14

Levinson G, Gutman GA (1987*b)* Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203-221

Li CD, Rossnagel BG, Scoles GJ (2000) The development of oats microsatellite markers and their use in identifying relationships among Avena species and oats cultivars. Theor Appl Genet 101: 1259- 1268

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21: 991–1007

Litt M, Lutty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44: 397-401

Lopez MT, Toojinda T, Vanavichit A, Tragoonrung S, (2003) Microsatellite markers flanking the tms2 gene facilitated tropical TGMS rice line development. Crop Science 43: 2267-2271

Majewski J, Ott J (2000) GT repeats are associated with recombination on human chromosome 22. Genome Res 10:1108-1114

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair MW (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35:89–99

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76: 815-829

McCouch SR., Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). DNA Res 9:199–207

Metzgar D, Bytof J, Wills C (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res 10: 72-80

Moore SS, Sargeant LL, King TG, Matick JS, Georges M, Hertzel DJS (1991) Conservation of dinucleotide microsatellites among mammalian genomes allows use of heterologus PCR primer pairs in closely related species. Genomics 10: 654-660

Morgante M, Hanafey M, Powell W (2002) Microsatellites are prefentially with nonrepetitive DNA in plant genomes. Nature Genet. (30) 194-200

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucl Acids Res 8:4321-4325

Nagao S, Takahashi T (1963) Trial construction of twelve linkage groups in Japanese rice. Genetical studies on rice plant, XXVII J Fac Agric Hokkaido Univ 53: 72-130

National research Centre on plant biotechnology (2006) IARI, New Delhi, India http://www.nrcpb.org

NCBI: Basic Local Alignment Search Tool (BLAST) www.ncbi.nlm.nih.gov/BLAST

Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Notredame C, Higgins D, Heringa J (2000) T-Coffee: A novel method for multiple sequence alignments. Journal of Molecular Biology 302: 205-217

Peng JH, Nora L, Capitan V (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integ Genomics 5: 80–96

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingley S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2: 225-238

Primmer CR,Ellegren H, Saino N, Moller AP (1996) Directional evolution in germline microsatellite mutations. Nat Genet 13:391-393

Pupko T, Graur D (1999) Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: Role of length and number of repeated units. J Mol Evol 48**:**313-316

Queney G, Ferrand N, Weiss S, Mougel F, Monnerot M (2001) Stationary distributions of microsatellite loci between divergent population groups of the European rabbit (*Oryctolagus cuniculus*). Mol Biol Evol 18:2169-2178

Ramsay L, Macaulay M, Ivanissevich DS, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh RA (2000) Simple sequence repeat-based linkage map of barley. Genetics 156 :1997-2005

Razin, A (1998) CpG methylation, chromatin structure and gene silencing—a three-way connection. EMBO J 17: 4905-4908.

Rice Chromosome 11 and 12 Sequencing Consortia (2005) The sequence of rice chr11 and 12, rich in disease resistance genes and recent gene duplication. BMC Biology 3:20

Rice Genome Research Program (2000) http://rgp.dna.affrc.go.jp/IRGSP

Robinson AJ, Love CG, Batley J, Barker G and Edwards D (2004). Simple Sequence Repeat Marker Loci Discovery using SSR Primer. Bioinformatics 9: 1475 – 1476

Roder MS, Korzn V, Wendehake K, Plaschke J, Tixier M, Leroy P, Ganal MA (1998) Microsatellite map of wheat. Genetics 149 :2007- 2023

Roder MS, Plaschke J, Konig S U, Borner A, Sorrells ME, Tanskley SD, Ganal M W (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246: 327-333

Rungis D Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (Picea spp.) from expressed sequence tags. Theor. Appl. Genet 109: 1283–1294

Saal B, Wricke G (1999) Development of simple sequence repeat markers in rye (Secale cereale L.). Genome 42 : 964-972

Saito A, Yano M, Kishimoto N, Nakagahra M, Yoshimura A (1991) Linkage map of restriction fragment length polymorphism loci in rice. Jpn J Breed 41: 665-670

Saji S, Umehara Y, Antonio BA, Yamane H, Tanoue H, Baba T, Aoki H, Ishige N, Wu K, Koike K, Matsumoto T, Sasaki T (2001) A physical map with Yeast Artificial(YAC) clones covering 63% of 12 rice chromosomes. Genome 44: 32-37

Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211-215

Scott KD (2001) Microsatellite derived from ESTs, and their comparison with those derived by other methods. In: Plant Genotyping: The DNA Fingerprinting of Plants, (Henry, R. J., Ed.), CABI Publishing, Oxon, U. K., 225-237

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000). Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100: 723–726

Serapion J, Kucuktas H, feng J, liu Z (2004) Bioinformatic mining of type I microsatellite from Expressed sequence tags of Channel catfish(*Ictalurus punctatus*. Mar Biotechnol. 6: 364-377

Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS (1996) Centromere mapping and orientation of the molecular linkage map of rice (Oryza sativa L.). Proc Natl Acad Sci USA 93: 6163-6168

Singh NK, Dalal V, Singh BK, Chitra G, Singh A, Ghazi IA, Yadav M, Pandit A, Dixit R, Singh PK**,** Singh H, Koundal KR, Gaikwad K, Mohapatra T, Sharma TR (2007). Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Functional & Integrative Genomics, DOI: 10.1007/s10142-006-0033-4

Singh NK, Raghuvanshi S, Srivastava SK, Gaur A, Pal AK, Dalal V, Singh A, Ghazi IA, Bhargav A, Yadav M, Dixit A, Batra K, Gaikwad K, Mohapatra T, Sharma TR Mohanty A, Bharti AK, Kapur A, Gupta V, Kumar D, Ravi V, Khurana P, Sharma S, McCombie, Messings J, Wing R, Sasaki T, Khurana P, Khurana JP, Tyagi A(2004). Single-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes. Functional & Integrative Genomics 4: 102-117

Smith S (1994) Recent advances in the use of PCR based technology for DNA fingerprinting in plants. Conservation of plant genes II: Utilization of ancient and modern DNA. *Missouri Botanical Garden*, St. Louis , MO. pp 101-112

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Desestabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274-276

Subramanian S, Mishra RK, Singh L (2002) Genome-wide analysis of microsatellite repeats in humans their abundance and density in specific genomic regions. Genome Biology 2003 4(2):R13

Swathi SP, Gupta VS, Agarwal RK, Ranjekar PK, Brar DS (2000) Genetic diversity and phylogenetic relationship as revealed by inter-simple sequence repeat (ISSR) polymorphismin the genus *Oryza*. Theor Appl Genet 100: 1311-1320

Tautz D (1989) Hypervariability of simple sequence as a general source of polymorphic DNA markers. Nucleic Acids Res 17: 6463-6471

Tautz D Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acid Res 12: 4127-4138

Tautz D, Trick M, Dover G. (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652-656

Temnykh S, DeClerck G, Lukashova A., Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (Oryza sativa L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11: 1441-1452

Temnykh S, Park WD, Ayers N, Cartinhour S, Cho YG, Hauck N, McCouch SR (1998) A microsatellite map for rice: Recent progress and prospectives. In: Plant and Animal Genome VI Conference. January 18-22. San Diego, CA.

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch, SR (2000) Mapping and genome organization of microsatellite sequences in rice (Oryza sativa L.). Theor Appl Genet 100: 697-712

The Institute for Genomic Research Rice Pseudomolecules (2007) http//:www.tigr.org

Thiel T, Identifizierung, Kartierung Chjarakterisierung (2001) cDNA basierter Mikrosatelliten-Marker zur Diversitätsanalyse bei gerste (Hordeum vulgare L.). Diploma thesis. Technische Universität Dresden 99

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106: 411–422

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10:967-981

Tsunematsu H Yoshimura A, Harushima Y, Nagamura Y Kurata N, Yano M, Sasaki T, Iwata N (1996) RFLP framework map using recombinant inbred lines in rice. Breeding Sci 46: 279-284.

Vanshanudhan, Rice gene database (2006) Rice Genome Project, National research Centre on plant biotechnology, IARI, New Delhi, India

Varshney RK, Kumar A, Balyan HS, Roy JK, Prasad M, Gupta PK (2000) Characterization of microsatellites and development of chromosome specific STMS markers in bread wheat. Plant Mol Biol Rep 18: 1- 12

Wang Z, Weber JL, Zhong G, and Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88: 1-6

Ware D, Jaiswal P, Ni J, Pan X, Chang K, Clark K, Teytelman L, Schmidt S, Zhao W, Cartinhour S, McCouch S, Stein L (2002) Gramene: a resource for comparative grass genomics. Nucleic Acid Res 30(1): 103-105.

Weber JL, (1990) Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. Genomics **7:** 524-530

Weber JL, May PE (1989) Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. Am J Hum Genet 44: 338-396

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123-1128

White PT (1994) Rice: The essential harvest. Natl Geogr 185: 48-79

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769-779

Wu KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241: 225-235.

Xu X, Peng M, Fang Z, Xu X (2000) The direction of microsatellite mutations is dependent upon allele length. Nature Genet 24:396-399

Yamamoto K, Sasaki Y (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35 (1-2): 135-144

Yashitola J Thirumurugan T, Sundaram RM, Naseerulla, MK, Ramesha MS, Sarma NP, Sonti RV (2002) Assessment of purity of rice hybrids using microsatllite and STS markers. Crop Sci 42: 1369-1373

Young ET, Sloan JS, Van Riper K (2000) Trinucleotide repeats are clustered in regulatory genes in *Saccharomyce cerevisiae*. Genetics 154: 1053–1068

Yu J, Lu H, Bernardo R (2001) Inconsistency between SSR groupings and genetic backgrounds of white corn inbreds. Maydica 46: 133-139

Zhu, Y., Strassmann, J.E., and Queller, D.C. 2000. Insertions, substitutions, and the origin of microsatellites. Genet Res 76: 227-236

Annexure I

Paper Published:

1. **Harvinder Singh**, Rupesh Deshmukh, Abhinay Singh, Ashok K Singh, Kishor Gaikwad, Tilak R Sharma, Trilochan Mohapatra, Nagendra K Singh Highly variable SSR markers suitable for rice genotyping using agarose gels (2009). Molecular Breeding (Accepted)
2. **Harvinder Singh,** Rupesh Deshmukh, Vikas Khandelwal, Rakesh Singh, P K Naik, Nagendra K Singh Genetic diversity and comparative polymorphism studies in genic and non genic regions of rice (*Oryza sativa* L.) Journal of Biosciences (2009) (under revision)
3. Nagendra K. Singh , Dalal Vivek , Kamlesh Batra, Binay K. Singh, G. Chitra, Archana Singh, Irfan A. Ghazi, Mahavir Yadav, Awadhesh Pandit, Rekha Dixit, Pradeep K. Singh**, Harvinder Singh**, Kirpa R. Koundal, Kishor Gaikwad, Trilochan Mohapatra and Tilak R. SharmaSingle-copy genes define a conserved order between rice and wheat for understanding differences caused by duplication, deletion, and transposition of genes (2006) Functional & Integrative Genomics ISSN: 1438-793X (Paper) 1438-7948 (Online) DOI: 10.1007/s10142-006-0033-4, Pages: 1 – 19
4. Rice genome consortia for 11* and 12 chromosome *(Indian Initiative for Rice Genome Sequencing (IIRGS), Indian Agricultural Research Institute: Nagendra K Singh, Trilochan Mohapatra , Tilak R Sharma, Kishor Gaikwad, Archana Singh, Vivek Dalal, Subodh K Srivastava, Anupam Dixit, Ajit K Pal, Irfan A Ghazi, Mahavir Yadav, Awadhesh Pandit, Ashutosh Bhargava, K Sureshbabu, Rekha Dixit, **Harvinder Singh**, Suresh C Swain, Sumita Pal, M Ragiba, Pradeep K Singh, Vibha Singhal, Sangeeta D Mendiratta, Kamlesh Batra The sequence of rice chromosome 11 and 12, rich in disease resistance genes and recent gene Duplication (2005)"*BMC Biology***",** 3:20 doi: 10.1186/1741-7007-3-20.
5. **Harvinder Singh**, Subodh Srivastava and A.K. Singh Path analysis of component traits of yield in chilli (Capsicum annuum) (2005). Progressive HorticultureVol. 2: 17-22

Poster Presented:

1 **Harvinder Singh** N.K.Singh (2007) Relationship between the Length of SSR and level of Polymorphism in Rice (*Oryza sativa* L.) Abstr ID: 102, Convergence of Genomics and the Land-Grant Mission: Emerging Trends in the Application of Genomics in Agricultural Research, A National Scientific Conference at Purdue University West Lafayette, Indiana September 10-12, 2007, Purdue**.** (USA)

2 N.K. Singh, V. Dalal, K. Batra, B. Singh, A. Singh, M. Yadav, R. Dixit, I. A. Ghazi, A. Pandit**, H. Singh**, P.K. Singh, K. Gaikwad, T. Mohapatra, and T.R. Sharma. (2006). Comparative analysis of rice and wheat genomes as a tool for gene discovery. Abstr ID: 5395, 2nd International Rice Congress, Oct. 9th - 13th, 2006, New Delhi.

3 I. A Ghazi, **H. Singh**, A. Singh, M, Yadav, A. Pandit, R. Dixit, V. Dalal, P.K. Singh, K. Batra, P.S. Srivastava, K. Gaikwad, T.R. Sharma, T. Mohapatra, and N.K. Singh (2006). Strategies for the Gap closure used in

the MTP of the long arm of chromosome 11 of rice. Abstr ID: 2574, 2nd International Rice Congress, Oct. 9th - 13th, 2006, New Delhi.

4 V. Dalal, G. Chitra, A. Singh, M. Yadav, R. Dixit, I. A Ghazi, A. Pandit, P.K. Singh**, H. Singh**, K. Batra, K. Gaikwad, T.R. Sharma, T. Mohapatra, and N.K. Singh (2006). 'Vanshanu Dhan' - a rice genes database for mapping and cloning of agronomically important genes. Abstr ID: 5312, 2nd International Rice Congress, Oct. 9th - 13th, 2006, New Delhi.

5 N.K. Singh, K. Batra, G. Chitra, V. Dalal, A. Singh, S.K. Srivastava, I. A. Ghazi, M. Yadav, R. Dixit, A. Pandit, **H. Singh**, P.K. Singh, K. Gaikwad, T.R. Sharma and T. Mohapatra (2005). 'Vanshanu Dhan' - a rice gene database for mapping and cloning of genes for agronomic traits. International Conference on Plant Genomics & Biotechnology : Challenges & Opportunities, Oct. 26th - 28th 2005, Raipur, India.

6 N.K. Singh, V. Dalal, A.K. Pal, K. Batra, A. Singh, A. Bhargava, S.K. Srivastava, A. Dixit, I.A. Ghazi, M. Yadav, A. Pandit, R. Dixit, B.K. Singh, P.K. Singh, **H. Singh**, S.D. Mendiratta, G. Chitra, K. Gaikwad, T.R. Sharma and T. Mohapatra (2005). Highly conserved synteny of single copy rice genes with wheat. International Conference on Plant Genomics & Biotechnology : Challenges & Opportunities, Oct. 26th - 28th 2005, Raipur, India.

7 Raghuvanshi, S, A. Mohanty, A. Bharti, A. Gaur, A Kapur, V. Gupta, D.Kumar, V Ravi, S Vij, S Sharma, Parul Khurana, P. Khurana, JP Khurana, AK Tyagi, AK Pal, A Dixit, A Singh, A Bhargava**,** A Pandit, IA Ghazi, **H Singh**, M Yadav, M Ragiba, R Dixit, S Srivastava, S. Pal KS Babu, SC Swain, V Dalal, K Batra, K Gaikwad. TR Sharma, T. Mohapatra, NK Singh. (2004). Indian Initiative for Rice Genome Sequencing. 9th National Rice Biotechnology Network Meeting April 15-17. NASC Complex, New Delhi.

8 N.K. Singh, K. Gaikwad, K. Batra, M.H.M. Ammar, Y. Amarawati, S. Anand, S.K. Srivastava, A. Pandit, **H. Singh**, V. Dalal, A. Singh, I.A. Ghazi, M. Yadav, R. Dixit, P.K. Singh, S.D. Mendiratta, T. R. Sharma, T. Mohapatra, A.K. Singh, V.P. Singh, R.K. Singh. (2004). Functional genomics of rice: potential and prospects. International Symposium on Rice: "From Green Revolution to Gene Revolution", October 4-6, 2004, DRR, Hyderabad.

9 N.K. Singh, S. Raghuwanshi, S.K. Shrivastava, A. Gaur, A.K. Pal, V. Dalal, A. Singh, I.A. Ghazi, A. Bhargav, M. Yadav, A. Dixit, A. Pandit, **H. Singh**, K. Batra, K. Gaikwad, T.R. Sharma, T. Mohapatra, A. Mohanty, A.K. Bharti, A. Kapur, V. Gupta, D. Kumar, S. Vij, V. Ravi, Parul Khurana, S. Sharma, McCombie, J. Messing, R. Wing, T. Sassaki, Paramjit Khurana, J.P. Khurana, A.K. Tyagi. Sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny : 9th National Rice Biotechnology Network meeting at NASS Campus, New Delhi(India). April 15-17, 2004.

10 Ramachandran S**, Singh H**, Verma SK (2003) Role of extracellular proteins in bioremediation of heavy metals/radionuclides. International Symposium on Emerging Trends in Genomics and Proteomics, Education

and Research. Jan 12 – 13, Birla Institute of Technology and Science, Pilani. p. 37.

Annexure II

Biography of Dr. N.K.Singh

Dr. N.K. Singh, presently holding the position of Principal Scientist in this institute has a distinguished and impressive record all throughout his long research career spanning over two decades beginning 1981 when he enrolled for his Ph.D. programme in the university of Adelaide, Australia. He obtained his M.Sc in Genetics & Plant Breeding from BHU in 1980. He holds the credit for guiding more than 30 students towards the accomplishment of their M.Sc./Ph.D. Degrees. He specializes in plant molecular biology, Transgenic wheat, Improvement of Technological & Nutritional quality of wheat through Biotechnological tools, Genomics, Proteomics etc. Dr. Singh has more than 50 publications to his credit in premier international journals. He also holds the honour of developing the first transgenic variety of wheat in India and has developed several innovative tools for the analysis of wheat proteins. He also has the distinction of receiving numerous honours and awards in recognition to his excellent academic and research contributions. These include : Gold medal and merit scholarships from BHU during B.Sc. and M.Sc., Best student award at University of Adelaide, CSIRO post-doctoral award, Australia(1986), National Research fellowship, QE II award, Australia(1988) and the National Bioscience award, DBT, Govt. of India (2001), Dr Rafi Ahmad Kidwai Award (2007).

Biography of Mr. Harvinder Singh

Mr. Harvinder Singh completed M.Sc. Genetics and Plant Breeding from C.S.A.University (Agri and Tech) and M.E. Biotechnology from Birla Institute of Technology and Science, Pilani. He has continued for Ph.D programme of this institute from 2004-2009. Harvinder Singh has an active interest in Genomics and Plant Molecular Biology. He has published research articles in renowned International journals and presented posters in various National and International conferences.

Annexure III

SSR frequency for di-, tri-,tetra-,penta- and hexanucleotide at different repeat length in the rice genome

	05 – 15X*	16 – 25X	26 – 35X	36 – 45X	46 – 55X	56 – 65X	66 – 75X	76 – 85X	86 – 95X	>=96X	Total
Dinucleotide repeats											
Chr1	3823	367	182	82	19	2	0	0	0	0	4475
Chr2	2973	271	157	95	19	0	0	1	0	0	3516
Chr3	2980	317	161	83	21	3	3	0	0	1	3569
Chr4	2768	250	113	35	6	1	2	1	1	0	3177
Chr5	2453	244	123	56	14	2	0	0	0	1	2893
Chr6	2857	241	153	88	23	0	0	0	0	0	3362
Chr7	2342	202	117	53	13	0	2	0	0	0	2729
Chr8	2451	218	144	68	16	2	0	0	0	1	2900
Chr9	1966	173	116	75	14	1	0	0	0	0	2345
Chr10	1961	187	94	58	9	1	0	0	0	2	2312
Chr11	2147	412	206	101	30	6	0	1	0	0	2903
Chr12	2479	244	110	62	9	2	2	2	1	0	2911
Total	**31200**	**3126**	**1676**	**856**	**193**	**20**	**9**	**5**	**2**	**5**	37092
Trinucleotide repeats											
Chr1	3497	38	14	5	1	1	0	0	0	0	3556
Chr2	3136	33	16	4	2	0	0	0	0	0	3191
Chr3	3179	29	14	5	1	1	0	1	0	0	3230

	05 – 15X*	16 – 25X	26 – 35X	36 – 45X	46 – 55X	56 – 65X	66 – 75X	76 – 85X	86 – 95X	>=96X	Total
Chr4	2465	12	6	3	2	0	1	0	0	1	2490
Chr5	2620	25	5	1	0	0	0	0	0	0	2651
Chr6	2545	30	8	2	0	2	0	0	0	0	2587
Chr7	2251	21	11	6	0	0	0	0	0	0	2289
Chr8	2255	18	4	2	2	0	1	0	0	0	2282
Chr9	1743	15	11	0	0	0	0	0	0	0	1769
Chr10	1758	18	4	1	0	1	0	0	0	0	1782
Chr11	1765	38	20	3	0	0	0	0	0	0	1826
Chr12	2000	18	16	3	0	0	0	0	0	0	2037
total	**29214**	**295**	**129**	**35**	**8**	**4**	**2**	**1**	**0**	**1**	29689
Tetranucleotide repeats											
Chr1	296	4	2	1	0	0	0	0	0	0	303
Chr2	271	7	1	0	0	0	0	0	0	0	279
Chr3	237	5	0	0	0	0	0	0	0	0	242
Chr4	173	2	2	0	1	0	0	0	0	0	178
Chr5	202	4	1	0	0	0	0	0	0	0	207
Chr6	197	5	3	0	0	0	0	0	0	0	205
Chr7	183	10	3	1	1	0	0	0	0	0	198
Chr8	186	3	1	1	0	0	0	0	0	0	191
Chr9	178	6	1	0	1	0	0	0	0	0	186
Chr10	128	3	0	0	0	0	0	0	0	0	131
Chr11	221	16	2	0	0	0	0	0	0	0	239
Chr12	205	3	0	0	0	0	0	0	0	0	208

	05 – 15X*	16 – 25X	26 – 35X	36 – 45X	46 – 55X	56 – 65X	66 – 75X	76 – 85X	86 – 95X	>=96X	Total
Total	**2477**	**68**	**16**	**3**	**3**	**0**	**0**	**0**	**0**	**0**	2567
Pentanucleotide repeats											
Chr1	77	0	0	0	0	0	0	0	0	0	77
Chr2	59	0	0	0	0	0	0	0	0	0	59
Chr3	65	0	0	0	0	0	0	0	0	0	65
Chr4	51	0	0	0	0	0	0	0	0	0	51
Chr5	64	1	0	0	0	0	0	0	0	0	65
Chr6	49	0	0	0	0	0	0	0	0	0	49
Chr7	44	0	0	0	0	0	0	0	0	0	44
Chr8	45	0	0	0	0	0	0	0	0	0	45
Chr9	30	0	0	0	0	0	0	0	0	0	30
Chr10	28	0	0	0	0	0	0	0	0	0	28
Chr11	41	0	0	0	0	0	0	0	0	0	41
Chr12	47	1	0	0	0	0	0	0	0	0	48
Total	600	2	0	0	0	0	0	0	0	0	602
Hexanucleotide repeats											
Chr1	42	0	1	0	0	0	0	0	0	0	43
Chr2	25	0	0	0	0	0	0	0	0	0	25
Chr3	23	1	0	0	0	0	0	0		0	24
Chr4	26	0	0	0	0	0	0	0	0	0	26
Chr5	**28**	0	0	0	0	0	0	0	0	0	28
Chr6	23	0	0	0	0	0	0	0	0	0	23

	05 – 15X*	16 – 25X	26 – 35X	36 – 45X	46 – 55X	56 – 65X	66 – 75X	76 – 85X	86 – 95X	>=96X	Total
Chr7	12	0	0	0	0	0	0	0	0	0	12
Chr8	22	0	0	0	0	0	0	0	0	0	22
Chr9	15	0	0	0	0	0	0	0	0	0	15
Chr10	10	0	0	0	0	0	0	0	0	0	10
Chr11	10	0	0	0	0	0	0	0	0	0	10
Chr12	22	0	0	0	0	0	0	0	0	0	22
Total	258	1	1	0	0	0	0	0	0	0	260

*** X depicts the repeat number**

Annexure IV

SSR frequency for di-, tri-, tetra-, penta- and hexanucleotide at different repeat range in intergenic, introns and UTR region in rice genome.

