GUAR LEAF TRANSCRIPTOME ANALYSIS AND EVALUATION OF GUAR GUM DRUG DELIVERY SYSTEM

Ph.D. THESIS

by

UMESH KUMAR TANWAR

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247667 (INDIA)
FEBRUARY, 2016

GUAR LEAF TRANSCRIPTOME ANALYSIS AND EVALUATION OF GUAR GUM DRUG DELIVERY SYSTEM

A THESIS

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

of

DOCTOR OF PHILOSOPHY

by

UMESH KUMAR TANWAR

DEPARTMENT OF BIOTECHNOLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247667 (INDIA)
FEBRUARY, 2016

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **"GUAR LEAF TRANSCRIPTOME ANALYSIS AND EVALUATION OF GUAR GUM DRUG DELIVERY SYSTEM"** in partial fulfillment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the Department of Biotechnology of the Indian Institute of Technology Roorkee is an authentic record of my own work carried out during a period from December, 2009 to February, 2016 under the supervision of Dr. G.S. Randhawa, Professor and Dr. Vikas Pruthi, Professor, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee.

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

(UMESH KUMAR TANWAR)

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

(Vikas Pruthi)
Supervisor

(G.S. Randhawa)
Supervisor

The Ph. D. Viva-Voce Examination of **Mr. Umesh Kumar Tanwar**, Research Scholar, has been held on .....

Chairman, SRC

Signature of External Examiner

This is to certify that the student has made all the corrections in the thesis.

Signature of Supervisors

Head of the Department

Dated: _____

ABSTRACT

Guar or clusterbean is an economically important crop because of galactomannan gum obtained from its seed endosperm. Genetic improvement in industrially important guar (*Cyamopsis tetragonoloba* [L.] Taub.) crop has been hindered due to the lack of sufficient genomic or transcriptomic resources. In this study, RNA-Seq technology was employed to sequence and characterize the transcriptome of leaf tissues from two guar varieties, namely, M-83 and RGC-1066. Approximately 30 million high-quality pair-end reads of each variety were generated by Illumina HiSeq platform and used for *de novo* assembly by Trinity program. A total of 62,146 non-redundant unigenes with an average length of 679 bp were obtained. The quality assessment of assembled unigenes revealed 87.50 % complete and 97.18 % partial core eukaryotic genes (CEGs). Sequence similarity analyses and annotation of the unigenes against non-redundant protein (Nr) and Gene Ontology (GO) databases identified 175,882 GO terms. The species distribution analysis of the unigenes showed highest similarity with *Glycine max* genes. The comparison of assembled unigenes with the sequence data of closely related sequenced species on the basis of meta annotation, gene family and functional annotation showed 63 %, 55.9 % and 56.7 % similarity of unigenes to the sequences of *Glycine max*, *Medicago truncatula* and *Lotus japonicus*, respectively. A total of 11,308 guar unigenes were annotated with various enzyme codes (EC) and categorized in six categories with 55 subclasses. The annotation of biochemical pathways resulted in a total of 11,971 unigenes assigned with 145 KEGG maps and 1,759 enzyme codes. The differential gene expression analysis showed ~80 % similar gene expression in both the varieties; 2,863 unigenes were found to express in variety M-83 only and 2,120 unigenes in RGC-1066 variety only. Approximately 175 and 158 unigenes overexpressed in RGC-1066 and M-83 varieties, respectively. A total of 5,773 potential simple sequence repeats (SSRs) and 3,594 high-quality single nucleotide polymorphisms (SNPs) were identified. Out of 20 randomly selected SSRs for wet laboratory validation, 13 showed consistent PCR amplification in both guar varieties. *In silico* analysis of SSRs resulted in identification of 145 polymorphic SSR markers in two varieties. In addition, 2,930 and 3,984 Insertion-Deletion (InDel) variations were found in the varieties M-83 and RGC-1066, respectively. Guar gum based drug delivery system was formulated by two methods and evaluated for colonic drug release using 5-fluorouracil as a model drug. The microparticles were spherical in shape with a size range of 7-30 micrometer. The study of drug release was carried out in PBS buffer (pH 7.4). The drug was found to be released after 6 h in both preparations of microparticles.

Keywords: Guar gum, Next generation sequencing, Transcriptome analysis, Molecular markers, Microparticles, Colon specific drug delivery

ACKNOWLEDGEMENTS

It is the contribution of one or more persons towards the task, which makes it successful. It gives me immense pleasure in acknowledging all the help that I have received during the period of research.

First and foremost I express my deep sense of gratitude to my supervisors ***Prof. G. S. Randhawa*** *and* ***Prof. Vikas Pruthi****, Department of Biotechnology, Indian Institute of Technology, Roorkee. They were not only a teacher and guide to me, but my mentors and well-wishers too. Their persistent encouragement, perpetual motivation, everlasting patience, constructive criticism and valuable technical inputs in research have benefitted me to an extent which is beyond expression. They have not only trained me in science but in all aspects of life. I would like to take this opportunity to also thank* ***Dr. Mrs. Surinder Randhawa*** *who has been supportive in all the efforts during this time period.*

I wish to acknowledge my deep sense of gratitude for ***Prof. Partha Roy****, HOD, Department of Biotechnology, IIT Roorkee, for providing all the necessary resources and lab facility. I also express my gratitude to* ***all the faculty members*** *of Department of Biotechnology and Institute Instrumentation Facility, IIT Roorkee for their support, guidance and timely help for my research.*

I would like to thank ***Prof. Sudesh Kumar****, RARI, Durgapura, Jaipur and* ***Dr. Surender Pahuja,*** *CCS HAU, Hissar for providing the research materials. I also take this opportunity to thank* ***Dr. Nagesh K.A.*** *for his timely support, encouragement, goodwill and all helping efforts during this time period.*

I would like to take this opportunity to extend my gratitude to my seniors ***Dr. Durga Prasad Panigrahi, Dr. Megha Agrawal, Dr. Pranita Bhatele Tiwari, Dr. Naincy Girdharwal, Dr. Shailender Verma and Dr. Satish Khatkar*** *for their timely advices, encouragement, guidance and goodwill which enabled me to start and successfully complete my work.*

I would like to express profound thanks and make a special mention of my colleagues ***Navneet Kaur, Deepa Dewan, Reeku Chaudhary and Dhramendra Singh*** *for assisting me at every stage of this research work. I also express my gratitude to the other lab members* ***Dr. Shilpi, Manisha, Dr. Swati, Dr. Shalini, Pallavi, Nishu, Poonam, Manjeshree, Preeti, Omika, Archana, Tsering and Prabhjot*** *for maintaining a friendly environment in the lab.*

I express my thanks to my friends ***Dr. Prabhat, Dr. Shon Suantak, Dhruv, Dr. Bibekananda, Pranav, Rajat, Madhusudhan, Vishvajeet, Rakesh and Brij Kishor*** *for supporting me at every stage of my work.*

I would also like to thank ***non-teaching staff*** *at Department of Biotechnology, IIT Roorkee for their supportive nature and help in processing all the documents.*

I express my gratitude to my previous teachers ***Ms. Saroj Rathi, Mr. Sanjeev Kumar, Mr. Vijay Yadav, Dr. Sumitra Singh, Prof. S. K. Singh, Dr. Manish Ahuja, Prof. U. C. Banerjee and Dr. A. K. Pandey*** *who inspired me to be in the field of science and research.*

I express my gratitude to ***Department of Biotechnology*** *(****DBT****), Government of India for the financial support in the form of JRF and SRF during the period of this research.*

I express my unbound gratitude to ***my parents, Sh. D. R. Tanwar and Smt. Somwati Devi,*** *for their blessings and dedicated efforts to educate me to this level. I also express my gratitude to all my* ***family members*** *for their constant support and encouragement.*

Finally I thank the ***Almighty GOD*** *for leading me all the way towards successful completion of this work.*

Date: ***(Umesh Kumar Tanwar)***

ABBREVIATIONS

˚C	:	Degree centigrade
%	:	Percent sign
µl	:	Microliter
µm	:	Micrometer
5-FU	:	5-fluorouracil
AFLP	:	Amplified fragment length polymorphism
AFM	:	Atomic force microscopy
$AgNO_3$	:	Silver nitrate
BAM	:	Binary vision of a SAM file
bp	:	Base pair
BLAT	:	BLAST-like alignment tool
CAPS	:	Cleaved amplified polymorphic sequences
CCSHAU	:	Chaudhary Charan Singh Haryana Agricultural University
cDNA	:	Complementary DNA
CD-HIT	:	Cluster database at high identity with tolerance
CEGMA	:	Core eukaryotic genes mapping approach
CEGs	:	Core eukaryotic genes
CTAB	:	Cetyl trimethyl ammonium bromide
cm	:	Centimeter
CMHPG	:	O-carboxymethyl-O-hydroxypropyl guar gum
CMHTPG	:	O-carboxymethyl-O-2-hydroxy-3-(trimethylammonio) propyl guar gum
dbEST	:	Database EST
DNA	:	Deoxyribose nucleic acid
dNTPs	:	Deoxy nucleotide tri-phosphates
EC	:	Enzyme code
EDTA	:	Ethylenediaminetetracetic acid
e.g.	:	For example
EST	:	Expressed sequence tags
et al.	:	et alia
FR	:	Forward to reverse
F	:	Forward
FPKM	:	Fragments per kilobase per million reads
g	:	Gram
G/Gal	:	Galactose
GB	:	Giga bases
GIT	:	Gastrointestinal tract
GO	:	Gene ontology
HPLC	:	High performance liquid chromatography

HTPG	:	O-2-hydroxy-3-(trimethylammonio) propyl guar gum
IBD	:	Irritable bowel diseases
IBS	:	Irritable bowel syndrome
IGV	:	Integrative genome viewer
IITR	:	Indian Institute of Technology Roorkee, Roorkee
ISSR	:	Inter simple sequence repeats
IUPAC	:	International union of pure and applied chemistry
Kb	:	Kilo bases
KEGG	:	Kyoto encyclopedia of genes and genomes
L.	:	Linnaeus
m	:	Meter
M/G ratio	:	Mannose/Galactose ratio
M/Man	:	Mannose
ManS	:	Mannan synthase
mg	:	Milligram
$MgCl_2$	:	Magnesium chloride
min	:	Minute
MISA	:	MIcroSAtellite identification tool
ml	:	Millilitre
mm	:	Millimeter
mM	:	Millimolar
mRNA	:	Messenger RNA
NaOH	:	Sodium chloride
NCBI	:	National Center for Biotechnology Information
ng	:	Nanogram
NGS	:	Next generation sequencing
nm	:	Nanometer
Nr	:	Non redundant
PBS	:	Phosphate-buffered saline
PCR	:	Polymerase chain reaction
PHGG	:	Partially hydrolysed guar gum
p-value	:	Calculated probability
R value	:	Coefficient of correlation
RAPD	:	Random amplified polymorphic DNA
RARI	:	Rajasthan Agricultural Reaserch Instititute, Jaipur
RF	:	Reverse to forward
R	:	Reverse
RPKM	:	Reads per kilobase per million reads
RNA	:	Ribose nucleic acid
RNA-Seq	:	High-throughput RNA-sequencing
rRNA	:	Ribosomal ribose nucleic acid
RNaseA	:	RibonucleaseA

RNaseH	:	RibonucleaseH
rpm	:	Revolutions per minute
RSEM	:	RNA-Seq by Expectation-Maximization
SAGE	:	Serial analysis of gene expression
SAM	:	Sequence alignment/map
SCAR	:	Sequence characterized amplified regions
sec	:	Second
SEM	:	Scanning electron microscopy
SNP	:	Single nucleotide polymorphism
SSCP	:	Single stranded conformational polymorphism
SSR	:	Simple sequence repeats
Taq	:	*Thermus aquaticus*
TBE	:	Tris borate EDTA
$TE/T_{10}E_1$	:	Tris EDTA
TRAPID	:	Tool for rapid analysis of transcriptome data
U	:	Unit
UniRef	:	UniProt reference clusters
USA	:	United States of America
UV	:	Ultraviolet
v	:	Volt
WEGO	:	Web gene ontology annotation plotting tool
w/v	:	Weight/volume

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Guar (*Cyamopsis tetragonoloba* [L.] Taub.), also known as clusterbean, is an annual drought-tolerant legume crop belonging to the family Leguminosae. It is grown mainly in semiarid regions of India, Pakistan, and the United States. Guar has been traditionally used as a forage, green manure and vegetable crop [73]. In recent times, it has attained the status of an economically important crop because of the gum contained in endosperm of its seeds. Guar gum contains about 90% galactomannan and is one of the most cost-effective natural thickeners [67]. It is used in textile, paper, petroleum, explosives, cosmetics and pharmaceutical industries [300]. Additionally, guar gum is used in the treatment of diarrhea, irritable bowel syndrome, diabetes and high cholesterol [37, 86, 249].

Guar gum has been highly exploited in pharmaceutical industry. It is mainly used as a gelling, viscosifying, thickening, suspending, stabilizing and emulsifying agent in many dosage forms. Guar gum and its chemically-modified derivatives are used as a binder and disintegrating agent in tablets and as a suspending, thickening, gelling and stabilizing agent in liquid/semisolid oral and topical dosage forms [200, 227]. These are also used in tablet formulation to mask the unpleasant taste and odor of drugs, and improve stability and drug release properties [200]. Guar gum is hydrophilic in nature and swells in cold water forming viscous colloidal dispersions or solutions. It is a non-ionic polymer and hence the viscosity of dispersion is unaffected by pH and is same in acidic and alkaline media. Guar gum and its derivatives have been extensively used in colon specific drug delivery due to their drug release retarding property and susceptibility to microbial degradation in the large intestine. These shield the drug from the environments of stomach and small intestine and deliver the drug in the colon. There are several reports on guar gum as a colon specific drug carrier, however a little information is available on the guar gum-based multiparticulate drug delivery system.

India is the largest producer of guar and accounts for 80 percent of total guar production in the world [1]. Total production of guar in India was estimated to more than 2.7 million metric tons during the agricultural year 2013-14. The increased demand of guar gum globally in recent years has led to crop introductions in several countries including South Africa, Australia and Brazil having varied climates and seasons [269]. Therefore, improved guar varieties for wide range of climatic conditions are needed to be developed through breeding programs.

Molecular markers have been found useful in breeding programs with marker-assisted selection, resulting in reduced time and effort for developing improved varieties [124]. These markers are used as a tool to detect genetic polymorphism at specific locus and whole-genome levels as they facilitate marker-based gene tagging, genetic mapping, map-based cloning, genetic diversity studies and phylogenetic analysis [124, 176]. Five molecular markers, namely, random amplified polymorphic DNA (RAPD), ribosomal DNA (rDNA), inter simple sequence repeat (ISSR), simple sequence repeat (SSR) and sequence characterized amplified region (SCAR) have been used in the study of molecular diversity in guar [143, 145, 146, 203, 204, 216, 243]. Among various markers, SSR and single nucleotide polymorphism (SNP) markers are very useful for genetic and plant breeding applications [100] but only a limited number of SSR markers are available in guar [143, 146] and no SNPs have been reported in this crop.

Next generation sequencing (NGS) offers novel opportunities in functional genomics, discovery of genes and development of molecular markers in non-model plants [287]. The massively parallel sequencing of RNA (RNA-Seq or transcriptome profiling) represents a powerful tool for transcription profiling, providing a rapid access to a collection of expressed sequences (transcriptome), as compared with traditional expressed sequence tag (EST) sequencing. The RNA-Seq technology has been successfully applied in several organisms including model and non-model plants [178, 185, 197, 289]. This technology can be used for cost-effective development of molecular markers such as SSRs and SNPs [288, 289]. These transcriptome-derived markers are expected to show greater transferability among closely related species than the genomic markers because of their presence in more-conserved transcribed regions of the genome [58]. These markers can also be used for comparative mapping and evolutionary studies [278].

Genetic resources of crop plants are essential for crop breeding programs [29]. At present complete genome sequences of several legumes including soybean, *Lotus*, *Medicago*, pigeonpea and chickpea are available [112, 234, 238, 274, 279, 307]. The genome and transcriptome sequencing of guar has not been yet done. Only 16,476 ESTs from developing guar embryos are available in National Center for Biotechnology Information (NCBI) database [143, 146]. Though guar is an industrially important legume crop, the breeding programs have been hindered due to the limited availability of genomic resources in this crop. The development of genomic resources

for guar is needed to support various breeding programs and molecular genetic studies at different levels.

With the above information in view, the current research work was planned with the following objectives:

1. Sequencing and *de novo* assembly of guar leaf transcriptome
2. Functional annotation and differential gene expression in guar leaf transcriptome
3. Identification of molecular markers in guar leaf transcriptome
4. Studies on guar gum based drug delivery system

2. REVIEW OF LITERATURE

The literature on relevant aspects of guar crop, guar gum and transcriptome analysis has been reviewed below under suitable headings.

2.1 Species of the genus *Cyamopsis*

Gillette (1958) recognized three *Cyamopsis* species, namely, *C. tetragonoloba, C. senegalensis and C. serrate* [88]. A fourth species *C. denate* was added to this genus after 2 years [266].

2.2 History of guar cultivation

There are two brief review reports and a book available on the history of guar cultivation [105, 141, 209]. Recently, another book has been published in 2015 on the physiology, genetics and cultivation of guar [201]. The genus *Cyamopsis* probably originated in Africa. The species *C. tetragonoloba* is considered to be originated by domestication of the African wild species, *C. senegalensis* which was brought to India probably through trades between 9^{th} to 13^{th} centuries A.D. [88]. The domestication process could have been taken place in the dry areas of the northwestern region of the Indo-Pakistan subcontinent [105, 201]. Guar was earlier grown in India mainly as a manure, forage and vegetable crop [245]. It was introduced in USA from India in 1903 for experimentation [106]. This crop gained the attention during World War II when there was shortage of galactomannan gum obtained from carob seeds. The search for domestic source of galactomannan gum found guar as an alternative source for galactomannan [12]. Now guar is grown mainly in India, Pakistan and the United States. It is also cultivated to a limited extent in Italy, Morocco, Spain, France, Greece, and Germany [95, 215].

2.3 Germplasm of guar

Germplasm is an important resource for any crop improvement programme. Natural variation is a huge and largely untapped resource, which has been subjected to selection over millions of years of evolution, with both basic and practical value, as well as the potential to break yield barriers of agricultural plants [114, 261, 309]. The variation in germplasm could be used as a useful source of important genes. It has been estimated that for most crop species, less than 5% of the biodiversity known to exist has been utilized in agriculture, particularly in the case of self-pollinated crops [261]. Much of the diversity present in living systems is probably adaptive [114]. In India, about 4,827 accessions of guar are available in National Bureau of Plant

Genetic Resources (NBPGR), New Delhi [66]. These accessions have been catalogued on the basis of accession numbers and characterized for important phenotypic traits like days to maturity, gum content, branching, pods per plant, seed size and disease resistance. Classical approaches using the donor cultivars as sources have led to the development of certain elite guar cultivars for cultivation [61].

2.4 Genetics and breeding of guar

The research work on breeding and genetics of guar has been recently reviewed by two groups [14, 141]. Guar has not been genetically well characterized as compared to other legume crops. The haploid chromosome number of guar is 7 [201, 206]. Genetic crossing is difficult in guar because of size and morphology of its flowers [141]. Guar flower is only 8 mm long and magnifying lens is required to emasculate the flowers. The flowers are cleistogamous which leads in self-pollination [87]. Although crossing techniques have been developed in guar [47, 87], yet very little success has been achieved in terms of efficiency of hybridization. Mutation breeding could be a useful tool in crops lacking the useful genetic variability to enrich the variation [14]. Using this approach, many mutants of guar have been produced with useful traits like gum content, increased yields and early flowering [247].

The inheritance of 5 characteristics namely growth habit, branching behavior, clustering pattern, leaf size and hairiness has been studied in guar [46]. The study revealed that each of these traits except branching behavior was controlled by a single pair of genes. The branching behavior exhibited digenic inheritance. The studies on partial male sterility in guar showed that this character might be controlled by two genes [255]. The genetic diversity and interrelationship in 40 genotypes of guar were studied by Pathak *et al.* [202]. In this study a considerable variation was found for 7 characters in the genotypes. It was concluded that the genotypic variation was due to high additive genetic effects. Table 2.4.1 shows the genetic studies in guar.

Table 2.4.1 Genetic studies in guar

S.No.	Description	References
1.	Analysis of mannan synthase (ManS) and cellulose synthase (Csl) genes in plant cell wall hemicellulose polysaccharide synthesis	[67]
2.	Functional genome studies of cellulose synthase-like A (CslA) gene family among guar and other plant species through microarray analysis	[155]
3.	Development of cDNA libraries and EST datasets of developing guar seeds	[188]
4.	Expression studies of mannan synthase gene of guar in *Medicago* through Affymatrix genome chip.	[189]
5.	Effect of alternative splicing on MADS box domain of plants including guar	[240]
6.	Analysis of self-incompatibility RNAse gene in guar and other plants through expression profiling	[3]

2.5 Guar gum

The guar gum molecule is a linear and highly anisodimensional carbohydrate polymer with a molecular weight on the order of 220,000. It is a galactomannan made of linear chains of a β-1,4-mannan as backbone to which galactosyl residues are attached through α-1,6 linkages [55]. The mannose to galactose (M/G) ratio determines the structure of the galactomannans that affects various properties and applications of the galactomannans [254]. The M/G ratio is ~2 in guar gum [67]. Figure 2.5.1 shows the chemical structure of guar gum.

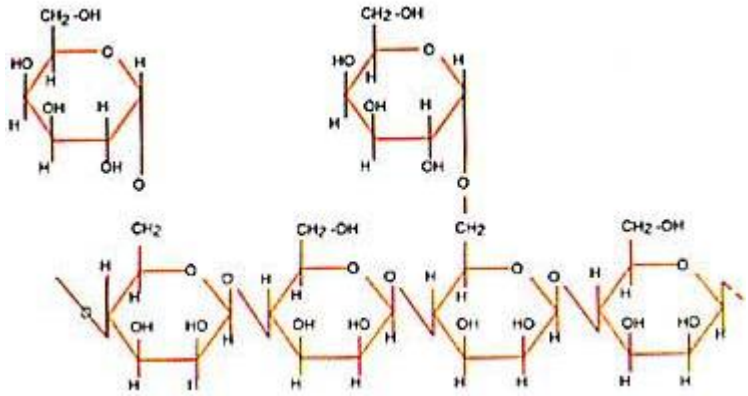


Figure 2.5.1 Chemical structure of guar gum.

2.6 Production and applications of guar gum

Guar gum is obtained by grinding the seed endosperms [300]. Fig 2.6.1 shows the flowchart of guar gum processing from its seeds. Depending on the degree of purification, various amounts of other seed tissues, such as residues of seed coat, germ and endosperm cell walls might be present in the obtained guar gum.

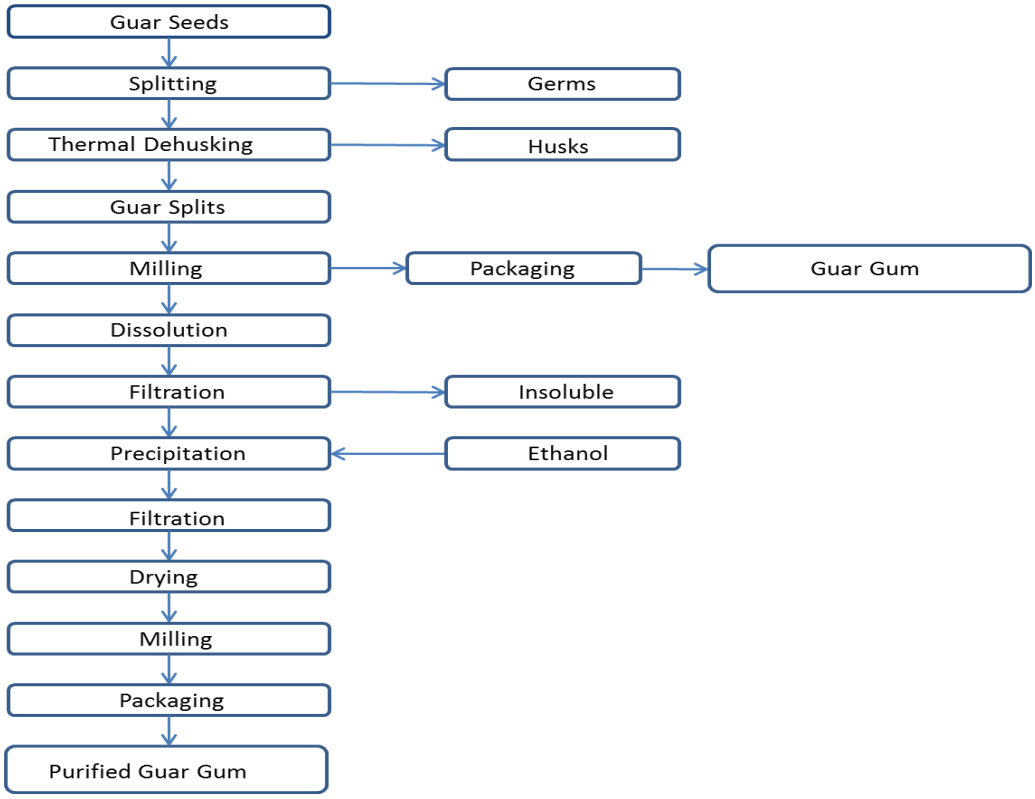


Figure 2.6.1 Flow chart of guar gum processing from its seeds. (Modified from Kawamura, 2008 [123])

Carob seeds were the main source of galactomannan gum used extensively in textile, paper and food industries in the USA before World War II [293]. The shortage of carob seed in the USA during World War II initiated the search for domestic source of galactomannan gum which found guar crop as an alternative source for galactomannan [12]. This information helped in the adoption of guar gum in different industries and guar has thus gained the status of an important industrial crop [73, 147, 201]. The applications of guar gum in different industries have been given in Table 2.6.1.

Table 2.6.1 Applications of guar gum and its derivatives in various industries

(On the basis of Kumar and Singh; Kuravadi *et al.* [141, 147])

S. No.	Industry	Functions/applications
1	**Petroleum Industry**	Thickener and gelling agent Provides viscosity stability Reduces friction and increases permeability Water loss improvement For plugging leakages Stable super-elastic liquid with lessened temperature sensitivity for control of lost circulation in oil field drilling operation
2	**Textiles Industry**	Thickener and stabilizer Printing paste thickener Foam composition Pigment retention aid
3	**Explosives Industry**	Water remover and plasticity improver Gelling and thickening agent Increases viscosity
4	**Paper Industry**	Improves retention for filter barrier Enhance surface and sizing Flocculent and sizing agent Imparts dry strength
5	**Coal mining Industry**	Stabilizer and dispersant For shock impregnation of coal seams
6	**Ore refining/metal Industry**	In settling fine particles colloidal Flocculent and binder
7	**Analytical Industry**	Separation of metal ions Support for immobilization of ligands Purification of lectins Chromatographic separation and selective resin for boron
8	**Food Industry**	Thickener and binder, Stabilizing and gelling agent Stable thixotropic stabilizer, emulsifier system Improves sedimentation To create pudding & cream dressing of cakes
9	**Pharmaceutical Industry**	Stabilizing and suspending agent, Binder/disintegrant and emulsifying agent Drug targeting to colon Insulinogenic and blood glucose lowering agent Cholesterol lowering agent
10	**Cosmetics Industry**	Suspending agent Binder and thickener Gelling and emulsifying agent

11	**Agriculture Industry**	Increased water retention capacity of soil Anticrusting agent and adhesive of azotobacter Prevention of granules
12	**Tobacco Industry**	Adhesive Reduces irritation & strengthening agent
13	**Others**	Thickener, binder, stabilizer and Adsorbent Flocculating and exchanging agent Gelling agent, foam stabilizer and thickening agent Water proofing Lubricant for installation of electric and telephone cable Electrical insulator

2.6.1 Guar gum in pharmaceuticals

Several workers have reviewed the applications of guar gum in pharmaceutical industry [55, 200, 211, 213, 242]. Guar gum is an extensively exploited polysaccharide in pharmaceutical industry. It is accepted for use as a food additive in Europe and included in the US-FDA inactive ingredients database (oral suspensions, syrups, tablets, topical preparations), non-parenteral medicines licensed in the UK and Canadian list of acceptable non-medicinal ingredients [227].

Guar gum finds limited use in natural form due to various concerns about its alcohol solubility, thermal stability and uncontrolled rate of viscosity and hydration [200]. The most specific property of guar gum is its chemical structure that has many hydroxyl groups suitable for chemical modifications [52]. A lot of research has been carried out on guar gum for development of its derivatives, with modified physical and chemical properties, by grafting, blending and compositing with synthetic\natural polymers. Some of the reported derivatives are carboxymethyl guar gum, hydroxymethyl guar gum, hydroxypropyl guar gum, O-carboxymethyl-O-hydroxypropyl guar gum (CMHPG), O-2-hydroxy-3- (trimethylammonia propyl) guar gum (HTPG), acryloyloxy guar gum, methacryloyl guar gum, sulfated guar gum, and guar gum esters [200]. Guar gum and its derivatives are used as a binder and disintegrating agent in tablets and as a suspending, thickening, gelling and stabilizing agent in liquid/semisolid oral and topical dosage forms [200, 227]. In tablet formulation it is also used to mask the unpleasant taste and odor of drugs, and to improve stability and drug release properties [200]. Guar gum hydrogels, prepared by crosslinking with different monomers, are also useful in various drug delivery systems. Guar gum is a prospective hydrophilic matrix carrier for oral controlled delivery of drugs [134]. It is mainly investigated for colon-specific drug delivery applications [26, 211]. Pharmaceutical applications of guar gum have been given in Table 2.6.1.1.

Table 2.6.1.1 Pharmaceutical applications of guar gum

Sr. No.	Name of drug	Formulation	Description	Reference
Colon specific drug delivery				
1	Dexamethasone and budesonide	Guar gum matrix tablets	• Negligible drug release in simulated gastric and intestinal fluid • Significant increase in drug release in simulated colonic fluid • Galactomannanase dependent drug dissolution	[297]
2	Indomethacin	Guar gum matrix tablets	• Only 21 % of the drug release in 0.1 M HCl for 2 h and Sorensen's phosphate buffer (pH 7.4) for 3h • About 91 % drug release in 4% cecal content medium after the enzyme induction of rats	[213]
3	5-amino salicylic acid	Guar gum compression coat tablets	• Less than 2 % drug release in simulated gastric and intestinal fluids • About 93 % drug release in pH 6.8 buffer containing rat cecal contents • High release rate in 11- 26 h	[137]
4	Mebendazole	Guar gum matrix tablets	• Only 8–15 % drug release in the physiological environment of stomach and small intestine *in vitro* • About 83 % drug release in simulated colonic fluids	[135]
5	Metronidazole	Matrix, multilayer and compression coated tablets of guar gum	• Comparative study of three types of guar gum tablets • Compression coated tablets found most suitable • Less than 1% of drug release in stomach and small intestine environment • About 61 % drug release in simulated colonic fluids	[136]
6	5-fluorouracil	Compression coated tablets of xanthan and guar gum mixture granules	• Highly retarded drug release • Xanthan gum: guar gum (10:20) found most suitable • About 83 % drug release in the presence of 4 % cecal content after 19 h of incubation	[248]

7	5-fluorouracil	Guar gum compression coat tablets	• Only 2.5-4 % drug release from tablets (80 % guar gum) in simulated GI fluids • About 70 % drug release in simulated colonic fluids (4%, w/v, rat caecal content medium)	[138]
8	Hydrocortisone	Phosphated cross-linked guar gum hydrogels	• Very little drug release in phosphate buffer pH 6.4 for 6 h • About 80 % drug release after addition of α-galactosidase and β-mannanase enzymes to buffer	[92]
9	Ketprofen	Guar gum/poly(acrylic acid) hydrogels	• Positive effect of hydrogel composition on drug release • Strong effect of pH on drug transport mechanism • Increased drug release simulated medium	[104]
10	Methotrexate	Cross-linked guar gum microspheres	• Less than 10 % drug release in phosphate-buffered saline (pH 7.4) and gastrointestinal • About 90 % drug release in rat cecal content medium	[48]
Antihypertensive drug delivery				
11	Diltiazem hydrochloride	Guar gum matrix tablets	• Slow and prolonged *in vivo* drug release • Better controlled release as compared to commercial diltiazem tablets	[8]
12	Trimetazidine dihydrochloride	Three layer guar gum matrix tablets	• A potential hydrophilic carrier in design of oral controlled drug delivery systems for highly soluble drugs	[133]
13	Verapamil hydrochloride and nifedipine	Cross-linked poly(acrylamide) -graft-guar gum hydrogels	• *In vitro* drug release dependence of extent of cross-linking, amount of drug and nature of drug molecule • Diffusion based controlled drug release	[253]
14	Diltiazem hydrochloride and nifedipine	Cross-linked hydrogels of polyacrylamide-grafted guar gum	• Controlled drug release process closely related to swelling of hydrogels in response to the pH changes	[252]
15	Diltiazem hydrochloride	Modified guar gum matrix tablets	• Maximum drug release in intestinal pH conditions • Controlled release up to 12 h	[268]
Protein delivery system				
16	Bovine serum albumin (BSA) as a model drug	Alginate–GG hydrogels cross-linked with glutaraldehyde	• Minimal protein release at pH 1.2 (~20 %), and significantly higher (~90 %) at pH 7.4 • Controlled release at higher pH of the intestine	[83]

Transdermal drug delivery system				
17	Terbutaline sulphate	Carboxymethyl-guar gum solution	• Good film forming property and zero order diffusion rate across human cadaver epidermis	[183]
18	l-DOPA and l-Tyrosine	Acryloyl guar gum hydrogels	• Sustained release of both l-tyrosine and l-DOPA even after 12 h • Efficient drug delivery carriers for transdermal applications	[263]
Therapeutic appllications				
19	Partially hydrolyzed guar gum (PHGG)		• Water-soluble dietary fiber • Treatment of Irritable bowel syndrome (IBS) • Useful in lowering serum cholesterol and glucose levels • Effective in reducing the cardiovascular disease risk, diabetes and weight loss programs • Management of diarrhea and constipation	[37, 86, 249]

2.6.1.1 Guar gum in colonic drug delivery

The goal in drug delivery research is to develop formulations to meet therapeutic needs relating to particular pathological conditions. Colon, as a site, offers several advantages as near neutral pH, reduced digestive enzymatic activity, much longer transit time and greater enhancers for absorption [221]. The delivery of drugs to the colon is highly desirable for local action in a variety of conditions like inflammatory bowel diseases (IBD), infectious diseases and colon cancer. It is also useful for protein and peptide drugs, because of less hostile environment as compared to stomach and small intestine [302]. The various approaches for targeting orally administered drugs to the colon have been shown in Table 2.6.1.2. The microflora-activated systems have been found highly reliable because of colon specific degradation of non-starch polysaccharides.

Table 2.6.1.2 Approaches for colon specific drug delivery (On the basis of Yang *et al*. [302])

Approach	Design strategy	Drug release triggering-mechanisms	Comments
Prodrugs	Chemically modified derivatives of drug	Release of active parent drug by enzymatic processes in colon	Highly site specific Considered a new chemical entity from regulatory perspective
pH-dependent systems	Using combination of polymers with pH-dependent solubility	pH-dependent solubility of polymers along the GI tract	Unpredictable site-specificity of drug release
Time-dependent systems	Incorporation of a time factor/polymer in formulation	Stimulation of delivery system transit in upper GI tract	Complicated prediction of accurate location of drug release due to variations of gastric retention times
Microflora-activated systems	Incorporation of non-starch polysaccharides into delivery system	Primarily degradation of non-starch polysaccharides by anaerobic bacteria in colon	Highly promising because of colon specific degradation of non-starch polysaccharides

The microflora of colon is in the range of 10^{11} -10^{12} CFU/mL, consisting mainly of anaerobic bacteria, e.g. *Bacteroides, Bifidobacteria, Eubacteria, Clostridia, Enterococci, Enterobacteria, Ruminococcus* etc. [139]. Colonic bacterial enzymes are capable of degrading a variety of polysaccharides present in the diet that are not affected either in the stomach or in the small intestine. Polysaccharides are resistant to the digestive action of gastrointestinal enzymes, hence retain their integrity in the upper gastro intestinal tract (GIT) [48]. These non-toxic and biodegradable polysaccharides have the potential of delivering drugs specifically to the colon. Rubinstein *et al.* have demonstrated the usefulness of pectin, calcium pectinate and chondroitin sulphate as potential colon-specific drug delivery carriers [229]. Studies were also carried out on pectin formulations by Ashford *et al.* [17] using pectinolytic enzymes. A suspension of natural polygalactomannans in polymethacrylate solution was used to form a degradable coating which was found to delay the drug release in the small intestine by forming a swellable layer around the drug core. The coating was found to be degraded by colonic bacterial enzymes thereby releasing the drug in the colon [150]. These findings formed the basis to investigate the usefulness of guar gum, which also contains polygalactomannans, as a colon-specific drug delivery carrier.

Guar gum is hydrophilic in nature and swells in cold water forming viscous colloidal dispersions or sols. It is non-ionic and hence the viscosity of dispersion is unaffected by pH and is same in acidic and alkaline medium. This gelling property retards the release of the drug from the dosage form as well as it is susceptible to degradation in the colonic environment [122, 211, 213]. Guar gum shields the drug from the environments of stomach and small intestine and is

able to deliver the drug to the colon. On reaching the colon, it undergoes degradation by colonic bacteria leading to a subsequent reduction in their molecular weight and thereby loss of mechanical strength [139]. They are then unable to hold the drug entity any longer and release the drug molecule in the colon. The anaerobic bacteria that are responsible for the degradation of guar gum in the colon are *Bacteroides species (B. fragilis, B. ovatus, B. variabilis, B. uniformis, B. distasonis and B. thetaiotaomicron*) [26, 139]. To investigate the degradation of polysaccharide by intestinal microflora, homogenized and diluted feces from human source were incubated with the guar gum. It produced a rapid decrease in viscosity and fall in pH while no such results were observed when it was incubated with autoclaved fecal homogenates [54]. It was concluded that guar gum in the form of either a matrix or coat over the drug core might have been degraded to a larger extent by the action of anaerobic microbial population of large intestine. Guar gum and its derivatives are extensively used in colonic drug delivery due to their drug release retarding property and susceptibility to microbial degradation in the large intestine [211]. Colonic drug delivery applications of guar gum have been given in Table 2.6.1.1.

The single unit colon targeted drug delivery system may suffer from the disadvantage of unintentional disintegration of the formulation due to manufacturing deficiency or unusual gastric physiology that may lead to drastically compromised systemic drug bioavailability or loss of local therapeutic action in the colon. For this reason and considering the selective uptake of micron or submicron particles by cancerous and inflamed cells/tissues a multiparticulate approach based on pellets, micro/nanoparticle type formulation is expected to have better pharmacological effect in the colon [15]. At present, more emphasis is being laid on the development of multiparticulate dosage forms in comparison to single unit systems because of their potential benefits like increased bioavailability, reduced risk of systemic toxicity, reduced risk of local irritation and predictable gastric emptying [15, 132]. There are several reports on guar gum as a colon specific drug carrier, however a little information is available in literature on the guar gum based multiparticulate drug delivery system.

2.7 Molecular markers

The research work on molecular markers and their application in plant science has been reviewed by several workers [2, 28, 70, 74, 84, 117, 175]. The researchers have utilized a diverse set of molecules in an effort to study the genetic inheritance and relationships among different plant species [59]. Gregor Mendel used phenotype-based genetic markers for understanding the

inheritance of traits in 19th century [2]. Study of phenotype-based markers in *Drosophila* led to the establishment of genetic linkage theory [256]. In 1960s and 70s, secondary metabolites, such as flavonoids and terpenoids, were mainly used for resolving the plant relationships at many taxonomic levels [59, 250]. The phenotypic and biochemical markers showed certain limitations which led to the development of direct DNA-based markers, also known as molecular markers [70, 175].

The DNA-based molecular markers unfold natural variation at the genome sequence level. These markers have been extensively utilized for plant diversity studies, genotyping, quantitative trait mapping, genetic-linkage studies and marker-assisted selection in plant breeding [116]. The DNA-based molecular markers are stable and can be detected in all tissues regardless of differentiation, development, growth or defense status of the plant. These markers offer various advantages over phenotypic-markers because they are not confounded by the environment, pleiotropic and epistatic effects [2]. The DNA-based molecular markers can be detected in genomic DNA sequences and are located at specific regions in the genome. These markers are transmitted by the standard laws of inheritance from one generation to the next. Since the markers and the genes they mark are close together on the same chromosome, they tend to stay together in each generation of plants produced. Hence these markers can be used to create a genetic linkage map [239]. The first theory about the construction of genetic linkage map in man using restriction fragment length polymorphism (RFLP) was given in 1980 [34]. Today several molecular markers are available with different principles and methodologies which require careful consideration in choosing one or more of such marker types. Fig 2.7.1 shows the development of various molecular markers over last three decades [2]. Molecular markers are considered ideal if, they have the following characteristics: (i) simple, quick and inexpensive; (ii) polymorphic and evenly distributed throughout the genome; (iii) generate multiple, independent and reliable markers; (iv) need small amounts of tissue and DNA samples (v) provide adequate resolution of genetic differences; (vi) have linkage to distinct phenotypes and (vii) require no prior information about the genome of the organism [2]. The commonly used molecular markers in various studies are isozyme markers, amplified fragment length polymorphism markers, random amplified polymorphic DNA markers, restriction fragment length polymorphism markers, cleaved amplified polymorphic sequences, simple sequence repeats and single nucleotide polymorphisms.

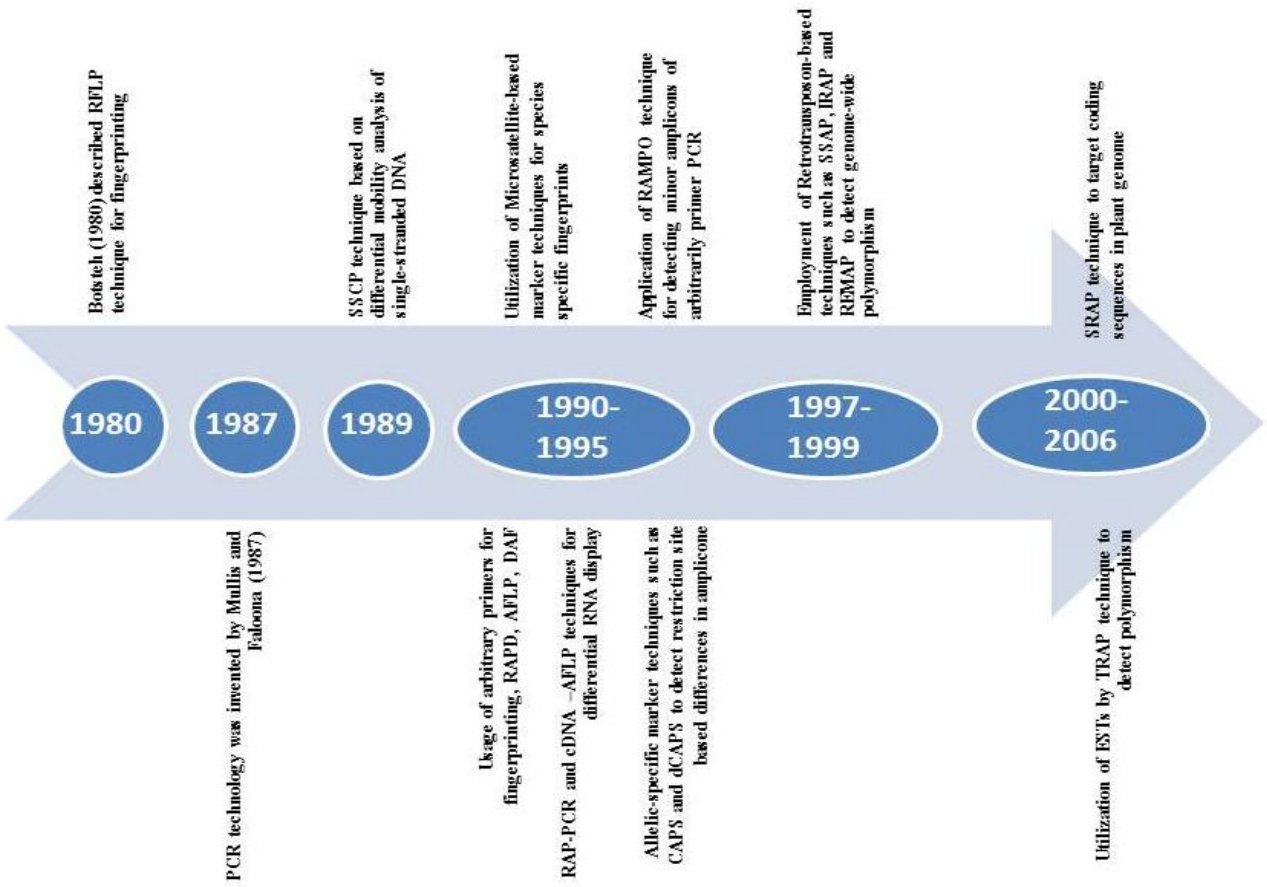


Figure 2.7.1 Schematic diagram describing the development of molecular markers over last three decades. (Modified from Agarwal *et al.* [2])

2.7.1 Isozyme markers

Isozymes are proteins with same enzymatic function but different structural, chemical, or immunological characteristics. These are the multiple forms of the same enzyme coded by different genes. Isozyme markers were the first used in the 1960s [149]. With the help of these markers the term biochemical markers, often referred as allozyme or isozyme markers, was introduced as a general tool for mapping QTL for the first time in early 1980s [291]. Isozyme markers have been widely used in different genetic studies in several plants, including studies of population genetics aiming at characterizing the diversity and genetic structure [7, 23, 80, 81, 91, 177, 280]. These markers are relatively simple and cheap, and present a codominant nature [121]. However, the use of these markers is limited due to the low number of loci and alleles per locus detected, post-translational modifications, tissue-specific forms, modifications in response to environmental conditions and the developmental stage of the individual [182].

2.7.2 Restriction fragment length polymorphism (RFLP) markers

In RFLP, DNA polymorphism is detected by hybridizing a chemically labeled DNA probe to a Southern blot of DNA digested by restriction endonucleases, resulting in differential DNA fragment profile [2]. The RFLP markers are highly polymorphic, co-dominantly inherited and highly reproducible. RFLPs are present throughout the genome, heritable and locus specific. The method can also be used to simultaneously screen numerous samples. The DNA blots can be analyzed repeatedly by stripping and re-probing (usually eight to ten times) with different RFLP probes [2]. RFLP was used widely in early 1980's for a wide range of plant species [210]. This technique is not very widely used now because it is time consuming and requires large amount of high-quality DNA, and involves expensive and toxic reagents [218]. In this technique prior sequence information is also required for probe generation which increases the complexity of the methodology. These limitations led to the development of a new set of molecular markers [2].

Polymerase chain reaction (PCR) technology was invented in 1980s [180, 181, 230, 231] that led to the development of various molecular markers for genetic diversity studies. Following are the various PCR-based markers mainly used in plants.

2.7.3 Amplified fragment length polymorphism (AFLP) markers

AFLP markers are generated by complete restriction endonuclease digestion of total genomic DNA, followed by selective PCR amplification and electrophoresis of a subset of the fragments. This results in a unique, reproducible fingerprint/profile for each individual. The fingerprint allows an assessment of genome-wide variation. These anonymous markers consist largely of non-coding DNA [168].

AFLP markers are useful in a wide range of applications including population genetics [159], genetic diversity studies [6, 75], linkage mapping [232, 272, 301], identifying hybrids [244] and cultivars [85], and developing single-locus sequence-characterized amplified region (SCAR) markers [118, 192].

2.7.4 Random amplified polymorphic DNA (RAPD) markers

Randomly amplified polymorphic DNA markers were introduced in 1980s [295]. They are powerful and effective tool for determining the genetic variation and also capable in generation of genome spanning markers without prior knowledge of the sequence [30, 109].

The RAPD markers are based upon differential PCR amplification of the genomic DNA. They deduce DNA polymorphisms produced by rearrangements or deletions at or between oligonucleotide primer binding sites in the genome [2]. These markers have been utilized for genetic diversity studies in many plant species [89]. They have been used for genome fingerprinting and characterization [220, 233, 292] and construction of genetic maps [128, 140, 194]. Markers linked to important plant genes like *Uromyces appendiculatus* (Pers.) resistance in common bean [98], *Pseudomonas* resistance gene in tomato [161], etc. have been developed using RAPD markers. This techniques can be useful in development of characterized amplified regions (SCARs) markers [68, 167].

DNA amplification fingerprinting (DAF) was introduced in 1980s [113]. It is a modification of RAPD marker technique in which one or more very short (less than or equal to 5 nt) arbitrary oligonucleotides are used for enzymatic amplification of DNA [39, 228, 295]. Polyacrylamide gel electrophoresis and silver staining is used to resolve the complex patterns. It is suitable for DNA fingerprinting [38]. DAF has been used for relationships studies among closely related plant species and cultivars [45] and bacteria [24].

2.7.5 Cleaved amplified polymorphic sequences (CAPS) and derived CAPS markers

CAPS are also known as PCR-RFLP markers. CAPS decipher the RFLPs caused by single base differences like SNPs, insertions/deletions, which modify the recognition sites for restriction endonuclease in PCR amplicons [2]. These are highly reproducible markers with different DNA extraction methods. The simple analysis and its stability make CAPS an important tool for the identification of plant cultivars [144]. These markers are developed by the comparison of sequence differences between two known regions and designing a combination of restriction enzymes and primers for assay [144]. They can also be developed by arbitrary marker techniques [284].

CAPS markers are the most commonly used markers for SNP detection assays using the gel based methods [264]. However, the existence of a restriction site difference spanning the SNPs between varieties/lines to be analyzed is essential for converting SNPs to CAPS markers. Michaels and Amasino [169] and Neff *et al.* [191] demonstrated that single-base changes generating restriction site difference could be employed for the development of PCR-based markers by the derived CAPS (dCAPS) method. In this method, a restriction enzyme recognition

site which includes the SNP is introduced into the PCR product by a primer containing one or more mismatches to template DNA [241]. The PCR product modified in this manner is then subjected to restriction enzyme digestion, and the presence or absence of the SNP is determined by the resulting restriction pattern. Like CAPS, the dCAPS markers are simple and relatively inexpensive to identify [191].

2.7.6 Inter simple sequence repeat (ISSR) markers

Microsatellites are usually more or less proportionally dispersed in the genome. However, regions with a greater abundance of these sequences have been found and are named as "SSR hot spots" [31, 32, 315]. Such regions can serve as a source of ISSR markers.

The ISSR technology is based on the amplification of regions (100-3000 bp) between inversely oriented closely spaced microsatellites. Single primer (16-18 bp) consisting of several simple sequence repeats is used for amplification of these regions. Primers can be based on any SSR motif along with 5' or 3' anchored bases usually 2-4 bases which are arbitrary selective nucleotides. However, nonanchored primers have also been used [32].

2.7.7 Single nucleotide polymorphisms (SNPs)

Single nucleotide polymorphisms (SNPs) and insertions/deletion (InDel) variations, are considered as the basis of most differences between alleles and have been simplified by many developments in sequencing technologies. They are the most abundant molecular markers in the genome and widely distributed throughout genomes although their occurrence and distribution varies among species [2]. SNP discovery in many crop species, such as corn and soybean, is relatively straight forward because of the presence of high level of intraspecific nucleotide diversity, and the availability of many gene and expressed sequence tag (EST) sequences [217]. SNP identification using next-generation sequencing technologies has been applied in many plant species [22, 25, 193]. This approach has been used for the full re-sequencing of the first *Arabidopsis* accessions [195]. There are several other strategies for the discovery of SNPs. These include the re-sequencing of PCR amplicons with or without pre-screening, electronic SNP (eSNP) discovery in shotgun genomic libraries, and eSNP discovery in expressed sequence tag (EST) libraries [217]. The SNPs can be used in various studies regarding linkage mapping, genetic variation, population structure analysis, map-based gene isolation, and plant breeding

[79]. SNPs are the popular molecular markers but their use in crops is still limited because of the development process cost and time consumption factors.

2.7.8 Simple sequence repeat (SSR) markers or microsatellites

Simple sequences repeats or short tandem repeats are also called microsatellites. These are repetitions of very short nucleotide motifs (1-6), which are found as interspersed repetitive elements in all eukaryotic genomes [262]. The variation in the number of repeated units is mainly due to strand slippage during DNA replication where the repeats allow matching via excision or addition of repeats [235]. As slippage in replication is more than a point mutation, microsatellite loci tend to be hypervariable. Microsatellites show extensive length polymorphisms among the individuals during PCR analysis of unique loci using specific primer sets. SSRs are considered as the most efficient molecular markers but their use is still limited because of the development process cost and time consumption factors [219].

SSRs are much popular genetic-markers because of their high abundance, co-dominant inheritance, enormous extent of allelic diversity and ease of assessing the size variation by PCR technique with SSR flanking primer pairs [2, 4]. Their potential for automation is an additional advantage when compared with other types of molecular markers. SSRs are generally highly polymorphic, genome specific, abundant and co-dominant, and have recently become important genetic markers [90, 107]. These have been used in studying genetic diversity [4, 10, 29, 111, 171, 187], identifying hybrids [44], linkage mapping [96, 115], parentage analysis of clones [125] and marker assisted selection [222].

2.8 Studies on molecular markers in guar

Five molecular markers, namely, RAPD, rDNA, ISSR, SSR and SCAR, have been used in the study of genetic diversity in guar [143, 145, 146, 203, 204, 216, 243]. Punia *et al.* used RAPDs as genetic markers to study genetic diversity in 34 commercially important guar cultivars [214]. In this study a significant polymorphism was found among the cultivars and all the genotypes were grouped in two major and three minor sub-clusters. Pathak *et al.* made an evaluation of genetic diversity using RAPD markers among 32 genotypes of guar collected from different geographical regions of India. The results revealed a considerable genetic diversity among the tested genotypes [204]. In another study, genetic diversity in 18 genotypes of guar was assessed using nuclear rDNA and RAPD markers [203]. The results showed that most of the total

genetic variation exists within population rather than among populations. The genetic diversity in 29 landraces and 19 commercial varieties of guar collected from different regions of India was studied using RAPD and ISSR markers [145]. The results revealed the presence of more variation within the populations than between the populations. Further separate clustering of landraces from commercial varieties was obtained. In another study, 35 genotypes of guar collected from different states of India analyzed for genetic diversity analysis using RAPD, ISSR and SCAR markers. This study developed the first set of sequence-based SCAR markers in guar which were found to be more useful than RAPD and ISSR markers. The RAPD markers were also used by Kumar *et al.* to study the extent of genetic variation in 23 released and elite genotypes of guar from different parts of India [142]. The results categorized the genotypes in various groups while any significant relationship between genetic diversity and geographical location was not observed. Kuravadi *et al.* used 224 EST-SSR markers to study of genetic diversity in three commercial varieties of guar and two wild species, namely, *C. senegalensis and C. serrate* [146]. The results showed very low genetic variation in the guar varieties. In another study using EST-SSR markers, 32 genotypes of guar with different regions of origin were tested for genetic variation analysis [143]. The study showed the existence of very low genetic diversity in guar. The results indicated that more SSR and SNP markers should be generated by high-throughput sequencing technology.

2.9 Transcriptome analysis

Several workers in last few years have reviewed the transcriptome analysis technique [102, 166, 184, 296]. A set of all RNA molecules which includes rRNA, mRNA, tRNA and other non-coding RNA transcripts in a cell or tissue is referred as transcriptome. The term 'transcriptome' was first proposed by Charles Auffray in 1996 [208]. It was used in a scientific paper for the first time in 1997 [281]. The study of the transcriptome of an organism/ biological sample is known as "Transcriptomics". Transcriptome has a long pedigree and meets all the requirements of a true 'omics' technology [166]. The study of transcriptomics is known as expression profiling or transcriptome analysis which helps to examine the expression of RNA molecules (usually mRNA) in a given biological sample [296]. The advancements and progress in "omics" technologies and availability of high throughput techniques has accelerated the transcriptome analysis of an organism. These high throughput techniques have provided new tools for universal detection of the entire mRNA with utmost accuracy in a biological sample

[184]. These techniques have further helped in targeted as well as non-targeted analysis of a transcriptome of an organism in a non-biased manner [296]. The technological advances have been further elaborated to analyse the genome, proteome and metabolome. In addition with other "omics" technologies, transcriptomics plays impeccable role to analyze the whole biological system also referred as systems biology via interactomics and studies related to network biology [27, 246].

2.9.1 Next generation sequencing (NGS) in transcriptome analysis

Transcriptome analysis can be performed through either microarray or next generation sequencing technologies [273]. Since mid-1990, microarray has been technique of choice for genome wide expression analysis of many genes in an organism i.e. high throughput transcriptome analysis. In the last couple of years intense development of transcriptomic applications have been made that has resulted in supplanting of microarrays by next generation technology (NGS) as the technology of choice for gene expression analysis [166].

The first NGS-based transcriptome analysis (RNA-Seq) utilizing 454/Roche technology was reported in the year 2006 [19]. The era of RNA-seq dominance started in 2008 with the publishing of three papers on a new short-read technology developed by Solexa (now Illumina) [178, 257, 294]. The Illumina/Solexa technology generated data has been raised rapidly from 1 GB per run in 2006 to 600 GB per run for the HiSeq 2500 in 2012 [166]. While the Roche/454 technology has always generated reads long enough for RNA-Seq it has been hampered by the relatively low throughput and high cost of the libraries compared to the more popular Illumina technology.

NGS and new complementary computational tools have made high-throughput sequencing and assembly a more commonplace [184]. NGS can produce billions of short reads in parallel, generally 50 – 800 base pairs (bp) depending on technology. Sequence data that may have once taken years to be generated may now be produced in a matter of days or even hours. NGS is replacing capillary sequencing in many applications due to its lower cost per base pair of DNA and its lack of a subcloning requirement [273]. In particular, the study of transcriptomics has been greatly advanced by this technology [184]. At present, RNA-Seq, is one of the most popular topic areas in NGS, as shown by the 'SEQanswers' search tag cloud (http://seqanswers.com/forums/search.php). Other methods of studying gene expression such as

microarrays and serial analysis of gene expression (SAGE) are being replaced in many applications with RNA-Seq [166]. RNA-Seq can show the repertoire of expressed sequences found in a particular tissue at a specific point of time, even rare transcripts, due to the great depth of sequencing. In this way, it can produce a nearly complete picture of transcriptomic events in a biological sample [273].

2.9.2 RNA-Seq experiment design

A typical RNA-Seq experiment starts with mRNA that is subsequently converted into cDNA to form an RNA-Seq library. By sequencing the millions of DNA fragments in the library (known as 'reads') using next-generation sequencing, an accurate measure of the relative abundance of each transcript and splice variants can be obtained [184]. In recent years, a wide array of bioinformatics tools have been developed to process the individual steps required to translate next-generation sequencing output into information on gene expression levels [273]. The workflow of RNA-Seq experiment is shown in Figure 2.9.1.

2.9.3 Applications of transcriptome analysis in plant sciences

Transcriptional expression analysis in plants has been found useful to identify the genes involved in certain functions [172]. Advances in biological techniques lead to high throughput analysis of whole transcriptome of an organism through microarray and RNA-Seq. RNA-Seq technique has been extensively used for *de novo* assembly of the organisms whose genome information is not known [166, 273]. *De novo* assembly generated by transcriptome analysis could serve as a reference for further genomic/transcriptomic studies. There are several other applications of RNA-Seq such as functional annotation, differential gene expression and development of genic-markers in various organism [184, 273]. Transcriptome analysis has been widely used in many different plant species and showed numerous applications. *Arabidopsis thaliana* is the first plant in which RNA-Seq was applied to identify the expressed sequence tags, protein coding sequences and expression analysis [290]. The applications of transcriptome analysis in plants have been summarized in Table 2.9.1.

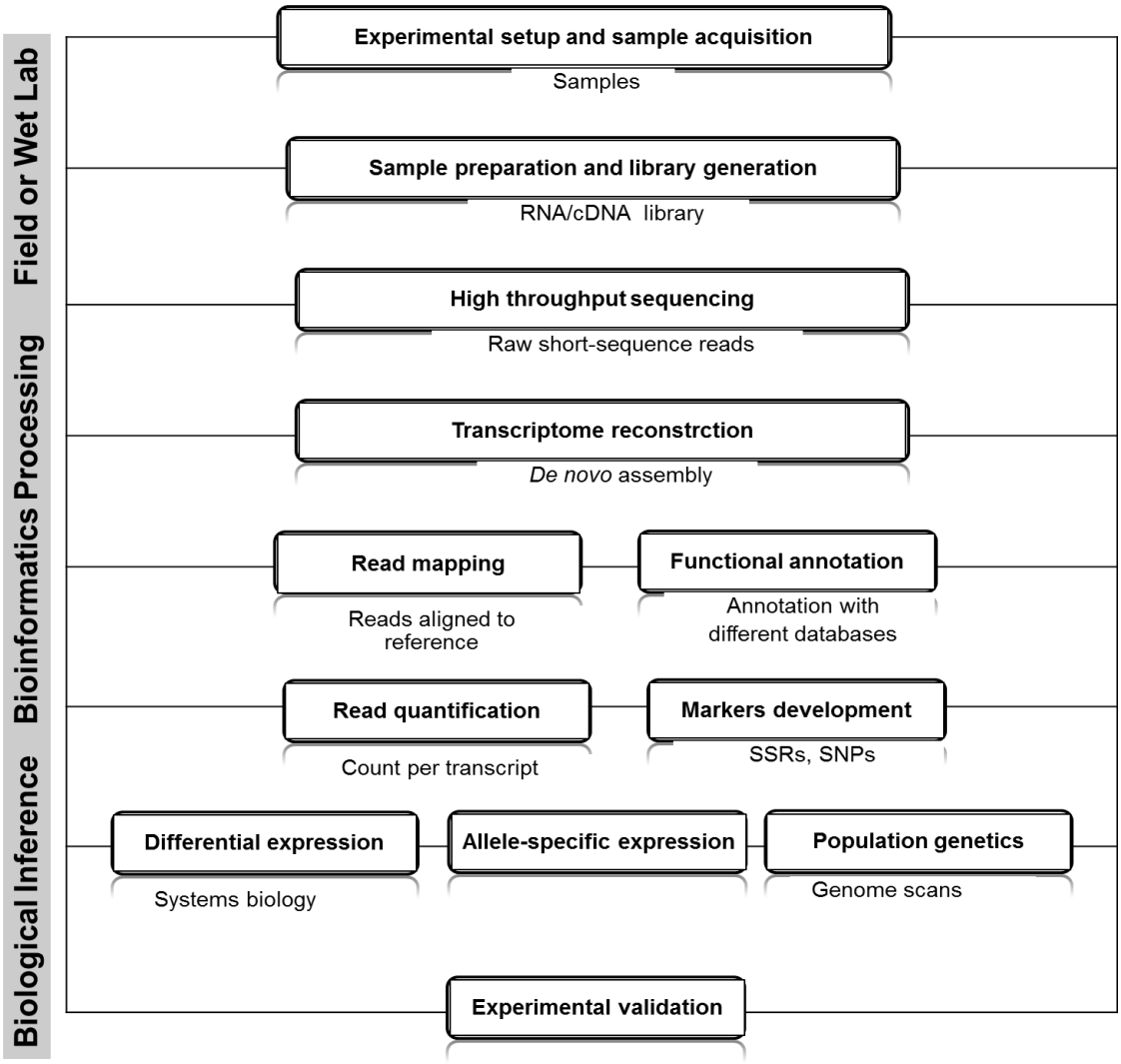


Figure 2.9.1 A typical workflow of RNA-Seq. (Modified from Wolf [296])

Table 2.9.1 Applications of transcriptome analysis in plant science

S. No.	Name of organism	Description	Reference
Functional genomics, genome annotation and alternative splicing			
1	*Arabidopsis thaliana*	Genome-wide mapping of alternative splicing	[76, 314]
2	*Olea europaea*	Functional annotation and identification of DEGs with potential relevance in regulating the fruit metabolism and phenolic content during ripening in olive	[9]
3	*Pisum sativum*	First comprehensive reference set for the model legume pea using flowers, leaves, cotyledons, epi- and hypocotyl, and etiolated and light treated seedlings tissues	[78]
4	*Glycine max*	Analysis of seed proteome in maintenance of the protein homeostasis by selective accumulation of other proteins to	[236]

		preserve the overall protein content in storage protein knockdown (SP-) seeds	
Genomic and proteomic resources of non-model species including marker development			
5	*Pteridium aquilinum*	*De novo* assembly, functional annotation, comparative evolutionary and functional genomics analysis, and development of 548 SSRs and 689 expressed transposable elements in fern	[64]
6	*Cicer arietinum*	Development of chickpea transcriptome database including *de novo* assembly, functional annotation, differential gene expression during progression of seed development and development of SSRs	[82] [212]
7	*Hippophae rhamnoides*	Identification of 6790 SSRs from *de novo* transcriptome assembly of seabuckthorn	[110]
8	*Carthamus tinctorius*	*De novo* assembly, functional annotation and identification of genes involving in control flower and seed quality in safflower	[158]
9	*Secale cereale*	Development of SNPs in rye	[99]
10	*Platycodon grandiflorum*	Development of genic-simple sequence repeats (SSRs), single nucleotide polymorphism (SNP) and cleaved amplified polymorphic (CAPS) in bellflower	[126]
11	*Sesamum indicum*	Development of 2164 genic SSR markers RNA-Seq data of 24 cDNA libraries of Sesame	[311]
12	*Prunus avium*	Generation of gene-linked SNP and haplotype markers from RNA-Seq data of sweet cherry	[129]
13	*Quercus pubescens*	*De novo* assembly, functional annotation and identification of 14202 microsatellite markers and 18425 in European oak	[267]
Transcriptional dynamics, evolutionary transcriptomics and host-pathogen interactions			
14	*Vitis vinifera*	Differential gene expression, splicing and related genes identification for three developmental stages of grapes i.e. post setting, veraison and fruit ripening	[310]
15	*Raphanus sativus*	Differential gene expression of early and late developmental stages of radish and development of EST-SSR markers	[286]
16	*Arabidopsis thaliana*	Identification of more than 1000 gene related to meiosis-specific male meiocyte development	[49]
17	*Solanum lycopersicum*	Tissue- and cell-type specific transcriptome profiling for metabolic and regulatory specialization and cuticle formation in five different tissues of tomato	[164]
18	*Zea mays*	Identification of genes related to early embryogenesis, function, maintenance of cell and organogenesis in shoot apical meristems of maize	[260]
19	*Sorghum bicolor*	Identification of various genes related to genes and networks involved in osmotic stress and hormonal treatment for the crops subjected to drought tolerance	[71]
20	*Sorghum bicolor*-	Identification of differential gene expression of sorghum under infection of target leaf spot	[174]

	Bipolaris sorghicola		
21	*Glycine max-Xanthomonas axonopodis*	Identification of genes related to damage-associated molecular patterns (DAMP) and pathogen-associated molecular patterns (PAMP) and their involvement in plant basal defense mechanism in near isogenic lines of soybean due to infection of bacterial leaf pustule disease	[127]
22	*Castanea spp.*	Functional annotation and identification of defense-related genes and the genes involved in resistance to *Cryphonectria parasitica* between American chestnut and Chinese chestnut	[20]
23	*Gylcine max*	Evolutionary dynamics study of homoeologous genomes of allopolyploid soybean and its related diploid species	[108]
24	*Brachypodium distachyon, Sorghum bicolor and Oryza sativa*	Identification of conserved expression patterns among several orthologous genes in reproductive tissues of three Poaceae species including *Brachypodium*, *Sorghum* and rice	[62]
25	*Eleusine indica*	Optimization of transcriptome references for studying herbicide resistance and evolutionary process in the three allotetraploid *E. indica* offspring	[50]
Identification and characterization the novel non-coding RNA molecules			
26	*Zea mays*	Identification and characterization of several protein coding and non-coding RNA molecules in developing maize endosperm	[312]
27	*Arabidopsis thaliana*	Identification of several non-coding RNAs (ncRNAs) including non-protein coding RNAs (npcRNAs), ORF encoding peptides and miRNAs as well as methylome in *Arabidopsis*	[36, 101, 103]
28	*Solanum lycopersicum*	Identification and characterization of conserved and non-conserved miRNA in leaf and fruit of tomato	[179, 313]
29	*Vitis vinifera*	Identification and characterization of 24 conserved and 26 known and non-conserved miRNA families involved in growth, development, abiotic stress tolerance and host-pathogen interactions in grapes	[196]
30	*Sorghum bicolor*	Identification of novel miRNA having significant role in stem sugar accumulation and flower development	[40]
31	*Oryza sativa*	Identification of transposable elements, methylation site and differential gene expression in rice Functional annotation, alternative splicing, deferential gene expression and SNPs detection	[53, 157]
Gene ontology, network and integrative data analysis			
33	*Arabidopsis thaliana*	Analysis of topological characteristics, biological networks and organization of functional gene modules to generate gene to gene associations and co-expression network	[160]
33	*Papaver*	Identification of cytochrome P450 involvement in formation of	[65]

	somniferum	noscapine and papaverine by studying the bezylisoquinoline alkaloids (BIAs) related pathways with metabolite profiling using mass spectrophotometry	
34	*Zea mays*	Analysis of involvement of sucrose metabolism, abscisic acid (ABA) related genes and phospholipase C-mediated signaling pathways in drought response in maize	[119]
35	*Elaeis guineensis*	Identification of WRI1 transcription factors and several other genes and enzymes involved in oil synthesis in seed mesocarp of oil palm	[35]
36	*Sorghum bicolor*	Network analysis for functional annotation of transcriptome in response to abscisic acid metabolism and osmotic stress by utilizing expression and co-regulation of identified genes	[71]
37	*Brassica rapa*	Detection of alternative splicing events, single nucleotide polymorphisms and refinement of annotated gene for regulating the leafy head formation in Chinese cabbage	[285]
38	*Eucalyptus grandis*	*De novo* assembly and digital expression profiling of diverse xylogenic and non-xylogenic tissues in eucalyptus	[173]
39	*Prunus persica*	Transcriptome functional annotation, digital expression analysis and identification of informative candidate genes associated with variegation in peach flowers	[51]
40	*Allium sativum*	*De novo* assembly, functional annotation and identification of genes involved in organic sulfur biosynthesis in garlic	[258]

Since RNA-Seq showed impeccable and significant role in studies related to plant biology, recent advances in RNA-Seq have opened new avenues and opportunities in plant transcriptomics [162].

The above literature review showed that guar is an economically important legume crop. Guar gum is essential for various industrial applications including pharmaceutical industries. It has been found useful as a colon specific drug carrier, however a little information is available in the literature for the guar gum based multiparticulate drug delivery systems. All the genetic improvement programs in guar have been carried out using conventional breeding without the involvement of molecular markers because of limited availability of genetic resources in this crop. As a result, only a limited success has been achieved in obtaining improved guar varieties. NGS technologies provide novel opportunities not only in functional genomics and gene discovery but also in developing huge genetic resources in non-model plants. Therefore, this study was done for development of large genetic resource and identification of molecular markers using NGS technology for quick and easy execution of crop improvement programs in guar.

3. MATERIALS AND METHODS

3.1 Materials

3.1.1 Plant material

The two commercially grown elite varieties of guar (*C. tetragonoloba),* namely, M-83 and RGC-1066 were chosen for the leaf transcriptome analysis. The seeds of these varieties were kindly provided by Rajasthan Agricultural Research Institute (Sri Karan Narendra Agriculture University, Jobner) Durgapura, Jaipur (India). The M-83 and RGC-1066 are vegetable and gum producing varieties, respectively. The plants of M-83 have glabrous leaf surface and white flower color whereas RGC-1066 plants have hairy leaf surface and purple flower color. Both the varieties have 0-3 branches.

3.1.2 Chemicals/biochemicals and buffers used

All the chemicals/biochemicals used were procured from standard companies and were of analytical grade.

3.1.2.1 TE/$T_{10}E_1$ buffer

TE buffer was composed of 1 ml of 1 M Tris-HCl (pH 8) and 200 µl of 0.5 M EDTA dissolved in 100 ml MilliQ water.

3.1.2.2 TBE buffer (1X)

The composition of the TBE buffer was as follows:

Name of the reagent	**Composition**
Tris	10.9 g
Boric acid	5.56 g
EDTA	0.98 g
Distilled water	1 L

3.1.2.3 DNA extraction buffer

The composition of DNA extraction buffer was as follows:

Name of the reagent	**Concentration**
Tris-HCl (pH 8.0)	100 mM
EDTA (pH 8.0)	20 mM
NaCl	1.4 M
CTAB	5 %
β-mercaptoethanol	0.01 %

3.1.3 Polymerase chain reaction (PCR) mixture

The composition of PCR reaction mixture was as follows:

Contents	**Amount per 20 µl**
10X *Taq* buffer	2.0 µl
dNTP mix (10 mM) (Biotool, USA)	2.0 µl
$MgCl_2$ (50 mM)	1.0 µl
Taq polymerase (1U/µl) (Biotool, USA)	2.0 µl
Forward primer (20 pM)	0.5 µl
Reverse primer (20 pM)	0.5 µl
DNA (100 ng/µl)	0.5 µl
MilliQ water	11.5 µl

3.1.4. DNA gel loading dye

The gel loading dye was composed of bromophenol blue (0.25 g) and xylene cyanol (0.25 g) dissolved in 50 % glycerol (6 ml), and MilliQ water (4 ml).

3.1.5 Acrylamide-bis acrylamide solution (40 %)

Acrylamide-bis acrylamide solution (40 %) was composed of acrylamide (38 g) and bis acrylamide (2 g) dissolved in double distilled water (final volume up to 100 ml).

3.1.6 PAGE gel

The composition of PAGE gel was as follows:

Name of the reagent	**Composition for 150 ml**
40% acrylamide-bis acrylamide solution	30 ml
10X TBE	7.5 ml
Ammonium persulphate	0.105 g
TEMED	125 µL
Double distilled water	92.38 ml

3.1.7 Silver staining solutions

The compositions of silver staining solutions were as follows:

A. Fixative solution (Solution 1):

Name of the reagent	**Composition**
Methanol	20 ml
Glacial acetic acid	1 ml
Distilled water	179 ml

B. Staining solution (Solution 2):

Name of the reagent	**Composition**
Methanol	20 ml
Glacial acetic acid	1 ml
$AgNO_3$	0.2 g
Distilled water	179 ml

C. Developing solution (Solution 3):

Name of the reagent	**Composition**
NaOH	5.1 g
Formaldehyde	600 µL
Distilled water	199.4 ml

3.2 Methods

3.2.1 Collection of guar leaves

The seeds of guar plants were grown in field conditions at Indian Institute of Technology Roorkee, India. The plants were grown in rows; the space between two rows was 0.5 m. The plant to plant distance was maintained at 0.2 m within a row. The healthy leaves were collected from 3-week-old plants for RNA extraction. The guar variety M-83 was named as Sample-1 and the variety RGC-1066 as Sample-2 for further studies.

3.2.2 Sequencing of guar leaf transcriptomes

The sequencing of the leaf transcriptomes of guar varieties M-83 and RGC-1066 was outsourced to SciGenom Labs Pvt. Ltd., Cochin (India) and done by the following procedure.

The total RNA of the plant leaves was extracted by using SIGMA Spectrum^{TM} Plant Total RNA Kit (Sigma-Aldrich, USA). The leaves were ground to a fine powder in liquid nitrogen and lysed in a lysis solution that released RNA and at the same time inactivated ribonucleases and interfering secondary metabolites such as polyphenolic compounds. After the removal of cellular debris, RNA was captured onto a binding column using a unique binding solution which effectively prevented polysaccharides as well as genomic DNA from clogging the column. Residual impurities and genomic DNA were removed by washing the column and purified RNA was eluted in RNase-free water. The quality assessment of the purified RNA was carried out on the Agilent 2100 Bioanalyzer system (Agilent Technologies, USA). The purified RNA was stored at -20 ^{o}C until use.

