Comparative study of capsaicinoid complex and transcriptomes between different *Capsicum* species and their hybrids

Thesis submitted to Jawaharlal Nehru University
for the partial fulfillment of the award of

DOCTOR OF PHILOSOPHY

Sushil Satish Chhapekar

SCHOOL OF LIFE SCIENCES
JAWAHARLAL NEHRU UNIVERSITY
NEW DELHI 110067

2017

School of Life Sciences
Jawaharlal Nehru University
New Delhi – 110067
India

CERTIFICATE

This is to certify that this thesis entitled "**Comparative study of capsaicinoid complex and transcriptomes between different *Capsicum* species and their hybrids**" submitted to the Jawaharlal Nehru University, New Delhi, by **Mr. Sushil Satish Chhapekar** is based on the studies carried out in School of Life Sciences, Jawaharlal Nehru University, New Delhi. This work is original and has not been submitted so far, in part or in full, for any degree or diploma in this or any other university or institute.

Sushil Satish Chhapekar
(Candidate)

for **Prof. Shyamal K Goswami**
(Dean)

Dr. Nirala Ramchiary
(Supervisor)

Dedicated to

Aai & Baba

List of Abbreviations

%	Percent
℃	Degree Celsius
Acc	Accession
A. thaliana	*Arabidopsis thaliana*
AFLP	Amplified fragment length polymorphism
bp	base pair
cDNA	Complementary DNA
C. annuum	*Capsicum annuum*
C. chinense	*Capsicum chinense*
C. frutescence	*Capsicum frutescence*
cm	Centimeter
cM	Centimorgan
CTAB	hexadecyl trimethyl ammonium bromide
dNTP	deoxynucleotide tri-phosphate
DNA	Deoxyribonucleic acid
DPA	Days after anthesis
DW	Distilled water
EtBr	Ethidium bromide
F1	first filial generation
F2	second filial generation
G	Gram
µg	Microgram
GC-MS	Gas Chromatography coupled with Mass Spectrometry
HPLC	High performance liquid chromatography

Kb	kilo base
µl	microliter
µM	micromolar
mM	milimolar
mins	minutes
mg	miligram
ng	nanogram
no.	number
PCR	Polymerase Chain Reaction
qRT-PCR	quantitative Real-Time Polymerase Chain Reaction
Pre-miRNA	precursor miRNA
Pri-miRNA	primary RNA
QTL	quantitative trait locus (loci)
RFLP	restriction fragment length polymorphism
RISC	RNA induced silencing complex
RNA	ribonucleic acid
RT	room temperature
Sec	second
SD	standard deviation
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
Tm	melting temperature
U	units
UTR	untranslated region

Acknowledgment

The success of this research depends largely on the encouragement of many others apart from the efforts from my side. I take this opportunity to appreciate those who have been instrumental directly or indirectly in the successful completion of my doctoral program.

First and foremost, with a deep sense of gratitude and immense respect, I thank my guide Dr. Nirala Ramchiary. He has guided me with his invaluable suggestions, lightened up the way in my darkest times and encouraged me a lot in the academic life. His benevolent attitude, scholarly guidance, constructive criticism and encouragement at every step helped me a lot while on work and enabled me to perform fruitfully. He left no stone unturned for making the facilities easily accessible.

I feel highly obliged in expressing my deep sense of gratitude to the Dean & Ex-dean, School of Life Sciences, Prof. S.K. Goswami, Prof B.C. Tripathi, Prof. B.N. Mallick, Prof. N.B. Sarin conceding me to avail the best of facilities at the elite school of the University. I want to thank University grant commission (UGC) for providing me Non-NET research fellowship during my Ph. D.

My heartfelt gratitude and thanks go out to especially to Dr. Rashmi Gaur for cooperating me in experimental and innumerable other ways. She was always there to solve my research & personal problems. My special appreciation goes to my friends Sarpras, Ajay and Vineet for all their scientific help and support throughout this study. I sincerely admire the contribution of all my lab mates Ajay kumar, Meenakshi (for special lunch), Abdul, Nitin, Dr. Vandana, Dr. Amita, Dr. Muhsin, Mukesh Nitin for extending their unstinted support, sympathetic attitude and unfailing help during study especially in thesis writing. Special thanks to Dr. Vandana, Abdul, Nitin, Ilyas & Meenakshi for critical inputs & help in thesis work. The help of lab staff Naveen, Vijender, RamkisanJi & VinodJi is deeply appreciated.

Besides, I would also like to convey my gratitude all the respected teachers and administrative staff of SLS for their kind help and support throughout my research period in SLS. I would like to thank all the CIF members for helping me in my experiments.

I would like to thank my friends Suyog, Pritam, Amol, Sachin, Pranav, Pavan and Vivek sir for their help in various ways. I do not know how to extend my gratitude towards you people. You guys were always there with me through the toughest moments, you saw my laugh in happy time and you saw me cry in my misery. Your friendly advices, your soothing words and your big hearts helped me face all the obstacles and continue with my work. I will never forget your kindness.

Apart from that, I am head over heels indebted to my sister, Sayli who is the one with me throughout my research work with her intense warmth, caring nature, affection and support with a view to making a ship fast to shore of success for me. Words prove a meager media to write down my feelings for Deven, Mama, Mami, Mavshi, Atya, Kaku, Dada & Vahini and all my relatives for providing me constant encouragement, divine presence and supporting me spiritually throughout.

Last but not least, I sought inspiration and owe a great deal to my mother Smt. Nisha and my father Late Shri Satish Chhapekar who have given unquestionable given their ears for the primrose path in my life. They formed part of my vision and taught me the good things that really matter in life. The happy memory of my father still provides persistent inspiration for my journey in this life. This thesis is dedicated to my beloved father, and my mother. It is your shining example that I try to emulate in all that I do. The successful completion of my doctoral program is an answer to their prayers and a fruit of their blessings.

I am not left with enough and appropriate words to thank Sonam. During the period of my trial and tribulation and moments of self-doubt she acted as harbor in the tempest. She has always provided optimistic company, encouragement, sound advice, and lots of excellent initiatives throughout my research period especially during the thesis-writing. I would have been lost without her.

Finally, unfailingly I pray and thank to the Aai Jagdamba & Varad Ganesha for his shower of blessings and giving me the strength to do my duty to the pinnacle satisfaction of my superiors. God, it is under your compassion that I stride to live, learn and flourish.

Sushil

CONTENTS

List of Figures

List of Tables

CHAPTER - I

INTRODUCTION

Introduction

Chilli peppers or *Capsicum* (2x = 2n = 12) belonging to the family Solanaceae plays an important role in global agriculture and horticulture as they are grown both for vegetables and spices. The genus *Capsicum* shows wide diversity within and between 38 species (USDA-ARS, 2011), and among them only *Capsicum annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*, *C. pubescens*, and *C. assamicum* are cultivated (Mega and Todd, 1975, Ramchiary et al., 2014). In India, chilli crop was introduced from Brazil in around 1584 by the Portuguese (Post harvest profile of chilli, Govt. of India, 2009). Several studies reported the presence of wide variety of beneficial metabolites such as carotenoids (provitamin A), vitamins (C and E), flavonoids, capsaicinoids, etc. in the fruits (Anandakumar et al., 2013); of which, the pungency (heat) content is the unique property of chilli peppers. These bioactive compounds in *Capsicum* fruits are also used in traditional medicines. The chilli pepper fruits also contain a wide variety of color due to the variation in carotenoids and pigments; which are also used as a coloring agent in food (Ramchiary et al., 2014 and Review of Literature Chapter).

Different land races and traditional cultivars of chilli-peppers have evolved and adapted in various parts of India i.e. Bhut jolokia (*Capsicum chinense*), and Bird eye chilli (*C. frutescens*) in North East India, pungent round chilli in Sikkim, Guntur chilli in Andhra Pradesh, etc. Bhut jolokia (*C. chinense*), reported as the naturally occurring highest pungent chilli pepper (Guinness Book of World Records, 2006) is widely used as spice crop. It is also used in folk remedies for dropsy, toothache, diarrhea, colic, asthma, muscle cramps, arthritis and headache by the native people of North East India (Meghvansi et al., 2010). Although, conventional breeding could develop desirable chilli varieties, such as Guntur chilli (Andhra Pradesh), Mundu chilli (Tamil nadu and Andhra Pradesh), Jwala chilli (Gujrat), Kanthari chilli (Kerala), Kashmiri chilli (Jammu & Kashmir), Byadagi chilli (Karnataka), Lavangi chilli (Maharashtra); however, progress has been slow, and there is a need to utilise molecular tools and techniques for development of desired varieties through precision breeding.

Of the many economically important traits in *Capsicum*, pungency and carotenoids contents, fruit traits (number, size and shape), and resistance to diseases are highly studied, mostly in *C. annuum* (Reviewed in Ramchiary et al., 2014). However, Bhut jolokia and *C. frutescens* from North East India is reported to be the highest and medium pungent containing *Capsicum* species, respectively, till now are not subjected to detail systematic study at both

phenotypic and genotypic level. The pungency property of *Capsicum* fruits, which is due to the presence of alkaloids, known as capsaicinoid complex, remains a core interest to researchers as till date all the genes involved in regulation of capsaicinoid biosynthesis could not be identified. Furthermore, Bhut jolokia or Ghost chilli, being considered as mysterious chilli pepper due to the presence of fiery hot pungency property in fruits, could be an important model system for studying the capsaicinoid biosynthesis. Till date, variations in expression level and promoter region of *Pun1* and *pAMT* genes are shown to be responsible for pungent and non-pungent phenotypes of the *Capsicum* fruit (Stuwart et al., 2005; Lang et al., 2009). Therefore, identification of more structural genes of this pathway and factors controlling the capsaicinoid content will help to decode the molecular basis of pungency development.

Apart from pungency content, *C. chinense* and *C. frutescens* are having important agronomic traits in a desirable form. These traits could be transferred to elite germplasm to ultimately enhance crop production. The discovery of molecular markers associated with important agro-economical traits is a prerequisite for breeding programme in plants. Although several molecular markers have been developed in *C. annuum* (Ramchiary et al., 2014), the development of species specific molecular markers in both species would help in accelerating breeding programme. Among the markers develop, simple sequence repeats (SSR) marker being highly polymorphic, easy for genotyping and cost effective, are highly preferred, and several SSR markers have been developed in *Capsicum* species mostly in *C. annuum* (Ramchiary et al., 2014; Chen et al., 2016). However, no such study reported in Bhut jolokia/Ghost chilli and very few in *C. frutescens*. Furthermore, the diversity analysis of crop germplasm including *Capsicum* have been widely done to identify genetically distinct genotypes/varieties so that those genetic resources could be used in successful breeding programme, since the more genetically diverse, more genetic gain in breeding programme have been (Rai et al., 2013, Reviewed in the following Chapter). The genetic diversity analysis will assist to identify the genetically distinct germplasms of *C. chinense* (Bhut jolokia) and *C. frutescens* which could later be used for breeding important traits like disease resistance, elevated pungency, high yield, etc.

Enhancement of yield potential of a crop is the ultimate aim of the breeder. With decreasing percentage of agriculture land throughout the world, managing crop production for billions of people is at serious risk. Heterosis breeding is one of the quick alternatives to overcome this, which has been used to gain an extraordinary stable increase in crop yield for

more than 90 years (Duvick, 2005). Generally, the phenomenon of enhanced biomass and fertility, high development rate and stress tolerance than either of the parents in hybrid is known as hybrid vigour or heterosis, and has been exploited to produce higher yields in numerous crop species (Birchler, 2016). The conventional genetic model explains heterosis is regulated by dominance and/or over-dominance gene expression (Lippman and Zamir, 2007). However, in rice, the phenomenon of epistasis is reported to have played a major role in heterosis than dominance and over-dominance (Yu et al., 1997, 2002). Brichler et al., (2010) suggested that effect of all three, epistasis, dominance and over-dominance on heterosis. However, till now, in chilli peppers, no such study has been performed. The phenotypic, physiological, and molecular manifestation of heterosis offers interesting and promising ways to enhance crop yield potential including in *Capsicum*.