	Intergenic						Intron						UTRs						
Chr.	**5-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**05-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**05-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**Total**
Dinucleotide repeats																			
Chr1	2633	297	149	69	15	1	723	76	42	17	4	1	320	19	3	0	0	0	4369
Chr2	1910	203	123	78	17	1	557	59	43	24	2	0	307	21	1	0	0	0	3346
Chr3	1861	229	125	70	14	8	450	62	24	5	3	2	396	27	2	0	0	0	3278
Chr4	1837	185	90	31	5	3	450	62	24	5	3	2	182	16	4	0	0	0	2899
Chr5	1574	185	103	52	12	2	413	54	23	6	2	2	249	15	1	0	0	0	2693
Chr6	1927	187	123	75	19	1	428	47	31	15	5	0	216	14	2	1	0	0	3091
Chr7	1488	147	98	41	11	2	371	47	18	13	1	0	186	13	3	0	0	0	2439
Chr8	1592	166	109	59	12	3	372	45	36	11	4	0	191	13	1	0	0	0	2614
Chr9	1234	126	87	76	12	0	302	38	20	35	2	0	136	12	0	1	0	0	2081
Chr10	1251	155	80	48	8	4	284	38	17	8	3	0	111	4	3	0	0	0	2014
Chr11	1630	211	109	59	10	1	406	61	28	12	3	0	95	5	2	0	0	0	2632
Chr12	1609	180	96	52	9	7	406	61	28	12	3	0	179	7	0	0	0	0	2649
total	**20546**	**2271**	**1291**	**710**	**144**	**33**	**5162**	**650**	**330**	**163**	**35**	**7**	**2568**	**166**	**22**	**2**	**0**	**0**	34100
Trinucleotide repeats																			
Chr1	1059	22	10	4	0	1	420	11	4	1	1	0	648	5	1	0	0	0	2187
Chr2	873	24	11	3	1	0	350	3	2	1	0	0	577	3	0	0	0	0	1848
Chr3	845	24	11	5	0	1	368	3	5	0	4	1	722	0	0	0	0	0	1989
Chr4	799	7	4	3	2	0	322	7	2	0	0	0	412	1	0	0	0	0	1559
Chr5	834	14	4	0	0	0	267	6	1	2	0	0	421	1	0	0	0	0	1550
Chr6	828	25	6	2	0	2	262	5	1	0	0	0	378	2	0	0	0	0	1511
Chr7	656	13	9	6	0	0	226	4	1	0	0	0	382	0	1	0	0	0	1298
Chr8	688	11	3	2	2	0	287	8	0	0	0	1	345	0	1	0	0	0	1348
Chr9	558	11	9	0	0	0	221	4	1	0	0	0	261	0	1	0	0	0	1066
Chr10	569	11	3	1	0	0	199	4	0	0	0	2	241	2	0	0	0	0	1032
Chr11	589	18	10	0	0	0	205	7	2	0	0	0	212	1	0	0	0	0	1044
Chr12	643	14	14	3	0	0	236	5	7	0	0	1	288	1	0	0	0	0	1212
total	**8941**	**194**	**94**	**29**	**5**	**4**	**3363**	**67**	**26**	**4**	**5**	**5**	**4887**	**16**	**4**	**0**	**0**	**0**	17644
Tetranucleotide repeats																			

	Intergenic						Intron						UTRs						
Chr.	**5-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**05-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**05-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**Total**
Chr1	187	3	2	1	0	0	71	3	0	0	0	0	35	0	0	0	0	0	302
Chr2	153	7	1	0	0	0	91	0	0	0	0	0	26	0	0	0	0	0	278
Chr3	133	4	0	0	0	0	75	1	0	0	0	0	18	0	0	0	0	0	231
Chr4	110	2	2	0	0	0	38	0	0	0	0	0	6	0	0	0	0	0	158
Chr5	114	3	1	0	0	4	56	1	0	0	0	0	25	0	0	0	0	0	204
Chr6	109	5	3	0	0	0	57	0	0	0	0	0	24	1	0	0	0	0	199
Chr7	110	9	3	1	1	0	38	1	0	0	0	0	14	0	0	0	0	0	177
Chr8	99	3	1	1	0	0	60	0	0	0	0	0	21	0	0	0	0	0	185
Chr9	112	6	0	1	1	0	34	0	0	0	0	0	14	0	0	0	0	0	168
Chr10	63	3	0	0	0	0	30	0	0	0	0	0	15	0	0	0	0	0	111
Chr11	156	3	0	0	0	0	48	1	0	0	0	0	8	0	0	0	0	0	216
Chr12	119	3	0	0	0	0	58	0	0	0	0	0	17	0	0	0	0	0	197
total	1765	51	13	4	2	4	656	7	0	0	0	0	223	1	0	0	0	0	2726
Pentanucleotide repeat																			
Chr1	55	0	0	0	0	0	13	0	0	0	0	0	13	0	0	0	0	0	81
Chr2	41	0	0	0	0	0	13	0	0	0	0	0	10	0	0	0	0	0	64
Chr3	44	0	0	0	0	0	11	0	0	0	0	0	14	0	0	0	0	0	69
Chr4	42	0	0	0	0	0	7	0	0	0	0	0	4	0	0	0	0	0	53
Chr5	44	0	0	0	0	0	14	1	0	0	0	0	5	0	0	0	0	0	64
Chr6	37	0	0	0	0	0	5	0	0	0	0	0	4	0	0	0	0	0	46
Chr7	36	0	0	0	0	0	3	0	0	0	0	0	4	0	0	0	0	0	43
Chr8	27	0	0	0	0	0	13	0	0	0	0	0	8	0	0	0	0	0	48
Chr9	22	0	0	0	0	0	6	0	0	0	0	0	7	0	0	0	0	0	35
Chr10	20	0	0	0	0	0	4	0	0	0	0	0	4	0	0	0	0	0	28
Chr11	30	0	0	0	0	0	6	0	0	0	0	0	1	0	0	0	0	0	37
Chr12	33	1	0	0	0	0	10	0	0	0	0	0	8	0	0	0	0	0	52
total	431	1	0	0	0	0	105	1	0	0	0	0	82	0	0	0	0	0	620
Hexanucleotide repeat																			
Chr1	23	0	0	0	0	0	4	0	1	0	0	0	4	0	0	0	0	0	32
Chr2	9	0	0	0	0	0	10	0	0	0	0	0	2	0	0	0	0	0	21
Chr3	11	0	0	0	0	0	3	1	0	0	0	0	5	0	0	0	0	0	20
Chr4	17	0	0	0	0	0	4	0	0	0	0	0	2	0	0	0	0	0	23
Chr5	14	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	18
Chr6	10	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	13
Chr7	3	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0	8
Chr8	12	0	0	0	0	0	4	0	0	0	0	0	1	0	0	0	0	0	17
Chr9	11	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	12

	Intergenic						Intron						UTRs						
Chr.	**5-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**05-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**05-15X**	**16-25X**	**26-35X**	**36-45X**	**46-55X**	**>=56X**	**Total**
Chr10	4	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	5
Chr11	4	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	6
Chr12	11	0	0	0	0	0	3	0	0	0	0	0	2	0	0	0	0	0	16
Total	129	0	0	0	0	0	37	1	1	0	0	0	23	0	0	0	0	0	191

Annexure V

SSR frequency for di-, tri-,tetra-,penta- and hexanucleotide at different repeat length in the predicted exonic region of rice genome

	Repeat range				
	5X-15X*	**16X-25X**	**26X-35X**	**36X-45X**	**Total**
Dinucleotide					
Chr01	241	14	3	0	258
Chr02	221	12	2	0	235
Chr03	314	20	0	0	334
Chr04	173	7	3	0	183
Chr05	206	6	0	0	212
Chr06	179	11	2	1	193
Chr07	159	9	2	0	170
Chr08	177	10	1	0	188
Chr09	122	4	0	1	127
Chr10	118	4	3	0	125
Chr11	103	2	2	0	107
Chr12	140	7	0	0	147
Total	2153	106	18	2	2279
Trinucleotide					
Chr01	1773	7	1	0	1781
Chr02	1590	4	3	0	1597
Chr03	1864	1	0	0	1865
Chr04	983	1	0	0	984
Chr05	1194	4	0	0	1198
Chr06	1079	1	0	0	1080
Chr07	923	3	2	0	928
Chr08	832	0	1	0	833
Chr09	528	2	2	0	532
Chr10	585	4	1	0	590
Chr11	582	3	1	0	586
Chr12	693	0	0	0	693
Total	12626	30	11	0	12667
Tetranucleotide					
Chr01	24	0	0	0	24
Chr02	19	0	0	0	19
Chr03	21	0	0	0	21

Chr04	9	0	0	0	9
Chr05	24	0	0	0	24
Chr06	20	0	0	0	20
Chr07	12	0	0	0	12
Chr08	15	0	0	0	15
Chr09	11	0	0	0	11
Chr11	9	0	0	0	9
Chr12	14	0	0	0	14
Total	189	0	0	0	189
Pentanucleotide					
Chr01	5	0	0	0	5
Chr02	8	0	0	0	8
Chr03	17	0	0	0	17
Chr04	3	0	0	0	3
Chr05	5	0	0	0	5
Chr06	4	0	0	0	4
Chr07	5	0	0	0	5
Chr08	6	0	0	0	6
Chr09	4	0	0	0	4
Chr10	4	0	0	0	4
Chr11	1	0	0	0	1
Chr12	5	0	0	0	5
Total	67	0	0	0	67
Hexanucleotide					
Chr01	16	0	0	0	16
Chr02	3	0	0	0	3
Chr03	11	0	0	0	11
Chr04	5	0	0	0	5
Chr05	11	0	0	0	11
Chr06	11	0	0	0	11
Chr07	6	0	0	0	6
Chr08	6	0	0	0	6
Chr09	3	0	0	0	3
Chr10	5	0	0	0	5
Chr11	6	0	0	0	6
Chr12	8	0	0	0	8
Total	91	0	0	0	91

*** X depicts the repeat number**

Annexure VI

Details of primers used for amplification of 201 SSR loci of different repeat motifs and lengths howing variable number of alleles in 8 rice genotypes

S. No	Primer Id#	SSR motif	Repeat length (bp)	Forward	Reverse	Product	No. of alleles	PIC
1	nksrssr01_20	TATC	60	CAACCAATTCGCTATGGTAT	CCAGTTTAAAGTTTCGTTCG	344	4	0.72
2	nksrssr01_25	GA	25	TAGGCTTTGTGATGGAAC	GAGGAGATGGCGGAGTAG	262	1	-
3	nksrssr01_240*	CTG	30	TAGCTAATTAACCCACCACC	TCCTACTCCACTCTACCTGG	345	1	-
4	nksrssr01_5940	ATT	69	TACTTTCTGTCTCCCTCCATT	TTTAGATGACTCGTTAGGTTCT	359	4	0.72
5	nksrssr01_9791	AGA	94	ATTTGAGCATTGACAGCAAC	AAAGATGAAGGGTTCTCCTG	383	2	0.49
6	nksrssr01_10282	ATT	114	TCTCCTTTGTCTCTCTTCACT	TTGTAGCATTTCACCCAGA	315	4	0.56
7	nksrssr01_19351	GA	95	TGGCAACGAAGACACATC	GGAGTAGAACCACGAAGCC	262	3	0.41
8	nksrssr01_20380	ATA	93	TTACACCTTGTGCCTAACA	CCTCCAGCGATAATGAAGTAG	378	1	-
9	nksrssr01_20427	TA	209	GCTCCAAGGTTCCAAAGT	TTTCTCCCTCAATGTCCTC	350	1	-
10	nksrssr01_24287*	GTT	50	AAAGAATCTCTTTGGGAAGG	AACAACATCAGCATCAACAA	256	3	0.65
11	nksrssr01_25171*	CCT	21	CAAAGCGGCAATGTAATC	GAGGAGGACGAGATGGTG	393	2	0.32
12	nksrssr01_25279	TA	105	CATTTAGGGCATCCACAAG	TACACACACACACACACACAC	374	2	0.49
13	nksrssr01_27499	GCAC	24	GCCAACACGAACTCCCGAAGG	CTGCTCCGTCTGCATCGTCTGC	367	2	0.22
14	nksrssr01_28113*	AAG	54	GACTCACTCGTCTCGTGG	CTGGCATCAACTTCTCATT	290	2	0.41
15	nksrssr01_31482*	CCT	38	GACGGTGTCGGTGGGCTT	TCGCTTCTCCTTCTCCCT	400	1	-
16	nksrssr01_31878	CT	69	CCTTCCCTTGTGATAGAGATT	TTATTAGCGTGTGTTTGCC	237	1	-
17	nksrssr01_36481*	CAA	59	CGGTCGATTATGCAGTCTAT	ATCGGATTGACAGATTGAAC	254	2	0.38
18	nksrssr01_38671*	AAT	122	AATAATACTCCCTCCGTTTCA	AGCGTGTCCCATAATCAGT	360	2	0.5
19	nksrssr01_38964	GA	25	GCGTCGTCTCCTCACTCT	ACAAAGCCCATAACAACATC	230	1	-
20	nksrssr01_42559*	GAC	25	CGAGCCATCACCAACTAC	GTCCAGCCAGCCAACACC	167	2	0.5
21	nksrssr02_83	AGAC	13	GACTCGCTACTGCTTCTTGT	TGGTGGATGGATGATTTATT	375	2	0.5
22	nksrssr02_125	TCT	63	AAGAGATGAGAAGAGCAATGA	CAACTTAGAGGAAGAAGGAGG	312	4	0.69
23	nksrssr02_806	TA	74	TAAGCCGCAAATGAAAGA	ATGTGTCCCTTCTCCATACT	234	4	0.67
24	nksrssr02_1078	TAT	78	TTAGGCACAATAAACACAGA	GAGATGGGATGGAGGAGT	350	2	0.47
25	nksrssr02_1124	ATAG	52	ATCACTTGATAGCTTGGGAA	TCCTGCTATAGGCTAGGTTG	350	1	-
26	nksrssr02_3423*	AAG	89	ATCTATTCAGGGTTCAGGGT	CCCAAGTTCGTTACAGCA	188	1	-

S. No	Primer Id#	SSR motif	Repeat length (bp)	Forward	Reverse	Product	No. of alleles	PIC
27	nksrssr02_5203	ACG	24	CCAAGTTCCCATTCTCTGTAA	GCCAACATCCTCGCCTTC	274	1	-
28	nksrssr02_5324	ACAG	16	ATTAGACCACATCCATCCACT	TCATCATCAATCTCAGCACTT	330	2	0.49
29	nksrssr02_5357*	AT	88	GATGACGAGGAGAGAGAGAGT	CAAGTGTGGATTTGAGCAG	304	1	-
30	nksrssr02-9769	ATAG	59	AGCTATCTACGTCTTGCGTG	ACAGTTACAAGCACGGATTC	390	1	-
31	nksrssr02_11027	TA	97	ATCAGTCGGAATCTCAGG	AGTCACATCTAAACAGGGTCT	363	2	0.44
32	nksrssr02_12192	ATA	24	CTAAACATTCAAGGCAAGGT	GAAAGCAGCAAGACAGATAG	213	2	0.47
33	nksrssr02_12270	AT	77	AGGATGAGAGAAGATGGAAGA	TGTGAACCTATTTGACTACCC	199	1	-
34	nksrssr02_19743*	CTT	82	ATGGCCAACTCTTCAACTAA	GGAGAAGAGAAGGAGGAAGA	225	3	0.59
35	nksrssr02_24589	AGG	24	GCATCGATCTTAATCTGCCC	CATGGCCATCCCTAAAACAC	153	1	-
36	nksrssr02_30376*	CAG	32	TGAGAAACAAGAGGTTGAAGA	CCTGATGAATACTGCCTATGA	399	1	-
37	nksrssr02_32994	ATCT	66	CCGATTGATCGGAATATCT	GGAGCGATTGTAGAAGTGAC	229	1	-
38	nksrssr02_35496*	AAC	68	CTGCATCAATATAATTGCGA	GCTACTTACACCACCACCAT	280	2	0.47
39	nksrssr03_105	TA	79	TGGAGGACACAAATACAGTGA	GCTCAACTACAAGGAAACAAA	373	2	0.49
40	nksrssr03_1022	AGAC	15	ATTGTCGTCCAGCATCAG	ACGGAGAGAGTAGGAGAGTTG	331	2	0.5
41	nksrssr03_5219	AT	70	AATCTGTTGCCATTTCTT	TGACCACCTGACTATGATG	355	2	0.41
42	nksrssr03_5219	AT	70	TGACCACCTGACTATGATG	AATCTGTTGCCATTTCTT	355	2	0.41
43	nksrssr03_5379	GCG	29	GGAGACAGCACGAAATCC	AGGAGCAAGAGAACGACC	339	1	-
44	nksrssr03_5694	ACAG	16	CAAGAAACAAGAACATCAAGG	AAAGTGGGAAAGAAAGAGAAA	325	2	0.49
45	nksrssr03_5942	GTA	80	ATTGACACGAAGAGGACAAA	GTGCCCGAGGTGAGTGAGT	261	2	0.22
46	nksrssr03_11479	TA	25	CATTCTCTCCTCTCCGCT	GATTGTGTCTATCTCCCGTC	381	1	-
47	nksrssr03_11652	TA	53	GGGTATTTGTAAGGTGAGGTG	AACCGAGAGAGAGATGTGTG	365	4	0.72
48	nksrssr03_12043	TAA	115	AAATGATAAAGTTGCTTACCC	AGTGCTGCTACCTCTTCTAAAC	264	4	0.69
49	nksrssr03_12086	GCT	23	CGGACCCACACATCAGTC	GTATCACACCCTGCTTATCCT	238	1	-
50	nksrssr03_12086	TA	60	GCCTGGACACATCAGTC	CATAGTCACCCTGCTTATCCT	238	2	0.22
51	nksrssr03_14678	TCTA	57	TTTCAATCCAGAAACCCTAA	GAAGTTCAAAGCGAGAATTG	328	2	0.22
52	nksrssr03_20495	AT	62	GATTGAAAATTAGAGTTTGCAC	GATTGAAAATTAGAGTTTGCAC	392	3	0.531
53	nksrssr03_26474	GATA	51	CACACCAACTCACTCTTGAA	CCGTTTCGTCTATGTTCATT	396	2	0.5
54	nksrssr03_27842	AT	166	CTAAAGAAGAACACGGTCA	GCAAGATAAGATGTCGCAAA	395	2	0.22
55	nksrssr03_28280	AAT	60	GTATTCCAGATTCCTGCGT	GTGTTATCCTCTCTTTGCTCTC	249	2	0.22
56	nksrssr03_28762	CTT	62	TCTTTACTCGTTTCCTTTGTTT	TAGATGAAGTTGTAGATGCCC	195	2	0.5
57	nksrssr03_36286	GTAT	28	GAGAGAGAGAGGAGAGGCG	GATGCACATCTCGACAGCTC	255	1	-
58	nksrssr04_168	CCT	25	GAAAGCCTCAAAGTAGTAATGG	GCAAGGATAAGTAGGTGGAAG	211	1	-
59	nksrssr04_558*	GA	21	GACGGTGAGAGAAGAGGG	ATCCAAATCCAAACTCCAA	240	1	-

S. No	Primer Id[#]	SSR motif	Repeat length (bp)	Forward	Reverse	Product	No. of alleles	PIC
60	nksrssr04_4030	ATCT	32	GCTGCACTATGCAGTTGCTGAGG	TGATCGATGGTGACGATGATACGC	437	2	0.5
61	nksrssr04_15571*	AAT	53	GTAGGAAGAGAAGAAGGGAAG	ATGACTTATGGGATGGAATGT	293	3	0.59
62	nksrssr04_15710	TCCA	24	CCATACATCCATCGAACCATGC	GCCTCATCCTCGCACTAGAGG	99	1	-
63	nksrssr04_16014	GA	39	CGGAAGAGACAGCAACAC	GAGGACACAAGAGCCAGTAG	312	2	0.5
64	nksrssr04_16255	TAA	65	ATTCTCTGTTTGCCTCCATT	CCCTGTCTTGTTTAGTGTCTTT	254	4	0.69
65	nksrssr04_16256	AT	171	ATGTGGAGCAGAAACGAA	GGAGAGTGGCATTTGAGA	279	1	-
66	nksrssr04_19472	TAT	154	TAACGAGTGGGTTCAGTTCA	AGTTGCGATGTTGCTTGT	312	3	0.45
67	nksrssr04_19845	AG	50	AACGAATTCTATTTTGCGTC	TTCTTCTCATTTCAATTCGC	-159	3	0.40
68	nksrssr04_25032	AT	33	ACTCAATCGGTCACGGTC	GAAGGTTGTCTCTCTTTAGCA	296	3	0.61
69	nksrssr04_25723	TA	50	TACATTAGGTGAGCCATCGT	AAGAAAGCCGTTTAGGACA	361	3	0.61
70	nksrssr04_30674	TC	96	ACAAACAGAAACGGACAGAG	CTGGAGAGGATGAAGGATTAC	333	2	0.49
71	nksrssr04_34518	TA	61	CCAAATCCAACGGTATGTAG	AGGTATGGTCTCTGTGGGT	230	1	-
72	nksrssr04_34639*	GAG	36	ACAATCAAGAACAGAAATCCA	TCAGCGAGACATCACTTC	262	2	0.41
73	nksrssr04_34658	AT	77	GAAATCAACTGTGTCCTCAAC	CTTCTCGCCTGTCTCTCA	397	1	-
74	nksrssr04_35125	ACAG	15	ACTTGGGACGGAGGTAGTAG	AGGCTGAAAGAATAAGAGGAA	242	1	-
75	nksrssr05_15	AAG	46	TGATAAGGGTTAGGATTGTTG	GGAGGAGGATAGAAGTTAGGG	274	1	-
76	nksrssr05_29	CTT	21	ACCATCCCTCCCTTGTAGT	AGCCTCATCTTGCTCCTC	348	1	-
77	nksrssr05_172	ATGA	63	GGCACATATATTCTCCCAAA	GCTTGATGTGGTTATCCATT	152	1	-
78	nksrssr05_328	AT	92	GCATAGTAGCAAATGGGATTAC	CCAAGAGGCAAGAGAAGA	278	3	0.61
79	nksrssr05_3046	AT	85	TCAAGCATAAACAAACATAGAAA	GTGACAAACAAGAACAATAAA	375	1	-
80	nksrssr05_3420	CT	23	ACGGTTTGGTAGGGTGTC	AGGGTGGCAGGGATGTAA	256	2	0.24
81	nksrssr05_3620	TTA	80	ACGGAGGGAGTAGGTCATT	TGTGTGGAAGAATAGAGAGTCA	366	3	0.61
82	nksrssr05_4943	CT	24	AACGCCTATATTAACGCGCC	CGACGTGGAGGAGAGAAGAG	298	1	-
83	nksrssr05_5010*	AAG	19	GGATCAACCAAGAAAGATCA	ATTATTGGCAAATCGATCAC	177	2	0.22
84	nksrssr05_7506	ATAC	24	CACATGGTCACATGGTGTGTAGC	GAGTAGTTCCTGCCGAGCAACC	551	2	0.22
85	nksrssr05_9030	ATA	67	GCAAGGAAGGGATTATTGA	GTAGGTGGAGAGGAGAACAAC	390	2	0.41
86	nksrssr05_9226	AT	73	GAAATCCAACACACGAAGAG	CACGGACAAGAGTAGGGTATT	367	3	0.65
87	nksrssr05_9458	ATAC	63	TGCACACATCCATTACATTT	ATGGGCATGTTTGTAGTCTC	364	2	0.38
88	nksrssr05_19770	ATAG	57	TAAATGTGGTGCCTTCTTTC	TGTGCAATTTGCTTCTGTAG	346	2	0.38
89	nksrssr05_21398*	GAA	75	AATGGTGGGTTGTACATCAT	CATTTGGATCATAAGTTGGG	390	1	-
90	nksrssr05_23557*	CTT	76	CAGGTTCTGCTTCTACTGCT	TCGCCTCTCAATCTTTGTAT	324	2	0.5
91	nksrssr05_23679	AT	104	ATAACTTGAAATGGAGGAAG	CAACGGCTACAACGAATG	396	3	0.61
92	nksrssr05_24848	TA	47	CGGGACAACATTCTACTACTC	TGCCAAAGATAACGAAACTG	321	1	-