Each library was constructed using 1µg of total leaf RNA. Library preparation was carried out by using Illumina® TruSeq® Standard Total RNA Sample Preparation kits (Illumina, Inc., USA). The first step involved the removal of ribosomal RNA (rRNA) using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads. The RNA was fragmented into small pieces using divalent cations under elevated temperature and further purified. The cleaved RNA fragments were used for first strand cDNA synthesis using reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA polymerase I and RNase H. These cDNA fragments were adenylated with a single 'A' base at 3’ end and ligated with the

adapters. The products were purified and enriched with PCR to create the final cDNA library. The library was stored at -20 ^{o}C until use.

The sequencing of libraries was carried out on Illumina HiSeq 2500 machine to get paired end sequence reads of 100 bp length. The data from the instrument was converted to FASTQ format using Illumina pipeline.

3.2.3 Transcriptome analysis of guar leaves

The raw reads of leaf transcriptome of guar varieties M-83 and RGC-1066 were obtained in compressed files FASTQ format. The reads were extracted and the transcriptome analysis was done using the pipeline shown in Figure 3.2.3.1.

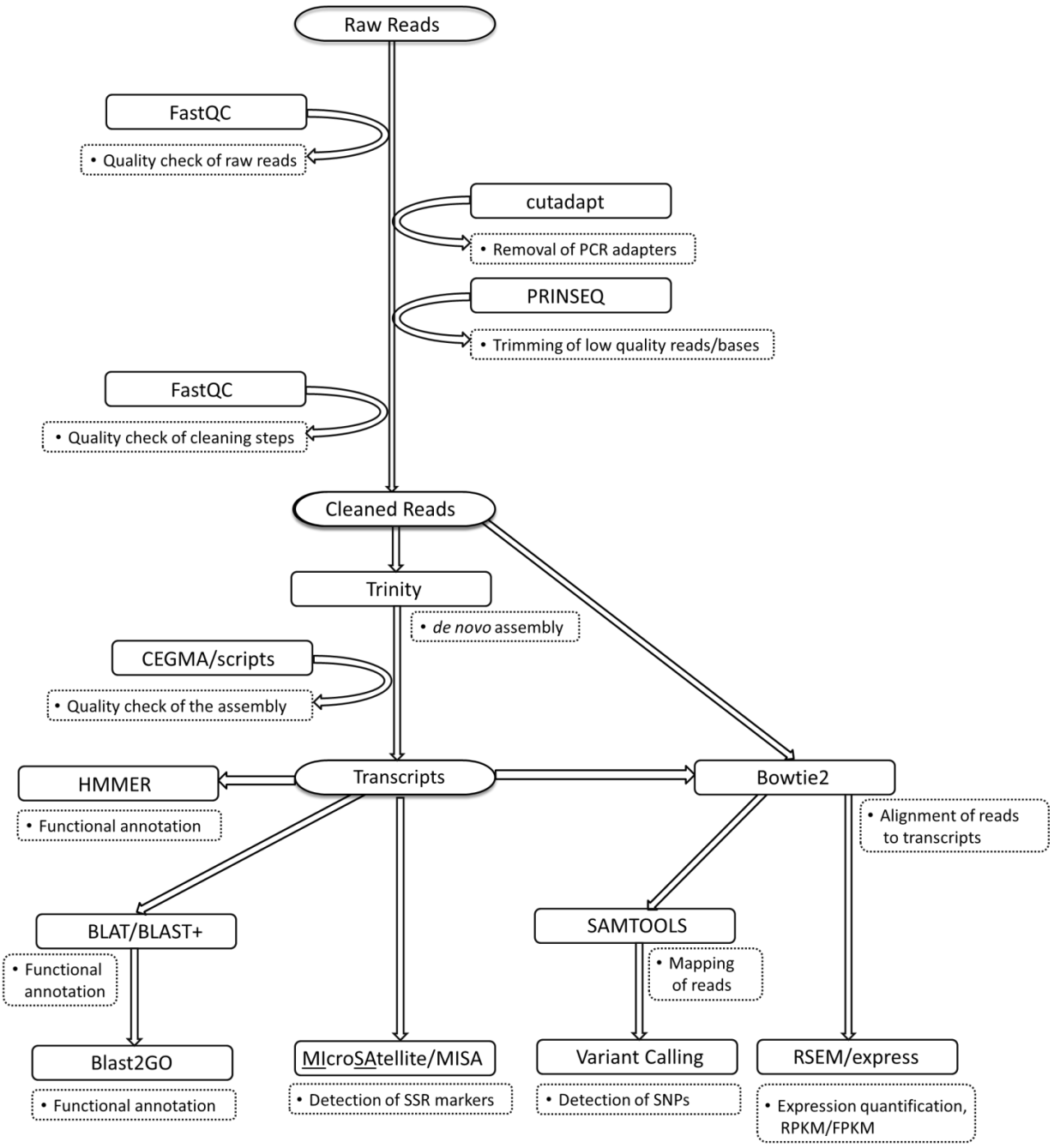


Figure 3.2.3.1 Workflow of guar leaves transcriptome analysis. (Modified from Kornobis [131])

3.2.3.1 Quality assessment of raw reads

The quality assessment of raw reads was carried out using FastQC version 0.11.4 [13] software. The paired end raw reads of sample-1 and sample-2 transcriptome obtained from Illumina HiSeq 2500 machine were analyzed for quality checking by assessment of average

sequence length, sequence quality, sequence length distribution, GC content, duplication level and *K-mer* content.

3.2.3.2 Cleaning of raw reads

The paired end raw reads of sample-1 and sample-2 were cleaned for removal of sequences of PCR adapters, duplicate sequences and other sequence contaminants by similarity search. These reads were further processed for trimming of low quality reads/bases with ambiguous sequences "N". The cutadapt 1.9.1 [163] software was used to remove the adapter sequences from paired end raw reads. The bar code and adapter sequences used were as follows:

Truseq adapters

Sample 1 - GTTTCG

5'ATCGGAAGAGCACACGTCTGAACTCCAGTCAC**GTTTCG**GAATCTCGTATGCCGTCTTCTGCTTG3'

Sample 2 - GAGTGG

5'GATCGGAAGAGCACACGTCTGAACTCCAGTCAC**GAGTGG**ATATCTCGTATGCCGTCTTCTGCTTG3'

Universal adapter

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT 3'

The read orientation based pooling of the clean reads of both varieties was carried out to generate a single end read of R1 S12 and R2 S12 reads as shown in Figure 3.2.3.2. PRINSEQ version 0.20.4 [237] software was used for removal of low quality reads/bases. The following parameters were set on the PRINSEQ:

Minimum length in bp: 40

Minimum mean quality: 20

Trim quality window: 1

Trim 5' (in bp): 10

Trim tail 3' (in bp): 6

Low complexity threshold: 32

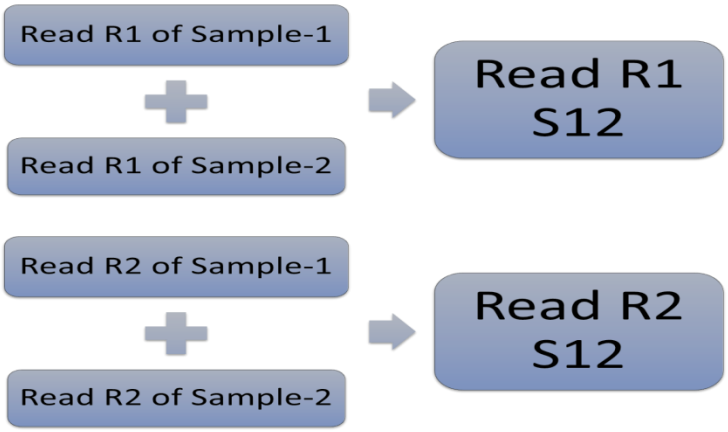


Figure 3.2.3.2 Merging pattern of raw reads of guar transcriptome.

The trimmed filtered reads obtained after cleaning were further analyzed for quality check using FastQC version 0.11.4 [13] software.

3.2.3.3 *De novo* transcriptome assembly

Transcriptomes User-Friendly Analysis web server (TRUFA) was used as cluster computing for *de novo* transcriptome analysis [131]. Transcriptome *de novo* assembly was carried out using the cleaned merged reads of sample-1 & sample-2 (Figure 3.2.4.2). The Trinity program [94] was used to assemble the clean reads for obtaining unigene contigs. Clean reads, with certain overlaps, were combined to form longer contigs (contiguous sequences), which were joined in to scaffolds that were further assembled by gap filling to acquire unigenes. Transcriptome coverage is highly variable owing to differential gene expression in cells, so there is no single absolutely optimal *k-mer* length for Transcriptome assembly. In this study, 25 was set as the default *k-mer* size for *de novo* assembly of the transcriptome, while other parameters used default values as given below:

Minimum contig length: 200
Strand Specific read orientation: RF, FR, F or R
Path reinforcement distance: 75
Minimum kmer coverage: 1
Minimum glue: 2
Group pairs distance: 500

The quality check of the transcriptome assembly was carried out by assessing the presence of 248 ultra-conserved CEGs (Core Eukaryotic Genes) in the assembly using CEGMA (Core Eukaryotic Genes Mapping Approach) computational method [198, 199]. The assembled transcripts were further clustered using CD-HIT version 4.5.4 (Sequence identity threshold 0.95) software to remove redundant transcripts [154]. This assembly and clustering generated the final dataset of clustered non-redundant unique sequences ("unigenes") for *C. tetragonoloba* leaf transcriptome.

3.2.3.4 Transcriptome annotation

Functional annotations were performed by the comparison of sequences of clustered assembly with public databases. The sequence similarity search of unigenes was performed by BLASTX [11] tool. All assembled unigenes were compared with the NCBI non-redundant protein (Nr), UniProt Reference Clusters (UniRef) [259], and Pfam [77] databases with default parameters to search for homologs. The BLAST+ [41] results against the nr database results were imported to Blast2GO suite [57] for mapping and retrieving Gene Ontology (GO) annotation of assembled unigenes, and further annotated with unique enzyme codes (EC). These retrieved GO terms were allocated to query sequences and the extensive groups of genes present in transcriptome were classified into three categories - cellular component, molecular function, and biological process. The WEGO tool was used for functional classification and graphical representation of GO terms at macro level [306]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways annotations were done by mapping assembled unigene sequences against KEGG metabolic pathway database [120]. Comparison of the assembled unigenes with the most closely related species was carried out using TRAPID online tool [271] with similarity search *E*-value $10e^{-5}$.

3.2.3.5 Differential gene expression

Both end paired R1 and R2 reads of each sample were merged individually as shown in Figure 3.2.3.3. The generated R12 S1 and R12 S2 reads were aligned against the reference (clustered assembly) using Bowtie2 version 2.2.6 [148], and mapped using express/RSEM [42, 151, 223] softwares with default parameters to obtain the FPKM values for each sample. The FPKM values were statistically analyzed to obtain the differential gene expression between the guar varieties M-83 and RGC-1066. The differentially expressed genes (DEGs) were identified

with a log-fold expression change (log FC) greater than 2 or less than 2 using a threshold of false discovery rate (FDR<0.001).

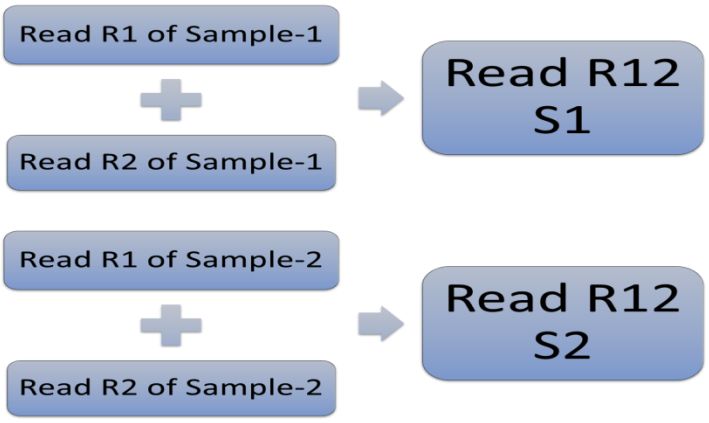


Figure 3.2.3.3 Merging pattern of clean reads of guar transcriptome.

3.2.3.6 Identification of simple sequence repeats (SSRs)

The clustered assembly from transcriptome data was used to mine the SSR motifs to obtain the molecular markers information in both the guar varieties. All the non-redundant transcript sequences were screened for repeat motifs using the PERL script MIcroSAtellite analyzer (MISA). The SSRs were considered to include six motifs: mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats. The following default definements were set for microsatellites (unit size/minimum number of repeats): (1/10) (2/6) (3/5) (4/5) (5/5) (6/5) in MISA. All motifs containing continuous uninterrupted repeats were classified as perfect and the motifs having two or more classes of repeats were classified as compound microsatellite. Maximal number of bases interrupting 2 SSRs in a compound microsatellite were set to 100. Statistical analysis was performed to summarize the number of SSRs with each type of motif and the length distribution of repeat units. A total number of 20 SSR markers representing all the motif types (except mononucleotide) were selected randomly for validation of these markers in both varieties. The primer designing was done by using Primer3 tool [130, 270].

3.2.3.7 Validation of SSR markers

3.2.3.7.1 DNA extraction and purification

Leaves were collected from the field grown guar plants of both varieties M-83 and RGC-1066, and DNA was extracted by the slightly modified CTAB method [69]. Approximately 0.5 g leaves were ground in liquid nitrogen to make a fine powder using pre-chilled, sterile mortar and pestle. The powder was added to a 2 ml Eppendorf tube containing 1 ml of pre-warmed DNA extraction buffer. The contents were gently mixed and incubated in a water bath at 65 ^{o}C for 1 h followed by incubation at room temperature for 10 min. An equal volume of chloroform: isoamyl alcohol (24:1) mixture was added to the Eppendorf tube. The contents were mixed gently and centrifuged at 8000 x g for 10 min at room temperature. The upper aqueous layer was pipetted out carefully in to another Eppendorf tube. The DNA was precipitated by adding an amount of ice cold isopropanol and incubating at 4 ^{o}C for 2 h. The precipitated DNA was pelleted by centrifugation centrifuged at 8000 x g for 10 min. The pellet was washed with 70 % ethanol, air dried and dissolved in 50 µl TE buffer. RNA was removed by treating with 2 µl of RNaseA (10 mg/ml) and incubating at 37 ^{o}C in a water bath for 1 h. An equal volume of chloroform: isoamyl alcohol (24:1) was added and the contents were centrifuged at 8000 x g for 10 min at room temperature. The pellet was discarded and the supernatant was transferred to another Eppendorf tube. The DNA was precipitated by adding an equal amount of 100 % ethanol and incubating at 4 ^{o}C for 2 h. The contents were centrifuged at 8000 x g for 10 min. The supernatant was discarded and the pellet was air dried. Finally the pellet was dissolved in 50 µl TE buffer.

3.2.3.7.2 Quantification and dilution of DNA

The quality of extracted DNA was determined by gel electrophoresis on 0.8 % agarose gel. The isolated DNA was quantified by measuring the absorbance at 260 nm in a UV-visible Varian spectrophotometer, model Cary 100 and diluted with TE buffer to approximately 100 ng/ µl. The diluted DNA samples were stored at -20 ^{o}C until use.

3.2.3.7.3 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was carried out in a Mastercycler gradient programmable thermal cycler (Eppendorf). The PCR program was run as shown in figure 3.2.3.4.

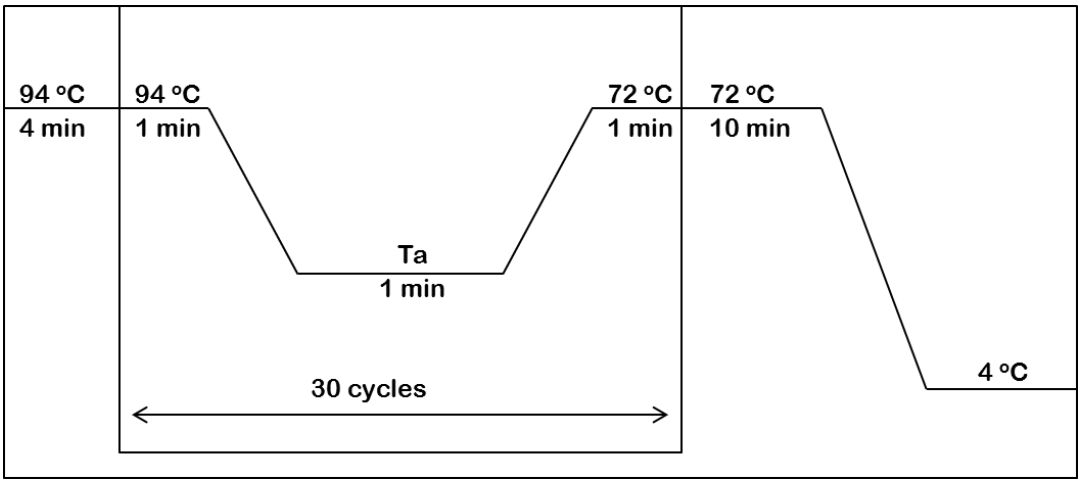


Figure 3.2.3.4 PCR program used for DNA amplification.

3.2.3.7.4 Polyacrylamide gel electrophoresis (PAGE)

PCR amplified products were electrophoresed on 8 % PAGE gels and silver stained. The gel was run at 5 V/cm for 5 to 6 h. A 100 bp DNA ladder was used as a molecular marker to determine the approximate size of the fragments. The gel was visualized under white light and documented in gel documentation unit (Bio-Rad).

3.2.3.7.5 Silver staining of poly-acrylamide gel

After gel electrophoresis, PAGE plates were disassembled. The gel was carefully placed in the staining tray. The gel was treated with fixative solution for 5 min with gentle rocking. After fixing the DNA, the fixing solution was decanted. Similarly the gel was incubated in staining solution for 5 min. The staining solution was decanted and the gel was washed gently with distilled water to remove excess silver nitrate on the gel and tray. The gel was then treated with developing solution for visualizing the bands. After the visualization of bands the developing solution was replaced by fixative solution for increasing the depth and sharpness of bands.

3.2.3.8 *In silico* analysis of SSR polymorphism

In silico identification of SSR polymorphism was carried out using Integrative Genome Viewer (IGV 2.3) software [97, 225]. The merged reads of each variety were mapped to the assembly using Bowtie2 version 2.2.6 [148] software to obtain the sorted transcripts as BAM files. The pairwise alignment of the sorted transcripts of both varieties was done against the

assembly using IGV 2.3 software and inspected manually to identify the SSR differences of 2 or more bp in guar varieties M-83 and RGC-1066.

3.2.3.9 Detection of single nucleotide polymorphisms (SNPs)

The analysis for single nucleotide polymorphisms (SNPs) was carried out in the guar varieties M-83 and RGC-1066. The detection of SNPs was done by using Bowtie2 version 2.2.6 [148] and SAMtools 1.3 [152, 153] programms in Integrated SNP Mining and Utilization (ISMU) pipeline [18]. Both R12 S1 and R12 S2 reads (Figure 3.2.4.3) were aligned against the reference (*de novo* assembly) using Bowtie2 version 2.2.6 program to generate the BAM (binary version of a SAM file) files of sorted transcripts. The BAM files were analyzed using SAMtools 1.3 program against the reference using single end mapping, and variant calling was run for SNPs and InDel variations mining. A position was called a putative SNP if any of the guar variety had a different allele against the reference. The putative SNPs were then filtered for a minimum read depth of 5 in each variety and at least 50 bp vicinity for an adjacent SNP. Furthermore high-confidence SNPs were obtained by screening for the homozygous allele types.

3.3 Studies on guar gum based drug delivery system

3.3.1 Purification of guar gum

Guar gum was purified using the method described by Cunha *et al*. [60] with slight modifications. Crude guar gum (5 g) was treated with 100 ml of boiling 80 % (v/v) ethanol for 10 min. The slurry obtained was collected on a glass filter (number 3) and washed successively with ethanol, acetone and ether. This material was added to 500 ml of distilled water and allowed to hydrate for 1 h in a conical flask. It was then mixed using a magnetic stirrer (15 min) and centrifuged at 1500 x g (Remi, Mumbai, India) for 15 min. The supernatant was precipitated in two volumes of cold acetone. After re-dissolving in hot water (80 ^{o}C), the guar gum solution was centrifuged at 6000 x g for 30 min at room temperature. The supernatant was precipitated with two volumes of ethanol. The precipitates were collected on a glass filter (number 4) and washed successively with ethanol and acetone and then dried overnight in an oven.

3.3.2 Hydrolysis of guar gum

Guar gum was partially hydrolyzed by acid catalyzed degradation. The guar gum solution (0.1 % w/v) was prepared in a beaker by dissolving appropriate amount of guar gum powder in deionized water. Acid hydrolysis was initiated by adjusting the solution to pH 1.0 using concentrated hydrochloric acid (HCl). The solution was kept at room temperature and continuously stirred on a magnetic stirrer. Samples were periodically removed and the reaction was terminated after 2 h by adjusting the solution pH 7 with 1 N NaOH.

3.3.3 Preparation of guar gum microparticles

The preparation of guar gum microparticles was carried out by two methods.

A) Precipitation and cross linking method: Guar gum (GG) was depolymerized by acid hydrolysis to get low molecular weight fractions. Estimation of the reduced end group was carried out by dinitro-salicylic acid method [170] to ensure the hydrolysis. The hydrolyzed suspension of GG was sonicated for 10 min and then filtered through 0.2 µm Millipore membranes. The precipitation was carried out by adding non-solvents like acetone, methanol, ethanol, isopropyl alcohol and ethylacetate, to the aqueous solutions of depolymerized GG previously mixed with surfactants, cross linkers and drug under constant magnetic stirring at 2000 rpm (Eltek, Mumbai, India). Triton X-100 and Tween 20 (non-ionic) were used as surfactants. Finally 25 % glutaraldehyde was added to affect cross linking. The composition of reaction mixture was as follows:

Name of reagent	**Quantity**
Depolymerized Guar gum (1 %)	10 ml
5-fluorouracil	5 mg
Triton X-100	20 µl
Tween 20	10 µl
Non solvent or Isoprpyl alcohol (IPA)	10 ml
Glutaraldehyde (25 %)	0.5 ml

B) Emulsification and cross linking method: Five mg of the drug was taken in 10 ml of hexane (drug loading solvent), this formed the oil phase in a conical flask. To this 5 mg of Span 80 was added under stirring at 800 rpm on a magnetic stirrer (Eltek, Mumbai, India). Then the oil phase was added to a 0.5 % aqueous guar gum solution under constant magnetic stirring at 2000 rpm. After mutual saturation of oil and continuous phase, the mixture was rapidly stirred at very high 4000 rpm. Glycerol (as stabilizer) was added followed by addition of 25 % glutaraldehyde to affect cross linking and kept overnight. The composition of reaction mixture was as follows:

Name of reagent	**Quantity**
5-fluorouracil	5 mg
Hexane	10 ml
Span 80	4 mg
MilliQ water	25 ml
Guar gum	0.5 %
Glutaraldehyde (25%)	0.5 ml
Glycerol	10 ml

The characterization of microparticles was carried out by using scanning electron microscopy (SEM) FEI Quanta 200F and atomic force microscopy (AFM) NT-MDT NTEGRA.

3.3.4 Drug release studies from guar gum microparticles

3.3.4.1 Standard curve of 5-fluorouracil

The phosphate buffer saline (PBS) pH 7.4 was prepared in a volumetric flask of 1,000 ml. The standard solution 5-fluorouracil (5-FU) in phosphate buffer (pH 7.4) was obtained by dissolving a quantity of 0.1 g of 5-FU in a volumetric flask of 100 ml. Stock solution of 5-FU in PBS was subjected to the UV spectrophotometric analysis and the absorption spectrum between 190-400 nm was recorded, and compared with the phosphate buffer pH 7.4 in a 1 cm cell. A peak was noticed at λ=265 nm. Three sets of solutions containing 5-FU were prepared in the concentration range of 1.0 - 10 µg/ml. For each set the absorbance was separately measured at λ = 265 nm and mean value was calculated. The data obtained were analyzed by linear regression (the calibration curve equation for this concentration range, the regression coefficient (r^2), intercept and the slope were calculated) and afterwards the calibration curve was drawn.

The composition of phosphate buffer saline (PBS) was as follows:

Name of reagent	**Concentration**
NaCl	137 mM
KCl	2.7 mM
Na_2HPO_4	4.3 mM
KH_2PO_4	1.47 mM

3.3.4.2 HPLC analysis of drug release from guar gum microparticles

The *in vitro* drug release study was carried out by using the method described by Krishnaiah *et al.* [138]. The quantitative determination of 5-fluorouracil was performed by high performance liquid chromatography (HPLC). A Shimadzu HPLC system (Shimadzu, Japan) with two LC-10AT VP pumps, a SPD-10A VP variable wavelength UV/Vis detector, a CTO-10AS VP column oven, a SCL-10A VP system controller (Shimadzu) and a RP C-18 column (250 mm×4.6 mm I.D.; particle size 5 μm; YMC, Wilmington, NC, USA) was used. The HPLC system was equipped with the software 'Class-VP series version 5.03 (Shimadzu)'. The mobile phase used was a mixture of methanol and sodium acetate buffer (pH adjusted to 4.0) in the ratio of 30:70. The filtered mobile phase was pumped at a flow rate of 0.8 ml/min. The column temperature was maintained at 40 °C. The eluent was detected by UV detector at 265 nm and the data were acquired, stored and analysed with the software Class-VP series version 5.03 (Shimadzu). The retention time of 5-fluorouracil using the present HPLC method was found to be 4.52 min. The low detection limit was found to be 100 ng/ml. The required studies were carried out to estimate the precision and accuracy of this HPLC method for analysis of 5-fluorouracil.

4. RESULTS

4.1 Sequencing of guar leaf transcriptomes

4.1.1 RNA extraction and quantification of guar leaves

The total RNA was extracted from the leaves of guar varieties M-83 and RGC-1066. The extracted RNA was found to be good in quality and concentration with the RNA integrity number (RIN) value above 6 for each sample. The results of quality control analysis of RNA extraction have been presented in Table 4.1.1.1 and Figure 4.1.1.1.

Table 4.1.1.1 Quality measures of total RNA extracted from leaves of guar varieties M-83 and RGC-1066

Name of Guar Variety	Sample Name	Nano Drop Concentration (ng/µl)	Qubit Concentration (ng/µl)	Absorbance 260/280	RIN Value
M-83	Sample-1	1310.4	>1000	2.13	6.0
RGC-1066	Sample-2	393.3	532	2.11	6.5

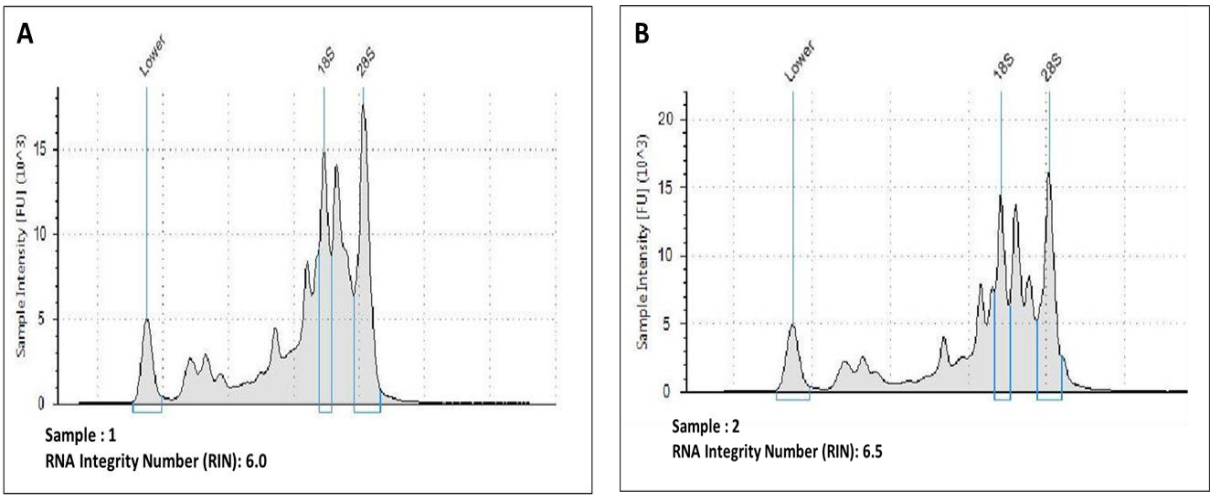


Figure 4.1.1.1 Quality control of total RNA extracted from the leaf samples; A) guar variety M-83 and B) guar variety RGC-1066.

4.1.2 Library preparation of guar leaf transcriptomes

The results of library preparation of leaf transcriptome of guar varieties M-83 and RGC-1066 are shown in Figures 4.1.2.1.

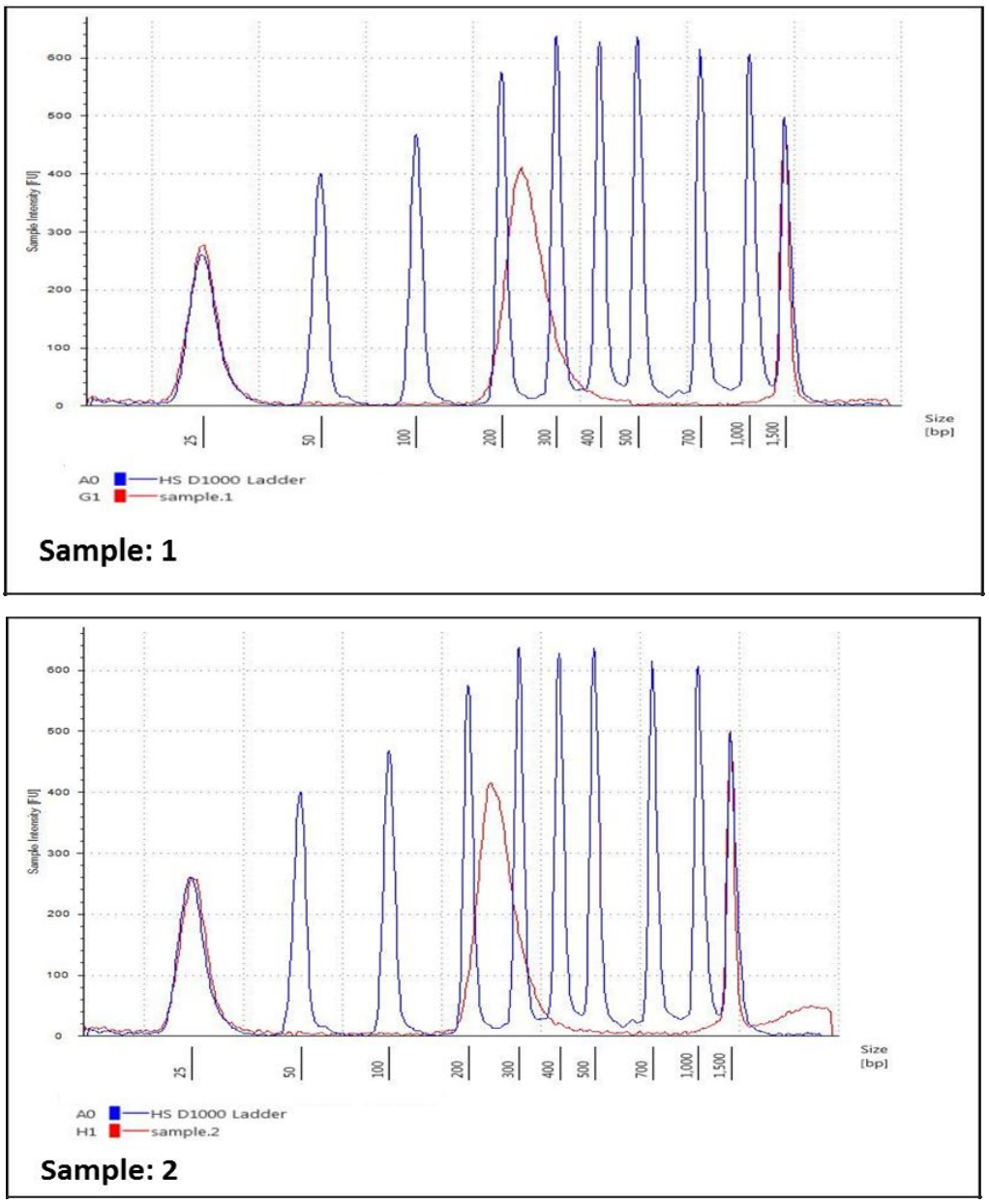


Figure 4.1.2.1 Library preparation results of leaf transcriptome of guar varieties; Sample-1: M-83 and Sample-2: RGC-1066.

4.1.3 Sequencing of guar leaf transcriptome libraries

A total of 28,688,024 and 33,018,878 pair-end raw reads (generated on Illumina Hiseq platform) for the leaves of guar varieties M-83 and RGC-1066, respectively, were obtained. These reads accounted for approximately 12 Gb of sequence data. Raw reads were further subjected to quality check.

4.2 Analysis of guar transcriptome data

4.2.1 Quality assessment of raw reads

The total numbers of bases in the pair-end raw reads of leaf trascriptomes of the guar varieties M-83 and RGC-1066 was found to be 28,688,024 and 33,018,878 bp respectively. The mean read quality (Phred Score) and % Q> 30 were approximately 35 and 90, respectively, for all the reads. The read length was 100 bp for each variety. The GC content was found to be 49 % for M-83 and 45 % for RGC-1066 variety. The quality boxplot for the reads are presented in Figures 4.2.1.1 to 4.2.1.8 and have been summarized in Table 4.2.1.1.

Table 4.2.1.1 Summary of Quality assessment report of raw reads of guar leaf transcriptomes

Sl. No.	Sample	Name of Guar Variety	Read Orientation	Mean Read Quality (Phred Score)	Number of reads	% GC	% Q<10	% Q 10-20	% Q 20-30	% Q>30	Number of Bases (MB)	Mean Read Length
1	Sample-1	M-83	R1	35.36	28,688,024	49.81	2.2	1.12	6.01	90.67	2868.8	100.0
			R2	34.45	28,688,024	49.61	4.45	1.24	6.19	88.12	2868.8	100.0
2	Sample-2	RGC-1066	R1	35.92	33,018,878	45.86	1.52	0.91	5.04	92.54	3301.89	100.0
			R2	35.43	33,018,878	45.51	2.9	0.96	4.81	91.33	3301.89	100.0

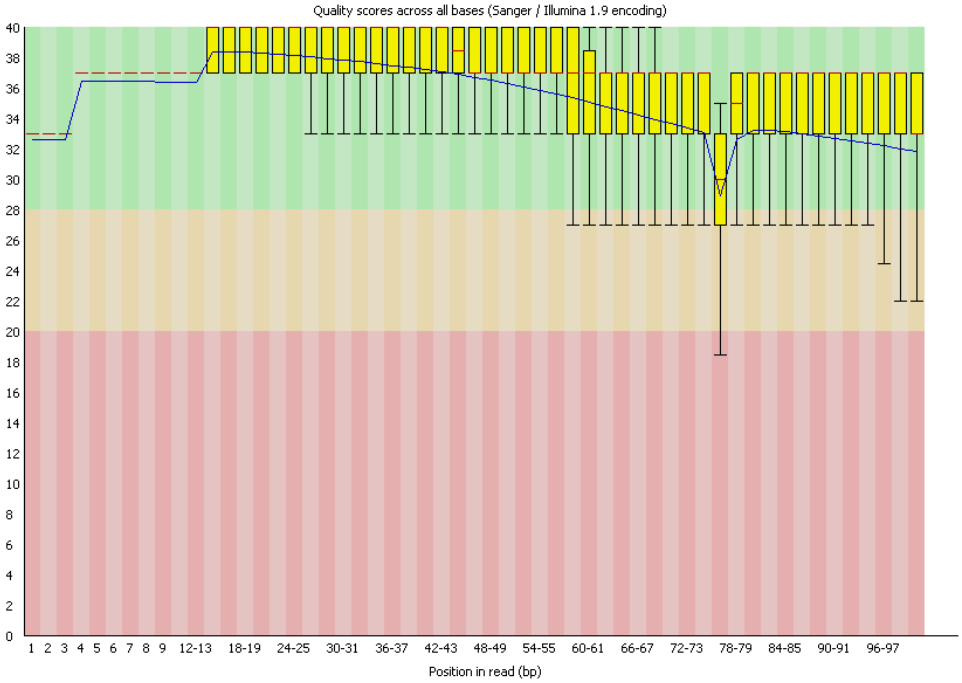


Figure 4.2.1.1 Per base sequence quality of Sample-1 R1 reads of guar leaf transcriptome.

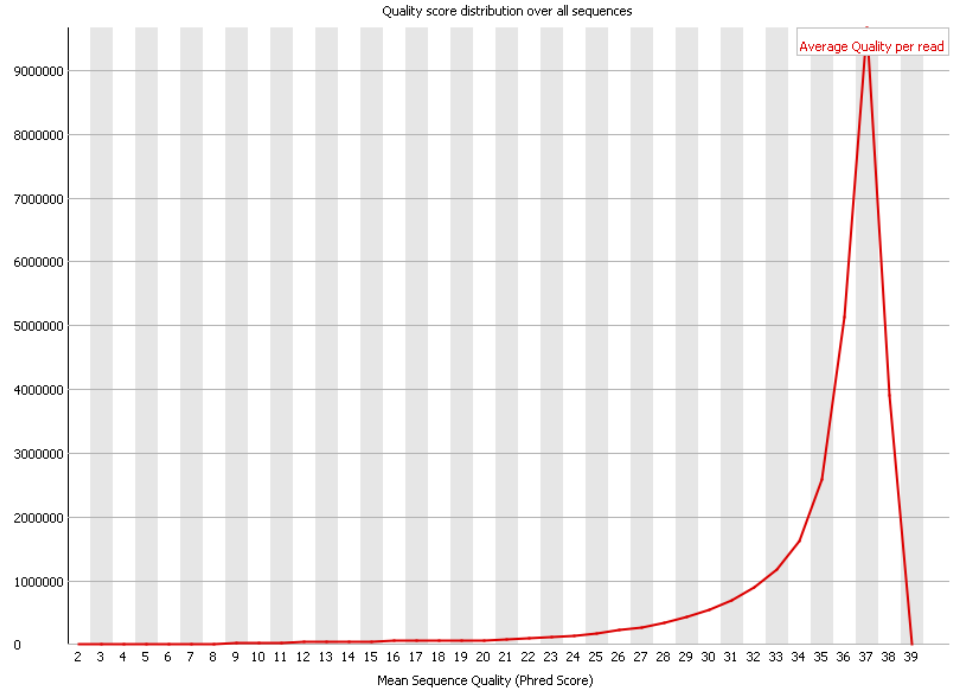


Figure 4.2.1.2 Per sequence quality score of Sample-1 R1 reads of guar leaf transcriptome.

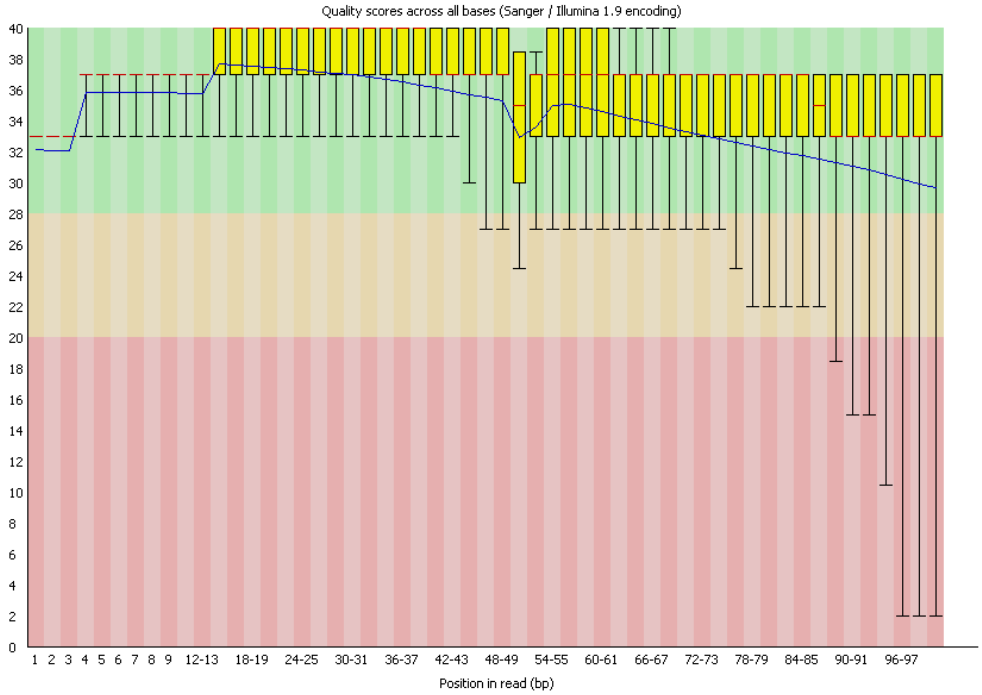


Figure 4.2.1.3 Per base sequence quality of Sample-1 R2 reads of guar leaf transcriptome.

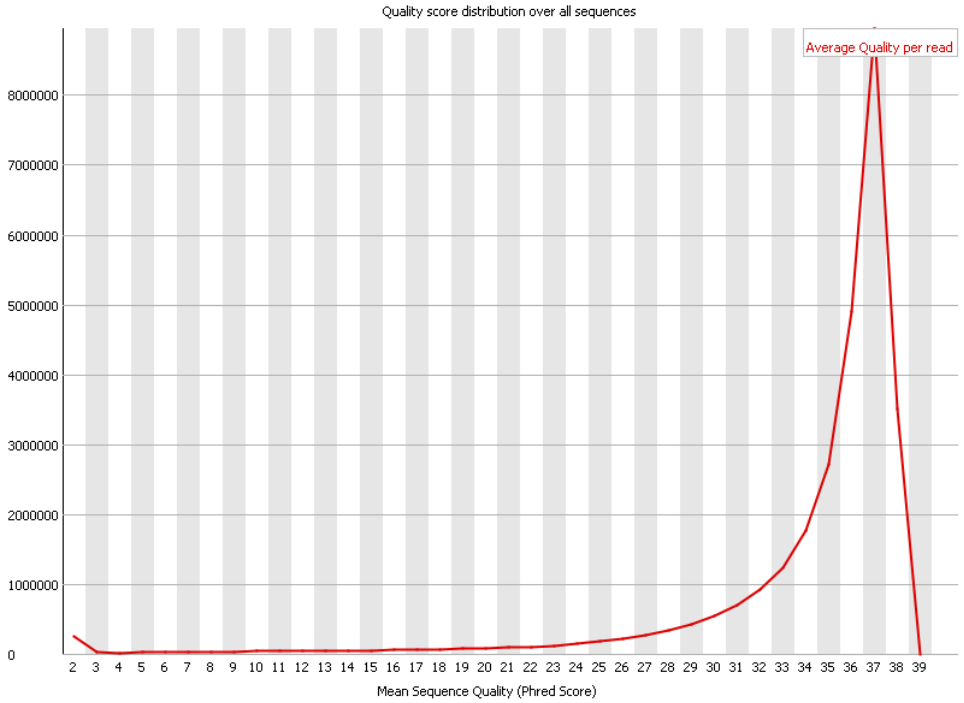


Figure 4.2.1.4 Per sequence quality score of Sample-1 R2 reads of guar leaf transcriptome.

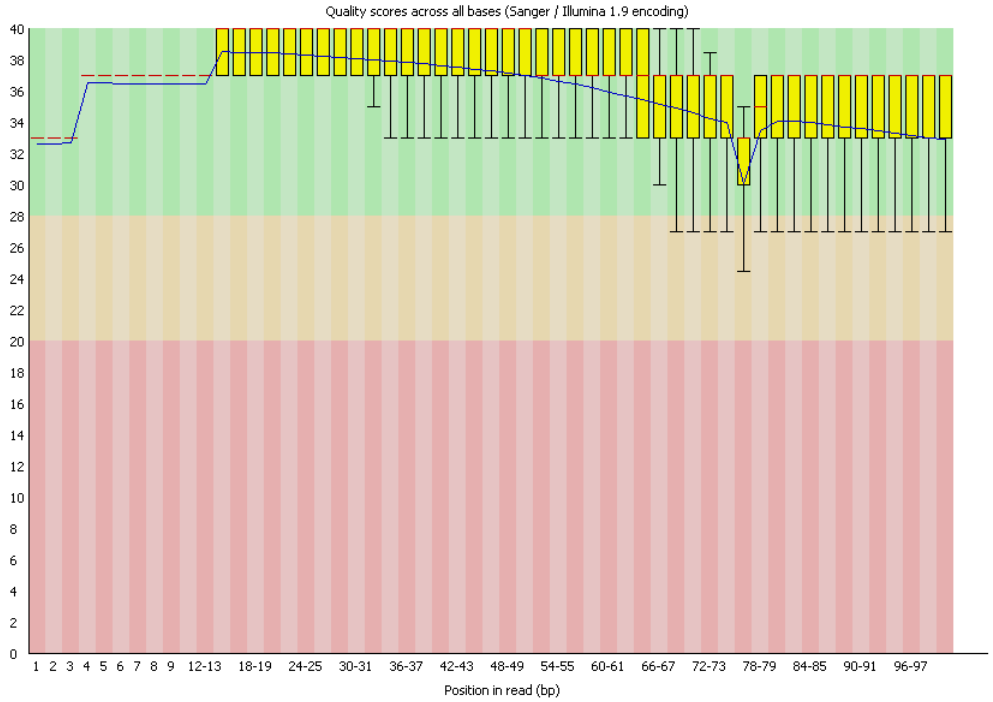


Figure 4.2.1.5 Per base sequence quality of Sample-2 R1 reads of guar leaf transcriptome.

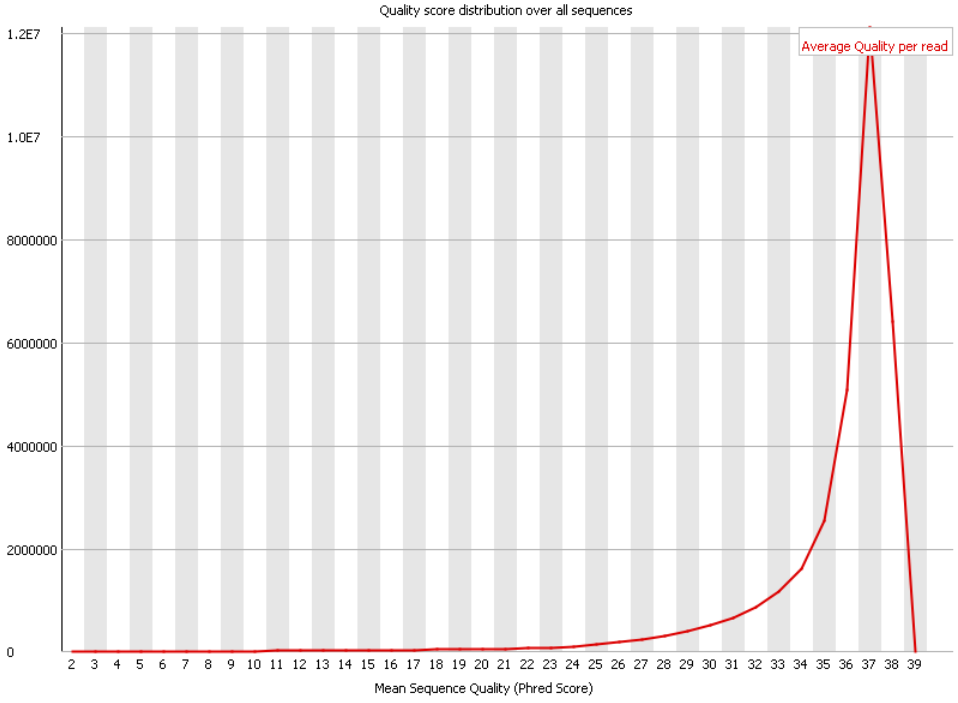


Figure 4.2.1.6 Per sequence quality score of Sample-2 R1 reads of guar leaf transcriptome.

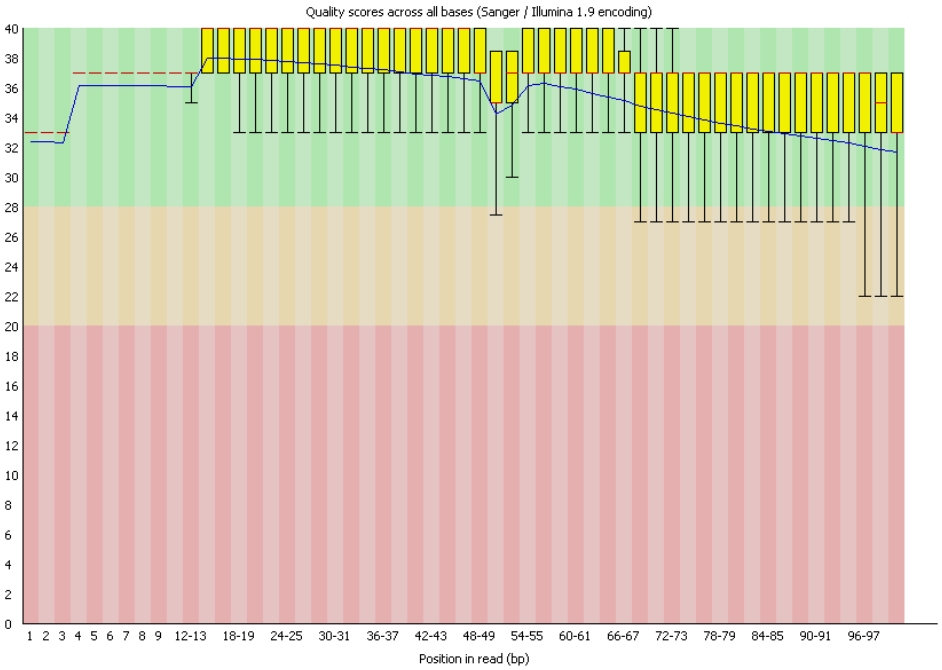


Figure 4.2.1.7 Per base sequence quality of Sample-2 R2 reads of guar leaf transcriptome.

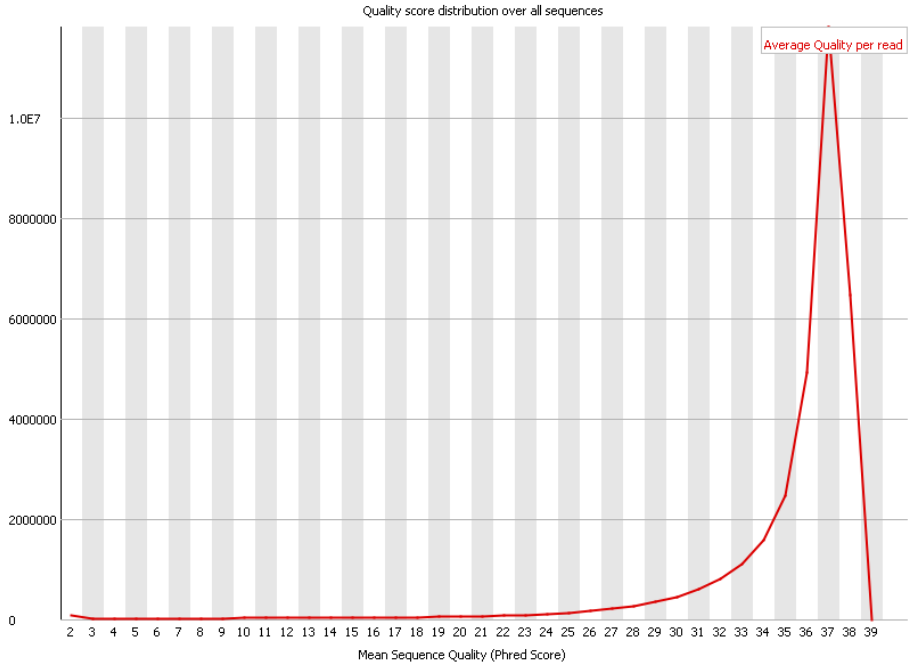


Figure 4.2.1.8 Per sequence quality score of Sample-2 R2 reads of guar leaf transcriptome.

4.2.2 Cleaning and merging of raw reads of guar leaf transcriptome

The reads were used for adapter clipping and trimming of low quality reads. The read orientation based pooling of the clean reads of both varieties was carried out to generate R1 S12 and R2 S12 reads. The clean and merged reads (filtered reads) were having the average read length of approximately 88 bp and 46 % GC content for all the reads. The figures 4.2.2.1 to 4.2.2.2 show the results of quality control analysis of the reads of guar leaf.

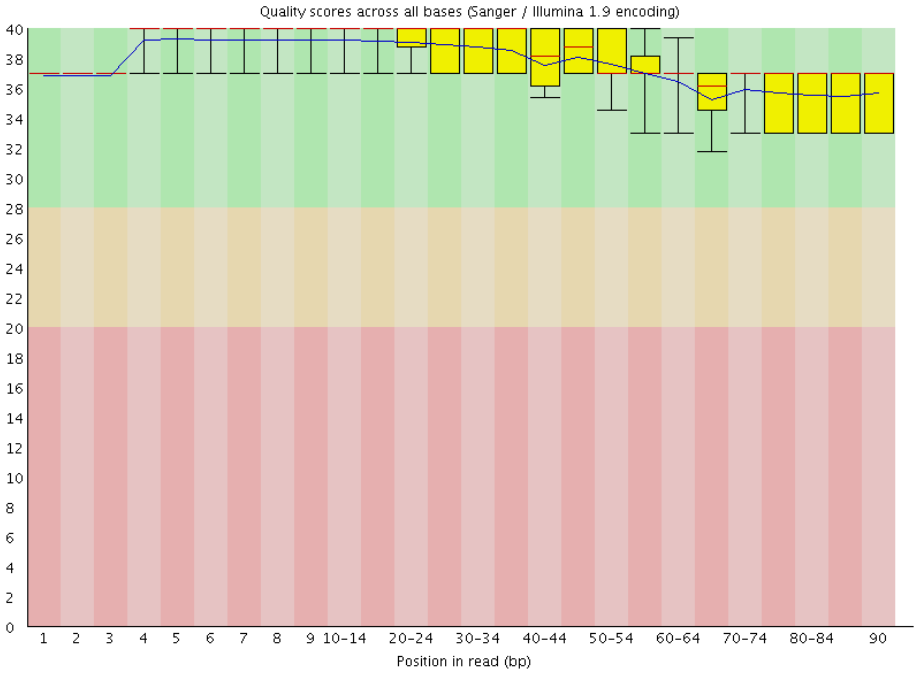


Figure 4.2.2.1 Per base sequence quality of R1 S12 reads of guar leaf transcriptome.

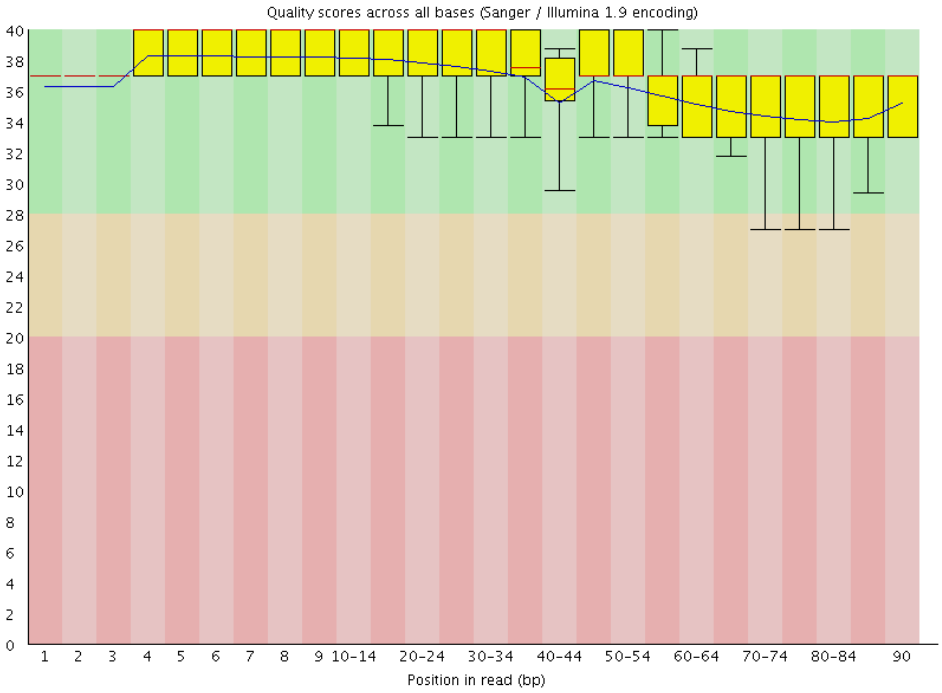


Figure 4.2.2.2 Per base sequence quality of R2 S12 reads of guar leaf transcriptome.

4.2.3 *De novo* transcriptome assembly of guar leaf

De novo assembly of all the clean reads by Trinity program [94] generated 79,355 contigs. The clustering of assembled sequences using CD-HIT version 4.5.4 [154] tool gave 62,146 non-reduntant unitranscripts having average length of 679 bp and an N^{50} of 1035 bp. The assembly statistics are shown in Table 4.2.3.1 and Figure 4.2.3.1. Among the contigs, the shortest and longest contigs are 201 and 29,056 bp, respectively. The length of 37,352 contigs was less than 500 bp whereas 24,794 contigs were having the length of more than 500 bp size. A total of 11,593 contigs were over 1000 bp and 237 contigs were over 5000 bp.

Table 4.2.3.1 Statistics of *de novo* assembly of guar leaf transcriptome

Characteristic	Details
Total number of Contigs	62,146
Min length	201
Max length	29,056
Average length	679.36
Standard deviation	792.86
Median length	394.0
Total bases in contigs	42,219,607
Number of contigs < 500 pb	37,352
Number of contigs ≥ 500 pb	24,794
Number of contigs ≥ 1000 pb	11,593
Number of contigs ≥ 2000 pb	3,292
Number of contigs ≥ 5000 pb	237
Number of contigs ≥ 10000 pb	38
N50	1035.0
Contigs in N50	11,028
GC content	43.68 %

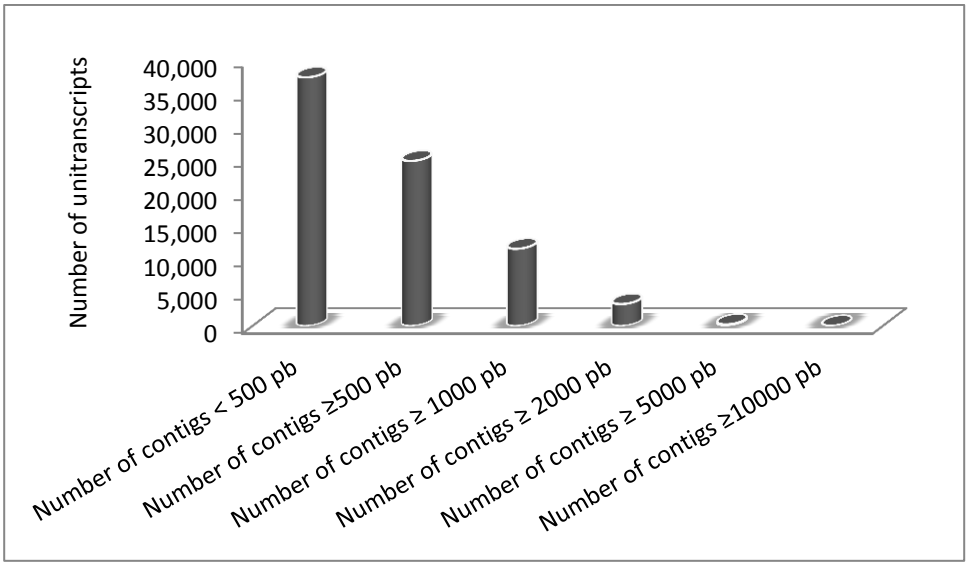


Figure 4.2.3.1 Sequence length distribution in *de novo* assembly of guar leaf transcriptome.

In order to assess the quality of *de novo* transcriptome assembly of guar, the clean reads were mapped to the assembled contigs. The overall alignment rate was 71 %. Among the mapped reads 74 % reads could uniquely map to the unigenes, while 11 % reads could map to multiple locations on unigenes. In addition, to check the quality of transcriptome assembly analysis was done to test the presence of 248 ultra-conserved Core Eukaryotic Genes (CEGs) in the assembly using CEGMA [198, 199]. The results revealed that assembly had 87.50 % of complete and 97.18 % partial CEGs. The statistics of CEGMA results have been depicted in Table 4.2.3.2.

Table 4.2.3.2 Statistics of CEGMA results of guar leaf transcriptome assembly

	Prots	%Completeness	Total	Average	%Ortho
Complete	217	87.50	490	2.26	70.51
Group 1	55	83.33	131	2.38	78.18
Group 2	45	80.36	101	2.24	73.33
Group 3	55	90.16	129	2.35	67.27
Group 4	62	95.38	129	2.08	64.52
Partial	241	97.18	661	2.74	80.50
Group 1	64	96.97	176	2.75	82.81
Group 2	54	96.43	158	2.93	85.19
Group 3	58	95.08	164	2.83	72.41
Group 4	65	100.00	163	2.51	81.54

#These results are based on the set of genes selected by Genis Parra

Prots represents the number of 248 ultra-conserved CEGs present in genome, % Completeness represents percentage of 248 ultra-conserved CEGs present, Total represents total number of CEGs present including putative orthologs, Average represents the average number of orthologs per CEG and % Ortho represents percentage of detected CEGS that have more than 1 ortholog.

4.2.4 Functional annotation of guar leaf transcriptome

The assembled leaf unigenes were annotated with multiple databases by using BLASTX against the NCBI non-redundant protein (Nr) database, Uniref90 database, Pfam database and Nt database with an *E*-value cut off of $1e^{-6}$. Figure 4.2.4.1 depicts the distribution frequency of sequences in transcriptome unigenes based on Blast2GO results. The total numbers of hits obtained in Uniref90 and Nr database were 44,992 and 45,972, respectively. Among the 62,146 unigenes, 44,268 (71.23 %) had at least one significant match in blast hit results with an *E*-value less than$1e^{-6}$ (Figure 4.2.4.2). The results indicated that most of the unigenes are protein coding genes. Based on the annotation results of the Nr database, the *E*-value distribution analysis showed that 72.29 % of the matched sequences had strong homology with the *E*-value < $1\ e^{-30}$, and 56.65 % of the matched sequences showed strong homology with the *E*-value <$1\ e^{-45}$,

whereas only 27.70 % of the matched sequences had high similarity with the *E*-value from$1e^{-30}$ to $1e^{-6}$ (Figure 4.2.4.3). The similarity distribution analysis revealed that 66.34 % of the BLAST hits sequences had a similarity higher than 80 %, whereas 33.65 % of BLAST hits sequences had a similarity ranging from 35 to 80 % (Figure 4.2.4.4). Figure 4.2.4.5 shows the distribution of Blast2GO three step processes including BLASTX, mapping and annotation of guar leaf transcriptome. The species distribution analysis revealed that leaf tissue in guar has a number of homologous sequences in many plant species (Figure 4.2.4.6). *Glycine max* genes have the highest similarity (41.91 %) with guar unigenes among the various plant species, followed by *Phaseolus vulgaris* (14.85 %), *Cicer arietinum* (13.30 %), *Sphingomonas melonis* (9.89 %) and *Medicago truncatula* (6.34 %) as shown in Figure 4.2.4.7.

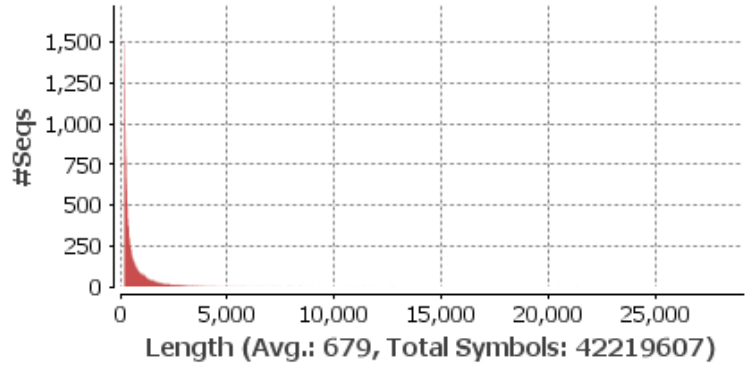


Figure 4.2.4.1 Sequence distribution in guar leaf transcriptome.

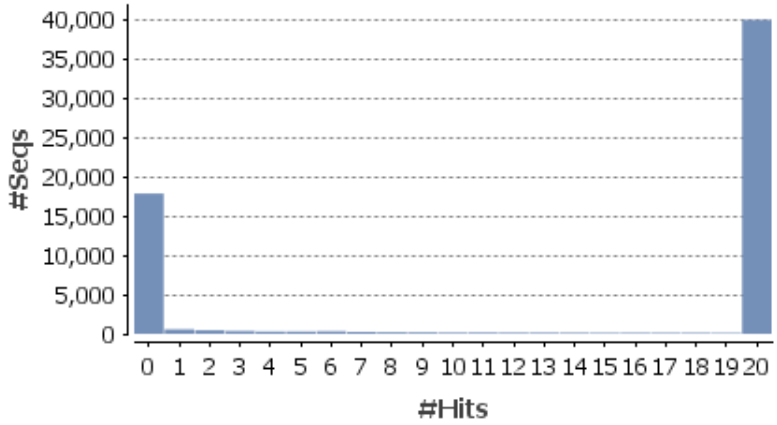


Figure 4.2.4.2 Blast hit distribution in guar leaf transcriptome.

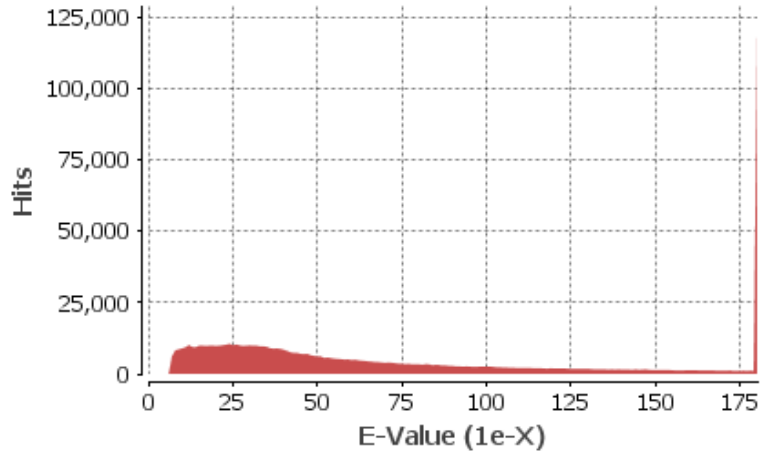


Figure 4.2.4.3 E-value distribution of BLAST hits for each unique sequence against the Nr database in guar leaf transcriptome.

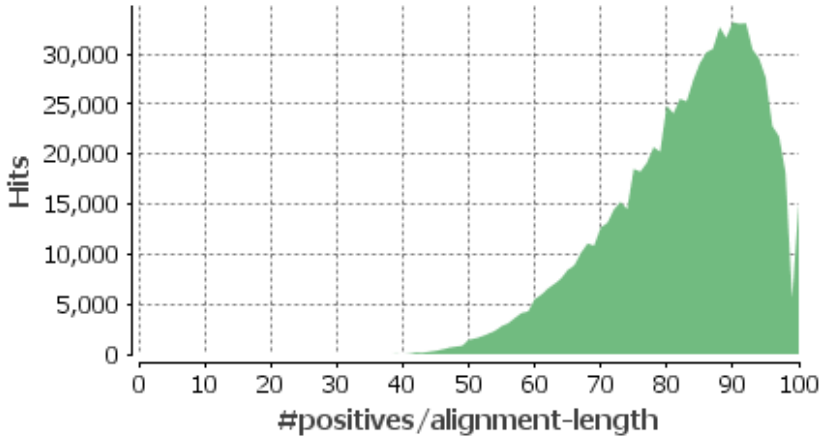


Figure 4.2.4.4 Similarity distributions of the top BLAST hits for each sequence against the Nr database in guar leaf transcriptome.

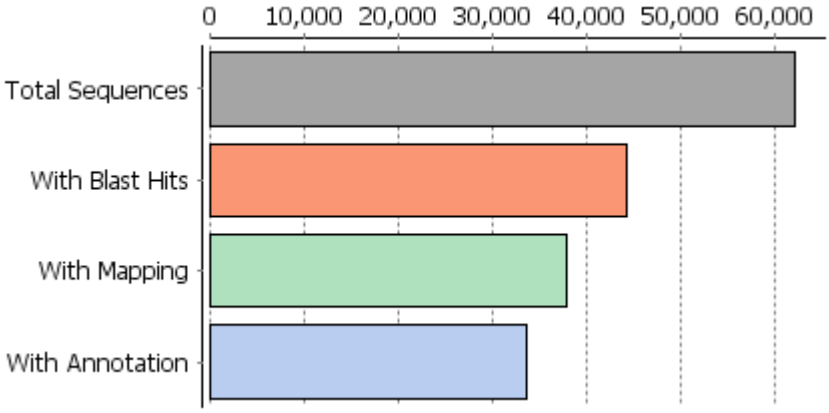


Figure 4.2.4.5 Distribution of Blast2GO three step processes including BLASTX, mapping and annotation of guar leaf transcriptome.

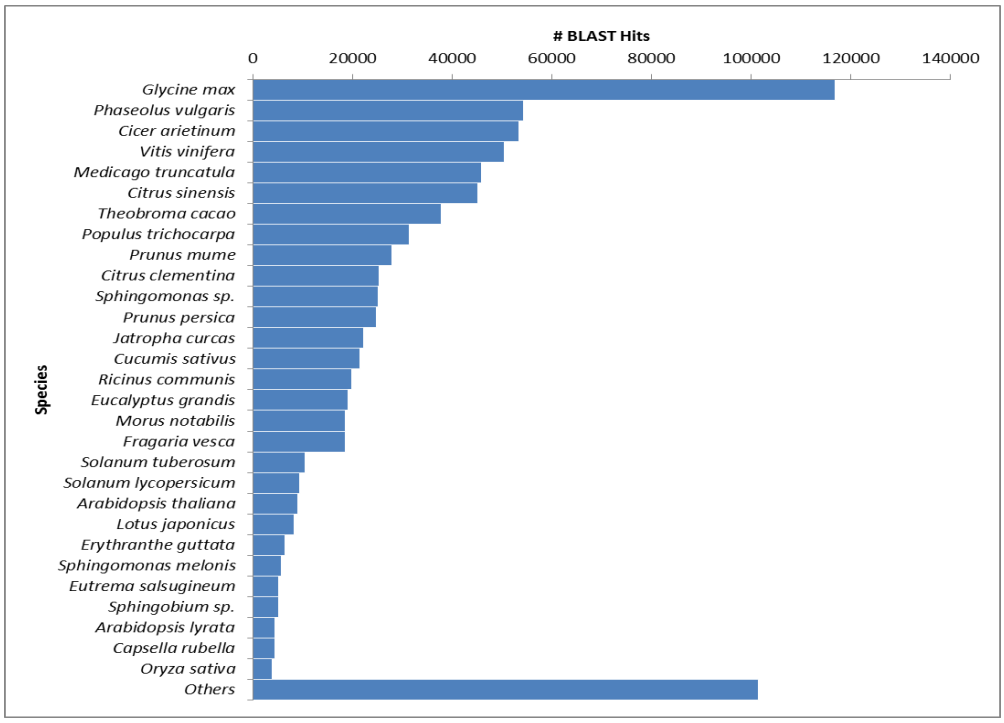


Figure 4.2.4.6 Species distribution by accounting all BLASTX hits in guar leaf transcriptome.

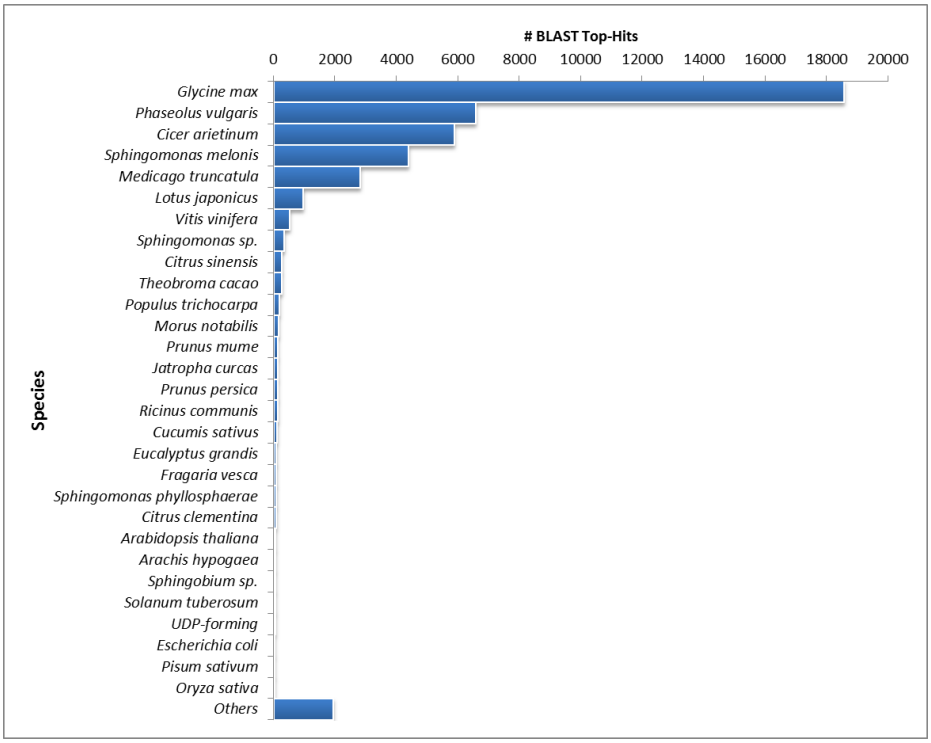


Figure 4.2.4.7 Top hit species distribution based on BLASTX alignments in guar leaf transcriptome.

4.2.4.1 Comparison of guar leaf transcriptome with closely related species

The comparison of assembled unigenes with closely related sequenced species was carried out by TRAPID analysis. Out of total 62,146 assembled unigenes 39,123 (63 %), 34,744 (55.9 %) and 35,263 (56.7 %) showed similarity to *Glycine max, Medicago truncatula* and *Lotus japonicus,* respectively. The detailed results of comparison with three species showing the meta annotation, gene family and functional annotation information have been presented in Table 4.2.4.1.

Table 4.2.4.1 Comparison of assembled guar leaf transcriptome with closely related sequenced species using TRAPID Analysis

Transcript Information			
Number of unigenes	62,146		
Average unigene length	679.4 bp		
	Glycine max	***Medicago truncatula***	***Lotus japonicus***
Meta Annotation Information			
Meta annotation full-length	7999 (12.9%)	7259 (11.7%)	6239 (10%)
Meta annotation quasi full-length	14948 (24.1%)	13472 (21.7%)	11058 (17.8%)
Meta annotation partial	15764 (25.4%)	13839 (22.3%)	11411 (18.4%)
Meta annotation no information	23435 (37.7%)	27576 (44.4%)	33438 (53.8%)
Similarity Search Information			
Similarity	39123	34744	35263
Gene Family Information			
Gene families	6454	5307	8382
Unitranscripts in GF	39123 (63%)	34744 (55.9%)	35263 (56.7%)
Functional Annotation Information			
Unitranscripts with GO	29981 (48.2%)	26607 (42.8%)	24079 (38.7%)
Unitranscripts with protein domain	34407 (55.4%)	30593 (49.2%)	28611 (46%)

4.2.4.2 Functional classification of guar leaf transcriptome by gene ontology (GO)

Based on sequence homology, 62,146 Trinity-assembled guar leaf unigenes were assigned with GO terms. A total of 175,882 annotations were found on the basis of BLAST+ results (Figure 4.2.4.8). The annotated GO terms were distributed into 46 functional groups, which were further classified under the three main categories, namely, biological process, molecular function and cellular component (Figure 4.2.4.9). Within the biological process, "metabolic process (23,214)", "cellular process (21,230)", "single-organism process (17,550)" and "biological regulation (7,295)" were the top GO terms. Among the molecular function category, "catalytic activity (18,275)", "binding (16,528)" and "transporter activity (2,164)" were major GO terms. In the cellular component, "cell (15,743)", "membrane (13,110)", "organelle (10,345)" and "macromolecular complex (4,985)" were mainly enriched. Furthermore, only a few unigenes were classified in terms of "cell killing", "behavior", "protein tag", "translation regulator activity", "nutrient reservoir activity" and "extracellular matrix". The GO terms were retrieved from Blast2GO and further analyzed using WEGO [306] tool. Similar results were obtained for the above mentioned categories as shown in Figure 4.2.4.10.