Crop yield is a multigenic trait and influenced by the action of both coding and non-coding genes in the genome. The non-coding RNAs (ncRNA) are identified and characterised as key regulatory elements controlling diverse developmental processes in plants (Bartel et al., 2003). Among ncRNA, micro-RNAs (miRNA) are well known for inhibition of gene expression at a transcriptional and post-transcriptional level (Chen et al., 2009) thereby regulating target gene expression. The miRNAs play a vital role in diverse developmental processes such as leaf morphogenesis, flower and root development, the transition of vegetative phase to reproductive, pathogen invasion, hormone signalling pathways, response to stress and regulation of own biogenesis (Rubio-Somoza and Weigel, 2011). Various studies have been carried out in diverse plant species indicating its crucial role in plant development including Solanaceae plants (Reviewed in Djami-Tchatchou et al., 2017). In *Capsicum* species, discovery and profiling of few miRNAs are reported which is limited to *C. annuum* (Hwang et al., 2013; Liu et al., 2017), thereby suggesting the need of comprehensive profiling and identification of miRNAs from different tissues to understand their role in regulation of genes expression governing economically important traits in *Capsicum* species including Bhut jolokia.

Therefore, the present study is designed to dissect molecular mechanism of capsaicinoid biosynthesis pathway, develop molecular markers, understand the molecular basis of heterosis and miRNA based gene regulation in Bhut jolokia/Ghost chilli (*C. chinense*) and *C. frutescens*, as none of these studies has been done in these two *Capsicum*species using germplasm from the North East India. The study is divided into the following four objectives :-

1. Comparative study of genes involved in capsaicinoid biosynthesis pathway in different *Capsicum* species

2. Development of genome-wide molecular markers in *C. chinense* and *C. frutescens* based on transcriptome data

3. Physiological and transcriptomic study of heterotic intra-specific and inter-specific *Capsicum* hybrids

MicroRNA (miRNA) identification, target prediction and expression analysis.

CHAPTER - II

REVIEW OF LITERATURE

Review of literature

1. History,origin and cultivation of *Capsicum* species

Chilli pepper is considered to be first spice used by human and believed to be originated in the Central and South America (Bolivia), (Bosland 1996; Perry et al., 2007). The microfossil records from several archaeological (excavation) sites located in Bahamas to South America revealed that chili pepper was domesticated around 6,000 years ago, and it is one of the first cultivated crops in the Central and South Americas (Perry et al., 2007). Historical data suggests that in 1584 the Portuguese introduced the chilli crop to the Indian continent from Brazil. By the early 17th century, chilli had spread throughout the many parts of the world (Govindrajan and Salzer, 1985). The *Capsicum* is having rich diversity, constitutes of about 38 species (USDA-ARS, 2011) of which 6 species, namely *Capsicum annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*, *C. pubescens*, and *C. assamicum* are cultivated (Ramchiary et al., 2014). A combination of several data from archaeology, hereditary analyses, and contemporary plant distributions indicated that *C. annuum* was primarily domesticated in the regions of Mexico or near to northern Central America, *C. chinense* in Amazonia, *C. frutescens* in the Caribbean region, *C. pubescens* and *C. baccatum* in the southern Andes (Bolivia and peru); (Eshbaugh, 1993;. Pickersgill, 1984). Among them recently, the *C. assamicum* has been identified as a distinct domesticated species in the North Eastern part of India (Purukayastha et al., 2012); which is closely related to *C. frutescens* and *C. chinense* but can be differentiated due to its unique morphological characteristics.

2. Area, yield and production of chilli pepper

Chilli peppers are cultivated in many parts of the world. In 2014, world's annual production was 3.8 million tons for dry chilli and 32.3 million tons for fresh green chilli (FAOSTAT, 2016). At present, India, China, Bangladesh, Peru and Ethiopia are the leading *Capsicum* producing countries in the world (Fig. 2.1). India is the largest *Capsicum* producer, exporter and consumer country contributing to 36% of the global *Capsicum* production with value of 287 million Indian rupees (FAOSTAT, 2016).In India, the average productivity of chilli pepper is 2010 kg per hectare. *Capsicum* is grown in all the states of India, led by Gujarat, Rajasthan, Andhra Pradesh and followed by Telungana, Karnataka and Maharashtra (Table 2.1; Spice board, Govt of India, 2016).

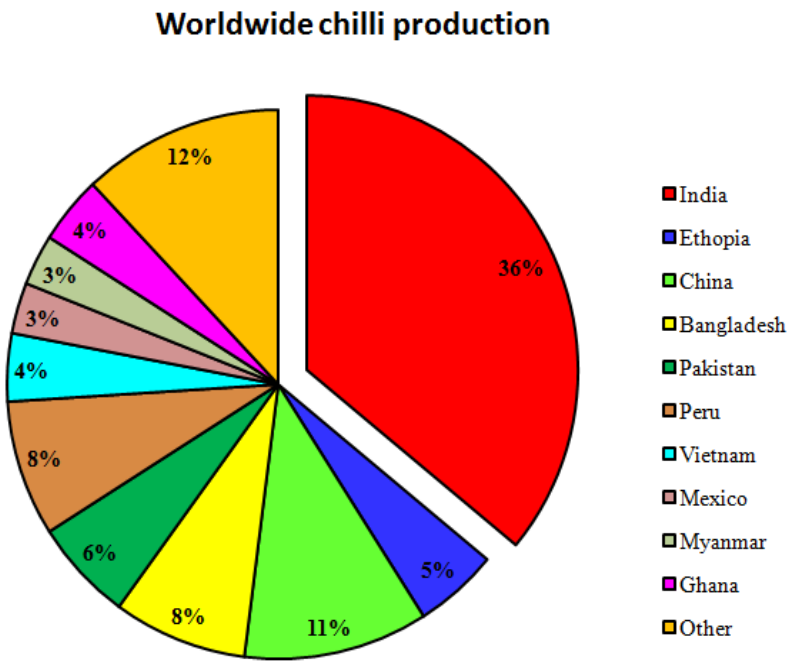


Fig 2.1 Summary of global production of Chilli peppers(FAOSTAT, 2016)

Table 2.1 Total harvested area and the average production of 10 major chilli pepper growing states in India for 2015-16.

States	Area (in Ha)	Production (in tons)
Gujarat	508750	1077560
Rajasthan	1014540	1056170
Andhra Pradesh	212730	920809
Telengana	128870	443400
Karnataka	225560	401110
Maharashtra	41820	371710
West Bengal	118170	329980
Uttar Pradesh	58590	217670
Tamil Nadu	113200	202530
Orissa	123320	181500

Source: Spice Board, Government of India, 2016.

3. Nutritional importance of chilli pepper

Chilli pepper is both consumed as fresh green vegetables, in the form of dry powder (as a spice) and sauce in diet. It is rich in vitamins, antioxidants and several vital minerals. The nutritional values of green and dry chilli pepper are mentioned in Table 2.2

Table 2.2 Nutritional values of chillies.

		Value (Per 100 gm)
Parameters	**Dry chillies (in gm)**	**Green chillies (in gm)**
Moisture	10	85.7
Protein	15	2.9
Fat	6.2	0.6
Minerals	6.1	1
Fiber	30.2	6.8
Carbohydrates	31.6	3
Calcium	160	30
Phosphorus	370	80
Iron	2.3	4.4
Vitamins		
Carotene	0.345	0.175
Thiamine	0.93	0.19
Riboflavin	0.43	0.39
Niacin	9.5	0.9
Vitamin C	50	111
Minerals & Trace elements		
Sodium	14	nd
Potassium	530	nd
Phytin Phosphorus	71	7
Magnesium	nd	272
Copper	nd	1.4
Manganese	nd	1.38
Molybdenum	nd	0.07
Zinc	nd	1.78
Chromium	nd	0.04
Oxalic acid	nd	67
Caloric values		
Dry chilli	297 Kcal	
Green chilli	229 Kcal	

Source – National Institute of Nutrition, Hyderabad
Abbreviations: nd- not detected

4. Medicinal importance of *Capsicum*

Apart from the dietary spices, *Capsicum* plays a major functional role in medicinal field. According to ancient literature around the world, chilli peppers are involved into a variety of medicinal preparations. The pharmaceutical applications of *Capsicum*s are due the presence of their analgesic, anti-arthritic, and anti-oxidant properties. The fruits is used against skin disease, anti-fever, anti-hypersensitive and in cold coughs (Pieroni et al. 2004; Mesfi n et al. 2009; Pieroni and Quave 2005). The oil of *C. annum* cultivar is used in dog bite treatment and roots can cure snake bite (Samal and Dhyani 2006; Kadel and Jain 2008). Even blended with other herbs it may help in assisting memory, curing of eczema, asthma and other ailments of aging (Kavasch and Baar 1999). *C. frutescens* cultivar can treat hypertension, malaria, fever and vaginal infections; in addition, it is also used for treating impotence and sterility (Kamatenesi-Mugisha 2005; Coe 2008; Odugbemi et al. 2007). *C. chinense* used as a painkiller and works in almost every type of ache like tooth ache, stomach ache and muscles pain. It works for respiratory and pulmonary disorders and as a hallucinogenic "hangover cure" (Bhagowati and Changkija 2009; Williams 1995; Coe and Anderson 1997). Moreover, some studies suggest that chilli peppers works as an anti-cancer agents and are having chemopreventive and chemotherapeutic effects (Meghvansi et al., 2010).

5. Mysterious element of *Capsicum* - Pungency (Capsaicinoids)

The pungency trait, the unique property of *Capsicum*, which distinguishes it from other Solanaceae plants, is considered to be one among the most economically valuable quality traits. Capsaicinoids start accumulating at around 20 to 30 days after pollination and last till fruit maturation (Iwai et al., 1979; Stewart et al., 2005). In *Capsicum* fruits, over 22 different capsaicinoids are known to be synthesized and are found to accumulate in placental epidermal cells, where they are secreted towards the outer cell wall, and fi nally accumulate within structures named ''blisters'' located on the placenta surface of fruits (Bosland and Walker, 2010). The major proportion of capsaicinoids is capsaicin (trans-8-methyl-N-vanillyl-6-nonenamdie)and dihydrocapsaicin (8-methyl-N-vanillylononanamide) (comprised of 77 to 98% of total concentration) [(Table 2.3) (Govindrajan et al., 1987; Zewdie et al., 2001)]. The other capsaicinoid components like

nordihydrocapsaicin, homocapsaicin, nonivamide and homodihydrocapsaicin are detected in small fractions (Huang et al., 2013).

Capsaicin biosynthesis is a synchronization of two pathways which are:- (i) fatty acid metabolism, which provides precursors valine/leucine for 8-methyl-6 nonenol-CoA moiety and (ii) phenylpropanoid pathway, which provides precursor phenylalanine for vanillylamine (Fig.2.2). The final condensation of vanillylamine and enol-CoA leads to formation of capasaicin (Ochoa-Alejo and Gomez-Peralta 1993; Kim et al., 2014). Various studies have been performed to identify the key genes of this pathway but till now the molecular basis of pungency remain elusive. Initially, the isolation of *Pal*, *Ca4h*, and *Comt* were done from a cDNA library of placenta tissue of pungent Habanero (*C. chinense*) by Curry et al. (1999) to dissect the capsaicinoid biosynthesis pathway. It was observed that these genes were co-related with pungency character. Later it was found that, *Kas* encodes 3-ketoacyl-ACP synthase protein while, *pAmt* codes for putative aminotransferase and these two are expressed only in placental tissues. Kim et al. (2001) constructed a subtractive cDNA library from placental tissues of *C. chinense* cv. Habanero. They observed two clones namely, SB2-149 and SB1-158 that displayed a great resemblance to the *pAmt* and *Kas* genes which confirms their involvement in capsaicinoid pathway. Later, Aluru et al. (2003) identified and mapped three placental-specific genes *Acl*, *Fat* and *Kas,* and found to bepositively regulated with pungency trait. According to Stewart et al. (2005), the SB2-66 clone is co-segregated with the pungent nature of chili and mapped it in the vicinity of *Pun 1* (locus C), which is responsible for pungency level (Blum et al. 2002). The allele was named *pun1,* and it showed significant similarity with the acetyltransferase of the superfamily BAHD and named as *AT3*.