S. No	Primer Id#	SSR motif	Repeat length (bp)	Forward	Reverse	Product	No. of alleles	PIC
93	nksrssr05_29392	AT	90	TGCGATGAACTAATGGTAA	GAAGGACGACGACAAAGAC	218	2	0.24
94	nksrssr06_293	AG	32	AAGAGATGGCTGAGGAAGA	CGAGTAGACAAAGAAAGCAAA	183	2	0.22
95	nksrssr06_534	TCT	32	AAACCCAACAGGAAGCAC	ATACTCGCATCGCAAACC	163	3	0.41
96	nksrssr06_535	GAA	67	TAGGGAATCAGCGGTTAG	GCTCTCTTGTCCTTCTTCTTC	216	3	0.45
97	nksrssr06_6354*	CCG	31	AGTCGGTATCCTCCTGCT	GCCATTGTTGTTTATGTGATT	170	1	-
98	nksrssr06_1086	TATG	63	CTGCAGAATGCATGTGAATAAA	AATATTGAAACGGATGGAGTA	218	2	0.47
99	nksrssr06_23125*	GGT	49	CCTTTGTGTTGCTCGTAGT	AGTCTTCCTCCACTTCCTCT	346	3	0.65
100	nksrssr06_27846	AAT	95	AAACAAGCGGCAGGTCAA	ATCTACCCAACCATACAGA	358	1	-
101	nksrssr06_28038	CT	25	ATCAGTTCAGTTCCCGCA	AAGTTCGCATCAGCAAAG	312	1	-
102	nksrssr06_30840	TA	50	TTAGATTGAGTTGTTTACGGG	AAGTGATGGTGAAGATTATTG	168	3	0.63
103	nksrssr06_30941	TC	39	TTAGTGCCTGTTTGTTTCCT	GAGCAAGTTAGTGTGGTCTTC	382	2	0.44
104	nksrssr07_19349	TCGA	19	CAACTGACGGAGTGGGTG	TCGTGTCGTGTCTTTGTTT	370	1	-
105	nksrssr07_19359	GT	52	ATGCCACCGACACCAAGT	GAAGCACGCAAATGATAAA	347	2	0.41
106	nksrssr07_19463	AT	98	ACATTTGGTTATTTGTTTATCT	ACCTCCTTGTGGTCCTTG	400	1	-
107	nksrssr07_20547*	AT	85	TCCTACTTCCGCCTTCTATC	CTACGGGCTTGTTCCATT	380	2	0.32
108	nksrssr07_21222*	AT	78	GCTGTCTGGTAATAGTCGTTG	ACTTCAATGGTGTCGCTTT	391	3	0.57
109	nksrssr07_21630	CT	39	CTGGTCTCTATTTCATTCCTG	TATGCTAACTACCTGCCAACT	314	3	0.63
110	nksrssr07_22533	ATGC	19	TAGGGACGATGGACGAAC	ACTTGGAACGAGAGAGAGATT	377	2	0.38
111	nksrssr07_22775	ATA	97	GGAGGAGGAGTTGTAGGGT	AATCTTCGTGGACTTCATTTC	271	4	0.72
112	nksrssr07_23109	CTT	33	GAGGTAGGCGAGTGGTAGTAG	AAGAGTTCGTCATCGTCGT	395	2	0.22
113	nksrssr07_23468	CTCC	35	GTGGGATGGCAACGGGCAG	AGGAGACAGGAGTCAGAAGAG	143	2	0.47
114	nksrssr07_23827	TA	62	AGAACAACAGCGGAAGATAA	AATAACCATAGCCACTCCTTG	298	1	-
115	nksrssr07_23994	ATAG	37	TTCTAAACCACGAAGATAGCA	AGACTACAAACCAAACCAACC	135	1	-
116	nksrssr07_25062*	GAA	48	GAGTAAAGATGTTGGGTGGA	AATATTTAACATGCCGAGGA	243	1	-
117	nksrssr07_27576*	GAC	33	GTAGACACCGACGACGAC	TCCACCTCCTTCAGCCAC	232	1	-
118	nksrssr08_3246	TC	32	ATGCAATACAGCACACTCGC	CTTATGCTCTCATGGCTCCC	179	1	-
119	nksrssr08_5039	TA	111	GTATGCGATTCTTATTTGAGTT	GAGGACGCTTTGACCGAG	310	1	-
120	nksrssr08_5267	GGC	32	CTGCTGCTGTTGGTGCGG	GAGCCCTGATGTGAACTTATT	370	1	-
121	nksrssr08_5371	TTTTC	28	ATTACGGATAGATGGGAAGAG	TTTGTTGGTTTGTGGTGTT	260	1	-
122	nksrssr08_5696	CAG	35	TTGAGGAAGTAACAGAAACAAA	GAGACAGGAAATGGAAATGA	164	3	0.41
123	nksrssr08_5791	AT	130	ACTCCACTACGATACTCCTCC	AAACACCAAACAAATTTACCA	298	3	0.5
124	nksrssr08_5844	AAG	43	CATGCCTATGAAACTCATCTC	TAGCTCACATCCAAATCCTTA	275	2	0.24
125	nksrssr08_5968	ATA	66	AAAGTCCACTTTACATCGTCA	AAAGTCCACTTTACATCGTCA	299	2	0.5

S. No	Primer Id[#]	SSR motif	Repeat length (bp)	Forward	Reverse	Product	No. of alleles	PIC
126	nksrssr08_5992	AGAC	27	AGTCCTTGAGAGAGATTGGAC	CGTATCCTCATCATCTCTTTG	389	1	-
127	nksrssr08_6088	AT	107	TACCACCGATTATTGTCGTAT	GAAGTGAGGAAGGGAAGAATA	346	2	0.32
128	nksrssr08_6121	GAATT	24	ATAGGTGAATCCAAACAGTGA	TAGGTCATCTCTCTGACATGG	349	1	-
129	nksrssr08_6208	GGC	35	ACACAACTGCAAGAGAAGAGA	CTTCTCCTCCTTCCTAACCTA	228	1	-
130	nksrssr08_7563	AAAG	28	CTATGTTGGCATTTCATCAAT	ACCATCCTTCCAGGTATGTAT	349	1	-
131	nksrssr08_7796	GGCTG	28	GAGAGAGAAGAGAAGCTCCAG	CAGCTACATTCTGCAACAAAT	325	2	0.41
132	nksrssr08_8250	TA	130	GAACGCACTCGTTATATTGAC	ACACCGTTTGTGATAAGATGT	276	1	-
133	nksrssr08_8753	CCCAG	29	AATAGCAGTAGTACACCGGAA	GCGAAGGAAGGAGTAGTAAAT	396	1	-
134	nksrssr08_9910	TATAT	25	TGGCATGTCAATTTCTAATCT	GGAGTACGTACGTGCAACTAT	283	3	0.59
135	nksrssr08_10937	AAT	45	CGTACATACTTGGGATGAGC	GCCATGATCAAAGCATTATT	399	1	-
136	nksrssr08_12289*	CTT	45	ATACATTTGCAGCCAGTCTT	GAGGATGGCAATACACAAAT	362	2	0.38
137	nksrssr08_13867	AAAGG	30	AAACAAGAGCGGATTTCAC	GCGATTAGATTAGGACACTCA	385	1	-
138	nksrssr08_20398	TTGT	26	ATGGGACTGAACATTATTGC	CAAACACAATCCCAAGTTTC	393	2	0.47
139	nksrssr08_20540	TAA	200	AACATCCATGAGTGCTACCT	CTTCTTCTTCTTCCTTGGCT	393	1	-
140	nksrssr08_20682	AATA	47	ATCAAACGCAGCTAATAACC	GAGCTTCGTCTCGTTTAGAA	194	1	-
141	nksrssr08_20841	AT	83	CCAATTCCACTTCCATTGTA	CCAAGGTAGAGGGAGAGTTT	281	1	-
142	nksrssr08_22040	GGT	32	AAGTCTCTCGACCTGAACAA	CCATGCAAATGCAAGTAGTA	204	1	-
143	nksrssr08_22077	TCAC	29	GGCTACCATGACTGAATGAT	GGCCTAATTGGTTTAGGTTT	315	1	-
144	nksrssr08_22078	AGCAG	24	GTTTGCAATAATCAAGGCTC	CTACCTCGCTTCTTGCTG	301	1	-
145	nksrssr08_22570	GAG	34	ATGAACCTGACGAGAGATTT	GAGAGGAAGGTGATGAGCA	377	1	-
146	nksrssr08_22813	TA	87	CAGCTTTACCAGCTGAATTT	GAATACCTCACCCATTGCTA	212	2	0.38
147	nksrssr08_23418	AT	92	TACATTTCCAATCAGACGGT	GACAAGCCGTTCAATATTTC	272	2	0.28
148	nksrssr08_23426	TTC	45	CCGACATGTACTACTGCTCA	ATTGATCTCGATCACCACAT	380	1	-
149	nksrssr08_24109	TTC	29	ATCTGAGCACGATTTCATTT	TTTAAGCCTAATTGCACCAT	368	1	-
150	nksrssr09_593*	TA	124	GGGAGGTTCCATCTTACA	CGGTTATTTATGTTCGGTC	330	1	0
151	nksrssr09_1576	GTATA	25	ATCTGCTTGTTTGATTTGTGT	TGTCCTCTGCCATTGATAAG	220	1	-
152	nksrssr09_2291	TA	70	GAACGGAGGGAGGTTGTT	AAAGTGTCCTAAAGCCAAGTC	272	2	0.5
153	nksrssr09_2463	AT	89	AGGTCTCCTAACAACCCTAAA	GATAAAGCCGAAAGCAATCT	274	2	0.5
154	nksrssr09_2682	AAT	86	TTCAATGTCACTTCCAGACC	ATGGCTCCTAATCTTTCAGTT	246	3	0.57
155	nksrssr09_3177	TA	102	GAGATTGGTGTGACCCTT	CCTACGGCTCCTGACATT	302	1	-
156	nksrssr09_4172	TATC	82	GCGATGACTTTGTTACTCT	GGCGGTTTAGGAGCGTTT	301	1	-
157	nksrssr09_4672	TA	99	ACGGATGGAAGGTGGAAG	ACAGGAGCAGCACAATCA	399	1	-
158	nksrssr09_5127	TA	55	ACCAAGAGAAACTATGAACGG	AACGGAGACCTAACCATCTAA	356	1	-

S. No	Primer Id[#]	SSR motif	Repeat length (bp)	Forward	Reverse	Product	No. of alleles	PIC
159	nksrssr09_5365	TA	39	GATGATTGAGACTGCCTGTT	GACGGTTTACTTGTGAGTGTG	384	2	0.44
160	nksrssr09_6284*	GGC	12	CAGTCATTGCTACCGCTC	GATTCCAACACAAACATTAGG	320	1	-
161	nksrssr09_9205	AG	40	GAGAATGGGATTAGACGATTT	GGCACAGAGAGAGAGAGAGA	337	2	0.38
162	nksrssr09_9921	AC	67	AGAATCACACACAGACACACA	CGCTGAGGAGACTATGACTTT	358	3	0.64
163	nksrssr09_10470	GATA	32	TGAGAGCTCGTAGGAAGTGTCC	CAGAGTCAGCAATCGCTAAGG	244	2	0.22
164	nksrssr09_12351	ATAG	66	GATTTACGTACCTGACGGAG	ACTCCTCTCCAACAATGAA	374	1	-
165	nksrssr09_14566	GAA	53	TGTGAAGAGAGGAAGAACAGA	AAGGTTAGGAAAGTAAGGCAA	219	1	-
166	nksrssr09_16935	TTAT	32	CGTGAAAGTGACGAGTTTCAGTCC	CGAAGTGAACATGGCACAAACC	242	2	0.38
167	nksrssr09_19195*	CAA	72	CAAGAAACCACCATTAGAGC	AAATCATCATCATCATCTTCTTC	254	1	-
168	nksrssr09_22356*	TCT	69	CAGCAAGAACCATTGGAG	TGAACGTATCACCGAACATA	388	2	0.49
169	nksrssr10_9	CATCA	15	TTCAGAACACACCAAACTGA	GATTGAAAGTGCATCTAGGC	316	1	-
170	nksrssr10_12	AG	11	TCAGTGGGAGAGAGAGAGAA	TTTCTGGGATTTATTTGAGC	208	1	-
171	nksrssr10_14*	CCA	12	CCAATCATACACTTGGTGAA	TTTGGCTTCCATACTTGATT	335	2	0.22
172	nksrssr10_17995	GA	15	TCACCTCAGTTTATCCCAAC	GGATGGATGCACTACTGTCT	326	1	-
173	nksrssr10_22114*	GAA	88	ACTACGAATTCGGACAGAGA	ATGTAACGGGTGACGAACT	387	3	0.61
174	nksrssr10_22560	CATCA	19	GACATATGGCCGTGTTTAGT	AGGCTGTGTTTAGTTCCTGA	239	1	-
175	nksrssr10_22560*	CATCA	14	TTAAACTTCCAACTTTCCCA	AGGCTGTGTTTAGTTCCTGA	204	1	-
176	nksrssr11_7483	CT	32	tctctcctcttgtttggctc	acacaccaacacgaccacac	271	2	0.32
177	nksrssr11_14014	ATAG	56	GCGGCGTATTAGCGTTGTAC	CCATAAGGTCTCAGCCCATG	245	3	0.59
178	nksrssr11_16750*	CTT	66	GGTTCCCAATGCAGTATAGA	TAACCAACGTTGATCATTCA	392	2	0.49
179	nksrssr11_19358*	TCT	90	CGTAATGCTTCGAAGTTCTC	GATGAAGTGGTCAGTAGGGA	308	3	0.41
180	nksrssr11_19417	CATCA	24	GCTACATTTGGTTGTGTGATACT	ATTGCTAAAGTGAGGCTGTGT	248	1	-
181	nksrssr11_21630	AT	130	TATGTGTCGTGCCGTGTAG	GAGATTTGGAGGGAGAGGT	320	1	-
182	nksrssr11_21743	CTCTC	40	ACACCATTCCCACATCAG	GCCGAGACACCCACAGAG	300	2	0.38
183	nksrssr11_23495	TA	85	AGTTCCTCCAACATCATCTCT	GCTAAATCACACAAACCAAA	281	2	0.22
184	nksrssr11_23831	AT	92	AAGCCTCATCGGCAACCC	AGGGACACTTCTATCTCCAGT	351	2	0.28
185	nksrssr11_24183	CCT	55	AGGGCTCACCAGTTGCGG	GAGAGATGTTTGATGCGG	319	2	0.24
186	nksrssr11_26523	ATCC	24	ACGACACGACCACCATAA	ACACGAAGAAGAGCACGA	391	1	-
187	nksrssr11_26981	AT	82	ATCTTGGAACAAACCGCA	TTCTCTAAATCAATCAATCAACT	281	1	-
188	nksrssr11_27039	TA	110	TACTCTTGCTCTTCTGCTT	TATCGTCTTGTAGGCTTGTTG	301	3	0.63
189	nksrssr11_27170	ATCT	68	TCCTCGTGTCATCGTGTT	TAAGAGCGTTTGTAGGGAGTT	272	1	-
190	nksrssr11_27617*	TA	91	TACAGGAGAGCCTTGAACAC	AATACTACCAAATGAGCCAAA	388	2	0.22
191	nksrssr11_27776	TA	69	TATTAGAGAAAGGGCGAAGAA	AGGGTCAAGAGACACACAGA	256	1	-

S. No	Primer Id[#]	SSR motif	Repeat length (bp)	Forward	Reverse	Product	No. of alleles	PIC
192	nksrssr12_0.3	AACAG	14	CAGTCAGCACCTCTATCCTC	GAACATTCGAAACCACAGAT	371	1	-
193	nksrssr12_1.1	ACATC	19	TGGTACTGTAGCAGCATGAG	TAGCTACAGACAGATTCGCA	340	1	-
194	nksrssr12_1.4	AGA	15	TGATGCGAATCTGTCTGTAG	CTTTCTCCTCCTCCTTGTTT	237	1	-
195	nksrssr12_1.7	CCG	13	AAACAAGGAGGAGGAGAAAG	GTTGGAGATGTAGAGCTTGG	365	1	-
196	nksrssr12_5104	TTTA	20	TGCAGTGAATGGAGACCACTAGC	CGGTGAGTCCCATATCTTCAACC	197	1	-
197	nksrssr12_12202	CCA	31	TTTAACAGCCCTCCTGTAGA	AAGTGTAGGAGAAGCGGG	326	1	-
198	nksrssr12_21961	TTTTA	14	GGCTCAACCTAGTTCTTCCT	CCCTGATTAGCCATACTTTG	350	1	-
199	nksrssr12_23443	GAAA	24	TCACCGTCACCTCTTAAGTC	GGTGGTTGTGTTCTGTTTGG	194	1	-
200	nksrssr12_23927*	GGA	18	GATGATGCAGAATCTCTTCG	GCTTCAGCTTCACTTTCTTC	270	1	-
201	nksrssr12_26396	CGC	13	AAACAGCAACCGAATTAAAG	ACACGTCTATCGTTTCGTTT	298	1	-

Annexure VII

Details of primers used for the amplification of 45 SSR loci with repeat lengths in the range of 51-70 bp from chromosome 11, designed and used for the validation of polymorphism level in eight rice genotypes

S. no	Primer Id#	SSR Motif	Repeat length (bp)	Forward primer	Reverse primer	Product size (bp)	No. of Alleles	PIC value
1	nksrssr11_19009	TA	54	AACACAGATGGAAAGAGAAGA	AGTGGTTACAAAGCAGTATTT	395	2	0.375
2	nksrssr11_19021	AT	65	ACGGTGGATAAGAGCAGG	TGAAAGACATTACACTACTTGA	214	2	0.219
3	nksrssr11_19201	AT	62	AGCCTCTCACCACCATCT	ACCGCCTACCTCCTGTAA	311	2	0.219
4	nksrssr11_19377	AAG	56	CCGTCACCTTCACTTACTTT	CCTACGAATACAAGCCGTT	231	3	0.531
5	nksrssr11_19631	TA	70	CTCATCCTCAGTTTGTGTCTG	CGGTCTATGTATTGTTGTTCTT	394	2	0.219
6	nksrssr11_19686	CCT	55	AGGGCTCACCAGTTGCGG	GAGAGATGTTTGATGCGG	319	2	0.469
7	nksrssr11_19876	TA	51	ATGCTATTTCTTCCGTTTCA	ATACTACCTCCGTTTCAAGT	151	2	0.469
8	nksrssr11_20180	AT	56	TTTCTCTCTTGTTTCTCCTTT	TGACTATCCTACCCTTGGTTT	320	NA	-
9	nksrssr11_20340	AT	64	ATGTATTTGGGTGCTGTC	AGACGCTTGTGGTGACTTT	316	3	0.531
10	nksrssr11_20461	AT	68	TTGAACTTAGGCAAGGCA	TACTTATTGGATTGGTCTCC	207	NA	-
11	nksrssr11_20467	AT	70	ATCTTACCTTGCTTGCTCTTG	TTTCGGACACACTACACAACT	373	2	0.375
12	nksrssr11_20564	ATCT	58	TAGCAACACAGAAAGGAAGAA	TAGGAGCGTTTGTAGGAAGT	209	2	0.375
13	nksrssr11_20598	TTA	51	CGTTCGTTGTAATGGGTG	TATGTGTTTGAGTCCGTTT	338	2	0.375
14	nksrssr11_20836	AT	68	GCTTACGACATTGGATTATTT	CCCGCCGTGTTCTGTATC	341	1	-
15	nksrssr11_20954	AT	64	GCGTGATGAAATGGTGTC	TGCCAAGGAAATAAACTAAA	361	2	0.219
16	nksrssr11_21004	CT	53	CAGCCTTCTATTTCTTTCACA	ATCTGGGTCTTATGGAGGAGT	320	1	-
17	nksrssr11_21049	TATC	59	AACATTTCTGACGGTGAA	ACAACTCTCTCAATCAATCTAC	382	2	0.219

S. no	Primer Id#	SSR Motif	Repeat length (bp)	Forward primer	Reverse primer	Product size (bp)	No. of Alleles	PIC value
18	nksrssr11_21464	TA	63	ATCACGCACCAGCCATTC	TACATCCCATAGTAGCAGGGT	398	2	0.469
19	nksrssr11_21466	TC	68	GTGAAAGAAAGCCACCAAG	GGCACCAGCACACTACAC	268	3	0.594
20	nksrssr11_22326	AT	70	CGTTCACACTCTCACAAATAC	ACGCTACCATCAAGCAGA	397	2	0.469
21	nksrssr11_23235	TAT	58	TGTAAGTCTGTAACCATCCTACT	TAACATCGGAAGAGGTGAA	377	1	-
22	nksrssr11_23400	AT	70	TCTATCTCTCATCCAACTCCC	TTACAAAGAAGTGCTCCCTC	362	2	0.375
23	nksrssr11_23471	AT	68	CGTGTCAGTCATCCATCT	GGAATACAAAGTGGTCTGCTC	222	2	0.375
24	nksrssr11_23703	TA	58	CTTCTTTGTTTCGGTTCTATG	TCTACACCTTCAGTCCTTCCT	213	2	0.469
25	nksrssr11_23742	AT	70	AGAGAATCATCAAAGGACAAG	AATCACTACACAAACATACAAGA	400	2	0.375
26	nksrssr11_23806	AT	56	CTTGTCTTCTCATTCCTCTTG	ACACACACACACACACACAC	325	NA	-
27	nksrssr11_23806	GT	53	GAAAGTTGGTGGTGTAATGG	GCTGCGTAAGTTCAGATT	263	2	0.219
28	nksrssr11_23840	TA	69	GCAAGTCATAGAGATTGAGGA	GTTTAGGAGCCAGAGGCA	394	2	0.219
29	nksrssr11_23901	AT	64	TCTGCTCAAATGTTCGTGT	TTGTCTTGTTACTGTCCGATG	334	NA	-
30	nksrssr11_24084	AT	66	GTCAGTCATTCCTCTTCTG	CGAACCATCTTTAGTCCC	400	3	0.531
31	nksrssr11_24137	TA	53	ATTGACACTTCTGCTGCC	TAAGGATTAGCGTGGAGA	398	2	0.375
32	nksrssr11_24183	ATCT	68	TCCTCGTGTCATCGTGTT	TAAGAGCGTTTGTAGGGAGTT	272	2	0.375
33	nksrssr11_24467	AT	70	ACCTCTAAGTTGTATGATGGTAG	GCTTCTTCTTTCCTTTCTCTT	275	2	0.219
34	nksrssr11_24526	AT	56	CACCTGAATACTACCTCCGTT	AGAACCACACACACACACAC	389	2	0.219
35	nksrssr11_24626	TA	65	AGATAAATAAGGGAGAGCCAA	CTACTTCAACAAGAGAGGCAA	284	2	0.469
36	nksrssr11_24855	GAT	53	TCAGAGCGTGTCAGTTAGTTT	ACAAGATTTAGGTTTAGCGTG	281	2	0.219
37	nksrssr11_25076	AT	70	TGCCTTCTTTATCATCACATT	TTGCCATTTGTTCTTTATCC	388	3	0.406
38	nksrssr11_25271	TTA	59	GGGATTCTTCTTCTTGTGTTT	CAGGATTATGGATGGGTTT	285	2	0.219
39	nksrssr11_25327	TTC	65	CGACCTCCGAACACAGCC	TTCTTCTCCCAACCCTTC	328	2	0.500
40	nksrssr11_25773	AT	56	GAGGACGGCGATGATTGA	TACTAATGGACGGTGGATG	378	2	0.219

S. no	Primer Id#	SSR Motif	Repeat length (bp)	Forward primer	Reverse primer	Product size (bp)	No. of Alleles	PIC value
41	nksrssr11_25834	ATCT	76	ACATCTTTCTGTTTAGGGTAGT	GAGGTGGATTTGGTTGTTT	348	2	0.219
42	nksrssr11_26256	ATAC	75	ACTGCTCTGTATCTGAACCTG	AAATGTGTGTGTGTGTGTGTC	310	2	0.375
43	nksrssr11_27170	TA	69	TATTAGAGAAAGGGCGAAGAA	AGGGTCAAGAGACACACAGA	256	2	0.469
44	nksrssr11_27772	ATAG	66	CAGTGAATGTGGTGCTTT	AAGTGAGGCAATAGAGATTAG	384	2	0.219
45	nksrssr11_27873	CTAT	53	AATCAGGAGAAAGAAGAGCAG	GTAGGAGCGTTTGTAGGAAGT	214	NA	-

Primer Id, e.g. "nksrssr11_19009" includes lab Id (nks), rice simple sequence repeat (rssr), chromosome number (12), and base pair position in the rice chromosome pseudomolecule build 4 (19009 kbp).