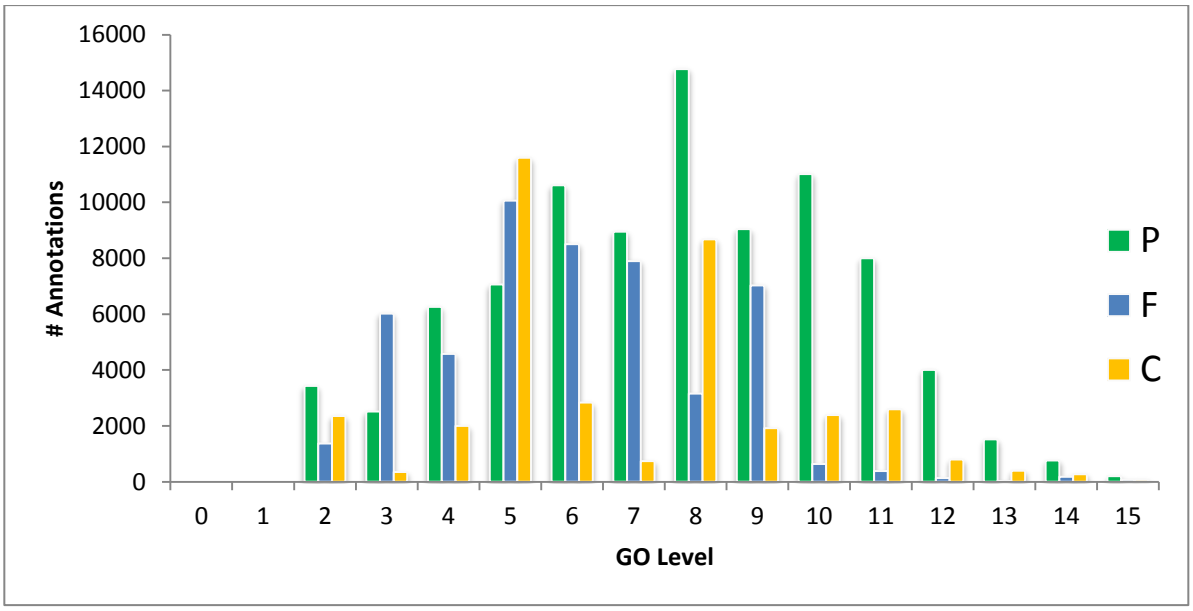


Figure 4.2.4.8 GO-level distributions in guar leaf transcriptome. P, F and C represent the biological process, molecular function and cellular component, respectively. Total Annotations = 175,882, Mean Level = 7.011

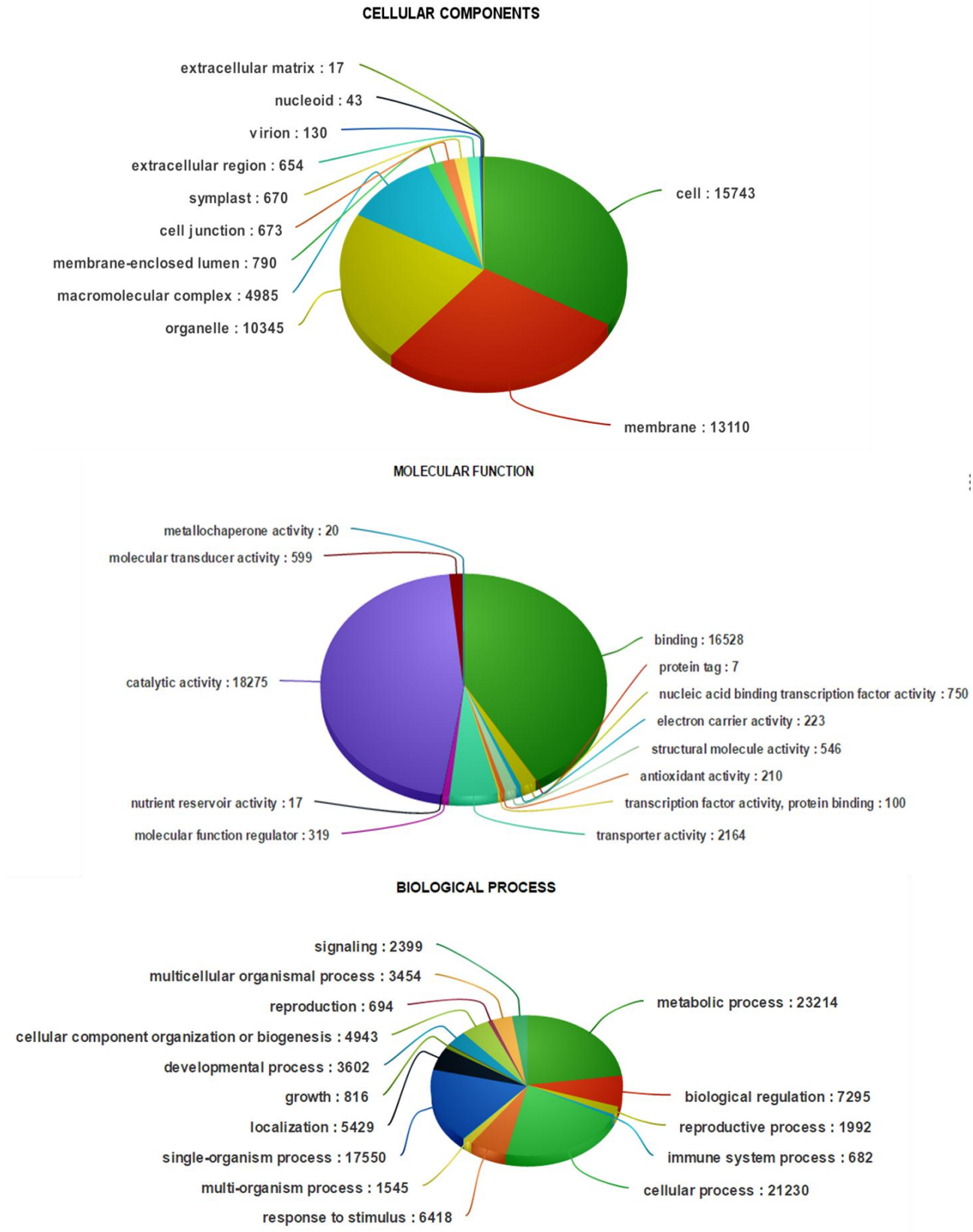


Figure 4.2.4.9 Classification of guar leaf transcripts into functional categories according to GO-terms.

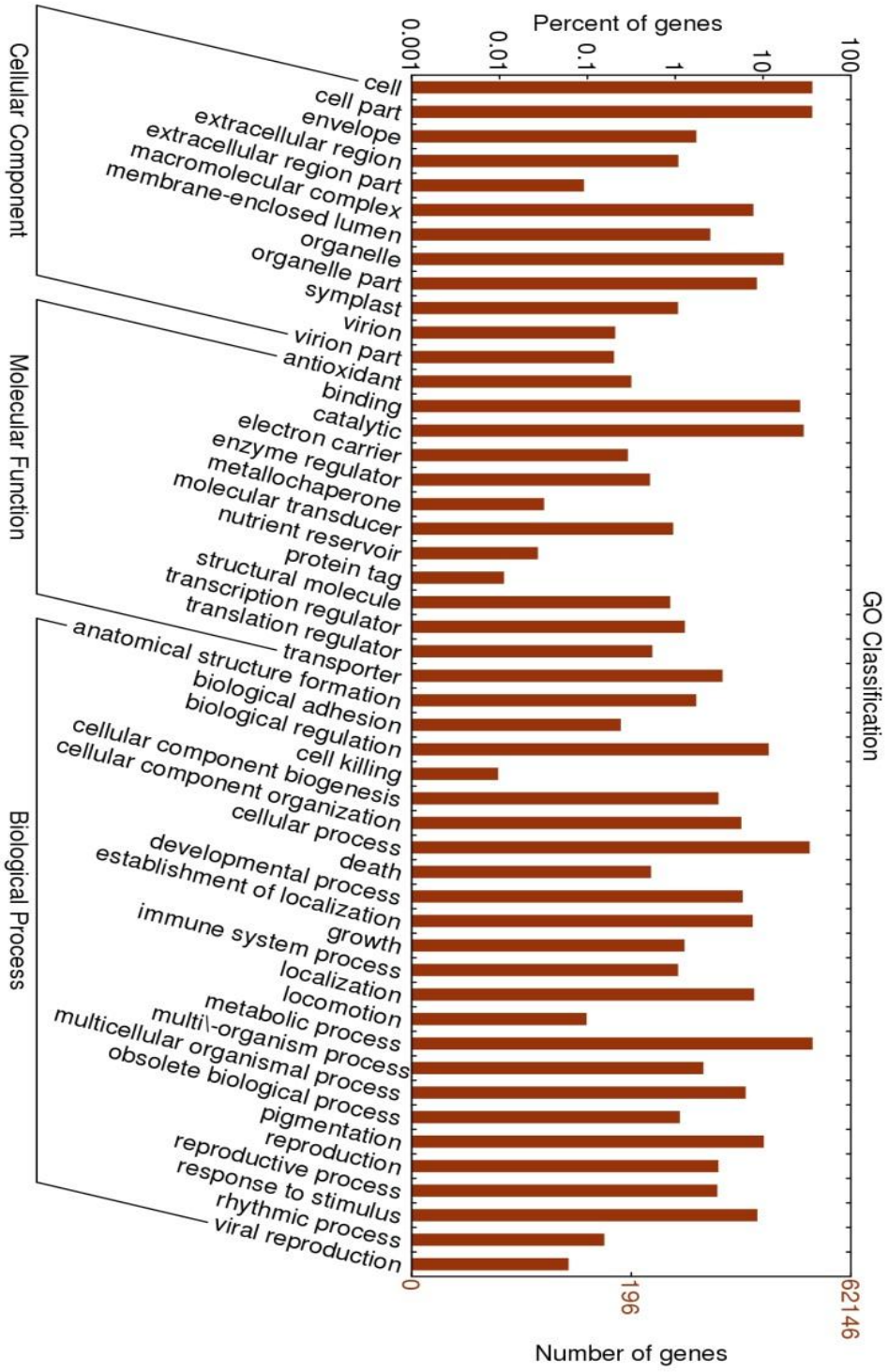


Figure 4.2.4.10 Classification of guar leaf transcripts into functional categories according to GO-terms on the basis of WEGO tool.

4.2.4.3 Enzyme detection in guar leaf transcriptome

By searching against the available database, a total of 11,308 guar unigenes were annotated with 1,759 enzyme codes (EC). The annotated enzyme codes included six classes: Oxidoreductases (2032 sequences), Transferases (4,641 sequences), Hydrolases (2,998 sequences), Lyases (521 sequences), Isomerases (408 sequences), and Ligases (708 sequences). The enzyme codes in guar leaf transcriptome were further classified in 55 subclasses. The detailed results are shown in Figure 4.2.4.11 and Table 4.2.4.2.

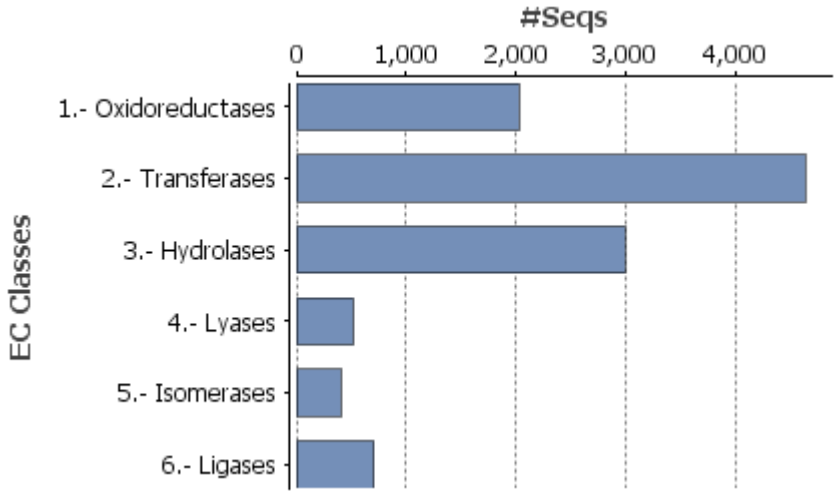


Figure 4.2.4.11 Enzyme code distributions in guar leaf transcriptome.

Table 4.2.4.2 Enzyme subclass distribution in guar leaf transcriptome

Sr. No.	Enzyme subclass	Number of unigenes
OXIDOREDUCTASES		
1	1.1.- Acting on the CH-OH group of donors	382
2	1.2.- Acting on the aldehyde or oxo group of donors	177
3	1.3.- Acting on the CH-CH group of donors	514
4	1.4.- Acting on the CH-NH(2) group of donors	53
5	1.5.- Acting on the CH-NH group of donors	59
6	1.6.- Acting on NADH or NADPH	168
7	1.7.- Acting on other nitrogenous compounds as donors	16
8	1.8.- Acting on a sulfur group of donors	91
9	1.9.- Acting on a heme group of donors	43
10	1.10.- Acting on diphenols and related substances as donors	39
11	1.11.- Acting on a peroxide as acceptor	169
12	1.12.- Acting on hydrogen as donors	3
13	1.13.- With NADH or NADPH as one donor	96

14	1.14.- With NADH or NADPH as one donor	211
15	1.15.- Acting on superoxide as acceptor	11
16	1.16.- Oxidizing metal ions	13
17	1.17.- Acting on CH or CH(2) groups	31
18	1.18.- Acting on iron-sulfur proteins as donors	4
19	1.20.- Acting on phosphorus or arsenic in donors	5
20	1.21.- Catalyzing the reaction X-H + Y-H = 'X-Y'	8
TRANSFERASES		
21	2.1.- Transferring one-carbon groups	641
22	2.2.- Transferring aldehyde or ketonic groups	43
23	2.3.- Acyltransferases	342
24	2.4.- Glycosyltransferases	664
25	2.5.- Transferring alkyl or aryl groups, other than methyl groups	152
26	2.6.- Transferring nitrogenous groups	102
27	2.7.- Transferring phosphorus-containing groups	2689
28	2.8.- Transferring sulfur-containing groups	42
HYDROLASES		
29	3.1.- Acting on ester bonds	809
30	3.2.- Glycosylases	673
31	3.3.- Acting on ether bonds	9
32	3.4.- Acting on peptide bonds (peptidases)	685
33	3.5.- Acting on carbon-nitrogen bonds, other than peptide bonds	181
34	3.6.- Acting on acid anhydrides	676
35	3.7.- Acting on carbon-carbon bonds	1
36	3.8.- Acting on halide bonds	10
37	3.13.- Acting on carbon-sulfur bonds	1
LYASES		
38	4.1.- Carbon-carbon lyases	213
39	4.2.- Carbon-oxygen lyases	209
40	4.3.- Carbon-nitrogen lyases	39
41	4.4.- Carbon-sulfur lyases	49
42	4.6.- Phosphorus-oxygen lyases	2
43	4.99.- Other lyases	10
ISOMERASES		
44	5.1.- Racemases and epimerases	79
45	5.2.- Cis-trans-isomerases	94
46	5.3.- Intramolecular oxidoreductases	63
47	5.4.- Intramolecular transferases	110
48	5.5.- Intramolecular lyases	14
49	5.99.- Other isomerases	48
LIGASES		
50	6.1.- Forming carbon-oxygen bonds	122
51	6.2.- Forming carbon-sulfur bonds	54

52	6.3.- Forming carbon-nitrogen bonds	463
53	6.4.- Forming carbon-carbon bonds	34
54	6.5.- Forming phosphoric ester bonds	31
55	6.6.- Forming nitrogen-metal bonds	18

4.2.4.4 KEGG pathways analysis in guar leaf transcriptome

A systematic analysis of high-level gene function by assigning guar leaf unigenes to the biochemical pathways in the KEGG database was performed. As a result, a total of 11,971 unigenes were assigned with 145 KEGG maps and 1,759 EC. These EC were used retrieve and color the KEGG pathway maps to represent the identified putative genes associated to several biochemical pathways. The annotated unigenes were categorized in to five major pathways in KEGG database (Figure 4.2.4.12-A) - "metabolism" (11,421), "genetic information processing" (132), "environmental information processing" (207), "organismal systems" (208) and "human diseases" (3). The "metabolism" was the most highly represented category which led to in-depth analysis of this group (Figure 4.2.4.12-B). The top five enriched pathways were "carbohydrate metabolism" (2,933), "amino acid metabolism" (1,754), "lipid metabolism" (1,297), "nucleotide metabolism" (1,094) and "energy metabolism" (1,070). The entire functional KEGG pathway categorization of guar leaf transcriptome unigenes have been shown in Appendix 1. An instance of a KEGG map for galactose metabolism has been shown in Figure 4.2.4.13**.**

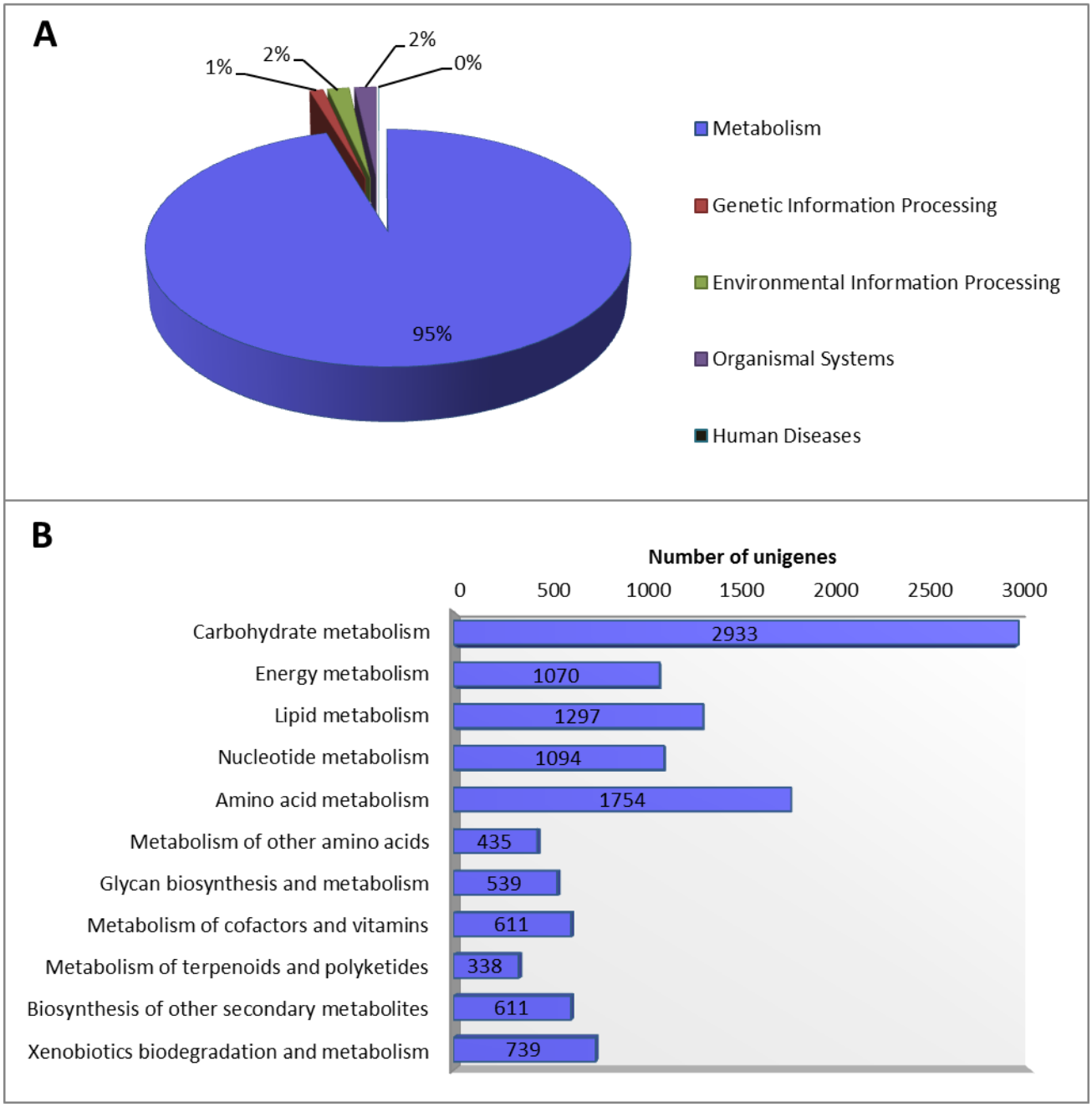


Figure 4.2.4.12 Annotation of guar leaf transcriptome in KEGG database. (A) Distribution of unigenes into KEGG biological categories. (B) Classification of unigenes in KEGG "metabolism" category.

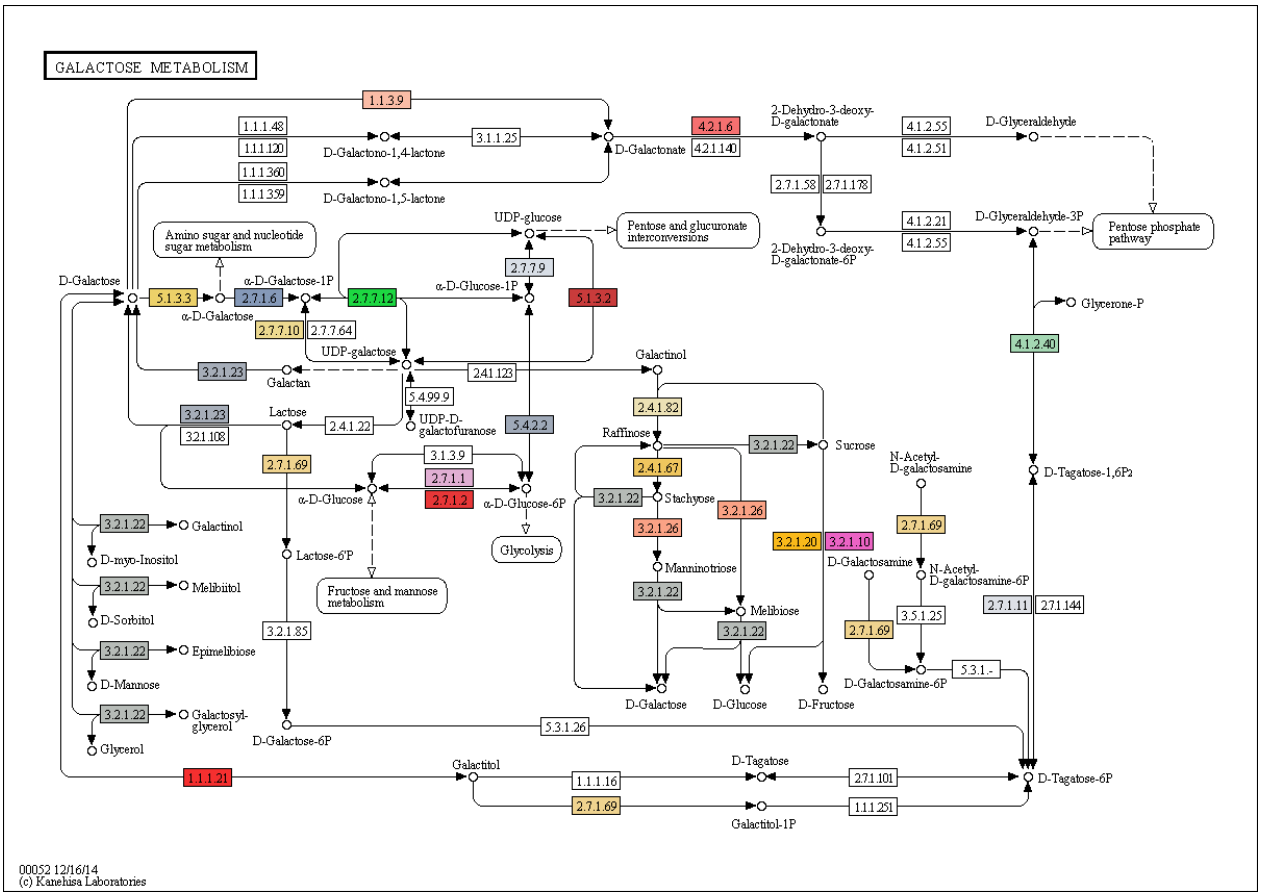


Figure 4.2.4.13 An instance of a KEGG map for galactose metabolism pathway. Each box represents the enzyme code involved in each section of the pathway. The colored boxes depict the enzymes identified in guar leaf transcriptome.

4.2.5 Differential gene expression in leaf transcriptomes of guar varieties M-83 and RGC-1066

A total of 62,146 transcripts were assessed for differential gene expression in the leaf transcriptome assembly of both guar varieties M-83 and RGC-1066. The guar M-83 leaf transcriptome was found to express 2,863 unigenes which were not expressed in RGC-1066, and 2,120 unigenes were found to be express only in guar RGC-1066 variety. Both the varieties showed ~80 % similar gene expression in leaf transcriptome, while, approximately 175 unigenes were found to be overexpressed in guar M-83 leaf transcriptome with at least 30 folds overexpression (Figure 4.2.5.1). These unigenes were further annotated against KEGG database and 36 KEGG maps with 49 enzyme codes (EC) were found to be present in transcripts with more than 30 folds expression (Table 4.2.5.1). A total of 158 unigenes were found in RGC-1066

variety with overexpression of 20 folds. Only two KEGG maps with five EC were annotated. A representative KEGG map of the biochemical pathway is shown in Figure 4.2.5.2.

Table 4.2.5.1 Differential gene expression on the basis of KEGG maps in guar varieties M-83 and RGC-1066

Pathway	**Pathway ID**
Guar variety M-83	
Aminobenzoate degradation	map00627
Steroid hormone biosynthesis	map00140
Glycolysis / Gluconeogenesis	map00010
Butanoate metabolism	map00650
Arginine and proline metabolism	map00330
Chlorocyclohexane and chlorobenzene degradation	map00361
Pentose and glucuronate interconversions	map00040
Methane metabolism	map00680
Fatty acid degradation	map00071
Caffeine metabolism	map00232
Purine metabolism	map00230
Glycine, serine and threonine metabolism	map00260
Valine, leucine and isoleucine biosynthesis	map00290
Carbon fixation in photosynthetic organisms	map00710
Pantothenate and CoA biosynthesis	map00770
Limonene and pinene degradation	map00903
Porphyrin and chlorophyll metabolism	map00860
Arginine biosynthesis	map00220
Alanine, aspartate and glutamate metabolism	map00250
Valine, leucine and isoleucine degradation	map00280
Thiamine metabolism	map00730
beta-Alanine metabolism	map00410
Drug metabolism - other enzymes	map00983
Glyoxylate and dicarboxylate metabolism	map00630
Lysine degradation	map00310
Starch and sucrose metabolism	map00500
Glycerolipid metabolism	map00561
Linoleic acid metabolism	map00591
Arachidonic acid metabolism	map00590
Vitamin B6 metabolism	map00750
Starch and sucrose metabolism	map00500
Glycerolipid metabolism	map00561
Linoleic acid metabolism	map00591
Arachidonic acid metabolism	map00590

Carbon fixation pathways in prokaryotes	map00720
Vitamin B6 metabolism	map00750
Guar variety RGC-1066	
Purine metabolism	map00230
Thiamine metabolism	map00730

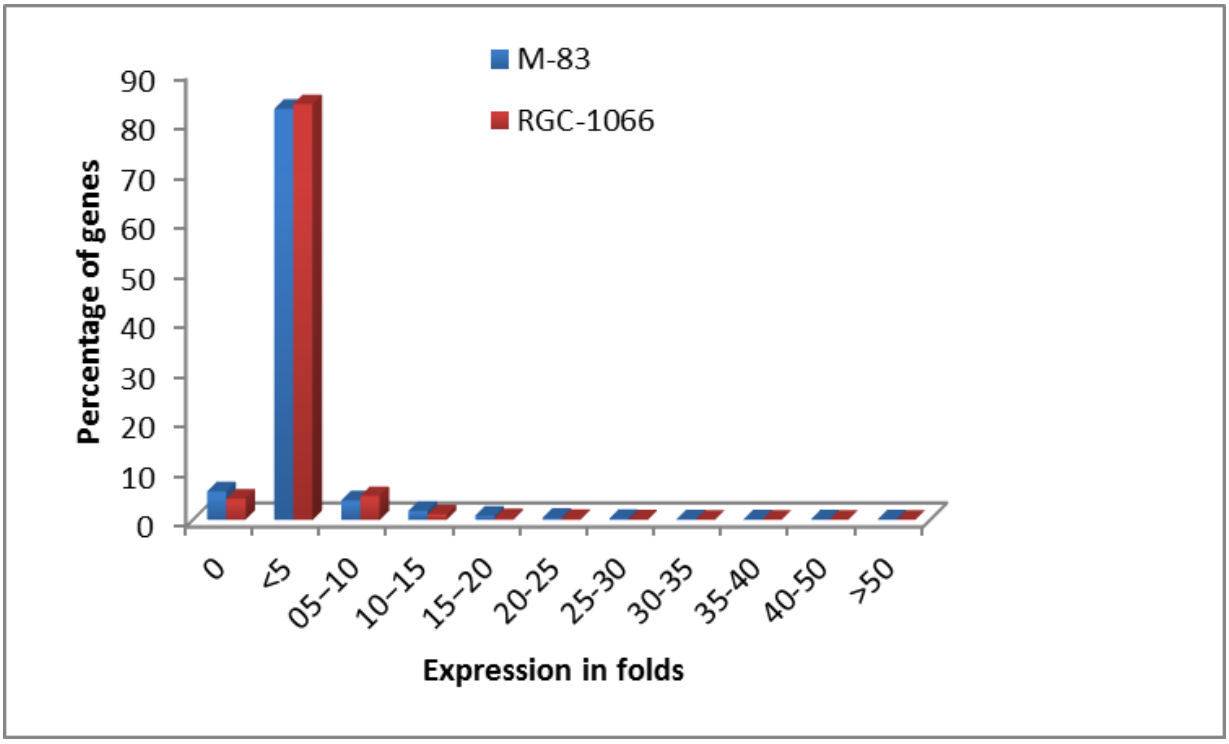


Figure 4.2.5.1 Differential gene expression profile in leaf transcriptomes of guar varieties M-83 and RGC-1066.

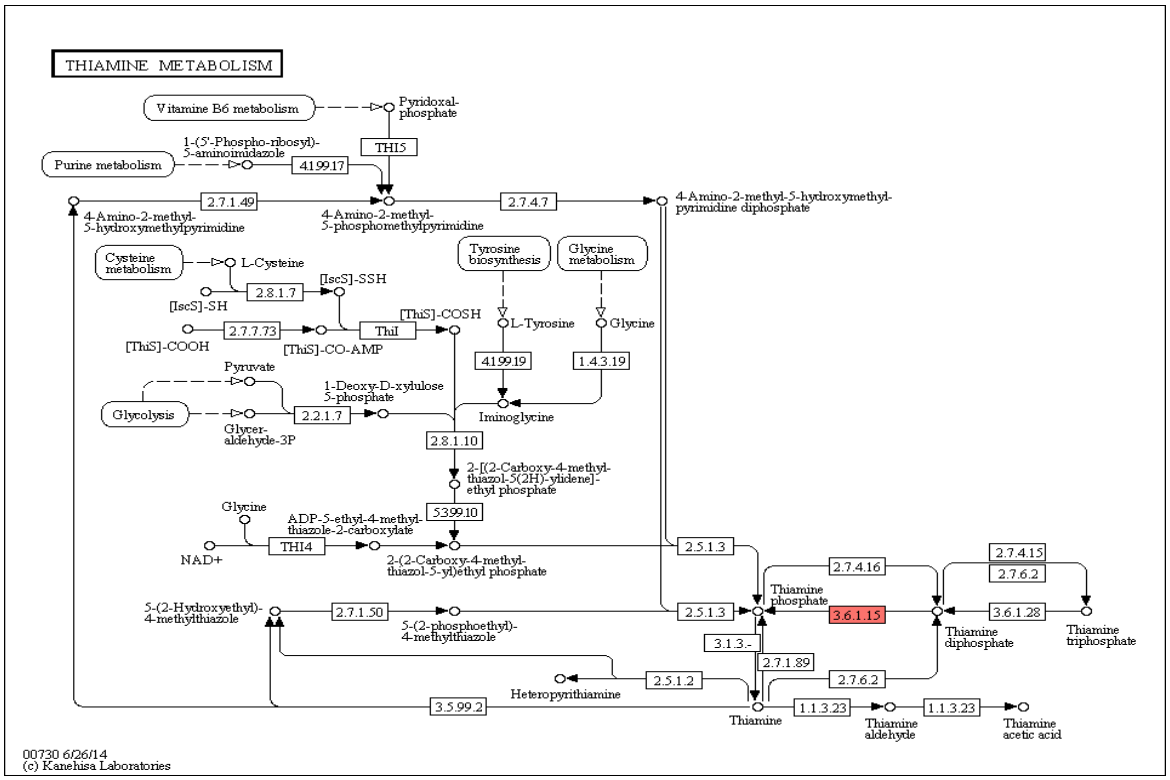


Figure 4.2.5.2 KEGG map for metabolism of thiamine metabolism pathway. Boxes represent the enzyme code involved in each part of the pathway. Colored boxes represent the enzyme codes present in guar assembly.

4.2.6 Identification of molecular markers in guar leaf transcriptome

In the present study, two types of molecular markers, namely, SSRs and SNPs were identified in the guar leaf transcriptome.

4.2.6.1 Mining of SSRs

SSRs mining was carried out on the assembled unigenes using MISA perl script. Out of total 62,146 unigenes assembled in guar leaf transcriptome, 4,970 unigenes were found to contain 5,773 SSRs. More than one SSR was found on 593 unigenes. On an average, one SSR could be found every 7.31 kb in the unigenes. A comprehensive analysis was performed to describe the type, frequency and distribution of all the potential SSRs. The SSRs included 2,624 (45.45 %) mononucleotide motifs, 1,179 (20.42 %) dinucleotide motifs, 1,856 (32.14 %) trinucleotide motifs, 97 (1.68 %) tetranucleotide motifs, 7 (0.12 %) pentanucleotide motifs, and 10 (0.17 %) hexanucleotide motifs (Figure 4.2.6.1). The number distributions of repeat units in all SSRs have been depicted in Figure 4.2.6.2 and Table 4.2.6.1. The result showed that the repeat number of most SSRs was not more than 10, and only a limited number of SSRs with more than 20 repeat

sequences were observed. In addition, the repeat number and the total repeat length (= type number * type length * type average repeat number) of each SSR type were analyzed. For most dinucleotide type, the repeat number was distributed between 6 and 11, with an average value of 9.92. While all the repeat number of pentanucleotide and hexanucleotide type was only 17. The mononucleotide SSRs are excluded, because of the frequent homopolymer errors found in sequencing data. There was a large proportion of both di- and trinucleotides (96 %) while the rest amounted to less than 4 %. The trinucleotide repeats were found to be the maximum (1856).

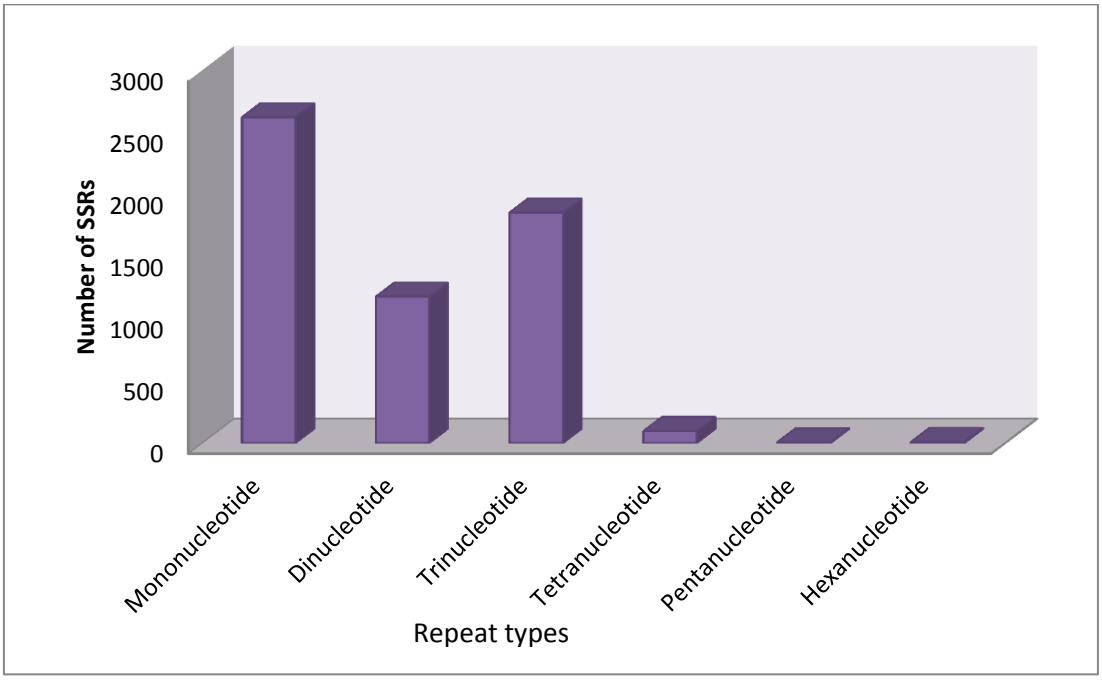


Figure 4.2.6.1 Distribution of different SSR repeat types in guar leaf transcriptome.

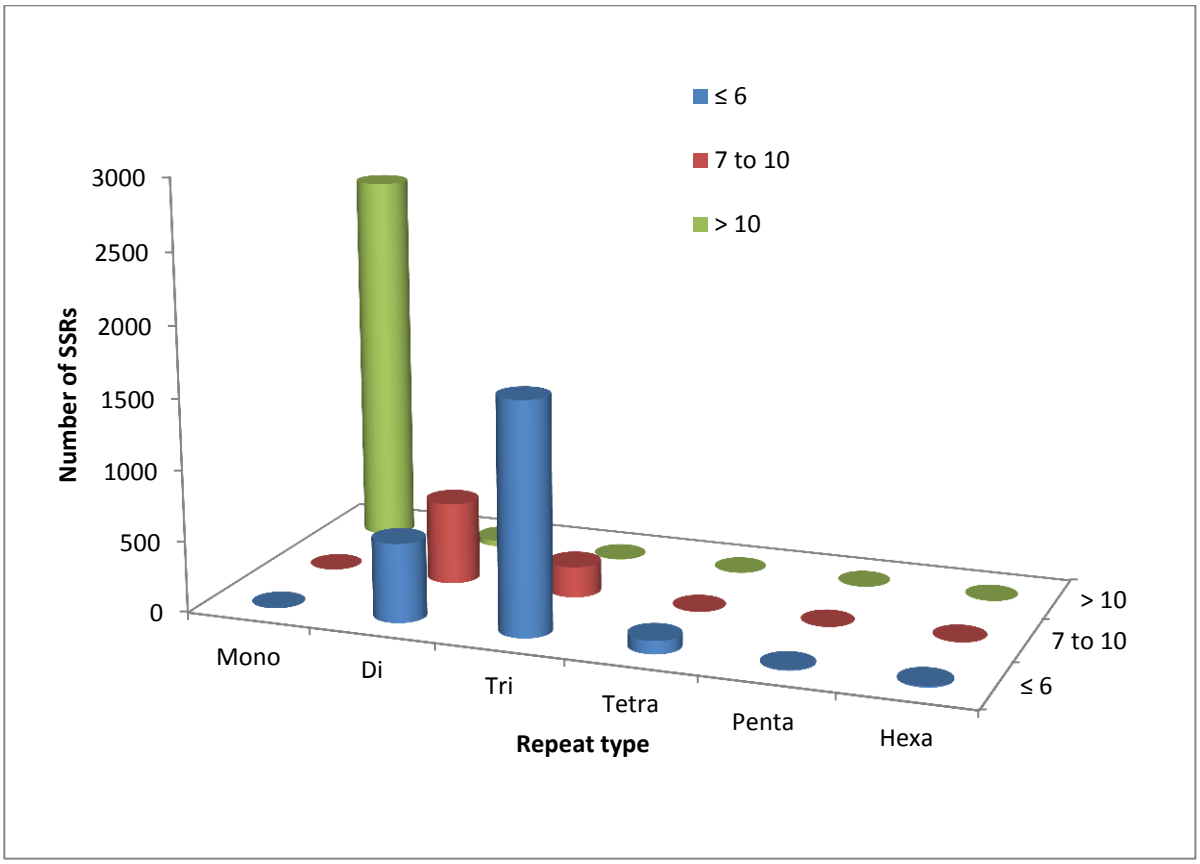


Figure 4.2.6.2 Frequency distribution of different SSR repeat types in guar leaf transcriptome.

Table 4.2.6.1 Frequency of classified SSR repeat types in guar leaf transcriptome

Repeats	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Total
A/T	-	-	-	-	-	1143	568	309	217	134	77	62	25	14	8	8	6	3	2	2576
C/G	-	-	-	-	-	16	10	6	1	3	4	6	2							48
AC/GT	-	71	38	31	6	7		2	1	1										157
AG/CT	-	426	190	118	78	47	32	2									3			896
AT/AT	-	54	31	13	14	4	1													117
CG/CG	-	8	1																	9
AAC/GTT	79	35	9	4																127
AAG/CTT	487	230	104	1																822
AAT/ATT	65	24	12	2			2													105
ACC/GGT	113	46	17	3		1														180
ACG/CGT	22	7	2	1																32
ACT/AGT	17	6	1																	24
AGC/CTG	91	32	16	5																144
AGG/CCT	97	50	21	3																171

ATC/ATG	124	39	11	2																176
CCG/CGG	45	27	2	1																75
AAAC/GTTT	2	1																		3
AAAG/CTTT	28	7					1													36
AAAT/ATTT	2																			2
AACC/GGTT					1															1
AACT/AGTT	1																			1
AAGG/CCTT	5	1	1																	7
AATC/ATTG	1																			1
AATG/ATTC	2																			2
AATT/AATT	12	1																		13
ACAT/ATGT		1																		1
ACGC/CGTG			1																	1
ACGG/CCGT	1																			1
ACTC/AGTG	3	3																		6
AGAT/ATCT	16	1																		17
AGCC/CTGG	1																			1
AGCT/AGCT	2																			2
ATCC/ATGG	2																			2
AAAAG/CTTTT	1																			1
AAACG/CGTTT		1																		1
AAGAG/CTCTT		1																		1
AATTC/AATTG	1																			1
ACGGC/CCGTG	1																			1
AGAGG/CCTCT			1																	1
AGATG/ATCTC	1																			1
AAAAAG/CTTTTT	2																			2
AAAAGG/CCTTTT		1																		1
AAATTC/AATTTG	1																			1
ACAGCT/AGCTGT	1																			1
ACCGGG/CCCGGT				1																1
ACTCGC/AGTGCG	1																			1
AGATGC/ATCTGC	1																			1
AGCTCC/AGCTGG	1																			1
AGGCGG/CCGCCT		1																		1

4.2.6.2 Detection of SNPs

The reads of guar varieties M-83 and RGC-1066 were mapped against the 62,146 sequences of assembled transcripts. A total of 53,402 putative SNPs (~1 SNP per transcript) were identified. The SNPs were screened by calculating the SNP density. A total of 8,416 SNPs were obtained with read depth of >5 in both verities. The results suggest that there is about one SNP in every 5.01 kb of leaf transcriptome in guar. High-confidence differences of 3,594 SNPs were obtained after filtering for homozygous allele types (Appendix- 3). The statistical analysis of SNP loci was done for each variety against the assembled transcripts. This resulted in 65.25 % transition nucleotide substitutions and 34.75 % transversions in guar variety M-83. In variety RGC 61.36 % transitions and 38.64 % transversions were found. The statistical information of SNPs in guar varieties M-83 and RGC-1066 against the reference is shown in Figure 4.2.6.3. In addition, 2,930 and 3,984 Insertion-Deletion (InDel) variations having read depth at least 10 were found in the varieties M-83 and RGC-1066, respectively.

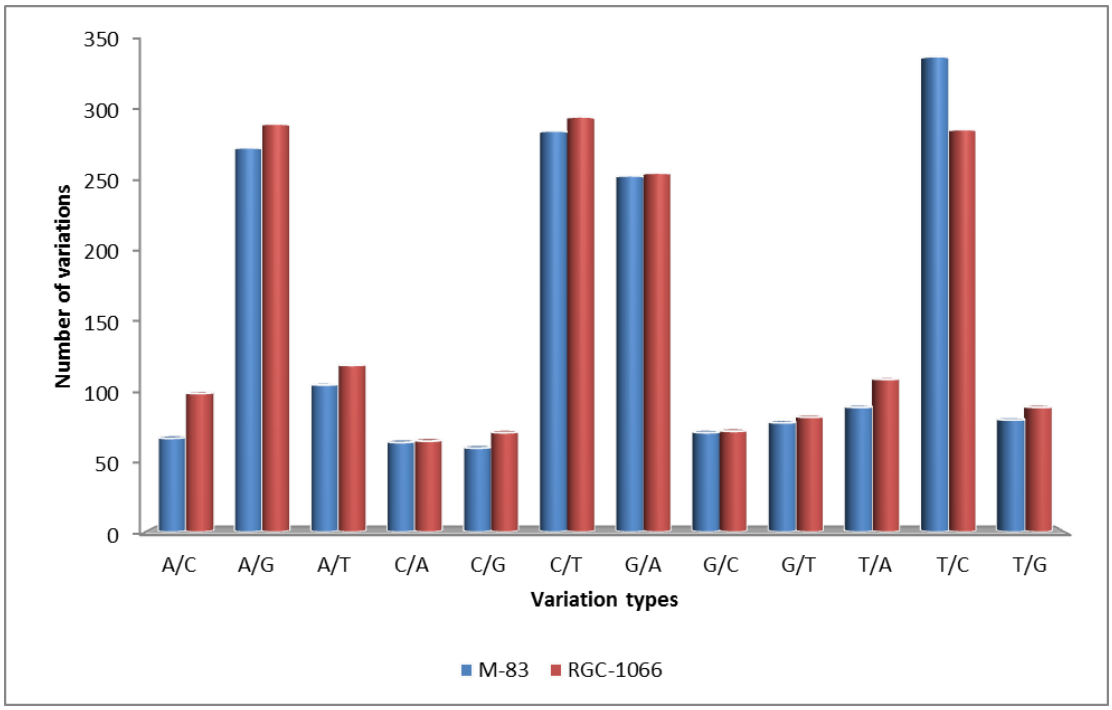


Figure 4.2.6.3 Statistical information of SNPs in guar varieties M-83 and RGC-1066 against *de novo* assembly.

4.3.6.3 Validation of SSR markers

A total of 20 primer pairs were designed and synthesized for randomly selected SSR markers (except mononucleotide repeats) of guar. The flanking primers were designed for SSR containing sequences using the online tool Primer3. Five primers for each dinucleotide, trinucleotide and tetranucleotide repeats, three primers for each pentanucleotide repeats and two primers for each hexanucleotide repeats, were designed and synthesized. The details of the transcriptome sequence ID, motif type and SSR length have been given in Table 4.2.6.2. Table 4.2.6.3 shows the names and sequences of the primers synthesized.

Table 4.2.6.2 Details of SSR markers used for validation in guar varieties M-83 and RGC-1066

ID	SSR No.	SSR Type	SSR	Size	Start	End
Dinucleotide						
comp5733_c0_seq1_len=241	1	p2	(GA)6	12	92	103
comp15714_c1_seq1_len=293	1	p2	(CT)8	16	206	221
comp23394_c0_seq1_len=298	1	p2	(CT)6	12	155	166
comp23819_c0_seq1_len=301	1	p2	(TC)6	12	175	186
comp1834_c0_seq1_len=232	1	p2	(AG)9	18	122	139
Trinuceotide						
comp19070_c0_seq1_len=655	1	p3	(GCC)5	15	407	421
comp23428_c0_seq1_len=1506	1	p3	(TGG)7	21	1386	1406
comp18850_c0_seq1_len=1876	1	p3	(TTC)6	18	434	451
comp17598_c0_seq1_len=432	1	p3	(TCT)5	15	197	211
comp19082_c0_seq1_len=464	1	p3	(CAT)5	15	302	316
Tetranucleotide						
comp23741_c0_seq1_len=2956	1	p4	(CTTT)5	20	236	255
comp50026_c0_seq1_len=304	1	p4	(ATAG)5	20	187	206
comp48573_c0_seq1_len=1530	1	p4	(TCAC)6	24	181	204
comp274511_c0_seq1_len=339	1	p4	(TGGT)9	36	222	257
comp33177_c0_seq1_len=5191	4	p4	(TATC)5	20	2934	2953
Pentanucleotide						
comp18299_c0_seq1_len=983	1	p5	(TTCTT)5	25	842	866

comp29744_c0_seq1_len=2424	1	p5	(AGAGA)6	30	2323	2352
comp33888_c0_seq1_len=582	2	p5	(TCTCA)5	25	472	496
Hexanucleotide						
comp148748_c0_seq1_len=369	1	p6	(AATTCA)5	30	60	89
comp27051_c0_seq1_len=1109	1	p6	(TGGAGC)5	30	696	725

Table 4.2.6.3 Details of primer pairs designed and synthesized for validation of SSR markers in guar varieties M-83 and RGC-1066

Sr. No.	ID	NAME	OLIGO	Start	Length	Tm	GC%	Sequence	Product Size
Dinucleotide									
1	comp5733	**GT-01**	FORWARD PRIMER	32	22	57.68	40.91	ACTTCATGGTG ATGGATTTGGA	210
			REVERSE PRIMER	241	23	57.63	43.48	CGCTCTTCCGA TCTAGAATTAG T	
2	comp15714	**GT-02**	FORWARD PRIMER	51	20	60.18	55	GGCAAGGAAA AGGGGTGAGT	229
			REVERSE PRIMER	279	20	59.58	55	AGAGAAGGAA GGAGGGAGCA	
3	comp23394	**GT-03**	FORWARD PRIMER	30	20	58.7	50	ACTGGGATCTG AATTGGGCT	231
			REVERSE PRIMER	260	20	60.18	60	TCCTCTGCTAG CCTAGTCCG	
4	comp23819	**GT-04**	FORWARD PRIMER	38	22	59.49	50	TCCACAGCCTC TTTCTTATCCC	200
			REVERSE PRIMER	237	20	58.22	60	GAGAGGGACA GGGAGAAGAG	
5	comp1834	**GT-05**	FORWARD PRIMER	3	23	58.98	43.48	ACGCTTAGATT AGTGGGTTCTC T	202
			REVERSE PRIMER	204	20	59.4	50	GGTTGCGTGCA TTTTCCTCT	
Trinucleotide									
6	comp19082	**GT-06**	FORWARD PRIMER	183	21	59.2	47.62	TGACTTTGTGA ATGCTACCGC	214
			REVERSE PRIMER	396	20	60.11	55	TGATGCTCTCA ATGCTGGGG	
7	comp19070	**GT-07**	FORWARD PRIMER	244	20	59.68	60	GGATGGATCG GAGGAAGACG	236
			REVERSE PRIMER	479	20	59.88	55	TTACCCCTACC CAGTGAGCA	
8	comp23428	**GT-08**	FORWARD PRIMER	1249	20	59.75	60	GGGAGCTGAA GACAAGAGGG	200
			REVERSE PRIMER	1448	20	58.95	50	TCGACCAACA ATGTCCCAGA	
9	comp18850	**GT-09**	FORWARD PRIMER	205	21	59.38	47.62	TGCGATCTGGG AGTTTCAAGA	454
			REVERSE PRIMER	658	20	59.97	55	CTTGCCCACCT TGAAACTGC	
10	comp17598	**GT-10**	FORWARD PRIMER	82	20	60.04	55	CTTCTCCGCGG TTTCTTCCT	310
			REVERSE PRIMER	391	20	59.62	50	GTCAACAGGT GCGTCGTTTT	
Tetranucleotide									
11	comp23741	**GT-11**	FORWARD PRIMER	129	20	60.04	60	GGACACCGGA GTAAACAGGG	380

			REVERSE PRIMER	508	20	59.82	60	GGCTTATCCTC CCACCCTTG	
12	comp50026	**GT-12**	FORWARD PRIMER	61	20	59.9	55	GATGCCCAAT GATGCACCAC	229
			REVERSE PRIMER	289	25	59.71	40	TCATAGCTTAG AACAAATCAC GCAG	
13	comp48573	**GT-13**	FORWARD PRIMER	157	20	60.32	60	CCAGCCACCA CACTCTTCTC	234
			REVERSE PRIMER	390	20	60.03	55	AAGGGCAGCT CTAGAGACGA	
14	comp274511	**GT-14**	FORWARD PRIMER	60	20	59.46	60	GTCCTCTGTCT TGGCTACCC	232
			REVERSE PRIMER	291	20	60.32	60	CTCCTTTACCA CCTTGCCCC	
15	comp33177	**GT-15**	FORWARD PRIMER	2568	20	60.03	60	TGGGATGGTG AGAGGAGAGG	451
			REVERSE PRIMER	3018	20	60.04	50	ATACGGCGGT GTTGGACATT	
Pentanucleotide									
16	comp18299	**GT-16**	FORWARD PRIMER	590	20	60.11	55	CCCCTGCACGA ATTGGATCT	314
			REVERSE PRIMER	903	20	59.97	55	GGAACGGCAA CACACTGAAC	
17	comp29744	**G-17**	FORWARD PRIMER	2113	20	59.97	50	AAATGGAAGC GTGGTTTGGC	261
			REVERSE PRIMER	2373	20	58.48	60	TCCTCTCCTCT CCTCTCCTC	
18	comp33888	**GT-18**	FORWARD PRIMER	139	20	59.82	60	CTCATGTCCCC TGAACTCGG	443
			REVERSE PRIMER	581	20	60.86	60	CACGACGCTCT TCCGATCTG	
Hexanucleotide									
19	comp148748	**GT-19**	FORWARD PRIMER	40	20	60.45	55	CTTCTCTTTGC GTCGCGTTG	255
			REVERSE PRIMER	294	20	60.04	55	ACGACGTTCCC TCCATCAAC	
20	comp27051	**GT-20**	FORWARD PRIMER	566	20	59.96	55	TGCCACCATTG TCAGGTCTC	377
			REVERSE PRIMER	942	20	60.18	55	TGGTCCTCTTC CTCAGCCTT	

The primers were tested for amplification and polymorphism on two guar varieties M-83 and RGC-1066. Out of the 20 primers tested 13 primers showed amplification in both varieties. The SSR primers GT-2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 18 resulted in amplification of the genomic DNA of both varieties. Figure 4.2.6.4 shows the banding pattern of SSR primers amplification on genomic DNA of guar. The SSR primers GT-16, 17 and 19 showed amplification only in the variety RGC-1066 and SSR primer GT-15 showed amplification on guar variety M-83. The SSR primer GT-17 showed amplification at higher size than the theoretical amplicon size. Some of the tested markers showed more than one band that might be due to the presence of multiple sites complementary to the primers in the genomic DNA. Only 65 % of the 20 tested SSR primers resulted in amplification in the target guar varieties M-83 and RGC-1066. However, no polymorphism was detected in the tested SSR primers.

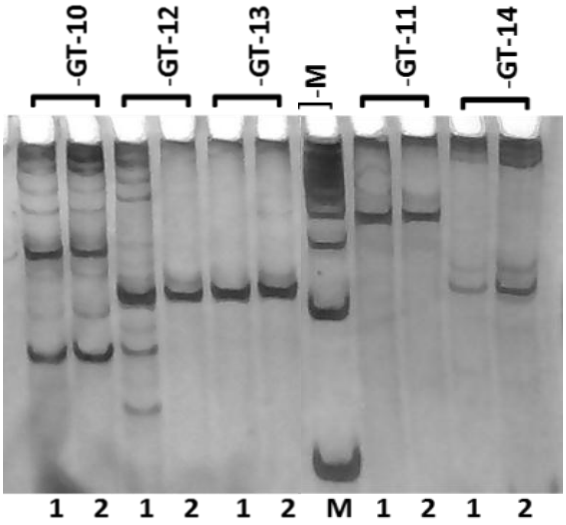


Figure 4.2.6.4 Polyacrylamide gel image of SSR primers (GT-10 to GT-14) amplification on genomic DNA of guar. M represents 100 bp marker, Lane 1 and 2 represents guar varieties M-83 and RGC-1066, respectively.

4.2.6.4 *In silico* analysis of SSR polymorphism

The reads of guar varieties M-83 and RGC-1066 were mapped against the 62,146 sequences of assembled transcripts to obtain the sorted alignment for each variety. The overall alignment rates were found to be 89.44 % and 91.69 % for M-83 and RGC-1066, respectively. Ten percent reads showed an alignment of >1 times for both varieties as shown in Table 4.2.6.4. The sorted alignments (BAM files) were further aligned against the reference using IGV 2.3 software [97] and observed manually to get the nucleotide differences surrounding the SSR region in both varieties. As a result, a total number of 145 SSRs were found to be polymorphic with 2 or more base differences between the guar varieties M-83 and RGC-1066. The details of *in silico* identified SSR markers have been presented in Appendix 2. Figures 4.2.6.5 and 4.2.6.6 show the instances of *in silico* identification of SSR polymorphism. The *in silico* identified polymorphic SSRs included 97 mononucleotide, 17 dinucleotide and 23 trinucleotide repeats (Figure 4.2.6.7).

Table 4.2.6.4 Alignment rates of leaf transcriptomes of guar varieties M-83 and RGC-1066

Alignment	RGC-1066	M-83
Aligned 0 times	8.31%	10.56%
Aligned 1 time	81.36%	78.63%
Aligned >1 times	10.32%	10.81%
Overall alignment rate	91.69%	89.44%

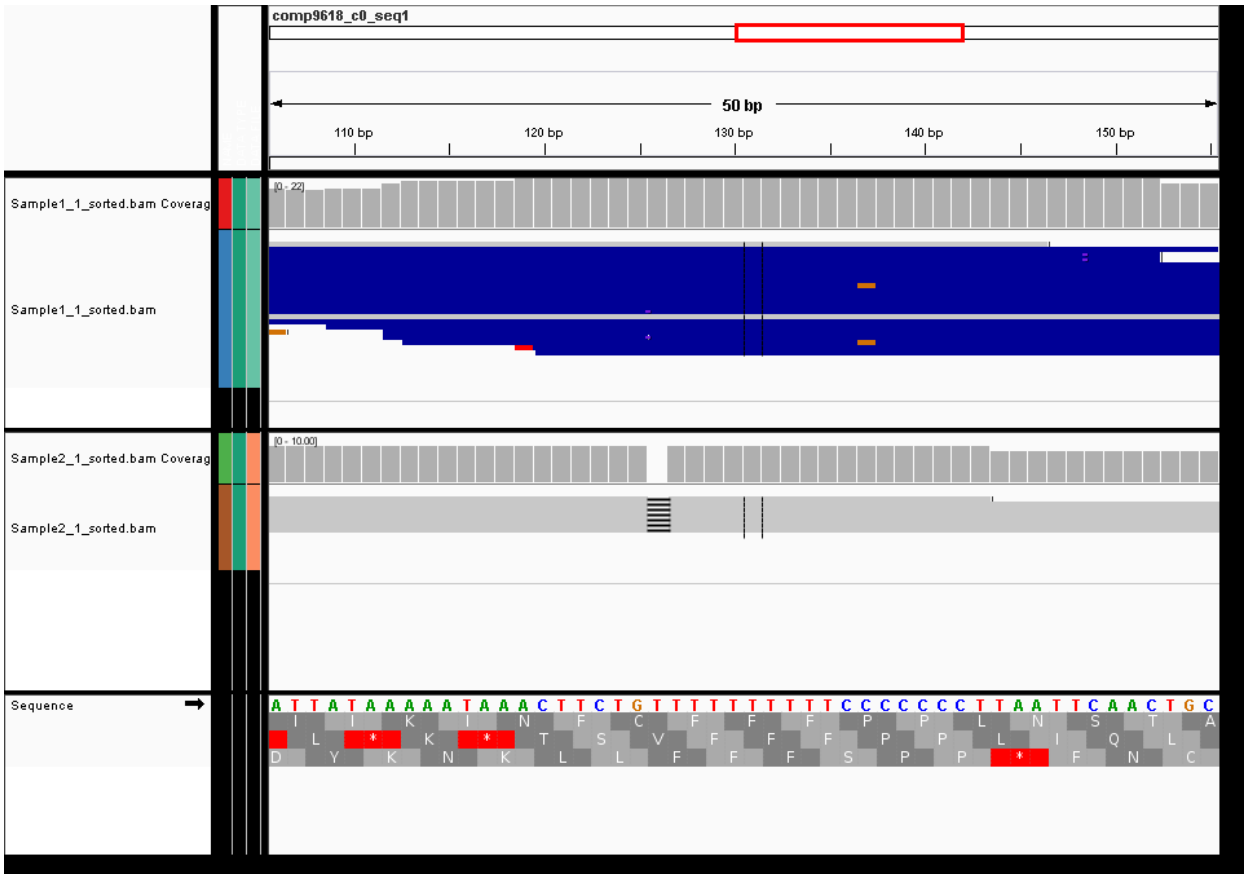


Figure 4.2.6.5 An instance of *in silico* identified SSR marker (comp9618_c0_seq1:106-155).

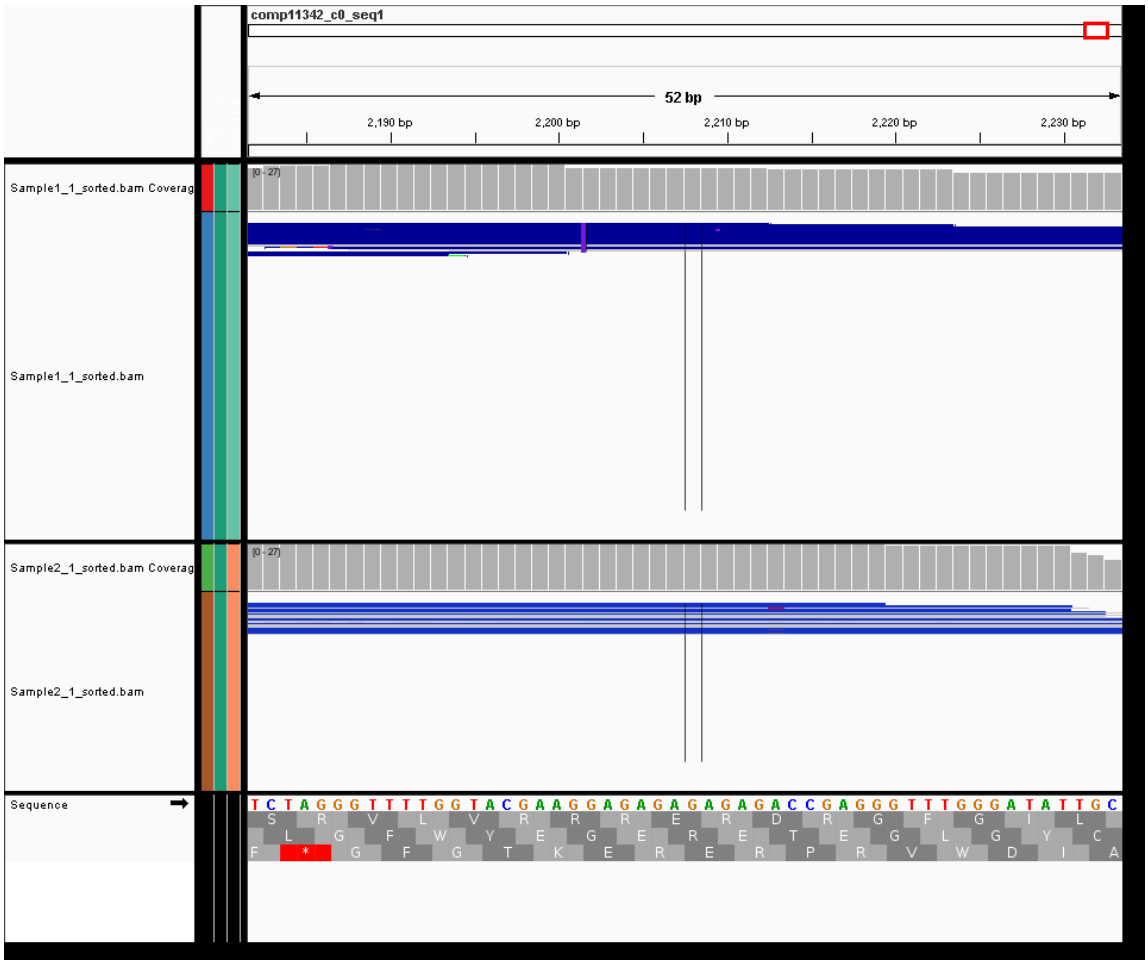


Figure 4.2.6.6 An instance of *in silico* identified SSR marker (comp11342_c0_seq1:2,182-2,233).

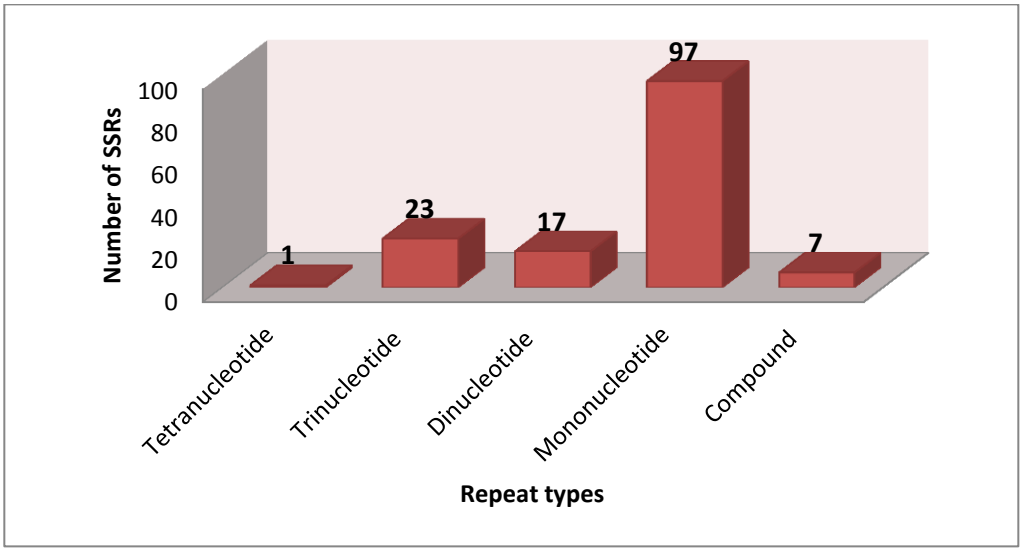


Figure 4.2.6.7 Distribution of *in silico* identified polymorphic SSR repeat types in guar leaf transcriptome.

4.3 Studies on guar gum based drug delivery system

4.3.1 Preparation of guar gum microparticles

Guar gum microparticles were prepared by two methods, namely, precipitation and cross linking method, and by emulcification and cross linking method. The drug loaded microparticles were named as GMF-1 and GMF-2 according to the method of preparation. The size, shape and distribution of guar gum microparticles were evaluated by using scanning electron microscopy (SEM) and atomic force microscopy (AFM). SEM images of guar gum microparticles have been presented in Figure 4.3.1.1. Figure 4.3.1.2 shows the AFM results of guar gum microparticles. The microparticles were found to be spherical in shape with a size range of 7-12 µm and 20-30 µm in GMF-1 and GMF-2, respectively.

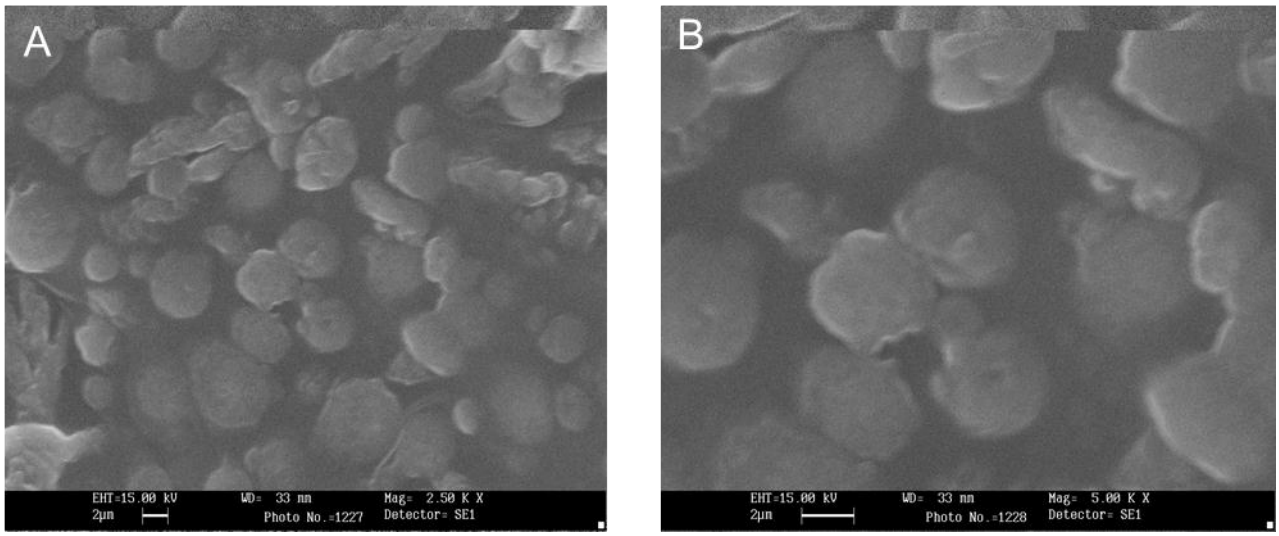
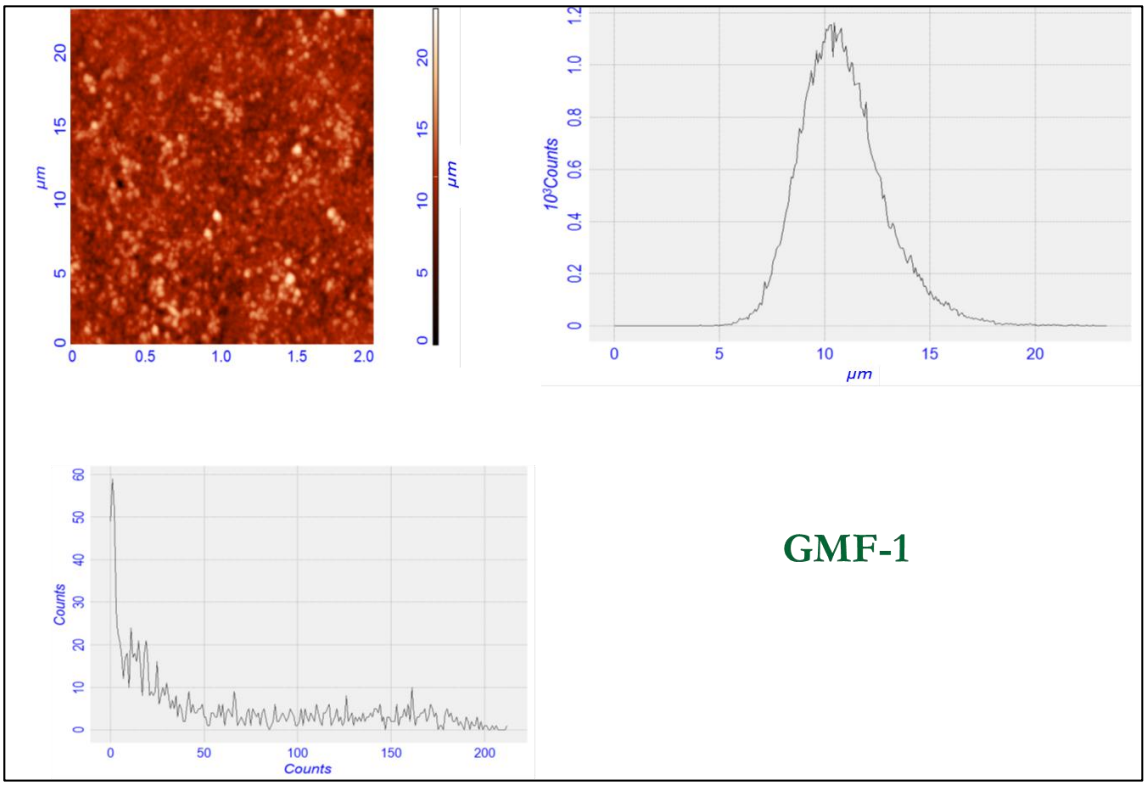


Figure 4.3.1.1 SEM images of guar gum microparticles. (A) at 2.5 KX and (B) at 5.0 KX.



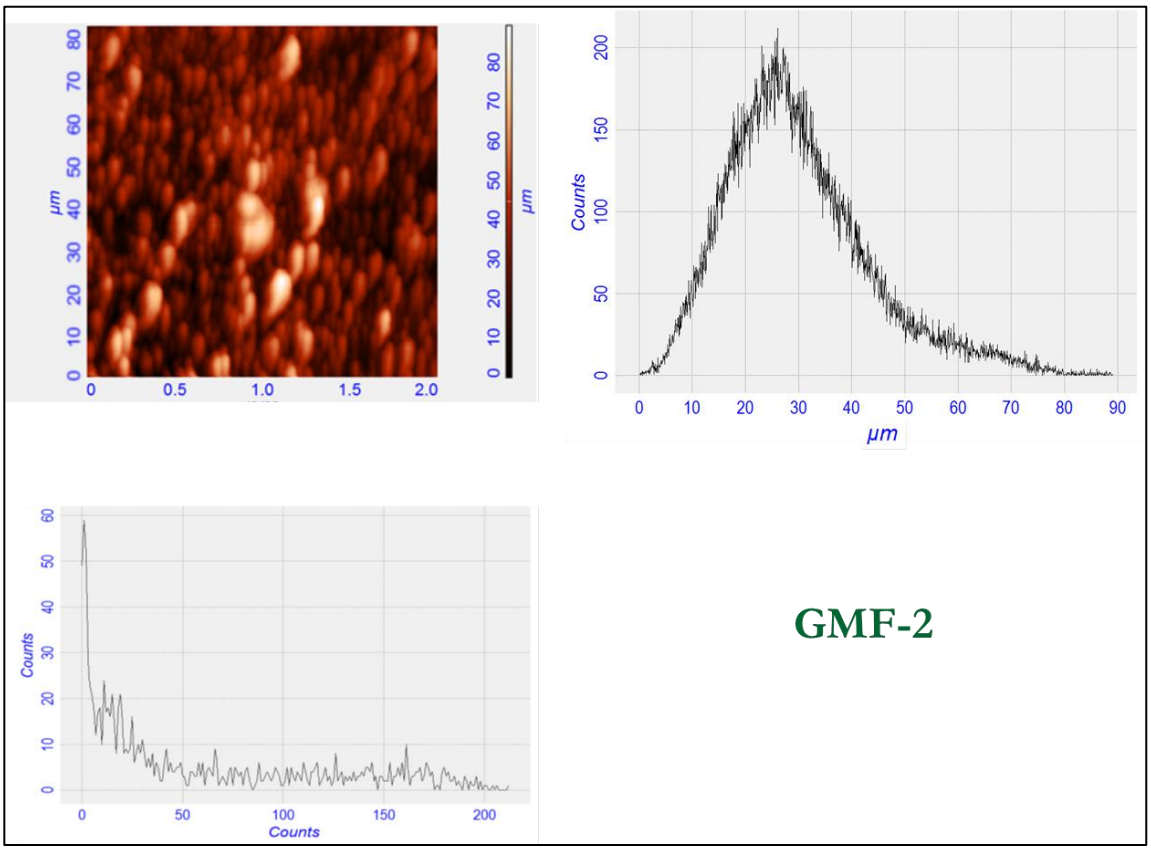


Figure 4.3.1.2 Atomic force microscopy results of guar gum microparticles.

4.3.2 Standard curve for 5-Flurouracil

The standard curve of 5-fluorouracil was prepared using phosphate buffer (pH 7.4). Figures 4.3.2.1 and 4.3.2.2 show the absorption spectra and standard curve of 5-fluorouracil, respectively.