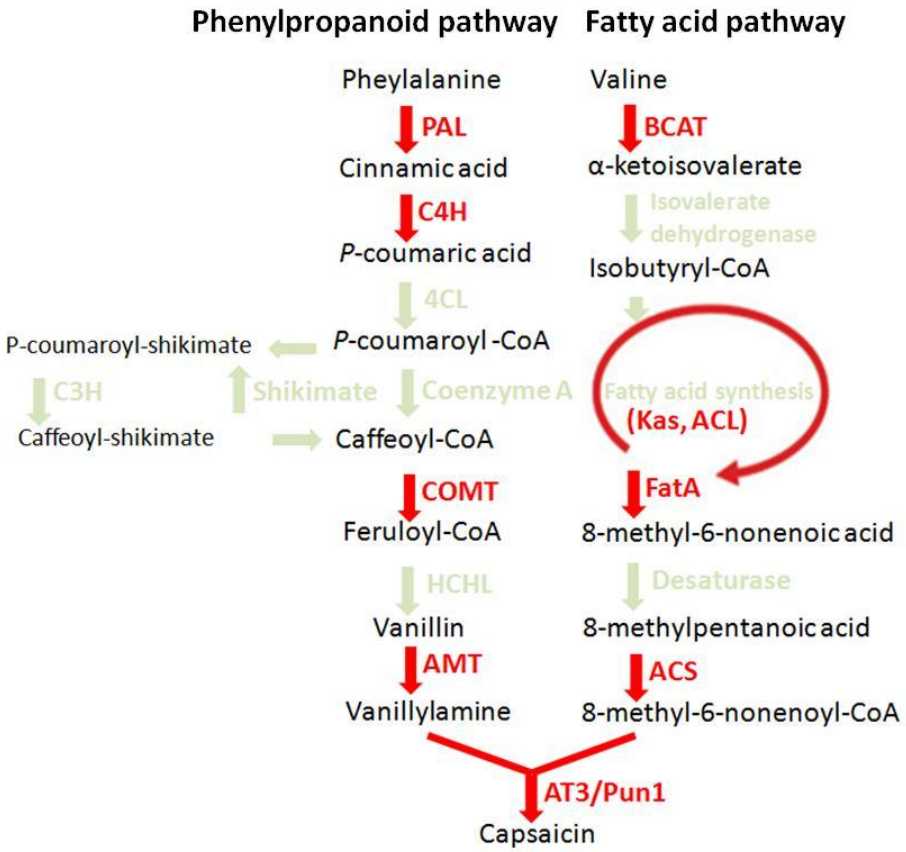
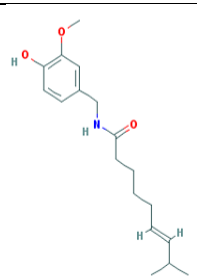
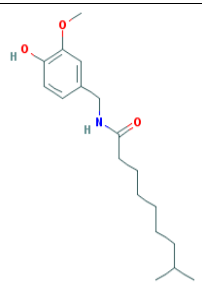
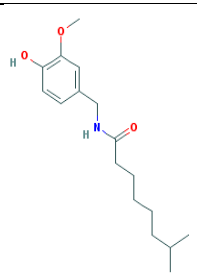
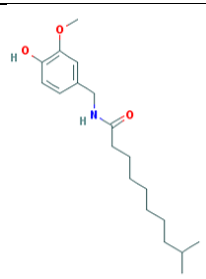
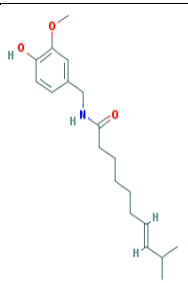
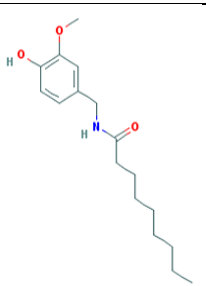


Figure 2.2 Capsaicinoid biosynthesis pathway (Adapted from Kim et al., 2014)

Table 2.3 Proportion of different capsaicinoid compounds in *Capsicum* fruits

Capsaicinoid name	Capsaicin	Dihydrocapsaicin	Nordihydrocapsaicin	Homodihydrocapsaicin	Homocapsaicin	Nonivamide
Abbrevation	C	DHC	NDHC	HDHC	HC	PAVA
Relative concentration (%)	69	22	7	1	1	NA
Scoville heat unit	16,000,000	15,000,000	9,100,000	8,600,000	8,600,000	9,200,000
Molecular formula	$C_{18}H_{27}NO_3$	$C_{18}H_{29}NO_3$	$C_{17}H_{27}NO_3$	$C_{19}H_{31}NO_3$	$C_{19}H_{29}NO_3$	$C_{17}H_{27}NO_3$
Chemical structure						

*The capsaicinoid structures were adapted from Pub Chem.

Further, by virus-induced gene silencing (VIGS), it was proved that *At3* was related to capsaicinoid production. Stewart et al. (2007) analyzed the *At3* gene in a non-pungent *C. chinense* NMCA 30036 chilli pepper and observed a 4-bp deletion in the first exon of *At3* gene and this allele was named *pun*^{2}. The mutation in putative aminotransferase gene (p*AMT*) results into formation of capsinoid, a sweat analogue of capsaicinoid (Lang et al., 2009) which is found in non-pungent chilli pepper cultivar. This confirms crucial role of p*AMT* gene in capsaicinoid biosynthesis. Further, 'CapCyc' a *Capsicum*-specific database was developed in which collections of the scattered metabolic information of capsaicinoid biosynthesis pathway were done. More than 30 candidate genes that plays crucial role in capsaicinoid biosynthesis pathways were included in this model and their sub-cellular locations was further predicted (Mazourek et al., 2009; Aza-González et al., 2011). Liu et al. (2013b) using transcriptome sequencing in *C. frutescens* identifiedthree candidate genes involved in capsaicinoids biosynthesis pathway. These three genes are Dihydroxyacid dehydratase (DHAD), Thr deaminase (TD) and Prephenate aminotransferase (PAT). The whole genome sequencing along with comparative analysis of pungency in hot pepper revealed that most of the capsaicinoid biosynthesis genes are specific to (with high expression) fruit development stages of pepper; however, the orthologous genes of tomato and potato showed very low expression. The results confirm the specificity of capsaicin biosynthesis pathway for development of pungent flavour in pepper fruit (Kim et al. 2014b and Qin et al. 2014). Recently, Reddy et al., (2014) performed an association mapping study in diverse *C. annuum* varieties. They identified SNPs in *Pun1*, *KAS*, *HCT*, *CCR*, genes and revealed *PUN1*, *CCR* and *KAS* act as important candidates in capsaicinoid production. Moreover, *Pun1* serves as the main controller of capsaicin pathway that is responsible for capsaicinoids and their precursor molecules. They also identified 6 SNPs in the upstream promoter region of *Pun1* and proposed that the capsaicinoids accumulation correlate with the degree of expression of *Pun1* gene. Tanaka et al., (2017) investigated that, increased expression of various capsaicinoid biosynthesis genes in pericarp tissue results into enhancement of total capsaicinoid concentration in *Capsicum* fruit. However, the comprehensive characterization of chilli pepper cultivars with variation in capsaicinoid content and study of molecular mechanism behind such variation has not been reported till date.

6. Development and application of microsatellite markers in *Capsicum* species

Molecular markers are important genetic tool for plant breeders to detect the genetic variation available in the germplasm collection. During the last two decade, a variety of molecular markers have been developed in large numbers crop plants. According to Gupta et al., (2002), molecular markers can be grouped in three main categories: (1) Hybridization-based markers: Restriction Fragment Length Polymorphism (RFLP), (2) PCR-based markers: Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP) and Microsatellite or Simple Sequence Repeat (SSR), and (3) sequence or chip-based markers: Single Nucleotide Polymorphism (SNP), Diversity Array Technology (DArTs) and Single Feature Polymorphism (SFP) markers. These markers have been utilized extensively for the construction of saturated molecular maps (genetic and physical) and their association with genes/QTLs controlling the traits of economic importance has been utilized marker assisted selection (MAS) (Ramchiary et al., 2014; Varshney et al., 2005b, 2006a).

Among the available markers, microsatellites of 2-6 nucleotides long tandem repeats are highly preferred because of their locus specificity, PCR-based, reliable, co-dominant, multi-allelic, and chromosome specific and highly informative genetic markers properties. Due to their abundance and inherent potential for variation, SSR markers have become a valuable source of genetic markers and are amenable to high-throughput genotyping. A variety of other names have been used for microsatellites such as simple sequence repeats (SSR) (Tautz et al., 1986). Historically, the term microsatellite has been applied solely to repeats of dinucleotide motif CA (GT) (Litt & Luty, 1989; Weber & May, 1989). However, they are mainly classified as perfect, imperfect and compound according to the type of microsatellite present and termed as di-, tri-, tetra-, penta-, hexa- etc. with respect to the number of SSR motif arranged on the nucleotide chain (Alghanim and Almirall et al., 2003). SSRs or microsatellites are suitable for construction of high-density linkage maps, QTL mapping, gene cloning, germplasm diversity analysis, cultivar identification, and marker-assisted selection. Moreover, they are useful for integrating the genetic, physical, and sequence-based maps which simultaneously provide breeders and geneticists an efficient tool to link phenotypic information at genome level.

Microsatellites are observed in coding and non coding regions of the genome; however, the coding region has low density because a high mutation rate (in coding region) would lead

loss of function (Oliviera et al., 2006). Although, SSRs derived from coding regions are lesser in number in comparison to non-coding regions, but it is sufficiently high numbers to permit development of a large number of SSR markers. About 13.6% of genes in *Arabidopsis thaliana* have been found to contain SSR within the coding regions (Gemayel et al. 2010). Genic SSRs are beneficial over genomic SSR as they have the potential of serving as functional markers if they have been designed from candidate genes and have higher cross transferability across species as the coding regions tend to be more conserved across different species (Varshney et al., 2005). With the emergence of NGS technology, it is becoming increasingly popular to perform transcriptome sequencing to develop genic SSRs (Zalapa et al., 2012). Because of several advantages, numerous studies have reported development of genic SSR in different plant species like soybean (Li et al., 2010), *Brassica rapa* (Ramchiary et al., 2011), sweet potato (Wang et al., 2011), potato (Dutta et al., 2011), pineapple (Wohrmann and Weising 2011), sorghum (Reddy et al., 2012), faba bean (Gong et al.,2011; Rodeny et al., 2014), chickpea (Parida et al., 2015) etc. using sequence data available in various databases or developed by transcriptome sequencing.

Development of SSR markers have also been reported in *Capsicum* species (Tam et al., 2005; Kwon et al., 2005; Yi et al., 2006; Kong et al., 2011; Ashrafi et al., 2012; Liu et al., 2012; Ahn et al., 2013; Shirasawa et al., 2013; Shirasawa et al., 2014; Ibarra-Torres et al., 2015; Tsaballa et al., 2015; Chen et al., 2016). Since the development of high through put sequencing technology, numerous transcriptome data are available in public with the ongoing multinational efforts of transcriptome sequencing, providing new sources for developing fast and efficient gene based SSR markers in a rapid and cost effective manner as compared to the SSRs developed form the genomic sequences. In *Capsicum* species, Portis et al., (2007) designed 348 SSRs primer pairs from 576 non-redundant EST sequences. Yi et al., (2006) identified a total of 1,201 SSRs which corresponds to one SSR in every 3.8 kb of the ESTs. Huan-huan et al., (2011) developed 755 SSR markers of which 127 SSR primer pairs showed polymorphism in 8 populations of *Capsicum*. Recently, genome-wide characterization of SSRs from *C. annuum* was performed by Chen et al., (2016) which revealed distribution of microsatellites in nuclear, chloroplast and mitochondrial genome. Mostly, the SSR markers are developed in *C. annuum* and no genic SSRs have been reported in Bhut jolokia (*C. chinense*) and very limited in *C. frutescens* till date. Therefore, knowing the importance of genic SSRs in genetic mapping and QTL studies, development of genic SSR is highly

desirable in *C. chinense* and *C. frutescens*. For, the development and application of other markers i.e RAPD, RFLP, AFLP, SNP and COS and their application for genetic map construction and identification of QTLs/genes for agronomically important traits in *Capsicum* kindly see reviews written by Ramchiary et al 2014 and Chhapekar et al., 2016b.