Annexure VIII

List of 832 potential hypervariable SSRs in the repeat length range of 51-70bp mined from rice genome.

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
1	HvSSR01-01	TATC	60	20117	CAACCAATTCGCTATGGTAT	CCAGTTTAAAGTTTCGTTCG	344
2	HvSSR01-02	AAC	67	1210258	TCAATCCTATGTTCAATCCC	GCAGTCGACATAATGCATAC	298
3	HvSSR01-03	AT	52	2523641	AAGGAAACGTGTTAGTGGAA	GTAGATCTCAAACGTTCCCA	302
4	HvSSR01-04	TAA	57	3800284	GCGATCGACATAGTTACACA	AGAAGACCGAGACGTAGATG	337
5	HvSSR01-05	AT	63	4191380	GAATAGCAAGCATATACGGG	TTTACAGGGAAAGCGAGTAG	218
6	HvSSR01-06	AT	62	4219456	CCCATTCTAGCTATTTGCAT	TTTGATCATTACAGTTTGCATT	366
7	HvSSR01-07	TA	65	4351318	ACGGGAGATTCACACATAGA	ATTATGGTATGCTTTGCCAG	136
8	HvSSR01-08*	CT	63	4392323	CATGACCAACGAGGTGTT	CAGAACACGAGGAGAATGAG	354
9	HvSSR01-10	TA	57	5691062	CAGGTAGGCACTCCATCTAC	ATAATTAACACAGGCGCTTC	292
10	HvSSR01-11	AGA	62	5799229	ATTGTCCAACAAGTACGCTT	AAGTAGAATCAACTCGAGCG	319
11	HvSSR01-12	ATT	69	5939992	TACTTTCTGTCTCCCTCCATT	TTTAGATGACTCGTTAGGTTCT	359
12	HvSSR01-13	AT	70	6011667	ACGACCATCTGCTTTACATT	ACAAACAATGCTAGAATGCC	344
13	HvSSR01-14	CT	69	6153806	ATAGACGCCTCCTCCTTATC	AGGACCGACATGCTATTAAC	183
14	HvSSR01-15	TC	53	6217731	AGGTAGTGTCCATGTGGAAG	TCTTACATGCTCTTGCATTG	250
15	HvSSR01-16	CT	60	6271294	TCAGATGTCAGAACCAAACA	GAATCAACCTCACGAACTGT	385
16	HvSSR01-17	AT	68	6462873	CTCAACCTCAGTGGAAAGAA	ATAATCACTAGCCAGCAGGA	392
17	HvSSR01-18	TA	58	6514581	TTTATTCCTTGTGGATGGTC	TCCTTGTACTCATTTCGTCC	327
18	HvSSR01-19	AT	55	6995278	ACTGTAATGTGGCGCTTAAT	CCCTAATCACTCACTTCGTT	318
19	HvSSR01-20	AT	60	7057671	CCTGATATGGACGAATTGAT	CTCGAATAGACGTCCACCTA	259
20	HvSSR01-21	TA	53	7243945	TACCGTTACGCCTATTCAGT	AGGACATATGTGCAGCTTCT	194

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
21	HvSSR01-22	TA	52	7281494	TTTACCAGCTCAAACCTAGC	TTAGAAACGGAGGGAGTACA	247
22	HvSSR01-23	TTA	60	7867371	AGACGATCATACGAGACACC	CGTCCGGTTTATATAGTTGG	325
23	HvSSR01-24	AT	61	8128197	ACGCAGTTACACCCACTACT	ATAAGTGCCTATGCATGGTT	252
24	HvSSR01-25	AT	65	8134745	CTTGCATATAATGCACAGGA	ACGTGTGCACTAACTCATCA	298
25	HvSSR01-26	CTT	60	8326760	GGCTTATGTAGTATTGCGCT	AACGTAGATCCCAGCAGAC	345
26	HvSSR01-27	TA	63	8638050	CTACACTCCTCCAAGTCCAA	CAACCACCACTACGTTCTCT	395
27	HvSSR01-28	TA	58	8675717	ACCTTGCCACCAGTAAGTAA	AATCCGAACATCACAAGAAC	287
28	HvSSR01-29	TA	59	8770337	ATTGCCTCTTCATCGAGATA	ATTAAGTTCAGCCAAATCCA	355
29	HvSSR01-30	AT	63	8913798	ACGAATCCAAGTTTGATGAC	TCAGTCGTTCCTCTCACTTT	383
30	HvSSR01-31	AT	55	9433640	ATGGTTGTCGCTAACTTTGT	ATGAGAATCAGTTGTGGAGG	303
31	HvSSR01-32	TTA	58	9461426	GATATCATGGATATGCCACC	CCATGGACAGATAAAGGAGA	392
32	HvSSR01-33	TA	63	9487756	AACTTGGGCTCTTTAATTCC	CAGAGTTCGAGAGAGACCAG	309
33	HvSSR01-34	CTT	57	9603525	AAACTGGAGATGAACTCGAA	GTAACGAACTAGAGCATGGG	250
34	HvSSR01-35	ATCT	66	10438217	GAGAAAGTACGTGTGTGCAG	AGAGATTTGGATTTGGGATT	400
35	HvSSR01-36	TA	53	10448489	GTTGGTAAGATGGCTCAGTC	GGGATTTATCAACGAACAAA	388
36	HvSSR01-37	AT	66	10565611	CTTAGTTCGGGCACAATTAG	GGTGACTTCGTCAATCTCTC	304
37	HvSSR01-38	AT	57	10696194	CGTAGCACAAGCTGAGATTA	ATCTAGAGACTGCCAAATGC	313
38	HvSSR01-39	AT	54	11172388	GGGAAGAGGAAATATGGATT	CTAGGAAAGAACCGCATAAA	329
39	HvSSR01-40	TA	68	11392583	ATGTTGATTAGCAGTTTGGC	TGGACTGGCTGACTTAGTTT	305
40	HvSSR01-41	TA	60	11901479	GTATCCCGTACGTATCCAAA	AATAAGTAGCGCCTTAGGGT	265
41	HvSSR01-42	AT	52	12919076	GATACATGCAAATCATGCAG	TGGAAACAGATGAATGTTGA	332
42	HvSSR01-43	TAA	61	14030027	ACACCTCTTAGCGAACACAC	AGAGGAGGCTTCTTCCTTC	344
43	HvSSR01-44	GT	55	14235715	TGAGTGAGACTTGACAGTGC	AGTTAACACCAATGCTGACC	348
44	HvSSR01-45	CT	56	14254308	TCAATGCAACTACTGCAAAG	CATTCTTTATCCAGCCTACG	369
45	HvSSR01-46	ATA	58	15874214	TATTTGATGGGTTGTCTTCC	TGGTTGATCCTATCCATTTC	287
46	HvSSR01-47	TTA	61	16574275	GTGTGTGTGATATGGTGGAA	TGATAAACCAGTGCAATGAA	266

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
47	HvSSR01-48	AAT	67	17608692	GTTAGCGCTTCTGGGTATAA	ACTGCTATAGCTTGATGGGA	342
48	HvSSR01-49	AT	54	17842739	TCCTAAAGTTCACACCAACC	TGTCGATTCTCCTTCACTTT	348
49	HvSSR01-50	TA	60	18136284	GACACTCCGGTAAGTTGAAA	TCAAGACTATTTCTAGCGCC	292
50	HvSSR01-51	AT	64	18804769	ATACACCTAGAGGCCTGGAC	TGACGTACCTAACCACACAA	276
51	HvSSR01-52	TA	65	19404529	CTTTCCCTCCTCTTCAAACT	AGCGACCTCAGATGAACTTA	364
52	HvSSR01-53	AT	60	19715245	CACAAGTGGCTGACAAAGTA	TCCCTTCACATGAACTTCTC	374
53	HvSSR01-54	TA	69	19831188	CACAACAGAACAATGCCTAA	TTGGCAACATGAATTGTAGA	292
54	HvSSR01-55	TTA	56	20979191	ACACCATACCAATACGAAGG	ACACCGTACTGTTTATTGGG	288
55	HvSSR01-56*	AGA	57	21384304	TGTCGTCCACGTAGTAGGAG	ACACTCCTCCTCTGTTCTCA	274
56	HvSSR01-57	AT	58	21554178	TCTCGCAGATTACGTGATTA	GAAAGACGAGTCTACAGTCCA	293
57	HvSSR01-58	AT	56	23483194	ATTGATGTGTGAGAATGCAA	GACATGACATTGGTGTCAGA	321
58	HvSSR01-59	TA	62	23536962	GTGTCCTATCAGGTGAGGAA	ATAGCTACTCCCTCTGTCCC	254
59	HvSSR01-60	AT	64	24364674	GGTTCTGAGTGGATCACCTA	ACAAGTATCACGGACCTCAT	260
60	HvSSR01-61	AT	58	24735132	TAAGAACATCCAATTCCACC	TTTCATCTTGGGTTCGAGTA	149
61	HvSSR01-62	AT	68	24854187	CATGCAACAGTAGAGAAGCA	TATGGTGCCACATACATACG	216
62	HvSSR01-63	AT	67	25217961	CTATTCCTTATGCGACGAAG	TGACATATTTGCAAACGAAA	336
63	HvSSR01-64	AT	68	26203496	CCCTATCTAAGCACGATGAC	TTTAATCCACGGTCAAGAAT	269
64	HvSSR01-65	TTA	60	26813771	TTTCTTTAAGGTGGTGGAAA	GAAGACCGTAGTAGCAGTGG	348
65	HvSSR01-66	AT	68	26840612	CTCTAGCACGAATCACAACA	TGAGCAACTCCCTCATTTAT	354
66	HvSSR01-67	ATA	66	27838340	CGATAGGCTAATCAATCCAC	GTAGTGGGAAGATGTGTGGT	286
67	HvSSR01-68	TC	55	27887991	TGACATACAACAAATGGGTG	CTAGCTAGCTGTTGCACGTA	324
68	HvSSR01-69	AAG	54	28112690	GACTCACTCGTCTCGTGG	CTGGCATCAACTTCTCATT	290
69	HvSSR01-70	GATA	63	28461969	GTCGATGAGTGGTTGAAGTT	GCACAAATATCGCCTAATTC	346
70	HvSSR01-71	TA	61	28642362	GCAGCTCAAATAGCAAGAGT	CAACTGCCAGCTATCATACA	293
71	HvSSR01-72*	ACT	61	28862962	GCTCCTACGACAAGGACTC	TTCTGGGATCCTAATGCTAA	377
72	HvSSR01-73	AT	57	30220581	TTTGATCAACTACCCTGTCC	GAGAAGTTTGAGATGCGTTT	223

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
73	HvSSR01-74	AT	59	30555611	CAGTTCTCCAGCAGTGTGTA	GAACATCGATCACACTCCAC	275
74	HvSSR01-75	CT	69	32224855	TCATCTTACCTTCCCTTGTG	GAGAGCTAACCATGAGCAAC	194
75	HvSSR01-76	TA	68	33255431	TCTAATCCAATTGGTGTTCC	AATCAAATGGTCTTCAGTGC	394
76	HvSSR01-77	AAT	53	34098875	ACGGATGATGACATGGTATT	TTAACCATGGTCTTGCCTAC	286
77	HvSSR01-78	TA	54	34453153	AATACCAGACGGAGATGTTG	TGCGTTTCCTCAGAATAAGT	325
78	HvSSR01-79	TA	67	34764634	ACCAAACTCCTAACTACCGTT	TTTATTGGAGCTTGTTGACC	254
79	HvSSR01-80	GA	54	34875822	TTTGAGCAAATAAACTTGAGG	GCTTCTACTTCCACAAGGC	258
80	HvSSR01-81	GA	53	35189963	AATATCTGTTTCGCCAGATT	TTGGGAAAGTTGTGTTCTTT	153
81	HvSSR01-82	CAA	59	36481467	CGGTCGATTATGCAGTCTAT	ATCGGATTGACAGATTGAAC	254
82	HvSSR01-83*	TGT	60	37910904	AATGTATGGGAATATCGTGC	TCTTCACCACATTGCTGTTA	384
83	HvSSR01-84	GA	57	38224005	AGGAACAGCTACACCAGAAA	ATTAGCCAAAGCATCCAAAC	398
84	HvSSR01-85	TA	60	38412916	AGCACTACCGTAATTACCCA	TAGGTTGCGTTCTTTCTAGC	314
85	HvSSR01-86	AT	58	38532908	CTCCCTACGAGTTGATTGAC	TCCGATCTTATTTAGCAACC	224
86	HvSSR01-87	CT	70	38943138	TTGGTACACGACCATGATTA	ATGGATCTGTGTGTGCGT	374
87	HvSSR01-88	CT	65	40150971	TCTCTCGTCGATCTTAGCAG	TCGTTAATGAAGTCGTTCGT	232
88	HvSSR01-89	AG	62	40580604	TGCGACGGATAGAGTACATA	GGATGCAAAGAAAGAACAAG	359
89	HvSSR01-90	AT	54	40628636	CCACGTACGCACACATATAC	CCAATCACAACCATTCTTCT	343
90	HvSSR01-91	CT	64	40779925	GCTCTTCGATTCACTTCATC	AGCAGAAGGTAACATGGAGA	351
91	HvSSR01-92	AT	70	40863900	ATATCATGCTATGCTGGGAC	AATTTCGGCATCTATATCCA	382
92	HvSSR01-93	AT	53	40898183	CAGATAGAATCCCATAGCCA	TTGCTCTTCCTCTACCTGTC	387
93	HvSSR01-94	AT	63	41080787	AATTAAATGCTGAAGTGCAAG	AATCATGGAGGTAATGCAAC	334
94	HvSSR01-95	AT	70	42019937	TTATGTGAATCAGCACCAAA	TCAAATAACCAATGATGCAA	341
95	HvSSR01-96	AT	66	42141523	GTCCTTTGTTTCTGAAGACG	CCTTTCTTTCACTTGATTGG	376
96	HvSSR01-97	TA	52	42400448	TTTATGGGAAAGACTCGGTA	GGTATGTAATCATGGATTGTGA	280
97	HvSSR01-98*	GAA	69	43499662	TAGTCTTCTTCTGCTGCTCC	ACACTACCAAGACTCCGCTA	313
98	HvSSR02-01	TCT	63	124679	AAGAGATGAGAAGAGCAATGA	CAACTTAGAGGAAGAAGGAGG	312

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
99	HvSSR02-02	AT	60	829563	ACCACATAATGGAACAGCTC	AGGCAATTAGTCGTCAGAGA	326
100	HvSSR02-03	TA	57	1325817	AAATTCACGACAGAGTCCAC	TTTAAGCAAATGCTGTGATG	158
101	HvSSR02-04	TA	57	1693222	GATACCAACAGGAGAACCAA	TTTGAATTGATTGGGATTTC	310
102	HvSSR02-05	TA	69	1705077	GCACAACATTAGACCCTGTT	TACTTGCCCTAAGATTCGTG	333
103	HvSSR02-06	AT	60	2775149	AAATGACATTACTGGATGGC	CTCTTGATTTGACGATCCAT	301
104	HvSSR02-07	AT	62	2798194	ATCAAATGTTTCAATCGTCC	TACTTCCTCCGTTTCACAGT	303
105	HvSSR02-08	AT	63	3570963	CTAGATGTCGAAGACCCTTG	TTTGTGCATTTGAAACCAT	176
106	HvSSR02-09	TA	68	4155526	AAATCATTAGCATACGGAGG	TACATGGCTTACCTCGATTT	288
107	HvSSR02-10*	AT	52	4297183	GAACTGGAGTTGCAGATTTC	GTTCATGATGCTTGTTGCTA	259
108	HvSSR02-11*	AT	53	4355427	GACCATGTCAGCGAACTC	AGGTACTTCCTCCGTTTCAT	180
109	HvSSR02-12	GA	59	4414124	TCTCCAATTCTCCATCAAAC	CTTGCTTGAGCGAGTCTAAT	329
110	HvSSR02-13	AT	60	5085817	AGCTTCAGGCTTCTTCCTAT	TTCATAGTCATTTGAATTGGG	284
111	HvSSR02-14	GAA	57	5477565	CTTTGAGATTGATCGAGAGG	ACGGAATGAGCAGTATCTGT	292
112	HvSSR02-15	AT	68	5957808	CTAATTGTTAGCGGCATTCT	CGTCGACAAACGTATGATTA	398
113	HvSSR02-16	AT	56	6322800	AACTAGTACTCCCTCCACCC	ATAAGATGGGTGGTCAAACA	345
114	HvSSR02-17	TA	53	6580532	TTTGACTTTGCATATCCCTC	CCCGTAGGTTCTACATAAAGG	336
115	HvSSR02-18	TA	62	6696412	AGGTGAAGACTCAGAAGTGG	ATGTAGCCCATGTAGGATTG	333
116	HvSSR02-19	AT	66	6746831	CTTTCTTCGTCACTCACCTC	TATATTTGGATTCCGTTCCA	254
117	HvSSR02-20	TA	67	6858564	CCGAGCTCAGAGAAAGATAA	ATGAGTTTGTCAGGGATGAG	352
118	HvSSR02-21	AT	56	7074066	CAGCATACGCATGAAAGATA	GTATGGTCTCCCGAATGTTA	357
119	HvSSR02-22	TA	70	7347730	CCGCTCAACAACGTACTAAT	ACACATTCGTATTTGAGTCG	315
120	HvSSR02-23	TA	53	7367388	AGCTAGCTACACACTTTCCG	ATAGATGCATGGCGATATTT	364
121	HvSSR02-24	TC	62	7706971	AGTCTCCACCTCCTCCTC	ACATCTCCGTCAGAATCAAG	351
122	HvSSR02-25	AT	69	8048434	GACTCCTGCATTGTTGAGAT	CTAATCCCGTCGTCTAAATG	329
123	HvSSR02-26	AT	67	8105992	GCTACGAGAAACGTACCAAC	TCAGTTCAAACTTGCTTTCA	303

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
124	HvSSR02-27	AGA	67	8186014	GGTGCAATTCTTATTCCTTG	AATTTGGATTCAGATGTTGC	233
125	HvSSR02-28	TA	53	8223488	GAAGAGGTCTAATCGGGTTT	GTGGCGTCATATATTTCCAT	234
126	HvSSR02-29	TA	58	8816783	GAAAGAATGAAATGGACCAA	CTACTCTTGGTGTGGTGGTT	334
127	HvSSR02-30	AT	59	8910628	AGTGTTCTCTCCTCCCTCTC	TCAATTTCAACAATGGTGC	346
128	HvSSR02-31	ATT	60	8981479	CGTTTGTCCCAAATAAGTTT	ACATTTCTATCGTTTCGCAT	344
129	HvSSR02-32	ATAG	59	9769607	AGCTATCTACGTCTTGCGTG	ACAGTTACAAGCACGGATTC	390
130	HvSSR02-33	TA	52	10098043	GTGCTTCTCCTCTCCCTCT	TCTTGTTACTGCACGATGAC	323
131	HvSSR02-34	TA	65	10289242	TAATGCACGCACAACTTTAC	TATAGAATGCTGACTGGGCT	355
132	HvSSR02-35	AT	70	10755117	TGGTTGCTCGTATTCACATA	TTATCCAAGAATTGTCCCAC	347
133	HvSSR02-36	ATAG	52	11245857	ATCACTTGATAGCTTGGGAA	TCCTGCTATAGGCTAGGTTG	350
134	HvSSR02-37	AT	68	11341279	AGATGGCCACTAGTTTGAAG	CTTGCATCTCTAACCTTGCT	365
135	HvSSR02-38	TA	57	11674071	GCAATGTAGATAGGACGAGC	GTTTCAAACGTGCTATCCTT	289
136	HvSSR02-39	CT	57	12517480	TTGGCCATATCTCCTATCAC	AGAAGCGACGAGGAGAAG	367
137	HvSSR02-40	TA	61	12938773	TGCTAACTTGACATGACTGC	CAACGGGTAAAGGCTATTT	186
138	HvSSR02-41	AT	64	13204599	CAAACGGTAAGAAATTCAGC	TGGACCAGAAGGTTAAGAGA	393
139	HvSSR02-42	AT	59	14328858	GTAGGTGGATGAGCAGATGT	GATGTGAGCGCCGAAGAC	210
140	HvSSR02-43	TA	65	14496187	GTTGGAGGGAGGGAATATAG	ATCCACTACCACCACATCAT	258
141	HvSSR02-44	AT	57	14618171	TACGGTTTGGAAGAGATTGT	GAAGGGAGTTGCTGATGG	358
142	HvSSR02-45	AT	64	15313278	TAGAATGTGGTGCATGTGTT	TTTCTCAAATCAAATTTCGG	397
143	HvSSR02-46	AAT	68	16960060	TCGATACCAGCTACCAAAGT	TGGTTACCATCCTCCTATTG	309
144	HvSSR02-47	AT	68	17089534	CACACATCCAGATGAAACAA	TATTATGGAACGGACTCCAC	303
145	HvSSR02-48	TA	70	18177461	AGAGAGGAGATCAGCAGTGA	AAACTTTATGCGCAATTTGT	338
146	HvSSR02-49	TA	52	18529839	TGCTACCACATGGACAATAA	ACTTCGTATGCGATTCTGTT	396
147	HvSSR02-50	AT	54	18805529	AAGTAAACGAAGCAACCAAA	CAATTAGAGGAGGACAGGTG	319
148	HvSSR02-51	TA	56	19059233	TTACTGCGGAAATTAGTGGT	CAAATTCAGTTCAAACTTGCT	369
149	HvSSR02-52	TC	52	19182948	TTTCAGGAATCTGATGCTTT	TTAATCAAAGCCCTAACAGC	195