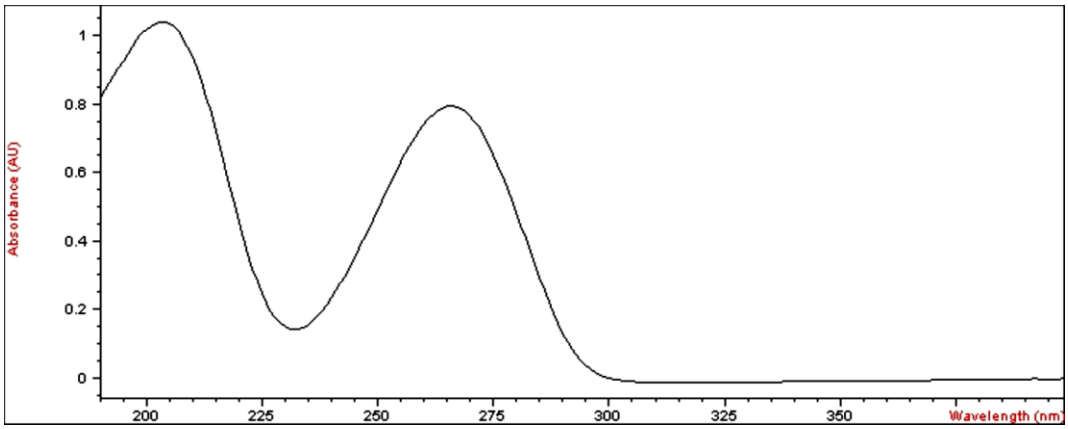


Figure 4.3.2.1 Absorption spectrum of 5-fluorouracil in phosphate buffer (pH 7.4).

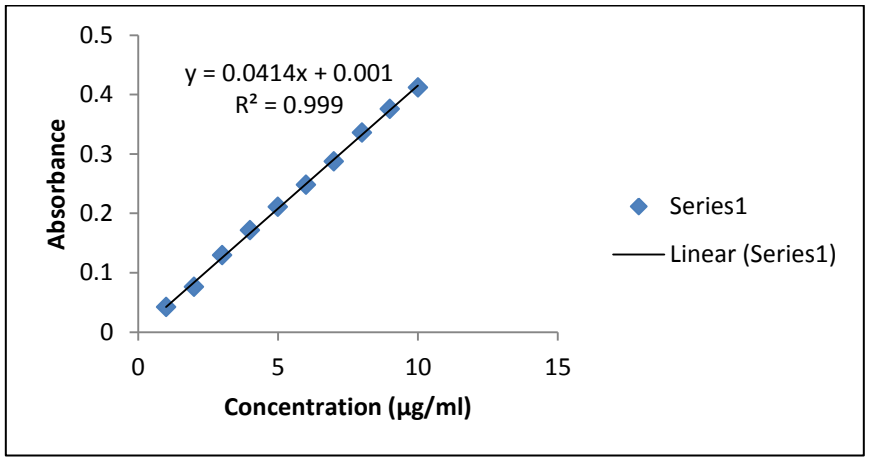


Figure 4.3.2.2 Calibration curve and equation calibration of 5-fluorouracil.

4.3.3 Drug release study from guar gum microparticles

The results of the drug release studies carried out on 5-fluorouracil loaded guar gum microparticles in phosphate buffer (pH 7.4) have been presented in Figure 4.3.3.1. The cumulative drug release from both the microparticle types was found to be very limited till 4 h and approximately 40 % cumulative drug release was observed after 6 h. After 7 h, a rapid increase cumulative drug release was observed. Approximately 80 % of drug was released after 12 h.

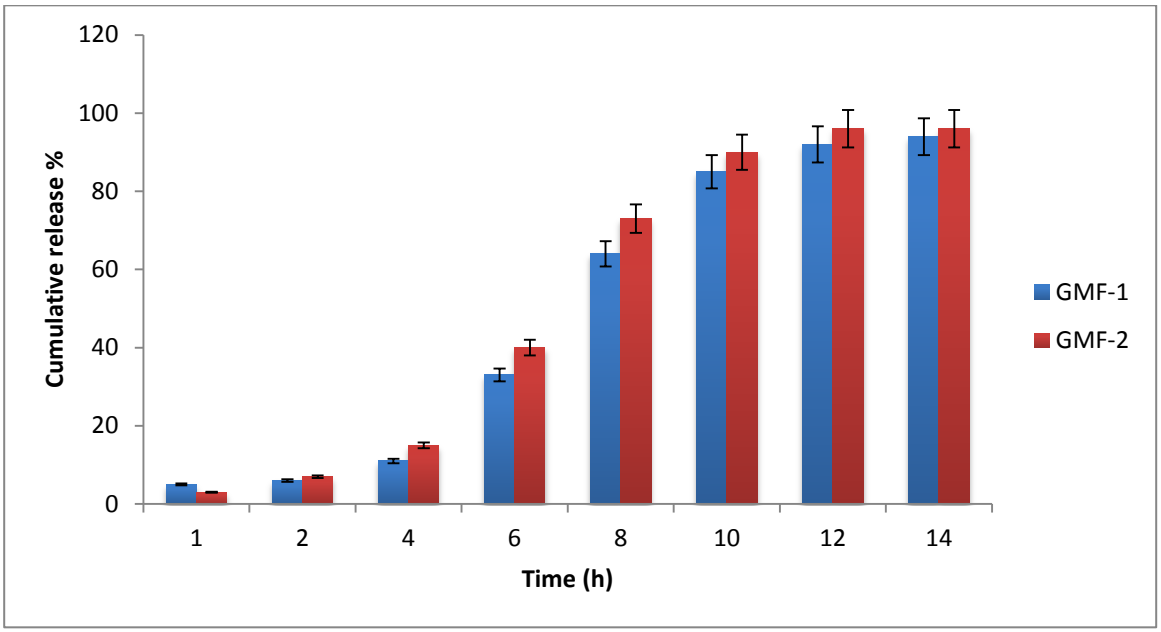


Figure 4.3.3.1 Cumulative release percentage of 5-FU from guar gum microparticles in phosphate buffer of pH 7.4.

4.3.3.1 Effect of glutaraldehyde concentration on drug release from guar gum microparticles

Figure 4.3.3.2 shows the effect of glutaraldehyde concentration on release of 5-fluorouracil from guar gum microparticles. The glutaraldehyde concentration was found to affect the rate and extent of drug release from guar gum microparticles. The drug release was found to be 63 % after 8 h with 0.5 ml glutaraldehyde in GMF-1. A significant decrease was observed with increase in the glutaraldehyde concentration. Approximately 20 % drug release was observed in GMF-1 after 8 h with 2 ml glutaraldehyde concentration. Similar drug release pattern was observed in GMF-2.

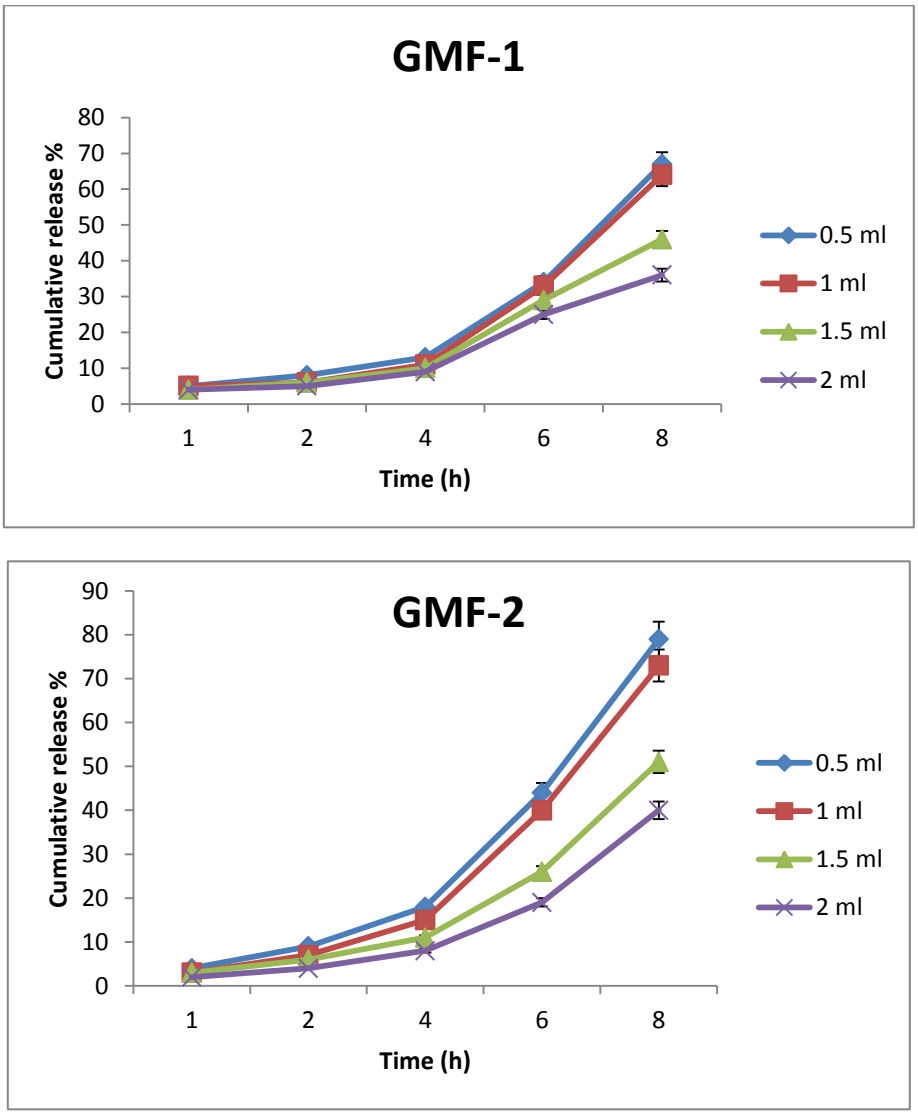


Figure 4.3.3.2 Effect of glutaraldehyde concentration on release of 5-fluorouracil from guar gum microparticles.

5. DISCUSSION

Initially the thesis research work was started on guar gum based drug delivery system, but there was some apprehension that the desired results may not come so the transcriptome analysis in guar was also started simultaneously. Finally the results of both aspects were retained in the thesis as per the suggestions of my SRC (student research committee).

All genetic improvement programs in guar have been carried out using conventional breeding without the involvement of molecular markers. As a result, only a limited success has been achieved in obtaining improved guar varieties. Marker-assisted breeding, especially with SSRs and SNPs, has given excellent results in several other crops [100, 124, 217]. Such breeding programs have not been possible in guar due to the lack of sufficient number of SSRs [143, 146] and the complete absence of SNPs. This has happened due to the limited availability of genetic resources in this crop. The NGS technology helps in building huge genetic resources for various organisms including non-model plants and provides novel opportunities in functional genomics and gene discovery [287]. This technology has been widely used for the development of molecular markers through transcriptome analysis in several plant species [72, 288, 289].

The present study on transcriptome sequencing by Illumina sequencing platform, for *de novo* transcriptome assembly, functional annotation and identification of genic markers was done on two guar varieties, namely, M-83 and RGC-1066. A total of 42,777,004 and 59,940,380 orientation-based high-quality sequence reads from the leaf tissue of both guar varieties were assembled to generate 62,146 unigene contigs which represented a large fraction of the guar transcriptome and helped in identification of a comprehensive set of genic-markers. The *de novo* assembly with shortest and longest unigenes sequence lengths of 201 and 29,056 bp, respectively, and an average length of 679 bp indicated the good coverage as well as the depth of sequencing data. Similar results were obtained in the *Phyllanthus amarus* leaf transcriptome analysis in which 14,608,389 high quality paired reads were assembled into 85,927 unitranscripts with maximum and minimum read lengths being 13,600bp and 200bp respectively, and the average unitranscript size 1548 bp [33]. In an another study of *Raphanus sativus* leaf transcriptome, a total of 70,879,904 clean reads generated a total of 68,086 unigenes with an average length of 576 bp and an N50 of 773 bp [298]. The CEGMA software was used for assessment of completeness of a transcriptome assembly by evaluating the presence and completeness of a widely conserved set of 248 CEGs. These CEGs represent the proteins that are mostly coded by

housekeeping genes and therefore can be expected to be expressed [186, 198, 199]. The CEGMA analysis results revealed that the assembly had 87.50 % of complete and 97.18 % partial CEGs. Similar results were obtained in *de novo* transcriptome assembly of *Nicotiana benthamiana* [186]. Hence, the *de novo* assembly obtained in this work is appropriate for the functional annotation and identification of genic markers.

As guar is a non-model plant and without any prior genome information, sequence similarity search and comparison for the assembled unigenes of guar leaf transcriptome were carried out by BLASTX against several databases. The total numbers of hits obtained in Uniref90 and Nr databases were 44,992 and 45,972, respectively. Among the 62,146 unigenes, 71.23 % had at least one significant match in blast hit results with an E-value less than $1e^{-6}$. The results indicated that most of the unigenes are protein coding genes. The unigenes showing no significant matches may be lacking known conserved functional domains or these may represent non-coding RNAs. Alternatively, these unigenes may contain a known protein domain but are too short to display sequence matches, resulting in false-negative results [298]. Moreover, the genomic and transcriptomic information is limited for guar; therefore many guar lineage specific genes might not be present in the database. The part of sequences showing no hits might be of great interest for further research for alternative splice variants, novel gene products and differentially expressed genes. The species distribution analysis revealed that the leaf tissue in guar has a number of homologous sequences in many plant species. Among the various plant species *Glycine max* genes have the highest similarity (41.91 %) with guar unigenes. The comparison of the guar leaf transcriptome was done with the sequences of closely related sequenced species using TRAPID tool. The results suggested that the genome of *Glycine max* can be used as a reference for the transcriptome analysis in guar.

The gene ontology (GO) database is an important resource as GO terms provide a set of dynamically controlled and structured vocabularies for describing the roles of genes in any organism [16]. Based on sequence homology, all the guar leaf transcriptome unigenes were assigned with GO terms. The annotated GO terms were found to be distributed into 46 functional groups, which were further classified under the three main categories, namely, biological process, molecular function and cellular component. Within the biological process, “metabolic process, “cellular process”, “single-organism process” and “biological regulation” were the top GO terms. Among the molecular function category, “catalytic activity”, “binding” and “transporter activity” were major GO terms. In the cellular component, “cell”, “membrane”, “organelle” and

"macromolecular complex" were mainly enriched. Only a few unigenes were classified in terms of "cell killing", "behavior", "protein tag", "translation regulator activity", "nutrient reservoir activity" and "extracellular matrix". These results are consistent with those from the other plant leaf transcriptome studies [33, 298]. By searching against the available database, a total of 11,308 guar unigenes were annotated with various enzyme codes. The maximum numbers of unigenes were assigned to tranferases followed by hydrolases, oxidoreductases, lagases, lyases and isomerases. As enzymes are a type of large biological molecules responsible for thousands of metabolic processes that sustain life, the identification of important enzyme codes might provide clues to reveal some important functional pathways and metabolic activities of leaf tissues in guar [298].

The biological pathway studies play a key role in gaining insight into the advanced studies of genomics. KEGG is a highly integrated database providing information of the biological systems and their relationships at the molecular, cellular and organism levels particularly via the KEGG pathway maps [120]. Therefore, a systematic analysis of high-level gene function was performed to identify the biochemical pathways by assigning guar leaf unigenes in the KEGG database. As a result, a total of 145 KEGG maps (pathways) were found to be associated with 11,971 unigenes. Enzyme codes were used to represent the putatively identified genes involved in several metabolic pathways. It was observed that more than one unigene in the guar transcriptome dataset was annotated as the same enzyme. Bose Mazumdar and Chattopadhyay also found the similar pattern in *P. amarus* leaf transcriptome [33]. In guar leaf transcriptome, maximum number of unigenes fell under metabolism pathways. Among various metabolic pathways lipid metabolism is associated with plants [165]. The leaf specific pathways included nitrogen metabolism, carbon fixation in photosynthetic organisms, photosynthesis and glycolysis/gluconeogesis. Nitrogen metabolism, glycolysis/gluconeogenesis and carbon fixation in photosynthetic organisms are all associated with photosynthesis [298]. These results indicated that these leaf-specific unigenes are mainly involved in the photosynthesis pathway. Further, the differential gene expression in the guar varieties M-83 and RGC-1066 was studied using *de novo* assembly of leaf transcriptome. Both the varieties showed ~80 % similar gene expression in leaf transcriptome. A total of 175 unigenes were found to be overexpressed, with at least 30 folds overexpression, in guar M-83 leaf transcriptome. These unigenes were further annotated against KEGG database and 36 KEGG maps with 49 enzyme codes were found. A total of 158 unigenes were found in RGC-1066 variety with overexpression of 20 folds and

only two KEGG maps with five EC were annotated. These results are expected to be helpful for functional genomics studies to identify the genes involved for leaf pubescence in these two guar varieties.

One main goal in this study was to identify molecular markers that can be readily used in breeding programs. Among various molecular markers, SSRs and SNPs are the most useful markers for genetics and plant breeding applications [100]. In the present study, two sets of molecular markers, SSR and SNP were identified using the transcriptome dataset of guar leaves for genetic and breeding analyses in guar crop. Transcriptome based markers are gene based markers which are advantageous as compared to the genomic markers [21]. Gene based markers or genic markers are becoming more popular as compared to traditional random genomic markers due to rapid and inexpensive method of isolation and their cross species portability [110]. Therefore, interest has been shifted from genomic to genic markers owing to their high inter-species transferability as they are developed from conserved coding regions of the genome [226]. In this study, a total of 5,773 potential SSRs were identified with an average of one SSR per 7.31 kb in the unigenes. This result was consistent with the previous EST-SSR report in guar with occurrence (kb/SSR) of 7.9 [143], however, Kuravadi *et al*. reported the occurrence of 4.1 using the same dataset [146]. The occurrence of genic-SSR was also comparable to 8.4 in pigeonpea, 3.4 in rice, 5.4 in wheat and 7.4 in soybean [43, 72, 207]. The differences in genic-SSR abundance may be due to the size of EST or unigene assembly dataset, and different data mining tools and criteria used [275]. A comprehensive analysis was performed to describe the type, frequency and distribution of all the potential SSRs. The frequency distribution of SSR markers is in agreement of previous reports in guar [143, 146]. If the mononucleotide SSRs are excluded because of the frequent homopolymer errors found in sequencing data, a large proportion was covered by di- and trinucleotides (96 %) while the rest amounted to less than 4 %. This is consistent with the EST-SSRs distributions reported in many legumes [289]. A similar trend was observed in other plant species [5, 251]. The trinucleotide repeats were found to be the maximum which are more frequently detected in coding regions [308]. The possible reason for abundance in trinucleotide motifs may be due to expansion or contraction of di-nucleotide repeat length in exons to suppress deleterious effects of the frame-shift mutation in translated regions [299]. These repeats are generally more robust since they are reported to give fewer “stutter bands” than the dinucleotide repeats. The trinucleotide repeats have been reported as highly polymorphic and stably inherited [303]. The 5,773 potential SSRs identified from *de novo*

transcriptome sequencing data of guar leaf represent a significant addition to the limited set of genic-SSR markers available in guar.

A total of 20 SSR markers (representing all the motif repeat types) were randomly selected for validation by wet laboratory analysis. The SSR markers validation results showed that 13 of the 20 tested SSR primers resulted amplification in the target guar varieties M-83 and RGC-1066. The lack of amplification could be because of the flanking primers extending across a splice site with a large intron or chimeric cDNA contigs [276]. Some of the tested markers showed more than one band that might be due to the presence of multiple sites complementary to the primers in the genomic DNA. None of the tested markers showed distinct polymorphism. The possible reason may be due to the small product size difference or actual lack of polymorphism as earlier reported in pigeonpea [72]. Overall 65 % of the tested SSRs were validated successfully by wet laboratory analysis. These results are consistent with barley, where 67-70 % of the primers showed amplification [264, 277]. The amplification success rate was higher than that reported in sugarcane (48 %) and lower than flax (92 %) [56, 58]. *In silico* polymorphism analysis of the SSR markers was done by IGV software [265]. A total of 145 out of 5,773 SSR markers were identified as *in silico* polymorphic in the guar varieties M-83 and RGC-1066. This result is in agreement with the reports in pigeonpea [72]. Taken together with the previous SSR polymorphism studies, it can be concluded that genetic diversity in the guar gene pool is very low [143, 146].

The detection of SNPs resulted in a total number of 53,402 putative SNPs (~1 SNP per transcript) in the guar varieties M-83 and RGC-1066. The putative SNPs were screened for a minimum depth of 5 reads with same homozygous allele and set to be at least 50 bp apart from adjacent SNPs. The screening process might have reduced the sensitivity in detecting rare SNPs, but the specificity of true SNP detection is increased due to the reduced chances of false variants inclusion that arise by sequencing errors. High-confidence differences were composed of 3,594 SNPs after screening for the SNP density. Statistical analysis of SNP loci resulted in 65.25 % transition nucleotide substitutions and 34.75 % transversions in guar variety M-83. In RGC-1066 variety, 61.36 % transitions and 38.64 % transversions were found. This finding is in agreement with red pepper transcriptome profiling [156]. These results are in accordance of transition/transversion rate bias. In virtually all DNA sequences, from any genome examined, transitions (T↔C and A↔G) have been found to occur at higher frequencies than transversions

or all other changes [93, 282, 283, 304]. Detection of transition/transversion rate bias is important to understand the patterns of DNA sequence evolution and phylogeny reconstruction [305]. As a result of various sequencing projects in plants, large number of SNPs has been reported. SNPs are genetic markers which are bi-allelic in nature, besides being highly abundant and less prone to mutations as compared to SSRs. They can contribute directly to a phenotype or can be associated with a phenotype as a result of linkage disequilibrium [191]. In plants, SNPs are particularly useful in the construction of high resolution genetic maps, the positional cloning of target loci, marker assisted breeding of important genes, genome wide large-scale linkage disequilibrium associate analysis, DNA fingerprinting, and species origin, relationship and evolutionary studies [241]. Most conventional molecular markers, such as restriction fragment length polymorphism (RFLP) and cleaved amplified polymorphic sequence (CAPS), are based on SNPs, i.e., nucleotide substitutions or insertions/deletions [190]. Many of the SNP assays rely on SNP genotyping by sequencing of the PCR amplified products that requires highly sophisticated instrumentation. These specialized instruments are very potent but come at prices which are prohibitive for most laboratories. Moreover, many assays require expensive probes or other reagents that escalate genotyping costs beyond that which is affordable [241]. With the influx of various SNP genotyping assays in recent years, there has been a need for an assay that is not only robust, but also cost effective, simple and highly accurate. The available SNP genotyping methods can be classified into non-gel and gel based detection systems. Most of the non-gel based systems, such as SnaPshot, pyrosequencing and biplex invader, are based on known sequence information, and tend to require a relatively large initial automated investment [205]. In contrast, the gel based methods are relatively low in cost and moderate in throughput. CAPS markers are the most commonly used markers in the gel based methods [264]. The existence of a restriction site difference spanning the SNPs between varieties/lines to be analyzed is essential for converting SNPs to CAPS markers. However, Michaels and Amasino [169] and Neff *et al*. [191] demonstrated that single-base changes generating no restriction site difference could be employed for the development of PCR-based markers by the derived CAPS (dCAPS) method. In this method, a restriction enzyme recognition site which includes the SNP is introduced into the PCR product by a primer containing one or more mismatches to template DNA [241]. The PCR product modified in this manner is then subjected to restriction enzyme digestion, and the presence or absence of the SNP is determined by the resulting restriction pattern. Like the CAPS markers, the dCAPS markers are simple and relatively inexpensive to identify [191].

Guar gum is widely used in many dosage forms for controlled drug delivery, however a little information is available in literature on the guar gum based multiparticulate drug delivery system. In this study, guar gum microparticles were prepared by two methods, namely, precipitation and cross linking method, and by emulsification and cross linking method. The 5-fluorouracil, an anticancer drug, was used as a model drug in this study. The microparticles as observed by scanning electron microscopy (SEM) and atomic force microscopy (AFM) were found to be spherical in shape with a size range of 7-12 µm and 20-25 µm. Hardening of microparticles was performed by chemical cross-linking with glutaraldehyde. The effect of various process variables such as stirring speed, glutaraldehyde concentration and temperature was studied in order to optimize the formulation. The drug release studies were carried out in PBS (pH 7.4). The drug release was found to be enhanced after 6 h in both types of microparticles. Similar results were obtained in previous studies using guar gum for coating of drug in tablet formulation [138, 248]. This controlled drug release might be due to the hydration of gum after exposure to dissolution fluids which forms a viscous gel layer that slows down further seeping-in of dissolution fluids towards drug. After guar gum swells up the drug release takes place by diffusion process. The 5-fluorouracil release from cross-linked microspheres was found to be glutaraldehyde concentration dependent. The drug release profiles from both guar gum microspheres clearly indicate that glutaraldehyde slows the drug release from microspheres. This result is consistent with the previous study on guar gum microspheres using methotrexate drug by Chaurasia *et al.* [48]. Glutaraldehyde causes cross-linking by reacting with the hydroxyl group of galactose and the mannose unit of guar gum, thus interfering with the free access of water to the hydroxyl group of guar gum. This significantly reduces the swelling rate of the microparticles and consequently the penetration of the solvent into the microparticles. Cross-linking also reduces polymer chain mobility, increases glass transition temperature, and decreases diffusivity [63, 224]. Based on these results it may be suggested that guar gum microparticles can be used for carrying the potent chemotherapeutic agents specifically to the site of action in case of colon cancer.

This study is the first report on the SNPs detection and transcriptome analysis in guar crop. The large number of SSRs and SNPs identified in this study provide a wealth of potential markers in this crop. These results open up new opportunities for population genetics, linkage mapping, comparative genomics and marker-assisted breeding in guar.

6. REFERENCES

1. Yadav, H. and Shalendra. An analysis of performance of guar crop in india. In, Report prepared by NIAM, Jaipur for United States Department of Agriculture (USDA), New Delhi (2014).

2. Agarwal, M., Shrivastava, N. and Padh, H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27(4):617-631 (2008).

3. Aguiar, B., Vieira, J., Cunha, A. E. and Vieira, C. P. No evidence for fabaceae gametophytic self-incompatibility being determined by Rosaceae, Solanaceae, and Plantaginaceae *S-RNase* lineage genes. *BMC Plant Biology* 15(1):129 (2015).

4. Ahmad, S., Singh, M., Lamb-Palmer, N. D., Lefsrud, M. and Singh, J. Assessment of genetic diversity in 35 *Pisum sativum* accessions using microsatellite markers. *Canadian Journal of Plant Science* 92(6):1075-1081 (2012).

5. Ahn, Y. K., Tripathi, S., Cho, Y. I., Kim, J. H., Lee, H. E., Kim, D. S., Woo, J. G. and Cho, M. C. *De novo* transcriptome assembly and novel microsatellite marker information in *Capsicum annuum* varieties Saengryeg 211 and Saengryeg 213. *Botanical Studies* 54(1):1-10 (2013).

6. Ajmone Marsan, P., Castiglioni, P., Fusari, F., Kuiper, M. and Motto, M. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. *Theoretical and Applied Genetics* 96(2):219-227 (1998).

7. Akimoto, M., Shimamoto, Y. and Morishima, H. Population genetic structure of wild rice *Oryza glumaepatula* distributed in the amazon flood area influenced by its life-history traits. *Molecular Ecology* 7(10):1371-1381 (1998).

8. Al-Saidan, S. M., Krishnaiah, Y. S. R., Patro, S. and Satyanaryana, V. *In vitro* and *in vivo* evaluation of guar gum matrix tablets for oral controlled release of water-soluble diltiazem hydrochloride. *AAPS PharmSciTech* 6(1):E14-E21 (2005).

9. Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., Giuliano, G., Chiusano, M. L., Baldoni, L. and Perrotta, G. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* 10(1):399 (2009).

10. Ali, M. L., Rajewski, J. F., Baenziger, P. S., Gill, K. S., Eskridge, K. M. and Dweikat, I. Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers. *Molecular Breeding* 21(4):497-509 (2008).

11. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25(17):3389-3402 (1997).

12. Anderson, E. Endosperm mucilages of legumes. *Industrial & Engineering Chemistry* 41(12):2887-2890 (1949).*

13. Andrews, S. FastQC: A quality control tool for high throughput sequence data. *Reference Source* http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).

14. Arora, R. N. and Pahuja, S. K. Mutagenesis in guar [*Cyamopsis tetragonoloba* (L.) Taub.]. *Plant Mutation Reports* 2(1) (2008).

15. Asghar, L. F. A. and Chandran, S. Multiparticulate formulation approach to colon specific drug delivery: Current perspectives. *Journal of Pharmacy and Pharmaceutical Sciences* 9(3):327-338 (2006).

16. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S. and Eppig, J. T. Gene ontology: Tool for the unification of biology. *Nature Genetics* 25(1):25-29 (2000).

17. Ashford, M., Fell, J., Attwood, D., Sharma, H. and Woodhead, P. Studies on pectin formulations for colonic drug delivery. *Journal of Controlled Release* 30(3):225-232 (1994).

18. Azam, S., Rathore, A., Shah, T. M., Telluri, M., Amindala, B., Ruperao, P., Katta, M. A. and Varshney, R. K. An integrated SNP mining and utilization (ISMU) pipeline for next generation sequencing data. *Plos One* 10:1371 (2014).

19. Bainbridge, M. N., Warren, R. L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M. and Magrini, V. Analysis of the prostate cancer cell line lncap transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* 7(1):246 (2006).

20. Barakat, A., DiLoreto, D. S., Zhang, Y., Smith, C., Baier, K., Powell, W. A., Wheeler, N., Sederoff, R. and Carlson, J. E. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biology* 9(1):51 (2009).

21. Barbara, T., Palma-Silva, C., Paggi, G. M., Bered, F., Fay, M. F. and Lexer, C. Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology* 16(18):3759-3767 (2007).

22. Barbazuk, W. B., Emrich, S. J., Chen, H. D., Li, L. and Schnable, P. S. SNP discovery via 454 transcriptome sequencing. *The Plant Journal* 51(5):910-918 (2007).

23. Barbier, P. Genetic variation and ecotypic diferentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. *The Japanese Journal of Genetics* 64(4):273-285 (1989).

24. Bassam, B. J., Caetano-Anolles, G. and Gresshoff, P. M. DNA amplification fingerprinting of bacteria. *Applied Microbiology and Biotechnology* 38(1):70-76 (1992).

25. Batley, J., Barker, G., O'Sullivan, H., Edwards, K. J. and Edwards, D. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiology* 132(1):84-91 (2003).

26. Bayliss, C. E. and Houston, A. P. Degradation of guar gum by faecal bacteria. *Applied and Environmental Microbiology* 48:626-632 (1986).

27. Berger, B., Peng, J. and Singh, M. Computational solutions for OMICS data. *Nature Reviews Genetics* 14(5):333-346 (2013).

28. Bernardo, R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science* 48(5):1649-1664 (2008).

29. Bhullar, N. K., Mackay, M. and Keller, B. Genetic diversity of the *Pm3* powdery mildew resistance alleles in wheat gene bank accessions as assessed by molecular markers. *Diversity* 2(5):768-786 (2010).

30. Bocianowski, J. and Seidler-Lozykowska, K. The relationship between RAPD markers and quantitative traits of caraway (*Carum carvi* L.). *Industrial Crops and Products* 36(1):135-139 (2011).

31. Bornet, B., Goraguer, F., Joly, G. and Branchard, M. Genetic diversity in european and Argentinian cultivated potatoes (*Solanum tuberosum* subsp. Tuberosum) detected by inter-simple sequence repeats (ISSRS). *Genome* 45(3):481-484 (2002).

32. Bornet, B., Muller, C., Paulus, F. and Branchard, M. Highly informative nature of inter simple sequence repeat (ISSR) sequences amplified using tri-and tetra-nucleotide primers from DNA of cauliflower (*Brassica oleracea* var. Botrytis L.). *Genome* 45(5):890-896 (2002).

33. Bose Mazumdar, A. and Chattopadhyay, S. Sequencing, *de novo* assembly, functional annotation and analysis of *Phyllanthus amarus* leaf transcriptome using the Illumina platform. *Frontiers in Plant Science* 6:1199 (2016).

34. Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32(3):314 (1980).

35. Bourgis, F., Kilaru, A., Cao, X., Ngando-Ebongue, G.-F., Drira, N., Ohlrogge, J. B. and Arondel, V. Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proceedings of the National Academy of Sciences* 108(30):12527-12532 (2011).

36. Breakfield, N. W., Corcoran, D. L., Petricka, J. J., Shen, J., Sae-Seaw, J., Rubio-Somoza, I., Weigel, D., Ohler, U. and Benfey, P. N. High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in arabidopsis. *Genome Research* 22(1):163-176 (2012).

37. Butt, M. S., Shahzadi, N., Sharif, M. K. and Nasir, M. Guar gum: A miracle therapy for hypercholesterolemia, hyperglycemia and obesity. *Critical Reviews in Food Science and Nutrition* 47(4):389-396 (2007).

38. Caetano-Anolles, G. and Bassam, B. J. DNA amplification fingerprinting using arbitrary oligonucleotide primers. *Applied Biochemistry and Biotechnology* 42(2):189-200 (1993).

39. Caetano-Anolles, G. and Brant, B. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Nature Biotechnology* 9(6):553-557 (1991).

40. Calviao, M., Bruggmann, R. m. and Messing, J. Characterization of the small RNA component of the transcriptome from grain and sweet sorghum stems. *BMC Genomics* 12(1):356 (2011).

41. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. Blast+: Architecture and applications. *BMC Bioinformatics* 10(1):421 (2009).

42. Cappa, O. and Moulines, E. On-line expectation-maximization algorithm for latent data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71(3):593-613 (2009).

43. Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D. and Waugh, R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156(2):847-854 (2000).

44. Carlos de Oliveira, A., Novac Garcia, A., Cristofani, M. and Machado, M. A. Identification of citrus hybrids through the combination of leaf apex morphology and SSR markers. *Euphytica* 128(3):397-403 (2002).

45. Cerny, T. A., Caetano-Anolles, G., Trigiano, R. N. and Starman, T. W. Molecular phylogeny and DNA amplification fingerprinting of *Petunia* taxa. *TAG Theoretical and Applied Genetics* 92(8):1009-1016 (1996).

46. Chaudhary, B. S. and Lodhi, G. P. Studies on the inheritance of five qualitative characteristics in clusterbean (*Cyamopsis tetragonoloba* (L.) Taub). *Euphytica* 30(1):161-165 (1981).*

47. Chaudhary, B. S., Paroda, R. S. and Solanki, K. R. New crossing technique in clusterbean (*Cyamopsis tetragonoloba* (L.) Taub.). *Current Science* (1974).*

48. Chaurasia, M., Chourasia, M. K., Jain, N. K., Jain, A., Soni, V., Gupta, Y. and Jain, S. K. Cross-linked guar gum microspheres: A viable approach for improved delivery of anticancer drugs for the treatment of colorectal cancer. *AAPS PharmSciTech* 7(3):143-151 (2006).

49. Chen, C., Farmer, A. D., Langley, R. J., Mudge, J., Crow, J. A., May, G. D., Huntley, J., Smith, A. G. and Retzel, E. F. Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated arabidopsis male meiocytes. *BMC Plant Biology* 10(1):280 (2010).

50. Chen, S., McElroy, J. S., Dane, F. and Peatman, E. Optimizing transcriptome assemblies for leaf and seedling by combining multiple assemblies from three *de novo* assemblers. *The Plant Genome* 8(1):1-10 (2015).

51. Chen, Y., Mao, Y., Liu, H., Yu, F., Li, S. and Yin, T. M. Transcriptome analysis of differentially expressed genes relevant to variegation in peach flowers. *Plos One* 9(6):e90842 (2014).

52. Cheng, Y. and Prud'homme, R. K. Enzymatic degradation of guar and substituted guar galactomannans. *Biomacromolecules* 1(4):782-788 (2000).

53. Choi, H.K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J.H., Kalo, P., Penmetsa, R. V., Seres, A., Kulikova, O., Roe, B., Bisseling, T., Kiss, G. B. and Cook, D. R. A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* 166(3):1463-1502 (2004).

54. Chourasia, M. K. and Jain, S. K. Pharmaceutical approaches to colon targeted drug delivery systems. *Journal of Pharmacy and Pharmaceutical Sciences* 6(1):33-66 (2003).

55. Chudzikowski, R. J. Guar gum and its applications. *Journal of the Society of Cosmetic Chemists* 22:43-60 (1971).

56. Cloutier, S., Niu, Z., Datla, R. and Duguid, S. Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). *Theoretical and Applied Genetics* 119(1):53-63 (2009).

57. Conesa, A., Göt, S., Juan Miguel García-Gómez, J. M., Terol, J., Talón, M. and Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674-3676 (2005).

58. Cordeiro, G. M., Casu, R., McIntyre, C. L., Manners, J. M. and Henry, R. J. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Erianthus* and *Sorghum*. *Plant Science* 160(6):1115-1123 (2001).

59. Crawford, D. J. Plant molecular systematics: Macromolecular approaches. New York: John Wiley & Sons (1990).

60. Cunha, P. L. R., de Paula, R. C. M. and Feitosa, J. P. A. Purification of guar gum for biological applications. *International Journal of Biological Macromolecules* 41(3):324-331 (2007).

61. Dabas, B. S., Mandal, S., Phogat, B. S., Bisht, I. S. and Agrawal, R. C. Guar (*Cyamopsis tetragonoloba*) - a resume of research at NBPGR. New Delhi: National Bureau of Plant Genetic Resources (2001).

62. Davidson, R. M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S. H., Jiang, N. and Robin Buell, C. Comparative transcriptomics of three poaceae species reveals patterns of gene expression evolution. *The Plant Journal* 71(3):492-502 (2012).

63. Deasy, P. B. Microencapsulation and related drug processes: Marcel Dekker Incorporated (1984).

64. Der, J. P., Barker, M. S., Wickett, N. J. and Wolf, P. G. *De novo* characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* 12(1):99 (2011).

65. Desgagn-Penix, I., Farrow, S. C., Cram, D., Nowak, J. and Facchini, P. J. Integration of deep transcript and targeted metabolite profiles for eight cultivars of *Opium poppy*. *Plant Molecular Biology* 79(3):295-313 (2012).

66. Dhillon, B. S., Varaprasad, S., Singh, M., Archak, S., Srivastava, U. and Sharma, G. D. National bureau of plant genetic resources: A compendium of achievements. *National Bureau of Plant Genetic Resources, New Delhi* (2001).

67. Dhugga, K. S., Barreiro, R., Whitten, B., Stecca, K., Hazebroek, J., Randhawa, G. S., Dolan, M., Kinney, A. J., Tomes, D. and Nichols, S. Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. *Science* 303(5656):363-366 (2004).

68. Dnyaneshwar, W., Preeti, C., Kalpana, J. and Bhushan, P. Development and application of RAPD-SCAR marker for identification of *Phyllanthus emblica* Linn. *Biological & Pharmaceutical Bulletin* 29(11):2313-2316 (2006).

69. Doyle, J. J. and Doyle, J. L. Isolation of DNA from small amounts of plant tissues. *BRL Focus* 12:13-15 (1990).

70. Dudley, J. W. Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. *Crop Science* 33(4):660-668 (1993).

71. Dugas, D. V., Monaco, M. K., Olson, A., Klein, R. R., Kumari, S., Ware, D. and Klein, P. E. Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. *BMC Genomics* 12(1):514 (2011).

72. Dutta, S., Kumawat, G., Singh, B. P., Gupta, D. K., Singh, S., Dogra, V., Gaikwad, K., Sharma, T. R., Raje, R. S. and Bandhopadhya, T. K. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biology* 11(1):17 (2011).

73. Dwivedi, N. K., Bhandari, D. C., Dubas, B. S., Agrawal, R. C., Mandal, S. and Rana, R. S. Catalogue on cluster bean (*Cyamopsis tetragonoloba* (L.) Taub.) germplasm part III. New Delhi: NBPGR (1995).

74. Eathington, S. R., Crosbie, T. M., Edwards, M. D., Reiter, R. S. and Bull, J. K. Molecular markers in a commercial breeding program. *Crop Science* 47 :S-154 - S-163 (2007).

75. Ferriol, M., Pico, B. and Nuez, F. Genetic diversity of a germplasm collection of *Cucurbita pepo* using SRAP and AFLP markers. *Theoretical and Applied Genetics* 107(2):271-282 (2003).

76. Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., Wong, W. K. and Mockler, T. C. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Research* 20(1):45-58 (2010).

77. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L. and Mistry, J. Pfam: The protein families database. *Nucleic Acids Research*: 1223 (2014).

78. Franssen, S. U., Shrestha, R. P., Brautigam, A., Bornberg-Bauer, E. and Weber, A. P. M. Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics* 12(1):227 (2011).

79. Ganal, M. W., Altmann, T. and Rader, M. S. SNP identification in crop plants. *Current Opinion in Plant Biology* 12(2):211-217 (2009).

80. Gao, L., Ge, S., Hong, D., Lin, R., Tao, G. and Xu, Z. Allozyme variation and conservation genetics of common wild rice (*Oryza rufipogon* Griff.) in Yunnan, China. *Euphytica* 124(3):273-281 (2002).

81. Gao, L. Z. and Hong, S. G. D. Allozyme variation and population genetic structure of common wild rice *Oryza rufipogon* Griff. in China. *Theoretical and Applied Genetics* 101(3):494-502 (2000).

82. Garg, R., Patel, R. K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., Bhatia, S., Chattopadhyay, D., Tyagi, A. K. and Jain, M. Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiology* 156(4):1661-1678 (2011).

83. George, M. and Abraham, T. E. pH sensitive alginate-guar gum hydrogel for the controlled delivery of protein drugs. *International Journal of Pharmaceutics* 335(1):123-129 (2007).

84. Gepts, P. The use of molecular and biochemical markers in crop evolution studies. In: *Evolutionary Biology*. 51-94. Springer, (1993).

85. Geuna, F., Toschi, M. and Bassi, D. The use of AFLP markers for cultivar identification in apricot. *Plant Breeding* 122(6):526-531 (2003).

86. Giannini, E. G., Mansi, C., Dulbecco, P. and Savarino, V. Role of partially hydrolyzed guar gum in the treatment of irritable bowel syndrome. *Nutrition* 22(3):334-342 (2006).

87. Gill, S. L. Evaluation of reciprocal hybrid crosses in guar: Texas Tech University; (2009).

88. Gillett, J. B. *Indigofera* (*microcharis*) in tropical Africa with the related genera *Cyamopsis* and *Rhynchotropis*. *Kew Bulletin* 1:1-166 (1958).*

89. Ginwal, H. S., Maurya, S. S. and Chauhan, P. Genetic diversity and relationship between cultivated clones of *Dalbergia sissoo* of wide geographical origin using RAPD markers. *Journal of Forestry Research* 22(4):507-517 (2011).

90. Ginwal, H. S., Mittal, N. and Barthwal, S. Development and characterization of polymorphic chloroplast microsatellite markers in sweet flag (*Acorus calamus* L.). *Indian Journal of Genetics and Plant Breeding* 69(3):256-259 (2009).

91. Glaszmann, J. C. Geographic pattern of variation among Asian native rice cultivars (*Oryza sativa* L.) based on fifteen isozyme loci. *Genome* 30(5):782-792 (1988).

92. Gliko-Kabir, I., Yagen, B., Baluom, M. and Rubinstein, A. Phosphated crosslinked guar for colon-specific drug delivery: II. *In vitro* and *in vivo* evaluation in the rat. *Journal of Controlled Release* 63(1):129-134 (2000).

93. Gojobori, T., Li, W. H. and Graur, D. Patterns of nucleotide substitution in pseudogenes and functional genes. *Journal of Molecular Evolution* 18(5):360-369 (1982).

94. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. and Zeng, Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29(7):644-652 (2011).

95. Gresta, F., Sortino, O., Santonoceto, C., Issi, L., Formantici, C. and Galante, Y. M. Effects of sowing times on seed yield, protein and galactomannans content of four varieties of guar (*Cyamopsis tetragonoloba* L.) in a mediterranean environment. *Industrial Crops and Products* 41:46-52 (2013).

96. Gupta, P., Balyan, H., Edwards, K., Isaac, P., Korzun, V., Roder, M., Gautier, M. F., Joudrier, P., Schlatter, A. and Dubcovsky, J. Genetic mapping of 66 new microsatellite (SSR) loci in bread wheat. *Theoretical and Applied Genetics* 105(2):413-422 (2002).

97. H Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 017 (2013).

98. Haley, S. D., Miklas, P. N., Stavely, J. R., Byrum, J. and Kelly, J. D. Identification of RAPD markers linked to a major rust resistance gene block in common bean. *Theoretical and Applied Genetics* 86(4):505-512 (1993).

99. Haseneyer, G., Schmutzer, T., Seidel, M., Zhou, R., Mascher, M., Schan, C. C., Taudien, S., Scholz, U., Stein, N. and Mayer, K. F. X. From RNA-Seq to large-scale genotyping-genomics resources for rye (*Secale cereale* L.). *BMC Plant Biology* 11(1):131 (2011).

100. Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., Whaley, A. M., Carrasquilla-Garcia, N., Gaur, P. M. and Upadhyaya, H. D. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnology Journal* 10(6):716-732 (2012).

101. Hirsch, J., Lefort, V., Vankersschaver, M., Boualem, A., Lucas, A., Thermes, C., d'Aubenton-Carafa, Y. and Crespi, M. Characterization of 43 non-protein-coding mRNA genes in *Arabidopsis*, including the MIR162a-derived transcripts. *Plant Physiology* 140(4):1192-1204 (2006).

102. Hoeijmakers, W. A. M., Bartfai, R. and Stunnenberg, H. G. Transcriptome analysis using RNA-Seq. In: *Malaria*. 221-239. Springer, (2013).

103. Hsieh, L. C., Lin, S. I., Shih, A. C. C., Chen, J. W., Lin, W. Y., Tseng, C. Y., Li, W. H. and Chiou, T. J. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. *Plant Physiology* 151(4):2120-2132 (2009).

104. Huang, Y., Yu, H. and Xiao, C. pH-sensitive cationic guar gum/poly (acrylic acid) polyelectrolyte hydrogels: Swelling and *in vitro* drug release. *Carbohydrate Polymers* 69(4):774-783 (2007).

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

106. Hymowitz, T. and Matlock, R. S. Guar in the United States. *Oklahoma Agricultural Experiment Station Technical Bulletin* 611:1-34 (1963).*

107. Ijaz, S. Microsatellite markers: An important fingerprinting tool for characterization of crop plants. *African Journal of Biotechnology* 10(40):7723-7726 (2011).

108. Ilut, D. C., Coate, J. E., Luciano, A. K., Owens, T. G., May, G. D., Farmer, A. and Doyle, J. J. A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-Seq in plant species. *American Journal of Botany* 99(2):383-396 (2012).

109. Irzykowska, L. and Bocianowski, J. Genetic variation, pathogenicity and mycelial growth rate differentiation between *Gaeumannomyces graminis* var. Tritici isolates derived from winter and spring wheat. *Annals of Applied Biology* 152(3):369-375 (2008).

110. Jain, A., Chaudhary, S. and Sharma, P. C. Mining of microsatellites using next generation sequencing of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. *Physiology and Molecular Biology of Plants* 20(1):115-123 (2014).

111. Jain, A., Ghangal, R., Grover, A., Raghuvanshi, S. and Sharma, P. C. Development of EST-based new SSR markers in seabuckthorn. *Physiology and Molecular Biology of Plants* 16(4):375-378 (2010).

112. Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., Shah, N., Singh, V. K., Garg, R., Jeena, G., Yadav, M., Kant, C., Sharma, P., Yadav, G., Bhatia, S., Tyagi, A. and Chattopadhyay, D. A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *The Plant Journal* 74(5):715-729 (2013).

113. Jeffreys, A. J., Wilson, V. and Thein, S. L. Hypervariable-minisatellite-regions in human DNA. *Nature* 314(6006):67-73 (1985).

114. Johal, G. S., Balint-Kurti, P. and Weil, C. F. Mining and harnessing natural variation: A little magic. *Crop Science* 49:2066-2073 (2008).

115. Jones, E., Dupal, M., Dumsday, J., Hughes, L. and Forster, J. An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics* 105(4):577-584 (2002).

116. Jones, N., Ougham, H., Thomas, H. and Pasakinskiene, I. Markers and mapping revisited: Finding your gene. *New Phytologist* 183(4):935-966 (2009).

117. Joshi, S. P., Ranjekar, P. K. and Gupta, V. S. Molecular markers in plant genome analysis. *Current Science* 77(2):230-240 (1999).

118. Julio, E., Verrier, J. L. and Dorlhac de Borne, F. Development of SCAR markers linked to three disease resistances based on AFLP within *Nicotiana tabacum* L. *Theoretical and Applied Genetics* 112(2):335-346 (2006).

119. Kakumanu, A., Ambavaram, M. M. R., Klumas, C., Krishnan, A., Batlang, U., Myers, E., Grene, R. and Pereira, A. Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. *Plant Physiology* 160(2):846-867 (2012).

120. Kanehisa, M. and Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28(1):27-30 (2000).

121. Karasawa, M. M. G., Vencovsky, R., Silva, C. M., Cardim, D. C., Bressan, E. d. A., Oliveira, G. C. X. and Veasey, E. A. Comparison of microsatellites and isozymes in genetic diversity studies of *Oryza glumaepatula* (Poaceae) populations. *Revista de Biología Tropical* 60(4):1463-1478 (2012).

122. Kaushik, D., Sardana, S. and Mishra, D. N. *In vitro* cytotoxicity analysis of 5-fluorouracil loaded guar gum microspheres on HT-29 colon cancer cell line. *International Journal of Pharmaceutical Sciences and Drug Research* 1(2):83-84 (2010).

123. Kawamura, Y. Guar gum chemical and technical assessment. *Prepared for the 69th Joint FAO/WHO Expert Committee on Food Additives* (2008).

124. Kesawat, M. S. and Kumar, B. D. Molecular markers: It's application in crop improvement. *Journal of Crop Science and Biotechnology* 12(4):169-181 (2009).

125. Khasa, D. P., Nadeem, S., Thomas, B., Robertson, A. and Bousquet, J. Application of SSR markers for parentage analysis of Populus clones. *Forest Genetics* 10(4):273-282 (2003).

126. Kim, H. J., Jung, J., Kim, M.-S., Lee, J. M., Choi, D., Yeam, I. and Lukens, L. Molecular marker development and genetic diversity exploration by RNA-Seq in P*latycodon grandiflorum*. *Genome* 58(10):441-451 (2015).

127. Kim, K. H., Kang, Y. J., Kim, D. H., Yoon, M. Y., Moon, J.-K., Kim, M. Y., Van, K. and Lee, S.-H. RNA-Seq analysis of a soybean near-isogenic line carrying bacterial leaf pustule-resistant and susceptible alleles. *DNA Research* 18(6):483-497 (2011).

128. Kiss, G. B., Csanadi, G., Kalman, K., Kaló, P. and Ökrész, L. Construction of a basic genetic map for alfalfa using RFLP, RAPD, isozyme and morphological markers. *Molecular and General Genetics* 238(1):129-137 (1993).

129. Koepke, T., Schaeffer, S., Krishnan, V., Jiwan, D., Harper, A., Whiting, M., Oraguzie, N. and Dhingra, A. Rapid gene-based SNP and haplotype marker development in non-model eukaryotes using 3'UTR sequencing. *BMC Genomics* 13(1):18 (2012).

130. Koressaar, T. and Remm, M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23(10):1289-1291 (2007).

131. Kornobis, E., Cabellos, L., Aguilar, F., Fras-Lapez, C., Rozas, J., Marco, J. S. and Zardoya, R. TRUFA: A user-friendly web server for de novo RNA-Seq analysis using cluster computing. *Evolutionary Bioinformatics* 11:97 (2015).

132. Kramar, A., Turk, S. and Vreaer, F. Statistical optimisation of diclofenac sustained release pellets coated with polymethacrylic films. *International Journal of Pharmaceutics* 256(1):43-52 (2003).

133. Krishnaiah, Y. S. R., Karthikeyan, R. S., Sankar, V. G. and Satyanarayana, V. Three-layer guar gum matrix tablet formulations for oral controlled delivery of highly soluble trimetazidine dihydrochloride. *Journal of Controlled Release* 81(1):45-56 (2002).

134. Krishnaiah, Y. S. R., Karthikeyan, R. S. and Satyanarayana, V. A three-layer guar gum matrix tablet for oral controlled delivery of highly soluble metoprolol tartrate. *International Journal of Pharmaceutics* 241(2):353-366 (2002).

135. Krishnaiah, Y. S. R., Raju, P. V., Kumar, B. D., Bhaskar, P. and Satyanarayana, V. Development of colon targeted drug delivery systems for mebendazole. *Journal of Controlled Release* 77(1):87-95 (2001).

136. Krishnaiah, Y. S. R., Reddy, P. R. B., Satyanarayana, V. and Karthikeyan, R. S. Studies on the development of oral colon targeted drug delivery systems for metronidazole in the treatment of amoebiasis. *International Journal of Pharmaceutics* 236(1):43-55 (2002).

137. Krishnaiah, Y. S. R., Satyanarayana, S. and Rama Prasad, Y. V. Studies of guar gum compression-coated 5-aminosalicylic acid tablets for colon-specific drug delivery. *Drug Development and Industrial Pharmacy* 25(5):651-657 (1999).

138. Krishnaiah, Y. S. R., Satyanarayana, V., Kumar, B. D. and Karthikeyan, R. S. *In vitro* drug release studies on guar gum-based colon targeted oral drug delivery systems of 5-fluorouracil. *European Journal of Pharmaceutical Sciences* 16(3):185-192 (2002).

139. Krishnaiah, Y. S. R., Seetha Devi, A., Nageswara Rao, L., Bhaskar Reddy, P. R., Karthikeyan, R. S. and Satyanarayana, V. Guar gum as a carrier for colon specific delivery; influence of metronidazole and tinidazole on in vitro release of albendazole from guar gum matrix tablets. *Journal of Pharmacy and Pharmaceutical Science* 4:235-243 (2001).

140. Kubisiak, T. L., Nelson, C. D., Nance, W. L. and Stine, M. RAPD linkage mapping in a longleaf pine x slash pine F1 family. *Theoretical and Applied Genetics* 90(7):1119-1127 (1995).

141. Kumar, D. and Singh, N. B. Guar in India. Jodhpur: Scientific Publishers (India) (2002).

142. Kumar, S., Joshi, U. N., Singh, V., Singh, J. V. and Saini, M. L. Characterization of released and elite genotypes of guar [*Cyamopsis tetragonoloba* (L.) Taub.] from India proves

unrelated to geographical origin. *Genetic Resources and Crop Evolution* 60(7):2017-2032 (2013).

143. Kumar, S., Parekh, M. J., Patel, C. B., Zala, H. N., Sharma, R., Kulkarni, K. S., Fougat, R. S., Bhatt, R. K. and Sakure, A. A. Development and validation of EST-derived SSR markers and diversity analysis in cluster bean (*Cyamopsis tetragonoloba*). *Journal of Plant Biochemistry and Biotechnology*:1-7 (2015).

144. Kunihisa, M., Fukino, N. and Matsumoto, S. Development of cleavage amplified polymorphic sequence (CAPS) markers for identification of strawberry cultivars. *Euphytica* 134(2):209-215 (2003).

145. Kuravadi, N. A., Tiwari, P. B., Choudhary, M. and Randhawa, G. S. Genetic diversity study of cluster bean (*Cyamopsis tetragonoloba* (L.) Taub.) landraces using RAPD and ISSR markers. *International Journal of Advance Biotechnology and Research* 4(4):460-471 (2013).

146. Kuravadi, N. A., Tiwari, P. B., Tanwar, U. K., Tripathi, S. K., Dhugga, K. S., Gill, K. S. and Randhawa, G. S. Identification and characterization of EST-SSR markers in cluster bean (spp.). *Crop Science* 54(3):1097-1102 (2014).

147. Kuravadi, N. A., Verma, S., Pareek, S., Gahlot, P., Kumari, S., Tanwar, U. K., Bhatele, P., Choudhary, M., Gill, K. S. and Pruthi, V. Guar: An industrial crop from marginal farms. *Agricultural sustainability: Progress and prospects in crop research,* Elsevier Science, Academic Press. (2013).

148. Langmead, B. and Salzberg, S. L. Fast gapped-read alignment with BOWTIE 2. *Nature Methods* 9(4):357-359 (2012).

149. Lcwontin, R. C. and Hubby, I. L. A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of hcterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* 54:595-609 (1966).

150. Lehmann, K. O. R. and Dreher, K. D. Methacrylate-galactomannan coating for colon-specific drug delivery. In: *Proc. Int. Symp. Controlled Release Bioact. Mater*. 331-332, (1991).

151. Li, B. and Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12(1):323 (2011).

152. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987-2993 (2011).

153. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. The sequence alignment/map format and SamTools. *Bioinformatics* 25(16):2078-2079 (2009).

154. Li, W. and Godzik, A. CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658-1659 (2006).

155. Liepman, A. H., Nairn, C. J., Willats, W. G. T., Sa¸rensen, I., Roberts, A. W. and Keegstra, K. Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. *Plant Physiology* 143(4):1881-1893 (2007).

156. Lu, F. H., Cho, M. C. and Park, Y. J. Transcriptome profiling and molecular marker discovery in red pepper, *Capsicum annuum* L. Tf68. *Molecular Biology Reports* 39(3):3327-3335 (2012).

157. Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., Feng, Q., Zhao, Y., Guo, Y. and Li, W. Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-Seq. *Genome Research* 20(9):1238-1249 (2010).

158. Lulin, H., Xiao, Y., Pei, S., Wen, T. and Shangqin, H. The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. *Plos One* 7(6):e38653 (2012).

159. Maheswaran, M., Subudhi, P. K., Nandi, S., Xu, J. C., Parco, A., Yang, D. C. and Huang, N. Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. *Theoretical and Applied Genetics* 94(1):39-45 (1997).

160. Mao, L., Van Hemert, J. L., Dash, S. and Dickerson, J. A. Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics* 10(1):346 (2009).

161. Martin, G. B., Williams, J. G. and Tanksley, S. D. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proceedings of the National Academy of Sciences* 88(6):2336 (1991).

162. Martin, L. B. B., Fei, Z., Giovannoni, J. J. and Rose, J. K. C. Catalyzing plant science research with RNA-Seq. *Frontiers in Plant Science* 4 (2013).

163. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* 17(1): 10-12 (2011).

164. Matas, A. J., Yeats, T. H., Buda, G. J., Zheng, Y., Chatterjee, S., Tohge, T., Ponnala, L., Adato, A., Aharoni, A. and Stark, R. Tissue-and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. *The Plant Cell* 23(11):3893-3910 (2011).

165. Mazliak, P. Lipid metabolism in plants. *Annual Review of Plant Physiology* 24(1):287-310 (1973).

166. McGettigan, P. A. Transcriptomics in the RNA-Seq era. *Current Opinion in Chemical Biology* 17(1):4-11 (2013).

167. Melotto, M., Afanador, L. and Kelly, J. D. Development of a SCAR marker linked to the I gene in common bean. *Genome* 39(6):1216-1219 (1996).

168. Meudt, H. M. and Clarke, A. C. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science* 12(3):106-117 (2007).

169. Michaels, S. D. and Amasino, R. M. A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. *The Plant Journal* 14(3):381-385 (1998).

170. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Analytical Chemistry* 31(3):426-428 (1959).

171. Mir, R. R., Rustgi, S., Sharma, S., Singh, R., Goyal, A., Kumar, J., Gaur, A., Tyagi, A. K., Khan, H. and Sinha, M. K. A preliminary genetic analysis of fibre traits and the use of new genomic SSRs for genetic diversity in Jute. *Euphytica* 161(3):413-427 (2008).

172. Mirza, N., Taj, G., Arora, S. and Kumar, A. Transcriptional expression analysis of genes involved in regulation of calcium translocation and storage in finger millet (*Eleusine coracana* L. Gartn.). *Gene* 550(2):171-179 (2014).

173. Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F. and Myburg, A. A. *De novo* assembled expressed gene catalog of a fast-growing eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics* 11(1):681 (2010).

174. Mizuno, H., Kawahigashi, H., Kawahara, Y., Kanamori, H., Ogata, J., Minami, H., Itoh, T. and Matsumoto, T. Global transcriptome analysis reveals distinct expression among duplicated genes during *Sorghum-Bipolaris sorghicola* interaction. *BMC Plant Biology* 12(1):121 (2012).

175. Mohan, M., Nair, S., Bhagwat, A., Krishna, T. G., Yano, M., Bhatia, C. R. and Sasaki, T. Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding* 3(2):87-103 (1997).

176. Morgante, M., Hanafey, M. and Powell, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* 30(2):194-200 (2002).

177. Morishima, H. and Barbier, P. Mating system and genetic structure of natural populations in wild rice *Oryza rufipogon**. *Plant Species Biology* 5(1):31-39 (1990).

178. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621-628 (2008).

179. Moxon, S., Jing, R., Szittya, G., Schwach, F., Pilcher, R. L. R., Moulton, V. and Dalmay, T. Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Research* 18(10):1602-1609 (2008).

180. Mullis, K. B. and Faloona, F. A. specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155:335-350 (1987).

181. Mullis, K. B., Faloona, F. A., Scharf, S. J., Saiki, R. K., Horn, G. T. and Erlich, H. Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. *Biotechnology Series*:17-17 (1992).

182. Murphy, R. W., Sites, J. W., Buth, D. G. and Haufler, C. H. Proteins: Isozyme electrophoresis. *Molecular Systematics* 2:51-120 (1996).

183. Murthy, S. N., Hiremath, S. R. R. and Paranjothy, K. L. K. Evaluation of carboxymethyl guar films for the formulation of transdermal therapeutic systems. *International Journal of Pharmaceutics* 272(1):11-18 (2004).

184. Mutz, K. O., Heilkenbrinker, A., Lönne, M., Walter, J.-G. and Stahl, F. Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology* 24(1):22-30 (2013).

185. Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320(5881):1344-1349 (2008).

186. Nakasugi, K., Crowhurst, R. N., Bally, J., Wood, C. C., Hellens, R. P. and Waterhouse, P. M. *De novo* transcriptome sequence assembly and analysis of RNA silencing genes of *Nicotiana benthamiana*. *Plos One* 8(3):59534 (2013).

187. Nandha, P. S. and Singh, J. Comparative assessment of genetic diversity between wild and cultivated barley using gSSR and EST-SSR markers. *Plant Breeding* 133(1):28-35 (2014).

188. Naoumkina, M., Torres-Jerez, I., Allen, S., He, J., Zhao, P. X., Dixon, R. A. and May, G. D. Analysis of cDNA libraries from developing seeds of guar (*Cyamopsis tetragonoloba* (L.) Taub.). *BMC Plant Biology* 7(1):62 (2007).

189. Naoumkina, M., Vaghchhipawala, S., Tang, Y., Ben, Y., Powell, R. J. and Dixon, R. A. Metabolic and genetic perturbations accompany the modification of galactomannan in seeds of *Medicago truncatula* expressing mannan synthase from guar (*Cyamopsis tetragonoloba* L.). *Plant Biotechnology Journal* 6(6):619-631 (2008).

190. Nasu, S., Suzuki, J., Ohta, R., Hasegawa, K., Yui, R., Kitazawa, N., Monna, L. and Minobe, Y. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Research* 9(5):163-171 (2002).

191. Neff, M. M., Neff, J. D., Chory, J. and Pepper, A. E. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: Experimental applications in *Arabidopsis thaliana* genetics. *The Plant Journal* 14(3):387-392 (1998).

192. Negi, M. S., Devic, M., Delseny, M. and Lakshmikumaran, M. Identification of AFLP fragments linked to seed coat colour in *Brassica juncea* and conversion to a SCAR marker for rapid selection. *Theoretical and Applied Genetics* 101(1):146-152 (2000).

193. Novaes, E., Drost, D. R., Farmerie, W. G., Pappas, G. J., Grattapaglia, D., Sederoff, R. R. and Kirst, M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9(1):312 (2008).

194. Oh, T. J., Gorman, M. and Cullis, C. A. RFLP and RAPD mapping in flax (*Linum usitatissimum*). *Theoretical and Applied Genetics* 101(4):590-593 (2000).

195. Ossowski, S., Schneeberger, K., Clark, R. M., Lanz, C., Warthmann, N. and Weigel, D. Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Research* 18(12):2024-2033 (2008).

196. Pantaleo, V., Szittya, G., Moxon, S., Miozzi, L., Moulton, V., Dalmay, T. and Burgyan, J. Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *The Plant Journal* 62(6):960-976 (2010).

197. Parchman, T. L., Geist, K. S., Grahnen, J. A., Benkman, C. W. and Buerkle, C. A. Transcriptome sequencing in an ecologically important tree species: Assembly, annotation, and marker discovery. *BMC Genomics* 11(1):180 (2010).

198. Parra, G., Bradnam, K. and Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061-1067 (2007).

199. Parra, G., Bradnam, K., Ning, Z., Keane, T. and Korf, I. Assessing the gene space in draft genomes. *Nucleic Acids Research* 37(1):289-297 (2009).

200. Patel, J. J., Karve, M. and Patel, N. K. Guar gum: A versatile material for pharmaceutical industries. *International Journal of Pharmacy and Pharmaceutical Sciences* 6(8):13-19 (2014).

201. Pathak, R. Clusterbean: Physiology, genetics and cultivation. Springer, (2015).

202. Pathak, R., Singh, M. and Henry, A. Genetic diversity and interrelationship among clusterbean (*Cyamopsis tetragonoloba*) genotypes for qualitative traits. *Indian Journal of Agricultural Sciences* 81(5):402 (2011).

203. Pathak, R., Singh, S. K. and Singh, M. Assessment of genetic diversity in clusterbean using nuclear rDNA and RAPD markers. *Journal of Food Legumes* 24(3):180-183 (2011).

204. Pathak, R., Singh, S. K., Singh, M. and Henry, A. Molecular assessment of genetic diversity in cluster bean (*Cyamopsis tetragonoloba*) genotypes. *Journal of Genetics* 89(2):243-246 (2010).

205. Pati, N., Schowinsky, V., Kokanovic, O., Magnuson, V. and Ghosh, S. A comparison between Snapshot, Pyrosequencing, and Biplex invader SNP genotyping methods: Accuracy, cost, and throughput. *Journal of Biochemical and Biophysical Methods* 60(1):1-12 (2004).

206. Patil, C. G. Nuclear DNA amount variation in *Cyamopsis* D.C. (Fabaceae). *Cytologia* 69(1):59-62 (2004).

207. Peng, J. H. and Lapitan, N. L. V. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. *Functional & Integrative Genomics* 5(2):80-96 (2005).

208. Pietu, G. v., Mariage-Samson, R. g., Fayein, N. A., Matingou, C., Eveno, E., Houlgatte, R. M., Decraene, C., Vandenbrouck, Y., Tahi, F. and Devignes, M.-D. The genexpress image knowledge base of the human brain transcriptome: A prototype integrated resource for functional and computational genomics. *Genome Research* 9(2):195-209 (1999).

209. Poats, F. J. Guar, a summer row crop for the southwest. *Economic Botany* 14(3):241-246 (1960).*

210. Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. and Rafalski, A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* 2(3):225-238 (1996).

211. Prabaharan, M. Prospective of guar gum and its derivatives as controlled drug delivery systems. *International Journal of Biological Macromolecules* 49(2):117-124 (2011).

212. Pradhan, S., Bandhiwal, N., Shah, N., Kant, C., Gaur, R. and Bhatia, S. Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. *Frontiers in Plant Science* 5:698 (2014).

213. Prasad, Y. V. R., Krishnaiah, Y. S. R. and Satyanarayana, S. *In vitro* evaluation of guar gum as a carrier for colon-specific drug delivery. *Journal of Controlled Release* 51(2):281-287 (1998).

214. Punia, A., Arora, P., Yadav, R. and Chaudhury, A. Optimization and inference of PCR conditions for genetic variability studies of commercially important cluster bean varieties by

RAPD analysis. *Asian Pacific Journal of Molecular Biology and Biotechnology* 17:33-38 (2009).

215. Punia, A., Arora, P., Yadav, R. and Chaudhury, A. Optimization and inference of PCR conditions for genetic variability studies of commercially important cluster bean varieties by RAPD analysis. *Asia Pacific Journal of Molecular Biology and Biotechnology* 17:33-38 (2009).

216. Punia, A., Yadav, R., Arora, P. and Chaudhury, A. Molecular and morpho-physiological characterization of superior cluster bean (*Cymopsis tetragonoloba*) varieties. *Journal of Crop Science and Biotechnology* 12(3):143-148 (2009).

217. Rafalski, A. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology* 5(2):94-100 (2002).

218. Rafalski, J. A. and Tingey, S. V. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends in Genetics* 9(8):275-280 (1993).

219. Rakoczy-Trojanowska, M. and Bolibok, H. Characteristics and a comparison of three classes of microsatellite-based markers and their application in plants. *Cellular and Molecular Biology Letters* 9(2):221-238 (2004).

220. Ranade, S. A., Verma, A., Gupta, M. and Kumar, N. RAPD profile analysis of betel vine cultivars. *Biologia Plantarum* 45(4):523-527 (2002).

221. Ravi, P., Rao Kusumanchi, R. M., Mallikarjun, V., Babu Rao, B. and B., R. N. Formulation and evaluation of guar gum microspheres of aceclofenac for colon targeted delivery. *Journal of Pharmacy Research* 3(7):1510-1512 (2010).

222. Ribaut, J. M., Hu, X., Hoisington, D. and González-de-León, D. Use of STSS and SSRs as rapid and reliable preselection tools in a marker-assisted selection-backcross scheme. *Plant Molecular Biology Reporter* 15(2):154-162 (1997).

223. Roberts, A. and Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods* 10(1):71-73 (2013).

224. Robinson, J. and Lee, V. H. L. Controlled drug delivery: Fundamentals and applications: Informa Health Care (1987).

225. Robinson, J. T., H Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. Integrative genomics viewer. *Nature Biotechnology* 29(1):24-26 (2011).

226. Roorkiwal, M. and Sharma, P. C. Mining functional microsatellites in legume unigenes. *Bioinformation* 7(5):264-270 (2011).

227. Rowe, R. C., Sheskey, P. J. and Quinn, M. E. Handbook of pharmaceutical excipients. 6 ed. London, UK: Pharmaceutical Press (2009).

228. Rozee, K. R. and Johnson, W. M. DNA amplification fingerprinting: Another diagnostic tool? *The Canadian Journal of Infectious Diseases* 2(4):165 (1991).

229. Rubinstein, A., Radai, R., Ezra, M., Pathak, S. and Rokem, J. S. *In vitro* evaluation of calcium pectinate: A potential colon-specific drug delivery carrier. *Pharmaceutical Research* 10(2):258-263 (1993).

230. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239(4839):487 (1988).

231. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230(4732):1350-1354 (1985).

232. Saliba-Colombani, V., Causse, M., Gervais, L. and Philouze, J. Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* 43(1):29-40 (2000).

233. Sangwan, R. S., Sangwan, N. S., Jain, D. C., Kumar, S. and Ranade, S. A. RAPD profile based genetic characterization of chemotypic variants of *Artemisia annua* L. *Biochmistry and Molecular Biology International* 47(6):935-944 (1999).

234. Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A. and Kawashima, K. Genome structure of the legume, *Lotus japonicus*. *DNA Research* 15(4):227-239 (2008).

235. Schlötterer, C. and Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20(2):211-215 (1992).

236. Schmidt, M. A., Barbazuk, W. B., Sandford, M., May, G., Song, Z., Zhou, W., Nikolau, B. J. and Herman, E. M. Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. *Plant Physiology* 156(1):330-345 (2011).

237. Schmieder, R. and Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863-864 (2011).

238. Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J. and Cheng, J. Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278):178-183 (2010).

239. Semagn, K., Bjornstad, A. and Ndjiondjop, M. N. An overview of molecular marker methods for plants. *African Journal of Biotechnology* 5(25):2540-2568 (2006).

240. Severing, E. I., van Dijk, A. D., Morabito, G., Busscher-Lange, J., Immink, R. G. and van Ham, R. C. Predicting the impact of alternative splicing on plant MADS domain protein function. *Plos One* 7(1):e30524 (2012).

241. Shahinnia, F. and Sayed-Tabatabaei, B. E. Conversion of barley SNPs into PCR-based markers using dCAPS method. *Genetics and Molecular Biology* 32(3):564-567 (2009).

242. Shaikh, T. and Kumar, S. S. Pharmaceutical and pharmacological profile of guar gum an overview. *International Journal of Pharmacy and Pharmaceutical Sciences* 3:38-40 (2011).

243. Sharma, P., Kumar, V., Raman, K. V. and Tiwari, K. A set of scar markers in cluster bean (*Cyamopsis tetragonoloba* L. Taub.) genotypes. *Advances in Bioscience and Biotechnology* 5(2):131-141 (2014).

244. Shasany, A. K., Darokar, M. P., Dhawan, S., Gupta, A. K., Gupta, S., Shukla, A. K., Patra, N. K. and Khanuja, S. P. S. Use of RAPD and AFLP markers to identify inter-and intraspecific hybrids of mentha. *Journal of Heredity* 96(5):542-549 (2005).

245. Singh, J. V. and Dahiya, B. S. Guar breeding-global scenario. *Guar Indian Society of Forage Research, Hisar and APEDA (India)*:10-33 (2004).

246. Singh, O. V. and Nagaraj, N. S. Transcriptomics, proteomics and interactomics: Unique approaches to track the insights of bioremediation *Briefings in Functional Genomics & Proteomics* 4(4):355-362 (2006).

247. Singh, V. P. Induced high yielding mutants in clusterbean. *Indian Journal of Agriculturl Sciences* 56:695-700 (1986).

248. Sinha, V. R., Mittal, B. R., Bhutani, K. K. and Kumria, R. Colonic drug delivery of 5-fluorouracil: An *in vitro* evaluation. *International Journal of Pharmaceutics* 269(1):101-108 (2004).

249. Slavin, J. L. and Greenberg, N. A. Partially hydrolyzed guar gum: Clinical nutrition uses. *Nutrition* 19(6):549-552 (2003).

250. Soltis, D. E., Moore, M. J., Burleigh, G. and Soltis, P. S. Molecular markers and concepts of plant evolutionary relationships: Progress, promise, and future prospects. *Critical Reviews in Plant Sciences* 28(1-2):1-15 (2009).

251. Sonah, H., Deshmukh, R. K., Sharma, A., Singh, V. P., Gupta, D. K., Gacche, R. N., Rana, J. C., Singh, N. K. and Sharma, T. R. Genome-wide distribution and organization of microsatellites in plants: An insight into marker development in *Brachypodium*. *Plos One* 6(6):e21298 (2011).

252. Soppimath, K. S., Kulkarni, A. R. and Aminabhavi, T. M. Chemically modified polyacrylamide-g-guar gum-based crosslinked anionic microgels as pH-sensitive drug delivery systems: Preparation and characterization. *Journal of Controlled Release* 75(3):331-345 (2001).

253. Soppirnath, K. S. and Aminabhavi, T. M. Water transport and drug release study from cross-linked polyacrylamide grafted guar gum hydrogel microspheres for the controlled release application. *European Journal of Pharmaceutics and Biopharmaceutics* 53(1):87-98 (2002).

254. Srivastava, M. and Kapoor, V. P. Seed galactomannans: An overview. *Chemistry & Biodiversity* 2(3):295-317 (2005).

255. Stafford, R. E. Inheritance of partial male-sterility in guar. *Plant Breeding* 103(1):43-46 (1989).*

256. Strutevant, A. The linear arrangement of six sex-linked factors indrosophila as shown by their mode of association. *Molecular and General Genetics MGG* 10(1):293-294 (1913).

257. Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A. and Parkhomchuk, D. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321(5891):956-960 (2008).

258. Sun, X., Zhou, S., Meng, F. and Liu, S. *De novo* assembly and characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. *Plant Cell Reports* 31(10):1823-1828 (2012).

259. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H. and UniProt, C. UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 739 (2014).

260. Takacs, E. M., Li, J., Du, C., Ponnala, L., Janick-Buckner, D., Yu, J., Muehlbauer, G. J., Schnable, P. S., Timmermans, M. C. P. and Sun, Q. Ontogeny of the maize shoot apical meristem. *The Plant Cell* 24(8):3219-3234 (2012).