SSR markers has been widely used for the analysis of genetic diversity in different plants such as, blackgram (Chaitieng et al., 2006), chickpea (Upadhyaya et al., 2009), rice (Lapitan et al., 2007), peanut (Naito et al., 2008), mungbean (Somta et al., 2009), pigeonpea (Odeny et al., 2009) and eggplant (Vilanova et al., 2012). In recent years, various diversity studies have been reported based on a variety of markers, but these were predominantly focused with *C. annuum*. Over the years, *Capsicum* genetic diversity has been analysed using restriction fragment length polymorphism (Lefebvre et al. 1993), direct amplification of minisatellite DNA (DAMD-PCR; Ince et al., 2009), amplified fragment length polymorphism (Aktas et al., 2009; Baba et al., 2015), random amplified polymorphic DNA (RAPD; Adetula 2006), SSRs (Portis et al., 2007; Stágel et al., 2009; Pacheco-Olvera et al., 2012; Rai et al., Carvalho et al., 2015).

7. Heterosis breeding in plants

A. Concept of heterosis and heterotic traits

The word ‘heterosis’ was originated from the Greek word ‘heteroiosis’ which means ‘different in kind’. The first scientist who systematically investigated the heterosis or hybrid vigour was Charles Darwin in 1876, when he observed cross-pollinated maize progeny was 25 % taller than inbred maize progeny (Darwin, 1876). In 1908, the term ‘heterosis’ was coined by Shull and East in independent study to explain the superiority of F_1 hybrids relative to parents in terms of vigour, size and yield (East, 1908; Shull, 1908). Later, this phenomenon was widely accepted and used by evolutionary biologists to incorporate heterosis for survival to be precise; adaptive, selective, and reproductive gain (Dobzhansky, 1950) or dominance of quantitative traits like yield (Griffin, 1953), development rate (Rao et al., 1992) and biomass (Liu et al., 2002). According to Swanson-Wagner et al., (2006), in the late 1990s, about 95% and 65% of the total maize area in the USA and in world was planted with hybrids. Heterosis could be detected in all known traits of plant and can be examined in embryo development (Meyer et al., 2007), mature plants (Hochholdinger and Hoecker, 2007) and seedling (Hoecker et al., 2007).

The percentage of heterosis for each trait in a hybrid varies greatly (Springer and Stupar, 2007) and for complex traits the gain in heterosis percentage has been usually observed (Becker, 1993). The amount of phenotypic difference of a trait in a F_1 hybrid relative to parents (P_1 and P_2) can be explained as mid-parent heterosis (MPH) or best-parent heterosis (BPH). MPH indicates that value of the particular trait for F_1 hybrids is significantly higher than the average value of the parents. BPH indicates trait value of the F_1 hybrids is significantly higher than the better parent. A higher amount of heterosis was detected in progeny re-produced by cross-pollination or cross-fertilization plants while lower heterosis in self fertilizing plant species (Barth et al., 2003).

B. Genetic basis of heterosis

Although heterosis phenomenon was rediscovered over a century ago and several genetic models were described underlying its mechanism, till now little consensus has been achieved about the molecular basis of heterosis (Lamkey and Edwards, 1998; Stuber, 1999; Birchler, 2006; Birchler, 2013).The most predominant and accepted quantitative genetic hypotheses to elucidate heterosis are 'dominance' and 'over-dominance'. Both of these hypotheses explain non-additive phenotypic behaviour as a result of genetic variations between divergent homozygous parents and their heterozygous hybrids. The dominance hypothesis describes heterosis by the action of superior dominant alleles from each parent at multiple loci that complement corresponding slightly deleterious (or unfavorable) alleles resulting in improved vigour of hybrids over the parents (Devenport, 1908; Keeble, 1910; Bruce, 1910; Jones, 1917) (Figure 3). The complementation due to hybridization may possibly lead to generation of such characteristics which are either equal to or superior than two parents.The over-dominance hypothesis explains heterosis by the action of allelic interactions at one or several loci in hybrids which lead to superior traits in F_1 plants over to the homozygous parents (Figure 3) (Shull, 1908; Powers, 1945). Other hypotheses to elucidate heterosis are also proposed such as pseudo-overdominance and epistatis. Pseudo-overdominance is genetically intermediate stage between dominance and over-dominance. It can be recognized as a case of dominance where, two recessive loci such as 'a' from P1 and 'b' from P2 are linked 'in repulsion' or '*in trans*' manner (Figure 3). Such case of complementation in hybrids relates to over-dominance phenomenon due to tight linkage. Moreover, the epistasis explains heterosis as the epistatic interactions of two

non-allelic genes at multiple loci as principal factor for the better phenotypic expression of a character in hybrids over parents (Semel, 2006). However, the relative contribution of each mechanism in heterosis remains ambiguous till date (Lippman and Zamir, 2006; Birchler et al., 2006; Hochholdinger and Hoecker, 2007; Schnable and Springer, 2013).

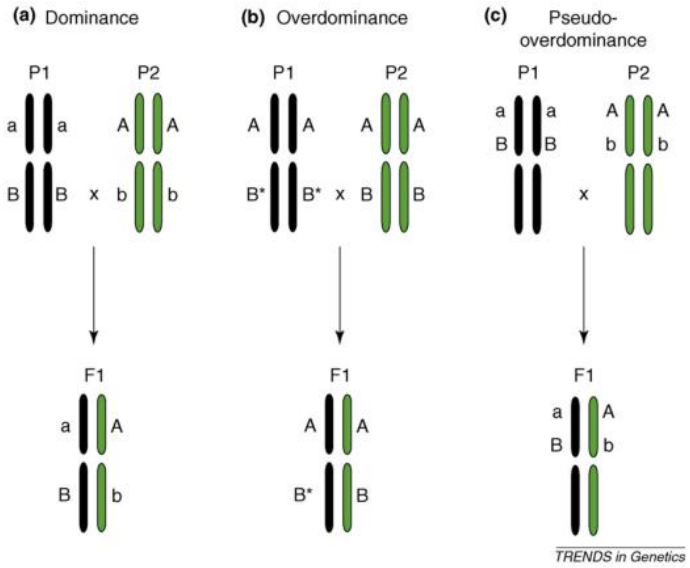


Figure 3. Genetic models of heterosis (Adapted from Hochholdinger and Hoecker, 2007)

Additionally, heterosis could be explained by molecular and/or physiological hypotheses such as:

1. Ashby (1937) indicated that heterosis is influenced with higher embryo weight and size due to increased cell number in tomato. While, in higher plants Srivastava (1983) observed a faster cell division rate leads to heterosis in higher plants
2. Heterosis may be an outcome of complex interactions of genetic and environmental impulse (Griffing and Zsiros, 1971; Griffing, 1990; Lippman and Zamir, 2006).
3. Heterosis might be regulated by parental genetic distance (Melchinger, 1999; Barbosa et al., 2003).
4. Heterosis could be result of epigenetic mechanism (Swanson-Wagner et al., 2006).
5. Heterosis may be a result of differential expression of allele-specific transcripts in F_1 hybrids (Swanson-Wagner et al., 2006).
6. Heterosis can be influenced by small RNAs. The small RNAs are known to regulate gene expression and epigenetic regulation leads to hybrid vigour (Ng et al., 2012).

C. Molecular tools and technologies to investigate heterosis

Over the years numerous genetics and functional genomics tools and methods have been developed which improved our understanding of fundamental plant processes, including heterosis. Recently, various studies reported the dissection of heterosis at the gene expression and genome organization level. The heterosis has been explained by various other approaches such as transcriptome, methylome, small RNA, and proteome studies. The Table 4 lists the heterosis related gene expression studies in plants.

Table 2.4. Heterosis related gene expression studies in plant species (Modified from Hochholdinger and Hoecker, 2007)

Plant organ	Developmental Stage	Approach	Genetic background	Global expression trends	References
Maize					
Endosperm	10, 14, 21 DAP	GeneCalling	7 Pioneer inbred lines	Nonadditivity	Guo et al., (2003)
Endosperm	18 DAP	RT-PCR	B73 and BSSS53		Song et al., (2003)
Embryo	6 DAP	cDNA microarray, SSH, qRT-PCR	UH005 and UH301	Additivity	Meyer et al,. (2007)
Embryo	19 DAP	Microarrays	Mo17	Additivity	Stupar et al., (2006)
Seedling	11 DAG		B73		
Adult leaves of di- and triploids	-	Quantitative Northern blotting	Mo17 and B73	Nonadditivity	Auger et al., (2005)
Immature ear		GeneCalling	17 Pioneer inbred lines	Additivity and Nonadditivity	Guo et al., (2006)
Seedling	14 DAG	cDNA microarray	Mo17	Additivity	Swanson-Wagner et al,. (2006)
Shoot apical meristem	21–23 DAP	cDNA microarray	UH002	Nonadditivity	Uzarowska et al., (2007)
		qRT-PCR	UH005,UH250 & UH301		
Arabidopsis					
First leaves	21, 24 DAG	cDNA microarray	Col, Ler and Cvi	Nonadditivity	Vuylsteke et al., (2005)
Seedling	15 DAG	DNA methylation	Ler and C24	Additivity and Nonadditivity	Shen et al., (2012)
Seedling	14 DAG	DNA methylation and small RNA analysis	C24 and Ler	Additivity in small RNA analysis and Nonadditivity in epigenome	Groszmanna et al., (2011)
Rice					
Panicle	Stage III, IV, V	cDNA microarray	Zhenshan97,Minghui63	Additivity	Huang et al., (2006)
Seedling	28 DAG	DNA methylation and transcriptome	Nipponbare and 93-11	Additivity and Nonadditivity	He et al., (2010)

Abbreviations: DAP, days after pollination; DAG, days after germination; SSH, suppressive subtractive hybridization.

8. MicroRNA in plant species

MicroRNA (miRNA) constitutes a large family of endogenous, small (19 to 25 nucleotides), single stranded, non-coding RNAs, which are broadly conserved across species. The first miRNA was discovered in 1993 by two different group (Lee et al., 1993; Wightman et al., 1993) while working on the control of development timing in a nematode worm, *Caenorhabditis elegans*. It gives the first insight into the identification and the mechanism of miRNA biogenesis. However, the mechanism of action of miRNA was defined in 1998, after five year of its discovery (Fire et al., 1998), where double-stranded RNA was found to be more effective than single stranded RNA to produce the interference in the expression of genes. In miRNA biogenesis, a RNA polymerase II transcribed MIR gene forms capped and poladenylated Pri-miRNA which further forms a stem-loop containing imperfectly folded structure known as primary miRNA (pre-miRNA). Later, Dicer-like 1 (DCL1) protein cleaves this pre-miRNA into precursor sequence with hairpin structure. This structure subsequently processed to form miRNA:mRNA* duplex by DCL1 protein, where after elimination of complementary sequence (miRNA*) yields single-stranded mature miRNA. Further, this mature miRNAs are integrated with RNA-induced silencing complex (RISC) leading to either target mRNA degradation and/or its translation inhibition (Jones-Rhoades et al., 2006). The term "miRNA" was coined for these small non-coding RNAs after their number started increasing exponentially during and after the year 2000.

The miRNA structure comprises of two parts: (i) Seed region; and (ii) 3' tail. Seed region is defined by 2^{nd} and 8^{th} nucleotide of miRNA sequence, starting from 5' end. The seed sequence shows perfect complimentary (Watson-Crick match) to target mRNA and nucleates the miRNA-mRNA interaction (Filipowicz et al., 2008; Lewis et al., 2003). Seed region defines the target mRNA repertoire and the functional identity of given miRNA (Krol et al., 2010). Nucleotides other than seed region towards 3’ end represent the 3' tail. This region shows variable complementarity to target mRNA and increase binding efficiency to target mRNA which can be expressed in terms of Gibbs free energy (ΔG). By predicting the hybridization pattern of miRNA and its cognate mRNA, the free energy (ΔG) release is calculated and if the interaction has lower free energy releases, higher is the interaction efficiency.