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
150	HvSSR02-53	AT	54	20259793	ATCGGATGGAATTTGTACTG	ATTTGCTTTGACTTGGATGT	204
151	HvSSR02-54	TCT	59	20980286	ATTTGATTCTTCTTCCTCCC	TGTTCTTGGCTTCTTGAGAT	207
152	HvSSR02-55	TA	57	21760749	ACGAGGTACACGCAGTAAAT	TAAGGAATCGAACGAAGAAA	334
153	HvSSR02-56	TA	61	21841720	TCACTTGATGGTACACCAGA	AGCGACACGGTAGTATGTTT	327
154	HvSSR02-57	TA	53	21876111	CCATCAATTTATGTGCATGT	ATATCATGGGTCCCACCTA	337
155	HvSSR02-58	AT	57	21968355	GGATCTCCTGTCGATTCATA	AGTTCGTTAAGCAGCAAGTC	299
156	HvSSR02-59	AT	56	21987737	TACCTTAATACCCTGGCTGA	AGAAAGTACTCCCTCCGTTC	310
157	HvSSR02-60	AT	57	22086749	GGCTTTGAGGAATCAAACTA	TTAACACGGAGCTTCCTTAG	220
158	HvSSR02-61	AT	53	22144733	GGTTGAATGTGGCATTTACT	TAGATGGTAAACCTGCCACT	328
159	HvSSR02-62	AT	63	23638618	ACTCAGAGCCATTAGCCATA	TCCTGTGTAGCCCATAGTTT	322
160	HvSSR02-63	TA	52	23891390	TCAGAACCAATGGCTAAAGT	CAAATTCAGTTCAAACTTGCT	248
161	HvSSR02-64	CT	56	24016627	GGTCGTCACTCGTCAGTAGT	ACGAAACAACACACAGCATA	370
162	HvSSR02-65	TTA	67	24827435	CATTTCATGTTTATGGCAGA	CCTAGACAATGGGACAGAGA	363
163	HvSSR02-66	AT	66	25313160	TAAACAAACTACCGAACGGA	ATTGTCCACTGAAATGTTCC	371
164	HvSSR02-67	TA	52	25620251	GACAGCCACTATCACACACA	TCAGTTCAAACTTGCTTTCA	293
165	HvSSR02-68	TC	69	26102578	GAAGATGACAAGGTGAGAGC	TTACACTAGCGATCACATGC	372
166	HvSSR02-69	AT	69	26964776	TACTCACTCCGTGCTCATAA	CATTACTCCCTTCGTACTCG	310
167	HvSSR02-70	CT	54	27059497	TGCTACTCCTAGCTGCTACC	TTACAGGCGGAATCTGTAAT	298
168	HvSSR02-71	GA	53	27077020	CCAAATTCACTCTCCGATAC	ACAGGTAGCAGAGCCAGAC	388
169	HvSSR02-72	AT	58	27305902	ATGTGTTAGACATCATCGCA	AAACAACGATTAATTTCGGA	140
170	HvSSR02-73	CT	57	27315018	ATACATGCATACTCCGATCC	ATGCTCGAGTTGTGAAAGAT	199
171	HvSSR02-74	TA	52	27425505	TTTGTGTTGTAACGGATGAA	TATTTGATATCATGGTGCCC	275
172	HvSSR02-75*	TA	63	28209262	CGTAATGCAGACCATTAACA	GACATGCTGGTTCCAATACT	196
173	HvSSR02-76*	TA	70	28534808	CGGAATTTGTCTTAGTGAGG	GGCAATTTGTATCAGGACAT	297
174	HvSSR02-77	AT	59	29046964	ATGATGAGTGAGGGAGTACG	TATGTACCATTGTCCAAGCA	247
175	HvSSR02-78	TA	54	29497793	GTTTCCTTGCAAACAGACAT	AGTCATTCTAGCATTTCCCA	306

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
176	HvSSR02-79	AG	59	30052384	ACAGCAACGACAAGGAGATA	GACAGACACAGCGACTCC	333
177	HvSSR02-80	AT	55	30890059	TGATGCCCTGTTCTTAGAGT	GCAGAAAGAATCAAACCAAG	210
178	HvSSR02-81	AT	61	32250183	ATTTCATTGGTGTGGCTAAG	GTGTGGCCTTTGTACTGATT	334
179	HvSSR02-82*	AT	59	32693711	TGATGGATATAGAGCGACCT	AATATGTTTCATCAAACCCG	363
180	HvSSR02-83	AG	54	32774391	GATTTCGTTCAAGGTTCAAA	CAACAACACTATAAACGCCA	310
181	HvSSR02-84	ATCT	66	32994680	CCGATTGATCGGAATATCT	GGAGCGATTGTAGAAGTGAC	229
182	HvSSR02-85	GA	59	33677017	TGATCATCTGCTGCTATCTG	TGCCTACTTTACCTAGCTGC	141
183	HvSSR02-86	ATAG	60	34293946	TAGATAGACTGCTCGCACAC	TGATTCAGCGTAGGAGAACT	339
184	HvSSR02-87	AT	55	34329071	GCGGGACGAAGTAGATTT	CTTCGGCACAAAGGTAATC	355
185	HvSSR02-88	TA	59	34976735	TCCACAATTGTTTGTTCTCA	GGGTCAAACCATTCTAACAA	348
186	HvSSR02-89	AG	58	35369568	AACCATGAAATAACCATTCG	CGTGTTCTAATCCAGCAACT	313
187	HvSSR02-90	AAC	68	35495748	CTGCATCAATATAATTGCGA	GCTACTTACACCACCACCAT	280
188	HvSSR02-91	TA	64	35538132	GCTTGAGTTCATTTCAGCTT	CGTGTAAGGATGAGAAGAGG	394
189	HvSSR03-01	AT	64	473813	TAGTTGCGTTACAAATCGTG	GATGAATGAAGGCTAGGTGA	325
190	HvSSR03-02	AT	52	1019485	TAGCGGAGTTGGAATAACAC	CTGCACTGCATACCTCATAA	228
191	HvSSR03-03	TA	56	1375577	GGTTAGTGCAATCACGTACA	ATAGTGAGGCACACACATGA	325
192	HvSSR03-04	AT	65	1966825	ACGAATCTAAGCCCACTACA	TATTATTGCTCCACCAACCT	295
193	HvSSR03-05	TA	62	2286723	CACATTTGTGCTGATATGGA	CAAATTCAGTTCAAACTTGCT	366
194	HvSSR03-06*	GT	56	2793119	AGATGAGCTTCAGTGCTAGG	TTCACCACAAAGTTCACAAA	318
195	HvSSR03-07	TTA	56	2848623	GCCTGAATTCAACGTGTAGT	GAAGATGTGTAAGAGGGCAG	299
196	HvSSR03-08	AT	52	3433566	CTATTTGGAACAACGGTAGC	CAAATTTGTGTTGCTGCTTA	331
197	HvSSR03-09	AT	62	3797148	TGTATTCAAGGAGGGCTAGA	CAACTGTTTCCTGGAATGAT	204
198	HvSSR03-10	GA	56	3914049	GTACACAACGTCACAACAGC	ACTGTGGCATATGTTCGATT	280
199	HvSSR03-11	AT	68	4549683	TTACGAACCGGGACTAATAA	TCGTAGCCTCCCAGATAATA	207
200	HvSSR03-12	GA	60	4928103	TTGTACTTCCTTCCCTGAAA	GTCAGGATGTTACCTAAGCG	275

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
201	HvSSR03-13	AT	56	5173505	CTTGTCCTTGGGATAATTTG	CGTACACGTGTGGATTTCTA	361
202	HvSSR03-14	AT	70	5218941	AATCTGTTGCCATTTCTT	TGACCACCTGACTATGATG	355
203	HvSSR03-15	AT	70	5216525	TGACCACCTGACTATGATG	AATCTGTTGCCATTTCTT	**355**
204	HvSSR03-16	AT	54	5374714	ATGTTAGCTACCACCACCAC	TGAAATATTGGGTGTCACAA	347
205	HvSSR03-17	TA	57	5482383	CCAGTGTGTTGGAGTGTATG	TCACTCCGTGTTCATATTCA	290
206	HvSSR03-18	GA	58	5840048	TAAATGAAGCACACGATACG	CACGGATGTAAAGAAGAACC	366
207	HvSSR03-19	AT	55	7221114	AATTCAGTTCACGCATTCTT	AGCTGTTCGTCTGCATAGTT	238
208	HvSSR03-20	AT	57	8056351	GCAAAGGAATCAAAGCTAGA	TTTGACCGTCCAGTTATACG	282
209	HvSSR03-21	GA	59	9203435	TGCCTTATACCTGCTTTGAT	CCAGTTGTTTCATCAGCTTT	259
210	HvSSR03-22	AT	70	9211778	TAATCGCAACAAGCACAATA	ACTGCAACCATATAGAACGG	383
211	HvSSR03-23	TC	57	9921327	AGTCGTTGAGGTCTTTGAGA	TCTCTTCAGTTTCTGTCGGT	197
212	HvSSR03-24	TA	55	10025111	GAGAGGACTACGGTTAGGCT	TTCGCTGAGATCTACAACATT	382
213	HvSSR03-25	AT	62	10149231	ATGATTTGACTCCCTCTGTG	GAGGAAGAAGGAAAGGTGAT	253
214	HvSSR03-26*	CT	61	10799063	GTCGTCGATGTCCCTCTC	CTGCTCCAGCGACTATTAAA	208
215	HvSSR03-27	TA	59	11419279	ACCGACTACTGTAACTGCGT	TCCGATATCTCCTATTTCCA	387
216	HvSSR03-28	TA	53	11644548	ATTTGTAAGGTGAGGTGTCG	GGGTAATTCAAATCAGGG	171
217	HvSSR03-29	AT	54	11644548	AACTACTAGCTGAATGCACG	CGACACCTCACCTTACAAAT	223
218	HvSSR03-30	AT	54	11644731	TGCACGTACTTTAGTACAGATAG	GAGGGTGAAATATGTCTGGA	393
219	HvSSR03-31	TA	53	11652469	GGG TAT TTG TAA GGT GAG GTG	AAC CGA GAG AGA GAT GTG TG	365
220	HvSSR03-32	TA	61	12085563	GCCTGGACACATCAGTC	CATAGTCACCCTGCTTATCCT	238
221	HvSSR03-33	AT	67	12749863	CACTGTGTGGATGTGGAATA	TAAGCTTTGACCACACCAAT	262
222	HvSSR03-34	TAA	55	13261306	TCTTAGTTTCTCTCGGTCCA	ACTGAAAGCACCGTTAAATC	299
223	HvSSR03-35	AT	66	13867569	TTGATTACGTGAATAGCTCG	CATAGCTAACTTGTGCGTTG	358
224	HvSSR03-36	AT	63	14267216	AATTTCATCCAAACATACCG	ACATGCTTCCTCTCAAACAC	378
225	HvSSR03-37*	TA	53	14459750	GGAAATCGTCAAGAACGTC	TAATTGTATACCACTCCGCC	386
226	HvSSR03-38	TCTA	57	14678937	TTTCAATCCAGAAACCCTAA	GAAGTTCAAAGCGAGAATTG	328

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
227	HvSSR03-39	CTT	60	15117523	TGACTCGGTCTTAGAATCGT	ATCATTCATCCAAGTCCAAG	260
228	HvSSR03-40	AT	70	15363219	CGAGAGGTTCAGAGAGAATG	GCATTCACCCTAAGGATACA	351
229	HvSSR03-41	AGA	67	15370888	ATGCAATTACTTGTTGCCTT	AAGTTCTGAACAACCACCAC	203
230	HvSSR03-42	TA	52	15505710	TTGTGAACCAAGCTAATCCT	TCAGTTCAAACTTGCTTTCA	349
231	HvSSR03-43	AT	65	15551486	GAATTCCTCAAGTCCTCAGA	TGCTAATATAAGTGCGACGA	282
232	HvSSR03-44	AT	57	15563520	GTTGGTGATGTCTCTGGAGT	ACGGATTTGCAGTGTACTTT	361
233	HvSSR03-45	TA	53	15838207	ATTGATGAAGGACTCCATTG	ATGAATTTGAAGATGCGACT	373
234	HvSSR03-46	AT	56	16581422	CACACACGTGACAAACTACC	TCTTCAATGGAAATGATGAA	387
235	HvSSR03-47	AT	56	16881061	AAACTAGGTGATTCTCAGCG	ACAAACATGTCACCTCATCA	398
236	HvSSR03-48	TA	52	17249758	TGGATTAGCTCCCTCAATTA	CGCCTAAAGATGTCCTTCTA	199
237	HvSSR03-49	GA	52	17413560	ATATCTCGTTACCCAAACCA	GCAACTAGAGCGCCGATA	356
238	HvSSR03-50	TA	57	17864550	GCAGGTACATCTTTGATGGT	TGGCTGAAATAATCAGGACT	335
239	HvSSR03-51	TA	59	20756767	GACTTGATGGGAAGATTTGA	AAACAGTGCATCCTTCTTGT	332
240	HvSSR03-52	AT	62	20970759	ATAAAGGACGAACATGGAGA	TCAATTTCAACAATGGTGC	339
241	HvSSR03-53	TA	54	21146449	CTCCTTGCTTCCTTACTTCA	CAAATTCAGTTCAAACTTGCT	312
242	HvSSR03-54	AT	56	22408719	CGCATAATAAGCCTGAAACT	GTTCGAAACCTGATTCTTTG	349
243	HvSSR03-55	AT	68	22446846	CAAACTATTAATGATGGAGCG	CCATTGACCGGAGCTAGTA	398
244	HvSSR03-56	TTA	56	23080448	GCCTATCAGGCTATCATCAC	GTGATCGACATTGAGGAGTT	352
245	HvSSR03-57	AT	53	23393951	CCTTCAGCTATCTCGGTGTA	CTTAGCAGTTTGCTTACCGT	269
246	HvSSR03-58	TA	62	23625306	TCACTACTAGCATGGAGCCT	TCATAAGTGGAGAGGTTGCT	385
247	HvSSR03-59	TAT	69	23833219	TTTCTCTTGAATCTGCCACT	ACAGTTGTAGGTTTCCGTGT	393
248	HvSSR03-60	AT	67	23948943	CAGACTGCTGAGCTGGTAAT	GCCAAGAGTACTACAGGCAG	186
249	HvSSR03-61	AT	68	24185968	GACGGTCTACTGTACGGTTC	AAGACACAATTGATTGGACC	388
250	HvSSR03-62	AT	68	24198386	CTCGGTCGTACGTATTCTTC	ATGCATGCCCTAGCTACTTA	339
251	HvSSR03-63	TA	60	24780861	GTGATCGGGTTAATTAGCTG	ATACAAAGCTTCCCAAACAC	386
252	HvSSR03-64	AG	56	25068838	ACATGGCCTTGTAGTAGACG	GAAGGAATCCAATGTGTGTT	300

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
253	HvSSR03-65	AT	65	25153211	GAGAGAGATGCTTGAGATGG	CAAGAAGAAGTGGGAAACAG	280
254	HvSSR03-66	AT	52	25314097	TTAACTCATTCCATCCTGTGT	AGTTTCGCGTGTCCTAGATA	249
255	HvSSR03-67	AT	57	25444406	TACTCTACCCACCCTTGAAA	CCAAACTGACCCTACACACT	205
256	HvSSR03-68	TA	53	25482068	CTTACGTACAGTGAACAGCG	TACTAATTCCGATCCCTGAA	291
257	HvSSR03-69	AT	53	26103389	GGTATTGTAAGCATTGCACA	CCTAACCTCTAAATCCTGGG	350
258	HvSSR03-70	AAT	69	26207737	CTCTATCAATACAAAGGCGG	CCTGTACTTGCTTTAATCGG	208
259	HvSSR03-71	GATA	51	26474919	CACACCAACTCACTCTTGAA	CCGTTTCGTCTATGTTCATT	396
260	HvSSR03-72	TAT	61	27272787	GAATTCATGGAGGATGAAGA	TTTATCGCGTAAACACAATG	291
261	HvSSR03-73	TA	62	27435768	AGATGACCGTGAGAGAGAGA	ACGAATTACCCACATGAACT	297
262	HvSSR03-74	AG	63	27658912	GGCCATACAACATTCAAGAC	CTCCTTGTCCACATTCATCT	359
263	HvSSR03-75	AT	66	27792003	AAGGAATGTGGATTTACACG	AGCTAGTACTCCCTCCGTTT	270
264	HvSSR03-76	AAT	60	28278804	GTATTCCAGATTCCTGCGT	GTGTTATCCTCTCTTTGCTCTC	249
265	HvSSR03-77	TA	52	28566550	TCCTGTAGTACGAATCTGGTT	TGCTAATGATGCGTAAGATG	236
266	HvSSR03-78	AT	55	28602578	ATGACATTCGGTGATTTAGG	TATTGGGCCATCTTACAAAC	333
267	HvSSR03-79	AT	69	28610858	GTTCACTGTTGTGCACTTGT	TAAGACGATCGGACTCCTAA	325
268	HvSSR03-80	CTT	62	28760860	TCTTTACTCGTTTCCTTTGTTT	TAGATGAAGTTGTAGATGCCC	195
269	HvSSR03-81	TA	70	29202897	AGGCTTGCTGCTATTGATAC	GGTATGTAATCATGGATTGTGA	292
270	HvSSR03-82	TA	55	29515075	ATAACCACTGCAGAAATTGG	GTCTTGTTTGGTTTGTTGGT	299
271	HvSSR03-83	TA	66	29922785	CAATGTTGTCACGTCAGGTA	CGTTTGGATTTACAGTGGAT	362
272	HvSSR03-84	AT	56	29960790	CCCTTAAACTGGGTATAGGC	CTGGAGTACCCTTCACACAT	151
273	HvSSR03-85	AG	66	30344308	GCAAACGACACAAGTCATTA	ATAGTGCCCTTTCTTTCACA	310
274	HvSSR03-86	TA	61	30415105	ATCCGGTTTAAGTAGCATGA	GTCACGTAAGGAAAGAGCAC	277
275	HvSSR03-87	AT	68	30594028	TCACTGATCCAGATGACAAA	CCGTTTCAGGTTATAAGACG	352
276	HvSSR03-88	CT	61	31692418	CTCCTCCTCACTCGATCATA	CCTGGACTTCACTTAGGTTG	360
277	HvSSR03-89	AT	67	31943273	GCAGTCTTCCGAATTGATAC	TCAATGATTTATTATGGCCC	372
278	HvSSR03-90	GA	60	32034304	TAAGGGTTACGACCTTTGAA	GTGCGGGAACCCTATCAC	291

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
279	HvSSR03-91	CT	68	33655350	GCTATACCCAAATGTCCAAG	GCATTACTAAGGTGCTTTCC	309
280	HvSSR03-92	AT	59	33853708	TCAGATTCCAACAACAGTGA	GGATGCCAATGGTAATTAAA	289
281	HvSSR03-93	TA	65	34558087	GATCTAGGCACAAGGCATAC	TTTCCTTCTTCTGTAAACGC	258
282	HvSSR03-94	TTC	66	34728526	GACACGATCCTTACCTTGAA	AAGAAGCAATGTCTCCTCAA	315
283	HvSSR03-95	AT	65	35985247	CCATGATAAACTAGGCCAAC	ACAATAAGAGGGTTGAGGGT	300
284	HvSSR03-96	AAG	53	36006586	GACTGACTTCGGTGTTCATT	TGCGTTCTTATATATGGGCT	232
285	HvSSR04-01	AAT	57	236216	TAAAGCGTACATCGAGGTCT	GCACATCATCATCATCACAT	345
286	HvSSR04-02*	AT	56	515091	ATGGCTACACACACTGCTC	TTTCATTGCTTACTTGACAC	325
287	HvSSR04-03	TA	54	693889	AACACCGTAAATGTTTCACC	ATTGAGATAGAGGTGGCAAA	339
288	HvSSR04-04	TA	54	941149	TAGCCAATAGACTAGGAGCG	TTGTGAACCAACCTACTCCT	307
289	HvSSR04-05	TA	69	1300032	TTGTAGAGCAGGACCTTGTT	TCATACACAGATCCACCAAA	360
290	HvSSR04-06	AT	54	1688793	CGGGTTATAACTATCGTGGA	TAATGCTTCCTGACTTGGAT	314
291	HvSSR04-07	AT	64	3102542	GTTGTTGCATGGTTGTGTAT	GGTGAATTCTACTACGACGG	351
292	HvSSR04-08	TA	64	4166765	GGCAAATAATCAGAGATTCG	AGGAATTGAGCTCTGAGAAA	339
293	HvSSR04-09	TA	57	4753069	TCATTCAATTCAAAGCTACG	TAGTTCACCATCAGTGTCCA	360
294	HvSSR04-10	AT	60	4780770	GGGTCTTTCTTTCAGTTTGA	TAGCATCTCTTTCGACCATT	302
295	HvSSR04-11	TA	69	6371773	CAAATTTAAACCATGCCATT	TGAATCTTGGAACACATCAA	253
296	HvSSR04-12	TA	57	6384741	GCAACTTGCATTCACATTTA	CTTTCATGTGCTTTCCTCTC	393
297	HvSSR04-13	AT	56	7956041	GGAACAAGATTCAGGACTCA	TTAATTAGCTTTCATTCGCC	271
298	HvSSR04-14	TA	52	9544770	ACCCACCACGAGGTATATTT	AAATGGTAAATTAAATGGGTGT	296
299	HvSSR04-15	ATC	56	11315953	TAGGTTTGGGTCCTTAAATG	GCCATTCAAGCAATTGTAA	276
300	HvSSR04-16	AT	60	11354694	TGTGGTGTTTGATGAACTTG	GAAACATTCCACAGATCGAG	303
301	HvSSR04-17	ATAG	58	11446887	TACGGTCAGGATGCTTTAAT	TGTTTCAAACCATCAACAAA	285
302	HvSSR04-18	TA	55	12023381	GCTGATCAGTGCTGTAACAA	TAGAGATCGGCCAGAGATTA	355
303	HvSSR04-19	TA	53	12416675	GCCATCCTTGACTTTACTTG	AAACAACAAAGCAAACACCT	338

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
304	HvSSR04-20	AT	64	14720513	TCGTGGAGTATCCTGTATCC	TTATAACTTGGAGCTCAGGC	265
305	HvSSR04-21*	AAT	53	15571508	GTAGGAAGAGAAGAAGGGAAG	ATGACTTATGGGATGGAATGT	293
306	HvSSR04-22	TAA	65	16410449	ATTCTCTGTTTGCCTCCATT	CCCTGTCTTGTTTAGTGTCTTT	254
307	HvSSR04-23	AT	68	19204351	ATTGCGCTTACACAGGATAC	GGATTTCACCTACTTCCTCC	273
308	HvSSR04-24	TA	68	19275448	GTACCGGGATTCAATGATAG	TTAATTTGAATCTTGGAGCC	250
309	HvSSR04-25	AT	57	19308752	AAATGTGGCACTGATTTCTC	TTTCGAATTGATCATGTGTC	317
310	HvSSR04-26	AT	66	19339244	GAGGAATTCATTCATCATGC	ATTTCGTTATTTGCATTGGT	187
311	HvSSR04-27	AT	62	19766014	ACGGCATTAATCACCAGTAT	GATATGCAATCAAAGGTGGT	224
312	HvSSR04-28	AT	58	20234496	ATGGATTTAGGCTTGTTTGA	ATACTGCGAAGGTGAAGAGA	318
313	HvSSR04-29	TA	53	21289578	TCTTACCTTCGATTAGCTGC	GGTTGAAATTGAAACGATGT	368
314	HvSSR04-30	TA	56	21528884	TCCTTACTTACCGTTCCTGA	TTGCTCTTTAAATTCGGTGT	390
315	HvSSR04-31	TA	66	21597514	GTAGGCGGAAGTAAAGTTCA	TTTGTGCATTTGAAACCAT	200
316	HvSSR04-32	TA	65	22552329	GCACCTCAGCTGGTACTAAT	ACTCCGGGTACGCCTATAA	324
317	HvSSR04-33	AG	62	22674757	CGAACACTTCAACTTTGTCA	GAGAAATCTCCCTTCCTGAG	389
318	HvSSR04-34	AT	59	22849739	ACGAACCATGATTTATTTGG	ATCTGGGTTTATAATTGGGC	323
319	HvSSR04-35	AT	54	22878450	ACCAACCTAATACCGATGTG	CGCGAGTGTTGTAACTTTAAC	323
320	HvSSR04-36*	GA	56	23139459	CGCGTCATTAACAGACATAA	ATTGATCTTCTTGCCTCAGA	360
321	HvSSR04-37	AT	56	23164498	ATTCGTCGAAATCACATCTC	TCTTTCCCTAGTAGCAGCAG	368
322	HvSSR04-38	TA	56	24056561	CCAAGCACCTCTTAACTTGA	CCGTTCTTATTAGGTTGTGG	311
323	HvSSR04-39	TC	56	26823470	CAAATAAGATCGCTGAAACC	TTCGGAGTAAATTGGACATC	268
324	HvSSR04-40	TA	70	27485057	AACTAAGGTGTTTCCCATGA	GCTAATTGCTGCTTTGTTCT	397
325	HvSSR04-41	AT	54	27828479	GGGATGACCATGTGAGTTTA	GCATTAACAGATGATGCAAA	149
326	HvSSR04-42	AT	65	28547010	GATGGTGAATCTCGGTCTAA	TGTCCCATCATCACAAACTA	305
327	HvSSR04-43	AT	62	28695031	TTTGAGACCTCCAGTACACC	ATATGCAATCCCACAAGAAC	199
328	HvSSR04-44	AT	68	29395963	CCATACCTCATACCACATGA	TATTGGGTGTCAAAGAAAGC	318
329	HvSSR04-45	AT	57	29826161	TGATTCATTCGTAGCACTTG	TTTGCGAGTTATTTATGGCT	368