261. Tanksley, S. D. and McCouch, S. R. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277(5329):1063 (1997).

262. Tautz, D. and Renz, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* 12(10):4127-4138 (1984).

263. Thakur, S., Chauhan, G. S. and Ahn, J. H. Synthesis of acryloyl guar gum and its hydrogel materials for use in the slow release of L-dopa and L-tyrosine. *Carbohydrate Polymers* 76(4):513-520 (2009).

264. Thiel, T., Michalek, W., Varshney, R. and Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106(3):411-422 (2003).

265. Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14(2):178-192 (2013).

266. Torre, A. R. Taxa angolensia nova vel minus cognita -1. *Memoras do Junta de Investigazoes do Ultramar, Seria II* 19:23-66 (1960).*

267. Torre, S., Tattini, M., Brunetti, C., Fineschi, S., Fini, A., Ferrini, F. and Sebastiani, F. RNA-Seq analysis of *Quercus pubescens* leaves: *De novo* transcriptome assembly, annotation and functional markers development. *Plos One* 9:11 (2014).

268. Toti, U. S. and Aminabhavi, T. M. Modified guar gum matrix tablet for controlled release of diltiazem hydrochloride. *Journal of Controlled Release* 95(3):567-577 (2004).

269. Undersander, D. J., Putnam, D. H., Kaminski, A. R., Kelling, K. A., Doll, J. D., Oplinger, E. S. and Gunsolus, J. L. Guar. Alternative field crops manual. *University of Wisconsin cooperative extension service, University of Minnesota extension service* (1991).

270. Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. and Rozen, S. G. Primer3-new capabilities and interfaces. *Nucleic Acids Research* 40(15):e115 (2012).

271. Van Bel, M., Proost, S., Van Neste, C., Deforce, D., Van de Peer, Y. and Vandepoele, K. TRAPID: An efficient online tool for the functional and comparative analysis of *de novo* RNA-Seq transcriptomes. *Genome Biology* 14(12):1-10 (2013).

272. van der Voort, J. R., Van Zandvoort, P., Van Eck, H. J., Folkertsma, R. T., Hutten, R. C. B., Draaistra, J., Gommers, F. J., Jacobsen, E., Helder, J. and Bakker, J. Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. *Molecular and General Genetics* 255(4):438-447 (1997).

273. Van Verk, M. C., Hickman, R., Pieterse, C. M. J. and Van Wees, S. C. M. RNA-Seq: Revelation of the messengers. *Trends in Plant Science* 18(4):175-179 (2013).

274. Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., Donoghue, M. T. A., Azam, S., Fan, G., Whaley, A. M., Farmer, A. D., Sheridan, J., Iwata, A., Tuteja, R., Penmetsa, R. V., Wu, W., Upadhyaya, H. D., Yang, S. P., Shah, T., Saxena, K. B., Michael, T., McCombie, W. R., Yang, B., Zhang, G., Yang, H., Wang, J., Spillane, C., Cook, D. R., May, G. D., Xu, X. and Jackson, S. A. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nature Biotechnology* 30(1):83-89 (2012).

275. Varshney, R. K., Graner, A. and Sorrells, M. E. Genic microsatellite markers in plants: Features and applications. *Trends in Biotechnology* 23(1):48-55 (2005).

276. Varshney, R. K., Grosse, I., HÃ¤hnel, U., Siefken, R., Prasad, M., Stein, N., Langridge, P., Altschmied, L. and Graner, A. Genetic mapping and bac assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. *Theoretical and Applied Genetics* 113(2):239-250 (2006).

277. Varshney, R. K., Grosse, I., Hahnel, U., Siefken, R., Prasad, M., Stein, N., Langridge, P., Altschmied, L. and Graner, A. Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. *Theoretical and Applied Genetics* 113(2):239-250 (2006).

278. Varshney, R. K., Sigmund, R., Barner, A., Korzun, V., Stein, N., Sorrells, M. E., Langridge, P. and Graner, A. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Science* 168(1):195-202 (2005).

279. Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., Cannon, S., Baek, J., Rosen, B. D. and Tar'an, B. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology* 31(3):240-246 (2013).

280. Veasey, E. A., Cardin, D., Silva, R. r. M., Bressan, E. d. A. and Vencovsky, R. Assessing the genetic structure of *Oryza glumaepatula* populations with isozyme markers. *Brazilian Archives of Biology and Technology* 51(5):873-882 (2008).

281. Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B. and Kinzler, K. W. Characterization of the yeast transcriptome. *Cell* 88(2):243-251 (1997).

282. Wakeley, J. Substitution-rate variation among sites and the estimation of transition bias. *Molecular Biology and Evolution* 11(3):436-442 (1994).

283. Wakeley, J. The excess of transitions among nucleotide substitutions: New methods of estimating transition bias underscore its significance. *Trends in Ecology & Evolution* 11(4):158-162 (1996).

284. Wang, C. and Roberts, P. A. Development of AFLP and derived CAPS markers for root-knot nematode resistance in cotton. *Euphytica* 152(2):185-196 (2006).

285. Wang, F., Li, L., Li, H., Liu, L., Zhang, Y., Gao, J. and Wang, X. Transcriptome analysis of rosette and folding leaves in Chinese cabbage using high-throughput RNA sequencing. *Genomics* 99(5):299-307 (2012).

286. Wang, S., Wang, X., He, Q., Liu, X., Xu, W., Li, L., Gao, J. and Wang, F. Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing

and development of EST-SSR markers in radish. *Plant Cell Reports* 31(8):1437-1447 (2012).

287. Wang, Z., Gerstein, M. and Snyder, M. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10(1):57-63 (2009).

288. Wang, Z., Li, J., Luo, Z., Huang, L., Chen, X., Fang, B., Li, Y., Chen, J. and Zhang, X. Characterization and development of EST-derived SSR markers in cultivated sweet potato (*Ipomoea batatas*). *BMC Plant Biology* 11(1):139 (2011).

289. Wang, Z., Yu, G., Shi, B., Wang, X., Qiang, H. and Gao, H. Development and characterization of simple sequence repeat (SSR) markers based on RNA-sequencing of *Medicago sativa* and *in silico* mapping onto the *M. truncatula* genome. *Plos One* 9(3):e92029 (2014).

290. Weber, A. P. M., Weber, K. L., Carr, K., Wilkerson, C. and Ohlrogge, J. B. Sampling the arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiology* 144(1):32-42 (2007).

291. Weller, J. I., Soller, M. and Brody, T. Linkage analysis of quantitative traits in an interspecific cross of tomato (*Lycopersicon esculentum* x *Lycopersicon pimpinellifolium*) by means of genetic markers. *Genetics* 118(2):329 (1988).

292. Welsh, J., Chada, K., Dalal, S. S., Cheng, R., Relph, D. and McClelland, M. Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Research* 20(19):4965-4970 (1992).

293. Whistler, R. L. and Hymowitz, T. Guar: Agronomy, production, industrial use, and nutrition: Purdue University Press. (1979).

294. Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J. and Bahler, J. r. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453(7199):1239-1243 (2008).

295. Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. and Tingey, S. V. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* 18(22):6531-6535 (1990).

296. Wolf, J. B. W. Principles of transcriptome analysis and gene expression quantification: An RNA-Seq tutorial. *Molecular Ecology Resources* 13(4):559-572 (2013).

297. Wong, D., Larrabee, S., Clifford, K., Tremblay, J. and Friend, D. R. USP dissolution apparatus III (reciprocating cylinder) for screening of guar-based colonic delivery formulations. *Journal of Controlled Release* 47(2):173-179 (1997).

298. Wu, G., Zhang, L., Yin, Y., Wu, J., Yu, L., Zhou, Y. and Li, M. Sequencing, *de novo* assembly and comparative analysis of *Raphanus sativus* transcriptome. *Frontiers in Plant Science* 6:198 (2015).

299. Xin, D., Sun, J., Wang, J., Jiang, H., Hu, G., Liu, C. and Chen, Q. Identification and characterization of SSRs from soybean (*Glycine max*) ESTs. *Molecular Biology Reports* 39(9):9047-9057 (2012).

300. Yadav, H., Prasad, A. K., Goswami, P., Pednekar, S., Haque, E. and Shah, M. Guar industry outlook 2015. Report made for: National commodity & derivatives exchange limited. November, 2013. NIAM, Jaipur, (2013).

301. Yan, Z., Denneboom, C., Hattendorf, A., Dolstra, O., Debener, T., Stam, P. and Visser, P. B. Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers. *Theoretical and Applied Genetics* 110(4):766-777 (2005).

302. Yang, L., Chu, J. S. and Fix, J. A. Colon-specific drug delivery: New approaches and *in vitro/in vivo* evaluation. *International Journal of Pharmaceutics* 235(1):1-15 (2002).

303. Yang, T., Bao, S. Y., Ford, R., Jia, T. J., Guan, J. P., He, Y. H., Sun, X. L., Jiang, J. Y., Hao, J. J. and Zhang, X. Y. High-throughput novel microsatellite marker of faba bean via next generation sequencing. *BMC Genomics* 13(1):602 (2012).

304. Yang, Z. and Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* 15(12):496-503 (2000).

305. Yang, Z. and Yoder, A. D. Estimation of the transition/transversion rate bias and species sampling. *Journal of Molecular Evolution* 48(3):274-283 (1999).

306. Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R. and Bolund, L. WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research* 34:W293-W297 (2006).

307. Young, N. D., Debella, F. D. R., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F. X., Gouzy, J. R. M. and Schoof, H. The medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 480(7378):520-524 (2011).

308. Yu, J. N., Won, C., Jun, J., Lim, Y. W. and Kwak, M. Fast and cost-effective mining of microsatellite markers using NGS technology: An example of a Korean water deer hydropotes inermis argyropus. *Plos One* 6(11):e26933 (2011).

309. Zamir, D. Improving plant breeding with exotic genetic libraries. *Nature Reviews Genetics* 2(12):983-989 (2001).

310. Zenoni, S., Ferrarini, A., Giacomelli, E., Xumerle, L., Fasoli, M., Malerba, G., Bellin, D., Pezzotti, M. and Delledonne, M. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiology* 152(4):1787-1795 (2010).

311. Zhang, H., Wei, L., Miao, H., Zhang, T. and Wang, C. Development and validation of genic-SSR markers in sesame by RNA-Seq. *BMC Genomics* 13(1):316 (2012).

312. Zhang, M., Zhao, H., Xie, S., Chen, J., Xu, Y., Wang, K., Zhao, H., Guan, H., Hu, X. and Jiao, Y. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proceedings of the National Academy of Sciences* 108(50):20042-20047 (2011).

313. Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B. and Xiang, J. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nature Biotechnology* 31(2):154-159 (2013).

314. Zhu, W., Schlueter, S. D. and Brendel, V. Refined annotation of the arabidopsis genome by complete expressed sequence tag mapping. *Plant Physiology* 132(2):469-484 (2003).

315. Zietkiewicz, E., Rafalski, A. and Labuda, D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20(2):176-183 (1994).

* Original not seen

Appendix 1. KEGG pathway categorization of assembled guar leaf transcriptome unigenes

KEGG pathway	Number of unigenes	Pathway ID
1.Metabolism		
1.1 Carbohydrate metabolism		
Glycolysis / Gluconeogenesis	257	map00010
Citrate cycle (TCA cycle)	123	map00020
Pentose phosphate pathway	164	map00030
Pentose and glucuronate interconversions	180	map00040
Fructose and mannose metabolism	121	map00051
Galactose metabolism	239	map00052
Ascorbate and aldarate metabolism	101	map00053
Starch and sucrose metabolism	667	map00500
Amino sugar and nucleotide sugar metabolism	299	map00520
Pyruvate metabolism	236	map00620
Glyoxylate and dicarboxylate metabolism	177	map00630
Propanoate metabolism	124	map00640
Butanoate metabolism	89	map00650
C5-Branched dibasic acid metabolism	23	map00660
Inositol phosphate metabolism	133	map00562
1.2 Energy metabolism		
Oxidative phosphorylation	211	map00190
Photosynthesis	3	map00195
Carbon fixation in photosynthetic organisms	158	map00710
Oxidative phosphorylation	211	map00190
Carbon fixation pathways in prokaryotes	173	map00720
Methane metabolism	161	map00680
Nitrogen metabolism	72	map00910
Sulfur metabolism	81	map00920
1.3 Lipid metabolism		
Fatty acid biosynthesis	78	map00061
Fatty acid elongation	43	map00062
Fatty acid degradation	143	map00071
Synthesis and degradation of ketone bodies	27	map00072
Cutin, suberine and wax biosynthesis	27	map00073
Steroid biosynthesis	24	map00100
Primary bile acid biosynthesis	34	map00120
Steroid hormone biosynthesis	59	map00140
Glycerolipid metabolism	171	map00561
Glycerophospholipid metabolism	221	map00564
Ether lipid metabolism	84	map00565
Sphingolipid metabolism	141	map00600
Arachidonic acid metabolism	49	map00590
Linoleic acid metabolism	35	map00591
alpha-Linolenic acid metabolism	98	map00592
Biosynthesis of unsaturated fatty acids	63	map01040

1.4 Nucleotide metabolism		
Purine metabolism	713	map00230
Pyrimidine metabolism	381	map00240
1.5 Amino acid metabolism		
Alanine, aspartate and glutamate metabolism	143	map00250
Glycine, serine and threonine metabolism	208	map00260
Cysteine and methionine metabolism	225	map00270
Valine, leucine and isoleucine biosynthesis	43	map00290
Valine, leucine and isoleucine degradation	169	map00280
Lysine biosynthesis	48	map00300
Arginine biosynthesis	86	map00220
Lysine degradation	127	map00310
Arginine and proline metabolism	112	map00330
Histidine metabolism	79	map00340
Tyrosine metabolism	121	map00350
Phenylalanine metabolism	129	map00360
Tryptophan metabolism	151	map00380
Phenylalanine, tyrosine and tryptophan biosynthesis	113	map00400
1.6 Metabolism of other amino acids		
beta-Alanine metabolism	78	map00410
Taurine and hypotaurine metabolism	18	map00430
Phosphonate and phosphinate metabolism	11	map00440
D-Glutamine and D-glutamate metabolism	11	map00471
Selenocompound metabolism	62	map00450
Cyanoamino acid metabolism	75	map00460
D-Arginine and D-ornithine metabolism	3	map00472
D-Alanine metabolism	6	map00473
Glutathione metabolism	171	map00480
1.7 Glycan biosynthesis and metabolism		
N-glycan biosynthesis	48	map00510
Various types of N-glycan biosynthesis	36	map00513
Other types of O-glycan biosynthesis	5	map00514
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	28	map00532
Glycosaminoglycan biosynthesis - heparan sulfate / heparin	35	map00534
Glycosaminoglycan degradation	79	map00531
Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	13	map00563
Glycosphingolipid biosynthesis - lacto and neolacto series	5	map00601
Glycosphingolipid biosynthesis - globo series	27	map00603
Glycosphingolipid biosynthesis - ganglio series	71	map00604
Lipopolysaccharide biosynthesis	13	map00540
Peptidoglycan biosynthesis	29	map00550
Other glycan degradation	150	map00511

1.8 Metabolism of cofactors and vitamins		
Thiamine metabolism	55	map00730
Riboflavin metabolism	25	map00740
Vitamin B6 metabolism	27	map00750
Nicotinate and nicotinamide metabolism	84	map00760
Pantothenate and CoA biosynthesis	69	map00770
Biotin metabolism	46	map00780
Lipoic acid metabolism	9	map00785
Folate biosynthesis	41	map00790
One carbon pool by folate	71	map00670
Porphyrin and chlorophyll metabolism	140	map00860
Ubiquinone and other terpenoid-quinone biosynthesis	44	map00130
1.9 Metabolism of terpenoids and polyketides		
Terpenoid backbone biosynthesis	84	map00900
Monoterpenoid biosynthesis	6	map00902
Retinol metabolism	63	map00830
Sesquiterpenoid and triterpenoid biosynthesis	4	map00909
Diterpenoid biosynthesis	20	map00904
Biosynthesis of vancomycin group antibiotics	7	map01055
Carotenoid biosynthesis	20	map00906
Zeatin biosynthesis	11	map00908
Insect hormone biosynthesis	4	map00981
Limonene and pinene degradation	24	map00903
Geraniol degradation	44	map00281
Brassinosteroid biosynthesis	1	map00905
Biosynthesis of ansamycins	11	map01051
Tetracycline biosynthesis	19	map00253
Polyketide sugar unit biosynthesis	12	map00523
Biosynthesis of siderophore group nonribosomal peptides	8	map01053
1.10 Biosynthesis of other secondary metabolites		
Phenylpropanoid biosynthesis	174	map00940
Stilbenoid, diarylheptanoid and gingerol biosynthesis	10	map00945
Flavonoid biosynthesis	76	map00941
Flavone and flavonol biosynthesis	14	map00944
Anthocyanin biosynthesis	1	map00942
Betalain biosynthesis	11	map00965
Isoflavonoid biosynthesis	22	map00943
Indole alkaloid biosynthesis	19	map00901
Monobactam biosynthesis	39	map00261
Isoquinoline alkaloid biosynthesis	49	map00950
Carbapenem biosynthesis	7	map00332
Tropane, piperidine and pyridine alkaloid biosynthesis	41	map00960
Caffeine metabolism	22	map00232
Glucosinolate biosynthesis	7	map00966
Penicillin and cephalosporin biosynthesis	5	map00311

Streptomycin biosynthesis	49	map00521
Butirosin and neomycin biosynthesis	15	map00524
Novobiocin biosynthesis	31	map00401
Aflatoxin biosynthesis	19	map00254
1.11 Xenobiotics biodegradation and metabolism		
Benzoate degradation	60	map00362
Aminobenzoate degradation	132	map00627
Fluorobenzoate degradation	10	map00364
Chloroalkane and chloroalkene degradation	44	map00625
Chlorocyclohexane and chlorobenzene degradation	15	map00361
Toluene degradation	34	map00623
Ethylbenzene degradation	11	map00642
Styrene degradation	20	map00643
Polycyclic aromatic hydrocarbon degradation	4	map00624
Atrazine degradation	7	map00791
Xylene degradation	4	map00622
Caprolactam degradation	43	map00930
Naphthalene degradation	21	map00626
Steroid degradation	19	map00984
Metabolism of xenobiotics by cytochrome P450	110	map00980
Drug metabolism - cytochrome P450	133	map00982
Drug metabolism - other enzymes	72	map00983
2. Genetic Information Processing		
Translation		
Aminoacyl-tRNA biosynthesis	132	map00970
3. Environmental Information Processing		
Signal transduction		
Phosphatidylinositol signaling system	167	map04070
mTOR signaling pathway	40	map04150
4. Organismal Systems		
Immune system		
T cell receptor signaling pathway	208	map04660
5. Human Diseases		
Drug resistance		
beta-Lactam resistance	3	map00312

Appendix 2. Detailed list of *in silico* identified polymorphic SSR markers

ID	SSR type	SSR	Size	Start	End
comp8464_c0_seq1	p3	(TTC)5	15	314	328
comp9618_c0_seq1	p1	(T)10	10	126	135
comp11342_c0_seq1	p2	(GA)6	12	2202	2213
comp12865_c0_seq1	p2	(GA)6	12	2202	2213
comp13809_c0_seq1	p2	(AG)9	18	1569	1586
comp13950_c0_seq1	p1	(T)12	12	124	135
comp15218_c0_seq1	p1	(T)16	16	97	112
comp16454_c0_seq2	p1	(A)13	13	359	371
comp16540_c0_seq1	p1	(T)12	12	85	96
comp16997_c0_seq1	p1	(A)10	10	105	114
comp17604_c0_seq1	p1	(A)10	10	97	106
comp17651_c0_seq1	p1	(A)11	11	1990	2000
comp17819_c0_seq1	p3	(GGA)5	15	655	669
comp17919_c0_seq1	p1	(A)10	10	54	63
comp18034_c0_seq1	p2	(GA)6	12	39	50
comp18111_c0_seq2	p2	(AG)6	12	8	19
comp19155_c0_seq1	p2	(AT)7	14	60	73
comp19636_c0_seq2	p1	(A)10	10	96	105
comp19724_c0_seq1	p2	(AG)6	12	148	159
comp20242_c0_seq1	p1	(T)14	14	222	235
comp20346_c0_seq1	c	(A)10ttattggatgccacatgtcaacatctcattgagtggcgtctgctaggaggcgccattcaatgaggtgctaccacgtgacagcaccagattttc(A)11	115	69	183
comp20690_c0_seq1	p1	(T)13	13	485	497
comp21982_c0_seq1	p1	(A)10	10	142	151
comp22227_c0_seq1	p1	(A)10	10	1480	1489
comp22347_c0_seq1	p3	(CAC)5	15	707	721
comp22851_c0_seq1	p1	(T)13	13	47	59
comp22851_c0_seq1	p1	(T)13	13	47	59
comp23073_c0_seq1	p3	(GGA)5	15	437	451
comp23183_c0_seq1	p1	(A)13	13	73	85
comp23487_c0_seq1	p1	(A)10	10	627	636
comp24716_c0_seq1	p3	(TTC)5	15	56	70
comp24793_c0_seq2	p1	(A)10	10	189	198
comp25073_c0_seq1	p1	(T)10	10	708	717
comp25073_c0_seq2	p1	(T)10	10	630	639
comp25090_c0_seq1	p2	(GA)7	14	2611	2624
comp25313_c0_seq1	p1	(A)13	13	147	159
comp25594_c0_seq1	p1	(A)12	12	1186	1197

comp26054_c0_seq1	p1	(T)10	10	479	488
comp26054_c0_seq1	p1	(T)10	10	479	488
comp27018_c0_seq1	p1	(A)11	11	1758	1768
comp27195_c0_seq1	p3	(TCA)5	15	172	186
comp27233_c0_seq1	p1	(T)16	16	106	121
comp27233_c0_seq2	p1	(T)16	16	106	121
comp27569_c0_seq1	p1	(T)11	11	930	940
comp27670_c0_seq2	c	(GT)6ttggccgat(A)11	32	167	198
comp28095_c0_seq1	p1	(A)13	13	79	91
comp28123_c0_seq2	p1	(A)13	13	46	58
comp28364_c0_seq1	p1	(T)13	13	34	46
comp28512_c0_seq1	p2	(AG)8	16	317	332
comp29718_c0_seq4	p1	(A)12	12	927	938
comp29859_c0_seq2	p1	(A)10	10	65	74
comp29959_c0_seq2	p3	(TCA)6	18	103	120
comp29959_c0_seq3	p3	(TCA)6	18	103	120
comp30167_c0_seq1	p1	(A)11	11	1285	1295
comp30202_c0_seq2	p1	(A)13	13	47	59
comp30203_c0_seq1	p2	(AG)6	12	48	59
comp30204_c0_seq2	p1	(T)10	10	55	64
comp30307_c0_seq1	p1	(A)10	10	37	46
comp30576_c0_seq1	p1	(A)11	11	193	203
comp30770_c0_seq1	p2	(CT)6	12	169	180
comp30801_c0_seq3	p1	(T)10	10	2349	2358
comp31001_c0_seq2	p1	(T)11	11	174	184
comp31381_c0_seq4	p1	(A)16	16	182	197
comp31433_c0_seq1	p3	(GAA)6	18	3364	3381
comp31433_c0_seq2	p3	(GAA)6	18	3269	3286
comp31433_c0_seq3	p3	(GAA)6	18	3193	3210
comp31433_c0_seq4	p3	(GAA)6	18	3098	3115
comp32419_c0_seq1	p2	(AT)8	16	2278	2293
comp32514_c0_seq2	p1	(A)11	11	91	101
comp32762_c0_seq3	p1	(A)10	10	627	636
comp32802_c0_seq1	p1	(T)15	15	178	192
comp32846_c0_seq1	p1	(T)10	10	349	358
comp32856_c0_seq7	p1	(T)10	10	371	380
comp33025_c0_seq2	p3	(GAT)5	15	2124	2138
comp33036_c0_seq2	p1	(A)12	12	3337	3348
comp33091_c0_seq1	p1	(A)12	12	556	567
comp33120_c0_seq1	p1	(A)14	14	59	72
comp33127_c0_seq1	p1	(A)13	13	270	282
comp33127_c0_seq1	p1	(A)13	13	270	282
comp33139_c1_seq2	p1	(A)12	12	190	201

comp33145_c0_seq13	p1	(A)10	10	196	205
comp33193_c0_seq5	p2	(AT)9	18	101	118
comp33352_c0_seq1	p1	(T)14	14	109	122
comp33366_c0_seq1	p1	(A)10	10	461	470
comp33386_c0_seq1	p3	(GTT)5	15	87	101
comp33415_c0_seq18	p3	(TGG)5	15	256	270
comp33416_c0_seq1	c	(C)13(A)11	24	49	72
comp33419_c0_seq1	p1	(A)10	10	1664	1673
comp33419_c0_seq12	p1	(T)10	10	139	148
comp33450_c0_seq7	c	(A)14caaccaacgcg(CT)6	37	81	117
comp33451_c0_seq1	p1	(T)10	10	1119	1128
comp33482_c0_seq13	p1	(T)12	12	547	558
comp33514_c0_seq11	p1	(A)11	11	1759	1769
comp33521_c0_seq13	p2	(AT)6	12	2693	2704
comp33523_c0_seq2	p3	(CTG)5	15	995	1009
comp33545_c0_seq3	p1	(T)10	10	151	160
comp33601_c0_seq26	p1	(A)13	13	237	249
comp33615_c0_seq6	p1	(T)11	11	315	325
comp33637_c0_seq4	p3	(ATG)5	15	214	228
comp34411_c0_seq1	p1	(T)10	10	114	123
comp35672_c0_seq1	p1	(T)11	11	123	133
comp37705_c0_seq1	p1	(A)17	17	73	89
comp37830_c0_seq1	c	(A)15gagaatgcaattgtaatcgtcttcgcctgaaaaaacatgcgtgagtttct(TC)6	77	1772	1848
comp38408_c0_seq1	p1	(A)10	10	79	88
comp39203_c0_seq1	p1	(A)12	12	136	147
comp39524_c0_seq1	p1	(T)12	12	739	750
comp40342_c0_seq1	p4	(TCTT)6	24	1617	1640
comp50788_c0_seq1	p1	(T)11	11	467	477
comp65519_c0_seq1	p1	(A)14	14	417	430
comp72045_c0_seq1	p3	(AAG)6	18	439	456
comp79376_c0_seq1	p1	(T)10	10	411	420
comp86705_c0_seq1	p1	(T)10	10	116	125
comp87801_c0_seq1	p2	(TC)10	20	146	165
comp89834_c0_seq1	p1	(A)14	14	287	300
comp91887_c0_seq1	c	(A)11caaaaatggaagccagtggtacccatggcatgatttttgttcattccaagttcgcagtt(TG)8	87	820	906
comp95348_c0_seq1	p1	(T)10	10	116	125
comp106042_c0_seq1	p1	(A)10	10	48	57
comp111064_c0_seq1	p1	(A)11	11	235	245
comp121224_c0_seq1	c	(A)18cgtacgcaacttttg(T)13	47	188	234

comp121500_c0_seq1	p1	(T)11	11	273	283
comp122195_c0_seq1	p3	(TCT)6	18	228	245
comp128817_c0_seq1	p1	(T)12	12	313	324
comp135740_c0_seq1	p1	(A)10	10	63	72
comp146132_c0_seq1	p3	(GAG)5	15	238	252
comp150496_c0_seq1	p1	(T)11	11	325	335
comp184033_c0_seq1	p1	(T)12	12	302	313
comp185516_c0_seq1	p2	(GA)10	20	130	149
comp187787_c0_seq1	p1	(T)13	13	309	321
comp195314_c0_seq1	p3	(GTG)5	15	118	132
comp195698_c0_seq1	p1	(T)10	10	81	90
comp196550_c0_seq1	p2	(GT)6	12	135	146
comp200502_c0_seq1	p1	(T)11	11	82	92
comp206092_c0_seq1	p1	(T)11	11	163	173
comp217974_c0_seq1	p1	(A)10	10	75	84
comp218195_c0_seq1	p3	(GAA)6	18	135	152
comp276179_c0_seq1	p1	(T)10	10	61	70
comp279255_c0_seq1	p3	(ATC)5	15	248	262
comp283457_c0_seq1	p1	(T)11	11	127	137
comp289131_c0_seq1	p1	(T)10	10	120	129
comp312043_c0_seq1	p1	(T)10	10	97	106
comp320230_c0_seq1	p1	(T)10	10	284	293
comp362625_c0_seq1	p1	(A)10	10	170	179
comp390231_c0_seq1	p1	(T)12	12	113	124
comp546252_c0_seq1	p1	(A)11	11	79	89
comp556501_c0_seq1	p1	(T)11	11	224	234

Appendix 3: Details of single nucleotide polymorphisms (SNPs) detected in guar varieties. Genotype1 represents M-83 and Genotype2 represents RGC-1066 variety.

Sr. no.	Unigene_ID	Position	Reference	Genotype1	Genotype2	Genotype1_depth	Genotype2_depth
1.	comp33629_c0_seq10	85	C	T	C	6/6	6/6
2.	comp28475_c0_seq2	435	G	C	G	4/6	10/12
3.	comp28475_c0_seq2	441	C	A	C	4/6	10/12
4.	comp19489_c0_seq1	368	T	A	T	7/10	14/20
5.	comp19489_c0_seq1	374	C	A	C	7/10	14/20
6.	comp19489_c0_seq1	378	C	T	C	7/9	14/20
7.	comp19489_c0_seq1	383	C	T	C	9/9	12/16
8.	comp33555_c0_seq52	488	A	G	A	4/6	18/22
9.	comp33696_c0_seq6	333	G	C	G	12/12	20/28
10.	comp33142_c0_seq1	361	T	C	T	12/15	14/16
11.	comp33142_c0_seq1	367	A	G	A	10/14	14/18
12.	comp33142_c0_seq1	371	C	G	C	12/14	14/18
13.	comp25528_c0_seq1	138	A	A	G	4/6	4/6
14.	comp25528_c0_seq1	222	T	C	T	11/11	8/8
15.	comp31439_c0_seq2	1392	G	A	G	10/10	9/9
16.	comp31788_c0_seq3	93	G	G	A	42/63	4/4
17.	comp31788_c0_seq3	212	C	C	G	49/72	2/2
18.	comp31788_c0_seq3	244	C	C	T	74/80	4/4
19.	comp31788_c0_seq3	264	G	G	A	74/84	4/4
20.	comp31788_c0_seq3	267	G	G	A	63/84	4/4
21.	comp31788_c0_seq3	345	C	C	T	14/21	2/2
22.	comp33175_c0_seq20	668	C	T	C	4/6	12/12
23.	comp33175_c0_seq20	673	C	T	C	4/6	10/10
24.	comp33175_c0_seq20	853	A	A	G	6/8	6/8
25.	comp28436_c0_seq1	130	T	T	C	12/15	4/6
26.	comp19361_c0_seq1	4	G	A	C	4/6	2/2
27.	comp27348_c1_seq2	123	C	A	C	25/30	4/6
28.	comp30274_c0_seq2	576	T	T	A	8/8	14/15
29.	comp30274_c0_seq2	582	T	T	C	9/9	12/14
30.	comp33625_c0_seq3	223	T	A	T	15/22	27/27
31.	comp33625_c0_seq3	233	A	G	A	15/22	28/28
32.	comp18038_c0_seq1	124	A	C	A	4/6	2/2
33.	comp29350_c0_seq1	2571	T	G	T	18/24	18/20
34.	comp22728_c0_seq1	80	A	A	G	7/7	4/6
35.	comp24408_c0_seq1	517	C	A	C	14/14	140/140
36.	comp25714_c0_seq1	103	T	C	T	6/9	4/4
37.	comp31696_c0_seq2	40	A	A	G	27/36	2/2
38.	comp31696_c0_seq2	41	T	T	C	27/35	4/4
39.	comp4886_c0_seq1	130	G	A	G	4/6	8/8
40.	comp32419_c0_seq1	573	T	C	T	20/29	5/7
41.	comp32419_c0_seq1	581	C	T	C	20/29	5/5
42.	comp32419_c0_seq1	585	C	T	C	18/27	5/5
43.	comp32419_c0_seq1	615	G	T	G	16/23	6/6
44.	comp32419_c0_seq1	620	C	T	C	16/23	6/6
45.	comp32419_c0_seq1	741	T	T	A	11/11	10/11
46.	comp32419_c0_seq1	764	A	A	G	28/29	8/11
47.	comp32419_c0_seq1	1600	T	T	C	8/8	4/6
48.	comp30513_c0_seq2	175	C	C	T	30/45	6/8
49.	comp30513_c0_seq2	199	G	G	T	44/49	4/6
50.	comp30513_c0_seq2	275	A	A	C	20/30	2/2
51.	comp27401_c0_seq1	212	C	T	C	4/6	6/6
52.	comp39160_c0_seq1	365	A	A	C	102/104	22/22
53.	comp33433_c0_seq1	1738	T	C	T	6/6	18/18
54.	comp24492_c0_seq1	3	A	A	C	7/9	2/2
55.	comp30270_c0_seq3	244	A	G	A	6/8	16/23
56.	comp30270_c0_seq3	245	C	G	C	5/7	14/21
57.	comp30270_c0_seq3	248	A	C	A	6/8	15/20
58.	comp30270_c0_seq3	250	C	G	C	4/6	13/18
59.	comp30270_c0_seq3	253	C	A	C	5/7	14/20
60.	comp30270_c0_seq3	259	A	T	A	4/6	14/15
61.	comp27457_c0_seq1	289	G	G	C	4/6	2/2
62.	comp135165_c0_seq1	132	C	C	T	12/12	14/14
63.	comp18448_c0_seq1	57	T	A	T	6/6	14/14
64.	comp19065_c0_seq1	3	A	G	A	4/6	2/2
65.	comp30606_c0_seq2	170	A	A	G	217/270	4/4
66.	comp30606_c0_seq2	184	T	T	G	224/276	4/4
67.	comp30606_c0_seq2	195	T	T	G	237/279	4/6
68.	comp30606_c0_seq2	198	C	C	A	241/281	4/6
69.	comp30606_c0_seq2	299	A	A	T	80/115	2/2
70.	comp30606_c0_seq2	303	T	T	C	79/113	2/2
71.	comp29148_c0_seq1	935	T	T	A	11/11	10/12

72.	comp31316_c0_seq1	145	A	G	A	27/40	12/16
73.	comp31316_c0_seq1	483	T	T	C	9/13	6/6
74.	comp206188_c0_seq1	258	G	A	G	8/8	2/2
75.	comp25626_c0_seq1	418	G	A	G	8/8	22/22
76.	comp32688_c0_seq7	572	T	C	T	4/6	18/20
77.	comp33701_c1_seq1	4066	A	C	A	9/9	14/14
78.	comp33701_c1_seq1	5392	G	A	G	6/6	20/20
79.	comp32183_c0_seq3	243	G	C	G	6/6	12/12
80.	comp25297_c0_seq1	104	C	C	T	6/7	6/8
81.	comp39027_c0_seq1	199	T	T	G	23/23	8/12
82.	comp14783_c0_seq1	76	T	T	A	8/8	4/4
83.	comp17703_c0_seq1	371	A	G	A	16/16	20/20
84.	comp33443_c0_seq2	1033	A	G	A	45/63	42/63
85.	comp32948_c0_seq3	1370	T	C	T	6/6	12/12
86.	comp33261_c0_seq10	170	A	A	G	57/75	21/30
87.	comp19346_c0_seq1	3	C	C	A	4/6	6/9
88.	comp6320_c0_seq1	29	G	A	G	6/6	6/6
89.	comp32194_c0_seq2	388	A	A	T	11/11	4/6
90.	comp32194_c0_seq2	392	G	G	C	10/10	4/6
91.	comp32809_c0_seq3	25	A	T	A	6/9	16/18
92.	comp32809_c0_seq3	38	T	C	T	9/12	20/22
93.	comp33230_c0_seq2	134	A	C	A	26/26	16/16
94.	comp33230_c0_seq2	1936	C	G	C	12/12	24/24
95.	comp33230_c0_seq2	4566	A	G	A	6/6	9/9
96.	comp33451_c0_seq8	119	G	A	G	30/44	30/30
97.	comp13298_c0_seq1	131	G	G	A	5/7	6/6
98.	comp31711_c0_seq5	55	T	T	C	7/9	2/2
99.	comp31711_c0_seq5	56	A	A	G	7/8	2/2
100.	comp11530_c0_seq1	272	C	C	T	9/9	10/10
101.	comp7929_c0_seq1	251	G	G	C	15/15	6/6
102.	comp32725_c0_seq10	324	G	A	G	12/16	18/21
103.	comp32725_c0_seq10	334	G	C	G	10/15	18/21
104.	comp32725_c0_seq10	344	T	C	T	10/14	20/22
105.	comp32725_c0_seq10	349	T	A	T	9/13	18/18
106.	comp29710_c1_seq2	24	T	C	T	45/68	4/4
107.	comp32797_c0_seq3	55	G	G	C	10/11	4/6
108.	comp32797_c0_seq3	84	A	A	G	12/13	6/8
109.	comp32797_c0_seq3	123	T	C	T	11/11	6/6
110.	comp32797_c0_seq3	126	T	C	T	11/12	8/8
111.	comp32797_c0_seq3	134	A	G	A	10/10	6/8
112.	comp32797_c0_seq3	149	C	T	C	8/12	8/8
113.	comp32797_c0_seq3	269	C	T	C	14/21	8/11
114.	comp32421_c0_seq23	146	C	T	C	8/8	8/11
115.	comp32691_c0_seq2	1	C	C	G	8/8	4/4
116.	comp33051_c0_seq1	336	T	C	T	20/20	28/28
117.	comp33521_c0_seq1	6633	T	C	T	8/11	13/14
118.	comp33521_c0_seq1	6751	T	C	T	12/18	12/18
119.	comp30544_c0_seq1	679	T	T	G	4/6	4/4
120.	comp32018_c0_seq2	338	T	A	T	7/7	23/23
121.	comp32018_c0_seq2	1165	C	C	G	11/12	18/20
122.	comp7492_c0_seq1	126	T	A	T	7/10	6/8
123.	comp33724_c0_seq11	14	A	A	C	16/24	12/16
124.	comp33724_c0_seq11	155	A	A	C	61/85	50/66
125.	comp33724_c0_seq11	247	C	A	C	32/41	40/59
126.	comp32470_c0_seq1	721	A	A	C	11/13	6/6
127.	comp32470_c0_seq1	722	T	T	C	10/13	6/6
128.	comp32235_c0_seq1	534	A	G	A	6/6	20/20
129.	comp14531_c0_seq1	102	C	T	C	6/6	6/6
130.	comp27791_c0_seq1	4	G	C	G	4/6	4/6
131.	comp25257_c0_seq1	159	G	G	A	12/12	10/10
132.	comp30469_c0_seq1	247	A	A	C	25/25	24/24
133.	comp26343_c0_seq1	53	T	G	T	6/7	8/8
134.	comp26343_c0_seq1	57	A	G	A	6/7	8/8
135.	comp26343_c0_seq1	82	C	C	T	7/8	8/10
136.	comp26343_c0_seq1	89	G	G	C	7/8	6/8
137.	comp26343_c0_seq1	91	A	A	G	7/8	6/8
138.	comp26343_c0_seq1	92	T	T	A	7/8	6/8
139.	comp33026_c0_seq2	1037	T	T	C	16/19	4/6
140.	comp33026_c0_seq2	1043	T	T	C	16/19	4/6
141.	comp12618_c0_seq1	308	A	C	A	6/6	11/13
142.	comp22404_c0_seq1	296	A	G	A	7/7	10/12
143.	comp33510_c0_seq3	381	C	C	T	21/26	2/2
144.	comp27417_c0_seq1	420	A	A	C	4/6	4/6
145.	comp27417_c0_seq1	424	C	C	T	4/6	4/6
146.	comp27417_c0_seq1	435	A	A	G	4/6	4/6

X

147.	comp27392_c0_seq1	88	C	C	T	11/16	2/2
148.	comp27392_c0_seq1	94	C	C	T	10/15	4/4
149.	comp27392_c0_seq1	159	T	T	C	10/14	6/8
150.	comp27392_c0_seq1	160	G	G	A	12/16	6/8
151.	comp27392_c0_seq1	181	T	T	G	8/10	4/6
152.	comp33145_c0_seq15	510	A	G	A	11/15	6/8
153.	comp33145_c0_seq15	544	A	G	A	15/15	6/8
154.	comp33145_c0_seq15	545	G	A	G	12/14	6/8
155.	comp18052_c0_seq1	189	G	G	A	33/33	8/10
156.	comp18052_c0_seq1	216	T	T	A	7/8	2/2
157.	comp33493_c0_seq3	523	C	C	T	6/9	4/6
158.	comp33718_c0_seq14	129	A	T	A	7/10	8/12
159.	comp31215_c0_seq2	1136	T	C	T	7/10	11/13
160.	comp32665_c0_seq2	200	C	C	A	6/6	8/8
161.	comp14692_c0_seq1	122	C	C	T	19/19	12/12
162.	comp14692_c0_seq1	148	G	G	A	34/34	16/16
163.	comp28124_c0_seq1	26	G	C	G	23/25	2/2
164.	comp27295_c0_seq1	593	T	C	T	14/14	24/25
165.	comp33501_c0_seq27	5	G	G	C	5/6	4/6
166.	comp21049_c0_seq1	112	A	A	C	10/12	6/8
167.	comp21049_c0_seq1	116	G	G	A	10/14	6/8
168.	comp21049_c0_seq1	118	G	G	T	12/15	6/8
169.	comp28862_c0_seq1	452	A	G	A	15/22	22/22
170.	comp28862_c0_seq1	470	T	C	T	21/28	24/32
171.	comp147498_c0_seq1	32	C	C	G	6/8	6/6
172.	comp33679_c0_seq4	2440	T	T	G	6/8	6/9
173.	comp33679_c0_seq4	2450	T	C	T	4/6	7/7
174.	comp33107_c0_seq3	149	C	C	T	21/24	8/12
175.	comp27242_c0_seq1	731	T	C	T	4/6	16/16
176.	comp19781_c0_seq1	98	T	T	C	7/8	2/2
177.	comp19781_c0_seq1	105	T	T	A	7/8	2/2
178.	comp19781_c0_seq1	120	A	A	G	8/9	2/2
179.	comp19781_c0_seq1	121	G	G	A	8/9	2/2
180.	comp17479_c0_seq1	197	G	G	T	9/9	4/4
181.	comp78535_c0_seq1	5	T	T	A	7/7	2/2
182.	comp29658_c0_seq1	87	T	T	A	4/6	8/10
183.	comp32184_c0_seq2	563	A	G	A	6/6	14/20
184.	comp12804_c0_seq1	821	G	T	G	7/7	12/12
185.	comp32738_c0_seq11	175	C	C	T	6/9	16/20
186.	comp15773_c0_seq1	105	G	G	T	7/10	2/2
187.	comp15773_c0_seq1	117	C	C	T	10/13	2/2
188.	comp15773_c0_seq1	130	G	G	C	6/9	2/2
189.	comp15773_c0_seq1	160	T	T	G	7/9	2/2
190.	comp31499_c0_seq1	1356	C	A	C	13/13	18/18
191.	comp32767_c0_seq1	409	G	G	C	9/13	4/4
192.	comp33723_c0_seq1	165	A	C	A	8/10	34/40
193.	comp117521_c0_seq1	352	T	C	T	6/8	4/6
194.	comp31266_c0_seq1	314	T	C	T	13/19	26/39
195.	comp31266_c0_seq1	334	C	T	C	13/19	31/44
196.	comp9297_c0_seq1	25	G	G	C	10/10	8/8
197.	comp33521_c0_seq13	1095	T	C	T	6/8	7/9
198.	comp33555_c0_seq55	154	T	C	T	5/7	7/9
199.	comp17686_c0_seq1	17	T	A	T	7/9	2/2
200.	comp33555_c0_seq32	207	A	A	C	6/7	17/25
201.	comp33555_c0_seq32	646	C	C	G	6/6	10/14
202.	comp33596_c0_seq24	2141	T	T	C	10/12	11/16
203.	comp33232_c0_seq3	286	A	A	G	6/6	5/5
204.	comp31609_c0_seq1	244	C	C	G	8/8	4/6
205.	comp32927_c0_seq1	832	C	T	C	7/7	20/20
206.	comp32927_c0_seq1	1066	T	C	T	8/8	32/32
207.	comp33471_c0_seq94	111	T	T	C	20/20	14/18
208.	comp33025_c0_seq7	625	T	A	T	6/9	21/25
209.	comp33025_c0_seq7	631	C	A	C	6/9	19/22
210.	comp33025_c0_seq7	639	G	A	G	6/9	13/16
211.	comp27747_c0_seq1	574	A	A	T	8/9	6/6
212.	comp33594_c0_seq1	1743	G	A	G	6/9	5/5
213.	comp33621_c0_seq1	1270	T	G	T	6/9	19/26
214.	comp33521_c0_seq16	1095	T	C	T	6/8	14/17
215.	comp33521_c0_seq16	1106	T	C	T	6/8	14/15
216.	comp31075_c0_seq1	154	T	T	C	14/21	22/30
217.	comp26865_c0_seq1	429	T	A	T	8/8	18/18
218.	comp119200_c0_seq1	197	A	A	G	9/10	4/6
219.	comp19528_c0_seq1	219	A	A	T	10/10	6/6
220.	comp31724_c0_seq1	401	A	G	A	9/13	18/26
221.	comp33123_c0_seq11	381	G	G	A	11/13	4/6

222.	comp33123_c0_seq11	386	C	C	T	11/13	4/6
223.	comp33123_c0_seq11	392	C	C	T	10/12	4/6
224.	comp33123_c0_seq11	394	G	G	T	9/11	4/6
225.	comp33123_c0_seq11	401	A	A	G	6/7	4/6
226.	comp33123_c0_seq11	403	T	T	C	6/7	4/6
227.	comp32931_c0_seq1	204	T	T	G	41/60	42/53
228.	comp31082_c0_seq1	1629	A	A	T	21/21	14/14
229.	comp32794_c0_seq1	862	C	T	C	6/6	4/6
230.	comp32794_c0_seq1	863	A	C	A	6/6	4/6
231.	comp33327_c0_seq10	173	G	G	A	8/10	8/12
232.	comp30341_c0_seq2	325	C	G	C	10/11	11/11
233.	comp32935_c0_seq2	641	T	T	C	7/7	4/4
234.	comp26235_c0_seq1	163	A	G	A	4/6	12/16
235.	comp33653_c0_seq8	167	C	C	T	9/12	6/9
236.	comp33523_c0_seq2	448	G	G	C	6/8	2/2
237.	comp33523_c0_seq2	457	T	T	A	8/9	2/2
238.	comp33523_c0_seq2	524	C	T	C	14/20	4/6
239.	comp33523_c0_seq2	1157	G	A	G	7/8	8/11
240.	comp33523_c0_seq2	1166	G	T	G	7/8	11/11
241.	comp33523_c0_seq2	1168	A	C	A	7/8	11/11
242.	comp33523_c0_seq2	1169	G	T	G	7/8	11/11
243.	comp33523_c0_seq2	1175	G	A	G	7/7	11/11
244.	comp33523_c0_seq2	1463	A	G	A	5/7	16/24
245.	comp29842_c0_seq3	235	T	T	C	48/48	15/16
246.	comp32985_c0_seq1	1481	T	A	T	6/6	18/18
247.	comp33327_c0_seq8	224	A	A	T	6/9	2/2
248.	comp32403_c0_seq1	970	C	A	C	9/9	23/23
249.	comp33521_c0_seq11	1047	T	C	T	6/7	11/13
250.	comp33521_c0_seq11	1050	T	C	T	6/7	9/11
251.	comp33521_c0_seq11	1095	T	C	T	5/6	5/7
252.	comp33521_c0_seq11	1106	T	C	T	4/6	5/7
253.	comp31991_c0_seq6	405	G	A	G	6/7	4/4
254.	comp32375_c0_seq1	2619	T	C	T	9/13	14/14
255.	comp32660_c0_seq1	374	C	T	C	11/16	36/36
256.	comp33417_c1_seq11	138	G	G	A	10/12	2/2
257.	comp33417_c1_seq11	142	T	T	C	9/11	2/2
258.	comp33417_c1_seq11	144	C	C	T	7/10	2/2
259.	comp33417_c1_seq11	281	T	T	A	11/16	4/6
260.	comp33417_c1_seq11	427	C	C	A	77/82	4/4
261.	comp33417_c1_seq11	456	T	T	C	52/76	2/2
262.	comp31771_c0_seq4	439	G	C	G	6/6	12/12
263.	comp23260_c0_seq1	34	C	C	T	6/7	5/5
264.	comp23260_c0_seq1	38	A	A	T	7/7	4/5
265.	comp29420_c0_seq1	4	A	A	C	6/6	4/4
266.	comp21519_c0_seq1	131	G	G	A	8/8	10/14
267.	comp21519_c0_seq1	133	G	G	T	8/8	10/14
268.	comp21519_c0_seq1	135	T	T	A	8/8	10/14
269.	comp21519_c0_seq1	137	G	G	C	8/8	8/12
270.	comp5733_c0_seq1	102	G	C	G	23/30	38/38
271.	comp25173_c0_seq1	294	A	G	A	8/9	6/8
272.	comp32975_c0_seq1	354	G	A	G	17/22	24/36
273.	comp209670_c0_seq1	102	T	G	T	6/6	4/4
274.	comp29509_c0_seq6	447	T	C	T	5/6	2/2
275.	comp18514_c0_seq1	46	C	T	C	6/9	12/18
276.	comp33315_c0_seq1	1172	C	C	T	7/8	6/6
277.	comp24288_c0_seq1	726	C	C	T	11/11	4/4
278.	comp31341_c0_seq1	608	T	C	T	4/6	30/32
279.	comp31341_c0_seq1	613	G	T	G	5/7	27/35
280.	comp33384_c0_seq6	1340	C	T	C	7/7	15/15
281.	comp32723_c0_seq1	3072	T	T	G	14/14	16/18
282.	comp27722_c0_seq1	45	C	C	G	5/7	2/2
283.	comp27722_c0_seq1	54	A	A	G	5/7	2/2
284.	comp27722_c0_seq1	93	C	C	T	9/11	2/2
285.	comp27722_c0_seq1	96	C	C	G	9/11	2/2
286.	comp27722_c0_seq1	98	T	T	C	8/10	2/2
287.	comp27722_c0_seq1	243	C	G	C	10/14	2/2
288.	comp32421_c0_seq14	320	A	T	A	12/12	6/6
289.	comp33273_c0_seq10	725	A	G	A	4/6	8/8
290.	comp33273_c0_seq10	731	C	T	C	4/6	8/8
291.	comp33273_c0_seq10	732	A	G	A	4/6	8/8
292.	comp26994_c0_seq2	100	A	G	A	6/7	4/4
293.	comp33727_c0_seq16	162	A	A	G	21/28	2/2
294.	comp32911_c0_seq2	257	A	A	G	6/6	7/7
295.	comp31849_c0_seq6	219	T	T	C	5/6	3/4
296.	comp31849_c0_seq6	275	G	G	A	6/9	7/9

297.	comp31849_c0_seq6	732	C	T	C	7/10	6/7
298.	comp31849_c0_seq6	853	A	C	A	4/6	12/16
299.	comp85044_c0_seq1	103	T	T	C	34/34	4/4
300.	comp85044_c0_seq1	118	T	T	C	34/36	4/4
301.	comp85044_c0_seq1	137	T	T	C	55/61	4/4
302.	comp85044_c0_seq1	164	A	A	T	53/59	4/4
303.	comp30257_c0_seq1	1103	T	C	T	17/24	18/22
304.	comp30257_c0_seq1	1139	A	A	T	14/16	16/21
305.	comp30257_c0_seq1	1165	T	G	T	10/14	16/20
306.	comp30257_c0_seq1	1166	G	C	G	9/12	16/22
307.	comp30257_c0_seq1	1174	A	G	A	10/15	20/26
308.	comp41632_c0_seq1	4	A	A	G	6/7	2/2
309.	comp33718_c0_seq2	221	A	G	A	30/45	52/76
310.	comp33718_c0_seq2	829	G	A	G	6/8	16/20
311.	comp33718_c0_seq2	851	G	A	G	5/7	20/24
312.	comp33718_c0_seq2	852	A	G	A	5/7	20/24
313.	comp33718_c0_seq2	885	C	T	C	10/12	28/40
314.	comp32402_c0_seq1	560	G	G	A	7/7	6/6
315.	comp32402_c0_seq1	730	A	C	A	14/14	24/24
316.	comp32402_c0_seq1	751	G	T	G	14/14	26/26
317.	comp32402_c0_seq1	763	T	A	T	14/14	24/24
318.	comp32402_c0_seq1	764	G	A	G	14/14	24/24
319.	comp32631_c0_seq1	917	C	T	C	10/10	57/57
320.	comp113769_c0_seq1	139	C	G	C	4/6	10/10
321.	comp29707_c0_seq1	342	G	A	G	10/10	31/31
322.	comp9414_c0_seq1	820	T	T	C	14/17	4/4
323.	comp23884_c0_seq1	210	C	C	T	8/10	4/6
324.	comp31217_c0_seq3	556	T	T	G	13/13	24/26
325.	comp31217_c0_seq3	599	C	T	C	14/15	31/33
326.	comp110718_c0_seq1	50	C	G	C	7/7	8/8
327.	comp32831_c0_seq7	426	C	T	C	6/9	17/21
328.	comp33566_c0_seq2	327	T	C	T	7/7	21/21
329.	comp33566_c0_seq2	481	T	C	T	23/27	20/29
330.	comp33566_c0_seq2	485	A	T	A	21/25	18/27
331.	comp33566_c0_seq2	486	A	C	A	19/23	19/28
332.	comp33566_c0_seq2	489	G	C	G	18/22	19/25
333.	comp33566_c0_seq2	1742	T	T	C	22/32	30/42
334.	comp26073_c0_seq1	282	C	T	C	9/9	4/4
335.	comp33107_c0_seq2	149	C	T	C	11/16	12/18
336.	comp33700_c0_seq18	178	T	T	C	11/14	25/35
337.	comp32660_c0_seq2	175	G	G	A	6/8	6/9
338.	comp32660_c0_seq2	179	C	C	G	8/10	9/13
339.	comp32660_c0_seq2	190	A	A	G	7/9	9/13
340.	comp32660_c0_seq2	231	G	G	T	7/10	15/22
341.	comp32660_c0_seq2	456	T	G	T	7/8	22/24
342.	comp32660_c0_seq2	471	G	A	G	5/6	22/26
343.	comp25410_c0_seq1	345	G	G	A	5/6	12/18
344.	comp31587_c0_seq1	342	T	C	T	18/18	24/24
345.	comp23387_c0_seq1	394	C	T	C	9/9	34/34
346.	comp10638_c0_seq1	717	C	C	T	14/14	8/8
347.	comp33629_c0_seq19	113	A	A	C	16/20	18/22
348.	comp31238_c1_seq9	43	T	T	G	22/30	2/2
349.	comp33530_c0_seq16	929	C	T	C	6/6	4/4
350.	comp33624_c0_seq1	1036	A	A	G	9/9	4/4
351.	comp406490_c0_seq1	250	C	C	G	4/6	2/2
352.	comp26280_c0_seq1	74	A	G	A	4/6	6/6
353.	comp25835_c0_seq1	260	T	C	T	35/35	94/94
354.	comp26374_c0_seq1	544	C	T	C	5/6	8/9
355.	comp26374_c0_seq1	550	A	T	A	6/7	8/11
356.	comp195082_c0_seq1	76	A	A	G	10/10	4/4
357.	comp22382_c0_seq1	255	A	C	A	14/20	6/6
358.	comp22382_c0_seq1	256	T	C	T	14/20	6/6
359.	comp22382_c0_seq1	280	C	T	C	14/17	4/4
360.	comp22382_c0_seq1	281	T	A	T	13/16	4/4
361.	comp22382_c0_seq1	296	T	C	T	13/17	6/6
362.	comp33142_c0_seq2	370	G	A	G	14/21	17/23
363.	comp33142_c0_seq2	386	T	C	T	11/16	17/21
364.	comp33559_c0_seq1	435	G	A	G	7/9	26/26
365.	comp31104_c0_seq1	2681	A	A	C	15/15	2/2
366.	comp18567_c0_seq1	262	A	A	T	7/7	4/4
367.	comp25536_c0_seq1	174	C	T	C	6/9	16/18
368.	comp25536_c0_seq1	197	T	A	T	9/13	32/36
369.	comp28445_c0_seq1	117	T	T	C	110/161	16/20
370.	comp28445_c0_seq1	129	A	A	G	85/122	12/16
371.	comp31824_c0_seq6	2	A	C	A	4/6	2/2

372.	comp31260_c0_seq1	70	T	T	C	95/119	4/6
373.	comp31260_c0_seq1	84	A	T	A	110/151	4/6
374.	comp31260_c0_seq1	89	C	C	G	119/152	4/6
375.	comp31260_c0_seq1	90	A	A	C	130/145	4/6
376.	comp31260_c0_seq1	237	A	A	C	70/85	8/10
377.	comp31260_c0_seq1	249	C	C	T	60/90	6/8
378.	comp31260_c0_seq1	258	T	T	C	55/82	8/8
379.	comp9779_c0_seq1	393	A	A	T	56/66	10/12
380.	comp9779_c0_seq1	395	T	T	C	55/66	6/8
381.	comp9779_c0_seq1	404	G	G	T	80/88	4/6
382.	comp9779_c0_seq1	408	A	A	T	80/85	4/6
383.	comp9779_c0_seq1	411	A	A	T	80/85	4/6
384.	comp3719_c0_seq1	365	A	A	C	6/9	2/2
385.	comp32151_c0_seq3	778	A	T	A	9/10	8/8
386.	comp30606_c0_seq1	99	C	C	T	299/337	4/4
387.	comp30606_c0_seq1	100	T	T	G	321/355	4/4
388.	comp30606_c0_seq1	196	C	C	T	307/420	4/4
389.	comp30606_c0_seq1	217	G	G	T	293/423	6/8
390.	comp30606_c0_seq1	220	A	A	T	282/395	8/8
391.	comp30606_c0_seq1	225	A	A	G	285/401	12/12
392.	comp30606_c0_seq1	226	T	T	G	287/397	12/12
393.	comp30606_c0_seq1	234	G	G	C	288/402	12/12
394.	comp30606_c0_seq1	235	T	T	C	284/404	12/12
395.	comp30606_c0_seq1	460	A	A	G	79/116	5/5
396.	comp30606_c0_seq1	464	T	T	C	77/112	5/5
397.	comp30606_c0_seq1	481	A	A	T	50/74	5/5
398.	comp3510_c0_seq1	122	G	G	A	6/6	4/4
399.	comp6498_c0_seq1	3	G	T	A	6/6	2/2
400.	comp33026_c0_seq9	366	C	C	T	10/10	3/3
401.	comp33026_c0_seq9	1293	G	G	A	13/13	6/6
402.	comp33026_c0_seq9	1296	G	G	A	11/11	6/6
403.	comp33026_c0_seq9	1302	A	A	T	12/12	6/6
404.	comp33026_c0_seq9	1335	G	G	A	12/12	6/8
405.	comp33026_c0_seq9	1347	T	T	A	11/11	6/8
406.	comp29260_c0_seq2	523	T	A	T	14/14	29/29
407.	comp33555_c0_seq25	639	C	C	A	7/10	10/12
408.	comp33555_c0_seq25	655	T	T	A	4/6	5/7
409.	comp33384_c0_seq1	1366	C	T	C	14/14	16/16
410.	comp127479_c0_seq1	4	C	C	T	4/6	2/2
411.	comp17579_c0_seq1	442	G	G	A	62/65	4/6
412.	comp17579_c0_seq1	443	A	A	G	62/65	4/6
413.	comp17579_c0_seq1	466	A	A	G	58/63	4/6
414.	comp17579_c0_seq1	471	G	G	A	71/77	4/4
415.	comp17579_c0_seq1	494	A	A	C	89/96	4/5
416.	comp17579_c0_seq1	498	T	T	A	94/99	4/6
417.	comp30796_c0_seq1	222	A	G	A	11/14	6/8
418.	comp30796_c0_seq1	242	G	A	G	7/10	8/10
419.	comp30796_c0_seq1	243	T	C	T	7/10	10/10
420.	comp30796_c0_seq1	794	A	A	G	14/14	11/11
421.	comp23751_c0_seq1	201	C	T	C	6/6	38/38
422.	comp11840_c0_seq1	655	G	T	G	7/7	24/34
423.	comp31409_c0_seq4	76	A	A	G	5/6	4/4
424.	comp33694_c0_seq4	119	C	A	C	6/8	4/6
425.	comp33694_c0_seq4	122	A	G	A	6/8	4/6
426.	comp33694_c0_seq4	150	T	T	C	6/8	4/4
427.	comp33694_c0_seq4	151	C	C	T	6/8	4/4
428.	comp10773_c0_seq1	78	G	G	T	12/14	2/2
429.	comp10773_c0_seq1	80	A	A	G	12/15	2/2
430.	comp14422_c0_seq1	65	G	A	G	10/15	4/6
431.	comp14422_c0_seq1	76	T	G	T	10/15	6/8
432.	comp31264_c0_seq3	208	A	A	C	15/20	2/2
433.	comp33573_c0_seq4	775	T	C	T	7/7	16/16
434.	comp29666_c0_seq1	228	T	T	G	6/6	4/6
435.	comp29666_c0_seq1	233	G	G	A	7/7	4/6
436.	comp28567_c0_seq2	134	A	A	G	9/13	4/6
437.	comp28567_c0_seq2	186	A	A	G	14/16	4/4
438.	comp273848_c0_seq1	141	T	T	C	6/6	4/5
439.	comp24903_c0_seq1	86	C	C	T	14/15	4/6
440.	comp117132_c0_seq1	133	C	T	C	6/6	8/8
441.	comp33384_c0_seq5	1366	C	T	C	10/10	12/14
442.	comp17896_c0_seq1	247	A	C	A	20/20	30/30
443.	comp24966_c0_seq1	415	A	T	A	8/10	10/12
444.	comp24966_c0_seq1	426	C	T	C	7/10	8/10
445.	comp31231_c1_seq2	93	A	A	G	12/14	14/14
446.	comp24519_c0_seq1	80	A	T	A	28/28	20/20

447.	comp33596_c0_seq5	195	T	T	C	4/6	8/12
448.	comp33583_c0_seq11	106	C	T	C	8/11	8/10
449.	comp33515_c0_seq1	1764	C	T	C	24/24	52/52
450.	comp33515_c0_seq1	3543	A	A	G	8/8	6/8
451.	comp31238_c0_seq1	1171	A	A	G	161/211	6/6
452.	comp31238_c0_seq1	1253	T	C	T	281/382	8/8
453.	comp31238_c0_seq1	1263	A	G	A	280/372	6/6
454.	comp31238_c0_seq1	1269	C	T	C	281/370	6/6
455.	comp31238_c0_seq1	1281	C	T	C	283/360	6/6
456.	comp31238_c0_seq1	1283	T	C	T	281/354	6/6
457.	comp31238_c0_seq1	1285	G	A	G	282/355	2/2
458.	comp31238_c0_seq1	2105	A	A	T	485/485	4/4
459.	comp28845_c0_seq1	190	T	T	C	7/7	4/4
460.	comp31335_c0_seq1	631	A	G	A	6/6	20/20
461.	comp28905_c0_seq1	33	C	C	T	7/7	4/6
462.	comp32250_c0_seq1	1424	G	G	C	11/16	16/24
463.	comp32250_c0_seq1	1649	T	T	A	7/8	4/6
464.	comp33196_c0_seq1	627	G	G	T	6/7	4/4
465.	comp29655_c0_seq1	1140	C	T	C	14/14	38/38
466.	comp25439_c0_seq1	127	T	T	C	34/51	32/38
467.	comp33248_c0_seq3	7	G	G	A	6/9	10/12
468.	comp33248_c0_seq3	203	T	C	T	9/10	6/8
469.	comp33248_c0_seq3	205	G	A	G	7/8	6/8
470.	comp33248_c0_seq3	215	T	C	T	5/6	6/8
471.	comp32621_c0_seq8	408	A	A	T	9/11	2/2
472.	comp33686_c0_seq30	427	A	G	A	16/20	12/12
473.	comp30614_c0_seq1	91	T	T	C	8/8	9/9
474.	comp30614_c0_seq1	796	G	C	G	11/11	20/20
475.	comp30614_c0_seq1	800	C	A	C	13/14	26/26
476.	comp33142_c0_seq7	10	G	A	G	4/6	4/4
477.	comp33142_c0_seq7	14	G	A	G	4/6	4/4
478.	comp33142_c0_seq7	51	T	T	C	8/10	4/4
479.	comp33142_c0_seq7	67	A	A	G	9/11	4/4
480.	comp33142_c0_seq7	132	G	G	A	16/20	4/6
481.	comp33142_c0_seq7	147	T	T	C	17/22	4/6
482.	comp33142_c0_seq7	168	G	G	A	17/23	12/12
483.	comp33142_c0_seq7	189	A	A	G	16/23	14/14
484.	comp33142_c0_seq7	190	T	T	C	16/22	12/14
485.	comp33652_c0_seq38	187	T	T	C	6/6	7/9
486.	comp33356_c0_seq1	1483	T	C	T	4/6	10/12
487.	comp33695_c0_seq3	404	A	A	C	18/21	12/18
488.	comp33695_c0_seq3	405	A	A	G	21/22	12/18
489.	comp33695_c0_seq3	744	C	C	T	17/24	4/6
490.	comp24051_c0_seq1	249	G	T	G	4/6	26/26
491.	comp24051_c0_seq1	250	A	G	A	4/6	26/26
492.	comp24051_c0_seq1	253	T	G	T	4/6	22/22
493.	comp24051_c0_seq1	254	T	A	T	4/6	22/22
494.	comp27856_c0_seq1	559	T	T	C	7/7	4/4
495.	comp27856_c0_seq1	570	T	T	A	6/8	6/8
496.	comp27856_c0_seq1	580	T	T	C	4/6	8/10
497.	comp27856_c0_seq1	657	A	A	C	17/20	8/10
498.	comp27856_c0_seq1	778	A	A	G	18/21	11/15
499.	comp27856_c0_seq1	785	C	C	A	18/21	13/15
500.	comp27856_c0_seq1	789	T	T	C	15/18	13/15
501.	comp20640_c0_seq1	152	G	G	A	15/18	2/2
502.	comp20640_c0_seq1	153	C	C	T	16/19	2/2
503.	comp29417_c0_seq3	364	C	C	T	7/7	10/10
504.	comp22280_c0_seq1	250	T	A	T	8/11	22/32
505.	comp14022_c0_seq1	96	T	T	A	8/9	2/2
506.	comp14022_c0_seq1	97	C	C	G	8/9	2/2
507.	comp14022_c0_seq1	103	G	G	A	6/7	2/2
508.	comp32401_c0_seq1	67	C	C	G	33/45	18/20
509.	comp30765_c0_seq1	2881	C	G	C	12/18	24/36
510.	comp30968_c0_seq1	3	G	A	G	5/7	2/2
511.	comp33702_c0_seq8	564	T	C	T	7/9	12/12
512.	comp17547_c0_seq1	84	T	T	C	58/66	4/6
513.	comp17547_c0_seq1	87	A	A	G	64/72	4/6
514.	comp17547_c0_seq1	101	G	G	C	73/79	4/6
515.	comp175012_c0_seq1	104	C	T	C	7/7	18/18
516.	comp33285_c0_seq4	1000	G	A	G	7/9	12/16
517.	comp33285_c0_seq4	1008	A	C	A	9/10	12/14
518.	comp33285_c0_seq4	1030	G	C	G	19/20	6/8
519.	comp28410_c0_seq1	1227	A	T	A	12/12	15/15
520.	comp20610_c0_seq1	105	G	A	G	5/6	10/12
521.	comp33175_c0_seq45	361	C	C	T	5/6	2/2

522.	comp31243_c0_seq1	539	G	A	G	9/9	36/38
523.	comp109351_c0_seq1	193	T	G	T	11/11	22/22
524.	comp33052_c0_seq1	596	C	C	G	11/15	8/10
525.	comp33571_c0_seq2	209	T	T	A	12/12	6/6
526.	comp24843_c0_seq1	302	A	A	T	35/35	22/22
527.	comp9037_c0_seq1	326	C	C	T	16/19	6/8
528.	comp9037_c0_seq1	347	C	C	G	13/16	4/6
529.	comp29730_c0_seq1	99	T	A	T	5/7	4/6
530.	comp29730_c0_seq1	101	T	C	T	5/7	4/6
531.	comp29730_c0_seq1	102	T	G	T	5/7	4/4
532.	comp29877_c0_seq1	191	G	A	G	9/9	16/16
533.	comp33466_c0_seq4	17	T	G	T	6/7	2/2
534.	comp33466_c0_seq4	20	A	T	A	5/6	2/2
535.	comp260788_c0_seq1	116	C	T	C	6/6	12/16
536.	comp21531_c0_seq1	72	G	G	T	9/9	4/6
537.	comp33026_c0_seq6	1477	G	G	T	12/13	4/6
538.	comp20883_c0_seq1	48	C	T	C	6/8	2/2
539.	comp20883_c0_seq1	53	G	A	G	6/8	2/2
540.	comp20883_c0_seq1	66	A	G	A	8/11	4/4
541.	comp24659_c0_seq1	78	C	T	C	45/61	2/2
542.	comp24659_c0_seq1	79	C	T	C	45/62	2/2
543.	comp24659_c0_seq1	80	A	C	A	46/63	2/2
544.	comp33571_c0_seq6	209	T	T	A	15/15	20/20
545.	comp33595_c0_seq23	119	T	T	C	6/9	5/7
546.	comp33425_c0_seq1	828	C	G	C	14/21	14/21
547.	comp33425_c0_seq1	949	G	G	A	21/26	29/43
548.	comp33425_c0_seq1	959	C	C	T	20/26	36/53
549.	comp33425_c0_seq1	1651	C	C	G	8/8	24/32
550.	comp33425_c0_seq1	1654	T	T	C	7/7	24/33
551.	comp33425_c0_seq1	1694	G	C	G	7/7	30/40
552.	comp33425_c0_seq1	1700	G	A	G	5/6	30/38
553.	comp33425_c0_seq1	1701	T	A	T	5/6	30/38
554.	comp33425_c0_seq1	1708	G	A	G	5/6	32/40
555.	comp33425_c0_seq1	1713	T	C	T	5/7	32/40
556.	comp33425_c0_seq1	1718	A	C	A	5/7	32/40
557.	comp33425_c0_seq1	1744	T	A	T	6/8	23/33
558.	comp33425_c0_seq1	4419	A	A	T	5/6	12/18
559.	comp3380_c0_seq1	528	A	T	A	4/6	4/4
560.	comp3380_c0_seq1	530	T	A	T	4/6	4/4
561.	comp3380_c0_seq1	531	A	G	A	6/8	4/4
562.	comp19222_c0_seq1	1294	A	T	A	7/7	30/30
563.	comp33332_c0_seq12	957	G	G	A	20/20	16/16
564.	comp21456_c0_seq1	177	G	C	G	9/9	40/40
565.	comp13720_c0_seq1	414	C	C	A	8/8	4/4
566.	comp33530_c0_seq8	651	G	A	G	9/9	1/1
567.	comp32498_c0_seq1	165	A	A	G	9/9	3/3
568.	comp32498_c0_seq1	167	A	A	G	10/10	3/3
569.	comp32498_c0_seq1	183	A	A	G	9/10	5/5
570.	comp32498_c0_seq1	193	A	A	C	10/11	9/11
571.	comp32498_c0_seq1	202	G	G	A	9/10	7/9
572.	comp33415_c0_seq8	355	G	G	A	8/12	8/10
573.	comp18585_c0_seq1	383	A	T	A	13/19	66/76
574.	comp18585_c0_seq1	397	G	T	G	17/24	66/76
575.	comp18585_c0_seq1	407	G	C	G	18/26	64/74
576.	comp32526_c0_seq7	4	T	G	T	6/6	4/6
577.	comp33177_c0_seq3	4574	A	T	A	22/22	19/19
578.	comp17525_c1_seq1	3	T	T	G	6/8	2/3
579.	comp24786_c0_seq1	1495	T	T	A	26/27	10/10
580.	comp33695_c0_seq1	231	C	C	G	21/30	12/18
581.	comp33695_c0_seq1	404	A	A	C	19/24	4/4
582.	comp33695_c0_seq1	405	A	A	G	24/24	4/4
583.	comp33695_c0_seq1	1251	T	T	C	23/27	6/8
584.	comp33695_c0_seq1	1266	A	A	C	23/33	8/10
585.	comp33695_c0_seq1	1669	T	T	C	79/107	22/32
586.	comp6768_c0_seq1	614	T	C	T	16/16	30/30
587.	comp26591_c0_seq1	169	C	T	C	9/9	10/10
588.	comp29417_c0_seq2	343	C	C	T	15/15	11/11
589.	comp167396_c0_seq1	4	T	G	T	4/6	4/4
590.	comp31849_c0_seq8	654	G	G	A	7/7	4/4
591.	comp31849_c0_seq8	682	G	G	C	5/6	2/2
592.	comp31849_c0_seq8	689	A	A	C	6/7	2/2
593.	comp108222_c0_seq1	415	C	C	T	9/9	10/10
594.	comp33521_c0_seq4	5554	T	C	T	10/14	18/20
595.	comp33521_c0_seq4	5557	T	C	T	10/14	15/20
596.	comp33521_c0_seq4	5562	T	C	T	10/14	13/16