The miRNA regulates gene expression at post transcriptional level by directly binding to the 3'untranslated region (3' UTR), coding sequences, or 5' UTR of target messenger RNAs (mRNA), leading to inhibition protein expression by either translation inhibition or mRNA cleavage (Ambros 2004; Bartel, 2004; Filipowicz et al., 2008). Most of the miRNA genes are located in the intron of protein - coding genes; while they could also be embedded in exonic regions or exists as separate transcriptional unit (Jansson and Lund, 2012). More than 60% of mRNA are predicted to be target of miRNA (Bartel, 2009;Calin et al., 2004) hence miRNA-mediated regulation is the most abundant mechanism for post-transcriptional regulation of gene expression, with one exception where miRNA interacts with nascent miRNA to alter the chromatin of corresponding template DNA by affecting the methylation (Bao et al., 2004). A high level of redundancy for target gene increase the difficulty in understanding the role of a miRNA in given condition (Alvarez-Saavedra and Horvitiz, 2010); as the phenotypic outcome of deregulating even individual miRNA is unlikely to be mediated via a single target gene. The summary of miRNA mediated gene silencing has been illustrated in figure 4 (Rogers and Chen, 2013). Briefly, the pri-miRNAs are processed in the nucleus and mature

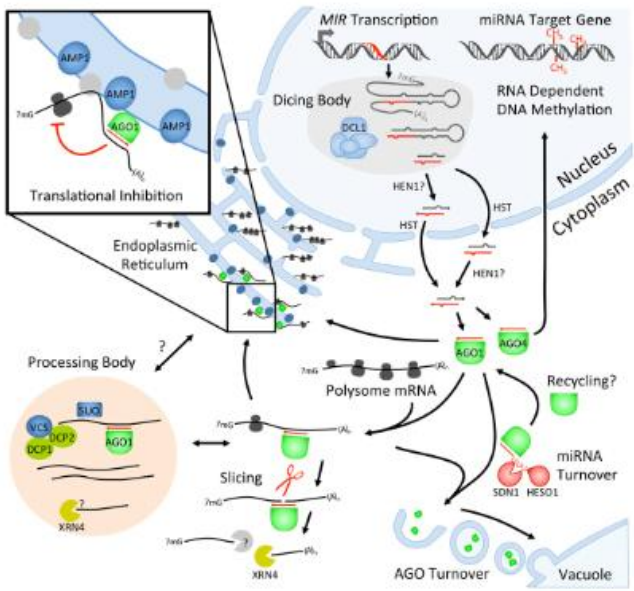


Figure 4. Overview of miRNA mediated gene silencing. Proteins with known function in miRNA target degradation pathway are coloured with yellow and non-AGO protein associated in translation inhibition are in blue color. The tentative molecular events, or site of the events, are designated by a question mark. (Adapted from Rogers and Chen, 2013).

miRNAs are exported to the cytoplasm. Further, miRNAs are integrated with AGO proteins and regulates post-transcriptional gene silencing through translational repression or slicing. The cytoplasmic locations of RISC pathway or miRNA target slicing are unidentified in plants, but a recent study indicated that translational repression occurs on the ER (Rogers and Chen, 2013). AGO import into the nucleus directs transcriptional silencing of target genes.

Several experimental studies in various plant species demonstrated the crucial involvement of miRNAs in a variety of biological and cellular processes such as in maintenance of genome architecture, hormone signaling pathways, signal transduction pathways, plants innate immunity, plant development and growth, abiotic and biotic stress response (Navarro et al., 2006; Sun, 2012; Zhang and Wang, 2015). The Table 5 summarises the functional role of numerous miRNAs in different plant species. However, very limited information is known about the role of small RNAs in *Capsicum* (Hwang et al., 2013; Liu et al., 2017). As of now small RNA studies were performed only in *C. annuum* (Hot pepper) in which, Hwang et al., (2013) identified and characterized 29 conserved and 35 novel miRNA families from ten different tissue libraries. Recently, about59 known miRNAs and 310 novel miRNAs were identified during fruit development in hot pepper (Liu et al., 2017). However, till now other cultivated species of *Capsicum* such as Bhut jolokia (*C. chinense*) and *C. frutescens* has been unexplored for miRNA profiling.

Table 2.5. Functional role of miRNAs in different plant species (Modified from Djami-Tchatchou et al., 2017)

Plant Family	miRNA	Target genes	Functional characterstics	References
Gramineae				
Rice (*Oryza sativa*)	miR393	Auxin receptor gene (TIR1 and AFB2)	Drought stress response	Zhou L. et al., (2010)
			Early flowering and high tillering	Jiao et al., (2010)
				Xia et al., (2012)
	miR397	L-ascorbate oxidase	Heat stress response and adaptation	Jeong et al., (2011)
	miR167	ARF transcription factors	Cold stress response	Jeong et al., (2011)
	miR820	DRM2	High temperature and salt stress	Sharma et al., (2015)a
Wheat (*Triticum aestivum*)	miR156d	Squamosa binding protein	Development, drought stress	Curaba et al., (2012)
	miR1435/miR51812		Ion transportation	
	miR395	ATP sulfurylase genes	Abiotic stress	Han et al., (2013)
	miR397, miR437	L-ascorbate oxidase	Plant development	
Maize (*Zea mays*)	miR156	Squamosa binding protein	Plant artichecture and development	Li et al., (2012)
	miR160	ARF transcription factors	Plant development	Gu et al., (2013)
	miR164	NAC1 (NAM, ATAF, CUC) - TFs	Ear and endosperm development	Ding et al., (2013)
	miR167	ARF transcription factors	Stress response	Sheng et al., (2015)
	miR169	NF-YA transcription factors	Drought stress reponse	Sheng et al., (2015)
	miR396	Growth factor		
Sugarcane (*Saccharum*)	miR156	SBP/SPL transcription factors	Plant development and stress response	Zanca et al., (2010)
	miR159	MYB protein	Plant development	
	miR164	NAC transcription factors	Drought stress	Ferreira et al., (2012)
	miR399	Inorganic pyrophosphatase 2		
	miR169	HAP12-CCAAT-box transcription factors	Salt stress response	Carnavale-Bottino et al., (2013)
	miR398	Serine/threonine kinase-like	Salt stress response and metabolism	
Barley (*Hordeum vulgare*)	miR396d	Growth factor	Cell differentiation & seed development	Shuzuo et al., (2012)
	miR399b	Phosphatase transporter	Drought stress	
	miR164	ARF transcription factors	Leaf and lateral root development	Deng et al., (2015)
Sorghum (*Sorghum bicolor*)	miR169	NFY	Plant development, drought response	Paterson et al., (2009)
	miR398	Selenium binding protein	Transportation	Du et al., (2010)
	miR170/171	GRAS domain transcription factors	Plant development	Zhang et al., (2011)
	miR156	SBP/SPL transcription factors	Development, increased biomass metabolism	Katiyar et al., (2012)
	miR397/398/408	Laccase	Response to Cu deficiency	
	miR399	UBC24 enzyme	Response to Phosphate deficiency	
	miR395	ATP, APS1 and Sultr1	Low Su response, development	
	miR396	Growth-regulating factor	Stress response, development	
Switchgrass (*Panicum virgatum*)	miR156	SPL, Cg1gene, MYB, HSP-binding	Drought stress, Biomass production,	Fu et al., (2012) Shen et al., (2012)
	miR167	ARF TFs, Glycosyl transferase-like protein	Biofuel recalcitrance and yield	Sun et al., (2012)
	miR172	AP2, SPL3	Plant development and stress response	
	miR159/319	MYB	Development, Biofuel yield	
	miR398	Fiber protein Fb2	Abiotic stress, Recalcitrance	
	miR396	Growth-regulating factor	Salt and Drought stress	Xie et al., (2014)

	miR397/408	Laccase		
Solanaceae				
Potato (*Solanum tuberosum*)	miR160	Auxin response factors	Plant growth and development	Din et al., (2014)
	miR172		Starch accumulation	
	miR473	Serine/threonine kinase-like	Plant metabolism	
	miR475	Thioredoxin	Plant metabolism	
Tomato (*Solanum lycopersicum*)	miR319	APETALA2	Leaf margins growth	Ori et al., (2007)
	miR169	NF-YA transcription factors	Drought response	Zhang X. et al., (2011)
	miR156/157	Colorless non-ripening	Fruit ripening	Xie et al., (2012)
	miR172	APETALA2	Leaf margin growth and development	Karlova et al., (2013)
	miR390	RNA-induced transcriptional silencing complex protein TAS3	Leaf morphology	
	miR167, miR169, miR172, miR393, miR397	ARF transcription factors, NF-YA transcription factors	Cold and drought stress response	Zhang X. et al., (2011), Koc et al., (2015)
Tobacco (*Nicotiana tabacum*)	miR160/167	ARF transcription factors	Stress response and development	Frazier et al., (2010)
	miR164	NAC transcription factors	Lateral root development	
	miR169	NFY	Drought response and development	
	miR171	GRAS domain transcription factors	Plant development and growth	
	miR172	AP2, SPL3	Plant development, Stress response	
	MiR319	MYB protein	Plant development	
	miR393	ARF and AFB	Plant development, Stress response	
	miR156	SPL	Plant development, Stress response	Guo et al., (2011)
	miR166	Leucine-rich (LRR) repeat family	Disease resistance	
	miR399	4-Coumarate-coenzyme A ligase	Stress response	Tang et al., (2012)
	miR408		Wounding and topping response	
Leguminosae				
Soybean (*Glycine max*)	miR156, miR160	Squamosa binding protein	Seed development	Song et al., (2011)
	miR164, miR166	ARF transcription factors		
	miR172, miR396	Growth factor		
Peanut (*Arachis hypogaea*)	miR159, miR171	Squamosa binding protein	Protein and lipid accumulation	Zhao et al., (2010)
	miR156	Squamosa binding protein	Growth and development	Chi et al., (2011)
	miR156, miR157	Lipid transfer protein,	Disease resistance	Zhao et al., (2015)
	miR159, miR396	Auxin response factors,		
Cowpea (*Vigna unguiculata*)	miR159, miR160, miR166, miR167	ARF transcription factors	Drought tolerance	Barrera-Figueroa et al., (2011)
	miR156b,f	Multicystatin gene	Protein degradationand and keep cellular proteins, drought response	Shui et al., (2013)
	miR169 miR319, miR390, miR393, miR396, miR403	Growth factor	Metabolic pathways of drought stress associated physiological changes	Shui et al., (2013)
Malvaceae				

Cotton (*Gossypium hirsutum)*	miR396	Callose synthase	Fiber development	Zhang et al., (2007)
	miR167a	ARF transcription factors	Salt stress response	Yin et al., (2012)
	miR395	APS1	Salt stress response	Wang et al., (2013)
	miR397a/b	Laccase		
	miR399a	UBC24 enzyme		
	miR156	SBP/SPL transcription factors	Plant development, Stress response	Wang and Zhang, (2015)
	miR172	AP2, SPL3	Phase transition,Flower development	
	miR319	MYB protein	Leaf development	
Convolvulaceae				
Sweet potato (*Ipomoea batatas*)	miR160, miR164, miR166, miR398	ARF, NAC1 transcription factors	Root development	Sun R. et al., (2015)
	miR156, miR162	Squamosa binding protein transcription factors	Storage root initiation and development	
	miR167	ARF transcription factors	Stamen development	
Rutaceae				
Orange (*Citrus sinensis*)	miR160	Auxin response factor 10	Stress response, Root development,	Song et al., (2012)
	miR165	Homeo-domain leucine zipper and HD-Zip protein	Plant development, Root absorption, Stress response	
	miR172	AP2	Plant growth, Stress response	
	miR393	TIR1, ARF, and AFB	Response of leaf to B- deficiency	Lu et al., (2014)
	miR408	Cu homeostasis, superoxide dismutase	Response to B-deficiency	Lu et al., (2015)
Euphorbiaceae				
Cassava (*Manihot esculenta*)	miR156, miR157, miR159, miR160	Transcription factors	Plant development and stress response	Patanun et al., (2013)
	miR414, miR473		Stress response	
	miR164	NAC transcription factors	Drought tolerance	
	miR172, miR319, miR395, miR396, miR397	Transcription factors, Growth factor	Starch biosynthesis and metabolism	Chen et al., (2015)
Vitaceae				
Grapevine (*Vitis vinifera*)	miR156	Squamosa binding protein	Fruit development	Pantaleo et al., (2010)
	miR160, miR167	Auxin response factor	Development, stress response	Wang et al., (2011)
	miR159, miR319	MYB transcriptionfactor	Transition of vegetative to reproductive phage, Stress response	Han et al., (2014)
	miR171, miR529	GRAS family transcription factors	Plant development, metabolism and photosynthesis	
	miR393, miR394	F-box	Stress response	
Rosaceae				
Apple (*Malus domestica*)	miR399	ARF transcription factors	Phosphate homeostasis, Long distance signaling,Shoot to root transport	Pant et al., (2008)
	miR156, miR159, miR166, miR167,miR172	Transcription factors	Plant growth and development, Stress response	Varkonyi-Gasicet et al., (2010)
	miR169a, miR160e miR167b,g,miR168a,b	ARF transcription factors	Fire blight resistance	Kaja et al., (2015)
Rubiaceae				