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
330	HvSSR04-46	AT	61	30499205	CTAGTGGGTAGCATGAGAGG	CAAATTCAGTTCAAACTTGCT	335
331	HvSSR04-47	AT	62	30640418	GGCGCGCTTATATATGTACT	CGATTGCGTGGTGTAACTAT	179
332	HvSSR04-48	AT	63	30833000	ACATGAGCTCTGATCGACTT	TAAATCTCCCAAACCACAAC	209
333	HvSSR04-49*	AT	64	32235029	CGTTCCTCTACAAGCAAGTC	GAAGTTGAATTTGAGCTTGT	275
334	HvSSR04-50*	TC	62	33273406	ATAGCTCAAGTGCAAGAACC	AAATTTGATGTACCGGCTAA	363
335	HvSSR04-51	TA	68	33989467	TGACTGAGTTGACAGGAACA	AAATGATTGAAGGATGATGG	393
336	HvSSR04-52	TAGA	56	34284987	ACTCTAGTCCCTCCGCTCTA	ACCACGAAATATAAAGCCAA	386
337	HvSSR04-53	TA	66	34466987	TTGTCTGTGGAACTAGGGAC	GGTTTACGGACACGAATAAA	328
338	HvSSR04-54	TA	60	34518242	AGTTGGAGAGTTGTTAGGCA	GGCGATCTACTGTTCTTGTC	230
339	HvSSR04-55*	TA	61	34738683	GGGTAGACCCGTTAACTTTC	TCTTCACTTCTTCAACGGAT	387
340	HvSSR05-01	TA	61	146190	TCCGTAAGACAGACACACAA	TGCATTCGTCACTACACTTT	374
341	HvSSR05-02	ATA	52	147947	TATTGGCCATTGATTACTCC	CATCTTACAAACTGAAACGGA	190
342	HvSSR05-03	AT	57	169938	TTTCCAGAGCAAAGGTTAAG	ACCAAATACAATCGATGGAG	299
343	HvSSR05-04	ATGA	63	172217	GGCACATATATTCTCCCAAA	GCTTGATGTGGTTATCCATT	338
344	HvSSR05-05	TA	57	209890	TATACATCAGCAATGCAAGC	TAATCATTCCTTCCAAATGC	152
345	HvSSR05-06	AT	65	588482	AAGGAATATTTGCGAGTCAA	GGCGAATTTGTGTATTTCAT	304
346	HvSSR05-07	AT	65	642293	CTTGAGGATGATGTGGACTT	GAGAGAGGAGGAGGAGGTTA	367
347	HvSSR05-08	AT	70	643983	AATCTCTAGGCCTTCTCCTG	CAACCGGGACTATTGTGTAT	191
348	HvSSR05-09	TA	57	1885011	TTGAAATTGAACAAACCCTT	ATAACTTTCCCAAATGCAAA	304
349	HvSSR05-10	AT	56	2123604	CTCTCCATCTTGCAATCTTC	TGCATGACTCTATCAACCAG	335
350	HvSSR05-11	AT	66	2194662	CATGCACCATTAATACATCG	CAGGAAGTCAGGAACGATAG	185
351	HvSSR05-12	CT	52	2893991	TTAGTCATGTCCCTCCTCAG	ACAAGATTGCAAGAACCCTA	307
352	HvSSR05-13	CT	59	3106874	TCCTCTACAGTTGTCTGCCT	CATTCCTCTCCACTTTCTTG	223
353	HvSSR05-14	AT	55	3180210	TAGGATGGGTTAGAATGGTG	GTATTGAGGCATGGTTTGAT	343
354	HvSSR05-15	TA	58	4471671	ACGATTCCTGCAAGATAAGA	GCCATTTAGTACGTAGCACC	249

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
355	HvSSR05-16	AT	67	4664936	CCATGTCAAACGGTTACTTT	GGGAGAAGTGAGAAAGAGGT	275
356	HvSSR05-17	AC	67	5335574	CAAAGAAGGTGGTTCCTATG	AAGCATTTCAAGGATGATTC	328
357	HvSSR05-18	AT	61	6112261	AACCTTTGGTGCTAATTGAA	GAGACCAACTGGAGTGGATA	366
358	HvSSR05-19	AT	68	6553798	AGCTTGTACTCCAACGGTTA	AGCAAGTTTGAACTGAATTTG	340
359	HvSSR05-20	AT	52	6563520	TTGGAATCCTTGAGAAGAAA	TCAAATACGATGACGTACCA	221
360	HvSSR05-21	TA	56	6614172	ACATAGGATACCCATACCCA	AATCCGGGACTAAAGATAGC	308
361	HvSSR05-22	GA	67	6806368	GGATCACCAGAGATGAAGAA	AAACTACTCCAATCTGCCAA	333
362	HvSSR05-23	TA	56	6996548	GCAGCCATCTATCATCTAGC	CTAGCTGCACCAGTTTGATT	392
363	HvSSR05-24	TA	53	7002277	TTGTCAAGATTTGAACCACA	TTCAGTAACCATGTACGCAG	225
364	HvSSR05-25	AT	53	8283540	AGCGTGGCTACATTCACTAT	GCTTGTCTTGGAAATTGTTC	362
365	HvSSR05-26	ATA	67	9128839	ACTCTATGGGCATTCAGCTA	GTGGAGAGGAGAACAACTCA	313
366	HvSSR05-27	TA	59	9457625	AGTATTAATTGCATGCCTCG	AAATTACGGCCTCATGTTTA	282
367	HvSSR05-28	ATAC	63	9458966	TGCACACATCCATTACATTT	ATGGGCATGTTTGTAGTCTC	364
368	HvSSR05-29*	AT	66	10654167	GTTCCCGGACTTTCTTTATC	CTCCATTCTCCAAACCATTA	309
369	HvSSR05-30	TA	59	10741121	TTGCCAAACAAGGATAGAGT	AAGAACACACGAACAAGTCC	171
370	HvSSR05-31	TA	65	13459245	TGGAGCTGTGTTGTTGATTA	ATTGTGACATGCTGATGTTG	385
371	HvSSR05-32	AT	69	13463656	TACGACGGACGATTAAAGTT	GCTAACTCATTCATCTCGCT	353
372	HvSSR05-33	TA	67	14246254	TTCTTCGCTTCTAGACGTTC	CATACTCATTTGGATTGGGT	371
373	HvSSR05-34	AT	63	14374612	ACGCGCGTATACAAACTT	CCCTCTATGCCTCTAATCTCT	317
374	HvSSR05-35	AAT	63	14496338	GAAGGCTACGTTCTAGCAAA	TCGGGCATCTAGTCATAGTT	193
375	HvSSR05-36	ATA	66	14799058	ACACCCATCACAATCTAACAG	CATGGTAGTACATGAGCACG	288
376	HvSSR05-37	AT	52	15791998	GAAATTCCAGCTATCGTGAG	ATGTTGATTACGGAAACCAG	295
377	HvSSR05-38	TA	66	16116040	GATGAGAGCGCTTGTAGAAT	CAATTTCAACAATAGTGCCC	246
378	HvSSR05-39	TAA	62	16823208	TGAGAGGATACTTGGGACTG	CCAGCATGCAACTGTAACTA	260
379	HvSSR05-40	AT	55	16902205	GTGAGGGCTGATGATTAGAT	GAGATATGATTGATTGGGCA	354
380	HvSSR05-41	AT	58	16942044	TCCAGGTGATAAGAAAGTGAA	AGTGCAACTGCTATATTCCC	209

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
381	HvSSR05-42	TA	63	17826744	AAACGATGCTATCCAAGAAA	ATTGGCTGGTGTTATCAGTC	303
382	HvSSR05-43	AT	67	18997773	CCATTCCCAAATATACCAGA	ACAAGGGTTCCCGATATAAT	333
383	HvSSR05-44	AT	58	19112592	TATCCTTCATCCGTTCCTTT	CGAACATGTGAATCTAACCTT	378
384	HvSSR05-45	GA	63	19181487	GTGCATTTGCAACTTAAACA	GGGAGATCAAGAAGAGGTTT	286
385	HvSSR05-46	TTA	53	19649637	TAGCCAAACCACATCACATA	AAGTTATGTCCAAATGCACC	288
386	HvSSR05-47	ATAG	57	19770828	TAAATGTGGTGCCTTCTTTC	TGTGCAATTTGCTTCTGTAG	346
387	HvSSR05-48	TA	54	19823109	GAATTGAAGGTGGGACATAA	GAAGATGGCATGTAAACGAT	124
388	HvSSR05-49	AT	52	20621514	ACGCAAGATTTGATCTAGGA	GGTGAATAAGAGCAATACCG	374
389	HvSSR05-50	TA	69	20630858	ACTCTTCTGATACAAACGCC	TACTGCTCGAGAGAACTGTG	380
390	HvSSR05-51	ATT	61	21175117	CCATGAAATAGTTCTAGGGAA	TAATTAATGCCTTCGTGGAT	276
391	HvSSR05-52	TCT	64	22582661	GCTTAGTACTTGCGGCTAAA	CCATCTTACATGTCCTCACC	194
392	HvSSR05-53	TC	64	22879755	GAAACAGGAAGCATTTAGGA	AATGGATCCCTTTATTCACA	343
393	HvSSR05-54	AT	65	23084460	TGAGATTATTTGTTGAGGACAA	ACCCTGATGTGAGTTGGATA	353
394	HvSSR05-55	TA	63	23105114	CAACTGTACGTACTCTCGCA	TGAGATCTTGTGTGCATAGG	396
395	HvSSR05-56	TA	58	24015384	AAACTATCCGCTTGTGAAAT	CCGGTTAAGGACTCCTATCT	203
396	HvSSR05-57	AT	54	24080305	CCGAAATACCATGTTAAGGA	GTCACTCTACGAGTTCGCAT	392
397	HvSSR05-58	AT	59	24143735	TGCTAGCTGTTTCGTGAGTA	TTTGTGCATTTGAAACCAT	321
398	HvSSR05-59	AT	56	24290012	TCTGCAAGTCAGGTTACACA	CGTTGTACTAGAAGATGCCA	296
399	HvSSR05-60	CT	65	24397979	GACCTCTCCTCGCCCTAC	CAGAGAGCACTTCCTGAATC	253
400	HvSSR05-61	AT	70	25435450	GCATATCTTGACACAACCCT	CATCTTTAGCCGTCATTTGT	336
401	HvSSR05-62	AAC	65	26078800	GATGAGGAGACGATCGAAA	AGGTCCATTGGTTTATGTGA	329
402	HvSSR05-63	TA	60	26202975	ATTAGGTGCTTGCTCAACAT	TGCACTTATTTCAGGATTCA	268
403	HvSSR05-64	TA	69	26597886	ACGAACACTACCTAAGTGGG	TAATAAATTCATCTCGCGGT	361
404	HvSSR05-65	TA	66	27217557	ATTAAACGCACACTGGAAGT	AAACGGAGGGAGTAGTTAGC	400
405	HvSSR05-66	TG	61	27292741	GTTATGCGCTTCTGCTTATT	AGTTGGCTTCTGGATTACAA	215
406	HvSSR05-67	AGA	59	27774554	GAGTACCCTGTTCCACCATA	CACAAACAACAGCATTTCAC	241

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
407	HvSSR05-68	TA	67	28456305	GTGCACAGGGTATGTTACCT	ATACCTCCTCTGGAATGGAT	308
408	HvSSR05-69	AT	52	28982918	GGATCGGTCATCTGAATCTA	TGTTCTGTTCATGTTGCAGT	257
409	HvSSR05-70*	TA	67	29667338	GTCCTCTACTCCTTCGTCCT	AGTAGACGGAGCACTAATCG	343
410	HvSSR06-01	TA	60	222545	CTTATTCCCAACCTCTCTCC	AACGGAACAGCTTATTATCG	219
411	HvSSR06-02	AT	58	483120	TTACACGCTAGCTGCTATGA	TAGCTAGCCACCTACACCAT	246
412	HvSSR06-03	GAA	67	534728	CTAGGGAATCAGCGGTTAG	GCTCTCTTGTCCTTCTTCTTC	212
413	HvSSR06-04	AT	66	718054	TATTGTTGATGTGGTCGCTA	TTGTTCTCTGTCGAAGTCCT	345
414	HvSSR06-05	AT	55	720892	TCTTCTTGTCGGATTAGCAT	TAGTTTCATGGAAAGGCTGT	310
415	HvSSR06-06	AT	58	1064440	CGTATGGAGATCCAAGTGAT	GTTTGATACGTACGCCAGAT	276
416	HvSSR06-07	TATG	63	1086331	CTGCAGAATGCATGAATAAA	AATATGAAACGGATGGAGTA	218
417	HvSSR06-08	AT	56	1550302	TGGAAAGTGACTGGTAGGTT	TTAGTACTCCCTCCGTTTCA	225
418	HvSSR06-09	AT	56	1551156	TGCTATCACATGCAGAAGAC	GCCTGTTTGGAAGAGTGTAG	159
419	HvSSR06-10	CT	54	1562658	CTTACATGGCAACGATCATA	CTAGACAACCTGAGAGTGCC	215
420	HvSSR06-11	TAT	69	1894192	ACTGGCACAGCCATACTACT	AAGAAGATGATGACGACGAC	328
421	HvSSR06-12	TA	63	1964028	TCCCTTTGACTAAAGCATGT	ATCCACGTACAAGACTGACC	353
422	HvSSR06-13	TA	60	2174519	CTTCAGTTTGAGACTTTGGG	TCAAATTACTGCAATGGACA	374
423	HvSSR06-14	CTT	54	2291887	CCCATCTGCACTACCATAAT	AGATGTGCTTTGCTACCAGT	265
424	HvSSR06-15	AT	64	3184600	GTTATGCTGACAGTGACACG	CATGCCAGTATAGACCCAAA	398
425	HvSSR06-16	AT	60	3723501	CAGTACAACACATGCTTTAATTT	GCTTTGGACTGATTATGAGC	290
426	HvSSR06-17	AGA	55	3945764	TCTGAAATGCTGTCATCAAG	GAGCAGAGTAGGACATGAGC	368
427	HvSSR06-18	AAT	61	3959976	TTGGAGTCTTGGATAAGTGG	GTTGAGTGGATGACCAGACT	280
428	HvSSR06-19	TA	66	3970474	GGCACGATACATAAACAACA	CTCACAGGGACAATTCAAAC	397
429	HvSSR06-20	TTA	56	5212414	CACTGACAAAGCTTCCGTAT	ATAGTTGCGGAGTGGATAGA	302
430	HvSSR06-21	AT	63	5233668	AGTGGGCGAGTATAAATCAA	AGGTTAACGCAGAGGATGTA	221
431	HvSSR06-22	TGT	51	5237589	TTCATTTGGTGTCAGAATCA	AATTAATCACATCTGCCCAC	334

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
432	HvSSR06-23	AT	62	5959083	CACCTTCAAGTACAGCCTCT	AGTGGTGTCTGACTGGTTTC	335
433	HvSSR06-24*	CT	57	6152229	CCCTCGGTGTTGTACAGG	ACACCTCCACTCCTCCAT	334
434	HvSSR06-25	AT	53	7278044	CTCTTTCCTTCCTCTTAGCC	ATTCCTGCATAGCATCATCT	360
435	HvSSR06-26	TAT	57	7812944	TCTAGCAAATTCCCATTGTT	ATATGGGCTCAGGGATTATT	336
436	HvSSR06-27	AAT	63	7857135	TGCAACCTTGATTTCTTCTT	ATTACACTTGCAGCACATGA	375
437	HvSSR06-28	AT	58	8166999	AGAAGTCATATGCAGAGCGT	CTGTTAAGCAAGCTCCAAAT	353
438	HvSSR06-29	TAA	65	8169444	GTCATCATCCACCCATAGTC	CCGTATCCATTTATTGTCGT	285
439	HvSSR06-30	TTA	52	8189755	GCGATTCAAAGTGACTGATT	CGCATGTGCAATATGCTA	349
440	HvSSR06-31	TA	61	8429828	GACCTAGACCAGATTGTGGA	ATCCGCTAGGAGAGGAAATA	183
441	HvSSR06-32	TA	57	8739617	GTTCTGATCGAGTGTGGTTT	ATGTGAGTCTTTGGATGCTT	385
442	HvSSR06-33	AT	56	8919015	CACGTTAACCGTAGGTCTTC	ATCCTATGACTGATCCAACG	223
443	HvSSR06-34	TA	59	9559659	CTAAGCTCCTGGTGGTAGTG	TTTGAACCAAGGTAGCAAAT	301
444	HvSSR06-35	TA	61	10822904	GAAAGGAAATCAGGTTGTGA	CCCATTAGACATTTCGGATA	279
445	HvSSR06-36	TA	63	10998822	GGATACCTGTGACGGATAAA	AGAGGAGAAGTGGAGGAGAG	285
446	HvSSR06-37	AT	68	11181731	CCACTACTATCGAACACAACC	ATTGATCCATTGTTGTTGGT	309
447	HvSSR06-38	TA	57	13423787	AAATTTGTCGATGAGCTGAA	ATCATTTGTTCAGATGGCTC	273
448	HvSSR06-39	AT	70	14543378	CTTGTGGTCAAGGTCTGATT	ATGAAATAATCAAAGCCAGC	315
449	HvSSR06-40	AT	52	16035015	TGATTGAAATAGCTGGGACT	AAAGGAGAGAGAGATCGGG	368
450	HvSSR06-41	ATA	65	16475032	CTCTTCCGTGGTTAAAGAAA	CACTGGTATGATCTCCGACT	385
451	HvSSR06-42	TA	70	16676847	TTTCCATAACGGATGAGATG	TAGTGGATAATGATGTGGCA	244
452	HvSSR06-43	AT	58	17067267	CTTAGCCTTGCCAATAGAGA	TCCACCGTTCTATACTCACC	385
453	HvSSR06-44	AT	57	17434487	GGAGCATCCATCACAATATC	GTAATTTCAGTCAGCCAAGC	301
454	HvSSR06-45	AT	64	19147086	TGACACTGTACCGTGCTAAC	CATAGCCTACTTCAAATGCC	377
455	HvSSR06-46	AT	63	20232583	AATCCCAATCAACTTTAGCA	CTAGGCTGTCCAGCTTGTAT	284
456	HvSSR06-47	AT	52	20970841	CATGTACACCAACTGACCAA	TCGATCTTCGATCTTCTAGC	261
457	HvSSR06-48	TA	70	21021132	AAACCGACTCATACACGATT	CTAACTTGAGACGATCGGAC	283

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
458	HvSSR06-49	TA	56	21588451	GGAAAGTCTTAGGTGGGTTT	TGTATTGAAACAAGCATTCG	372
459	HvSSR06-50	TA	65	22379464	TGCTATCAATTTCGTTTCAA	ACTAGTTATGGTCCGTTTGC	270
460	HvSSR06-51	ATCT	58	22682131	TATTCGCATGGGCTATTATT	GAGTTCACCGGATTAAATGA	322
461	HvSSR06-52	AT	52	23162026	AATCAAGCTTAAGTCGTTCG	CTACCGTAGAAGAGAAGCGA	216
462	HvSSR06-53	AT	70	23335176	TTCCATAGTTGGAAAGGAAA	GTGACATACTCCATCCGTTT	378
463	HvSSR06-54	AT	62	24079271	AAATTACTCCCTTCGGTTTC	ATCAAGCTGATCGAGAGATG	382
464	HvSSR06-55	AT	55	24307806	ATGTGCTGGATTGTTAGGAG	TACTTTATATGCATTGGCCG	314
465	HvSSR06-56	CT	66	24321014	AGCATTTGTGTGTGCAATAG	ATGCTTGCCTCATCAGTAGT	351
466	HvSSR06-57	AT	62	24666218	CAGCCATTTAAACTCAGGAC	TAATTATCACGCACCATGAA	368
467	HvSSR06-58	TC	56	25670749	GAACTAATCTGCTGACCTGG	CCTATACTGGTAATGGCAGC	334
468	HvSSR06-59	AG	54	25929671	GTTTAGGCGGTCTAGAGGTT	TAAATAAAGCAGTCGCCAGT	256
469	HvSSR06-60	TA	60	26379779	GGTTGGGAATACCCTTAATC	GGTATGTAATCATGGATTGTGA	374
470	HvSSR06-61	TA	61	26720133	TTAATGGAGTAATGGAACCG	AACTCACCGTGACATCTAGG	309
471	HvSSR06-62	AT	57	27062876	AGTTGCAGCCTTAAGCTATG	TTTGCCCTGTTCAATTTACT	209
472	HvSSR06-63	AT	68	27111555	TATGGAGTGTTGTGCATGTT	GAGGATGTCTTGAGCTTGTC	318
473	HvSSR06-64	AT	70	27121561	TTGTTCCATAGGATAGGCTG	AGGTTGGCGTAATCTCTACA	307
474	HvSSR06-65	TC	63	27828305	GTGTGGCAATTTAACATCCT	TTGTTGCTTGTTCTTCACTG	302
475	HvSSR06-66	TA	70	27889919	GCATGATTTCTTTGATAGGG	TGTCACATTGGATATACGGA	398
476	HvSSR06-67	TA	55	27920830	GTAATAACCTGCACCTGCTC	TCGGCACTAGTTACCCTAGA	388
477	HvSSR06-68	AT	57	28015724	TTTCCAATGCCCAATTATAC	CAGTACAGTGCTGCAGAGAA	378
478	HvSSR06-69	TA	69	28114473	ATCAGCTGCACTTTCAATTT	GGTATGTAATCATGGATTGTGA	323
479	HvSSR06-70	ATA	52	28166203	CATATCCAACTTTAGCCGTC	TCTCGATCCGTATGGTATTT	376
480	HvSSR06-71*	TC	57	28532534	ATCCAGATGGAGATGGTACA	TTGAGAGTGAAACGAGAACC	341
481	HvSSR06-72	AT	54	28779105	TTACTACTGCCTTCGTTTCAG	AGGTTCAGTAGTGTCGTGCT	250
482	HvSSR06-73	TAA	55	28999925	TTTCTCCCTGTTACTTCTCCT	TCGCGTGATTTAGATGCTAT	369
483	HvSSR06-74	AT	58	30156065	CCATGAATGACAAGGAAGAT	CTAGAACCTGTCAGTGCCTC	399

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
484	HvSSR06-75	TA	56	30175298	CCAACAGTTTCAAGAGGAAG	CTGCTTCATGTATCCTAGCC	254
485	HvSSR06-76	AT	53	30786736	CAAAGGGCAACAAAGATAAC	CTACCCACATGACTGGATTT	398
486	HvSSR06-77	AAT	56	31006153	TTTAGAGATCATGCCGAGTT	GACGGAGGGAATAACACTTT	170
487	HvSSR07-01*	TA	62	156287	TGCTCCAAGAGACTTGATCT	GAGTACCAATTCAGCGACAT	275
488	HvSSR07-02	TAA	63	257451	GCTGATAACTAGCGAAGGAA	GACAGTGAGGTGAATCAGGT	388
489	HvSSR07-03	AT	68	304707	CCAAAGATTGTCAACAGTGA	GAGATGAGCTTGTCGAAGTC	384
490	HvSSR07-04	AAT	68	805693	ATCCTCCCTCGACTCTACTC	AATTTGTTCCTTCCATGTTG	228
491	HvSSR07-05	AT	63	1091818	ATAGCTCAACATGAACACCC	CAGCTTGCCTTAGTTGAAAT	365
492	HvSSR07-06	AT	52	1291302	GATGGAATGTAGATGTGCCT	AATAGGTCCACATGGACAAC	393
493	HvSSR07-07	AT	64	1559317	GAACCGTAGGTTACAGTGGA	AGAGCTTGATTCAATCGGTA	365
494	HvSSR07-08	AT	65	1705461	AATATTTAATGTTTGACGGTTT	GTGGAAGCACTTTGACTAGC	371
495	HvSSR07-09	AT	55	1871836	AACGGTTAAGATGGTGTACG	ACTAGGAACAAAGGAGGAGG	334
496	HvSSR07-10	TA	70	2215932	TGAGCTTTCAAGTTTCACCT	ATGTCAATTTGATGGGATGT	361
497	HvSSR07-11	TA	68	2240984	CACAAGGAGAGATGTGGAAT	TTGAATCACCACAATTTAGC	355
498	HvSSR07-12	AT	63	2418037	GCCTCTAGCTAAACCTCAAA	ACAGTAGCAGTGAAGGTGCT	170
499	HvSSR07-13*	AT	54	2516167	CCACAGGGAAACTATCATGT	AACAACCACCACCACTACTC	337
500	HvSSR07-14	AGAA	53	2524217	ACCATCTTGGATCACACATT	GTTCGTACGGATTAAAGTGG	349
501	HvSSR07-15	TA	64	2622240	TGTCGATGATCAGGTGTAGA	AGGCATGCGAATACACTACT	395
502	HvSSR07-16	TA	58	2791806	CAATGTACCATTTCAACACG	AAGTACTTCCTCCGTCACAA	384
503	HvSSR07-17	AT	52	2877874	ATATTATGGGACGAGGGATT	GTGTTGCACACCTATACACG	172
504	HvSSR07-18	AT	67	4872379	TATTACCTTGAAGCCAAACC	GTGTGGAAACCCAGTAGGTA	395
505	HvSSR07-19	AT	63	5768137	GGTGTGTTGTCGAATCTCTC	ATGCCATTGTCCTTACATTC	343
506	HvSSR07-20	AT	56	6114254	TATGGTTAGGTGGCAGTCTC	TGATGACATGGATGAATCAC	190
507	HvSSR07-21	TA	57	6596980	ATGGTCCAAGAACATGCTAC	TCTGACATGAACAGGAGACA	398
508	HvSSR07-22	AT	66	6894492	GGCAGATAAACCAGAAACAG	ATCAACGACCAGTTCAATTC	251