597.	comp33521_c0_seq4	5563	T	C	T	12/15	13/14
598.	comp33521_c0_seq4	5611	T	C	T	8/9	4/4
599.	comp28065_c0_seq1	118	A	C	A	13/13	10/10
600.	comp33722_c0_seq9	93	G	G	A	16/24	22/30
601.	comp33722_c0_seq9	99	C	C	A	18/25	18/26
602.	comp31762_c0_seq1	1133	C	G	C	10/14	18/18
603.	comp33413_c0_seq1	2953	A	G	A	9/9	6/6
604.	comp33555_c0_seq28	650	T	G	T	6/9	4/4
605.	comp99130_c0_seq1	59	T	T	A	8/10	2/2
606.	comp99130_c0_seq1	78	A	A	T	12/14	2/2
607.	comp7986_c0_seq1	83	T	T	C	15/19	1/1
608.	comp33686_c0_seq2	4019	T	A	T	7/10	6/8
609.	comp33686_c0_seq2	4020	A	C	A	6/9	6/8
610.	comp33510_c0_seq10	519	G	T	G	160/235	6/6
611.	comp33510_c0_seq10	552	G	C	G	277/375	2/2
612.	comp33510_c0_seq10	559	G	A	G	268/362	2/2
613.	comp33510_c0_seq10	562	T	C	T	274/362	2/2
614.	comp23912_c0_seq1	763	G	G	A	20/22	4/6
615.	comp23912_c0_seq1	802	G	G	A	5/6	2/2
616.	comp23912_c0_seq1	850	C	A	C	6/7	2/2
617.	comp30145_c0_seq2	34	G	G	A	7/7	4/4
618.	comp31914_c0_seq1	659	A	A	G	54/54	34/34
619.	comp31914_c0_seq1	820	T	G	T	11/11	26/26
620.	comp31914_c0_seq1	1082	C	T	C	17/17	64/64
621.	comp31914_c0_seq1	1125	G	T	G	11/11	40/40
622.	comp1927_c0_seq1	182	C	C	A	11/13	2/2
623.	comp1927_c0_seq1	195	C	C	A	15/15	18/24
624.	comp28640_c0_seq1	115	G	A	G	4/6	4/4
625.	comp31003_c0_seq1	1117	C	G	C	5/7	12/12
626.	comp40733_c0_seq1	4	T	C	T	4/6	2/2
627.	comp10405_c0_seq1	446	T	C	T	7/7	16/16
628.	comp30293_c0_seq1	1523	G	G	T	5/7	2/2
629.	comp30293_c0_seq1	1648	T	T	G	21/28	2/2
630.	comp30293_c0_seq1	1652	A	A	C	20/26	2/2
631.	comp30293_c0_seq1	1663	T	T	C	21/28	2/2
632.	comp30293_c0_seq1	1677	T	T	A	18/25	2/2
633.	comp30293_c0_seq1	2163	A	A	G	7/10	2/2
634.	comp30293_c0_seq1	2484	T	T	G	15/22	2/2
635.	comp33550_c0_seq4	348	A	A	C	10/11	8/12
636.	comp30275_c0_seq4	49	A	A	G	18/19	8/12
637.	comp30275_c0_seq4	102	T	G	T	46/63	12/18
638.	comp30275_c0_seq4	146	G	G	A	47/50	6/8
639.	comp118988_c0_seq1	105	C	T	C	4/6	6/8
640.	comp301049_c0_seq1	226	G	G	C	6/6	4/4
641.	comp31779_c0_seq1	911	T	G	T	10/12	4/6
642.	comp30770_c0_seq1	2397	T	C	T	8/8	42/42
643.	comp33138_c0_seq20	134	G	C	G	22/27	38/48
644.	comp33138_c0_seq20	139	A	T	A	15/20	38/47
645.	comp33138_c0_seq20	141	A	T	A	13/18	32/38
646.	comp33138_c0_seq20	148	T	C	T	6/9	32/34
647.	comp33656_c0_seq26	69	A	G	A	4/6	10/10
648.	comp33656_c0_seq26	72	T	C	T	8/8	8/10
649.	comp33496_c0_seq14	263	C	C	T	9/13	4/6
650.	comp7671_c0_seq1	571	A	T	A	8/8	26/26
651.	comp33624_c0_seq3	1036	A	A	G	8/8	10/10
652.	comp30729_c0_seq1	148	G	C	G	8/8	4/4
653.	comp33640_c0_seq57	537	T	T	C	4/6	6/8
654.	comp29899_c0_seq1	272	G	T	G	16/16	24/24
655.	comp12314_c0_seq1	91	A	A	T	4/6	2/2
656.	comp18297_c0_seq1	4	T	T	C	5/6	2/2
657.	comp32646_c0_seq1	529	G	A	G	4/6	28/30
658.	comp205939_c0_seq1	112	G	G	C	11/14	4/4
659.	comp205939_c0_seq1	114	C	C	T	12/15	4/4
660.	comp205939_c0_seq1	150	C	C	T	11/13	4/4
661.	comp205939_c0_seq1	159	T	T	C	11/13	4/4
662.	comp205939_c0_seq1	164	A	A	T	11/13	4/4
663.	comp205939_c0_seq1	179	G	G	A	11/13	4/4
664.	comp22491_c0_seq1	137	G	G	T	4/6	9/13
665.	comp19552_c0_seq1	122	C	C	T	5/6	2/2
666.	comp5618_c0_seq1	325	T	T	A	8/9	8/8
667.	comp20150_c0_seq1	42	G	A	G	9/12	2/2
668.	comp20150_c0_seq1	50	G	T	G	9/13	2/2
669.	comp20150_c0_seq1	91	C	T	C	20/29	2/2
670.	comp20150_c0_seq1	186	G	G	C	24/33	2/2
671.	comp33257_c0_seq1	664	T	G	T	15/15	20/20

672.	comp33257_c0_seq1	3321	G	G	C	19/19	16/19
673.	comp33257_c0_seq1	3529	T	G	T	16/16	34/34
674.	comp4350_c0_seq1	109	C	C	T	6/6	6/6
675.	comp32286_c0_seq1	76	G	A	G	7/7	26/26
676.	comp29106_c0_seq1	503	G	G	A	7/7	4/6
677.	comp29344_c0_seq1	710	G	A	G	9/13	6/9
678.	comp25368_c0_seq1	228	T	C	T	8/11	12/12
679.	comp25368_c0_seq1	251	A	G	A	8/9	8/8
680.	comp31245_c0_seq1	1089	A	T	A	8/9	8/8
681.	comp33700_c0_seq23	187	A	A	T	5/7	10/11
682.	comp226503_c0_seq1	43	A	G	A	6/9	4/4
683.	comp31569_c0_seq15	102	A	A	C	6/7	8/11
684.	comp31569_c0_seq15	108	C	C	T	6/8	8/11
685.	comp31569_c0_seq15	115	C	C	T	6/8	10/11
686.	comp31569_c0_seq15	116	C	C	T	6/8	10/11
687.	comp31569_c0_seq15	133	C	C	G	6/8	10/10
688.	comp274887_c0_seq1	229	C	C	A	11/11	4/4
689.	comp2326_c0_seq1	174	G	G	A	6/6	4/6
690.	comp2326_c0_seq1	176	G	G	A	6/6	4/6
691.	comp31395_c0_seq6	351	T	T	C	6/6	16/16
692.	comp7495_c0_seq1	285	T	T	A	15/15	4/4
693.	comp32052_c0_seq1	1022	A	G	A	9/10	28/28
694.	comp28778_c0_seq3	405	G	C	G	14/14	8/8
695.	comp24574_c0_seq1	100	T	T	C	27/28	8/8
696.	comp24574_c0_seq1	226	C	C	G	24/31	4/6
697.	comp24574_c0_seq1	238	G	G	A	27/36	4/6
698.	comp24574_c0_seq1	346	A	A	G	11/15	2/2
699.	comp32767_c0_seq3	71	T	T	A	7/9	4/6
700.	comp32767_c0_seq3	115	G	G	C	8/10	4/4
701.	comp32767_c0_seq3	152	G	G	C	5/7	4/4
702.	comp33471_c0_seq108	271	C	T	C	62/76	8/12
703.	comp33471_c0_seq108	279	G	A	G	56/68	8/10
704.	comp30120_c0_seq1	933	T	T	C	6/6	4/4
705.	comp27656_c0_seq1	3	C	C	G	8/11	6/8
706.	comp21954_c0_seq1	366	T	T	A	31/31	8/8
707.	comp31663_c0_seq1	449	A	A	G	25/25	6/6
708.	comp6448_c0_seq1	4	A	A	C	6/8	4/6
709.	comp26961_c0_seq1	369	C	C	T	4/6	8/10
710.	comp163692_c0_seq1	106	C	C	T	9/9	8/8
711.	comp33719_c0_seq11	111	C	T	C	4/6	7/9
712.	comp25152_c0_seq1	4	T	A	T	4/6	2/2
713.	comp30572_c0_seq1	245	C	C	T	7/8	8/12
714.	comp30572_c0_seq1	304	A	G	A	8/10	14/16
715.	comp30572_c0_seq1	502	T	C	T	6/6	4/4
716.	comp30572_c0_seq1	698	A	A	G	15/19	20/30
717.	comp27203_c0_seq1	680	G	G	A	7/10	2/2
718.	comp27203_c0_seq1	686	T	T	C	7/10	2/2
719.	comp29233_c0_seq3	104	T	T	C	14/18	2/2
720.	comp29233_c0_seq3	106	A	A	G	13/17	2/2
721.	comp29233_c0_seq3	125	C	C	T	10/14	2/2
722.	comp29233_c0_seq3	143	C	C	G	11/15	2/2
723.	comp28538_c0_seq1	54	G	A	G	6/6	30/30
724.	comp175922_c0_seq1	100	T	A	T	10/12	8/10
725.	comp28190_c0_seq1	210	C	C	T	14/14	4/6
726.	comp29090_c0_seq1	118	A	T	A	4/6	5/5
727.	comp29090_c0_seq1	120	T	C	T	4/6	5/5
728.	comp24693_c0_seq1	829	G	G	T	13/15	16/16
729.	comp27543_c0_seq1	179	A	A	G	5/6	2/2
730.	comp27543_c0_seq1	183	T	T	C	5/6	2/2
731.	comp27543_c0_seq1	186	G	G	A	5/6	2/2
732.	comp25537_c0_seq1	389	G	G	T	23/23	16/16
733.	comp91388_c0_seq1	71	C	C	T	12/16	2/2
734.	comp91388_c0_seq1	277	C	C	T	43/43	6/8
735.	comp91388_c0_seq1	280	A	A	G	44/44	6/8
736.	comp91388_c0_seq1	290	G	G	A	42/42	6/8
737.	comp91388_c0_seq1	293	T	T	A	43/43	6/8
738.	comp91388_c0_seq1	295	C	C	T	42/42	6/8
739.	comp91388_c0_seq1	303	T	T	G	43/45	6/8
740.	comp91388_c0_seq1	329	C	C	G	39/40	4/6
741.	comp91388_c0_seq1	347	G	G	A	24/26	4/6
742.	comp91388_c0_seq1	349	G	G	A	25/26	4/6
743.	comp23515_c0_seq1	170	T	T	C	23/23	16/16
744.	comp33158_c0_seq1	4004	G	G	A	4/6	14/14
745.	comp139342_c0_seq1	356	T	G	T	4/6	12/16
746.	comp139342_c0_seq1	404	G	G	T	9/11	10/14

747.	comp21184_c0_seq1	132	C	T	C	7/7	4/6
748.	comp32470_c0_seq2	163	G	G	A	4/6	7/9
749.	comp32470_c0_seq2	606	G	A	G	12/18	8/11
750.	comp32470_c0_seq2	616	T	G	T	12/18	11/14
751.	comp32470_c0_seq2	647	A	A	G	11/16	9/12
752.	comp52547_c0_seq1	4	A	A	T	6/6	4/6
753.	comp173284_c0_seq1	53	T	C	T	6/6	4/4
754.	comp33626_c0_seq6	369	T	C	T	7/7	4/6
755.	comp33626_c0_seq6	374	A	G	A	7/7	4/6
756.	comp33626_c0_seq6	380	G	A	G	7/7	4/6
757.	comp33295_c0_seq14	760	G	A	G	10/15	2/2
758.	comp18536_c0_seq1	1436	C	T	C	14/14	14/14
759.	comp17620_c0_seq1	37	T	T	C	8/10	2/2
760.	comp208081_c0_seq1	153	C	C	T	6/6	4/6
761.	comp32789_c0_seq1	134	T	T	C	315/382	4/6
762.	comp32789_c0_seq1	147	A	A	T	275/378	4/6
763.	comp32789_c0_seq1	236	T	T	C	293/309	4/6
764.	comp32789_c0_seq1	251	G	G	T	258/365	8/10
765.	comp32789_c0_seq1	259	A	A	G	271/376	10/14
766.	comp132042_c0_seq1	35	G	A	G	7/7	3/3
767.	comp32020_c0_seq1	339	G	A	G	18/19	18/18
768.	comp33025_c0_seq2	1799	C	T	C	6/8	14/14
769.	comp29090_c0_seq4	813	G	G	T	4/6	26/35
770.	comp29090_c0_seq4	816	G	G	A	4/6	24/33
771.	comp29090_c0_seq4	819	G	G	T	4/6	24/35
772.	comp29090_c0_seq4	822	A	A	G	5/7	22/33
773.	comp32157_c0_seq1	89	A	G	A	7/10	3/3
774.	comp27163_c0_seq1	182	T	T	C	18/18	26/26
775.	comp32283_c0_seq3	141	T	C	T	9/9	11/11
776.	comp32283_c0_seq3	856	T	G	T	10/10	14/14
777.	comp33531_c0_seq2	1031	G	C	G	7/7	13/13
778.	comp32283_c0_seq11	141	T	C	T	9/9	20/20
779.	comp33722_c0_seq1	527	T	T	C	33/40	20/30
780.	comp33722_c0_seq1	529	C	C	T	32/41	21/30
781.	comp32313_c0_seq1	34	C	T	C	8/8	24/24
782.	comp32313_c0_seq1	986	A	G	A	12/12	24/24
783.	comp33668_c0_seq2	2267	T	C	T	11/11	8/12
784.	comp33668_c0_seq2	2268	T	C	T	10/10	8/12
785.	comp33668_c0_seq2	2297	G	G	A	11/11	8/12
786.	comp33668_c0_seq2	2301	G	G	A	11/11	8/12
787.	comp33668_c0_seq2	2332	A	A	G	6/6	8/12
788.	comp33668_c0_seq2	2336	A	A	G	6/6	8/12
789.	comp33668_c0_seq2	2339	T	T	C	6/6	10/14
790.	comp33668_c0_seq2	3000	T	C	T	8/8	31/37
791.	comp33668_c0_seq2	3003	T	C	T	8/8	28/34
792.	comp33668_c0_seq2	3116	T	G	T	6/6	4/6
793.	comp33668_c0_seq2	3134	G	A	G	6/6	6/8
794.	comp33668_c0_seq2	3160	A	A	G	6/6	8/12
795.	comp33668_c0_seq2	3188	C	C	A	13/13	8/12
796.	comp33668_c0_seq2	4182	C	C	T	32/32	14/20
797.	comp33668_c0_seq2	4183	T	T	C	32/32	14/20
798.	comp33668_c0_seq2	4213	G	G	A	24/24	12/16
799.	comp33668_c0_seq2	4242	A	G	A	10/10	14/18
800.	comp33668_c0_seq2	4260	T	C	T	8/8	12/14
801.	comp33668_c0_seq2	4306	T	T	C	8/9	8/12
802.	comp33668_c0_seq2	4308	A	A	C	8/9	10/14
803.	comp33668_c0_seq2	4310	G	G	A	8/9	10/14
804.	comp33668_c0_seq2	4355	C	C	T	16/18	16/22
805.	comp33668_c0_seq2	4371	T	T	C	14/16	22/28
806.	comp33668_c0_seq2	4373	G	G	A	14/16	24/30
807.	comp33668_c0_seq2	4392	G	G	A	15/17	30/34
808.	comp33668_c0_seq2	4399	T	T	C	18/20	32/36
809.	comp33668_c0_seq2	4406	A	A	G	17/18	30/32
810.	comp33668_c0_seq2	4413	T	T	C	16/17	30/32
811.	comp33668_c0_seq2	4514	T	T	C	26/26	14/18
812.	comp33668_c0_seq2	4515	G	G	A	26/26	14/18
813.	comp33668_c0_seq2	4518	A	A	G	25/25	14/18
814.	comp33668_c0_seq2	4525	C	C	T	26/26	10/14
815.	comp33668_c0_seq2	4560	G	G	A	11/11	10/12
816.	comp33668_c0_seq2	4577	T	T	C	8/8	12/16
817.	comp33668_c0_seq2	4607	A	A	G	11/11	8/12
818.	comp33668_c0_seq2	4669	C	C	T	18/18	12/16
819.	comp33668_c0_seq2	4691	C	C	T	26/27	16/18
820.	comp33668_c0_seq2	4693	C	C	T	28/29	16/18
821.	comp33668_c0_seq2	4733	T	T	G	29/31	14/16

822.	comp33668_c0_seq2	4769	A	A	G	25/27	12/12
823.	comp33668_c0_seq2	4780	T	T	C	15/18	8/8
824.	comp33668_c0_seq2	4784	C	C	T	14/16	8/8
825.	comp33668_c0_seq2	4787	T	T	C	13/14	8/8
826.	comp33668_c0_seq2	4793	C	C	T	11/13	6/6
827.	comp33668_c0_seq2	4795	A	A	C	10/11	6/6
828.	comp33668_c0_seq2	4806	T	T	C	9/10	10/10
829.	comp33668_c0_seq2	4809	C	C	T	9/10	8/8
830.	comp37658_c0_seq1	205	C	C	T	15/22	10/12
831.	comp37658_c0_seq1	218	T	T	C	26/27	4/6
832.	comp26506_c0_seq1	298	G	G	T	32/32	10/10
833.	comp23282_c0_seq1	118	C	T	C	6/6	12/12
834.	comp32157_c0_seq2	56	T	C	T	5/6	6/7
835.	comp32157_c0_seq2	57	C	T	C	6/8	7/7
836.	comp32157_c0_seq2	61	C	T	C	8/9	6/9
837.	comp32157_c0_seq2	69	C	A	C	8/9	12/12
838.	comp32157_c0_seq2	73	G	T	G	8/11	13/13
839.	comp27833_c0_seq1	826	T	T	C	12/12	18/18
840.	comp30550_c0_seq1	13	C	A	C	12/18	4/6
841.	comp30550_c0_seq1	498	C	G	C	10/13	12/17
842.	comp33571_c0_seq1	209	T	T	A	23/23	6/6
843.	comp25217_c0_seq1	1153	G	A	G	10/11	12/12
844.	comp25217_c0_seq1	1157	A	G	A	7/7	12/12
845.	comp33283_c0_seq4	936	G	T	G	5/7	7/9
846.	comp14720_c0_seq1	100	C	C	A	6/6	3/3
847.	comp24209_c0_seq1	4	G	G	T	4/6	2/2
848.	comp33703_c0_seq4	727	A	A	G	8/9	13/16
849.	comp33728_c0_seq1	306	C	T	C	45/65	99/101
850.	comp33728_c0_seq1	319	C	T	C	43/65	88/90
851.	comp33728_c0_seq1	345	A	G	A	45/66	108/112
852.	comp33728_c0_seq1	384	A	G	A	36/47	112/112
853.	comp33728_c0_seq1	387	C	T	C	36/47	106/106
854.	comp33728_c0_seq1	402	A	T	A	46/46	98/98
855.	comp33728_c0_seq1	429	A	A	G	41/48	73/74
856.	comp33728_c0_seq1	444	G	G	T	59/69	62/62
857.	comp33728_c0_seq1	452	A	A	G	61/68	29/29
858.	comp33728_c0_seq1	453	T	T	C	62/69	33/33
859.	comp33728_c0_seq1	459	A	A	G	64/74	35/35
860.	comp33728_c0_seq1	483	G	G	A	89/105	30/30
861.	comp33728_c0_seq1	516	G	G	A	198/223	33/33
862.	comp33728_c0_seq1	519	C	C	T	233/235	35/35
863.	comp33728_c0_seq1	549	A	A	G	223/244	56/56
864.	comp33728_c0_seq1	564	C	C	T	222/242	60/60
865.	comp33728_c0_seq1	576	G	G	A	225/252	63/64
866.	comp33728_c0_seq1	582	A	A	G	220/255	63/65
867.	comp33728_c0_seq1	602	G	G	A	173/199	70/71
868.	comp33728_c0_seq1	1106	G	A	G	22/24	34/34
869.	comp33728_c0_seq1	1165	C	T	C	81/81	116/116
870.	comp33728_c0_seq1	1449	A	G	A	57/58	74/75
871.	comp33728_c0_seq1	1566	G	A	G	130/130	152/152
872.	comp33728_c0_seq1	1629	A	G	A	46/64	92/92
873.	comp33728_c0_seq1	1641	G	A	G	31/44	96/98
874.	comp33728_c0_seq1	1644	A	G	A	32/45	96/96
875.	comp33728_c0_seq1	1698	G	A	G	48/66	122/122
876.	comp33728_c0_seq1	1706	G	A	G	48/65	122/122
877.	comp33728_c0_seq1	1746	T	C	T	41/61	126/128
878.	comp33728_c0_seq1	1758	G	A	G	50/55	119/119
879.	comp33728_c0_seq1	1761	T	C	T	55/55	120/120
880.	comp33728_c0_seq1	1899	A	T	A	41/42	173/173
881.	comp33728_c0_seq1	1911	C	T	C	49/50	186/186
882.	comp33728_c0_seq1	1950	A	G	A	80/80	192/193
883.	comp33728_c0_seq1	1980	T	C	T	102/103	193/193
884.	comp33728_c0_seq1	2097	T	C	T	157/160	278/283
885.	comp33728_c0_seq1	2183	T	C	T	159/160	256/256
886.	comp33728_c0_seq1	2235	A	G	A	123/170	278/278
887.	comp33728_c0_seq1	2328	T	C	T	139/143	230/231
888.	comp33728_c0_seq1	2379	T	C	T	58/60	118/122
889.	comp33728_c0_seq1	2403	G	A	G	50/50	101/101
890.	comp33728_c0_seq1	2457	A	G	A	33/33	78/78
891.	comp33728_c0_seq1	2480	C	T	C	31/35	78/78
892.	comp33728_c0_seq1	2490	A	G	A	29/37	90/90
893.	comp33728_c0_seq1	2545	T	G	T	20/26	72/72
894.	comp33728_c0_seq1	2607	T	C	T	22/22	102/104
895.	comp33728_c0_seq1	2613	A	G	A	25/25	109/109
896.	comp33728_c0_seq1	2688	T	C	T	56/69	160/160

897.	comp33728_c0_seq1	2742	T	C	T	59/59	132/133
898.	comp33728_c0_seq1	2793	G	A	G	24/35	106/106
899.	comp33728_c0_seq1	2796	G	A	G	22/33	102/102
900.	comp33728_c0_seq1	2811	T	C	T	33/41	103/103
901.	comp33728_c0_seq1	2889	A	G	A	58/79	136/136
902.	comp33728_c0_seq1	2895	G	A	G	62/84	156/156
903.	comp33728_c0_seq1	3096	C	T	C	20/30	136/138
904.	comp33728_c0_seq1	3099	G	A	G	18/27	138/138
905.	comp33728_c0_seq1	3219	C	T	C	60/62	97/99
906.	comp33728_c0_seq1	3282	A	G	A	79/80	108/108
907.	comp33728_c0_seq1	3300	G	A	G	70/70	106/106
908.	comp33728_c0_seq1	3429	G	A	G	15/15	96/97
909.	comp33728_c0_seq1	3440	A	G	A	13/13	104/104
910.	comp33728_c0_seq1	3506	G	A	G	8/8	76/76
911.	comp33728_c0_seq1	3519	G	A	G	6/7	74/74
912.	comp33728_c0_seq1	3552	A	C	A	26/31	68/68
913.	comp33728_c0_seq1	3570	G	A	G	29/39	72/72
914.	comp33728_c0_seq1	3573	T	C	T	28/38	74/76
915.	comp33728_c0_seq1	3576	A	G	A	29/39	74/74
916.	comp33728_c0_seq1	3579	C	A	C	29/39	78/78
917.	comp33728_c0_seq1	3612	G	A	G	34/45	85/86
918.	comp33728_c0_seq1	3624	C	T	C	31/42	90/90
919.	comp33728_c0_seq1	3738	G	A	G	54/55	160/162
920.	comp33728_c0_seq1	3825	T	C	T	51/52	106/106
921.	comp33728_c0_seq1	3879	G	A	G	25/25	66/68
922.	comp33728_c0_seq1	3930	T	C	T	28/34	52/54
923.	comp33728_c0_seq1	3932	T	C	T	34/36	54/54
924.	comp33728_c0_seq1	3963	T	T	C	67/67	60/60
925.	comp33728_c0_seq1	4053	C	C	T	89/131	32/32
926.	comp33728_c0_seq1	4062	A	A	G	144/190	30/30
927.	comp33728_c0_seq1	4071	C	C	T	239/281	32/32
928.	comp33728_c0_seq1	4077	T	T	C	249/287	32/32
929.	comp33728_c0_seq1	4079	T	T	G	252/290	34/34
930.	comp33728_c0_seq1	4080	C	C	T	253/284	30/30
931.	comp33728_c0_seq1	4083	G	G	C	248/276	32/32
932.	comp33728_c0_seq1	4107	A	A	G	301/301	50/52
933.	comp33728_c0_seq1	4143	T	T	C	251/271	108/108
934.	comp33728_c0_seq1	4191	C	T	C	45/46	130/130
935.	comp33728_c0_seq1	4341	G	A	G	10/10	74/74
936.	comp33728_c0_seq1	4362	T	C	T	7/7	90/90
937.	comp33728_c0_seq1	4387	T	A	T	10/13	82/82
938.	comp33728_c0_seq1	4404	C	T	C	21/23	74/76
939.	comp33728_c0_seq1	4486	A	G	A	46/47	120/124
940.	comp33728_c0_seq1	4528	G	A	G	43/43	154/154
941.	comp33728_c0_seq1	4597	C	T	C	43/45	214/214
942.	comp33728_c0_seq1	4605	G	T	G	43/45	212/212
943.	comp33728_c0_seq1	4623	T	C	T	40/40	200/200
944.	comp33728_c0_seq1	4638	G	A	G	34/34	226/226
945.	comp33728_c0_seq1	4652	A	G	A	29/30	198/200
946.	comp33728_c0_seq1	4704	A	G	A	23/23	167/167
947.	comp33728_c0_seq1	4717	G	A	G	20/20	152/152
948.	comp33728_c0_seq1	4762	G	A	G	19/19	75/109
949.	comp33728_c0_seq1	4786	C	T	C	10/14	140/140
950.	comp33728_c0_seq1	4790	C	T	C	9/13	143/143
951.	comp33728_c0_seq1	4798	C	T	C	15/15	148/148
952.	comp33728_c0_seq1	4824	G	A	G	15/15	154/156
953.	comp33728_c0_seq1	4840	A	G	A	17/17	157/157
954.	comp33728_c0_seq1	4876	G	A	G	17/17	90/90
955.	comp33728_c0_seq1	4880	T	C	T	13/17	88/88
956.	comp33728_c0_seq1	4885	A	G	A	19/19	84/84
957.	comp33728_c0_seq1	4896	A	T	A	14/17	62/64
958.	comp33728_c0_seq1	4946	G	T	G	7/7	26/26
959.	comp33728_c0_seq1	4956	A	G	A	7/9	34/34
960.	comp33728_c0_seq1	4971	C	T	C	7/7	44/44
961.	comp33728_c0_seq1	5090	G	T	G	44/51	165/168
962.	comp33728_c0_seq1	5121	G	T	G	65/84	176/176
963.	comp33728_c0_seq1	5160	A	G	A	81/122	176/176
964.	comp33728_c0_seq1	5301	T	A	T	80/121	262/266
965.	comp33728_c0_seq1	5352	T	C	T	61/85	292/292
966.	comp33728_c0_seq1	5364	G	A	G	61/83	284/286
967.	comp33728_c0_seq1	5382	A	T	A	38/57	228/228
968.	comp33728_c0_seq1	5409	A	G	A	53/53	172/178
969.	comp33728_c0_seq1	5454	A	G	A	44/52	122/122
970.	comp33728_c0_seq1	5487	G	A	G	51/74	151/151
971.	comp33728_c0_seq1	5522	A	G	A	72/72	138/138

972.	comp33728_c0_seq1	5610	A	G	A	19/19	108/108
973.	comp33728_c0_seq1	5697	C	T	C	20/20	102/102
974.	comp33728_c0_seq1	5730	A	G	A	19/21	80/82
975.	comp33728_c0_seq1	5775	A	G	A	46/46	80/82
976.	comp33728_c0_seq1	5871	C	T	C	72/72	212/213
977.	comp33728_c0_seq1	5879	G	T	G	57/59	208/210
978.	comp33728_c0_seq1	5898	G	A	G	66/69	190/190
979.	comp33728_c0_seq1	5961	A	G	A	70/72	122/127
980.	comp33728_c0_seq1	6084	T	C	T	38/52	73/73
981.	comp33728_c0_seq1	6111	G	A	G	42/45	89/94
982.	comp33728_c0_seq1	6174	C	T	C	74/74	124/124
983.	comp33728_c0_seq1	6231	A	G	A	56/56	110/111
984.	comp33728_c0_seq1	6234	A	G	A	59/59	101/101
985.	comp33728_c0_seq1	6303	C	A	C	32/32	118/118
986.	comp33728_c0_seq1	6336	C	T	C	23/23	152/153
987.	comp33728_c0_seq1	6342	C	T	C	23/23	159/159
988.	comp33728_c0_seq1	6357	G	A	G	18/18	174/175
989.	comp33728_c0_seq1	6378	A	G	A	17/19	144/144
990.	comp33728_c0_seq1	6561	C	C	T	131/169	53/53
991.	comp33728_c0_seq1	6573	T	T	C	154/188	63/63
992.	comp33728_c0_seq1	6588	T	T	C	306/342	53/53
993.	comp33728_c0_seq1	6674	G	G	A	380/390	28/28
994.	comp33728_c0_seq1	6693	C	C	T	152/166	28/28
995.	comp33728_c0_seq1	6702	A	A	C	112/125	32/34
996.	comp33728_c0_seq1	6707	T	T	C	99/114	38/38
997.	comp33728_c0_seq1	6708	G	G	A	94/109	36/36
998.	comp33728_c0_seq1	6738	T	C	T	50/67	54/54
999.	comp33728_c0_seq1	6747	G	A	G	50/69	64/64
1000.	comp33728_c0_seq1	6756	G	A	G	50/72	90/90
1001.	comp33728_c0_seq1	6765	A	G	A	70/72	94/94
1002.	comp33728_c0_seq1	6807	T	C	T	38/38	118/118
1003.	comp33728_c0_seq1	6876	T	C	T	17/24	134/134
1004.	comp33728_c0_seq1	6882	A	G	A	19/26	130/130
1005.	comp33728_c0_seq1	6891	T	C	T	22/33	142/142
1006.	comp33728_c0_seq1	6942	G	A	G	50/50	136/136
1007.	comp33728_c0_seq1	6963	A	G	A	59/61	121/122
1008.	comp33728_c0_seq1	7008	A	G	A	48/72	156/156
1009.	comp33728_c0_seq1	7182	G	A	G	23/33	40/40
1010.	comp33728_c0_seq1	7305	T	C	T	41/47	78/78
1011.	comp33728_c0_seq1	7308	A	G	A	39/41	63/63
1012.	comp33728_c0_seq1	7374	A	G	A	28/30	118/118
1013.	comp33728_c0_seq1	7388	A	G	A	31/31	143/143
1014.	comp33728_c0_seq1	7403	G	A	G	30/30	144/144
1015.	comp33728_c0_seq1	7431	C	T	C	14/14	166/166
1016.	comp33728_c0_seq1	7458	A	G	A	14/14	132/132
1017.	comp33728_c0_seq1	7463	C	T	C	17/17	140/140
1018.	comp33728_c0_seq1	7470	T	A	T	16/20	136/136
1019.	comp33728_c0_seq1	7488	G	A	G	18/21	130/130
1020.	comp33728_c0_seq1	7530	C	T	C	27/27	100/100
1021.	comp33728_c0_seq1	7542	C	T	C	26/26	98/98
1022.	comp33728_c0_seq1	7587	G	A	G	25/25	100/100
1023.	comp33728_c0_seq1	7593	A	T	A	26/26	100/102
1024.	comp33728_c0_seq1	7699	C	T	C	21/21	64/64
1025.	comp33728_c0_seq1	7707	T	C	T	21/21	66/68
1026.	comp33728_c0_seq1	7709	G	A	G	21/22	72/72
1027.	comp33728_c0_seq1	7713	T	C	T	23/23	74/74
1028.	comp33728_c0_seq1	7721	G	T	G	21/22	83/84
1029.	comp33728_c0_seq1	7734	G	T	G	22/22	94/94
1030.	comp33728_c0_seq1	7749	C	T	C	23/25	106/106
1031.	comp33728_c0_seq1	7755	C	T	C	22/28	96/96
1032.	comp33728_c0_seq1	7758	T	C	T	24/26	90/90
1033.	comp33728_c0_seq1	8019	G	C	G	40/59	192/192
1034.	comp33728_c0_seq1	8052	G	A	G	37/37	176/178
1035.	comp33728_c0_seq1	8112	A	G	A	17/20	97/97
1036.	comp33728_c0_seq1	8135	A	G	A	18/27	104/104
1037.	comp33728_c0_seq1	8154	T	C	T	25/32	102/104
1038.	comp33728_c0_seq1	8466	A	G	A	42/49	78/94
1039.	comp33728_c0_seq1	8487	G	A	G	46/46	74/110
1040.	comp33728_c0_seq1	8505	T	G	T	31/46	112/112
1041.	comp33728_c0_seq1	8514	A	G	A	36/52	134/134
1042.	comp33728_c0_seq1	8562	A	G	A	38/38	155/167
1043.	comp33728_c0_seq1	8577	T	C	T	46/46	172/172
1044.	comp33728_c0_seq1	8634	G	T	G	25/30	128/128
1045.	comp33728_c0_seq1	8658	A	G	A	31/36	120/120
1046.	comp33728_c0_seq1	8664	G	A	G	36/36	108/108

1047.	comp33728_c0_seq1	8941	T	C	T	73/101	164/164
1048.	comp33728_c0_seq1	8943	G	A	G	74/100	164/164
1049.	comp33728_c0_seq1	8991	T	C	T	41/41	104/106
1050.	comp33728_c0_seq1	9014	T	C	T	29/39	72/73
1051.	comp33728_c0_seq1	9159	C	T	C	41/44	120/120
1052.	comp33728_c0_seq1	9249	T	C	T	62/89	178/178
1053.	comp33728_c0_seq1	9303	A	G	A	53/53	111/112
1054.	comp33728_c0_seq1	9315	A	G	A	49/49	93/93
1055.	comp33728_c0_seq1	9372	C	G	C	65/65	116/116
1056.	comp33728_c0_seq1	9450	A	G	A	27/30	90/90
1057.	comp33728_c0_seq1	9471	A	G	A	27/28	62/62
1058.	comp33728_c0_seq1	9498	T	C	T	33/46	97/97
1059.	comp33728_c0_seq1	9567	T	C	T	54/55	118/118
1060.	comp33728_c0_seq1	9594	G	A	G	47/48	100/102
1061.	comp33728_c0_seq1	9642	T	C	T	35/48	100/100
1062.	comp33728_c0_seq1	9658	C	T	C	34/47	92/94
1063.	comp33728_c0_seq1	9660	A	G	A	35/48	97/97
1064.	comp33728_c0_seq1	9663	A	G	A	36/52	102/103
1065.	comp33728_c0_seq1	9845	A	G	A	52/77	68/69
1066.	comp33728_c0_seq1	9918	T	C	T	28/29	36/36
1067.	comp33728_c0_seq1	10095	G	A	G	17/18	94/94
1068.	comp33728_c0_seq1	10107	C	T	C	21/21	90/92
1069.	comp33728_c0_seq1	10281	C	T	C	55/83	156/158
1070.	comp33728_c0_seq1	10305	G	A	G	49/49	136/136
1071.	comp33728_c0_seq1	10350	A	G	A	22/22	102/102
1072.	comp33728_c0_seq1	10353	T	C	T	23/23	98/102
1073.	comp33728_c0_seq1	10368	C	T	C	25/25	118/120
1074.	comp33728_c0_seq1	10464	G	C	G	20/20	104/107
1075.	comp33728_c0_seq1	10542	G	A	G	21/21	92/92
1076.	comp33728_c0_seq1	10548	A	G	A	20/20	83/83
1077.	comp33728_c0_seq1	10566	T	C	T	23/23	62/62
1078.	comp33728_c0_seq1	10605	C	C	T	21/29	51/51
1079.	comp33728_c0_seq1	10607	C	C	T	22/31	50/51
1080.	comp33728_c0_seq1	10665	A	G	A	36/47	62/62
1081.	comp33728_c0_seq1	10683	A	G	A	30/40	66/66
1082.	comp33728_c0_seq1	10692	A	C	A	30/41	72/72
1083.	comp33728_c0_seq1	10695	G	A	G	30/43	68/68
1084.	comp33728_c0_seq1	10752	A	A	G	93/93	83/83
1085.	comp33728_c0_seq1	10872	A	G	A	29/41	142/142
1086.	comp33728_c0_seq1	10881	A	G	A	26/26	98/145
1087.	comp33728_c0_seq1	11019	T	C	T	10/10	108/108
1088.	comp33728_c0_seq1	11025	T	C	T	10/10	114/114
1089.	comp33728_c0_seq1	11085	T	C	T	20/22	156/158
1090.	comp33728_c0_seq1	11097	T	C	T	26/26	146/146
1091.	comp33728_c0_seq1	11181	T	C	T	27/27	103/103
1092.	comp33728_c0_seq1	11268	C	T	C	47/69	98/98
1093.	comp33728_c0_seq1	11272	T	C	T	49/71	96/96
1094.	comp33728_c0_seq1	11291	C	T	C	48/67	88/88
1095.	comp33728_c0_seq1	11772	G	A	G	31/44	122/122
1096.	comp33728_c0_seq1	11819	A	G	A	18/22	114/114
1097.	comp33728_c0_seq1	11838	T	C	T	16/18	90/90
1098.	comp33728_c0_seq1	11862	A	C	A	14/20	98/98
1099.	comp33728_c0_seq1	11889	A	G	A	19/25	70/70
1100.	comp33728_c0_seq1	11892	C	A	C	19/24	72/72
1101.	comp33728_c0_seq1	11910	T	C	T	29/29	56/56
1102.	comp33728_c0_seq1	11913	A	G	A	27/30	56/56
1103.	comp33728_c0_seq1	11952	G	A	G	23/28	57/57
1104.	comp33728_c0_seq1	11991	C	T	C	16/23	48/48
1105.	comp33728_c0_seq1	12021	A	T	A	24/24	42/42
1106.	comp33728_c0_seq1	12030	A	G	A	25/25	44/45
1107.	comp33728_c0_seq1	12099	T	C	T	58/58	94/94
1108.	comp33728_c0_seq1	12288	T	C	T	59/59	162/162
1109.	comp33728_c0_seq1	12309	T	C	T	40/40	155/155
1110.	comp33728_c0_seq1	12443	G	A	G	60/60	114/114
1111.	comp33728_c0_seq1	12576	T	C	T	34/36	83/83
1112.	comp33728_c0_seq1	12597	G	A	G	48/49	86/86
1113.	comp33728_c0_seq1	12615	C	T	C	51/52	106/106
1114.	comp33728_c0_seq1	12804	A	G	A	21/24	82/85
1115.	comp33728_c0_seq1	12810	T	C	T	28/34	104/104
1116.	comp33728_c0_seq1	12861	A	G	A	14/16	86/86
1117.	comp33728_c0_seq1	12876	G	C	G	10/12	76/76
1118.	comp33728_c0_seq1	12885	A	G	A	9/11	70/70
1119.	comp33728_c0_seq1	12888	T	C	T	11/11	70/70
1120.	comp33728_c0_seq1	12897	A	G	A	8/10	54/54
1121.	comp33715_c1_seq1	3186	A	G	A	16/24	32/40

1122.	comp33715_c1_seq1	3246	G	C	G	8/12	32/38
1123.	comp33261_c0_seq16	159	A	G	A	9/12	5/7
1124.	comp33261_c0_seq16	160	T	C	T	9/12	5/7
1125.	comp33261_c0_seq16	164	T	C	T	9/12	5/7
1126.	comp32289_c0_seq1	355	G	A	G	11/11	20/20
1127.	comp24846_c0_seq1	71	C	C	T	14/21	15/15
1128.	comp24846_c0_seq1	72	T	T	G	16/21	15/15
1129.	comp24846_c0_seq1	129	T	T	A	26/39	15/15
1130.	comp24846_c0_seq1	145	A	A	G	41/53	13/13
1131.	comp22031_c0_seq1	66	C	C	T	12/14	2/2
1132.	comp25879_c0_seq1	192	T	C	T	5/7	10/14
1133.	comp25879_c0_seq1	195	A	G	A	5/7	10/14
1134.	comp25879_c0_seq1	240	A	A	T	6/6	6/8
1135.	comp6841_c0_seq1	5	A	A	C	11/14	2/2
1136.	comp5208_c0_seq1	1480	T	C	T	28/28	32/32
1137.	comp31330_c0_seq1	370	A	A	G	7/7	4/6
1138.	comp31167_c0_seq2	172	T	C	T	6/9	16/18
1139.	comp33679_c0_seq1	2528	T	G	T	4/6	19/25
1140.	comp22941_c0_seq1	64	A	A	G	9/9	4/4
1141.	comp33033_c0_seq4	37	T	G	T	5/6	12/14
1142.	comp33033_c0_seq4	47	C	T	C	7/8	12/14
1143.	comp33033_c0_seq4	62	C	A	C	9/12	22/28
1144.	comp33033_c0_seq4	64	T	G	T	9/12	22/28
1145.	comp23832_c0_seq1	283	C	G	C	10/10	16/16
1146.	comp31360_c0_seq4	302	A	A	T	5/6	4/6
1147.	comp28923_c0_seq2	316	C	C	T	17/17	28/38
1148.	comp33493_c0_seq32	369	A	A	T	6/9	4/4
1149.	comp30469_c0_seq2	247	A	A	C	34/34	16/16
1150.	comp33584_c0_seq4	1412	T	A	T	4/6	9/13
1151.	comp33584_c0_seq4	1416	C	G	C	4/6	9/13
1152.	comp33584_c0_seq4	1419	G	T	G	4/6	9/13
1153.	comp12638_c0_seq1	56	C	C	T	5/7	4/4
1154.	comp12638_c0_seq1	62	T	T	C	5/7	4/4
1155.	comp12638_c0_seq1	74	T	T	A	6/8	4/4
1156.	comp12638_c0_seq1	84	A	A	G	6/8	4/4
1157.	comp30975_c0_seq6	71	C	C	A	9/9	8/12
1158.	comp33393_c0_seq4	893	A	A	T	7/8	6/6
1159.	comp33393_c0_seq4	899	C	C	A	12/12	6/8
1160.	comp184148_c0_seq1	139	C	C	T	8/8	4/4
1161.	comp33709_c1_seq1	4246	A	G	A	23/23	32/32
1162.	comp33177_c0_seq2	4606	A	T	A	12/12	27/27
1163.	comp33700_c0_seq1	1167	A	A	C	6/8	4/6
1164.	comp33700_c0_seq1	1168	G	G	T	6/8	4/6
1165.	comp33700_c0_seq1	1172	C	C	G	6/8	4/6
1166.	comp30930_c0_seq1	2	C	G	T	4/6	6/6
1167.	comp30930_c0_seq1	25	C	T	C	7/7	8/8
1168.	comp28147_c0_seq1	18	T	A	T	4/6	2/2
1169.	comp28951_c0_seq1	195	C	C	T	17/24	2/2
1170.	comp33659_c0_seq1	1980	T	T	G	6/9	18/26
1171.	comp31028_c0_seq1	46	C	C	T	15/20	2/2
1172.	comp31028_c0_seq1	64	C	C	A	20/25	6/6
1173.	comp31028_c0_seq1	303	G	G	C	28/41	10/12
1174.	comp31028_c0_seq1	304	A	A	G	28/41	10/12
1175.	comp31028_c0_seq1	334	C	G	C	21/29	8/8
1176.	comp16085_c0_seq1	81	G	A	G	4/6	2/2
1177.	comp33656_c0_seq27	84	A	G	A	9/13	8/12
1178.	comp33656_c0_seq27	91	C	A	C	12/16	8/12
1179.	comp33656_c0_seq27	93	G	A	G	12/16	8/12
1180.	comp33514_c0_seq37	296	G	G	A	5/7	4/6
1181.	comp15287_c0_seq1	769	G	T	G	6/6	16/18
1182.	comp33425_c0_seq11	1668	T	C	T	4/6	8/9
1183.	comp33425_c0_seq11	1673	G	T	G	4/6	8/8
1184.	comp32429_c0_seq3	339	A	G	A	6/7	6/8
1185.	comp32429_c0_seq3	366	A	T	A	6/6	7/7
1186.	comp32429_c0_seq3	648	C	T	C	4/6	5/7
1187.	comp32534_c0_seq7	358	A	T	A	16/21	11/15
1188.	comp33595_c0_seq9	83	T	T	C	13/16	10/14
1189.	comp33595_c0_seq9	105	C	C	T	13/17	10/14
1190.	comp33595_c0_seq9	141	G	G	A	18/24	20/28
1191.	comp7718_c0_seq1	257	T	T	C	20/20	22/22
1192.	comp24974_c0_seq2	115	A	G	A	9/9	10/10
1193.	comp29446_c0_seq1	585	A	G	A	6/7	20/20
1194.	comp32945_c0_seq6	541	A	A	G	9/10	12/18
1195.	comp32945_c0_seq6	569	T	A	T	8/9	12/16
1196.	comp32945_c0_seq6	574	C	T	C	7/8	15/19

1197.	comp32945_c0_seq6	577	T	A	T	7/8	16/21
1198.	comp32945_c0_seq6	583	T	C	T	4/6	16/18
1199.	comp32945_c0_seq6	584	C	T	C	4/6	16/18
1200.	comp32945_c0_seq6	586	A	G	A	4/6	17/17
1201.	comp26191_c0_seq1	557	A	T	A	4/6	40/42
1202.	comp32421_c0_seq13	320	A	T	A	10/10	6/6
1203.	comp28784_c0_seq1	623	A	G	A	6/6	18/18
1204.	comp30251_c0_seq3	3	C	C	A	6/6	4/4
1205.	comp30251_c0_seq3	4	T	T	A	6/6	4/4
1206.	comp52214_c0_seq1	4	A	T	A	6/6	4/4
1207.	comp150530_c0_seq1	142	T	C	T	4/6	4/6
1208.	comp18868_c0_seq1	1548	A	G	A	7/7	26/26
1209.	comp33644_c0_seq5	531	T	T	C	12/13	10/14
1210.	comp33644_c0_seq5	538	C	C	T	9/10	12/18
1211.	comp33644_c0_seq5	812	T	T	C	8/9	4/6
1212.	comp33644_c0_seq5	813	G	G	A	8/9	4/6
1213.	comp33644_c0_seq5	820	T	T	C	6/7	4/6
1214.	comp33644_c0_seq5	823	C	C	T	6/7	4/6
1215.	comp33644_c0_seq5	826	A	A	C	5/6	4/6
1216.	comp33644_c0_seq5	839	G	G	A	5/6	4/4
1217.	comp26961_c0_seq2	272	A	A	G	6/8	8/8
1218.	comp337097_c0_seq1	69	C	T	C	4/6	2/2
1219.	comp30447_c0_seq1	376	T	T	C	7/7	4/6
1220.	comp32404_c0_seq1	581	A	G	A	10/10	10/10
1221.	comp32404_c0_seq1	671	T	C	T	8/8	4/4
1222.	comp33175_c0_seq2	764	A	A	G	9/11	2/2
1223.	comp32052_c0_seq2	1022	A	G	A	7/10	10/10
1224.	comp29320_c0_seq1	4	C	C	T	69/92	2/2
1225.	comp20762_c0_seq1	338	T	T	C	6/6	2/2
1226.	comp33178_c0_seq5	865	C	T	C	6/8	8/8
1227.	comp33178_c0_seq5	868	T	G	T	6/7	6/6
1228.	comp33178_c0_seq5	874	T	G	T	7/8	4/4
1229.	comp33178_c0_seq5	883	A	T	A	8/11	2/2
1230.	comp33178_c0_seq5	919	G	T	G	11/16	6/6
1231.	comp33178_c0_seq5	928	A	G	A	11/16	6/6
1232.	comp33178_c0_seq5	937	T	C	T	10/15	6/6
1233.	comp33712_c0_seq4	86	A	A	G	7/10	19/25
1234.	comp27033_c0_seq1	363	A	G	A	11/11	7/7
1235.	comp32487_c0_seq2	2	C	C	G	6/6	4/6
1236.	comp23674_c0_seq1	958	T	C	T	6/6	16/16
1237.	comp33640_c0_seq42	686	T	T	G	6/7	2/2
1238.	comp33640_c0_seq42	689	T	T	A	6/7	2/2
1239.	comp27453_c0_seq1	104	C	T	C	6/6	4/4
1240.	comp11629_c0_seq1	18	C	C	T	5/6	2/2
1241.	comp33722_c0_seq6	334	T	C	T	43/61	41/59
1242.	comp33722_c0_seq6	394	A	A	G	48/67	32/40
1243.	comp33722_c0_seq6	427	T	T	C	36/51	36/36
1244.	comp33722_c0_seq6	436	C	C	A	30/39	30/42
1245.	comp31918_c0_seq1	543	T	T	C	5/6	8/10
1246.	comp30882_c0_seq1	1482	G	T	G	6/7	15/17
1247.	comp33596_c0_seq2	2286	T	G	T	5/7	4/4
1248.	comp31987_c0_seq1	1677	C	T	C	6/8	14/14
1249.	comp25740_c0_seq1	321	C	A	C	4/6	8/8
1250.	comp32461_c0_seq1	259	C	C	G	9/12	10/10
1251.	comp31415_c0_seq1	381	T	T	A	20/22	4/6
1252.	comp31415_c0_seq1	412	G	G	A	14/14	4/4
1253.	comp31415_c0_seq1	421	C	C	T	11/11	4/4
1254.	comp31415_c0_seq1	693	A	A	G	13/15	2/2
1255.	comp31415_c0_seq1	701	C	C	T	13/15	2/2
1256.	comp28134_c0_seq2	769	C	T	C	7/8	10/10
1257.	comp33418_c0_seq1	35	G	A	G	8/11	4/6
1258.	comp33724_c0_seq4	156	G	G	A	20/21	4/4
1259.	comp33724_c0_seq4	195	A	G	A	28/39	12/14
1260.	comp33609_c0_seq6	295	A	G	A	7/7	12/18
1261.	comp33609_c0_seq6	297	T	C	T	6/6	12/18
1262.	comp33609_c0_seq6	927	C	C	T	6/6	4/6
1263.	comp23160_c0_seq1	1912	T	A	T	23/23	32/32
1264.	comp32100_c0_seq1	3084	C	C	A	17/17	10/10
1265.	comp19019_c0_seq1	297	G	G	A	29/40	13/19
1266.	comp19019_c0_seq1	298	A	A	G	29/39	13/19
1267.	comp19019_c0_seq1	340	T	T	C	9/10	6/6
1268.	comp25912_c0_seq1	4	T	G	T	6/8	2/2
1269.	comp32568_c0_seq1	268	C	T	C	5/6	11/13
1270.	comp32568_c0_seq1	280	T	C	T	9/9	11/15
1271.	comp31258_c0_seq2	73	C	T	C	4/6	6/6

1272.	comp33206_c0_seq36	18	A	A	G	8/11	9/11
1273.	comp30928_c0_seq1	459	T	C	T	14/21	30/44
1274.	comp17542_c0_seq1	27	A	C	A	4/6	16/24
1275.	comp33723_c0_seq4	386	C	T	C	8/12	21/24
1276.	comp33723_c0_seq4	466	A	A	G	7/10	17/25
1277.	comp33723_c0_seq4	895	G	G	A	16/18	15/18
1278.	comp33723_c0_seq4	963	A	C	A	12/17	21/23
1279.	comp33723_c0_seq4	1129	A	G	A	8/10	20/26
1280.	comp33723_c0_seq4	1191	G	A	G	16/24	30/37
1281.	comp31839_c0_seq3	206	T	C	T	5/7	12/18
1282.	comp31839_c0_seq3	209	T	C	T	5/7	14/20
1283.	comp31839_c0_seq3	211	G	A	G	5/7	14/20
1284.	comp31839_c0_seq3	215	T	C	T	5/7	12/16
1285.	comp33501_c0_seq1	826	G	C	G	7/7	39/39
1286.	comp32665_c0_seq1	235	G	A	G	4/6	2/2
1287.	comp32665_c0_seq1	238	A	C	A	4/6	2/2
1288.	comp32665_c0_seq1	240	G	A	G	4/6	2/2
1289.	comp7891_c0_seq1	239	C	C	T	7/10	2/2
1290.	comp7891_c0_seq1	242	T	T	G	7/10	2/2
1291.	comp29953_c0_seq5	430	T	A	T	5/7	30/40
1292.	comp29953_c0_seq5	453	G	T	G	7/8	22/32
1293.	comp25136_c0_seq1	468	G	A	G	4/6	6/6
1294.	comp18084_c0_seq1	60	T	C	T	5/6	3/3
1295.	comp32874_c0_seq27	1342	T	T	G	6/7	3/3
1296.	comp32874_c0_seq27	1343	T	T	G	6/7	3/3
1297.	comp32874_c0_seq27	1347	T	T	C	6/7	3/3
1298.	comp33699_c0_seq1	237	C	C	G	21/21	10/14
1299.	comp23864_c0_seq1	90	T	T	C	5/6	2/2
1300.	comp6301_c0_seq1	189	G	G	A	49/49	20/20
1301.	comp19590_c0_seq1	219	G	T	G	8/8	14/14
1302.	comp47778_c0_seq1	45	A	A	C	26/26	4/4
1303.	comp27114_c0_seq2	515	C	T	C	13/14	16/16
1304.	comp30977_c0_seq7	56	A	A	T	4/6	4/6
1305.	comp30977_c0_seq7	57	A	A	G	4/6	4/6
1306.	comp33558_c0_seq1	1373	C	T	C	6/6	12/12
1307.	comp32451_c0_seq1	1988	A	A	T	6/6	6/6
1308.	comp28247_c0_seq1	236	G	G	A	34/34	26/26
1309.	comp26331_c0_seq1	199	T	T	C	74/74	10/10
1310.	comp175276_c0_seq1	4	T	T	A	10/10	2/2
1311.	comp30875_c0_seq1	4	T	A	T	5/7	6/8
1312.	comp38350_c0_seq1	5	T	A	T	9/10	4/4
1313.	comp29990_c0_seq2	238	G	A	G	6/7	13/19
1314.	comp29990_c0_seq2	250	A	G	A	5/6	16/22
1315.	comp29990_c0_seq2	253	A	T	A	5/7	17/25
1316.	comp29990_c0_seq2	256	C	A	C	5/7	21/27
1317.	comp32612_c0_seq2	932	A	A	T	8/8	7/9
1318.	comp32612_c0_seq2	938	C	C	G	7/7	7/9
1319.	comp32612_c0_seq2	952	T	T	A	6/6	9/11
1320.	comp23684_c0_seq1	157	C	C	T	10/14	2/3
1321.	comp23684_c0_seq1	158	A	A	G	10/15	2/2
1322.	comp23684_c0_seq1	159	T	T	G	10/15	2/3
1323.	comp23684_c0_seq1	175	G	G	A	13/18	2/2
1324.	comp21858_c0_seq1	114	A	A	G	12/14	2/2
1325.	comp33338_c0_seq9	378	A	A	G	5/7	4/6
1326.	comp31023_c0_seq7	866	A	G	A	5/7	4/6
1327.	comp26623_c0_seq1	18	G	G	T	7/7	4/6
1328.	comp33695_c0_seq8	261	T	C	T	11/11	32/32
1329.	comp4717_c0_seq1	137	C	C	A	12/12	4/4
1330.	comp31650_c0_seq2	284	C	T	C	8/10	2/2
1331.	comp8519_c0_seq1	190	G	T	G	8/8	6/6
1332.	comp31930_c0_seq1	566	T	T	C	8/8	10/14
1333.	comp31930_c0_seq1	569	C	C	A	9/9	14/18
1334.	comp31176_c0_seq1	2432	G	A	G	14/14	18/18
1335.	comp33030_c0_seq7	193	G	G	A	19/19	14/16
1336.	comp33267_c0_seq1	247	A	C	A	11/15	20/22
1337.	comp33267_c0_seq1	248	A	G	A	10/14	20/22
1338.	comp32671_c0_seq3	111	C	G	C	13/13	18/22
1339.	comp32671_c0_seq3	183	A	C	A	5/7	8/12
1340.	comp27452_c0_seq1	76	G	A	G	6/6	6/8
1341.	comp301892_c0_seq1	177	C	T	C	6/6	6/6
1342.	comp25944_c0_seq6	626	G	C	G	4/6	4/6
1343.	comp24416_c0_seq1	30	C	C	T	5/7	2/2
1344.	comp32729_c0_seq9	447	A	A	C	6/9	6/9
1345.	comp13072_c0_seq1	229	A	T	A	6/6	8/8
1346.	comp33206_c0_seq3	448	A	T	A	15/20	35/49

1347.	comp17645_c0_seq3	53	T	T	C	10/15	2/2
1348.	comp29155_c0_seq1	70	A	A	T	78/111	4/4
1349.	comp29155_c0_seq1	169	T	T	C	41/44	4/6
1350.	comp29155_c0_seq1	179	C	T	C	26/37	4/4
1351.	comp41783_c0_seq1	96	A	A	G	9/13	3/3
1352.	comp41783_c0_seq1	98	A	A	T	9/12	3/3
1353.	comp32582_c0_seq2	1814	C	C	G	17/17	16/16
1354.	comp23770_c0_seq1	5	C	C	T	9/9	2/2
1355.	comp59237_c0_seq1	207	A	A	C	10/12	2/2
1356.	comp31872_c0_seq1	227	A	A	G	5/6	2/2
1357.	comp55272_c0_seq1	184	A	A	G	42/42	18/18
1358.	comp27878_c0_seq1	80	T	T	A	6/8	8/10
1359.	comp9074_c0_seq1	94	C	G	C	6/9	4/6
1360.	comp32811_c0_seq1	192	A	A	G	18/26	12/18
1361.	comp32811_c0_seq1	788	A	A	G	35/53	12/18
1362.	comp18065_c0_seq1	4	C	C	T	8/10	20/29
1363.	comp32934_c0_seq1	982	C	A	C	7/7	34/34
1364.	comp33221_c0_seq1	213	A	G	A	4/6	8/8
1365.	comp33179_c0_seq1	868	C	T	C	20/29	26/32
1366.	comp25549_c0_seq2	534	G	G	A	23/23	18/20
1367.	comp33142_c0_seq5	64	A	A	G	6/6	10/10
1368.	comp33142_c0_seq5	214	T	G	T	6/6	8/10
1369.	comp18824_c0_seq2	1592	C	C	T	31/33	29/41
1370.	comp18824_c0_seq2	1607	G	G	T	33/36	25/37
1371.	comp13346_c0_seq1	686	A	A	G	7/10	6/6
1372.	comp13346_c0_seq1	692	C	C	A	6/9	6/6
1373.	comp13346_c0_seq1	693	A	A	G	6/9	6/6
1374.	comp13346_c0_seq1	698	T	T	G	6/9	6/6
1375.	comp31814_c0_seq1	214	G	A	G	6/6	16/16
1376.	comp31814_c0_seq1	413	T	C	T	7/9	8/10
1377.	comp28834_c0_seq1	84	C	T	C	5/7	8/10
1378.	comp33578_c0_seq1	540	A	A	G	17/21	32/46
1379.	comp33578_c0_seq1	2959	G	G	C	17/17	8/8
1380.	comp30815_c0_seq1	236	G	G	A	9/9	4/4
1381.	comp31711_c0_seq2	244	G	G	T	98/115	4/6
1382.	comp31711_c0_seq2	262	C	C	A	48/64	4/6
1383.	comp19041_c0_seq1	191	A	G	A	21/31	56/56
1384.	comp21036_c0_seq1	121	G	G	A	7/8	6/8
1385.	comp21036_c0_seq1	154	G	G	A	8/9	4/4
1386.	comp21036_c0_seq1	168	T	T	C	9/10	4/4
1387.	comp21036_c0_seq1	180	G	G	C	9/10	4/4
1388.	comp32688_c0_seq18	122	G	T	G	9/13	18/25
1389.	comp32291_c0_seq1	563	G	A	G	7/7	20/20
1390.	comp33501_c0_seq14	1342	T	T	C	6/6	4/6
1391.	comp9706_c0_seq1	118	G	A	G	6/6	16/16
1392.	comp33332_c0_seq11	1003	G	G	A	10/10	18/18
1393.	comp30468_c0_seq4	72	C	C	A	26/30	8/12
1394.	comp26655_c0_seq1	248	C	T	C	5/7	18/18
1395.	comp28067_c0_seq1	293	C	C	T	12/12	12/18
1396.	comp33523_c0_seq3	671	G	A	G	10/11	5/7
1397.	comp28650_c0_seq1	264	G	A	G	4/6	4/4
1398.	comp28650_c0_seq1	269	G	A	G	4/6	4/4
1399.	comp31863_c0_seq2	488	C	C	T	32/34	6/7
1400.	comp31863_c0_seq2	528	A	A	G	35/43	6/6
1401.	comp31863_c0_seq2	530	T	T	A	35/47	6/6
1402.	comp32820_c0_seq3	3620	A	G	A	6/6	8/8
1403.	comp155168_c0_seq1	134	G	T	G	8/8	8/8
1404.	comp28536_c0_seq1	754	G	C	G	10/10	18/18
1405.	comp27835_c0_seq1	109	A	A	G	12/12	12/12
1406.	comp76429_c0_seq1	101	C	C	T	22/25	4/4
1407.	comp76429_c0_seq1	114	A	A	C	31/34	4/6
1408.	comp20837_c0_seq2	198	A	T	A	7/7	7/9
1409.	comp20837_c0_seq2	201	C	A	C	6/6	7/7
1410.	comp27337_c0_seq1	120	A	G	A	12/15	16/24
1411.	comp27337_c0_seq1	121	A	G	A	12/15	16/24
1412.	comp288247_c0_seq1	86	C	T	C	7/7	6/6
1413.	comp288247_c0_seq1	101	A	G	A	8/8	2/2
1414.	comp288247_c0_seq1	102	T	C	T	8/8	2/2
1415.	comp29676_c0_seq1	673	G	G	C	4/6	24/36
1416.	comp29676_c0_seq1	690	T	T	A	4/6	20/30
1417.	comp33727_c0_seq14	132	C	T	C	6/9	4/6
1418.	comp33727_c0_seq14	468	G	G	A	8/12	2/2
1419.	comp24916_c0_seq5	122	G	A	G	15/16	4/4
1420.	comp29363_c0_seq1	1318	C	T	C	10/10	26/26
1421.	comp8915_c0_seq1	904	C	T	C	6/7	2/2

1422.	comp28338_c0_seq2	401	T	A	T	5/7	6/8
1423.	comp32793_c1_seq1	231	G	G	A	5/7	4/4
1424.	comp33668_c0_seq1	2332	A	A	G	8/8	6/8
1425.	comp33668_c0_seq1	2336	A	A	G	8/8	6/8
1426.	comp33668_c0_seq1	2339	T	T	C	8/8	6/8
1427.	comp33668_c0_seq1	2351	C	C	T	8/8	10/10
1428.	comp33668_c0_seq1	2353	G	G	A	8/8	10/10
1429.	comp33668_c0_seq1	2356	C	C	T	8/8	10/10
1430.	comp33668_c0_seq1	2365	T	T	A	7/7	10/10
1431.	comp33668_c0_seq1	2378	A	A	C	7/7	12/12
1432.	comp33668_c0_seq1	2384	T	T	A	7/7	12/12
1433.	comp33668_c0_seq1	2470	T	C	T	5/6	22/22
1434.	comp33668_c0_seq1	4796	G	C	G	5/6	12/14
1435.	comp31142_c0_seq1	550	A	G	A	6/6	17/17
1436.	comp26890_c0_seq1	819	G	C	G	6/7	12/14
1437.	comp33656_c0_seq20	309	G	A	G	5/7	12/14
1438.	comp33656_c0_seq20	312	A	G	A	10/13	18/20
1439.	comp31217_c0_seq1	556	T	T	G	17/18	10/12
1440.	comp24992_c0_seq1	133	T	C	T	9/9	16/18
1441.	comp28778_c0_seq1	405	G	C	G	7/7	21/21
1442.	comp32033_c0_seq1	26	G	T	G	4/6	2/2
1443.	comp32033_c0_seq1	27	C	G	C	4/6	2/2
1444.	comp32033_c0_seq1	29	A	G	A	4/6	4/4
1445.	comp28179_c0_seq1	81	A	T	A	6/6	16/16
1446.	comp31792_c0_seq1	134	A	A	G	12/12	4/4
1447.	comp31792_c0_seq1	146	T	T	C	17/17	4/4
1448.	comp19360_c0_seq1	193	A	A	G	23/23	4/4
1449.	comp27124_c0_seq3	192	A	A	G	8/11	2/2
1450.	comp32316_c0_seq1	353	A	G	A	14/14	28/30
1451.	comp26621_c0_seq1	617	C	A	C	4/6	8/10
1452.	comp26621_c0_seq1	618	A	G	A	4/6	8/10
1453.	comp32088_c0_seq1	471	A	A	C	6/6	10/12
1454.	comp31888_c0_seq6	19	A	A	G	7/10	4/6
1455.	comp33417_c1_seq2	7465	T	T	C	76/99	2/2
1456.	comp33417_c1_seq2	7497	A	A	T	107/155	2/2
1457.	comp33417_c1_seq2	7901	A	A	G	92/110	4/4
1458.	comp33417_c1_seq2	7942	C	C	T	59/72	4/4
1459.	comp12454_c0_seq1	60	G	G	A	7/7	6/6
1460.	comp24242_c0_seq1	103	A	G	A	7/10	10/14
1461.	comp24242_c0_seq1	104	C	A	C	7/10	10/14
1462.	comp15446_c0_seq1	159	C	T	C	14/14	24/24
1463.	comp33521_c0_seq15	317	C	C	T	5/6	5/7
1464.	comp33521_c0_seq15	2194	T	C	T	12/16	20/23
1465.	comp33521_c0_seq15	2202	T	C	T	12/16	12/17
1466.	comp33521_c0_seq15	2203	T	C	T	12/16	11/15
1467.	comp6811_c0_seq1	75	A	A	G	12/12	6/6
1468.	comp6811_c0_seq1	171	A	A	C	55/55	52/53
1469.	comp31671_c0_seq1	10	T	C	T	31/38	52/58
1470.	comp13739_c0_seq1	128	T	C	T	4/6	4/6
1471.	comp30339_c0_seq1	409	T	C	T	4/6	12/13
1472.	comp30339_c0_seq1	418	T	C	T	5/7	10/10
1473.	comp30339_c0_seq1	442	G	C	G	5/7	16/16
1474.	comp41629_c0_seq1	168	A	T	A	29/39	74/74
1475.	comp43882_c0_seq1	4	T	T	A	6/9	12/18
1476.	comp26782_c0_seq3	209	C	C	T	4/6	4/6
1477.	comp17779_c0_seq1	279	C	A	C	15/15	20/20
1478.	comp128705_c0_seq1	206	G	G	C	6/6	4/4
1479.	comp128705_c0_seq1	207	T	T	G	6/6	4/4
1480.	comp128705_c0_seq1	208	T	T	C	6/6	4/4
1481.	comp19401_c0_seq1	63	A	G	A	7/7	6/8
1482.	comp6972_c0_seq1	66	T	G	T	6/6	14/14
1483.	comp18667_c0_seq1	197	T	T	C	28/28	4/4
1484.	comp8312_c0_seq1	437	C	T	C	4/6	6/6
1485.	comp33521_c0_seq3	6289	T	C	T	7/7	10/13
1486.	comp32833_c0_seq1	32	G	C	G	5/6	6/8
1487.	comp33644_c0_seq1	220	A	G	A	5/7	10/10
1488.	comp33644_c0_seq1	225	C	A	C	6/8	12/12
1489.	comp33644_c0_seq1	232	G	A	G	6/7	10/10
1490.	comp33644_c0_seq1	235	C	G	C	5/7	6/6
1491.	comp33595_c0_seq14	115	T	C	T	6/8	10/14
1492.	comp32216_c0_seq4	694	A	A	G	48/53	9/9
1493.	comp28942_c0_seq1	45	A	C	A	10/10	2/2
1494.	comp33477_c0_seq1	1896	C	T	C	10/10	22/22
1495.	comp33477_c0_seq1	2585	T	C	T	8/8	6/6
1496.	comp32688_c0_seq13	129	C	A	C	6/6	12/16

1497.	comp33492_c1_seq18	473	C	C	T	10/11	8/12
1498.	comp33492_c1_seq18	513	T	C	T	15/20	54/70
1499.	comp33492_c1_seq18	528	T	C	T	13/19	66/90
1500.	comp31985_c0_seq1	16	T	A	T	11/15	2/2
1501.	comp22081_c0_seq1	276	A	A	G	12/12	14/18
1502.	comp31514_c0_seq1	945	T	G	T	9/11	14/20
1503.	comp31514_c0_seq1	954	C	T	C	7/9	18/24
1504.	comp31514_c0_seq1	967	G	T	G	6/8	16/20
1505.	comp551019_c0_seq1	148	C	C	T	4/6	2/2
1506.	comp31524_c0_seq1	3039	T	C	T	6/6	29/29
1507.	comp20598_c0_seq1	177	C	C	A	9/9	2/2
1508.	comp33722_c0_seq12	101	A	A	G	15/19	6/6
1509.	comp33722_c0_seq12	111	A	A	G	15/19	7/9
1510.	comp33722_c0_seq12	123	A	A	G	20/24	9/11
1511.	comp33722_c0_seq12	177	T	T	C	11/15	21/28
1512.	comp33722_c0_seq12	229	G	A	G	5/7	21/21
1513.	comp33437_c0_seq1	2095	A	C	A	7/9	34/42
1514.	comp33645_c0_seq24	973	T	C	T	14/17	12/17
1515.	comp28907_c0_seq1	115	G	G	T	7/10	10/12
1516.	comp28907_c0_seq1	122	A	A	G	13/13	12/14
1517.	comp28907_c0_seq1	125	G	G	A	13/13	12/14
1518.	comp28907_c0_seq1	149	T	T	C	14/15	20/24
1519.	comp28907_c0_seq1	152	C	C	T	14/15	18/22
1520.	comp28907_c0_seq1	155	A	A	T	15/16	20/24
1521.	comp28907_c0_seq1	160	G	G	T	15/16	20/24
1522.	comp28907_c0_seq1	173	C	C	G	15/17	18/22
1523.	comp28907_c0_seq1	199	C	G	C	13/17	18/22
1524.	comp33615_c0_seq11	112	A	G	A	6/6	8/8
1525.	comp105630_c0_seq1	25	T	T	C	8/9	6/6
1526.	comp31392_c0_seq3	166	A	A	T	7/7	8/10
1527.	comp32787_c0_seq1	1277	G	G	A	35/44	37/51
1528.	comp32787_c0_seq1	1286	G	G	A	34/43	35/49
1529.	comp33417_c1_seq1	7871	T	C	T	13/13	2/2
1530.	comp33417_c1_seq1	8297	T	T	C	128/129	2/2
1531.	comp33417_c1_seq1	8298	G	G	C	136/136	2/2
1532.	comp33417_c1_seq1	8360	A	A	G	82/103	8/8
1533.	comp33417_c1_seq1	8381	A	A	T	59/72	6/6
1534.	comp33417_c1_seq1	8382	T	T	C	58/72	6/6
1535.	comp33417_c1_seq1	8401	C	C	T	45/62	2/2
1536.	comp29141_c0_seq1	1066	T	G	T	19/19	24/24
1537.	comp31882_c0_seq1	420	T	C	T	7/10	4/6
1538.	comp33137_c0_seq1	255	T	T	C	144/198	16/24
1539.	comp23563_c0_seq1	4	G	G	T	5/6	2/2
1540.	comp25188_c0_seq1	792	C	C	T	8/8	8/8
1541.	comp31004_c0_seq1	698	C	G	C	9/9	20/20
1542.	comp31711_c0_seq4	199	C	T	C	5/6	2/2
1543.	comp31711_c0_seq4	204	A	G	A	5/6	2/2
1544.	comp32445_c0_seq7	220	T	C	T	9/11	26/28
1545.	comp33652_c0_seq6	1073	T	T	G	17/18	24/24
1546.	comp33652_c0_seq6	1853	G	A	G	4/6	10/10
1547.	comp32730_c0_seq1	100	C	T	C	8/12	33/38
1548.	comp30059_c0_seq1	447	T	T	C	7/8	7/9
1549.	comp30059_c0_seq1	465	C	C	T	10/12	7/9
1550.	comp30059_c0_seq1	467	T	T	C	10/12	8/10
1551.	comp30059_c0_seq1	504	T	T	G	8/9	6/8
1552.	comp33727_c0_seq12	487	G	G	A	27/35	4/6
1553.	comp33727_c0_seq12	493	C	C	T	23/32	6/6
1554.	comp33727_c0_seq12	499	T	T	C	23/32	6/6
1555.	comp29129_c0_seq1	125	C	C	G	10/11	2/2
1556.	comp29129_c0_seq1	141	A	A	G	8/8	2/2
1557.	comp30590_c0_seq1	252	A	A	C	8/12	6/8
1558.	comp33566_c0_seq3	327	T	C	T	7/7	12/12
1559.	comp33566_c0_seq3	1635	G	T	G	24/36	37/47
1560.	comp33566_c0_seq3	1676	T	T	C	27/33	18/24
1561.	comp31258_c0_seq1	351	C	T	C	12/15	16/16
1562.	comp19639_c0_seq1	143	A	A	C	7/7	4/4
1563.	comp22658_c0_seq1	347	C	G	C	9/10	16/18
1564.	comp69707_c0_seq1	3	G	G	T	12/15	2/2
1565.	comp32294_c1_seq5	1304	T	A	T	7/9	5/5
1566.	comp6751_c0_seq1	4	A	A	T	8/8	1/1
1567.	comp33371_c0_seq19	65	T	T	C	7/10	4/5
1568.	comp32829_c0_seq1	479	G	C	G	11/11	26/26
1569.	comp31133_c0_seq1	776	G	G	T	24/24	30/32
1570.	comp27170_c0_seq1	44	T	T	C	227/288	12/18
1571.	comp33521_c0_seq9	4774	T	C	T	9/12	17/23

1572.	comp11413_c0_seq1	112	A	G	A	7/7	12/13
1573.	comp30712_c0_seq1	677	A	G	A	7/7	8/8
1574.	comp5027_c0_seq1	254	T	C	T	6/6	4/4
1575.	comp27341_c1_seq1	754	A	A	G	6/8	2/2
1576.	comp27341_c1_seq1	766	A	A	G	6/9	2/2
1577.	comp27341_c1_seq1	775	G	G	C	6/9	2/2
1578.	comp24336_c0_seq1	266	C	C	T	9/9	12/12
1579.	comp168599_c0_seq1	120	T	T	G	4/6	6/8
1580.	comp25034_c0_seq1	19	G	C	G	6/6	12/12
1581.	comp33696_c0_seq4	546	T	C	T	5/6	13/17
1582.	comp125040_c0_seq1	439	C	T	C	4/6	2/2
1583.	comp103319_c0_seq1	4	A	A	T	6/7	8/10
1584.	comp26991_c0_seq2	324	C	C	A	4/6	4/4
1585.	comp26991_c0_seq2	325	A	A	C	5/6	4/4
1586.	comp19453_c0_seq1	63	C	T	C	4/6	4/6
1587.	comp31581_c0_seq2	190	A	C	A	4/6	4/4
1588.	comp15575_c0_seq1	2097	G	A	G	15/15	40/40
1589.	comp30405_c0_seq1	331	A	G	A	12/12	30/30
1590.	comp31083_c0_seq1	748	T	C	T	5/7	12/16
1591.	comp31083_c0_seq1	751	A	T	A	5/7	12/16
1592.	comp31083_c0_seq1	853	G	G	A	6/9	16/20
1593.	comp31083_c0_seq1	859	A	A	T	7/9	16/20
1594.	comp31083_c0_seq1	871	G	G	A	5/7	12/16
1595.	comp31083_c0_seq1	876	G	G	A	4/6	12/18
1596.	comp31083_c0_seq1	877	C	C	T	4/6	12/18
1597.	comp21969_c0_seq1	666	A	A	G	9/9	14/14
1598.	comp29417_c0_seq1	364	C	C	T	12/12	10/10
1599.	comp31700_c0_seq5	40	T	T	C	16/24	4/6
1600.	comp19909_c0_seq1	219	C	T	C	8/10	4/6
1601.	comp33644_c0_seq2	322	C	T	C	21/28	21/31
1602.	comp33644_c0_seq2	469	A	A	G	9/11	2/2
1603.	comp33644_c0_seq2	475	G	G	A	9/11	2/2
1604.	comp29732_c0_seq2	1143	C	A	C	17/20	20/26
1605.	comp29732_c0_seq2	1152	C	T	C	20/21	20/26
1606.	comp29732_c0_seq2	1155	C	T	C	20/21	22/28
1607.	comp29732_c0_seq2	1167	A	G	A	20/21	16/24
1608.	comp29732_c0_seq2	1176	A	G	A	20/21	16/22
1609.	comp29732_c0_seq2	1179	G	A	G	20/21	14/20
1610.	comp29732_c0_seq2	1191	A	G	A	20/21	12/18
1611.	comp29732_c0_seq2	1194	C	T	C	20/21	12/18
1612.	comp28401_c0_seq1	186	A	T	A	7/7	8/8
1613.	comp28401_c0_seq1	700	G	A	G	9/9	16/16
1614.	comp30727_c0_seq1	383	G	A	G	6/6	48/48
1615.	comp32419_c0_seq13	826	A	T	A	17/25	8/8
1616.	comp32419_c0_seq13	836	T	T	G	24/24	6/6
1617.	comp47419_c0_seq1	49	T	G	T	5/7	8/8
1618.	comp47419_c0_seq1	56	T	C	T	6/7	8/8
1619.	comp47419_c0_seq1	60	C	A	C	5/7	8/8
1620.	comp47419_c0_seq1	71	A	G	A	5/7	8/8
1621.	comp47419_c0_seq1	74	G	A	G	5/7	8/8
1622.	comp47419_c0_seq1	169	T	C	T	4/6	6/6
1623.	comp32430_c0_seq15	63	A	A	G	25/29	16/22
1624.	comp29784_c0_seq1	128	T	C	T	14/14	39/39
1625.	comp31260_c0_seq2	203	C	C	T	96/134	2/2
1626.	comp30183_c0_seq2	84	A	C	A	5/7	4/6
1627.	comp30183_c0_seq2	86	C	G	C	4/6	4/6
1628.	comp30183_c0_seq2	102	T	C	T	4/6	4/4
1629.	comp30183_c0_seq2	105	T	C	T	4/6	4/4
1630.	comp27975_c0_seq1	82	A	G	A	13/13	22/22
1631.	comp32179_c0_seq2	637	C	C	T	6/8	10/15
1632.	comp32179_c0_seq2	1219	C	C	T	29/39	28/28
1633.	comp5560_c0_seq1	119	A	G	A	7/10	18/26
1634.	comp33384_c0_seq2	1340	C	T	C	16/16	21/21
1635.	comp21759_c0_seq1	442	A	A	T	11/11	4/4
1636.	comp30805_c0_seq1	508	G	T	G	7/7	24/24
1637.	comp32570_c0_seq3	5	G	C	G	4/6	2/2
1638.	comp33702_c0_seq4	624	C	C	T	8/10	9/11
1639.	comp31222_c0_seq1	494	A	T	A	18/27	31/31
1640.	comp26501_c0_seq1	83	A	A	G	7/9	2/2
1641.	comp26501_c0_seq1	101	T	T	A	9/10	4/4
1642.	comp26501_c0_seq1	143	T	T	C	24/34	4/6
1643.	comp26501_c0_seq1	672	T	T	G	52/60	4/6
1644.	comp26501_c0_seq1	835	T	T	C	35/47	4/4
1645.	comp31030_c0_seq1	1295	A	A	G	10/15	8/10
1646.	comp31030_c0_seq1	1298	A	A	G	10/15	6/8