Coffee (*Coffea arabica* and *canephora*)	miR159e	Medium chain reductase/dehydrogenases	Plant development, Stress response	Loss-Morais et al., (2014)
	miR393	Transport inhibitor-like protein, DNA-binding proteins, GRR1-like protein	Chitin, cold, salt stress, and water deprivation	Akter et al., (2014)
	miR167	Auxin response factor	Plant development, stress response	Chaves et al., (2015)
	miR171	GRAS family transcription factors	Plant development and metabolism	
	miR390	TAS3	Cellular signaling pathways, Plant development	
Theaceae				
Tea plant (*Camellia inensis*)	miR156	SBP/SPL transcription factors	Plant growth, development, Cold Stress tolerance	Zhu and Luo, (2013)
	miR171	GRAS family transcription factors	Plant development, Stress response	Zhang et al., (2014)
	miR397	Laccase	Stress responses	
	miR399	Ubiquitin-conjugating enzyme	Stress responses	
	miR408	Plastocyanin-like	Cold stress	

CHAPTER - III

Comparative study of genes involved in capsaicinoid biosynthesis pathway in different *Capsicum* species

1. Introduction

Pungency, the most important economically trait of *Capsicum* species has been widely studied (Ramchiary et al., 2014 and Kim et al., 2014). Several studies reported intra as well inter-specific variation of pungency content in *Capsicum* fruits. QTL mapping study also reported identification of genomic region governing pungency traits (Ramchiary et al., 2014). However, of the many studies till date, the study where identification of *Pun1* could be done is the most significant study showing direct involvement of this gene in the pungency content (Curry et al., 1999; Aluru et al., 2003; Keyhaninejad et al., 2014). Mazourek et al., (2009) constructed a CapCyc model integrating a total of 55 candidate genes involved in capsaicinoids biosynthesis pathway. Whole genome sequencing identified sequence and expression of about nine geneinvolved in pungent and non pungent *C. annuum* (Kim et al., 2014, Qin et al., 2014). Several molecular markers has been reported for the pungency trait while, most of them are not reproducible in nature and are restricted to specific accessions of chilli peppers (Truong et al., 2009). However, study to completely dissect pungency at molecular level and identify all the genes involved in Capsaicinoid (pungency) biosynthesis pathway is underway.

From our study, we have identified that there is a wide variation with respect to pungency content, fruit traits (color, size, shape, etc) and other agronomic traits in Bhut jolokia (*C. chinense*) and other *Capsicum* germplasm collected from North East India (Sarpras et al., 2016). Therefore, the present study was designed for comparative study of pungency content and correlate with the expression of gene(s) involved in pugency biosynthesis pathway. The comparative study was done between extremely pungent Bhut jolokia and other *Capsicum* species i.e. moderate pungent *C. frutescens* and low pungent *C. annuum* genotypes. The biochemical analysis using GC-MS and expression analysis of pungency genes (reported by Kim et al., 2014) by qRT-PCR was performed in different stages of fruit development such as early (20 DPA), breaker (30-40 DPA) and mature (60 DPA) stage. Furthermore, our aim was to develop pungency gene(s) based molecular marker(s) which could be used in future breeding programme to enhance capsaicinoid content.

2. Materials and Methods

2.1 Sterilization protocols

Standard sterilization protocols were followed by autoclaving glassware's and culture media for 15 min at 121.6°C under 15 lb psi pressures. Heat labile reagents were filter sterilized using 0.22 µm pore size polyvinylidene difluoride filters (MillexTM, Millipore, USA) driven with a dispensable syringe. For work involving RNA, all required materials like glassware and mortar-pestle were incubated at 180°C for 5-6 hour (h). Precaution was taken to prevent RNase contamination of gel running unit. Diethyl pyrocarbonate (DEPC)-treated sterile Milli-Q (MQ) water was used to make all solutions. Working space and regular use items were cleaned with RNase-OUT reagent (G-Biosciences).

2.2 Plant material & growth conditions

A total of 20chilli pepper germplasm belonging to Bhut jolokia/Ghost chilli (*C. chinense*), *C. frutescens* and *C. annuum* collected from North-East (Assam, Nagaland, Manipur and Meghalaya) were used for the present study. Seeds of all the genotypes were sterilized before germination to prevent seed-borne diseases using the following protocol. Initially, these seeds were immersed in 0.1% of Bavistin for 5 minutes followed by soaking in 3% of sodium hypochlorite for 15-20 minutes with continuous shaking. Further, to remove the traces of surfactant, the seeds were washed 4-5 times using MQ water. These sterilized seeds were placed in petri plates containing the moist filter papers in dark condition. The germinated seeds were transferred in soil rite beds and further 2 weeks old seedlings were transferred in a clay pot at glass house of School of Life Sciences (SLS) at Jawaharlal Nehru University, New Delhi. The glasshouse was set at 24-26^{o}C, relative humidity of approximately 70%, and 14 h photoperiod with 250 µmol photons m^{-2} s^{-1} light intensity. The standard cultivation practice was followed for chilli pepper plant growth. In order to identify the days post anthesis (DPA) of fruits, the flowers were tagged with label. The samples from different developmental stages of fruit (preferably form second and third flush of flowers) were collected such at 5 DPA, 10 DPA, 20 DPA, 40 DPA, 60 DPA for the capsaicinoid extraction and molecular analysis.

2.3 Capsaicinoids extraction

The pre-weighed (1 gram)ripened chilli pepper fruits were homogenized with methanol (1:10, w/v) and further filtered by using Whatman paper No. 1 over anhydrous Sodium sulphate (Na_2SO_4). The filtered extract was evaporated in vacuum for complete drying, and the filtrate was suspended with 10 ml acetonitrile (Chinn et al., 2011, Sarpras et al., 2016). The solution containing samples were mixed and were then centrifuged at 14,000 rpm for 10 minutes and filtered through 0.45 µm Polytetrafluoroethylene (PTFE) membrane filter (Millipore). Further, this filtrate was injected to GC-MS. We used three replicates of samples for extraction and GC-MS analysis.

2.4 GC-MS analysis

The gas chromatography coupled with mass spectrometry (GCMS, Shimadzu QP2010 Plus) was used to detect and quantify the capsaicinoids. A Rtx- 5 MS capillary column (0.25 mm film thickness, 30 m in length and 0.25 mm internal diameter, Restek, USA) was used. The column temperature conditions are as follows. The temperature was kept at 100^{o}C for 2 min, further increased to 250^{o}C at a rate of 5^{o}C/minute, and finally set to 280^{o}C at a rate of 10^{o}C/ minute. About 1 µl of individual sample was infused to the column in split mode with split ratio 10. The carrier gas used in the instrument was helium with a flow rate of 1.21 ml/minute. Quadrupole mass spectrometer detector (in full scan mode) observed the occurrence of characteristic peak fragmentation patterns for capsaicin. To identify and quantify capsaicin and dihydrocapsaicin, we compared their peak retention times with mass and area with external reference standards analyzed under the same GC-MS conditions. The standard solutions were used with a different concentration of capsaicin and dihydrocapsaicin such as 0.5, 1, 2, 3, 4 and 5 mg/ml. The presence of metabolites was further confirmed by evaluating the spectral data of peaks with the analogous standard mass spectra from the NIST05 (National Institute of Standards and Technology library) and Wiley 8 library database. The capsaicinoid contents from the genotypes were measured in µg/g of fruits and Scoville heat unit (SHU) with multiplication of these values with the conversion factor for capsaicin and dihydrocapsaicin (16.0 x 10^6), nordihydrocapsaicin (9.3 x 10^6) and nonivamide (9.2 x 10^6) (Todd et al., 1977).

2.5 Nucleic acids extraction

A. Genomic DNA isolation by CTAB method

Genomic DNA was isolated from leaf tissues following the CTAB method as described by Porebski et al., (1997). About 1 gram (g) of sample was ground to a fine power using liquid nitrogen and 10 ml of preheated (at 65°C) CTAB extraction buffer was added. After incubation at 65°C for 1 hour (h), the suspension was cooled to room temperature and centrifuged at 13000 rpm for 15 min. To the supernatant, an equal volume of chloroform: isoamyl alcohol (24:1; v/v) solution was added, thoroughly mixed and centrifuged at 13,000 rpm for 15 min. The extraction was repeated once more, and subsequently, DNA was precipitated with isopropanol (600 µl/ml of the aqueous layer). After incubating overnight at 4°C/-20^{O}C for 1 h, DNA was pelleted by spinning at 13,000 rpm 4°C for 15 min followed by two washes with 70% ethanol (v/v) by centrifuging at 13,000 rpm for 10 min. After air-drying, the pellet was suspended in 500 µl nuclease free water.

To remove contamination with RNAs, the dissolved DNA was mixed with 5 µl RNaseA (100 µg/ml) and incubated at 37°C for 45 min. RNase-treated DNA was extracted with equal volume of Phenol: chloroform: isoamyl alcohol (24:24:1; v/v), and further with chloroform: isoamyl alcohol (24:1; v/v) followed by precipitation with equal volume of ice cold ethanol and centrifugation for 20 min at 8,000 rpm and 4°C. Pellets obtained was washed with 70% cold ethanol (v/v), air dried and dissolved in appropriate amount of nuclease free water.

B. RNA isolation

The RNA Extraction was carried out using Macherey-Nagel NucleoSpin RNA kit (MN, US) according to manufacturer's protocol. Briefly, about 500 mg plant tissue was uniformly homogenized with liquid Nitrogen and taken into 1.5 ml centrifuge tube. About 350 µl Lysis buffer (RA1) was added with 3.5 µl of β -mercaptoethanol to ground tissue and vortexed vigorously followed by centrifugation at 13,000 rpm for 3 min. The filtrate was transferred to a fresh tube, and 350 µl 70% ethanol was added and mixed properly. The solution was loaded into supplied column and centrifuged at 13,000 rpm for 1 min. The membrane desalting buffer (350 µl) was added and centrifuged at 13,000 rpm for 1 min to dry the membrane. To each column 10 µl rDNase and 90 µl reaction buffer was added and kept for incubation at room temperature for 15 min. The column is washed with series of

multiple wash buffers such as 200 µl wash buffer (RAW2), 600 µl wash buffer (RA3) and 250 µl wash buffer RA3 with centrifugation at 13,000 rpm for 1 min. The membrane was air dried followed by eluted with addition of 50 µl RNAse free water. Isolated RNA was stored in -80 ^{O}C for future use.

2.6 Spectrophotometric estimation

The quality and quantities of nucleic acids, i.e. DNAs and RNAs were checked by measuring absorbance at 230, 260 nm and 280 nm with NanoDrop 1000 Spectrophotometer (Thermo Scientific, USA). The RNA samples with A_{260}/A_{280} ratio of 1.8-2.0 were considered pure. DNAs, which showed A_{260}/A_{280} ratio in between 1.7-1.8, were considered as pure and taken for further experiments.

2.7 Gel electrophoresis

Horizontal agarose (Genetix, India) gel electrophoresis was used to separate RNA, DNA and PCR amplified products. For RNA, a denaturing 1.2% (w/v) formaldehyde gel was made by adding 10 ml 10X MOPS and 3 ml formaldehyde to 87 ml molten agar and cast in pre-set trays with fitted combs. About 1 µg of total RNA was mixed with formaldehyde gel loading buffer and incubated at 65°C for 10 min followed by quick chilling on ice. Denatured RNA samples were loaded after adding 2 µl of 10 X RNA gels loading dye and gel running was done in 1X MOPS buffer at low voltage until dyes were well separated. For DNA PCR samples appropriate percentage of gel was chosen based on product size (0.8-2.0%; w/v), molten in 1X TBE/TAE buffer and poured onto gel trays after adding 0.5 µg/ml EtBr. Two µl of 6X DNA gel loading dye was added to samples which were then run in 1X TBE/TAE along with appropriate size standards on an electric current of 3 V/cm and ended depending on the distance between the migrated bands of the dyes present in the DNA loading buffer. The nucleic acids were detected in a gel-documentation system on a UV-transilluminator.