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
509	HvSSR07-23	AT	57	7118703	TAGCTACCCTCAACAAGAGC	CCCTTACCTCACATCACCTA	217
510	HvSSR07-24	TA	58	7541676	AGTGCCGTTTAATACGTAGC	AAATAAGACGGACGATCAAA	183
511	HvSSR07-25	AT	62	8256609	CACTCCAGCCAACTAAGAAC	TTATATTTCCTTCGGTCCAA	358
512	HvSSR07-26	AT	53	8660711	CACGTGTACGTGTCTCTGAA	GTCCACAACGATTTCCTTTA	259
513	HvSSR07-27	AT	62	8775834	TAAGATTCGTGCAACCCTAT	AGTCATGAATTGTCATCGGT	231
514	HvSSR07-28	TA	69	8838918	CAAACCAGGTGAGAGAGAAG	ACTCAGGTCATCCATCAGTC	295
515	HvSSR07-29	GA	56	8876596	AATTTGATGGAGCTAGCAGA	GATTCATGGTGTCTTTAGCC	352
516	HvSSR07-30	TA	53	12998594	AAAGCGGTTATTCTTGTCAG	TGTGAATCAGATTTGTGAGC	334
517	HvSSR07-31	AT	52	13205012	TTCAAATCAACTTACCTCCC	TTTAAATAGAGGCTTGCACC	331
518	HvSSR07-32	AT	70	14375860	CCGGGTGTAGAATTGTCTTA	CGTCGAAGTCGTAGCTAAAT	310
519	HvSSR07-33	TCA	60	17469507	AGAGGAGAGAGGGAGAGTTG	GCTCAGTTATGGTTGAGGAG	336
520	HvSSR07-34	TA	62	17795078	ATCCGCTCCAAATATAAGGT	ATTGGGTGAACATCATCTCT	315
521	HvSSR07-35	AT	56	17879176	ATGCGATTATATCAAATGGG	TACTTCCTCCGTTTCACAAT	338
522	HvSSR07-36	AT	70	19181545	TTGTCTCTCACCAATCAACA	ACTTTCTAGCATTGCCCATA	340
523	HvSSR07-37	ATCT	66	19200554	GGCAGTAATTCCCAAGAGA	GGTCCACCAATTTAAATGAA	361
524	HvSSR07-38	GT	52	19359848	AGGATTTGAAGGTTGGGTAT	GGTATGTAATCATGGATTGTGA	353
525	HvSSR07-39	AG	62	21764293	TAGCCCTGGATTATTGTTCA	TTCAGCTTCTAAACCACTCC	396
526	HvSSR07-40	GAA	54	21870543	GATTTACTCGCAAGTTACCG	TGTTTCAGGTTCGTCTATCC	348
527	HvSSR07-41	AT	66	21932744	GGATCGATAGGTTTAGGGTT	GTGAAGTCGCGAGAGAATAC	387
528	HvSSR07-42	AT	52	22396384	AACGGAGGGAGTAATTCAGT	TTCCTGGTATTCTACGCACT	321
529	HvSSR07-43	TC	53	23789010	CAACTCAGTTCCAATCCCTA	TTGTGTGTTTCATATACGGC	163
530	HvSSR07-44	TA	62	23826962	AGAACAACAGCGGAAGATAA	TAACCATAGCCACTCCTTGT	296
531	HvSSR07-45	AT	61	24060449	GAACGACAGAGGTGGTGA	TCATCCCGATCGATTCTA	380
532	HvSSR07-46	CTT	55	24194491	ACAGCTGTAGAGGATGAGGA	TCCCTAATTCGAATCACAAC	281
533	HvSSR07-47	AT	70	24705886	TGACAGTATTTGACATCGGA	CGATTCATTTAGCATAGGAGA	333
534	HvSSR07-48	TA	67	24775934	CATTCTCATTTGAATTGGGT	TTGGCATAGAGGGAATAGAA	180

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
535	HvSSR07-49	TA	65	25400352	TAGTTTGGACTTTGGACACC	CTCCATGGTTCAAATTGAGT	367
536	HvSSR07-50	AG	53	26773356	TTAAGGCCTCGTTTGAAATA	TTTCCCTATGAGCTTGTTTG	326
537	HvSSR07-51	AT	62	26883012	GAATGATGCATAAATAACGGT	ATAGCTCGTAGCCATAACCA	347
538	HvSSR07-52	AT	52	27491979	TCATAAACATCACGGGTGTA	ATTGTGCTTTGTTATTGGCT	388
539	HvSSR07-53	TA	57	29406921	CGAGCATGTCTGTCAAGTAA	GTTCGAATGTAATGTTGGCT	341
540	HvSSR07-54	TAA	56	29482695	GAGCCTGTGATTCCATACAT	CCGCACAGCTTAGTAGTACC	179
541	HvSSR08-01	CATA	52	89499	TACGTACACAGATGAGACGC	TTATTCCTCAATTCCATCGT	185
542	HvSSR08-02	AT	65	422028	ATGCCATGCATCTTTAATTC	CAATTCATCGCTGCTACATA	278
543	HvSSR08-03	AT	56	460810	AGCAAGAAGCATCATGAAAT	TGTAACCACCAGATGCAATA	169
544	HvSSR08-04	TA	58	769798	AATAGGGACATCATCCATCA	CACGGAAACATCTTCAATTT	249
545	HvSSR08-05	AT	60	1854756	CCGATCCACTTTGAAATCTA	TAGTAGGTCATCCAACCACC	309
546	HvSSR08-06	TA	60	2953486	CTTCACATCACTCACACTGC	ATATTGTCGCTAGCTTGACC	228
547	HvSSR08-07	TA	55	3085075	CATGAATTGATGAGCTTAAA	TACTCCCTCCGTACTCGTAA	229
548	HvSSR08-08	TA	60	3151554	AACAGTATCCATGTCCTTGC	TGTTTATATGAGCGGTTGTG	200
549	HvSSR08-09	AT	56	3567557	TGGATGACTGATTGACAAGA	TGGGTTTGCTAGCTAGTGTT	342
550	HvSSR08-10	TA	66	4638364	TGAACCTCAAGAATGAATCC	AACAGCTCCTACTGTTGCAT	319
551	HvSSR08-11	AT	54	5183345	ATATCCAACAATATGTCCCG	GCATTAAGTGAAACGGTAGG	268
552	HvSSR08-12	TA	52	5363030	CCATCACTTTGGATGTTTCT	GATCTCACTGTGTGCATGTC	385
553	HvSSR08-13	AAT	60	5586081	TGAAGGTTCTTCATTTCAGG	AGTACGTCATTCTGCAACCT	352
554	HvSSR08-14	TA	52	5795410	ACATATCTGCATCCCTGAAC	TCAGTTCAAACTTGCTTTCA	234
555	HvSSR08-15	ATA	66	5967354	TCCACTTTACATCGTCACAA	CTACCTCTTAACCGCACATT	295
556	HvSSR08-16	AT	53	6317015	ATATGGAAGAACCTGTGCAT	AAGAAGTGGTAGACTTGCCA	243
557	HvSSR08-17	AT	56	6638968	GTAGCTCTTCCACTCAGGC	TTAACTATTTACGCTTCCGC	274
558	HvSSR08-18	AT	60	6715556	GCCTACGACAATTTAAGGGT	CAAATTCAGTTCAAACTTGCT	257
559	HvSSR08-19	AT	67	7322476	ATCTTAGAGCGGAATTGACA	TGTAATCATGGATTGTGAAA	400

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
560	HvSSR08-20*	AG	64	7611709	CATCTCTTGAGAAATCTGCC	TGTGCATTTCGTCTTTCATA	221
561	HvSSR08-21	AT	63	7631024	ATTACCAGGCAAGTGAAAGA	CAAATTCAGTTCAAACTTGCT	353
562	HvSSR08-22	AT	64	7896677	AGACGATCGGACTCCTAAGT	GTGAGACCCATTTGTCATTC	297
563	HvSSR08-23*	AT	64	7896677	GAGAACAATCGAACGAAGAC	TATTACGTCACTTGCCTCCT	288
564	HvSSR08-24	AT	69	8845455	GGCCAACATCTATCGTGTAT	CCCTGTCTGCTCATTTATGT	389
565	HvSSR08-25	AAT	52	9278657	CGACTACGTACTTCCTTCCA	ATGAACGGAGGTGTACTTGT	179
566	HvSSR08-26	AT	54	11404849	CATATACTCAATGGCGTTCC	CAAATTCAGTTCAAACTTGCT	359
567	HvSSR08-27	AT	54	13257863	TAGTGGTTAGCAGTAGCACG	TAGCAGAAGAATCCCACACT	299
568	HvSSR08-28	TA	65	13616298	ACACCAGTTCGATTCACTTC	TGAAACGACCAATCCTTAGT	305
569	HvSSR08-29	GA	66	15308050	AACTGAGAGGCTGCTTGTAT	TAAAGGGTTCACTCATGGAC	325
570	HvSSR08-30	TA	61	15832481	CTGTTTAACACGTAGCACCA	TTGTGCAGCACTATTGAGTC	315
571	HvSSR08-31	AT	68	17451680	CTTACCGCACAAACACTCTT	TTTGTGTATTTCCTCCGTTT	389
572	HvSSR08-32	AT	63	17501335	TTACCCAAATCAAATCCATC	TTCATATCTTGTTCGGTCCT	378
573	HvSSR08-33	ATT	66	17777808	CTACCATATGCTTTGCTTCC	CAGTGCAGCAGGTACAGTTA	346
574	HvSSR08-34	TAA	54	18130796	CCGTTGTCTGATCTATGGTT	AGTAACGGAAATGGAAGGAG	260
575	HvSSR08-35	TA	59	18233120	ACGGGACCAAACCTTATAGT	TTCGGACTTCCTCTGAGTTA	377
576	HvSSR08-36	AGAT	60	18630783	GAAGCGTTTATATGAGTGCC	GTCATTAAAGATGCGAAAGG	297
577	HvSSR08-37	AT	57	18876445	CTGCAACGTTTATAGCTCAA	TGTTTGTGAACAGATCGTGA	357
578	HvSSR08-38	TA	53	20243524	AATCACCTAAGAGCGAATGT	CCGATCTACGATCTGATTACA	358
579	HvSSR08-39	TA	61	20621158	TCGGTAATTCGTTATCCAGT	TGGTATCAAAGTGAACAGTACC	208
580	HvSSR08-40	AT	59	21097750	ACTCTCTCCGTTTGACAATG	AAATTAGTGGGAGTGTGGTG	347
581	HvSSR08-41	TA	54	21768826	ATTAGCTTAAACCACGCAAT	GATTGGTTTATTCCACGGTA	333
582	HvSSR08-42	TA	58	22816195	ATCATTCTTGTCGGTCTCAC	AGTGCTACATGGAGTATGGG	362
583	HvSSR08-43	AT	59	22823529	AGGACCATACTTATCCCTCTG	GATGATCGGACTCCTAACTG	303
584	HvSSR08-44	AT	58	23597791	AGCAGGTGCAATTTACTCAT	TAGAAATTGTTGCGTACGTG	156
585	HvSSR08-45	TA	55	23950220	TGTGAAATTCATGTCTGCAT	ATCAAACAAGGTTCTGGATG	339

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
586	HvSSR08-46	AT	66	24009498	TTTGAGGTTATTCCCAAATG	TTGTAGCGGTTGGTATCTTT	392
587	HvSSR08-47	CTT	57	24348607	TAGTTGAGCCCATTTACAGC	CTTCCCAAGTTAGACAAGGA	349
588	HvSSR08-48	TA	67	24703055	GCTTTGTGAAAGTTTGGAAC	ACCACCTTTGATTCCCTATT	348
589	HvSSR08-49	AT	54	26267270	GTTCGGTCGTAAACACATTT	TTCTAGGATTGCTCAAATTCA	369
590	HvSSR08-50	GA	53	26403473	ACAAGCAATATCCACATCCT	TCCATTGCTCTCGTTAATCT	324
591	HvSSR08-51	AT	58	27316297	TGGGTCAGTCAGGAAATTAG	CGGGCTAGTATAGTGCAATC	365
592	HvSSR08-52	GT	63	27772501	GTAACTGTTTCGCTTGCTTT	ATCTCATACCCAATTTCGTG	321
593	HvSSR08-53	TA	55	27935731	GAGTACGATCGACCACAAGT	CCCATGCATATAGATTGGTC	335
594	HvSSR09-01	AT	62	693008	AGGTAGGCATCTTTAGACCC	AGAACTCGTAAGTGCAGCTC	320
595	HvSSR09-02	TA	55	780072	CATCCAGTAACAGGGTTGAT	TTGTTGGATTTGTGATTCAG	374
596	HvSSR09-03	TTA	53	1715836	GCCACCTAAGCTCTTTGTTA	CATATGGAGGGTTGGATAAA	305
597	HvSSR09-04*	TA	55	2083370	GTTCACGTACGTCTTGGC	TGCTCTAGCTAGCTTCGTCT	339
598	HvSSR09-05	TA	70	2290648	GAACGGAGGGAGGTTGTT	AAAGTGTCCTAAAGCCAAGTC	258
599	HvSSR09-06	AT	64	2369521	ATCACGTGCATGTTCTATCA	AAATACCCAAATATTCCAAA	255
600	HvSSR09-07	ATT	57	4353753	CATCTCAGCAAACAAGAACA	GTAAAGACTCCAGCTTTCTCC	321
601	HvSSR09-08*	TA	61	4593433	TCCGAGCTTATTCAGATGAT	CTTTGGTGATGAAATAAGCC	236
602	HvSSR09-09	TA	55	5126913	ACCAAGAGAAACTATGAACGG	AACGGAGACCTAACCATCTAA	356
603	HvSSR09-10	AT	54	5816502	CGTATTGGTCTCCACTAAGC	TGGAACCGATAAAGTACACA	204
604	HvSSR09-11	AT	65	6938056	TGCAGAATTTCTTCCTTCAT	ACCAGAATCTCCCAAATGTA	366
605	HvSSR09-12	AT	67	7087191	GTGGAACATTGGGCCTTT	ATTATATGCCATGGACGAAA	162
606	HvSSR09-13	TA	57	7720094	GATCGATGTGAGTGAGATCC	TTCCTTTAATTTGCTGCTTC	343
607	HvSSR09-14	TA	57	9529538	CTGAGCCATCTTACTAACCG	ATTCATCAACGAACAAATCC	343
608	HvSSR09-15	AT	63	9734385	CTCTCAACCATGTACGGAAT	CACGTCGAACGTATGAGTAG	320
609	HvSSR09-16	TA	65	9905138	TAAGGTGGTAGCTGCTTCAT	GGACAATTCGAATAGCAGAA	294
610	HvSSR09-17*	AT	54	10066628	GATAGCTGGCTGAACAGAAC	TGTAGTTGCACTCTTGTTCG	283

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
611	HvSSR09-18	TA	67	10396429	TTTAACACCGTTACCAGACC	TCTATTCCTTCATGAGGTGG	359
612	HvSSR09-19	AT	64	10514116	TCGAATTTAGTCCAGGGTAA	GGTGAGAGATCTTGAGTTCG	296
613	HvSSR09-20	TA	57	10956729	TATCATACCACTTTGTCGCA	TTGATTTGAAGATCATTGGT	354
614	HvSSR09-21	TA	58	12072076	TACCAATCCGGAAATAAAGA	ACATATGCATTCGCATAACA	296
615	HvSSR09-22	TA	58	12087982	TCCGTGCTAATTTGAGATTT	TAGAGAGAGAGAAGGGAGGG	284
616	HvSSR09-23	ATAG	66	12351924	GATTTACGTACCTGACGGAG	ACTCCTCTCCAACAAATGAA	374
617	HvSSR09-24	AT	60	14110765	TTCAGTTAGATTCCATTGGG	TAGCATTGTCCACGTTCATA	288
618	HvSSR09-25	GAA	53	14566048	GATCGATCTCATCATCACCT	TAGCTTCCTACTGGGAGTGA	215
619	HvSSR09-26	TTC	68	14784930	GTCGATCGAGGAGTAAACTG	GGTGTCACTGGTTTGTTTCT	400
620	HvSSR09-27	TC	62	15505057	TGGGCATCTGGTACTATCTT	AGCTCATTCCACAGGTTAGA	331
621	HvSSR09-28	AT	67	15574920	TATTCCCAAGATTGTCCATT	TTCTTGGTAACAACAAGGCT	355
622	HvSSR09-29	AT	64	15577568	CACTGAGGAAGATTGGACAT	ACAAAGCTGGGACTAATGAA	264
623	HvSSR09-30	AT	60	15663120	TAATAAATGTTTGTGCCGTG	AATCCAACCTTGACTTTGTG	329
624	HvSSR09-31	TA	54	15812256	ACTGTAAGCATCCAATCACC	AACAACAGAGGCTGGAAATA	379
625	HvSSR09-32	TA	70	16116649	TCTGTGTTTCAGTGTTTGCT	TCTCCTGCTAGAGAGGATCA	290
626	HvSSR09-33	AT	60	16133935	CATTCCCTGTTCAGTTTCAT	GTTTGTCATTGGTTCTTGCT	394
627	HvSSR09-34	TA	63	16311025	ATGGTACTAGACCGTGGATG	TTTCGTTAGCGCTTCTAATC	199
628	HvSSR09-35	TA	59	16447183	TACACGAAGTGATGCTCTTG	TGTGACCAGGAACAAATACA	395
629	HvSSR09-36	AT	64	16809129	AAGCTAACTAGGATGGTGGC	TTGGAAGAAGAGCATAGAGTC	391
630	HvSSR09-37	CT	59	16815917	AATCTCAACTGCTCGGATTA	TTGATTGATTGATTGAACGA	373
631	HvSSR09-38	AT	54	17075034	ATTGATTTGATTGCAGGAAC	CTCTAGTCACGATACCCTCG	324
632	HvSSR09-39	TA	55	17356839	GCAAAGCAATCTCCTACAAG	TGTTGTGTAATGAGGCCATA	255
633	HvSSR09-40	GT	58	17405160	CGTTCTAGTATTCGCAGGTT	TCAGAACCTAACTATTCGCC	347
634	HvSSR09-41	CT	60	17719738	AACTTAAATCCAAACAGGCA	GATCTTTAGTCCCGGATTCT	275
635	HvSSR09-42	TA	66	18166174	TCCTAAACACAATCCAATCC	CTCAATTTGAATCTTGGAGC	325
636	HvSSR09-43	ATA	57	18312692	TTGAAGGTGACTGATGTGAT	GGATACTGCTGGTGCTCTAC	398

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
637	HvSSR09-44	AT	53	19037129	TCCACACAGTAGTCCACGTA	GCCCATATTAATACACGCAT	294
638	HvSSR09-45	TTA	56	19709432	TGGTCATTGTTGGTTGTTTA	GGTGCTGATGACTTCTGAAT	309
639	HvSSR09-46	AGA	69	19788287	ATCGCTTTCAGTGTCAACTT	TTAAGAAGAGATGAGCCAGG	378
640	HvSSR09-47	TA	58	20197523	CGAAATAAGTGGAACGACAT	GCAAAGTTGAAACCAAGAAA	151
641	HvSSR09-48	AT	67	20467144	TCCTGCGTGTAGTCAGATATT	CCATCCTAGTACAGCTCTCG	356
642	HvSSR09-49	TA	61	21122930	GAGAGATGAGGCATGACACT	ACATCGAATACGTCCTTGTC	274
643	HvSSR09-50	TA	56	21347711	ATTCTCTGCACCCAAAGTAA	CTCGCTAGCCTATTGTTGAT	382
644	HvSSR09-51	TA	57	21757561	AGTACTGCTGTCCTCTGTGC	TATTGGAATCAGTCGCTACC	379
645	HvSSR09-52	AT	66	21982399	GTTATCCCGGTGAACAACTA	TTGGTACAAAGAATTGGAGC	361
646	HvSSR09-53	TA	58	22343879	TTGATGACGTTACTGCTTTG	TTTATACCTTGGAGCCCG	299
647	HvSSR09-54	TCT	69	22356519	CAGCAAGAACCATTGGAG	TGAACGTATCACCGAACATA	388
648	HvSSR09-55*	AT	62	22392301	GAGAAGGCTGAAGGTGAAG	AAATGCATACCAACGTTCTC	351
649	HvSSR09-56	AT	66	22482806	TTACTCCGCATATATCCATGT	ATTTGACACCAAGTTGATCC	382
650	HvSSR09-57	AT	57	22819260	GGAGGTTGTTGTTACGTTGT	GGGAGGGTAATTCAGGTAAG	369
651	HvSSR10-01	TA	66	37013	ATGTATCGCTCGACAGATTT	CCGATTCATTGATGATTTCT	380
652	HvSSR10-02	AT	62	170341	TTAATTCCACTTCCCTTTCA	GATTGCTCCCTTTCCTAGAT	338
653	HvSSR10-03	TA	55	272650	TCTTTCCCAAATTCCAGATA	CATTAGTTGTTTGTGGCAGA	289
654	HvSSR10-04	AT	63	297382	GAATCAAATATATCCCGTGG	AACGACTTACAATCACACCC	365
655	HvSSR10-05	TAT	57	299671	TCTCGCTCACTACCAGACTT	AATTTCGCTTCACATCACTT	285
656	HvSSR10-06	AT	57	866084	TAACAGTGTGTGTGGATGCT	GGGAGATACAGGATTGATGA	283
657	HvSSR10-07	AT	69	1405578	CACATACGTTATGCAGGCTA	TTTGGTATGTCATCCGTGTA	394
658	HvSSR10-08*	ATG	51	2035022	CAAGAAAGCCGAGTTAAAGA	TCCTCCAAAGATGGTATGAC	350
659	HvSSR10-09	AT	52	2992239	TTGGCCTGGGTAGTTAATAG	TTTCACAGTCCATGATTACA	328
660	HvSSR10-10	AT	52	3992179	ATGCACACGTACCAAGTACA	TAAACTTATTCTCCTGCCCA	340
661	HvSSR10-11	AT	64	4637758	ATCCAGTTCAACAGGTCCTA	TAAATTTCAATCATGGTGCC	359
662	HvSSR10-12	AT	70	6248420	AAGGTGGTTCTGTGTAGTGG	TTTGCATATCATCAGTGGTAA	260