XXX

1647.	comp29962_c0_seq1	674	G	T	G	11/11	28/28
1648.	comp32664_c0_seq4	622	C	C	T	27/27	14/14
1649.	comp10346_c0_seq1	78	G	G	C	11/12	4/6
1650.	comp10346_c0_seq1	92	A	A	T	12/13	4/6
1651.	comp10346_c0_seq1	139	T	T	A	11/14	4/6
1652.	comp10346_c0_seq1	142	C	C	G	10/13	4/6
1653.	comp10346_c0_seq1	160	G	G	C	8/10	4/6
1654.	comp27954_c0_seq2	123	C	C	T	6/6	6/8
1655.	comp27954_c0_seq2	126	A	A	T	6/6	6/8
1656.	comp33506_c0_seq5	4	G	A	G	4/6	3/4
1657.	comp18732_c0_seq1	98	G	G	A	10/13	12/18
1658.	comp28410_c0_seq2	1227	A	T	A	10/10	7/7
1659.	comp33157_c0_seq1	200	A	G	A	5/7	15/21
1660.	comp33478_c0_seq10	140	T	A	T	6/7	10/14
1661.	comp33478_c0_seq10	141	G	T	G	6/7	10/14
1662.	comp26929_c0_seq1	9	C	T	C	5/6	5/6
1663.	comp178986_c0_seq1	12	G	G	C	6/7	2/2
1664.	comp28370_c0_seq1	131	C	C	T	12/17	2/2
1665.	comp32959_c0_seq2	1473	C	A	C	8/12	35/39
1666.	comp29779_c0_seq4	1430	T	T	A	18/18	36/36
1667.	comp10270_c0_seq1	111	A	T	A	9/12	22/22
1668.	comp32536_c0_seq2	5	C	C	A	10/10	8/10
1669.	comp29312_c0_seq1	146	A	C	A	6/6	6/6
1670.	comp29312_c0_seq1	499	A	A	G	7/7	4/6
1671.	comp31176_c0_seq5	2248	G	A	G	15/15	18/18
1672.	comp33727_c0_seq1	272	C	C	A	8/9	2/2
1673.	comp33727_c0_seq1	580	A	A	T	8/12	1/1
1674.	comp33727_c0_seq1	775	G	G	A	30/37	2/2
1675.	comp33727_c0_seq1	968	A	A	G	24/31	2/2
1676.	comp33727_c0_seq1	969	A	A	C	26/35	2/2
1677.	comp33727_c0_seq1	1360	C	C	T	26/31	2/2
1678.	comp33727_c0_seq1	1375	G	G	A	28/33	2/2
1679.	comp33727_c0_seq1	2003	C	C	A	31/45	2/2
1680.	comp33727_c0_seq1	2005	C	C	T	29/40	2/2
1681.	comp33727_c0_seq1	2008	G	G	A	32/42	2/2
1682.	comp24565_c0_seq1	117	T	T	G	6/6	4/4
1683.	comp18040_c0_seq1	296	T	T	G	6/6	8/12
1684.	comp31730_c0_seq3	1272	T	G	T	6/6	112/127
1685.	comp31730_c0_seq3	1963	G	T	G	7/9	90/90
1686.	comp31730_c0_seq3	1981	G	T	G	10/12	76/80
1687.	comp31730_c0_seq3	2008	T	C	T	8/10	86/93
1688.	comp31730_c0_seq3	2010	C	A	C	8/10	94/94
1689.	comp31730_c0_seq3	2035	T	C	T	5/6	73/78
1690.	comp31730_c0_seq3	2038	G	A	G	5/6	71/76
1691.	comp31730_c0_seq3	2041	A	G	A	5/6	73/80
1692.	comp136931_c0_seq1	205	T	T	G	15/17	2/2
1693.	comp19890_c0_seq1	485	T	A	T	7/7	10/10
1694.	comp19890_c0_seq1	490	G	T	G	8/8	10/10
1695.	comp19890_c0_seq1	492	T	G	T	7/7	8/8
1696.	comp19890_c0_seq1	493	A	C	A	8/8	8/8
1697.	comp19890_c0_seq1	525	C	C	T	10/10	6/8
1698.	comp19890_c0_seq1	528	T	T	C	10/10	4/6
1699.	comp19890_c0_seq1	529	G	G	A	10/10	4/6
1700.	comp19890_c0_seq1	531	C	C	A	11/11	4/6
1701.	comp19890_c0_seq1	534	C	C	T	11/11	6/8
1702.	comp19890_c0_seq1	537	C	C	T	11/11	4/6
1703.	comp19890_c0_seq1	546	G	G	T	9/9	4/6
1704.	comp19890_c0_seq1	549	T	T	C	8/8	4/6
1705.	comp9685_c0_seq1	118	A	A	C	72/77	24/24
1706.	comp27833_c0_seq2	512	G	C	G	5/6	13/17
1707.	comp27833_c0_seq2	515	T	C	T	5/6	15/19
1708.	comp27833_c0_seq2	518	T	C	T	5/6	15/19
1709.	comp27833_c0_seq2	754	T	T	C	12/12	8/8
1710.	comp33502_c0_seq2	2375	G	T	G	18/18	24/24
1711.	comp31788_c0_seq1	66	C	G	C	11/16	28/30
1712.	comp10424_c0_seq1	169	T	T	C	5/7	5/7
1713.	comp5633_c0_seq1	152	C	C	T	9/11	4/6
1714.	comp31872_c0_seq4	46	T	T	G	82/83	5/7
1715.	comp31872_c0_seq4	62	A	A	C	102/112	5/7
1716.	comp31872_c0_seq4	96	C	C	G	142/170	15/22
1717.	comp33608_c0_seq5	94	A	G	A	10/13	4/6
1718.	comp24449_c1_seq1	1157	T	T	C	18/18	8/8
1719.	comp24449_c1_seq1	1262	C	T	C	7/10	12/12
1720.	comp24449_c1_seq1	1271	G	A	G	8/12	18/18
1721.	comp33051_c0_seq12	336	T	C	T	6/6	20/20

1722.	comp32935_c0_seq9	641	T	T	C	12/12	16/16
1723.	comp27114_c0_seq1	968	C	T	C	14/15	6/6
1724.	comp23290_c0_seq1	903	G	A	G	11/11	1/1
1725.	comp32182_c0_seq1	1732	C	G	C	6/6	6/8
1726.	comp21878_c0_seq1	67	G	G	T	8/10	4/6
1727.	comp23983_c0_seq1	74	G	G	A	28/28	22/22
1728.	comp33157_c0_seq2	71	A	G	A	6/7	6/8
1729.	comp30713_c0_seq1	611	A	A	T	18/18	12/12
1730.	comp56851_c0_seq1	4	G	G	A	11/16	4/4
1731.	comp30508_c0_seq1	180	T	T	C	23/29	2/2
1732.	comp33068_c0_seq5	3	T	T	G	5/7	4/6
1733.	comp33727_c0_seq9	619	A	A	G	8/12	2/3
1734.	comp33727_c0_seq9	639	G	G	A	10/14	3/3
1735.	comp33727_c0_seq9	923	C	C	T	14/18	2/2
1736.	comp33727_c0_seq9	932	G	G	A	12/18	2/2
1737.	comp33727_c0_seq9	977	T	T	C	9/12	2/2
1738.	comp32831_c0_seq4	1944	C	T	C	5/7	32/48
1739.	comp28030_c0_seq1	268	T	A	T	9/9	58/58
1740.	comp32820_c0_seq2	3876	G	C	G	7/7	20/20
1741.	comp28791_c0_seq1	301	C	C	T	6/9	20/30
1742.	comp33501_c0_seq2	826	G	C	G	11/11	25/25
1743.	comp31810_c0_seq2	1528	T	C	T	5/6	8/8
1744.	comp33426_c0_seq1	475	C	C	A	15/21	30/40
1745.	comp33426_c0_seq1	479	G	G	A	16/22	30/38
1746.	comp33426_c0_seq1	491	G	G	A	19/25	32/44
1747.	comp33426_c0_seq1	509	G	G	C	17/24	34/42
1748.	comp33426_c0_seq1	510	C	C	A	17/24	33/43
1749.	comp33417_c1_seq14	117	T	T	A	19/22	2/2
1750.	comp33417_c1_seq14	127	A	A	C	24/35	4/4
1751.	comp33417_c1_seq14	255	C	C	T	61/72	4/4
1752.	comp33427_c0_seq26	175	G	G	A	6/7	8/10
1753.	comp28742_c0_seq5	1394	A	T	A	8/8	21/21
1754.	comp32308_c0_seq1	1750	C	T	C	18/26	56/72
1755.	comp26630_c0_seq1	307	C	G	C	29/37	8/10
1756.	comp41752_c0_seq1	482	C	C	T	20/20	28/28
1757.	comp33415_c0_seq20	91	A	C	A	7/10	2/2
1758.	comp12379_c0_seq1	142	A	A	G	5/7	4/4
1759.	comp33502_c0_seq1	2400	G	T	G	19/19	26/26
1760.	comp31733_c0_seq9	154	A	A	G	8/12	19/28
1761.	comp30246_c0_seq1	202	G	G	C	87/113	42/58
1762.	comp30246_c0_seq1	253	C	C	T	59/88	37/52
1763.	comp33055_c0_seq5	133	C	C	A	31/38	36/50
1764.	comp33604_c0_seq2	422	A	G	A	11/12	20/30
1765.	comp33604_c0_seq2	471	G	G	A	7/8	12/18
1766.	comp33604_c0_seq2	472	G	G	A	7/8	12/18
1767.	comp13158_c0_seq1	305	A	T	A	6/6	7/7
1768.	comp33555_c0_seq16	452	T	T	C	16/22	17/25
1769.	comp33555_c0_seq16	453	G	G	A	16/22	17/23
1770.	comp33555_c0_seq16	495	T	T	G	9/9	13/15
1771.	comp33555_c0_seq16	501	G	G	A	15/16	9/13
1772.	comp33555_c0_seq16	507	G	G	T	13/14	13/15
1773.	comp33555_c0_seq16	537	C	C	A	9/13	6/9
1774.	comp19930_c0_seq1	173	A	G	A	8/8	12/12
1775.	comp31232_c0_seq2	1153	G	A	G	7/8	10/14
1776.	comp31232_c0_seq2	1155	C	T	C	7/8	10/14
1777.	comp31232_c0_seq2	1160	G	C	G	11/12	10/14
1778.	comp31232_c0_seq2	1192	C	C	T	13/13	10/14
1779.	comp31232_c0_seq2	1195	A	A	G	13/13	10/14
1780.	comp31232_c0_seq2	1224	T	T	C	11/11	8/12
1781.	comp31232_c0_seq2	1225	G	G	T	11/11	8/12
1782.	comp32089_c0_seq4	284	A	A	G	18/27	2/2
1783.	comp17680_c0_seq1	297	A	A	G	48/48	4/6
1784.	comp17680_c0_seq1	336	G	G	A	20/26	4/6
1785.	comp25563_c0_seq1	460	C	C	T	32/45	38/53
1786.	comp25563_c0_seq1	466	A	A	T	34/46	34/51
1787.	comp31017_c0_seq4	1943	C	C	T	6/6	4/4
1788.	comp33485_c0_seq1	1217	T	T	C	8/9	6/9
1789.	comp33485_c0_seq1	1226	G	G	A	10/11	6/9
1790.	comp33485_c0_seq1	1232	A	A	T	12/13	6/9
1791.	comp33485_c0_seq1	1241	T	T	C	12/13	6/9
1792.	comp33485_c0_seq1	1242	G	G	A	12/13	6/9
1793.	comp33485_c0_seq1	1243	A	A	G	12/13	6/9
1794.	comp33485_c0_seq1	1246	G	G	A	13/14	6/9
1795.	comp12098_c0_seq1	112	A	A	G	4/6	4/6
1796.	comp32018_c0_seq1	1173	C	C	G	11/12	15/15

1797.	comp26767_c0_seq1	310	A	G	A	12/12	14/14
1798.	comp29886_c0_seq1	157	A	G	A	10/10	82/90
1799.	comp32820_c0_seq15	1785	G	C	G	7/7	21/21
1800.	comp27815_c0_seq1	3	G	G	C	8/10	2/2
1801.	comp23051_c0_seq1	357	G	T	G	12/16	6/8
1802.	comp23051_c0_seq1	385	C	T	C	9/13	6/8
1803.	comp23051_c0_seq1	386	C	C	A	9/13	4/6
1804.	comp23051_c0_seq1	395	T	C	T	6/6	4/6
1805.	comp23484_c0_seq1	387	G	A	G	6/6	48/48
1806.	comp33595_c0_seq1	218	T	T	C	8/9	8/10
1807.	comp33595_c0_seq1	255	A	G	A	5/7	10/14
1808.	comp33595_c0_seq1	259	T	C	T	5/7	12/14
1809.	comp33595_c0_seq1	262	C	T	C	5/7	14/16
1810.	comp18041_c0_seq1	49	A	A	T	18/19	4/4
1811.	comp33360_c0_seq1	399	C	G	C	8/8	24/25
1812.	comp15082_c0_seq1	186	T	A	T	6/6	2/2
1813.	comp145947_c0_seq1	94	T	T	C	11/12	14/18
1814.	comp32363_c0_seq4	275	T	A	T	5/6	13/13
1815.	comp32363_c0_seq4	308	G	A	G	6/8	18/20
1816.	comp32363_c0_seq4	311	G	A	G	6/8	18/18
1817.	comp32363_c0_seq4	313	A	G	A	6/8	18/18
1818.	comp32363_c0_seq4	317	C	T	C	6/8	18/18
1819.	comp32363_c0_seq4	323	T	C	T	6/8	16/16
1820.	comp32363_c0_seq4	326	G	A	G	6/8	16/16
1821.	comp32363_c0_seq4	347	T	C	T	6/8	14/14
1822.	comp242098_c0_seq1	2	G	T	G	6/6	2/2
1823.	comp20674_c0_seq1	198	G	A	G	6/9	10/14
1824.	comp28975_c0_seq2	196	T	T	C	17/20	4/4
1825.	comp33556_c0_seq25	905	A	A	C	9/10	4/6
1826.	comp33556_c0_seq25	923	A	A	G	10/13	4/6
1827.	comp33629_c0_seq1	182	G	C	G	4/6	6/9
1828.	comp33629_c0_seq1	195	C	G	C	7/9	4/4
1829.	comp33629_c0_seq1	200	G	A	G	7/9	2/2
1830.	comp33629_c0_seq1	209	A	T	A	7/9	2/2
1831.	comp33629_c0_seq1	210	T	C	T	7/9	2/2
1832.	comp54461_c0_seq1	4	G	G	T	11/13	14/20
1833.	comp32950_c0_seq1	2274	C	C	T	19/21	12/12
1834.	comp28596_c0_seq1	557	T	C	T	4/6	18/22
1835.	comp311354_c0_seq1	71	C	C	T	5/6	2/2
1836.	comp19823_c0_seq1	398	C	T	C	7/7	6/6
1837.	comp15502_c0_seq1	91	A	T	A	10/10	14/14
1838.	comp31260_c0_seq5	48	T	T	C	155/202	6/8
1839.	comp23767_c0_seq1	412	A	T	A	5/6	5/5
1840.	comp23767_c0_seq1	415	C	T	C	7/8	5/5
1841.	comp23767_c0_seq1	428	G	A	G	5/6	7/7
1842.	comp23767_c0_seq1	430	T	G	T	5/6	7/7
1843.	comp25594_c0_seq1	958	A	G	A	33/33	34/34
1844.	comp96888_c0_seq1	4	G	C	A	4/6	2/2
1845.	comp33724_c0_seq3	225	G	G	T	12/16	25/29
1846.	comp33724_c0_seq3	231	A	A	G	22/26	25/29
1847.	comp33724_c0_seq3	282	A	A	G	85/102	49/62
1848.	comp33555_c0_seq38	584	T	T	A	4/6	17/22
1849.	comp33555_c0_seq38	762	A	A	T	10/10	4/5
1850.	comp29834_c0_seq1	3084	G	G	C	4/6	2/2
1851.	comp33710_c0_seq1	5094	T	G	T	14/21	46/69
1852.	comp19936_c0_seq1	178	A	G	A	6/8	12/14
1853.	comp33719_c0_seq1	131	C	A	C	7/9	47/53
1854.	comp29432_c0_seq3	1038	C	C	G	5/7	2/3
1855.	comp29432_c0_seq3	1041	T	T	G	4/6	4/5
1856.	comp29432_c0_seq3	1050	G	G	T	4/6	4/5
1857.	comp29432_c0_seq3	1053	G	G	A	4/6	4/5
1858.	comp175643_c0_seq1	147	C	T	C	5/6	8/8
1859.	comp26490_c0_seq1	252	T	T	A	11/11	8/12
1860.	comp26490_c0_seq1	292	T	C	T	11/11	10/12
1861.	comp26490_c0_seq1	297	C	T	C	10/10	10/12
1862.	comp26490_c0_seq1	307	C	T	C	10/10	10/12
1863.	comp26490_c0_seq1	318	T	G	T	10/10	10/12
1864.	comp26490_c0_seq1	334	A	C	A	9/9	10/12
1865.	comp26490_c0_seq1	336	G	A	G	9/9	10/10
1866.	comp26490_c0_seq1	354	C	A	C	7/7	8/8
1867.	comp30833_c0_seq2	1198	T	C	T	8/8	12/12
1868.	comp32916_c0_seq2	2029	G	G	A	38/38	35/47
1869.	comp5125_c0_seq1	139	T	C	T	9/9	4/4
1870.	comp31231_c1_seq1	104	A	A	G	9/12	8/8
1871.	comp31231_c1_seq1	113	A	A	T	10/13	8/10

1872.	comp32751_c0_seq2	1056	T	T	G	10/13	4/6
1873.	comp32751_c0_seq2	1059	G	G	A	10/13	4/5
1874.	comp30280_c0_seq2	282	A	G	A	9/9	7/7
1875.	comp13801_c0_seq1	640	C	G	C	5/6	2/2
1876.	comp13801_c0_seq1	641	T	A	T	5/6	2/2
1877.	comp30664_c0_seq2	128	A	A	T	23/28	4/4
1878.	comp31925_c0_seq1	662	G	C	G	13/18	42/42
1879.	comp33496_c0_seq35	111	A	G	A	4/6	14/14
1880.	comp33427_c0_seq1	400	C	C	G	5/6	7/7
1881.	comp33555_c0_seq6	355	C	C	A	10/10	7/9
1882.	comp33555_c0_seq6	364	T	T	C	5/6	7/9
1883.	comp27555_c0_seq1	178	T	T	C	38/56	1/1
1884.	comp27555_c0_seq1	189	T	T	G	31/46	1/1
1885.	comp10025_c0_seq1	133	T	T	C	14/14	4/4
1886.	comp10025_c0_seq1	197	C	C	T	18/20	4/4
1887.	comp32283_c0_seq6	128	T	C	T	9/9	35/35
1888.	comp32283_c0_seq6	843	T	G	T	9/9	4/5
1889.	comp33636_c0_seq5	319	C	C	G	6/7	3/3
1890.	comp31423_c0_seq1	1693	G	A	G	10/10	18/18
1891.	comp31155_c0_seq1	32	G	C	G	5/7	8/12
1892.	comp21995_c0_seq1	175	G	G	A	25/33	2/2
1893.	comp21995_c0_seq1	202	A	A	G	21/31	2/2
1894.	comp29743_c0_seq1	62	C	C	T	22/28	10/10
1895.	comp29743_c0_seq1	79	A	A	C	28/39	10/12
1896.	comp29743_c0_seq1	82	A	A	C	35/42	10/12
1897.	comp29743_c0_seq1	107	T	T	C	47/58	6/8
1898.	comp33543_c0_seq7	74	T	T	C	10/14	14/14
1899.	comp29928_c0_seq1	268	T	T	A	8/9	10/14
1900.	comp29928_c0_seq1	275	A	A	G	8/9	10/14
1901.	comp33088_c0_seq13	533	A	A	G	10/10	13/13
1902.	comp32668_c0_seq1	1755	T	A	T	91/131	134/200
1903.	comp27776_c0_seq1	114	A	A	C	4/6	6/8
1904.	comp27776_c0_seq1	115	C	C	G	4/6	6/8
1905.	comp27776_c0_seq1	229	T	C	T	4/6	24/34
1906.	comp25162_c0_seq1	125	A	C	A	8/10	12/12
1907.	comp25162_c0_seq1	127	C	T	C	8/9	10/12
1908.	comp25162_c0_seq1	129	A	G	A	9/10	10/12
1909.	comp25162_c0_seq1	149	A	G	A	7/8	8/8
1910.	comp25162_c0_seq1	158	G	A	G	6/7	6/6
1911.	comp18596_c0_seq1	164	A	A	T	21/21	16/16
1912.	comp33396_c1_seq3	34	T	T	A	6/7	4/6
1913.	comp33396_c1_seq3	54	A	A	G	7/8	4/6
1914.	comp33396_c1_seq3	78	G	G	A	7/8	6/8
1915.	comp33396_c1_seq3	81	A	A	T	7/8	6/8
1916.	comp33396_c1_seq3	84	T	T	A	8/10	8/12
1917.	comp33396_c1_seq3	117	C	C	A	7/8	10/14
1918.	comp33396_c1_seq3	126	A	A	G	8/10	10/14
1919.	comp33396_c1_seq3	134	A	A	C	4/6	8/12
1920.	comp33396_c1_seq3	776	T	G	T	17/24	13/19
1921.	comp33396_c1_seq3	781	C	G	C	16/23	13/19
1922.	comp33396_c1_seq3	1519	G	G	A	12/17	12/16
1923.	comp33396_c1_seq3	1523	A	A	G	12/17	10/14
1924.	comp4895_c0_seq1	216	T	T	A	8/8	4/4
1925.	comp4895_c0_seq1	217	T	T	G	8/8	4/4
1926.	comp36365_c0_seq1	4	A	C	A	4/6	20/28
1927.	comp29372_c0_seq1	747	A	A	G	10/10	14/14
1928.	comp5527_c0_seq1	4	A	C	A	5/7	9/11
1929.	comp33721_c0_seq1	8320	G	G	T	13/19	21/28
1930.	comp33178_c0_seq4	1240	A	T	A	5/7	4/6
1931.	comp33178_c0_seq4	1294	T	C	T	6/9	6/6
1932.	comp20169_c0_seq1	94	T	C	T	10/10	10/10
1933.	comp3829_c0_seq1	344	T	C	T	6/8	6/8
1934.	comp31909_c0_seq3	1613	C	C	T	4/6	6/8
1935.	comp25236_c0_seq1	4	T	T	A	8/9	4/4
1936.	comp24142_c0_seq1	214	C	C	T	14/15	16/16
1937.	comp32878_c0_seq2	648	G	A	G	6/8	27/27
1938.	comp33327_c0_seq3	250	A	A	G	51/75	8/12
1939.	comp33327_c0_seq3	305	C	C	T	26/34	6/8
1940.	comp33327_c0_seq3	338	G	G	C	7/9	4/6
1941.	comp32783_c0_seq8	219	T	C	T	5/7	10/10
1942.	comp19103_c0_seq1	32	A	C	A	11/11	12/12
1943.	comp31217_c0_seq9	147	C	T	C	7/8	16/16
1944.	comp72726_c0_seq1	540	A	A	G	8/8	2/2
1945.	comp18249_c0_seq1	262	G	A	G	7/10	14/18
1946.	comp20830_c0_seq1	189	C	C	A	11/11	10/10

1947.	comp8034_c0_seq1	148	T	T	A	6/6	2/2
1948.	comp18193_c0_seq1	182	A	G	A	10/11	18/18
1949.	comp30538_c0_seq1	198	T	A	T	8/12	18/20
1950.	comp33719_c0_seq15	168	T	C	T	5/7	2/3
1951.	comp33719_c0_seq15	170	T	C	T	5/7	2/3
1952.	comp32018_c0_seq3	519	A	A	T	5/6	9/13
1953.	comp32018_c0_seq3	1152	C	C	G	11/12	6/6
1954.	comp21172_c0_seq1	95	G	C	G	6/6	10/14
1955.	comp30851_c0_seq1	784	A	A	T	7/7	8/12
1956.	comp30851_c0_seq1	789	A	A	G	7/7	8/12
1957.	comp30851_c0_seq1	792	T	T	A	7/7	8/12
1958.	comp33695_c0_seq5	779	C	C	T	27/37	2/2
1959.	comp33177_c0_seq4	4558	A	T	A	14/14	22/22
1960.	comp32797_c0_seq11	135	T	T	C	12/16	12/17
1961.	comp30204_c0_seq1	243	T	C	T	17/17	30/30
1962.	comp31734_c0_seq1	281	A	A	G	8/8	16/18
1963.	comp9432_c0_seq1	254	T	A	T	10/10	12/12
1964.	comp32935_c0_seq1	641	T	T	C	8/8	6/6
1965.	comp33687_c0_seq2	3061	T	T	G	9/13	14/20
1966.	comp160838_c0_seq1	25	G	G	C	10/11	1/1
1967.	comp32639_c0_seq1	2130	G	T	G	9/9	16/16
1968.	comp27729_c0_seq1	455	A	G	A	11/11	6/6
1969.	comp27729_c0_seq1	482	T	T	G	9/9	12/12
1970.	comp27729_c0_seq1	487	A	A	G	7/7	12/12
1971.	comp27729_c0_seq1	494	G	G	A	7/7	14/14
1972.	comp33530_c0_seq3	2914	A	T	A	7/10	7/9
1973.	comp29903_c0_seq1	367	G	A	G	4/6	12/12
1974.	comp29903_c0_seq1	446	C	C	T	9/9	6/6
1975.	comp32899_c0_seq4	480	T	T	G	6/8	8/10
1976.	comp32899_c0_seq4	492	A	A	T	8/10	8/12
1977.	comp31781_c0_seq1	103	A	A	C	7/7	10/10
1978.	comp25883_c0_seq2	4	T	A	C	4/6	2/2
1979.	comp26715_c0_seq1	4	A	A	T	8/12	4/4
1980.	comp30457_c0_seq1	227	G	G	C	22/30	23/23
1981.	comp31121_c0_seq1	1727	T	G	T	11/11	26/26
1982.	comp33308_c0_seq12	657	A	G	A	5/6	10/10
1983.	comp33308_c0_seq12	680	C	T	C	6/7	11/15
1984.	comp33308_c0_seq12	696	C	A	C	6/7	12/16
1985.	comp29315_c0_seq1	68	G	A	G	6/6	2/2
1986.	comp29315_c0_seq1	71	A	G	A	6/6	2/2
1987.	comp32889_c0_seq3	806	G	A	G	12/12	46/46
1988.	comp23367_c0_seq1	59	A	A	C	8/10	4/6
1989.	comp23367_c0_seq1	62	G	G	C	8/10	4/6
1990.	comp23367_c0_seq1	69	C	C	T	8/10	4/6
1991.	comp23367_c0_seq1	105	T	T	C	23/25	4/6
1992.	comp23367_c0_seq1	106	G	G	A	24/26	4/6
1993.	comp33546_c0_seq1	741	G	A	G	20/28	21/31
1994.	comp24202_c0_seq1	807	T	C	T	15/15	20/20
1995.	comp30881_c0_seq2	426	T	C	T	9/9	8/8
1996.	comp26597_c0_seq1	491	G	G	C	88/89	16/16
1997.	comp33062_c0_seq8	113	T	T	C	47/56	3/3
1998.	comp33062_c0_seq8	134	A	A	C	69/85	4/6
1999.	comp33062_c0_seq8	146	G	G	A	72/88	6/6
2000.	comp33062_c0_seq8	152	A	A	T	71/89	6/6
2001.	comp33062_c0_seq8	200	T	T	C	45/57	6/7
2002.	comp100794_c0_seq1	427	G	G	C	7/7	2/2
2003.	comp32312_c0_seq2	580	A	G	A	6/9	9/9
2004.	comp32312_c0_seq2	595	G	A	G	7/9	9/9
2005.	comp20019_c0_seq1	124	G	A	G	17/17	10/10
2006.	comp33560_c0_seq3	729	C	T	C	6/9	8/12
2007.	comp24077_c0_seq1	309	A	T	A	6/6	14/14
2008.	comp32332_c0_seq1	1596	A	T	A	13/15	14/14
2009.	comp231670_c0_seq1	314	G	T	G	7/7	2/2
2010.	comp23064_c0_seq1	99	T	A	T	4/6	30/32
2011.	comp23064_c0_seq1	102	A	G	A	4/6	28/32
2012.	comp23064_c0_seq1	132	G	A	G	4/6	32/38
2013.	comp30523_c1_seq1	892	T	T	C	6/8	12/16
2014.	comp30523_c1_seq1	895	T	T	C	7/10	12/16
2015.	comp33030_c0_seq1	136	A	G	A	4/6	8/8
2016.	comp33030_c0_seq1	258	A	T	A	15/18	4/6
2017.	comp33030_c0_seq1	516	C	T	C	8/12	6/8
2018.	comp33030_c0_seq1	517	A	G	A	8/9	6/8
2019.	comp33030_c0_seq1	541	C	T	C	7/8	4/6
2020.	comp21709_c0_seq1	173	G	A	G	4/6	20/22
2021.	comp21709_c0_seq1	188	T	A	T	11/13	20/21

2022.	comp21709_c0_seq1	194	G	A	G	8/10	16/17
2023.	comp21709_c0_seq1	211	G	A	G	8/10	16/16
2024.	comp21709_c0_seq1	214	G	A	G	8/10	14/14
2025.	comp21709_c0_seq1	218	C	G	C	9/11	12/12
2026.	comp21709_c0_seq1	238	A	G	A	9/13	10/10
2027.	comp21709_c0_seq1	249	C	T	C	10/14	8/8
2028.	comp21709_c0_seq1	250	C	T	C	10/14	8/8
2029.	comp21709_c0_seq1	251	C	T	C	10/14	8/8
2030.	comp32623_c0_seq1	2164	C	C	T	12/17	6/6
2031.	comp29086_c0_seq1	388	G	G	A	68/89	69/97
2032.	comp29086_c0_seq1	389	G	G	C	63/84	73/101
2033.	comp29086_c0_seq1	390	G	G	A	64/81	67/93
2034.	comp29086_c0_seq1	401	C	C	T	41/56	57/65
2035.	comp29086_c0_seq1	406	G	G	A	40/53	53/63
2036.	comp29086_c0_seq1	408	A	A	G	26/37	51/61
2037.	comp6787_c0_seq1	56	A	G	A	12/12	14/14
2038.	comp33415_c0_seq15	81	C	T	C	4/6	8/8
2039.	comp33415_c0_seq15	148	C	C	T	28/33	6/8
2040.	comp33645_c0_seq1	1192	C	C	T	27/34	18/25
2041.	comp33350_c0_seq1	4042	G	A	G	11/11	20/20
2042.	comp31991_c0_seq2	457	G	G	A	5/7	4/6
2043.	comp31991_c0_seq2	1262	T	T	A	4/6	5/5
2044.	comp26475_c0_seq1	208	G	A	G	7/9	14/18
2045.	comp31319_c0_seq1	65	C	C	T	7/8	4/6
2046.	comp23397_c0_seq1	177	T	A	T	19/28	4/4
2047.	comp22536_c0_seq1	4	G	G	C	8/8	2/2
2048.	comp29304_c0_seq1	16	T	T	A	6/6	2/2
2049.	comp32488_c0_seq2	1588	G	G	A	7/7	8/12
2050.	comp32488_c0_seq2	1590	A	A	T	8/8	8/12
2051.	comp33523_c0_seq1	344	C	A	C	4/6	4/6
2052.	comp33523_c0_seq1	448	G	G	C	5/6	2/2
2053.	comp13649_c0_seq1	96	C	C	T	6/6	4/4
2054.	comp13649_c0_seq1	107	T	T	A	6/6	4/4
2055.	comp33003_c0_seq1	438	G	A	G	18/18	46/60
2056.	comp31160_c0_seq1	1913	C	G	C	4/6	48/48
2057.	comp32434_c0_seq2	1225	G	C	G	17/17	18/18
2058.	comp282305_c0_seq1	187	C	C	G	6/6	4/4
2059.	comp33668_c0_seq4	185	T	C	T	7/7	28/30
2060.	comp33668_c0_seq4	199	T	C	T	8/8	28/30
2061.	comp33668_c0_seq4	202	G	A	G	8/8	28/30
2062.	comp33668_c0_seq4	207	T	C	T	8/8	22/24
2063.	comp33668_c0_seq4	227	C	T	C	9/9	18/20
2064.	comp9926_c0_seq1	41	C	C	T	7/8	4/6
2065.	comp172164_c0_seq1	76	C	A	C	10/10	10/10
2066.	comp32829_c0_seq2	479	G	C	G	8/8	24/24
2067.	comp28143_c0_seq1	576	T	C	T	11/11	14/14
2068.	comp33479_c0_seq1	4675	A	T	A	7/7	36/36
2069.	comp33555_c0_seq1	514	C	T	C	9/13	11/11
2070.	comp32707_c1_seq3	332	T	C	T	6/8	17/20
2071.	comp32195_c0_seq1	1682	C	C	T	4/6	18/24
2072.	comp21231_c0_seq1	125	C	C	A	10/10	6/6
2073.	comp9653_c0_seq1	148	G	G	A	14/14	3/3
2074.	comp31266_c0_seq2	502	C	C	G	10/11	6/8
2075.	comp27561_c0_seq1	1595	T	G	T	7/7	20/20
2076.	comp25285_c0_seq1	17	T	C	T	6/6	2/2
2077.	comp25285_c0_seq1	158	G	T	G	31/44	32/47
2078.	comp23433_c0_seq1	84	T	C	T	7/7	9/11
2079.	comp26648_c0_seq1	299	C	A	C	26/28	2/2
2080.	comp26016_c0_seq1	289	T	C	T	10/10	36/36
2081.	comp30761_c0_seq1	297	A	A	C	5/6	6/9
2082.	comp32157_c0_seq3	75	T	C	T	5/7	8/11
2083.	comp32157_c0_seq3	87	T	C	T	5/7	9/11
2084.	comp32583_c0_seq1	649	T	C	T	11/11	10/12
2085.	comp33718_c0_seq1	122	G	A	G	7/10	20/30
2086.	comp33718_c0_seq1	136	C	T	C	4/6	24/32
2087.	comp33718_c0_seq1	227	T	G	T	17/18	44/54
2088.	comp33718_c0_seq1	236	T	C	T	15/18	42/54
2089.	comp33718_c0_seq1	239	G	A	G	16/19	36/46
2090.	comp33718_c0_seq1	242	C	T	C	17/20	30/38
2091.	comp33718_c0_seq1	310	T	C	T	9/10	22/33
2092.	comp33718_c0_seq1	335	T	C	T	6/9	28/36
2093.	comp33718_c0_seq1	362	C	T	C	7/9	40/52
2094.	comp33718_c0_seq1	370	A	C	A	8/8	42/54
2095.	comp33718_c0_seq1	407	A	G	A	10/11	44/60
2096.	comp33718_c0_seq1	419	A	G	A	9/12	46/62

2097.	comp33718_c0_seq1	422	C	T	C	9/12	42/58
2098.	comp33718_c0_seq1	428	T	C	T	10/13	44/58
2099.	comp33718_c0_seq1	508	A	A	G	12/17	32/44
2100.	comp33718_c0_seq1	566	A	A	G	22/33	34/48
2101.	comp33718_c0_seq1	630	T	C	T	19/23	46/66
2102.	comp33327_c0_seq5	315	A	A	G	9/13	2/2
2103.	comp33327_c0_seq5	318	A	A	G	10/14	2/2
2104.	comp33327_c0_seq5	323	T	T	A	8/11	2/2
2105.	comp18975_c0_seq2	1	G	G	A	5/6	4/6
2106.	comp33719_c0_seq24	148	G	G	A	4/6	17/25
2107.	comp33530_c0_seq29	449	C	T	C	6/6	2/2
2108.	comp31493_c0_seq1	671	C	T	C	11/11	22/22
2109.	comp25093_c0_seq1	70	C	C	T	100/130	4/4
2110.	comp33142_c0_seq3	12	C	T	C	6/6	2/2
2111.	comp33142_c0_seq3	25	C	T	C	5/6	2/2
2112.	comp33142_c0_seq3	27	A	G	A	6/7	2/2
2113.	comp33142_c0_seq3	182	G	G	A	15/19	26/38
2114.	comp33142_c0_seq3	402	C	C	T	15/15	14/18
2115.	comp26139_c0_seq2	113	T	T	C	14/14	20/20
2116.	comp23630_c0_seq4	270	C	C	T	18/18	12/12
2117.	comp26352_c0_seq1	1778	C	T	C	9/9	16/16
2118.	comp33505_c0_seq3	261	C	C	T	14/14	16/16
2119.	comp29671_c0_seq1	83	C	T	C	4/6	24/24
2120.	comp29671_c0_seq1	84	C	T	C	4/6	24/24
2121.	comp33637_c0_seq4	136	C	C	T	5/6	9/13
2122.	comp33727_c0_seq6	124	G	G	A	34/47	2/2
2123.	comp33727_c0_seq6	445	C	C	T	47/59	2/2
2124.	comp33727_c0_seq6	517	T	T	C	43/57	4/4
2125.	comp33727_c0_seq6	986	C	C	T	14/19	4/6
2126.	comp33727_c0_seq6	997	G	G	A	19/25	4/6
2127.	comp33727_c0_seq6	1441	T	T	C	11/15	2/2
2128.	comp29736_c0_seq1	528	G	G	C	8/12	2/2
2129.	comp32702_c0_seq1	973	C	C	T	18/24	23/33
2130.	comp32702_c0_seq1	982	C	C	T	23/29	25/35
2131.	comp33711_c0_seq3	353	G	G	A	6/8	9/11
2132.	comp33711_c0_seq3	449	A	A	C	9/13	5/7
2133.	comp33711_c0_seq3	450	T	T	C	9/13	5/7
2134.	comp33711_c0_seq3	467	A	A	G	10/12	5/7
2135.	comp24647_c0_seq1	176	A	A	G	20/25	8/11
2136.	comp24647_c0_seq1	188	A	A	G	19/24	4/6
2137.	comp31620_c0_seq1	1774	T	T	C	11/12	6/8
2138.	comp32735_c0_seq5	240	C	T	C	16/22	14/18
2139.	comp32735_c0_seq5	252	C	G	C	15/18	10/14
2140.	comp32735_c0_seq5	259	C	T	C	11/13	8/10
2141.	comp228102_c0_seq1	77	C	T	C	15/21	4/6
2142.	comp31280_c0_seq1	3	A	T	A	6/9	6/8
2143.	comp33353_c1_seq1	9152	G	C	G	11/14	18/22
2144.	comp33626_c0_seq11	254	G	A	G	15/15	6/6
2145.	comp26052_c0_seq1	133	A	T	A	5/7	8/10
2146.	comp26052_c0_seq1	135	G	C	G	5/7	10/10
2147.	comp33140_c0_seq1	87	G	A	G	6/6	22/22
2148.	comp33140_c0_seq1	3663	G	G	A	11/11	18/18
2149.	comp33336_c0_seq1	88	A	A	T	10/10	8/10
2150.	comp33336_c0_seq1	90	A	A	G	10/10	8/10
2151.	comp33336_c0_seq1	102	C	C	A	11/11	6/8
2152.	comp33336_c0_seq1	113	C	C	T	11/12	6/8
2153.	comp33336_c0_seq1	958	C	C	T	20/20	24/34
2154.	comp32419_c0_seq3	573	T	C	T	31/40	11/13
2155.	comp32419_c0_seq3	581	C	T	C	31/40	9/11
2156.	comp32419_c0_seq3	585	C	T	C	30/40	13/15
2157.	comp32419_c0_seq3	588	A	T	A	29/38	13/15
2158.	comp32419_c0_seq3	598	G	A	G	28/37	15/17
2159.	comp32419_c0_seq3	603	G	A	G	29/38	15/17
2160.	comp32419_c0_seq3	615	G	T	G	29/37	12/14
2161.	comp32419_c0_seq3	620	C	T	C	26/33	12/14
2162.	comp32419_c0_seq3	1507	T	T	A	6/6	4/4
2163.	comp32419_c0_seq3	1528	C	C	T	6/6	4/4
2164.	comp32419_c0_seq3	1600	T	T	C	8/9	5/5
2165.	comp26236_c0_seq1	979	G	A	G	17/17	14/14
2166.	comp32899_c0_seq1	460	C	C	G	6/8	7/7
2167.	comp31180_c0_seq2	717	G	G	A	4/6	2/3
2168.	comp29103_c0_seq1	1006	G	G	T	6/6	4/4
2169.	comp30655_c0_seq1	140	T	C	T	8/8	16/16
2170.	comp30655_c0_seq1	141	G	C	G	8/8	16/16
2171.	comp29419_c0_seq3	70	C	C	T	6/6	8/12

2172.	comp30251_c0_seq2	537	T	G	T	4/6	14/16
2173.	comp26044_c0_seq1	165	T	T	C	16/20	12/16
2174.	comp26044_c0_seq1	176	C	C	T	15/19	18/20
2175.	comp26044_c0_seq1	177	A	A	G	15/19	18/20
2176.	comp26044_c0_seq1	183	A	A	G	15/21	18/20
2177.	comp26044_c0_seq1	191	G	G	T	19/21	18/26
2178.	comp26044_c0_seq1	193	A	A	T	18/20	18/26
2179.	comp26044_c0_seq1	231	G	A	G	18/23	28/34
2180.	comp26044_c0_seq1	237	T	G	T	18/23	26/34
2181.	comp26044_c0_seq1	255	T	G	T	13/17	24/30
2182.	comp26044_c0_seq1	270	T	A	T	11/13	16/20
2183.	comp26044_c0_seq1	272	G	A	G	11/13	16/20
2184.	comp26044_c0_seq1	292	T	A	T	5/7	10/12
2185.	comp26044_c0_seq1	296	C	T	C	5/6	8/10
2186.	comp17895_c0_seq1	130	C	C	T	48/48	34/38
2187.	comp33534_c0_seq1	1369	T	C	T	4/6	8/12
2188.	comp33534_c0_seq1	1373	A	G	A	4/6	8/12
2189.	comp33534_c0_seq1	1386	T	A	T	4/6	10/14
2190.	comp33534_c0_seq1	1427	C	T	C	4/6	16/22
2191.	comp26894_c0_seq1	361	C	T	C	12/12	10/10
2192.	comp32442_c0_seq2	2861	A	A	G	6/8	2/2
2193.	comp30963_c0_seq4	1177	C	T	C	4/6	5/7
2194.	comp23743_c0_seq1	72	G	A	G	6/6	4/4
2195.	comp27132_c0_seq1	577	C	A	C	38/42	10/14
2196.	comp27132_c0_seq1	578	G	A	G	37/41	10/14
2197.	comp27132_c0_seq1	670	A	A	C	4/6	4/4
2198.	comp31556_c0_seq1	464	C	C	T	5/7	10/12
2199.	comp31556_c0_seq1	467	A	A	G	4/6	10/12
2200.	comp31556_c0_seq1	542	A	G	A	6/9	9/9
2201.	comp20239_c0_seq1	126	T	T	G	25/25	8/8
2202.	comp33415_c0_seq12	539	T	T	C	7/8	14/16
2203.	comp33415_c0_seq12	638	A	A	C	8/11	12/16
2204.	comp33137_c0_seq3	55	A	G	A	8/12	2/2
2205.	comp33679_c0_seq39	123	C	C	T	5/7	2/3
2206.	comp33679_c0_seq39	124	C	C	T	5/7	2/3
2207.	comp10081_c0_seq1	100	A	C	A	8/12	6/9
2208.	comp32664_c0_seq1	733	T	T	G	4/6	8/12
2209.	comp32664_c0_seq1	734	A	A	C	4/6	8/12
2210.	comp32664_c0_seq1	741	A	A	T	4/6	8/12
2211.	comp32165_c0_seq1	2	T	G	T	4/6	4/4
2212.	comp74973_c0_seq1	782	G	G	A	41/41	8/8
2213.	comp72252_c0_seq1	4	C	A	C	4/6	12/16
2214.	comp33283_c0_seq7	66	T	T	A	4/6	4/4
2215.	comp33283_c0_seq7	68	T	T	G	5/7	4/4
2216.	comp33283_c0_seq7	73	C	C	G	8/10	4/6
2217.	comp143637_c0_seq1	145	T	A	T	4/6	10/12
2218.	comp32878_c0_seq1	704	G	A	G	6/8	31/31
2219.	comp33609_c0_seq4	991	G	A	G	41/58	32/34
2220.	comp39134_c0_seq1	840	G	C	G	5/6	2/2
2221.	comp26918_c0_seq1	403	C	C	A	22/22	10/12
2222.	comp11539_c0_seq1	128	C	G	C	14/16	8/10
2223.	comp19120_c0_seq1	1130	G	A	G	23/23	24/24
2224.	comp31415_c0_seq2	527	A	A	G	15/20	10/15
2225.	comp28880_c1_seq1	1352	G	A	G	18/18	20/20
2226.	comp32528_c0_seq1	2656	G	A	G	5/7	12/12
2227.	comp18246_c0_seq1	3	A	G	T	7/7	2/2
2228.	comp73422_c0_seq1	5	A	A	C	7/8	2/2
2229.	comp29664_c0_seq1	1095	G	T	G	12/15	22/22
2230.	comp31597_c0_seq4	1139	A	T	A	5/6	23/25
2231.	comp31597_c0_seq4	1148	T	C	T	5/7	23/25
2232.	comp31597_c0_seq4	1150	T	A	T	5/7	23/25
2233.	comp31597_c0_seq4	1151	T	C	T	5/7	23/25
2234.	comp22403_c0_seq1	81	G	G	A	30/30	14/14
2235.	comp33303_c0_seq2	839	G	T	G	5/7	15/19
2236.	comp33303_c0_seq2	840	T	A	T	5/7	13/17
2237.	comp37776_c0_seq1	1224	C	C	A	12/12	10/12
2238.	comp37776_c0_seq1	1225	A	A	T	12/12	10/12
2239.	comp124322_c0_seq1	91	G	T	G	7/7	6/6
2240.	comp23630_c0_seq1	373	C	C	T	15/15	4/4
2241.	comp32447_c0_seq2	204	T	C	T	7/8	4/4
2242.	comp33667_c0_seq1	3505	A	G	A	7/9	24/28
2243.	comp33667_c0_seq1	3523	C	T	C	8/10	30/34
2244.	comp25944_c0_seq5	611	G	G	T	7/10	10/12
2245.	comp25944_c0_seq5	626	G	G	C	8/10	8/10
2246.	comp25944_c0_seq5	635	C	C	A	11/13	2/2

2247.	comp27373_c0_seq1	3	T	T	G	6/9	2/2
2248.	comp26502_c0_seq1	521	G	T	G	18/25	16/24
2249.	comp33402_c0_seq12	642	A	T	A	6/6	8/8
2250.	comp33402_c0_seq12	646	T	G	T	6/6	8/8
2251.	comp33402_c0_seq12	652	G	A	G	6/6	8/8
2252.	comp30100_c0_seq6	15	T	T	C	6/9	4/6
2253.	comp31460_c0_seq1	1025	T	T	C	6/8	4/6
2254.	comp33177_c0_seq1	4736	A	T	A	13/13	36/36
2255.	comp31979_c0_seq1	778	A	A	G	26/32	2/2
2256.	comp31979_c0_seq1	793	G	G	A	22/27	2/2
2257.	comp31979_c0_seq1	904	T	T	G	22/29	2/2
2258.	comp31979_c0_seq1	906	C	C	T	21/27	2/2
2259.	comp31979_c0_seq1	919	G	G	A	28/36	2/2
2260.	comp31979_c0_seq1	931	C	C	T	31/39	2/2
2261.	comp10760_c0_seq1	634	T	A	T	10/10	14/14
2262.	comp33510_c0_seq5	380	G	G	A	14/17	2/2
2263.	comp25014_c0_seq1	195	G	G	A	4/6	8/12
2264.	comp24967_c0_seq1	414	T	A	T	4/6	22/22
2265.	comp31282_c0_seq1	477	T	T	C	9/10	4/6
2266.	comp31771_c0_seq1	1436	T	A	T	11/11	16/16
2267.	comp33596_c0_seq4	2229	T	G	T	4/6	11/11
2268.	comp33596_c0_seq4	2309	C	T	C	11/16	13/16
2269.	comp33596_c0_seq4	2321	G	C	G	10/15	13/18
2270.	comp33596_c0_seq4	2339	C	G	C	9/13	21/28
2271.	comp256052_c0_seq1	65	G	G	A	6/6	4/4
2272.	comp7317_c0_seq1	37	A	C	A	8/8	4/4
2273.	comp33264_c0_seq1	3	A	G	A	5/6	6/6
2274.	comp33154_c0_seq3	354	C	C	T	23/23	24/26
2275.	comp33154_c0_seq3	750	T	C	T	14/21	8/12
2276.	comp33510_c0_seq1	94	A	A	T	87/94	4/4
2277.	comp33510_c0_seq1	702	G	G	A	10/12	4/4
2278.	comp17207_c0_seq1	100	C	C	A	8/12	2/2
2279.	comp19763_c0_seq1	100	G	G	A	14/14	8/8
2280.	comp33521_c0_seq8	4898	T	C	T	12/16	18/26
2281.	comp33067_c0_seq1	274	A	C	A	10/15	6/8
2282.	comp33067_c0_seq1	276	C	T	C	10/15	8/10
2283.	comp33067_c0_seq1	280	T	C	T	11/16	8/10
2284.	comp28867_c0_seq4	1245	G	A	G	5/7	8/12
2285.	comp32570_c0_seq2	268	A	C	A	6/8	9/13
2286.	comp32570_c0_seq2	313	T	G	T	7/9	9/13
2287.	comp33719_c0_seq9	427	A	G	A	11/16	29/40
2288.	comp21247_c0_seq1	168	A	T	A	6/6	4/4
2289.	comp119327_c0_seq1	19	C	C	T	10/12	2/2
2290.	comp29937_c0_seq1	529	C	C	T	15/22	21/27
2291.	comp5157_c0_seq1	251	A	T	A	11/11	2/2
2292.	comp18807_c0_seq1	1203	G	C	G	7/7	20/20
2293.	comp18807_c0_seq1	1204	C	G	C	7/7	20/20
2294.	comp28438_c0_seq1	99	G	G	A	51/60	4/6
2295.	comp28438_c0_seq1	147	G	G	T	43/51	4/6
2296.	comp28828_c0_seq1	495	A	A	G	8/11	6/8
2297.	comp33496_c0_seq1	980	T	T	G	5/7	4/6
2298.	comp2604_c0_seq1	62	C	C	G	7/8	2/2
2299.	comp2604_c0_seq1	86	G	G	C	8/9	2/2
2300.	comp32661_c0_seq3	148	T	T	C	5/6	6/6
2301.	comp32661_c0_seq3	150	C	C	G	5/7	6/7
2302.	comp313288_c0_seq1	93	A	C	A	6/6	6/6
2303.	comp313288_c0_seq1	96	T	C	T	6/6	6/6
2304.	comp313288_c0_seq1	113	G	A	G	6/6	4/4
2305.	comp33510_c0_seq12	142	A	A	G	73/75	4/6
2306.	comp33510_c0_seq12	152	A	A	T	60/74	4/6
2307.	comp33510_c0_seq12	195	A	T	A	47/53	6/6
2308.	comp33510_c0_seq12	196	T	G	T	39/52	6/6
2309.	comp33510_c0_seq12	311	A	A	T	41/60	4/6
2310.	comp33510_c0_seq12	468	C	C	T	11/16	6/8
2311.	comp33695_c0_seq7	274	A	A	G	31/39	12/14
2312.	comp30604_c0_seq7	34	A	A	C	8/10	8/10
2313.	comp30604_c0_seq7	38	C	C	A	8/10	8/10
2314.	comp30604_c0_seq7	40	G	G	C	8/10	8/10
2315.	comp30604_c0_seq7	44	A	A	G	8/10	8/10
2316.	comp30604_c0_seq7	85	A	A	C	15/17	8/10
2317.	comp30604_c0_seq7	91	G	G	A	12/14	10/12
2318.	comp30604_c0_seq7	104	C	C	G	12/14	10/12
2319.	comp30604_c0_seq7	114	G	G	A	9/10	10/10
2320.	comp30604_c0_seq7	135	C	C	A	7/8	2/2
2321.	comp33261_c0_seq14	295	G	A	G	19/22	13/13

2322.	comp33261_c0_seq14	297	T	A	T	17/21	12/13
2323.	comp21556_c0_seq1	269	C	T	C	9/9	12/12
2324.	comp17897_c0_seq1	67	C	C	A	19/24	4/4
2325.	comp17897_c0_seq1	79	G	G	T	20/27	6/6
2326.	comp30345_c0_seq1	28	G	A	G	6/6	26/26
2327.	comp22482_c0_seq1	356	T	T	G	9/12	6/8
2328.	comp22482_c0_seq1	375	C	C	T	11/16	8/10
2329.	comp16580_c0_seq1	238	G	G	C	10/10	8/8
2330.	comp29279_c0_seq1	884	G	G	A	4/6	13/15
2331.	comp29279_c0_seq1	890	G	G	T	4/6	15/19
2332.	comp29279_c0_seq1	916	G	G	C	9/11	15/19
2333.	comp32342_c0_seq1	950	C	T	C	10/10	28/28
2334.	comp33415_c0_seq21	400	C	C	T	11/15	4/6
2335.	comp33415_c0_seq21	402	C	C	G	11/13	4/6
2336.	comp28580_c0_seq1	307	C	C	T	15/15	10/10
2337.	comp31849_c0_seq20	751	A	C	A	5/7	1/1
2338.	comp28020_c0_seq1	809	C	C	T	6/6	2/2
2339.	comp26396_c0_seq1	152	C	G	C	7/7	12/12
2340.	comp30480_c0_seq1	27	A	A	G	5/6	2/2
2341.	comp30480_c0_seq1	34	A	A	G	6/8	2/2
2342.	comp17653_c0_seq1	49	A	A	G	7/8	2/2
2343.	comp26427_c0_seq1	43	A	G	A	5/6	16/24
2344.	comp18774_c0_seq1	644	T	C	T	7/7	34/34
2345.	comp16077_c0_seq1	236	C	C	A	9/9	10/10
2346.	comp91513_c0_seq1	252	T	T	C	26/26	26/26
2347.	comp20051_c0_seq1	193	A	G	A	9/12	10/13
2348.	comp25135_c0_seq1	186	A	G	A	6/6	24/25
2349.	comp25135_c0_seq1	213	T	C	T	6/6	30/30
2350.	comp27595_c0_seq1	115	T	C	T	6/6	46/47
2351.	comp27595_c0_seq1	1403	A	T	A	6/6	12/12
2352.	comp33477_c0_seq2	2486	T	C	T	7/7	12/12
2353.	comp26794_c0_seq1	988	A	A	T	46/46	8/8
2354.	comp26794_c0_seq1	1426	C	T	C	16/16	16/16
2355.	comp31212_c0_seq1	91	C	T	C	8/8	27/27
2356.	comp26738_c0_seq1	1408	A	A	C	6/6	8/12
2357.	comp26738_c0_seq1	1419	C	C	A	6/6	8/12
2358.	comp20319_c0_seq1	33	A	A	G	11/15	2/2
2359.	comp33629_c0_seq2	724	A	A	G	11/14	13/15
2360.	comp10076_c0_seq1	1250	G	G	A	21/21	14/14
2361.	comp10327_c0_seq1	198	A	T	A	6/7	6/6
2362.	comp31544_c0_seq1	1274	A	C	A	8/8	8/8
2363.	comp32608_c0_seq2	849	T	T	C	10/10	10/12
2364.	comp32608_c0_seq2	865	T	T	C	10/10	10/14
2365.	comp32608_c0_seq2	870	T	T	A	9/9	10/14
2366.	comp33402_c0_seq10	717	G	A	G	6/6	6/6
2367.	comp33388_c0_seq5	208	T	T	C	63/95	52/73
2368.	comp31733_c0_seq5	424	C	T	C	20/24	8/12
2369.	comp30608_c0_seq2	131	T	G	T	16/19	10/10
2370.	comp33415_c0_seq10	487	G	T	G	4/6	2/2
2371.	comp33415_c0_seq10	658	A	A	C	6/9	11/15
2372.	comp33415_c0_seq10	684	T	T	G	18/23	10/15
2373.	comp33617_c0_seq1	1439	C	C	T	6/9	27/39
2374.	comp28604_c0_seq1	3	A	A	C	6/7	4/4
2375.	comp33573_c0_seq3	775	T	C	T	12/12	8/8
2376.	comp22118_c0_seq1	405	T	T	C	16/16	14/14
2377.	comp10943_c0_seq1	119	T	C	T	10/14	20/20
2378.	comp33659_c0_seq11	150	A	G	A	5/6	6/8
2379.	comp33519_c0_seq8	89	T	T	C	6/8	9/13
2380.	comp33519_c0_seq8	90	G	G	A	6/8	9/13
2381.	comp33519_c0_seq8	94	A	A	T	6/8	9/13
2382.	comp33519_c0_seq8	122	A	G	A	6/8	20/26
2383.	comp19645_c0_seq1	253	T	T	C	6/8	2/2
2384.	comp24237_c0_seq1	490	A	A	C	23/23	10/10
2385.	comp27050_c0_seq1	126	C	T	C	6/6	14/14
2386.	comp30993_c0_seq1	495	G	A	G	8/10	12/12
2387.	comp26612_c0_seq1	348	A	A	G	26/26	18/18
2388.	comp9137_c0_seq1	881	A	A	G	16/16	6/6
2389.	comp144582_c0_seq1	105	T	C	T	11/11	8/11
2390.	comp144582_c0_seq1	108	A	T	A	11/11	8/11
2391.	comp32395_c0_seq1	1016	C	T	C	9/10	18/18
2392.	comp32395_c0_seq1	1196	C	C	T	9/13	32/46
2393.	comp32395_c0_seq1	1214	T	T	C	13/16	28/37
2394.	comp32395_c0_seq1	1244	C	T	C	11/14	18/20
2395.	comp32395_c0_seq1	1286	A	G	A	9/12	16/22
2396.	comp33123_c0_seq3	202	G	G	T	10/12	4/6

2397.	comp28308_c0_seq2	190	A	G	A	7/10	11/14
2398.	comp28308_c0_seq2	198	T	A	T	7/9	10/13
2399.	comp28308_c0_seq2	202	A	T	A	7/10	11/14
2400.	comp23249_c0_seq1	287	G	G	C	5/7	10/10
2401.	comp27585_c0_seq1	255	G	A	G	7/7	13/13
2402.	comp32656_c0_seq6	3	G	G	T	6/8	4/6
2403.	comp30458_c0_seq1	395	A	G	A	7/8	12/16
2404.	comp33645_c0_seq7	975	T	C	T	4/6	16/16
2405.	comp30772_c0_seq1	101	G	G	A	24/24	24/24
2406.	comp104627_c0_seq1	52	T	T	A	6/8	4/4
2407.	comp104627_c0_seq1	54	T	T	G	6/8	4/4
2408.	comp104627_c0_seq1	55	T	T	A	6/8	4/4
2409.	comp104627_c0_seq1	57	A	A	G	6/8	4/4
2410.	comp29373_c0_seq1	199	T	G	T	6/6	4/4
2411.	comp33230_c0_seq1	134	A	C	A	12/12	16/16
2412.	comp33230_c0_seq1	1936	C	G	C	12/13	26/26
2413.	comp57942_c0_seq1	3	C	T	C	4/6	2/2
2414.	comp14647_c0_seq1	170	C	T	C	5/6	6/6
2415.	comp28966_c0_seq2	4	C	C	T	6/7	2/2
2416.	comp32539_c0_seq1	1560	A	G	A	14/14	21/21
2417.	comp30275_c0_seq5	148	A	A	G	79/93	6/8
2418.	comp30275_c0_seq5	161	T	T	C	58/63	6/8
2419.	comp30275_c0_seq5	170	G	G	A	35/50	6/6
2420.	comp30275_c0_seq5	175	C	T	C	40/48	6/6
2421.	comp2550_c0_seq1	108	A	A	G	12/15	2/2
2422.	comp31711_c0_seq1	264	C	C	T	80/106	14/18
2423.	comp30259_c0_seq2	60	G	G	C	6/7	2/2
2424.	comp29928_c0_seq3	101	T	A	T	7/7	28/29
2425.	comp28002_c0_seq1	276	T	C	T	8/12	24/28
2426.	comp28002_c0_seq1	282	T	A	T	9/12	32/36
2427.	comp28002_c0_seq1	286	A	T	A	8/11	40/42
2428.	comp28002_c0_seq1	293	A	T	A	8/12	46/46
2429.	comp7212_c0_seq1	247	A	G	A	4/6	20/20
2430.	comp28754_c0_seq1	16	A	A	T	7/8	3/3
2431.	comp25971_c0_seq1	155	T	C	T	4/6	8/10
2432.	comp33645_c0_seq18	973	T	T	C	6/9	11/15
2433.	comp32738_c0_seq1	502	C	T	C	7/9	13/19
2434.	comp25573_c0_seq1	133	G	A	G	6/6	6/6
2435.	comp33159_c0_seq1	2636	A	A	T	16/23	24/34
2436.	comp32416_c0_seq1	2515	C	T	C	5/7	11/13
2437.	comp33571_c0_seq4	209	T	T	A	15/15	22/22
2438.	comp33510_c0_seq2	53	A	A	G	86/106	4/4
2439.	comp33510_c0_seq2	68	T	T	C	87/117	4/4
2440.	comp33510_c0_seq2	81	A	A	G	105/126	4/4
2441.	comp33510_c0_seq2	83	A	A	G	100/122	4/4
2442.	comp33510_c0_seq2	101	A	A	G	127/142	4/4
2443.	comp33510_c0_seq2	209	G	G	A	108/148	2/3
2444.	comp33510_c0_seq2	730	A	A	T	7/10	2/2
2445.	comp33510_c0_seq2	731	T	T	A	7/10	2/2
2446.	comp33510_c0_seq2	744	A	A	C	6/8	2/2
2447.	comp33510_c0_seq2	745	T	T	G	6/7	2/2
2448.	comp33510_c0_seq2	746	G	G	C	6/8	2/2
2449.	comp33510_c0_seq2	1012	C	C	T	14/20	2/2
2450.	comp33497_c0_seq2	1598	T	C	T	17/25	28/40
2451.	comp32312_c0_seq1	965	A	G	A	6/9	3/3
2452.	comp32312_c0_seq1	966	C	A	C	6/9	3/3
2453.	comp33123_c0_seq1	907	C	C	A	11/13	18/20
2454.	comp31788_c0_seq2	154	C	C	T	69/98	2/2
2455.	comp66828_c0_seq1	81	G	A	G	33/33	26/26
2456.	comp3207_c0_seq1	103	T	T	A	9/9	4/4
2457.	comp3207_c0_seq1	104	T	T	A	10/10	4/4
2458.	comp3207_c0_seq1	105	G	G	A	10/10	4/4
2459.	comp3207_c0_seq1	109	C	C	A	6/6	6/6
2460.	comp33594_c0_seq3	1668	C	T	C	4/6	9/9
2461.	comp31029_c0_seq1	1	T	C	T	4/6	4/4
2462.	comp33476_c0_seq3	452	G	G	A	6/6	8/10
2463.	comp33476_c0_seq3	453	T	T	A	6/7	8/10
2464.	comp76757_c0_seq1	3	T	A	T	4/6	3/3
2465.	comp27427_c0_seq4	234	T	G	T	4/6	6/6
2466.	comp27380_c0_seq1	4	T	T	A	7/9	2/2
2467.	comp30273_c0_seq1	188	C	C	T	16/16	14/14
2468.	comp23781_c0_seq1	448	T	T	C	14/14	6/6
2469.	comp29458_c0_seq3	148	C	C	G	8/11	4/4
2470.	comp33656_c0_seq1	1319	A	G	A	6/9	25/30
2471.	comp31662_c0_seq2	619	C	G	C	10/10	11/11

2472.	comp33690_c0_seq1	210	G	T	G	5/7	15/17
2473.	comp33690_c0_seq1	221	T	A	T	6/9	12/16
2474.	comp27776_c0_seq2	25	G	G	A	9/12	6/6
2475.	comp27776_c0_seq2	32	A	A	G	9/13	8/8
2476.	comp27776_c0_seq2	38	C	C	T	12/16	8/8
2477.	comp27776_c0_seq2	54	T	T	A	13/15	2/2
2478.	comp27776_c0_seq2	64	A	A	G	13/19	10/10
2479.	comp27776_c0_seq2	198	C	T	C	4/6	6/6
2480.	comp27776_c0_seq2	203	A	C	A	4/6	4/4
2481.	comp23326_c0_seq1	19	A	A	T	20/22	4/4
2482.	comp29104_c0_seq1	1919	T	T	C	14/14	20/26
2483.	comp98477_c0_seq1	4	C	C	T	5/6	4/4
2484.	comp30681_c0_seq1	159	A	A	G	6/6	15/21
2485.	comp20081_c0_seq1	229	A	G	A	17/17	18/20
2486.	comp30325_c0_seq1	800	T	G	T	9/10	44/44
2487.	comp21853_c0_seq1	24	G	A	G	10/10	2/2
2488.	comp21853_c0_seq1	26	G	A	G	10/11	2/2
2489.	comp21853_c0_seq1	76	T	T	A	34/45	6/6
2490.	comp21853_c0_seq1	106	C	C	A	72/76	6/8
2491.	comp21853_c0_seq1	114	T	T	C	67/71	6/8
2492.	comp21853_c0_seq1	145	T	T	A	41/56	6/6
2493.	comp19100_c0_seq1	77	G	G	T	43/60	2/2
2494.	comp19100_c0_seq1	88	A	A	T	55/71	2/2
2495.	comp32179_c0_seq1	637	C	C	T	6/7	14/17
2496.	comp32179_c0_seq1	1308	C	C	T	34/41	34/34
2497.	comp31832_c0_seq1	536	C	T	C	6/6	10/12
2498.	comp12736_c0_seq1	198	A	G	A	7/7	12/12
2499.	comp16468_c0_seq1	56	C	C	T	4/6	2/2
2500.	comp32742_c0_seq4	380	A	A	T	16/16	4/4
2501.	comp4126_c0_seq1	136	C	C	G	7/7	4/4
2502.	comp280775_c0_seq1	124	T	T	C	6/9	2/2
2503.	comp32742_c0_seq1	380	A	A	T	25/26	4/4
2504.	comp32742_c0_seq1	454	C	C	T	23/23	4/4
2505.	comp32419_c0_seq15	930	T	T	A	12/12	3/3
2506.	comp32419_c0_seq15	951	C	C	T	6/6	3/3
2507.	comp33492_c1_seq3	5006	C	C	T	5/7	12/18
2508.	comp33492_c1_seq3	5009	A	A	C	5/7	10/15
2509.	comp33556_c0_seq39	176	A	G	A	6/8	2/2
2510.	comp33556_c0_seq39	190	G	A	G	4/6	4/4
2511.	comp33604_c0_seq1	818	C	T	C	7/10	24/24
2512.	comp27590_c0_seq1	229	G	A	G	19/19	18/18
2513.	comp31217_c0_seq8	84	C	T	C	8/9	18/22
2514.	comp33555_c0_seq4	477	C	T	C	10/14	9/13
2515.	comp31144_c0_seq1	152	G	G	A	12/12	4/4
2516.	comp31144_c0_seq1	155	C	C	T	12/12	4/5
2517.	comp31144_c0_seq1	156	A	A	G	12/12	5/5
2518.	comp31144_c0_seq1	515	A	G	A	31/46	12/12
2519.	comp31144_c0_seq1	811	T	T	C	22/22	16/16
2520.	comp32742_c0_seq9	108	A	A	T	39/48	2/2
2521.	comp27323_c0_seq2	221	A	G	A	52/74	6/8
2522.	comp28792_c0_seq1	445	T	A	T	6/7	30/42
2523.	comp28792_c0_seq1	492	G	C	G	9/11	20/30
2524.	comp28792_c0_seq1	493	G	A	G	9/11	20/30
2525.	comp28792_c0_seq1	504	G	A	G	9/10	22/26
2526.	comp28792_c0_seq1	521	C	T	C	9/10	16/18
2527.	comp28792_c0_seq1	531	T	A	T	9/12	25/27
2528.	comp29968_c0_seq4	824	C	T	C	6/6	13/13
2529.	comp13933_c0_seq1	169	A	G	A	5/7	12/12
2530.	comp13933_c0_seq1	184	C	G	C	4/6	10/10
2531.	comp29629_c0_seq3	80	A	A	G	4/6	3/3
2532.	comp33090_c0_seq3	2203	A	A	T	7/7	4/6
2533.	comp29630_c0_seq1	119	T	C	T	6/8	10/14
2534.	comp29630_c0_seq1	156	A	G	A	5/7	8/10
2535.	comp29630_c0_seq1	415	G	A	G	8/8	10/10
2536.	comp29630_c0_seq1	450	A	A	G	14/16	15/19
2537.	comp33702_c0_seq2	923	C	C	T	5/7	10/13
2538.	comp29492_c0_seq1	128	C	C	T	8/10	10/12
2539.	comp29492_c0_seq1	129	G	G	A	10/10	8/12
2540.	comp29492_c0_seq1	133	T	T	C	8/10	10/12
2541.	comp29492_c0_seq1	138	A	A	G	10/10	8/12
2542.	comp6206_c0_seq1	146	A	A	C	22/22	20/20
2543.	comp26087_c0_seq1	937	G	A	G	6/6	18/18
2544.	comp167355_c0_seq1	586	G	G	A	5/6	8/10
2545.	comp28829_c0_seq1	834	A	G	A	45/56	8/10
2546.	comp28829_c0_seq1	858	T	C	T	15/20	6/8

2547.	comp31711_c0_seq3	226	T	T	C	151/198	4/6
2548.	comp32934_c0_seq2	958	C	A	C	8/8	18/18
2549.	comp28921_c0_seq1	15	A	G	A	15/20	2/3
2550.	comp24642_c0_seq1	175	G	C	G	9/13	4/6
2551.	comp24642_c0_seq1	186	A	T	A	10/12	4/6
2552.	comp24642_c0_seq1	190	A	G	A	11/13	4/6
2553.	comp25049_c0_seq1	16	C	C	T	7/8	2/2
2554.	comp25049_c0_seq1	253	A	G	A	10/14	24/33
2555.	comp33114_c0_seq4	627	A	A	G	16/20	7/10
2556.	comp33332_c0_seq1	1857	G	G	A	6/6	12/12
2557.	comp33231_c0_seq1	2992	T	A	T	9/9	30/30
2558.	comp23923_c0_seq1	20	A	A	T	15/18	4/4
2559.	comp1044_c0_seq1	156	G	G	T	4/6	2/2
2560.	comp1044_c0_seq1	157	A	A	C	4/6	2/2
2561.	comp22548_c0_seq1	162	A	A	T	23/23	12/12
2562.	comp29675_c0_seq1	1598	G	A	G	22/23	33/39
2563.	comp29205_c0_seq1	113	T	C	T	17/19	16/16
2564.	comp29205_c0_seq1	515	C	T	C	9/10	28/28
2565.	comp35496_c0_seq1	987	T	T	C	7/7	4/4
2566.	comp33727_c0_seq5	1030	C	C	T	11/14	4/4
2567.	comp33687_c0_seq1	4106	A	G	A	7/7	21/21
2568.	comp33687_c0_seq1	4655	T	T	G	10/12	12/16
2569.	comp33687_c0_seq1	4668	T	T	G	15/16	12/16
2570.	comp28780_c0_seq2	123	A	A	G	240/344	22/26
2571.	comp32916_c0_seq1	385	C	C	A	36/36	25/32
2572.	comp22782_c0_seq1	135	T	T	A	16/20	4/6
2573.	comp32900_c0_seq1	910	G	G	A	6/9	26/37
2574.	comp25994_c0_seq1	146	T	C	T	4/6	8/10
2575.	comp29357_c0_seq2	319	T	C	T	15/21	29/29
2576.	comp31961_c0_seq4	190	C	C	T	26/38	49/65
2577.	comp31961_c0_seq4	197	A	A	G	28/39	41/57
2578.	comp31961_c0_seq4	198	C	C	T	30/39	39/57
2579.	comp12749_c0_seq1	49	A	T	A	6/6	10/10
2580.	comp33493_c0_seq21	523	C	C	T	11/16	10/13
2581.	comp33493_c0_seq21	562	C	T	C	7/8	7/8
2582.	comp33643_c0_seq2	705	C	A	C	4/6	10/12
2583.	comp33643_c0_seq2	1146	G	C	G	9/13	23/31
2584.	comp33643_c0_seq2	1149	A	G	A	9/13	22/31
2585.	comp33643_c0_seq2	1152	T	A	T	10/14	26/34
2586.	comp33643_c0_seq2	1164	G	A	G	8/10	28/34
2587.	comp31554_c0_seq1	3190	C	A	C	12/14	18/18
2588.	comp23236_c0_seq1	105	A	A	T	7/8	2/2
2589.	comp33094_c0_seq1	424	T	T	C	11/15	16/24
2590.	comp28272_c0_seq1	198	G	G	C	22/22	12/12
2591.	comp115548_c0_seq1	204	T	T	A	7/7	2/2
2592.	comp33566_c0_seq1	327	T	C	T	8/9	14/14
2593.	comp33566_c0_seq1	997	T	C	T	21/25	20/24
2594.	comp33566_c0_seq1	1844	T	T	C	30/41	16/22
2595.	comp26174_c0_seq1	351	G	G	T	24/34	2/2
2596.	comp23961_c0_seq1	662	A	A	G	16/16	4/4
2597.	comp30137_c0_seq2	301	T	T	C	8/9	4/5
2598.	comp30137_c0_seq2	306	A	A	T	8/9	4/5
2599.	comp33238_c0_seq12	120	T	T	C	10/15	6/6
2600.	comp33238_c0_seq12	143	G	G	C	11/14	2/2
2601.	comp131883_c0_seq1	575	T	A	T	10/13	6/6
2602.	comp33459_c0_seq1	2191	G	G	A	54/73	24/36
2603.	comp33459_c0_seq1	2201	T	T	A	61/89	28/42
2604.	comp33459_c0_seq1	2212	A	A	G	68/95	30/45
2605.	comp29341_c0_seq1	374	A	G	A	10/10	14/14
2606.	comp20171_c0_seq1	95	A	G	A	6/7	10/10
2607.	comp23400_c0_seq3	93	C	C	G	7/7	4/6
2608.	comp22273_c0_seq1	201	C	C	A	15/15	12/12
2609.	comp22273_c0_seq1	894	T	C	T	11/12	22/22
2610.	comp10841_c0_seq1	83	C	C	T	13/14	4/4
2611.	comp10841_c0_seq1	84	G	G	A	13/14	4/4
2612.	comp13738_c0_seq1	11	T	A	T	11/16	17/20
2613.	comp21259_c0_seq1	35	G	A	G	4/6	2/2
2614.	comp21259_c0_seq1	43	T	A	T	4/6	2/2
2615.	comp21259_c0_seq1	67	A	G	A	4/6	4/4
2616.	comp21259_c0_seq1	70	T	C	T	4/6	6/6
2617.	comp21259_c0_seq1	75	G	A	G	4/6	6/6
2618.	comp21259_c0_seq1	76	C	A	C	4/6	6/6
2619.	comp21259_c0_seq1	77	A	T	A	4/6	6/6
2620.	comp21259_c0_seq1	80	A	G	A	4/6	6/6
2621.	comp33516_c0_seq1	2036	C	T	C	17/17	22/22