2.8 Complementary DNA synthesis

RNA was reverse transcribed to synthesize the first strand of complementary DNA (cDNA) by using Verso cDNA synthesis kit (Thermo nuclear) using the manufactures instruction. Initially, 1 µg of RNA was takenfor 20 µl of reaction mixture and DEPC water was added together to make the final volume 11 µl. The mix was kept at 65°C for 5 min. In another vial 4 µl of 5X cDNA synthesis buffer, 2 µl of dNTP mix, 1 µl of 5 µM RNA Primer, 1 µl of RT Enhancer and 1 µl of Verso Enzyme Mix were added together to make a final

volume of 9 µl. This mixture was added to the cDNA- water mix and the final volume of the reaction mix was made 20 µl. The total reaction mixture was kept at 42°C for 1 h followed by termination at 95°C for 2 min.The synthesized cDNA was used for downstream PCR amplification reaction using respective gene-specific primers.

2.9 DNA and cDNA amplification

A. Polymerase Chain Reaction

All PCR reactions were performed in the Eppendorf (Germany) Thermal Cycler. Each PCR reaction mix contained 50 ng of DNA, 1X PCR buffer (50 mM KCl, 20 mM Tris-Cl pH 8.4), 1.5 mM $MgCl_2$, 2.5 mM of each dNTPs, 0.2 µM of each primer, and 0.5U of Taq DNA polymerase (New England BioLabs). The final reaction volume was made up to 20 µl with nuclease free water. Amplification was done using the touchdown-PCR. Amplifications were carried out in a thermal cycler with the following touch-down profile.

Table 3.1 The touch-down PCR amplification profile.

Step	**Temperature (^oC)**	**Time**	**Cycle**
Initial denaturation	94	2 minutes	1
Denaturation	94	20 sec	18 [0.5°C decrement in each subsequent cycle]
Annealing	60	50 sec	
Extension	72	50 sec	
Denaturation	94	20 sec	20
Annealing	50	50 sec	
Extension	72	50 sec	
Final extension	72	7 minutes	1
Hold	4	∞	

B. Quantitative Real Time PCR (qRT-PCR) analysis

The qRT-PCR was performed with SYBR Premix Ex Taq (Clontech, USA) as per the manufacturer's instruction. The qRT-PCR was conducted in ABI7500 Fast system (Applied Biosystems) with the following thermal protocol: 95°C for 2 min followed by 40 cycles of amplification of 15 s at 95°C and 1 min at 60°C. Immediately after final PCR cycle, a melt

curve analysis was performed from 60 to 95°C in increments of 0.5°C to confirm the specificity of PCR products. The reactions were carried out in triplicate, and the experiment was repeated at least twice. Also, the control reaction with the absence of template and reverse transcription were included for individual mRNA. Actin gene was used as an internal reference. The primer sequences of all the genes including the actin gene are listed in Annexure 1. After the completion of the reaction, the comparative Ct method $2^{-[\Delta\Delta Ct]}$ was used to quantify the relative expression of individual mRNA genes (Livak et al., 2001).

3. Results

3.1 Evaluation of morphological characters of *Capsicum* species

Among the 20 *Capsicum* germplasms belonging to three different *Capsicum* species (*C. chinense*, *C. frutescens* and *C. annuum*) wide variation in fruit length, fruit size and shape, fruit color and seed number was observed (Table 5). The average fruit weight (in g) was highest for *C. chinense* (4.9 g) followed by *C. annuum* (2.8 g), and *C. frutescens* (0.2 g) genotypes. Average fruit length (in cm) was found to be largest in *C. annuum* (5.5 cm) compared to that of *C. chinense* (4.7 cm) and *C. frutescens* (1.4 cm). The seed number was higher in *C. annuum* (46.7) compared to *C. chinense* (26.1) and *C. frutescens* (7.1). A wide variation in fruit color such as red, orange, yellow, chocolate, and purple was observed. A triangular fruit shape was mostly seen in *C. chinense* genotypes, while blocky shape, elongated fruit and short slender shape were observed in *C. chinense* and *C. frutescens*, respectively.

3.2 Comparative analysis of capsaicinoid content in different *Capsicum* species

The pungency (capsaicinoids) content in the matured fruits of the 20 *Capsicum* germplasm belonging to three different *Capsicum* species (*C. chinense*, *C. frutescens* and *C. annuum*) was estimated by using GC-MS. Among the reported capsaicinoids, the capsaicin was detected as the major fraction followed by dihydrocapsaicin, while nordihydrocapsaicin and nonivamide were found in smaller fractions in all six accessions. Other capsaicinoids could not be detected. The pungency content measured in SHU value of matured fruit of *C. chinense* varied from 913413 (0.9 million) to 1031624 (1 million), followed by *C. frutescens* from 200694 (0.2 million) to 452337 (0.4 million), and *C. annuum* from 3155 to 198055 (0.1 million). Furthermore, to see the pungency (capsaicinoids) level at different developmental stages of fruit, and correlate with corresponding gene expression in the three *Capsicum* species with varying pungency level, the pungency analysis was done in three developmental stages of *Capsicum* fruits in 6 different accessions belonging to *C. chinense*, *C. frutescens* and *C. annuum* (Fig. 3.1). The pungency content analysis revealed higher expression of capsaicinoids in Bhut jolokia compared to other *Capsicum* species at different stages (early, breaker and matured) fruit developmental as observed in the matured fruits (Table 3.2).

Our analysis showed that in *C. chinense*, the capsaicinoid content in fruits enhanced progressively from early to the matured stage (Fig. 3.1). In Acc 22 (Bhut jolokia), when the

fruit turns to maturity, the SHU value increased to 0.9 million from 0.3 million (early fruit stage). The same result was found in another *C. chinense* (Bhut jolokia) Acc 17 where, SHU increased from 0.5 million (early fruit stage) to 1 million at maturity stage. Similar pattern of increase of pungency content form early to maturity stage was observed in all the accessions of *C. frutescens* and *C. annuum* (Fig. 3.1), indicating highest pungency in matured stage of fruits.

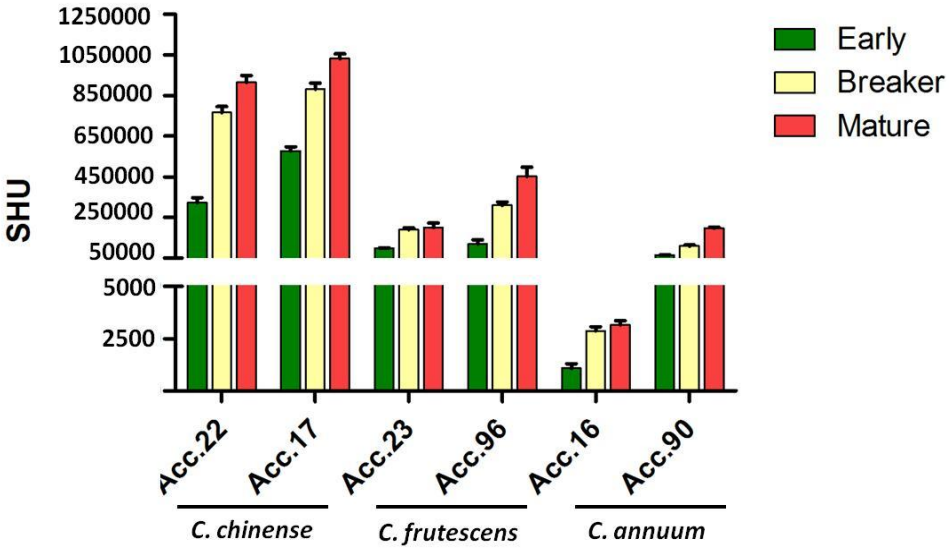


Figure 3.1 Pungency analyses of different accessions belonging to *C. chinense*, *C. frutescens*, and *C. annuum*. For GC-MS analysis, samples from different fruit developmental stages such as early (20 DPA), breaker (30-40 DPA) and mature (60 DPA) stages were selected.

Table 3.2 Concentration of different capsaicinoid constituents (in µg/g DW of fruit) detected at three fruit developmental stages of six accessions of three *Capsicum* species.

Accession	Species	Fruit stage	Capsaicin	Dihydro capsaicin in µg	Nordihydro capsaicin in µg	Nonivamide in µg	Capsaicinoids in µg	SHU
ACC 22	*C. chinense*	Early	14669± 463	5070± 331	37.7±2.3	9.33±0.12	19787.5±932.3	322143±25564
		Breaker	34859.4±1076	12050.1±654	90.2±9.9	23.3±1.2	47023.0±2019	765532±28991
		Mature	41593.35 ± 1506	14377.93 ± 978	107.64 ± 43.3	27.8395 ± 3.4	56106.7± 2530.7	913413.0 ± 40426
ACC 17	*C. chinense*	Early	23774±212	12080.64±389	nd	6.16± 0.09	35861.04± 698	577832±19983
		Breaker	36181±345	18385±676	nd	9.38± 0.45	54575±1282.6	879378±30992
		Mature	42446.0 ± 546.1	21568.0 ± 886.1	nd	11.01 ± 1.1	64024.0 ± 1433.1	1031624± 23065
ACC 23	*C. frutescens*	Early	3569.4± 212.2	2556.3±156.4	2.87±0.08	0.24±0.002	6159.3± 4252	99201±1030
		Breaker	6834.06±376.4	4965.1±312.9	5.56±0.1	0.47±0.05	11792.3± 789	189931±9001
		Mature	7221.34± 446.3	5238.41± 376.1	6.1 ± 0.1	0.50 ± 0.11	12461± 1101.3	200694.0± 18912
ACC 96	*C. frutescens*	Early	4293.9± 412.1	3115.04± 489.2	0.308±0.03	0.19±0.02	7409.51± 681.6	119332±20934
		Breaker	11248±786	8087.5±1021	0.80±0.01	0.51±0.002	19237.2±1652.1	309821±16244
		Mature	17546.22 ± 1671	12728.82 ± 1332	1.261 ± 0.7	0.81 ± 0.5	30277.1 ± 3004.2	487619.6 ± 48359
ACC 16	*C. annuum*	Early	48.45± 8.1	17.9± 1.3	nd	nd	66.8± 9.1	1092±212
		Breaker	127.84±16.2	47.46±3.9	nd	nd	176.23±22.1	2881±189
		Mature	140 ± 24	52 ± 6.1	0.6 ± 0.11	0.34 ± 0. 09	193 ± 30.3	3155 ± 323.1
ACC 90	*C. annuum*	Early	2857.02± 290.3	1055± 183	nd	15.8± 0.13	3982.4± 323	65370±944
		Breaker	4797.7± 567.2	1891±267.7	0.90±0.12	26.53±0.16	6687.68±543.1	109774±4220
		Mature	8656.12 ± 1004	3196.4 ± 546	1.63 ± 0.7	47.88 ± 0.8	12066.0 ± 1566.5	198055.6 ± 24968

Abbreviation: nd- Not detected. Early and breaker fruit is of 20 DPA and 30-40 DPA, respectively.

The accession numbers are the serial numbered given at the lab during the time of germplasm collection.

Table 3.3 Categorization of chilli peppers based on pungency level.