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
663	HvSSR10-13	TA	58	6424330	CAGGGAATCAACATCAAAGT	AGCAAGGCAAGTCATCTCTA	169
664	HvSSR10-14	TA	61	7017747	CTAGAGTTTGGCGTGCTAGT	CTCAATTTGAATCTTGGAGC	194
665	HvSSR10-15	TAT	51	7272167	AATCACTTACATGTGGGACC	ATATCCAAGCGAATTCACAC	176
666	HvSSR10-16	AT	54	7300898	AAGTTCTCCTTGGTCTTGGT	TAGCGATGCTTGTGCTATAA	398
667	HvSSR10-17	AT	66	7332139	CGTCTTGAATCAATTTCCAT	GATTGCCCGTAGAACTATTG	336
668	HvSSR10-18	AT	57	7360723	CCCTTGTCTCCTCACAGTAG	TCACACTCACACCTCTCAAA	296
669	HvSSR10-19	AT	66	8074507	TATGCACGAATCTTGAAACA	GCATTTCCTCAAACTGTAGG	318
670	HvSSR10-20	TA	58	8517310	AATCACGCACTGGTGTAAA	CACAAGGCTTATTCCATCAT	316
671	HvSSR10-21*	TTC	66	8589125	AGAAAGAGCTTCCCTAATGG	GCAAGATTCTCCAGCATAAC	322
672	HvSSR10-22	AT	56	8896081	ATTCAAATACTGGCACCAAC	TCTTCTCTCCTGCAGTCAAT	260
673	HvSSR10-23	TA	61	10651773	TATCAGGACGCAATTCTTCT	AGGACAGTATAGGCACATGG	376
674	HvSSR10-24	TA	61	10974687	GTAGTGCTGGTGGTCAATCT	CTAAACAAATGCCTGATTGC	399
675	HvSSR10-25	TA	53	11703682	ATGGACGTTGAGAGAAAGAA	TACGCACACTAAATTGCATC	294
676	HvSSR10-26	AT	53	11967024	ACTAGATCCAGTCCGTTTCA	TGTCAACAGACCATTAGCAA	190
677	HvSSR10-27	AT	52	12429023	GTAGAGCACAATGATGCAAA	CATATCCAACGACCTCACTT	361
678	HvSSR10-28	AT	54	12810129	TGGTTCTTTAGCCCAGTTAG	GCACAACTTCTGCTTTCTCT	302
679	HvSSR10-29	TTA	53	13223380	TAAGGGCAAACTACCATCAC	ACCGATGAGTCAGATTGAAG	307
680	HvSSR10-30*	AGA	57	13391127	ACATCTCCTTCAGCAGCTT	TGCTTCTCCAACAATAACCT	368
681	HvSSR10-31	AT	70	14084588	ATAAGAGTGGGAATGGGAAT	GATAGCGTTAGCCATAAACG	288
682	HvSSR10-32	AT	56	14391009	TTACAATCAGCAACGACAAG	TTGATCTATAGGTGGGTTCG	369
683	HvSSR10-33	TA	58	15587427	CGACCACGAACTAATAGGAG	GGCGAATTTGTGTATTTCAT	323
684	HvSSR10-34	TA	55	15905373	TAGACCGAGGAATTGAAAGA	TTTGGGCTTATTGTCAGTTT	202
685	HvSSR10-35	AG	62	16386862	AGCTTGTTGGCAGTTTGTAT	ACACTAAGTGGGCAAATCAG	287
686	HvSSR10-36*	TA	68	17501537	TACACGGTAGTCTCCCACTC	TGCTAAGGGTCTTAATTCCA	384
687	HvSSR10-37	TAG	65	17676178	CTAGAAAGCCAAGAACGAGA	AGCCTCTCAACGCTTAGTTA	322
688	HvSSR10-38	TC	70	17932763	CTCCTCCAGTTCATCTTCTG	AGGAGTGCAGTGATTTGTCT	334

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
689	HvSSR10-39	TA	60	18397449	TATAGGTGTCGGTTGGAGTT	AGTGTCAGCAAGATAGGGAA	370
690	HvSSR10-40	AGA	54	18821233	GACAAGTGAGGGTGGCTG	CTACCGAAGGACCTAATGTG	323
691	HvSSR10-41	TA	70	19854657	CATATTGAAGTTACCACGCA	AAACTTCGTGCTCTTTACGA	363
692	HvSSR10-42	AT	62	19872689	TGGTAAATACTCCATCCGTC	CATGAAAGTGGTCTACCGAT	283
693	HvSSR10-43	TC	61	21314945	AACACGATAGGTAGCTGGAA	ATGTCAGGATAAGCAAGGAA	380
694	HvSSR10-44	TA	60	22367615	GTGAACCACCAAGTACGAGT	CATCCCAACGATTTGTTTAT	300
695	HvSSR10-45	AT	57	22381868	GTGCATTAGATACCATGTGC	ATCGAAGGGTAAGACGAGAT	176
696	HvSSR11-01	AT	61	60291	TGTGTGTCCGCATACTTAAA	ATGTCAAAGTCCGAAAGTGT	214
697	HvSSR11-02	AAG	54	364257	TAGATTGGGTGATGGATAGC	CTACTTGATCCAGGGAAATG	372
698	HvSSR11-03	TA	65	576351	GGTTGACACCGTTAACATTT	TGGAACTACCTACCTAGCCA	351
699	HvSSR11-04	AT	68	780582	TGAATGCACACACAAGTTCT	GGCAGTTATTTCTGTATCACG	381
700	HvSSR11-05	GA	57	903714	ATCAAAGTCAAACAGCGATT	AAGCTACCTTCACTTTGCAG	298
701	HvSSR11-06	TA	66	1004384	GATGGTGGAGGTAAGTTGTG	GGTATGTAATCATGGATTGTGA	354
702	HvSSR11-07	TA	53	1338334	GTAGGGCACTTGTCAGTTCT	AGTTCGAATTACGCTTGAAC	281
703	HvSSR11-08	AAT	56	1494068	ACCAGAACCTCTAGGAAAGG	GTATGTCCTACTGCCACCAT	266
704	HvSSR11-09*	GA	53	1558712	TAGTTGAGGTACGCGAAGAT	AGCCTGTCTTGTTGTACTCG	306
705	HvSSR11-10	AT	57	1959653	TTCACCACTGAAAGAAAGGT	ATGTATGTTTCCCTCTCCCT	386
706	HvSSR11-11	CT	67	2257017	CTCCTCACTCCATCATCAAT	AAGAACTCCAGCAGTGAGAG	254
707	HvSSR11-12*	AGA	51	2480030	AACCTCATCGTATCCTCCTC	ACTAGTGAGTCCCACGTGTC	207
708	HvSSR11-13*	AG	56	5919351	TGAAACCACAATGAGTCAAA	GCCCTAAACCCAAATAGAAG	285
709	HvSSR11-14	AT	52	5952398	TCGAACACGGTACAATCATA	TCATGTTCTTTCTTCGCTTT	350
710	HvSSR11-15	AT	58	6079370	TCGAAGTCCAAATTTCCTTA	GATTGGAGTTTGAACGATGT	172
711	HvSSR11-16	TA	68	6256758	TATGGGAGGAGATGGGTTA	TGTTACTGCATAGTGGTCGT	365

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
712	HvSSR11-17	TA	66	6853555	GGCATTCTTCTTATTGGTGA	TTAACTTGGATTATTGGCGT	275
713	HvSSR11-18	AT	54	7631172	GAGTTTACCACTCACCTCCA	TGCTAGAGATTTCAGCTGTG	262
714	HvSSR11-19	TA	61	7977957	GGCTGTGCTAGCTGTTTAGT	TATACGCCAACGATGTATGA	367
715	HvSSR11-20	TA	60	8264894	TACTGGACCAGGCGTATAAT	TATGGACCGTCCTAGTTCTG	393
716	HvSSR11-21	ATCT	61	9100879	GAGATCAGTTGTTCCCTTGA	TTGAGAGCGTTTATAGGGAG	198
717	HvSSR11-22	AT	67	9233362	TACGCTATAACCATGAAGCA	CTCCCGTTATTGTCCTTACA	291
718	HvSSR11-23	AT	53	9963880	TTCTTGTTCGAGCTAGGAAG	ATCTACCAAATTTGTCACGC	151
719	HvSSR11-24*	CTT	66	16749699	GGTTCCCAATGCAGTATAGA	CACTAACGACCAAGGTAAGG	146
720	HvSSR11-25	AT	57	16872147	AACTGAGCACAAATCAGGTC	CCATTAGTTTGGTTCCACAT	367
721	HvSSR11-26	AT	55	17151433	ATACGCACTATGATCCCAAA	GCCATGAACCAAATCATAAC	365
722	HvSSR11-27	AAT	52	17236080	CTTAGAATCTGTCCCTCCCT	ACTACATACCGTGAACTGGG	305
723	HvSSR11-28	ATA	60	17544779	TTGCAATGGTCTAACTCCTT	TCCGTTATGAGATTGTCCTC	380
724	HvSSR11-29	GA	57	17847570	CCACAAGAGGAGTCCGTG	GGATAGGATAGGATACCGCT	206
725	HvSSR11-30	AT	62	18201199	ATAGTCTCCTCGCGTAGTTG	TCAATTTCAATTATGGTGCC	313
726	HvSSR11-31	ATT	55	18581326	GACACAGTTACAAAGCCACA	GGTGTGCTCCTACTTTGTGT	192
727	HvSSR11-32	TA	66	18751728	TTGTTGAGAGCATACGATTG	TATGCTGCCATAGCTTTCTT	203
728	HvSSR11-33	TA	54	19008710	AACACAGATGGAAAGAGAAGA	AGTGGTTACAAAGCAGTATTT	395
729	HvSSR11-34	AT	65	19021541	ACGGTGGATAAGAGCAGG	TGAAAGACATTACACTACTTGA	214
730	HvSSR11-35	AT	62	19201561	AGCCTCTCACCACCATCT	ACCGCCTACCTCCTGTAA	311
731	HvSSR11-36	TA	65	19375846	AGGACAGGACGCGTATAGTA	GGGACGAAAGAACTGTAGTG	285
732	HvSSR11-37	AAG	56	19377364	CCGTCACCTTCACTTACTTT	CCTACGAATACAAGCCGTT	231
733	HvSSR11-38	TA	70	19630866	CTCATCCTCAGTTTGTGTCTG	CGGTCTATGTATTGTTGTTCTT	394

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
734	HvSSR11-39	CCT	55	19685749	AGGGCTCACCAGTTGCGG	GAGAGATGTTTGATGCGG	319
735	HvSSR11-40	TA	51	19875688	ATGCTATTTCTTCCGTTTCA	ATACTACCTCCGTTTCAAGT	151
736	HvSSR11-41	AT	56	20179713	TTTCTCTCTTGTTTCTCCTTT	TGACTATCCTACCCTTGGTTT	320
737	HvSSR11-42	AT	64	20340114	ATGTATTTGGGTGCTGTC	AGACGCTTGTGGTGACTTT	316
738	HvSSR11-43	AT	70	20461102	TTGAACTTAGGCAAGGCA	TACTTATTGGATTGGTCTCC	207
739	HvSSR11-44	AT	68	20466968	ATCTTACCTTGCTTGCTCTTG	TTTCGGACACACTACACAACT	373
740	HvSSR11-45	ATCT	58	20563884	TAGCAACACAGAAAGGAAGAA	TAGGAGCGTTTGTAGGAAGT	209
741	HvSSR11-46	TTA	51	20597922	CGTTCGTTGTAATGGGTG	TATGTGTTTGAGTCCGTTT	338
742	HvSSR11-47	AT	70	20835702	GCTTACGACATTGGATTATTT	CCCGCCGTGTTCTGTATC	341
743	HvSSR11-48	AT	64	20953701	GCGTGATGAAATGGTGTC	TGCCAAGGAAATAAACTAAA	361
744	HvSSR11-49	CT	53	21004341	CAGCCTTCTATTTCTTTCACA	ATCTGGGTCTTATGGAGGAGT	320
745	HvSSR11-50	TATC	59	21049152	AACATTTCTGACGGTGAA	ACAACTCTCTCAATCAATCTAC	382
746	HvSSR11-51	TA	63	21464449	ATGCTTTAAGTTTCGCAAAG	GTTGGCCATTGTTATATGGT	253
747	HvSSR11-52	TC	68	21466513	GTGAAAGAAAGCCACCAAG	GGCACCAGCACACTACAC	268
748	HvSSR11-53	TC	67	21979552	TAGCTATGTACGTGCTGGTG	TCCATTTAATGCTCCATTTC	324
749	HvSSR11-54	AT	63	22206013	CCTCTACGATCACTTTGTCC	TCAGTTCAAACTTGCTTTCA	310
750	HvSSR11-55	AT	54	22289577	AATTTCTTGTCTCCACAACG	TTTCGTTGGTATGCACTATG	231
751	HvSSR11-56	AT	70	22326462	CGTTCACACTCTCACAAATAC	ACGCTACCATCAAGCAGA	397
752	HvSSR11-57	TAT	58	23235118	TGTAAGTCTGTAACCATCCTACT	TAACATCGGAAGAGGTGAA	377
753	HvSSR11-58	AT	70	23399950	TCTATCTCTCATCCAACTCCC	TTACAAAGAAGTGCTCCCTC	362
754	HvSSR11-59	AT	68	23471346	CGTGTCAGTCATCCATCT	GGAATACAAAGTGGTCTGCTC	222
755	HvSSR11-60	TA	58	23703162	CTTCTTTGTTTCGGTTCTATG	TCTACACCTTCAGTCCTTCCT	213
756	HvSSR11-61	AT	52	23733178	ACTGAATCCTTACTGGAGCA	GGAGATAAGCATTTGGAAGA	371
757	HvSSR11-62	AT	70	23742354	AGAGAATCATCAAAGGACAAG	AATCACTACACAAACATACAAGA	245

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
758	HvSSR11-63	AT	56	23805606	CTTGTCTTCTCATTCCTCTTG	ACACACACACACACACACAC	325
759	HvSSR11-64	GT	53	23806053	GAAAGTTGGTGGTGTAATGG	GCTGCGTAAGTTCAGATT	263
760	HvSSR11-65	TA	69	23840501	GCAAGTCATAGAGATTGAGGA	GTTTAGGAGCCAGAGGCA	394
761	HvSSR11-66	AT	64	23900705	TCTGCTCAAATGTTCGTGT	TTGTCTTGTTACTGTCCGATG	334
762	HvSSR11-67	AT	70	24083762	GTCAGTCATTCCTCTTCTG	CGAACCATCTTTAGTCCC	400
763	HvSSR11-68	AT	63	24130806	TCTTCCTTGAGCTCACTAGC	CAAATTCAGTTCAAACTTGCT	320
764	HvSSR11-69	TA	53	24136735	ATTGACACTTCTGCTGCC	TAAGGATTAGCGTGGAGA	398
765	HvSSR11-70	ATCT	68	24183454	TCCTCGTGTCATCGTGTT	TAAGAGCGTTTGTAGGGAGTT	272
766	HvSSR11-71	AT	70	25851157	ACCTCTAAGTTGTATGATGGTAG	GCTTCTTCTTTCCTTTCTCTT	275
767	HvSSR11-72	AT	56	24526291	CACCTGAATACTACCTCCGTT	AGAACCACACACACACACAC	389
768	HvSSR11-73	TA	65	24626314	AGATAAATAAGGGAGAGCCAA	CTACTTCAACAAGAGAGGCAA	284
769	HvSSR11-74	GATA	53	24855584	TCAGAGCGTGTCAGTTAGTTT	ACAAGATTTAGGTTTAGCGTG	281
770	HvSSR11-75	AT	70	25075529	TGCCTTCTTTATCATCACATT	TTGCCATTTGTTCTTTATCC	388
771	HvSSR11-76	TTC	68	25326643	CGACCTCCGAACACAGCC	TTCTTCTCCCAACCCTTC	328
772	HvSSR11-77	AT	56	25772721	GAGGACGGCGATGATTGA	TACTAATGGACGGTGGATG	378
773	HvSSR11-79	ATAC	66	26256429	ACTGCTCTGTATCTGAACCTG	AAATGTGTGTGTGTGTGTGTC	310
774	HvSSR11-80	TA	69	27170154	TATTAGAGAAAGGGCGAAGAA	AGGGTCAAGAGACACACAGA	256
775	HvSSR11-81	TAT	69	27447301	AATAGTCCCACGTGAACATC	AGAATTAGTTTCACGAAGCG	396
776	HvSSR11-78	ATCT	70	27581539	ACATCTTTCTGTTTAGGGTAGT	GAGGTGGATTTGGTTGTTT	348
777	HvSSR11-82	TTA	59	27752168	GGGATTCTTCTTCTTGTGTTT	CAGGATTATGGATGGGTTT	285
778	HvSSR11-83	ATAG	66	27772548	CAGTGAATGTGGTGCTTT	AAGTGAGGCAATAGAGATTAG	384
779	HvSSR11-84	CTAT	53	27872755	AATCAGGAGAAAGAAGAGCAG	GTAGGAGCGTTTGTAGGAAGT	214
780	HvSSR11-85	AT	70	28067767	CAAGTAATCCCTCCGTTCTA	ACATTGATGGCAGATTTGAT	363
781	HvSSR11-86	AT	66	28259196	GCTAATCACATATCGACGGT	AGCCATGAGAATCAGAAAGA	381
782	HvSSR12-01	AT	57	598151	GATTTGCAACACGTACGATA	GATCATCCACTCTGAGCAAT	271
783	HvSSR12-02	AT	55	1104532	AGTTGAATGTGTTCTGGACC	AGTCATTCTAGCATTTCCCA	240

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
784	HvSSR12-03	AT	52	1356493	AGTGGAACTGGAATGAACAC	GTGACTCATTTACCGCTGAT	387
785	HvSSR12-04	TA	56	2339606	AAGTAGTGCGTTTGTCAGGT	GATCCATACAAAGTCCGAAA	375
786	HvSSR12-05	TA	55	2340115	CTAGGATCGGATGGGATATT	CGACCACTATTTGCTAATCC	300
787	HvSSR12-06	AT	61	2938563	GTAACAGAATCTGCCTGCTC	CGAAGGGTCTTGTGATGTAT	315
788	HvSSR12-07	TA	59	2951388	GAAGGTTGGTTGATGTGAAT	CGACAGATTTGTTTGCTACA	312
789	HvSSR12-08	AT	52	3031885	ATTTGCAGTGTCTAGTGCCT	CTACTGAACATGCTCGATGA	252
790	HvSSR12-09	TA	69	3071546	GTTTGGAGTTGAGATTTGGA	CTAACTAGCAGAGCGTCACC	325
791	HvSSR12-10	AG	57	3185497	TACACGATACACCCACACAC	CGTGGTAGAAGTTGACCAAT	206
792	HvSSR12-11	AAT	70	3284181	TTGGTATTGTTATGTGCAGG	AAAGCCAACCATGTTTATTG	376
793	HvSSR12-12	AT	58	3435567	AAGGCTTTCAAGAATTGTCT	TACTTCGGCAAATTCGTTAC	356
794	HvSSR12-13	TA	66	3553768	ACCTTAGGGCTGAGTTCTTT	TTAGGCTTGTCTCTTCCTCA	388
795	HvSSR12-14	TA	53	3803527	AGGGAGTGAACACAAAGATG	GATGATCATGTAGCCATCAA	288
796	HvSSR12-15	AT	63	3813870	TGCATGTGCTATGCTTTACT	TAAATTTCAATCATGGTGCC	184
797	HvSSR12-16	TGTA	52	4828862	TTGGTGCTCTTGGATTATTT	AACATGAGAGGCAATAGGAG	373
798	HvSSR12-17	TA	59	5001194	TGGAATATCAATTGTGCGTA	AGTCCATGAACTTGCTCAAT	374
799	HvSSR12-18	TA	65	5187927	GAGACGCTAGAAGACGCTAA	TGGGATGGGAGTGTAGAATA	239
800	HvSSR12-19	AT	59	7565173	AGGCCTTCTTTAATTGATCC	CGATCTATCCATGAGCAAAT	298
801	HvSSR12-20	TA	65	9276214	CAAGTACGTCTTGGTGTTCA	GATGAAGATGATGTGGGAAC	315
802	HvSSR12-21	AT	52	9570313	ATATAACCGCATGGAGACAC	ATTATTCCTTGTTGCCGATA	287
803	HvSSR12-22	TA	69	9575518	GGAACCCTAATGTGAGAACA	GCGCACATATAAAGGCTAGT	387
804	HvSSR12-23	TA	60	10203639	GAGTACTTCAGATCCGGACA	TATTACAACGGGACGCTAAT	334
805	HvSSR12-24	ATT	70	10854308	TATTCCCTTCGGTTCTGTTA	GAGGGACATGATGGTTTAGA	308
806	HvSSR12-25	TA	57	12119972	AAGTTTGCAATGGAGGAATA	AAATCTTAGGCCAGGGTTAC	306
807	HvSSR12-26	TA	64	12237099	ATGAAATTGAGTTTGGAACG	CGTCATCTGTGTATGAGTCG	305
808	HvSSR12-27	TTA	57	12468939	TAAACAAGATGGCCAACAG	ACTACCAATGATTACCGTGC	368
809	HvSSR12-28	AT	54	12866441	AACCAGTAGGGTGGATAGGT	TTGGTTAATAAATTCAGCCC	372

S.no.	HvSSR ID#	Motif	Length	Start position on pseudomolecule (build 5)	Forward primer	Reverse primer	Product Size
810	HvSSR12-29	AT	63	16052671	GAAATATTTAGTTCCGGACG	TAATCAAACCTGGAAGAGGA	377
811	HvSSR12-30	TA	59	16517784	TGATTTGGGAATCTATCTCG	GGTATGTAATCATGGATTGTGA	378
812	HvSSR12-31	AT	66	19138619	GACATAAGGGACAATCCAAA	TCAGATCATGATGCCTACAA	340
813	HvSSR12-32	TA	61	19403091	TCAATTTCATGGTATCAGCA	AGTGTATTCAGTTGGGCTGT	364
814	HvSSR12-33	TC	63	19426521	AAATTCACTTGGTTGACACC	CTCAAGGGAAATCACAGAAG	374
815	HvSSR12-34	AT	66	19906994	GTTCTTATGGTTATCAGCCG	TAGCATTGCCCACATACATA	182
816	HvSSR12-35	GA	63	19929359	ATGACCATAATCCCAACAAA	GTCGTGGTGTTATTCTTGGT	302
817	HvSSR12-36	AT	64	21140995	ATCAGCGACTAAGGATCTCA	CTAATGTTGCCACATACGAA	368
818	HvSSR12-37	TTA	68	21166514	ACAAATGTTGGACTTCATCC	ACTCATTGCTTTAACATGCC	385
819	HvSSR12-38	TA	62	22195606	ATGAGGGAGAACATCATTTG	CATGTTTGACCGTTGTCTTA	336
820	HvSSR12-39	TA	57	22511745	TGTATGTTCATCCTCCGTTT	AATGTATCTGGACGGAAGG	228
821	HvSSR12-40	TA	61	22613088	ATCTAACAACAACAATCCCG	CATCTTCATCCCTCGTGTAT	289
822	HvSSR12-41	ATA	67	22708926	GCCATTGTCACTTTGATTCT	GAAGATCACGGATAGATGGA	346
823	HvSSR12-42	GA	53	23414403	GATCTGATGTCGTCTTCCAT	AGAGATAGAGACGGAGGTGG	339
824	HvSSR12-43	TA	62	23953958	GGTATAATGACAGAGCTGGG	TCGTAACGAGGTCAAGATTT	341
825	HvSSR12-44	TA	63	24473298	CTCTTTCCTTGGTTTGTCAG	AAATGGTGATGTGGCTTAAC	344
826	HvSSR12-45	AT	62	24859195	AAGTTGCAGACAGGACTGAT	GGTATGTAATCATGGATTGTGA	309
827	HvSSR12-46	AT	66	25352959	TAGCACACAGTGGAAGTACG	CCGATTATACCATTGTATTCGT	364
828	HvSSR12-47	TA	67	26153853	TGTTCTGGAGGTAGGTATGG	CAGCAACTCAAAGTATGCAA	261
829	HvSSR12-48	TA	67	26230080	AAACTCGATCAGACTTAGAGAAG	TCTCTGATGGCAATACAACA	326
830	HvSSR12-49	AT	53	26408057	TTTCCATGGCTTATTAGCTT	CTCAATTTGAATCTTGGAGC	161
831	HvSSR12-50	AT	52	26408501	TTTATACCTTAGAGTCCGGG	TAAGCTAGGGTATGGTTGGA	187
832	HvSSR12-51	TA	59	27025266	AATCATCATATTGCCGAAAG	ATCACCATCTATCATTGCAC	288

SSR id HvSSR represents hypervariable simple sequence repeats followed by (*) genic regions, chromosome no. (pseudomolecule build 5) and consecutive serial no. espectively