2622.	comp33516_c0_seq1	2714	C	T	C	7/7	24/24
2623.	comp208876_c0_seq1	123	A	A	G	9/9	4/4
2624.	comp32815_c0_seq3	588	G	G	T	7/7	4/5
2625.	comp32815_c0_seq3	647	G	G	A	4/6	6/7
2626.	comp18977_c0_seq1	172	A	T	A	12/18	18/24
2627.	comp18977_c0_seq1	182	T	C	T	12/16	16/22
2628.	comp18977_c0_seq1	191	T	C	T	12/15	14/20
2629.	comp33142_c0_seq4	210	C	T	C	4/6	16/20
2630.	comp33142_c0_seq4	239	C	T	C	5/6	18/26
2631.	comp21432_c0_seq1	114	T	T	C	13/15	6/6
2632.	comp21432_c0_seq1	125	T	C	T	11/14	6/8
2633.	comp21432_c0_seq1	126	G	T	G	11/14	6/8
2634.	comp21432_c0_seq1	128	C	T	C	11/14	6/8
2635.	comp26859_c0_seq1	143	T	T	C	7/8	16/24
2636.	comp33727_c0_seq11	455	C	C	T	36/48	2/2
2637.	comp33727_c0_seq11	565	G	G	A	31/41	4/6
2638.	comp33727_c0_seq11	566	A	A	G	28/38	4/6
2639.	comp22432_c0_seq1	159	G	A	G	6/6	12/12
2640.	comp20333_c1_seq1	524	C	C	T	12/18	2/2
2641.	comp20333_c1_seq1	626	C	C	T	21/31	4/4
2642.	comp20333_c1_seq1	637	T	T	C	20/25	2/2
2643.	comp25412_c0_seq1	683	C	C	T	21/21	16/16
2644.	comp31556_c0_seq3	312	A	A	C	6/9	17/23
2645.	comp30490_c0_seq5	4	G	G	C	6/7	2/2
2646.	comp32587_c0_seq2	390	T	A	T	5/7	13/15
2647.	comp32587_c0_seq2	399	A	C	A	5/7	13/15
2648.	comp32587_c0_seq2	428	C	G	C	4/6	17/19
2649.	comp32587_c0_seq2	429	A	G	A	4/6	17/19
2650.	comp32587_c0_seq2	437	C	T	C	4/6	19/21
2651.	comp32587_c0_seq2	447	G	T	G	4/6	19/21
2652.	comp32587_c0_seq2	453	T	C	T	5/6	17/19
2653.	comp32587_c0_seq2	456	C	A	C	4/6	17/19
2654.	comp32587_c0_seq2	457	C	T	C	4/6	17/19
2655.	comp32587_c0_seq2	549	A	C	A	4/6	4/6
2656.	comp232408_c0_seq1	233	T	T	G	6/9	2/2
2657.	comp232408_c0_seq1	234	A	A	G	6/9	2/2
2658.	comp232408_c0_seq1	250	G	G	A	6/8	2/2
2659.	comp232408_c0_seq1	253	C	C	G	6/8	2/2
2660.	comp232408_c0_seq1	259	C	C	T	6/8	2/2
2661.	comp2969_c0_seq1	406	T	C	T	6/6	6/6
2662.	comp24996_c0_seq1	108	A	A	C	5/6	4/6
2663.	comp30690_c0_seq1	593	C	G	C	6/8	36/36
2664.	comp30690_c0_seq1	633	G	A	G	5/7	42/42
2665.	comp23485_c0_seq1	376	C	C	T	8/8	4/6
2666.	comp32543_c0_seq2	54	T	T	C	25/30	2/2
2667.	comp2251_c0_seq1	135	G	A	G	4/6	8/8
2668.	comp31834_c0_seq1	143	C	C	T	12/12	6/6
2669.	comp32813_c0_seq2	1712	A	T	A	8/8	12/12
2670.	comp5449_c0_seq1	46	G	A	G	4/6	1/1
2671.	comp2747_c0_seq1	320	G	C	G	20/20	32/32
2672.	comp27528_c0_seq1	478	T	T	C	15/20	21/21
2673.	comp33495_c0_seq1	2185	T	C	T	8/8	18/18
2674.	comp33495_c0_seq1	2227	C	T	C	6/6	14/14
2675.	comp37394_c0_seq1	299	T	C	T	6/6	14/14
2676.	comp32961_c0_seq1	243	G	A	G	4/6	24/34
2677.	comp32961_c0_seq1	654	C	T	C	4/6	20/24
2678.	comp32961_c0_seq1	681	A	T	A	4/6	24/29
2679.	comp29559_c0_seq1	794	A	G	A	5/7	8/8
2680.	comp29495_c0_seq1	193	C	T	C	4/6	6/6
2681.	comp72180_c0_seq1	251	T	C	T	7/7	10/10
2682.	comp33521_c0_seq7	4921	T	C	T	8/11	29/39
2683.	comp33521_c0_seq7	4926	T	C	T	9/11	37/43
2684.	comp33521_c0_seq7	4927	T	C	T	8/11	33/41
2685.	comp33521_c0_seq7	4975	T	C	T	6/8	10/13
2686.	comp25748_c0_seq1	191	G	C	G	6/9	26/26
2687.	comp16686_c0_seq1	28	C	C	T	8/8	4/4
2688.	comp10256_c0_seq1	6	A	G	A	8/12	6/9
2689.	comp30370_c1_seq3	150	A	A	T	190/194	8/10
2690.	comp30370_c1_seq3	204	T	T	C	344/345	4/6
2691.	comp119390_c0_seq1	267	G	G	T	5/7	2/2
2692.	comp119390_c0_seq1	285	A	A	T	7/9	2/2
2693.	comp119390_c0_seq1	294	C	C	T	6/8	2/2
2694.	comp25768_c0_seq2	270	C	C	T	30/30	12/12
2695.	comp20165_c0_seq1	494	C	T	C	6/6	20/20
2696.	comp26523_c0_seq1	417	T	C	T	8/8	12/12

2697.	comp28160_c0_seq1	187	C	T	C	18/22	38/38
2698.	comp26711_c0_seq1	171	G	T	G	9/9	16/16
2699.	comp28524_c0_seq1	896	G	G	A	18/19	12/18
2700.	comp21868_c0_seq1	217	A	A	G	6/7	6/6
2701.	comp21868_c0_seq1	220	C	C	T	6/7	6/6
2702.	comp29284_c0_seq1	350	G	G	A	37/38	20/20
2703.	comp33308_c0_seq11	593	C	C	T	6/6	16/20
2704.	comp33308_c0_seq11	595	G	G	A	6/6	16/20
2705.	comp33607_c0_seq1	2150	G	G	T	10/15	32/44
2706.	comp33607_c0_seq1	2369	G	G	A	6/8	36/54
2707.	comp31089_c0_seq2	465	T	T	G	29/29	12/12
2708.	comp29530_c0_seq1	646	A	A	T	9/10	6/7
2709.	comp29530_c0_seq1	668	A	A	G	4/6	1/1
2710.	comp169621_c0_seq1	234	A	G	A	5/7	6/6
2711.	comp20312_c0_seq1	369	T	T	C	6/8	2/2
2712.	comp27541_c0_seq1	639	A	G	A	17/17	24/26
2713.	comp21780_c0_seq1	88	C	C	T	13/16	4/4
2714.	comp32300_c0_seq2	181	T	C	T	8/8	10/10
2715.	comp27224_c0_seq4	20	A	A	C	7/7	4/6
2716.	comp19365_c0_seq1	431	C	G	C	6/6	8/10
2717.	comp19365_c0_seq1	537	C	T	C	8/8	4/6
2718.	comp32151_c0_seq2	778	A	T	A	7/7	15/15
2719.	comp31260_c0_seq3	87	C	C	T	92/119	6/9
2720.	comp31260_c0_seq3	173	C	C	T	79/116	2/2
2721.	comp31260_c0_seq3	218	T	T	C	31/43	2/2
2722.	comp31260_c0_seq3	222	T	T	G	31/43	2/2
2723.	comp68372_c0_seq1	167	C	C	G	8/8	6/9
2724.	comp19803_c0_seq1	200	A	A	G	9/9	12/18
2725.	comp33371_c0_seq7	339	T	T	G	9/9	6/7
2726.	comp11903_c0_seq1	103	C	T	C	6/8	6/6
2727.	comp24838_c0_seq1	606	G	T	G	9/9	26/26
2728.	comp28902_c0_seq1	555	G	A	G	24/24	30/30
2729.	comp18640_c0_seq1	1071	G	A	G	4/6	10/10
2730.	comp31730_c0_seq6	772	C	A	C	7/9	98/110
2731.	comp31730_c0_seq6	773	C	T	C	7/9	98/110
2732.	comp31730_c0_seq6	788	C	A	C	9/11	108/122
2733.	comp31730_c0_seq6	790	C	T	C	9/10	108/122
2734.	comp15606_c0_seq1	281	T	T	G	10/10	4/4
2735.	comp32276_c0_seq4	294	G	T	G	6/6	40/46
2736.	comp26091_c0_seq1	277	A	G	A	6/9	6/8
2737.	comp26091_c0_seq1	315	T	T	C	6/9	4/6
2738.	comp26091_c0_seq1	316	G	G	A	6/9	6/8
2739.	comp26091_c0_seq1	325	A	A	T	6/8	4/6
2740.	comp33691_c0_seq11	4	G	G	A	6/6	2/3
2741.	comp32307_c0_seq3	2404	C	A	C	19/20	38/38
2742.	comp8067_c0_seq1	194	C	T	C	11/11	6/6
2743.	comp33629_c0_seq11	313	A	A	G	7/9	10/14
2744.	comp31965_c0_seq2	149	G	C	G	11/15	6/8
2745.	comp31965_c0_seq2	169	C	A	C	11/16	8/10
2746.	comp5782_c0_seq1	4	G	A	G	4/6	2/2
2747.	comp24294_c0_seq1	22	T	T	G	8/11	2/2
2748.	comp24294_c0_seq1	112	C	C	T	13/16	2/2
2749.	comp19893_c0_seq1	85	G	G	A	10/13	11/15
2750.	comp19893_c0_seq1	88	T	T	G	12/15	11/15
2751.	comp20578_c0_seq1	103	A	A	G	6/6	4/4
2752.	comp20578_c0_seq1	106	T	T	A	6/6	4/4
2753.	comp20578_c0_seq1	107	G	G	C	6/6	4/4
2754.	comp28446_c0_seq1	113	G	G	A	8/8	4/6
2755.	comp28446_c0_seq1	132	C	C	G	7/7	4/6
2756.	comp31826_c0_seq15	103	C	C	G	10/14	10/13
2757.	comp24857_c0_seq1	99	T	C	T	12/12	8/8
2758.	comp24857_c0_seq1	116	A	C	A	17/17	4/6
2759.	comp17645_c0_seq1	391	T	T	C	8/12	3/3
2760.	comp17645_c0_seq1	396	T	T	C	8/12	3/3
2761.	comp229363_c0_seq1	279	T	C	T	4/6	6/6
2762.	comp33719_c0_seq14	89	T	T	C	7/8	2/2
2763.	comp33719_c0_seq14	93	T	T	C	7/8	2/2
2764.	comp33719_c0_seq14	95	C	C	T	7/8	2/2
2765.	comp33719_c0_seq14	98	T	T	C	7/8	2/2
2766.	comp33719_c0_seq14	99	T	T	G	6/7	2/2
2767.	comp33372_c0_seq8	141	C	G	C	6/7	22/32
2768.	comp32040_c0_seq1	229	T	T	G	9/12	14/18
2769.	comp32040_c0_seq1	230	A	A	T	9/12	16/20
2770.	comp32040_c0_seq1	231	A	A	T	10/13	16/20
2771.	comp32040_c0_seq1	238	A	A	G	11/14	16/20

2772.	comp32040_c0_seq1	247	C	C	G	12/15	14/20
2773.	comp32278_c0_seq1	256	T	A	T	8/8	22/22
2774.	comp24667_c0_seq2	3	A	C	A	4/6	2/2
2775.	comp24667_c0_seq2	32	T	T	C	8/8	4/4
2776.	comp24667_c0_seq2	94	C	C	T	19/19	8/8
2777.	comp32238_c0_seq2	1176	A	A	G	19/27	12/18
2778.	comp32238_c0_seq2	1182	C	C	A	21/29	12/18
2779.	comp32238_c0_seq2	1362	C	T	C	9/13	10/14
2780.	comp32238_c0_seq2	1380	T	C	T	13/19	9/13
2781.	comp28660_c0_seq1	258	C	C	G	8/8	8/8
2782.	comp32089_c0_seq6	96	G	G	A	215/226	4/5
2783.	comp32066_c0_seq2	812	G	G	A	91/119	41/41
2784.	comp32287_c0_seq1	2338	T	C	T	31/39	65/98
2785.	comp32287_c0_seq1	2341	T	C	T	31/39	66/99
2786.	comp31880_c0_seq4	479	C	T	C	9/9	8/8
2787.	comp33177_c0_seq5	4444	A	T	A	11/11	33/33
2788.	comp27218_c0_seq2	171	G	G	A	5/6	2/2
2789.	comp9207_c0_seq1	320	A	A	G	15/15	4/6
2790.	comp33371_c0_seq5	896	C	G	C	6/8	14/15
2791.	comp30898_c0_seq1	1541	G	A	G	4/6	32/32
2792.	comp30214_c3_seq1	70	A	T	A	15/22	2/2
2793.	comp30214_c3_seq1	164	A	A	G	31/42	2/2
2794.	comp2309_c0_seq1	17	C	T	C	27/28	5/7
2795.	comp18016_c0_seq1	108	G	G	A	30/40	8/10
2796.	comp18016_c0_seq1	110	T	T	G	30/40	8/10
2797.	comp18016_c0_seq1	136	T	T	C	22/29	12/14
2798.	comp33510_c0_seq20	4	A	A	C	4/6	2/2
2799.	comp32906_c0_seq3	983	C	C	T	5/7	8/12
2800.	comp33636_c0_seq18	551	C	C	G	6/7	6/8
2801.	comp32737_c0_seq1	949	G	A	G	18/27	38/38
2802.	comp33495_c0_seq12	265	A	A	G	11/11	5/5
2803.	comp32419_c0_seq19	3	A	T	A	6/8	2/2
2804.	comp32288_c0_seq1	2127	T	C	T	6/8	16/16
2805.	comp20925_c0_seq1	116	G	G	T	6/6	8/10
2806.	comp27870_c0_seq1	85	T	T	C	22/22	6/6
2807.	comp24543_c0_seq1	90	G	G	A	15/19	2/2
2808.	comp33497_c0_seq1	936	T	C	T	8/11	28/30
2809.	comp33497_c0_seq1	937	G	A	G	8/12	26/28
2810.	comp33497_c0_seq1	941	A	G	A	10/14	24/26
2811.	comp33723_c0_seq8	480	G	G	A	49/61	36/48
2812.	comp32371_c0_seq1	5	T	C	T	4/6	7/9
2813.	comp26735_c0_seq1	195	T	C	T	7/10	26/26
2814.	comp30183_c0_seq1	44	G	A	G	11/12	2/2
2815.	comp30183_c0_seq1	45	G	T	G	11/12	2/2
2816.	comp30183_c0_seq1	49	C	G	C	14/15	2/2
2817.	comp33719_c0_seq19	41	T	C	T	6/7	16/22
2818.	comp32062_c0_seq5	107	G	A	G	7/8	4/4
2819.	comp32062_c0_seq5	113	C	T	C	5/6	4/4
2820.	comp29341_c0_seq3	199	A	A	C	13/13	5/5
2821.	comp32168_c0_seq1	878	C	T	C	6/7	10/14
2822.	comp32168_c0_seq1	995	A	T	A	5/6	16/22
2823.	comp32168_c0_seq1	1000	T	C	T	5/6	20/26
2824.	comp96101_c0_seq1	6	T	T	A	9/9	2/2
2825.	comp6242_c0_seq1	627	G	T	G	16/16	22/22
2826.	comp30339_c0_seq7	415	A	G	A	5/7	2/2
2827.	comp30339_c0_seq7	429	T	C	T	5/7	2/2
2828.	comp30370_c1_seq1	593	G	G	A	288/292	12/18
2829.	comp21141_c0_seq1	172	G	G	T	9/9	6/6
2830.	comp12002_c0_seq1	183	A	A	G	6/7	2/2
2831.	comp12002_c0_seq1	195	A	A	C	5/6	2/2
2832.	comp31909_c0_seq2	485	T	T	C	42/54	33/42
2833.	comp31909_c0_seq2	488	T	T	C	39/52	33/41
2834.	comp31909_c0_seq2	497	G	G	T	30/45	35/39
2835.	comp8492_c0_seq1	32	C	A	C	5/7	12/14
2836.	comp8492_c0_seq1	102	T	T	A	8/12	14/14
2837.	comp8492_c0_seq1	105	A	A	G	8/11	14/14
2838.	comp8492_c0_seq1	115	T	T	C	6/9	8/10
2839.	comp28025_c0_seq2	456	T	T	A	9/10	4/6
2840.	comp28025_c0_seq2	462	G	G	A	8/9	4/6
2841.	comp28025_c0_seq2	475	G	G	T	7/10	8/10
2842.	comp28025_c0_seq2	477	T	T	G	6/9	8/10
2843.	comp28025_c0_seq2	480	G	G	T	6/9	8/10
2844.	comp28025_c0_seq2	653	C	G	C	6/8	12/18
2845.	comp28025_c0_seq2	654	T	C	T	6/8	18/18
2846.	comp33694_c0_seq1	2064	C	T	C	19/28	132/146

2847.	comp32448_c0_seq1	695	C	T	C	18/18	34/34
2848.	comp67616_c0_seq1	6	A	A	C	6/6	2/2
2849.	comp33332_c0_seq9	1325	G	G	A	15/15	10/10
2850.	comp126265_c0_seq1	641	G	G	A	21/21	12/12
2851.	comp33496_c0_seq21	205	C	C	T	6/8	6/8
2852.	comp31053_c0_seq1	2	G	C	G	22/31	12/12
2853.	comp29546_c0_seq2	272	C	G	C	9/9	18/18
2854.	comp33530_c0_seq12	651	G	A	G	6/6	11/12
2855.	comp28345_c0_seq1	640	G	G	T	10/11	15/21
2856.	comp28345_c0_seq1	658	T	T	C	6/7	11/15
2857.	comp28345_c0_seq1	886	A	A	C	23/31	12/18
2858.	comp20018_c0_seq1	134	T	T	C	11/14	6/8
2859.	comp20018_c0_seq1	137	C	C	T	10/13	6/8
2860.	comp20018_c0_seq1	139	A	A	G	10/13	6/8
2861.	comp20018_c0_seq1	146	A	A	G	9/12	6/8
2862.	comp20018_c0_seq1	152	T	T	C	7/10	6/8
2863.	comp20018_c0_seq1	155	T	T	C	6/9	6/8
2864.	comp33703_c0_seq16	184	C	T	C	5/7	5/7
2865.	comp26575_c0_seq1	76	G	G	A	8/8	4/4
2866.	comp78190_c0_seq1	72	G	G	A	10/10	2/2
2867.	comp30206_c0_seq4	944	C	C	T	14/16	6/9
2868.	comp30206_c0_seq4	1478	C	C	T	8/12	16/22
2869.	comp32428_c0_seq1	401	C	A	C	6/9	34/43
2870.	comp32428_c0_seq1	404	A	G	A	6/9	32/41
2871.	comp30407_c0_seq1	695	A	T	A	6/7	11/15
2872.	comp30407_c0_seq1	704	A	T	A	7/9	11/15
2873.	comp30407_c0_seq1	706	A	G	A	7/9	11/15
2874.	comp30407_c0_seq1	710	C	A	C	5/7	11/15
2875.	comp30407_c0_seq1	731	C	A	C	5/7	11/13
2876.	comp30407_c0_seq1	734	T	C	T	5/7	13/15
2877.	comp33521_c0_seq17	2336	T	C	T	4/6	18/20
2878.	comp33521_c0_seq17	2369	T	C	T	5/6	16/18
2879.	comp17260_c0_seq1	159	A	A	G	5/6	2/2
2880.	comp17260_c0_seq1	219	C	C	G	8/9	2/2
2881.	comp27899_c0_seq1	229	G	G	A	34/38	6/6
2882.	comp27899_c0_seq1	247	T	T	C	33/36	4/6
2883.	comp27899_c0_seq1	270	A	A	G	26/29	4/6
2884.	comp27899_c0_seq1	591	G	G	A	13/13	6/8
2885.	comp27899_c0_seq1	861	T	T	A	21/21	8/12
2886.	comp27899_c0_seq1	864	T	T	A	21/21	8/12
2887.	comp27899_c0_seq1	884	A	A	G	17/17	4/6
2888.	comp31238_c0_seq3	582	T	T	A	481/482	6/6
2889.	comp5282_c0_seq1	28	G	G	A	14/17	38/38
2890.	comp23310_c0_seq1	112	A	A	G	112/157	4/4
2891.	comp23310_c0_seq1	125	T	T	C	143/198	4/4
2892.	comp32220_c0_seq2	104	A	A	G	94/104	4/4
2893.	comp32220_c0_seq2	114	G	G	A	125/132	6/6
2894.	comp32220_c0_seq2	150	A	A	G	141/158	8/12
2895.	comp32220_c0_seq2	203	A	A	T	102/103	8/12
2896.	comp32220_c0_seq2	207	C	C	G	90/93	8/12
2897.	comp20348_c0_seq1	509	T	T	C	13/13	6/6
2898.	comp33466_c0_seq10	269	C	C	T	19/19	18/18
2899.	comp31439_c0_seq1	163	C	G	C	11/16	27/28
2900.	comp31439_c0_seq1	1520	G	A	G	9/9	9/9
2901.	comp24143_c0_seq1	151	A	A	G	67/84	8/12
2902.	comp33415_c0_seq17	604	T	T	G	19/24	14/19
2903.	comp33415_c0_seq17	682	T	T	A	16/18	8/11
2904.	comp357226_c0_seq1	154	A	T	A	6/8	2/2
2905.	comp357226_c0_seq1	157	G	T	G	6/9	2/2
2906.	comp32089_c0_seq3	290	T	T	C	85/86	4/4
2907.	comp32255_c0_seq6	250	A	T	A	15/15	30/30
2908.	comp32255_c0_seq6	333	A	G	A	25/25	40/40
2909.	comp21456_c0_seq3	115	C	T	C	11/11	14/14
2910.	comp3656_c0_seq1	237	T	C	T	6/8	4/4
2911.	comp32531_c0_seq3	528	G	G	A	13/13	6/6
2912.	comp23001_c0_seq1	125	A	T	A	4/6	4/6
2913.	comp6850_c0_seq1	1	T	G	T	5/6	2/2
2914.	comp6850_c0_seq1	2	C	T	C	5/6	2/2
2915.	comp6850_c0_seq1	3	A	C	A	5/6	2/2
2916.	comp28775_c0_seq1	320	C	C	T	9/11	8/11
2917.	comp33633_c0_seq1	1608	A	T	A	6/6	6/8
2918.	comp144416_c0_seq1	126	G	A	G	5/6	6/8
2919.	comp32932_c0_seq1	1761	G	G	T	16/16	16/16
2920.	comp32932_c0_seq1	2601	T	C	T	12/12	18/18
2921.	comp28279_c0_seq1	391	C	A	C	6/7	24/24

2922.	comp27063_c0_seq1	878	A	A	G	6/7	4/6
2923.	comp27063_c0_seq1	897	G	G	A	6/7	4/6
2924.	comp33555_c0_seq58	119	T	G	T	10/13	15/17
2925.	comp33555_c0_seq58	124	A	G	A	10/15	14/16
2926.	comp33555_c0_seq58	125	T	A	T	10/15	14/16
2927.	comp32085_c0_seq4	137	C	T	C	6/9	6/8
2928.	comp32583_c0_seq2	292	T	C	T	6/6	10/10
2929.	comp24768_c0_seq1	266	A	A	G	15/15	6/6
2930.	comp31232_c0_seq1	1279	G	G	T	7/10	6/8
2931.	comp32621_c0_seq18	104	C	C	T	20/20	12/12
2932.	comp3943_c0_seq1	87	T	G	T	6/6	16/20
2933.	comp3943_c0_seq1	89	A	C	A	6/7	16/20
2934.	comp3943_c0_seq1	90	C	G	C	6/7	18/23
2935.	comp3943_c0_seq1	91	A	C	A	6/7	18/23
2936.	comp3943_c0_seq1	93	A	C	A	7/7	19/24
2937.	comp3943_c0_seq1	96	A	C	A	13/14	20/25
2938.	comp3943_c0_seq1	97	T	C	T	10/15	20/26
2939.	comp3943_c0_seq1	100	G	T	G	15/21	20/26
2940.	comp18717_c0_seq1	133	T	C	T	9/9	4/6
2941.	comp24343_c0_seq1	666	C	C	T	11/13	8/12
2942.	comp30149_c0_seq3	259	T	C	T	4/6	4/4
2943.	comp30149_c0_seq3	268	G	T	G	4/6	6/6
2944.	comp30149_c0_seq3	272	A	G	A	5/6	6/6
2945.	comp30149_c0_seq3	277	C	T	C	5/6	6/6
2946.	comp30149_c0_seq3	344	T	T	C	6/8	4/6
2947.	comp33530_c0_seq10	929	C	T	C	9/9	6/6
2948.	comp33723_c0_seq10	76	A	G	A	19/28	32/46
2949.	comp19133_c0_seq2	102	C	C	A	12/12	8/8
2950.	comp19133_c0_seq2	107	C	C	T	12/12	6/6
2951.	comp23312_c0_seq1	118	A	T	A	4/6	6/6
2952.	comp32526_c0_seq1	1792	T	T	C	10/10	14/14
2953.	comp28841_c0_seq2	696	G	G	T	6/9	20/29
2954.	comp28841_c0_seq2	702	A	A	G	6/7	20/29
2955.	comp33510_c0_seq13	68	T	T	C	80/107	2/2
2956.	comp33510_c0_seq13	322	C	C	T	24/32	2/2
2957.	comp36002_c0_seq1	52	A	A	G	24/24	16/16
2958.	comp127850_c0_seq1	137	C	T	C	7/7	10/10
2959.	comp25688_c0_seq1	3004	A	T	A	10/12	12/16
2960.	comp25688_c0_seq1	3017	T	C	T	5/7	12/16
2961.	comp25688_c0_seq1	3026	C	T	C	5/7	12/16
2962.	comp30201_c0_seq1	523	A	T	A	10/10	26/26
2963.	comp33425_c0_seq7	3217	T	A	T	4/6	15/19
2964.	comp33425_c0_seq7	3218	A	T	A	4/6	15/19
2965.	comp33425_c0_seq7	3219	C	T	C	4/6	15/19
2966.	comp33425_c0_seq7	3225	G	A	G	4/6	11/15
2967.	comp33425_c0_seq7	3229	C	A	C	4/6	11/15
2968.	comp4582_c0_seq1	226	G	G	A	8/10	4/4
2969.	comp31668_c0_seq2	198	T	T	C	6/6	4/6
2970.	comp31668_c0_seq2	203	G	A	G	4/6	6/6
2971.	comp33480_c0_seq1	1574	A	A	G	26/26	8/8
2972.	comp29620_c0_seq1	372	G	G	A	8/12	10/14
2973.	comp23353_c0_seq2	751	A	C	A	18/25	4/4
2974.	comp23353_c0_seq2	752	T	C	T	19/26	4/4
2975.	comp33021_c0_seq1	697	C	G	C	4/6	8/12
2976.	comp31849_c0_seq3	670	C	T	C	6/9	1/1
2977.	comp33722_c0_seq4	127	G	G	A	28/29	16/21
2978.	comp33722_c0_seq4	130	T	T	C	31/36	17/25
2979.	comp33722_c0_seq4	141	T	T	C	29/35	21/25
2980.	comp33722_c0_seq4	144	A	A	G	32/41	22/26
2981.	comp33722_c0_seq4	176	A	A	G	29/41	26/34
2982.	comp33722_c0_seq4	178	G	G	A	30/42	26/34
2983.	comp33722_c0_seq4	182	T	T	C	28/40	26/34
2984.	comp33722_c0_seq4	240	C	C	T	25/29	10/14
2985.	comp33722_c0_seq4	261	A	A	C	18/19	8/12
2986.	comp33722_c0_seq4	275	A	A	G	18/18	4/6
2987.	comp31238_c1_seq3	341	T	T	C	108/109	5/5
2988.	comp31238_c1_seq3	370	T	T	G	124/130	5/5
2989.	comp31238_c1_seq3	373	C	C	T	151/155	5/5
2990.	comp31238_c1_seq3	374	A	A	G	144/156	5/5
2991.	comp31238_c1_seq3	387	T	T	C	266/287	5/5
2992.	comp31279_c0_seq1	67	A	A	T	72/107	6/8
2993.	comp31279_c0_seq1	80	T	T	A	81/118	6/9
2994.	comp31279_c0_seq1	155	C	C	T	57/79	4/4
2995.	comp31279_c0_seq1	168	G	G	A	67/87	4/4
2996.	comp31279_c0_seq1	196	A	A	G	108/112	4/6

2997.	comp31279_c0_seq1	228	C	C	T	81/122	12/12
2998.	comp5920_c0_seq1	444	C	T	C	6/7	6/6
2999.	comp124218_c0_seq1	206	A	A	C	10/10	20/20
3000.	comp28417_c0_seq1	181	A	A	G	26/26	18/26
3001.	comp22204_c0_seq1	130	A	G	A	5/6	4/6
3002.	comp22204_c0_seq1	132	T	C	T	5/6	4/6
3003.	comp22204_c0_seq1	136	A	G	A	8/9	4/6
3004.	comp22204_c0_seq1	141	G	T	G	9/10	4/6
3005.	comp22007_c0_seq1	5	G	A	C	4/6	4/6
3006.	comp29782_c0_seq1	1239	G	A	G	14/14	66/71
3007.	comp30048_c0_seq1	537	T	C	T	13/16	10/12
3008.	comp33425_c0_seq9	1711	A	G	A	8/11	19/28
3009.	comp33425_c0_seq21	787	C	T	C	6/7	13/19
3010.	comp33425_c0_seq21	793	G	A	G	6/7	13/17
3011.	comp26597_c0_seq2	491	G	G	C	88/89	12/12
3012.	comp69009_c0_seq1	5	C	C	G	8/8	6/8
3013.	comp31257_c0_seq1	342	G	G	A	8/8	8/12
3014.	comp32355_c0_seq2	2065	G	T	G	9/12	6/8
3015.	comp184502_c0_seq1	180	G	G	A	15/15	10/10
3016.	comp3455_c0_seq1	83	C	T	C	8/11	16/16
3017.	comp31310_c0_seq1	730	A	G	A	20/20	18/18
3018.	comp33601_c0_seq33	79	T	T	C	5/6	2/2
3019.	comp19142_c0_seq1	268	A	A	G	12/12	4/4
3020.	comp33596_c0_seq27	1789	G	G	A	7/8	8/12
3021.	comp33596_c0_seq27	1887	A	A	G	10/15	4/6
3022.	comp33596_c0_seq27	2037	G	G	A	8/12	6/7
3023.	comp23111_c0_seq2	330	A	A	G	5/7	8/8
3024.	comp23111_c0_seq2	332	A	A	G	5/7	7/8
3025.	comp33478_c0_seq8	286	A	A	G	19/28	9/13
3026.	comp33478_c0_seq8	289	T	T	G	19/28	9/13
3027.	comp17518_c0_seq1	38	T	C	T	6/7	2/2
3028.	comp19277_c0_seq1	3	G	G	T	6/7	4/6
3029.	comp29547_c0_seq3	54	C	C	T	44/56	6/8
3030.	comp29547_c0_seq3	55	A	A	G	41/56	6/8
3031.	comp29547_c0_seq3	56	G	G	A	50/58	6/8
3032.	comp29547_c0_seq3	64	A	A	T	41/61	6/8
3033.	comp29547_c0_seq3	66	C	C	T	41/62	6/8
3034.	comp29547_c0_seq3	139	T	T	C	46/67	6/8
3035.	comp29547_c0_seq3	210	A	G	A	23/30	10/10
3036.	comp11661_c0_seq1	270	A	G	A	6/9	12/12
3037.	comp33356_c0_seq8	1300	T	C	T	9/12	7/7
3038.	comp33356_c0_seq8	1313	C	T	C	9/12	9/11
3039.	comp33356_c0_seq8	1315	G	A	G	10/13	9/13
3040.	comp33356_c0_seq8	1319	G	C	G	8/11	11/13
3041.	comp33341_c0_seq1	1093	C	T	C	13/18	28/38
3042.	comp33341_c0_seq1	1096	A	G	A	13/18	30/38
3043.	comp33341_c0_seq1	1105	T	C	T	12/17	26/36
3044.	comp26825_c0_seq1	1659	A	A	C	8/10	2/2
3045.	comp26825_c0_seq1	1660	A	A	T	8/10	3/4
3046.	comp31238_c1_seq10	155	A	G	A	24/30	4/6
3047.	comp31238_c1_seq10	163	A	G	A	16/22	2/2
3048.	comp31238_c1_seq10	264	C	C	T	6/9	6/8
3049.	comp31238_c1_seq10	268	T	T	G	9/9	6/8
3050.	comp31238_c1_seq10	277	T	T	G	9/9	6/8
3051.	comp31238_c1_seq10	301	T	T	C	9/9	8/10
3052.	comp31238_c1_seq10	332	C	T	C	7/9	10/12
3053.	comp11163_c0_seq1	79	C	C	T	13/13	10/10
3054.	comp12789_c1_seq1	45	A	A	G	168/195	4/4
3055.	comp12789_c1_seq1	47	T	A	T	143/206	4/4
3056.	comp12789_c1_seq1	88	A	A	T	266/345	10/10
3057.	comp12789_c1_seq1	89	C	C	T	276/354	10/10
3058.	comp12789_c1_seq1	90	A	A	G	279/354	10/10
3059.	comp12789_c1_seq1	104	G	G	T	324/373	8/8
3060.	comp12789_c1_seq1	105	C	C	T	337/378	8/8
3061.	comp12789_c1_seq1	106	T	T	C	336/375	8/8
3062.	comp12789_c1_seq1	117	T	T	G	319/357	6/6
3063.	comp12789_c1_seq1	126	T	T	C	306/341	6/6
3064.	comp12789_c1_seq1	132	T	T	G	268/299	6/6
3065.	comp12789_c1_seq1	134	T	T	A	266/297	6/6
3066.	comp32906_c0_seq1	1422	C	C	A	7/9	6/8
3067.	comp32201_c0_seq2	1380	T	C	T	16/22	32/32
3068.	comp26924_c0_seq1	359	T	C	T	8/11	8/12
3069.	comp26924_c0_seq1	374	T	C	T	7/7	10/13
3070.	comp26924_c0_seq1	389	T	C	T	6/7	10/12
3071.	comp26924_c0_seq1	395	T	A	T	6/8	10/13

3072.	comp26924_c0_seq1	401	T	A	T	6/8	10/13
3073.	comp26924_c0_seq1	410	G	A	G	5/7	8/9
3074.	comp26924_c0_seq1	428	A	T	A	5/7	12/13
3075.	comp31626_c0_seq2	888	A	G	A	12/16	15/15
3076.	comp31626_c0_seq2	2256	G	G	A	13/15	13/13
3077.	comp25225_c0_seq1	608	C	C	G	10/13	5/7
3078.	comp32474_c0_seq1	2532	T	A	T	5/6	24/36
3079.	comp32185_c0_seq1	341	C	C	T	8/8	6/8
3080.	comp33333_c0_seq2	1056	T	T	C	4/6	4/6
3081.	comp33336_c0_seq3	101	G	T	G	7/7	10/14
3082.	comp33336_c0_seq3	120	A	G	A	7/7	12/16
3083.	comp33711_c0_seq4	94	T	T	C	30/36	34/36
3084.	comp33711_c0_seq4	95	G	G	A	31/37	32/34
3085.	comp33711_c0_seq4	98	T	T	A	31/37	32/36
3086.	comp33711_c0_seq4	131	G	G	A	53/59	22/30
3087.	comp33711_c0_seq4	132	A	A	G	53/59	22/30
3088.	comp33711_c0_seq4	369	A	A	G	21/21	4/6
3089.	comp25868_c0_seq2	131	G	G	A	6/8	4/6
3090.	comp31656_c0_seq1	839	A	A	G	9/13	6/6
3091.	comp31656_c0_seq1	851	A	A	G	9/12	6/6
3092.	comp25167_c0_seq1	85	G	G	A	14/14	12/12
3093.	comp28423_c0_seq2	3	T	G	T	6/6	2/3
3094.	comp17264_c0_seq1	239	A	A	G	33/34	10/10
3095.	comp28467_c0_seq1	340	C	T	C	7/7	16/16
3096.	comp55478_c0_seq1	18	T	C	T	4/6	2/2
3097.	comp26575_c0_seq2	20	G	G	A	14/14	6/6
3098.	comp129353_c0_seq1	131	G	G	A	31/31	4/4
3099.	comp129353_c0_seq1	140	C	C	G	32/32	4/4
3100.	comp92230_c0_seq1	412	G	C	G	7/7	10/10
3101.	comp29705_c0_seq1	195	A	A	G	13/16	2/2
3102.	comp29705_c0_seq1	197	A	A	T	13/16	2/2
3103.	comp30660_c0_seq1	348	G	C	G	8/12	44/44
3104.	comp33492_c1_seq4	4882	C	C	G	5/6	8/12
3105.	comp33492_c1_seq4	4883	A	A	C	5/6	8/12
3106.	comp30992_c0_seq2	732	T	T	C	6/6	8/8
3107.	comp32082_c0_seq1	966	A	A	C	18/18	18/18
3108.	comp33535_c0_seq1	3132	T	C	T	32/48	106/149
3109.	comp16015_c0_seq1	158	T	T	C	8/8	12/12
3110.	comp24926_c0_seq1	683	C	T	C	14/14	30/30
3111.	comp33629_c0_seq12	152	G	G	C	6/6	14/18
3112.	comp161847_c0_seq1	463	A	A	T	4/6	2/2
3113.	comp10402_c0_seq1	71	A	T	A	7/7	14/14
3114.	comp124462_c0_seq1	221	T	C	T	4/6	6/6
3115.	comp30474_c0_seq1	933	C	T	C	6/8	6/6
3116.	comp29328_c0_seq1	223	T	T	C	14/14	14/14
3117.	comp32584_c0_seq1	57	A	A	T	5/7	10/10
3118.	comp18961_c0_seq1	1164	T	A	T	14/19	20/20
3119.	comp31626_c0_seq1	888	A	G	A	17/17	25/25
3120.	comp31626_c0_seq1	2340	G	G	A	16/18	15/15
3121.	comp23740_c0_seq1	196	G	T	G	6/6	6/6
3122.	comp195430_c0_seq1	156	C	T	C	7/10	4/4
3123.	comp33688_c0_seq2	2007	C	C	T	12/16	22/33
3124.	comp33688_c0_seq2	2043	T	T	G	7/9	24/35
3125.	comp33688_c0_seq2	2045	A	A	T	7/9	24/35
3126.	comp33688_c0_seq2	2115	C	C	T	5/7	12/18
3127.	comp33688_c0_seq2	2127	G	G	A	9/11	14/18
3128.	comp33688_c0_seq2	2140	C	C	G	11/13	14/20
3129.	comp33688_c0_seq2	2141	T	T	A	11/13	14/20
3130.	comp33688_c0_seq2	2161	C	C	T	13/15	14/20
3131.	comp33574_c0_seq8	558	T	C	T	10/15	17/21
3132.	comp33686_c0_seq4	2854	G	G	C	5/6	4/5
3133.	comp33686_c0_seq4	2863	A	A	T	5/6	4/5
3134.	comp33686_c0_seq4	2867	A	A	G	6/7	4/5
3135.	comp27737_c0_seq1	119	C	C	T	19/27	4/4
3136.	comp26660_c0_seq1	1115	C	T	C	10/14	10/14
3137.	comp31175_c0_seq2	589	A	A	C	7/7	4/6
3138.	comp31175_c0_seq2	595	A	A	T	7/7	4/6
3139.	comp31175_c0_seq2	607	T	T	G	7/7	4/6
3140.	comp33530_c0_seq34	449	C	T	C	9/9	8/8
3141.	comp31217_c0_seq4	147	C	T	C	6/6	15/15
3142.	comp25836_c0_seq2	443	G	G	A	9/9	5/7
3143.	comp12238_c0_seq1	120	T	C	T	4/6	10/10
3144.	comp32856_c0_seq4	115	A	G	A	14/18	18/18
3145.	comp32856_c0_seq4	119	T	C	T	14/18	16/20
3146.	comp33510_c0_seq18	82	T	G	T	7/10	2/2

L

3147.	comp33510_c0_seq18	84	A	G	A	10/12	2/2
3148.	comp33510_c0_seq18	264	G	G	A	14/20	1/1
3149.	comp28009_c0_seq1	350	A	A	G	7/10	12/16
3150.	comp28009_c0_seq1	358	T	T	C	8/11	12/16
3151.	comp10986_c0_seq1	106	C	C	T	22/30	2/2
3152.	comp10986_c0_seq1	119	G	G	T	21/28	2/2
3153.	comp10986_c0_seq1	170	A	A	C	8/11	2/2
3154.	comp33700_c0_seq30	268	G	G	A	5/7	4/6
3155.	comp33271_c0_seq1	230	G	C	G	8/9	16/17
3156.	comp27074_c0_seq1	389	A	A	G	16/21	14/20
3157.	comp27074_c0_seq1	398	C	C	T	21/26	14/18
3158.	comp28543_c0_seq1	51	T	T	C	32/35	5/5
3159.	comp28543_c0_seq1	77	T	T	C	36/38	7/7
3160.	comp28543_c0_seq1	78	T	T	A	34/38	7/7
3161.	comp28543_c0_seq1	79	A	A	T	35/39	7/7
3162.	comp28543_c0_seq1	89	T	T	A	31/36	9/9
3163.	comp28543_c0_seq1	90	A	A	C	37/38	9/9
3164.	comp28543_c0_seq1	91	A	A	G	36/38	7/9
3165.	comp28543_c0_seq1	207	A	A	G	17/24	2/2
3166.	comp32797_c0_seq4	218	C	T	C	6/9	16/18
3167.	comp32797_c0_seq4	219	C	T	C	8/10	18/20
3168.	comp32797_c0_seq4	236	C	T	C	7/10	19/21
3169.	comp32797_c0_seq4	280	C	T	C	8/11	36/40
3170.	comp33702_c0_seq19	142	G	A	G	6/7	16/20
3171.	comp33702_c0_seq19	144	T	C	T	6/6	14/18
3172.	comp33702_c0_seq19	176	A	G	A	6/8	8/8
3173.	comp33702_c0_seq19	191	G	T	G	5/7	8/8
3174.	comp33702_c0_seq19	193	A	G	A	5/7	6/6
3175.	comp33702_c0_seq19	208	G	A	G	5/7	4/4
3176.	comp32945_c0_seq5	49	T	T	C	9/13	6/8
3177.	comp32945_c0_seq5	597	G	A	G	4/6	19/27
3178.	comp28407_c0_seq2	417	G	T	G	6/6	2/2
3179.	comp21864_c0_seq1	240	A	G	A	14/14	14/14
3180.	comp27584_c0_seq1	349	G	G	A	14/14	16/16
3181.	comp30400_c0_seq1	53	A	A	G	4/6	4/6
3182.	comp30400_c0_seq1	302	T	T	C	47/70	4/6
3183.	comp26225_c0_seq1	1078	G	A	G	8/11	42/44
3184.	comp28393_c0_seq1	38	A	A	G	17/20	2/2
3185.	comp28393_c0_seq1	42	G	G	A	18/23	2/2
3186.	comp28393_c0_seq1	191	G	G	A	55/83	18/18
3187.	comp31627_c0_seq1	67	T	C	T	14/21	22/26
3188.	comp33136_c0_seq4	147	A	A	G	8/11	16/20
3189.	comp33595_c0_seq4	430	A	G	A	6/9	14/16
3190.	comp30540_c0_seq1	1023	G	G	A	36/36	24/24
3191.	comp28897_c0_seq1	215	G	G	C	13/14	3/3
3192.	comp19779_c0_seq1	56	T	T	C	12/12	6/6
3193.	comp19779_c0_seq1	90	T	T	C	13/14	12/14
3194.	comp19779_c0_seq1	94	C	C	T	11/12	12/14
3195.	comp19779_c0_seq1	98	T	T	C	12/13	14/16
3196.	comp19779_c0_seq1	153	A	A	G	6/7	8/12
3197.	comp33640_c0_seq80	567	T	T	C	6/6	4/6
3198.	comp32004_c0_seq1	15	T	T	C	50/51	12/16
3199.	comp32004_c0_seq1	17	T	T	C	52/54	12/16
3200.	comp32773_c0_seq8	274	C	C	T	11/12	4/4
3201.	comp32773_c0_seq8	278	A	A	G	9/12	4/4
3202.	comp32773_c0_seq8	293	A	A	G	12/14	4/6
3203.	comp29852_c0_seq1	115	G	G	A	17/17	6/7
3204.	comp33286_c0_seq1	135	T	A	T	4/6	12/18
3205.	comp6598_c0_seq1	4	C	C	G	5/7	2/2
3206.	comp31545_c0_seq3	346	C	T	C	4/6	18/18
3207.	comp344671_c0_seq1	103	G	A	G	4/6	2/2
3208.	comp30838_c0_seq1	563	T	C	T	53/53	12/12
3209.	comp60484_c0_seq1	24	A	A	T	11/15	2/2
3210.	comp31880_c0_seq1	341	G	G	C	13/13	19/21
3211.	comp31880_c0_seq1	513	C	T	C	9/9	10/10
3212.	comp32429_c0_seq1	438	A	T	A	4/6	11/13
3213.	comp31483_c0_seq1	60	T	G	T	4/6	10/14
3214.	comp15477_c0_seq1	154	G	G	A	15/15	16/16
3215.	comp174353_c0_seq1	167	T	T	C	8/8	6/8
3216.	comp174353_c0_seq1	176	G	G	A	8/8	6/8
3217.	comp32260_c0_seq2	2508	C	A	C	4/6	4/4
3218.	comp26734_c0_seq1	33	C	C	T	9/9	4/6
3219.	comp32874_c0_seq47	620	T	T	A	8/9	3/3
3220.	comp32874_c0_seq47	623	A	A	T	9/10	3/3
3221.	comp32874_c0_seq47	632	T	T	C	8/9	3/3

3222.	comp32874_c0_seq47	641	G	G	T	8/9	3/3
3223.	comp25911_c0_seq1	293	T	A	T	7/7	48/48
3224.	comp25911_c0_seq1	294	C	A	C	7/7	48/48
3225.	comp31730_c0_seq1	2116	A	T	A	4/6	55/67
3226.	comp31730_c0_seq1	2416	T	C	T	8/12	61/66
3227.	comp31730_c0_seq1	2443	T	C	T	9/10	51/58
3228.	comp31730_c0_seq1	2446	G	A	G	6/7	49/56
3229.	comp31730_c0_seq1	2449	A	G	A	6/7	55/63
3230.	comp33215_c0_seq16	226	A	G	A	8/8	10/10
3231.	comp32470_c0_seq11	228	T	G	T	7/8	4/4
3232.	comp32470_c0_seq11	550	A	C	A	8/11	8/8
3233.	comp32470_c0_seq11	551	T	C	T	8/11	8/8
3234.	comp6925_c0_seq1	361	G	G	T	47/47	10/10
3235.	comp30117_c0_seq2	392	G	G	T	25/29	12/17
3236.	comp30117_c0_seq2	407	C	C	A	23/28	18/21
3237.	comp30117_c0_seq2	419	C	C	T	20/27	20/23
3238.	comp28804_c0_seq1	326	T	A	T	10/11	26/26
3239.	comp17616_c0_seq1	484	C	T	C	9/9	12/12
3240.	comp8535_c0_seq1	801	A	G	A	4/6	4/4
3241.	comp30346_c0_seq1	248	G	G	A	7/10	16/24
3242.	comp30346_c0_seq1	256	T	T	A	7/10	20/24
3243.	comp24516_c0_seq1	384	A	A	C	9/9	14/20
3244.	comp15375_c0_seq1	323	T	C	T	12/12	12/12
3245.	comp32826_c0_seq1	332	A	G	A	15/16	9/13
3246.	comp32826_c0_seq1	346	C	T	C	11/15	17/21
3247.	comp33625_c0_seq4	394	A	A	C	17/24	20/28
3248.	comp33625_c0_seq4	679	G	G	A	18/21	12/18
3249.	comp33625_c0_seq4	697	G	G	A	14/18	16/20
3250.	comp33345_c0_seq4	1074	A	A	G	7/9	8/9
3251.	comp33345_c0_seq4	1079	T	T	C	7/9	8/9
3252.	comp33345_c0_seq4	1082	C	C	T	7/8	8/9
3253.	comp33345_c0_seq4	1104	A	A	T	5/6	4/5
3254.	comp41765_c0_seq1	12	A	A	G	48/48	26/26
3255.	comp10202_c0_seq1	254	A	A	G	6/6	6/8
3256.	comp10202_c0_seq1	256	A	A	G	6/6	10/10
3257.	comp10202_c0_seq1	262	G	G	T	6/6	10/10
3258.	comp155847_c0_seq1	93	G	A	G	8/10	18/18
3259.	comp328222_c0_seq1	125	C	C	T	7/9	2/2
3260.	comp328222_c0_seq1	132	A	A	G	7/9	2/2
3261.	comp328222_c0_seq1	133	T	T	C	7/9	2/2
3262.	comp32030_c0_seq1	74	T	C	T	13/18	9/13
3263.	comp33205_c0_seq1	268	G	A	G	9/9	36/36
3264.	comp124491_c0_seq1	303	T	T	A	11/13	2/2
3265.	comp27930_c0_seq2	620	G	G	T	25/25	12/12
3266.	comp30967_c0_seq1	911	A	C	A	6/6	8/12
3267.	comp30967_c0_seq1	914	A	C	A	6/6	8/12
3268.	comp30881_c0_seq1	426	T	C	T	11/11	18/18
3269.	comp32255_c0_seq1	225	G	C	G	9/9	41/41
3270.	comp32255_c0_seq1	250	A	T	A	10/10	44/44
3271.	comp25938_c0_seq1	183	G	T	G	4/6	12/12
3272.	comp28559_c0_seq1	219	A	C	A	11/12	18/18
3273.	comp28559_c0_seq1	702	C	G	C	7/9	20/20
3274.	comp24398_c0_seq1	609	T	G	T	12/12	14/14
3275.	comp24037_c0_seq1	401	A	A	T	14/14	10/10
3276.	comp32797_c0_seq2	224	T	C	T	8/9	10/10
3277.	comp32797_c0_seq2	236	T	C	T	5/7	10/10
3278.	comp33471_c0_seq104	37	A	A	G	18/18	3/3
3279.	comp33471_c0_seq104	102	A	A	G	27/27	12/14
3280.	comp33471_c0_seq104	111	T	T	C	26/26	14/14
3281.	comp33471_c0_seq104	290	T	T	C	26/26	10/14
3282.	comp33471_c0_seq104	297	C	C	T	23/24	10/14
3283.	comp33471_c0_seq104	309	C	C	T	19/20	12/14
3284.	comp33471_c0_seq104	311	T	T	C	20/20	11/14
3285.	comp33471_c0_seq104	318	C	C	T	19/20	13/14
3286.	comp33471_c0_seq104	323	T	T	C	19/19	10/14
3287.	comp33679_c0_seq5	1340	A	A	C	6/6	5/7
3288.	comp33679_c0_seq5	2169	A	A	T	4/6	9/9
3289.	comp23757_c0_seq1	76	A	A	G	8/8	4/4
3290.	comp30897_c0_seq1	896	G	A	G	8/8	18/18
3291.	comp33723_c0_seq3	386	C	T	C	16/23	34/38
3292.	comp33723_c0_seq3	453	T	T	C	10/10	13/15
3293.	comp33723_c0_seq3	895	G	G	A	15/17	9/12
3294.	comp33723_c0_seq3	952	G	A	G	19/27	15/19
3295.	comp33723_c0_seq3	963	A	C	A	19/26	23/27
3296.	comp33723_c0_seq3	1510	T	T	A	18/18	6/6

3297.	comp33723_c0_seq3	1518	G	G	A	13/13	6/6
3298.	comp33723_c0_seq3	1522	A	A	G	13/13	4/4
3299.	comp33723_c0_seq3	1569	T	T	A	6/6	4/4
3300.	comp147302_c0_seq1	110	T	T	C	14/16	4/4
3301.	comp147302_c0_seq1	119	G	G	T	14/16	4/4
3302.	comp147302_c0_seq1	153	A	A	G	13/15	4/4
3303.	comp147302_c0_seq1	159	A	A	G	12/14	4/4
3304.	comp33269_c0_seq7	52	T	C	T	4/6	4/4
3305.	comp33269_c0_seq7	54	C	T	C	4/6	4/4
3306.	comp33060_c0_seq1	80	C	T	C	7/9	16/18
3307.	comp21887_c0_seq1	86	A	A	G	8/8	14/18
3308.	comp21887_c0_seq1	91	T	T	A	8/8	14/18
3309.	comp21887_c0_seq1	92	G	G	T	8/8	14/18
3310.	comp21887_c0_seq1	96	T	T	G	7/7	14/18
3311.	comp21887_c0_seq1	103	T	T	A	6/6	14/18
3312.	comp307229_c0_seq1	4	C	C	G	5/6	2/2
3313.	comp28025_c0_seq1	357	G	G	A	7/7	4/6
3314.	comp28025_c0_seq1	366	T	T	A	7/9	6/6
3315.	comp32085_c0_seq1	233	T	T	A	4/6	8/12
3316.	comp32085_c0_seq1	292	T	T	C	7/9	8/12
3317.	comp33496_c0_seq12	157	C	T	C	9/12	8/12
3318.	comp29440_c0_seq2	123	C	T	C	11/11	43/43
3319.	comp44922_c0_seq1	2	C	G	C	14/19	2/2
3320.	comp44922_c0_seq1	3	T	G	T	48/52	2/2
3321.	comp24585_c0_seq1	284	T	C	T	5/6	2/2
3322.	comp100161_c0_seq1	95	G	G	A	32/33	4/6
3323.	comp29269_c0_seq1	940	A	A	G	24/31	18/18
3324.	comp26691_c0_seq1	870	T	C	T	6/6	12/12
3325.	comp33596_c0_seq26	1458	T	G	T	7/10	6/8
3326.	comp27015_c0_seq1	345	T	T	C	14/14	22/22
3327.	comp25972_c0_seq1	841	T	G	T	15/16	22/22
3328.	comp28778_c0_seq2	405	G	C	G	18/18	18/18
3329.	comp10380_c0_seq1	82	A	A	T	15/15	6/6
3330.	comp21957_c0_seq1	982	T	G	T	10/14	26/26
3331.	comp33333_c0_seq3	1974	A	A	C	6/6	2/2
3332.	comp33703_c0_seq22	150	T	C	T	12/13	30/41
3333.	comp33703_c0_seq22	154	A	G	A	12/13	31/42
3334.	comp31303_c0_seq1	871	C	G	C	12/12	12/12
3335.	comp33565_c0_seq9	136	C	A	C	7/10	9/13
3336.	comp33565_c0_seq9	143	T	G	T	6/9	9/13
3337.	comp450869_c0_seq1	185	A	A	G	4/6	2/2
3338.	comp32698_c0_seq1	605	A	T	A	25/25	32/32
3339.	comp33332_c0_seq7	1371	G	G	A	8/8	14/14
3340.	comp33026_c0_seq3	366	C	C	T	8/8	9/9
3341.	comp33026_c0_seq3	1431	A	A	T	14/14	6/8
3342.	comp33026_c0_seq3	1476	T	T	A	7/7	6/8
3343.	comp33026_c0_seq3	1503	A	A	T	6/6	4/6
3344.	comp33026_c0_seq3	1506	A	A	T	7/7	4/6
3345.	comp32317_c0_seq1	15	T	G	T	6/6	3/4
3346.	comp24757_c0_seq1	123	G	A	G	4/6	28/28
3347.	comp31222_c0_seq4	124	A	T	A	18/24	13/13
3348.	comp33550_c0_seq6	660	A	A	G	8/8	6/6
3349.	comp4809_c0_seq1	266	C	C	T	8/8	4/4
3350.	comp176814_c0_seq1	16	G	G	A	6/6	4/6
3351.	comp27293_c0_seq1	799	T	C	T	15/15	20/20
3352.	comp25992_c0_seq1	1729	T	C	T	10/10	52/52
3353.	comp31396_c0_seq1	160	T	T	A	19/22	16/24
3354.	comp31803_c0_seq1	1299	G	G	T	12/14	4/6
3355.	comp17928_c0_seq1	94	C	T	C	19/19	18/18
3356.	comp33203_c0_seq2	2234	C	A	C	12/16	8/8
3357.	comp33719_c0_seq5	431	A	G	A	4/6	36/46
3358.	comp26918_c0_seq2	403	C	C	A	20/20	10/12
3359.	comp20472_c0_seq1	24	T	T	C	10/15	4/4
3360.	comp20472_c0_seq1	25	C	C	T	10/15	4/4
3361.	comp31761_c0_seq1	523	T	G	T	15/21	8/12
3362.	comp32154_c0_seq1	761	C	G	C	11/11	20/20
3363.	comp33510_c0_seq7	68	T	T	C	100/131	2/2
3364.	comp33510_c0_seq7	201	C	C	T	199/222	5/7
3365.	comp33510_c0_seq7	226	T	T	G	218/275	7/10
3366.	comp33510_c0_seq7	261	T	T	G	258/318	7/7
3367.	comp33510_c0_seq7	263	T	T	A	262/321	7/7
3368.	comp33510_c0_seq7	267	C	C	T	286/347	7/7
3369.	comp33510_c0_seq7	296	C	C	T	299/329	5/5
3370.	comp33510_c0_seq7	702	A	A	T	12/15	4/4
3371.	comp33510_c0_seq7	723	C	C	T	9/13	4/4

3372.	comp32821_c0_seq4	580	A	G	A	4/6	7/10
3373.	comp33546_c0_seq7	568	T	C	T	7/8	9/13
3374.	comp33546_c0_seq7	576	G	A	G	6/6	11/15
3375.	comp3949_c0_seq1	127	T	T	C	11/13	2/2
3376.	comp29777_c0_seq1	185	A	A	G	128/164	4/6
3377.	comp29777_c0_seq1	186	T	T	G	115/157	6/6
3378.	comp28157_c0_seq1	142	A	A	G	9/9	8/9
3379.	comp28157_c0_seq1	214	T	G	T	4/6	12/16
3380.	comp286162_c0_seq1	79	C	C	T	7/7	4/4
3381.	comp406285_c0_seq1	28	A	C	T	5/7	1/1
3382.	comp27029_c0_seq1	6	C	C	G	14/18	1/1
3383.	comp30353_c0_seq1	1011	T	T	C	10/15	8/12
3384.	comp30353_c0_seq1	1046	A	A	G	7/10	8/12
3385.	comp9215_c0_seq1	232	T	T	A	4/6	6/6
3386.	comp9215_c0_seq1	263	A	A	G	4/6	6/8
3387.	comp29494_c0_seq3	337	C	T	C	7/7	30/30
3388.	comp29494_c0_seq3	371	G	G	A	6/8	17/20
3389.	comp26689_c0_seq1	649	A	G	A	11/11	12/12
3390.	comp13672_c0_seq1	137	C	T	C	4/6	8/8
3391.	comp13672_c0_seq1	139	A	G	A	4/6	8/8
3392.	comp13672_c0_seq1	146	T	A	T	4/6	8/8
3393.	comp13672_c0_seq1	147	T	C	T	4/6	8/8
3394.	comp13672_c0_seq1	162	G	A	G	4/6	8/8
3395.	comp33486_c0_seq7	53	T	C	T	10/10	26/26
3396.	comp32499_c0_seq2	1095	T	G	T	17/25	18/26
3397.	comp32499_c0_seq2	1428	G	A	G	5/7	6/8
3398.	comp192868_c0_seq1	158	A	A	G	5/7	2/2
3399.	comp29277_c0_seq1	489	G	G	A	19/19	18/18
3400.	comp33584_c0_seq1	1000	C	A	C	6/7	30/34
3401.	comp33665_c0_seq1	1046	T	T	C	11/11	4/5
3402.	comp33665_c0_seq1	1049	C	C	T	10/10	2/2
3403.	comp33727_c0_seq7	183	C	C	A	8/12	2/2
3404.	comp33727_c0_seq7	184	T	T	C	9/12	2/2
3405.	comp32686_c0_seq1	260	C	T	C	11/11	38/38
3406.	comp31771_c0_seq3	439	G	C	G	8/8	19/19
3407.	comp30564_c0_seq1	734	A	T	A	11/11	12/14
3408.	comp33492_c1_seq6	3975	C	C	T	6/6	10/14
3409.	comp33492_c1_seq6	3976	T	T	G	6/6	10/14
3410.	comp33492_c1_seq6	3977	C	C	G	6/6	10/14
3411.	comp33492_c1_seq6	3978	A	A	C	6/6	10/14
3412.	comp18547_c0_seq1	862	G	A	G	21/21	84/84
3413.	comp27440_c0_seq1	687	G	A	G	4/6	31/35
3414.	comp27440_c0_seq1	688	T	G	T	4/6	27/31
3415.	comp27440_c0_seq1	694	G	A	G	4/6	25/29
3416.	comp32707_c0_seq1	78	A	A	T	42/43	6/8
3417.	comp32707_c0_seq1	82	C	C	T	42/43	6/8
3418.	comp32707_c0_seq1	83	A	A	G	42/43	6/8
3419.	comp32707_c0_seq1	131	C	C	G	41/42	6/8
3420.	comp32707_c0_seq1	150	T	T	C	34/35	4/6
3421.	comp33569_c0_seq3	300	C	A	C	4/6	4/4
3422.	comp2643_c0_seq1	302	C	A	C	4/6	4/6
3423.	comp2643_c0_seq1	312	A	G	A	4/6	4/6
3424.	comp2643_c0_seq1	318	A	G	A	4/6	6/8
3425.	comp32680_c0_seq1	111	C	C	T	8/9	14/20
3426.	comp32680_c0_seq1	117	G	G	C	10/11	16/22
3427.	comp27076_c0_seq1	527	G	C	G	6/9	10/10
3428.	comp29622_c0_seq1	701	A	A	G	92/136	23/25
3429.	comp29622_c0_seq1	727	G	G	A	67/94	15/17
3430.	comp29563_c0_seq1	117	C	C	T	16/17	6/7
3431.	comp29563_c0_seq1	144	C	T	C	9/13	8/9
3432.	comp29563_c0_seq1	147	T	C	T	9/13	8/9
3433.	comp258739_c0_seq1	191	T	T	C	6/6	4/4
3434.	comp161200_c0_seq1	195	G	G	T	16/16	4/4
3435.	comp33712_c0_seq5	290	T	T	C	20/28	24/32
3436.	comp32931_c0_seq2	99	T	T	G	9/12	19/22
3437.	comp33573_c0_seq1	775	T	C	T	9/9	12/12
3438.	comp33703_c0_seq1	814	C	C	A	5/7	9/9
3439.	comp15909_c0_seq1	162	A	C	A	6/6	2/2
3440.	comp32105_c0_seq1	136	C	C	A	14/14	6/6
3441.	comp15841_c0_seq1	120	G	A	G	7/7	17/17
3442.	comp31231_c0_seq1	118	G	G	T	11/15	2/2
3443.	comp31231_c0_seq1	152	G	G	C	16/23	12/16
3444.	comp31231_c0_seq1	153	A	A	T	16/23	12/16
3445.	comp31231_c0_seq1	233	A	A	C	12/16	8/10
3446.	comp31231_c0_seq1	272	T	T	A	12/14	4/6

3447.	comp317731_c0_seq1	68	C	C	A	6/6	4/4
3448.	comp24157_c0_seq1	77	A	A	C	6/6	4/6
3449.	comp24157_c0_seq1	78	G	G	A	5/6	4/6
3450.	comp24157_c0_seq1	99	C	C	T	9/10	4/6
3451.	comp20453_c0_seq1	518	A	A	T	6/6	4/4
3452.	comp24020_c0_seq1	129	G	G	A	6/8	2/2
3453.	comp29303_c0_seq1	185	G	A	G	13/16	10/12
3454.	comp29303_c0_seq1	193	C	T	C	9/10	10/12
3455.	comp32773_c0_seq52	76	C	C	T	8/12	3/3
3456.	comp15892_c0_seq1	213	C	G	C	4/6	2/2
3457.	comp29605_c0_seq2	243	G	G	A	19/19	4/6
3458.	comp29785_c0_seq1	111	A	A	G	17/22	14/14
3459.	comp134941_c0_seq1	73	T	C	T	6/7	8/8
3460.	comp8014_c0_seq2	1	A	A	T	6/6	2/2
3461.	comp30149_c0_seq1	413	G	T	G	4/6	14/14
3462.	comp9395_c0_seq1	135	T	T	C	11/11	6/6
3463.	comp26341_c0_seq1	1251	G	A	G	12/12	44/44
3464.	comp32916_c0_seq20	385	C	C	A	34/34	26/34
3465.	comp31138_c0_seq1	1897	C	C	T	30/34	17/17
3466.	comp33483_c0_seq1	1765	G	G	C	23/28	2/2
3467.	comp33483_c0_seq1	1784	A	A	G	23/28	2/2
3468.	comp33483_c0_seq1	2066	T	T	C	29/39	4/4
3469.	comp33483_c0_seq1	3650	G	G	A	19/24	4/4
3470.	comp33483_c0_seq1	5459	G	G	T	29/33	4/4
3471.	comp33483_c0_seq1	5478	C	C	T	29/33	4/4
3472.	comp31849_c0_seq11	630	C	T	C	8/9	4/4
3473.	comp33640_c0_seq45	616	T	G	T	6/8	4/5
3474.	comp33640_c0_seq45	619	T	A	T	6/8	2/3
3475.	comp33371_c0_seq10	760	C	G	C	6/8	16/16
3476.	comp31462_c0_seq7	1826	G	T	G	7/7	11/11
3477.	comp83730_c0_seq1	6	C	C	G	8/8	4/6
3478.	comp32499_c0_seq1	1424	C	C	T	11/12	6/9
3479.	comp32499_c0_seq1	1457	T	T	A	9/11	10/15
3480.	comp32681_c0_seq1	2964	C	T	C	14/18	47/47
3481.	comp27851_c0_seq3	829	T	T	A	6/6	7/7
3482.	comp25888_c0_seq1	223	C	T	C	5/6	12/16
3483.	comp25888_c0_seq1	229	C	T	C	5/6	12/16
3484.	comp32487_c0_seq1	304	A	T	A	6/8	22/30
3485.	comp21429_c0_seq1	96	G	T	G	4/6	8/8
3486.	comp33681_c0_seq26	1917	A	A	G	12/12	6/9
3487.	comp33681_c0_seq26	1922	C	C	T	12/12	6/9
3488.	comp33681_c0_seq26	1924	T	T	C	11/11	6/9
3489.	comp33681_c0_seq26	1926	G	G	A	13/13	6/9
3490.	comp236048_c0_seq1	119	G	G	A	8/8	4/6
3491.	comp21300_c0_seq1	87	A	A	G	13/13	4/4
3492.	comp33555_c0_seq31	760	T	T	C	9/11	4/6
3493.	comp19084_c0_seq1	331	T	G	T	7/7	22/22
3494.	comp19084_c0_seq1	332	G	A	G	7/7	22/22
3495.	comp24096_c0_seq1	366	C	C	T	15/15	4/4
3496.	comp33425_c0_seq3	1718	A	C	A	6/9	39/44
3497.	comp29710_c1_seq1	17	A	A	G	49/53	4/4
3498.	comp29710_c1_seq1	76	C	C	T	185/214	4/6
3499.	comp29710_c1_seq1	83	C	C	T	178/218	4/4
3500.	comp29710_c1_seq1	87	T	T	C	200/225	8/8
3501.	comp29710_c1_seq1	96	A	A	G	220/243	6/6
3502.	comp29710_c1_seq1	104	T	T	C	233/243	4/6
3503.	comp31622_c0_seq1	790	A	G	A	16/17	27/39
3504.	comp28862_c0_seq3	56	C	C	T	8/9	2/2
3505.	comp28862_c0_seq3	101	T	T	C	10/12	2/2
3506.	comp33332_c0_seq2	1811	G	G	A	7/7	6/6
3507.	comp31694_c0_seq1	1509	G	G	C	16/16	18/26
3508.	comp27805_c0_seq1	1064	A	T	A	8/8	12/12
3509.	comp25507_c0_seq2	84	A	A	C	5/7	4/4
3510.	comp25507_c0_seq2	96	T	T	C	8/10	4/4
3511.	comp25507_c0_seq2	105	C	C	T	8/10	2/2
3512.	comp25507_c0_seq2	231	A	G	A	13/16	6/7
3513.	comp33573_c0_seq8	448	C	C	T	5/6	7/9
3514.	comp33573_c0_seq8	450	G	G	C	6/7	7/9
3515.	comp33206_c0_seq41	131	T	T	C	7/7	7/10
3516.	comp33206_c0_seq41	140	A	A	C	7/8	11/14
3517.	comp31771_c0_seq2	1206	G	C	G	6/6	20/20
3518.	comp32434_c0_seq1	1225	G	C	G	12/12	20/20
3519.	comp24220_c0_seq1	626	A	A	G	5/6	4/4
3520.	comp24220_c0_seq1	629	C	C	G	5/6	4/4
3521.	comp29231_c0_seq1	384	A	G	A	9/9	22/22

3522.	comp140486_c0_seq1	434	A	A	T	6/7	4/6
3523.	comp18920_c0_seq1	570	C	T	C	19/27	28/28
3524.	comp33578_c0_seq2	933	G	G	A	6/9	12/15
3525.	comp33578_c0_seq2	935	G	G	T	7/10	12/15
3526.	comp33578_c0_seq2	1219	G	G	C	12/12	10/10
3527.	comp31462_c0_seq4	1826	G	T	G	7/7	12/12
3528.	comp32205_c0_seq1	4	T	C	T	7/8	4/4
3529.	comp33065_c0_seq1	548	C	C	T	8/12	22/30
3530.	comp27358_c0_seq1	161	C	T	C	18/19	36/36
3531.	comp21962_c0_seq1	207	G	G	C	14/16	10/10
3532.	comp21962_c0_seq1	314	T	T	G	6/7	14/16
3533.	comp26000_c0_seq1	278	A	A	G	6/6	6/6
3534.	comp26000_c0_seq1	537	G	G	T	32/33	24/26
3535.	comp32699_c0_seq1	1308	G	A	G	13/13	14/14
3536.	comp29194_c0_seq2	82	C	C	T	31/40	2/2
3537.	comp29194_c0_seq2	243	C	C	T	33/39	2/2
3538.	comp32408_c0_seq1	531	A	G	A	16/23	18/24
3539.	comp32408_c0_seq1	900	A	G	A	5/7	6/6
3540.	comp32326_c0_seq2	2062	T	C	T	8/8	12/14
3541.	comp29286_c0_seq1	590	A	G	A	12/12	14/14
3542.	comp31023_c0_seq3	1506	C	C	T	11/12	8/12
3543.	comp32751_c0_seq1	1534	G	G	A	15/15	4/5
3544.	comp32751_c0_seq1	1540	T	T	C	13/13	4/5
3545.	comp29879_c0_seq3	1375	G	G	A	27/27	7/7
3546.	comp33595_c0_seq26	108	T	T	C	7/8	6/6
3547.	comp33595_c0_seq26	124	G	G	A	9/13	12/14
3548.	comp15677_c0_seq1	105	G	A	G	12/18	10/10
3549.	comp32063_c0_seq1	1049	A	A	C	6/6	2/2
3550.	comp24594_c0_seq1	2017	G	A	G	6/9	18/18
3551.	comp32170_c0_seq1	1496	T	C	T	32/42	16/20
3552.	comp21420_c0_seq1	165	A	A	G	4/6	2/2
3553.	comp33555_c0_seq26	577	A	G	A	4/6	7/8
3554.	comp28136_c0_seq1	218	T	C	T	4/6	34/34
3555.	comp31359_c0_seq2	487	T	G	T	20/26	10/14
3556.	comp33659_c0_seq9	77	G	A	G	7/7	9/11
3557.	comp26780_c0_seq1	178	T	T	C	6/7	8/12
3558.	comp26780_c0_seq1	181	A	A	G	6/7	8/12
3559.	comp26780_c0_seq1	213	T	T	C	12/15	4/6
3560.	comp26780_c0_seq1	214	G	G	A	14/18	4/6
3561.	comp32296_c0_seq1	403	G	A	G	10/13	14/18
3562.	comp32296_c0_seq1	412	G	A	G	10/13	12/16
3563.	comp140696_c0_seq1	3	G	C	A	6/7	2/2
3564.	comp33625_c0_seq1	622	T	T	A	7/7	4/6
3565.	comp33625_c0_seq1	623	T	T	G	7/7	4/6
3566.	comp33625_c0_seq1	717	T	T	A	8/12	8/12
3567.	comp33625_c0_seq1	723	A	A	G	8/11	10/14
3568.	comp30440_c0_seq1	1800	G	G	A	8/8	6/6
3569.	comp19224_c0_seq1	6	T	T	C	8/8	4/6
3570.	comp31768_c0_seq1	755	T	C	T	4/6	8/8
3571.	comp31556_c0_seq2	345	A	A	T	7/9	12/18
3572.	comp56679_c0_seq1	1	G	G	T	10/13	2/2
3573.	comp30708_c0_seq3	853	G	G	A	6/9	16/18
3574.	comp30708_c0_seq3	856	C	C	A	8/11	14/18
3575.	comp30708_c0_seq3	859	G	G	T	9/12	12/16
3576.	comp30796_c0_seq3	51	T	T	C	8/8	4/4
3577.	comp30796_c0_seq3	54	T	T	C	8/8	4/4
3578.	comp30796_c0_seq3	59	T	T	G	8/8	4/4
3579.	comp30796_c0_seq3	287	A	A	G	12/16	5/7
3580.	comp168757_c0_seq1	163	C	T	C	4/6	10/10
3581.	comp33606_c0_seq1	78	T	A	T	6/9	17/17
3582.	comp24141_c0_seq2	181	C	C	T	14/20	13/17
3583.	comp24141_c0_seq2	184	C	C	G	14/20	17/21
3584.	comp14812_c0_seq1	674	C	T	C	4/6	19/19
3585.	comp218154_c0_seq1	206	A	T	A	6/6	10/10
3586.	comp23509_c0_seq1	89	C	T	C	6/6	8/8
3587.	comp2830_c0_seq1	254	G	G	A	11/12	6/8
3588.	comp6282_c0_seq2	186	T	C	T	9/9	4/6
3589.	comp6282_c0_seq2	190	T	A	T	9/9	4/6
3590.	comp33431_c0_seq1	884	T	A	T	5/6	16/20
3591.	comp3436_c0_seq1	81	G	G	A	11/11	4/4
3592.	comp3436_c0_seq1	158	C	A	C	9/13	6/6
3593.	comp3436_c0_seq1	172	T	C	T	7/10	6/6
3594.	comp3436_c0_seq1	186	T	C	T	7/10	6/6