Scoville Heat Units	**Pungency**
>800,000	Extremely high pungent
500,000 to 800,000	High pungent
100,000 to 500,000	Moderate pungent
1,000 to 100,000	Low or mild pungent
0	Non pungent

3.3 Comparative expression analysis of capsaicinoid biosynthesis genes in contrasting *Capsicum* germplasm

To understand the molecular mechanism of pungency biosynthesis in the extremely pungent Bhut jolokia, moderately pungent *C. frutescens*, and low pungent *C annuum* (as shown by biochemical analysis with GC-MS), we have selected 10 already reported pungency (Capsaicinoids) biosynthesis genes for expression analysis (Fig. 3.2). These genes are; *PAL*, *BCAT*, *C4H*, *ACL*, *KAS*, *COMT*, *ACS*, *FAT*, *pAMT* and *AT3* (*Pun1*) (for details of the genes please refer Table 3.5). The qRT-PCR analysis for these genes was done in the six *Capsicum* accessions, i.e. two each from *C. chinense* (Acc 22 and 17), *C. frutescens* (Acc 23 and 96) and *C. annuum* (Acc 16 and 90), respectively in leaf, flower and three stages of fruit development (green, breaker and mature stages of the fruit (Fig. 3.3). The qRT-PCR analysis showed that the level of expression of the pungency genes varies with the pungency content. Among the 10 structural genes investigated in this study, four genes i.e *Pun1, pAMT, KAS and BCAT* were observed to be predominantly expressed in fruit tissues (Fig. 3.3). The expression of these four genes positively correlated with the capsaicinoid content by showing significantly higher expression in the fruit tissues of all the *Capsicum* species while expression of other structural genes was not seen to be fruit specific.

Furthermore, it was observed that, many of the pungency biosynthesis genes such as *PAL*, *BCAT*, *KAS*, *ACL*, *pAMT*, *Pun1* displayed significantly higher expression in extremely high pungent *C. chinense* genotypes followed by moderate pungent *C. frutescens* genotypes, while a very expression pattern was seen in *C. annuum* genotypes with low pungent phenotypes (Fig. 3.3). The expression analysis also showed that in both the *C. chinense* and *C. frutescens*,the expression of *Pun1, pAMT, KAS and BCAT* geneswas significantly high in early fruit development compared to later stage of fruit suggesting that the expression of these genes decreases with the fruit development from early to maturity stage (Fig. 3.3). However, in low pungent *C. annuum*, although similar trend might be follow, our analysis did not show expression of any of these genes.

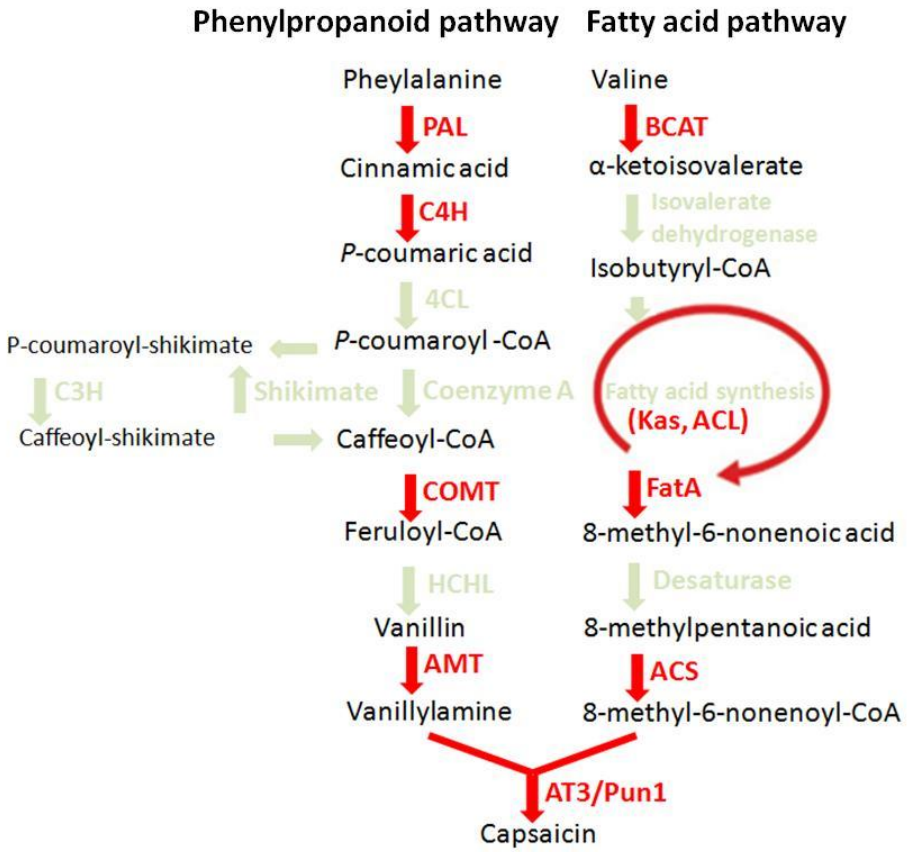


Figure 3.2 Capsaicinoid biosynthesis pathway. The genes responsible for enzymes which marked in red color are used in the expression analysis (Adapted from Kim et al., 2014).

Table 3.4 SHU range of different *Capsicum* species. (Sarpras et al., 2016)

Species Name	**SHU value**
C. chinense	0.27 million– 1.1 million
C. frutescens	0.1 – 0.48 million
C. annuum	0- 0.2 million

Table 3.5. List of the capsaicinoid biosynthesis pathway genes used for expression analysis

Gene Name	Gene ID	Chromosome number	Gene length (bp)	CDS ID	CDS length
PAL -Phenylalanine ammonia-lyase	Capana09g002199	9	3675	CA09g02410	2163
BCAT-Branched-chain amino acid transferase	Capana04g000751	4	2897	AY034379.1	1158
C4H-Cinnamic acid 4-hydroxylase	Capana06g000273	6	3275	CA00g30980	1518
Acl-Acyl carrier protein	Capana01g003112	1	2013	AF127796.1	399
Kas-β -Ketoacyl-ACP synthase	Capana01g000111	1	2158	CA01g00840	1290
CCoAOMT-Caffeic acid O-methyl transferase	Capana08g002351	8	1928	CA00g52190	744
ACS-Acyl-CoA synthetase	Capana00g003392	8	8923	CA00g04420	1977
FatA-acyl-ACP thioesterase	Capana06g000197	6	6923	CA00g30270	1116
pAMT- putative aminotransferase	CA03g07640	3	NA	CA03g07640	1395
AT3-acyltransferase	Capana02g002339	2	1996	CA02g18630	1323

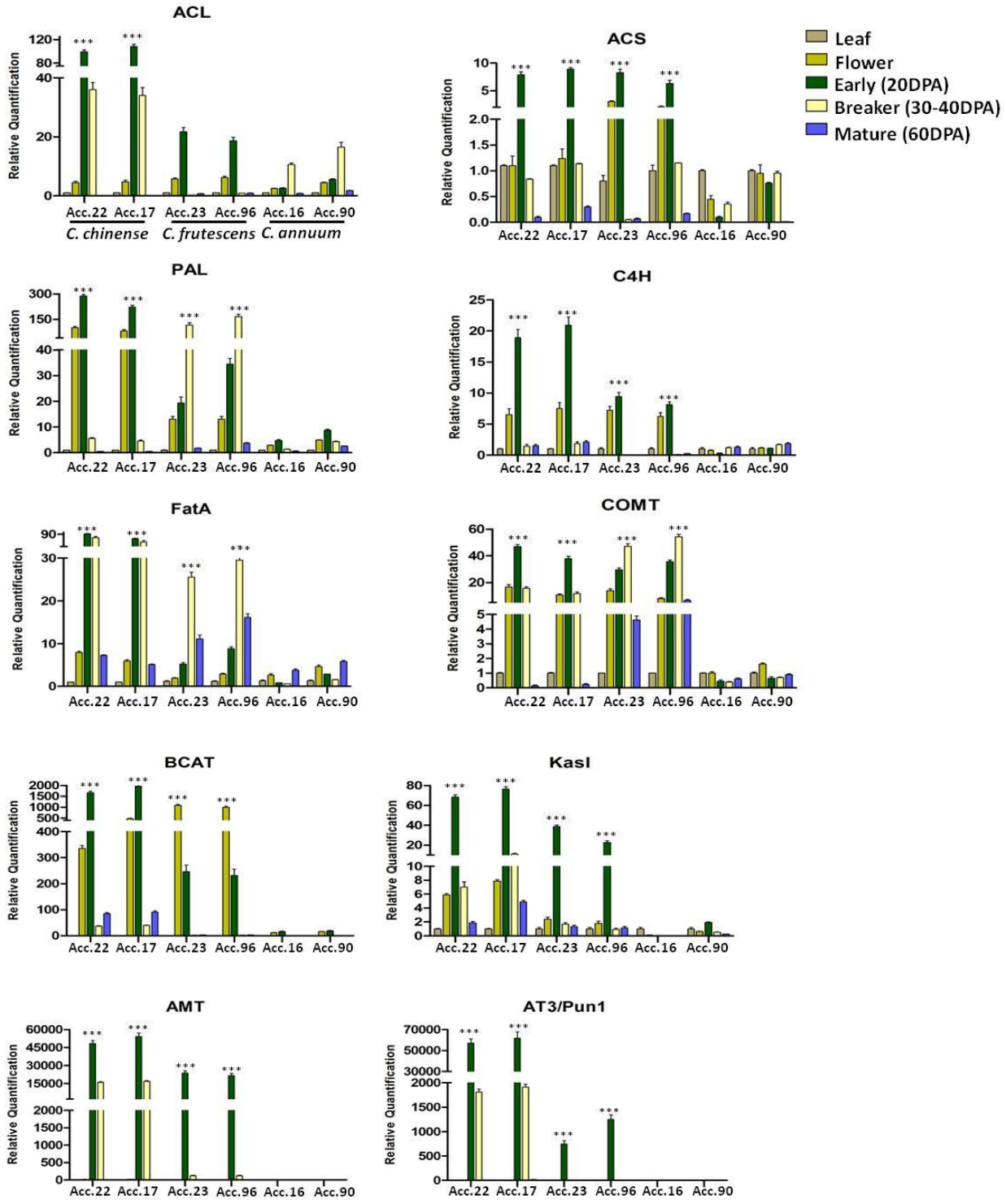


Figure 3.3 qRT-PCR analyses of genes involved in capsaicinoid biosynthesis pathway in high pungent *C. chinense* (Acc 22 & Acc 17), moderate pungent *C. frutescens* (Acc 23 & Acc 96) and low pungent *C. annuum* (Acc 16 & Acc 90) accessions. The expression analysis was performed with leaf, flower, and three developmental stages of fruit such as early, breaker and mature fruit of each accession of *Capsicum* species. Data represents means ± SD of three replicates (n=3); **P*<0.05, ***P*<0.01, ****P*<0.001.

3.4 Development of KAS gene-based marker for molecular breeding in *Capsicum*

We obtained the sequence of pungency biosynthesis gene *KAS* (keto acyl-ACP-synthase) from *C. annuum* Zunla-1 reference genome (Qin et al., 2014) as development of markers from *Pun1* and *AMT* has been already reported (Lang et al., 2009; Truong et al., 2009; Rodríguez-Maza et al., 2012; Wyatt et al., 2012). A total of four pairs of primers were designed based on the 2158 bp *KAS* gene sequence, and amplified the target sequence from the Bhut jolokia (*C. chinense*) and *C. frutescens* and *C. annuum* genotypes with varying pungency levels. After sequencing of the amplified products of the target regions, the alignment of the *KAS* gene of the three species showed single nucleotide polymorphisms (SNPs) as well as an insertion/deletions (InDels) along the sequences. Among this, a 12 bp deletion was identified at the moderate pungent *C. frutescens* between the 485–496 bp (Fig 3.4A) which constitutes the intronicregion of the gene, however, such InDel was absent in the other two pepper accessions from *C. chinense* and *C. annuum*. Since the InDel region is derived from the intronic region, there is no change in the coding of amino acids in *KAS* gene.

To validate the presence of InDel, primer pairs flanking the InDel region was designed to amplify 243 bp fragment of *KAS* gene in *C. chinense* and *C. annuum* accessions, and 231 bp fragment (12 bp deletion) in *C. frutescens* accession (Fig. 3.4B). As expected we observed a 243 bp fragment in all accessions of *C. annuum* and *C. chinense*, while a 231 bp fragment was detected in all the accessions of *C. frutescens* (Fig. 3.4C). This *KAS* gene based marker was named as *KAS*1 marker which can differentiate KAS allele of *C. frutescens* from *C. chinense* and *C. annuum* (Table 3.6). Although this gene based PCR based marker could not differentiate between pungent and non-pungent pepper accessions of *Capsicum* species, this marker could be used along with other pungency gene based markers in breeding programme.

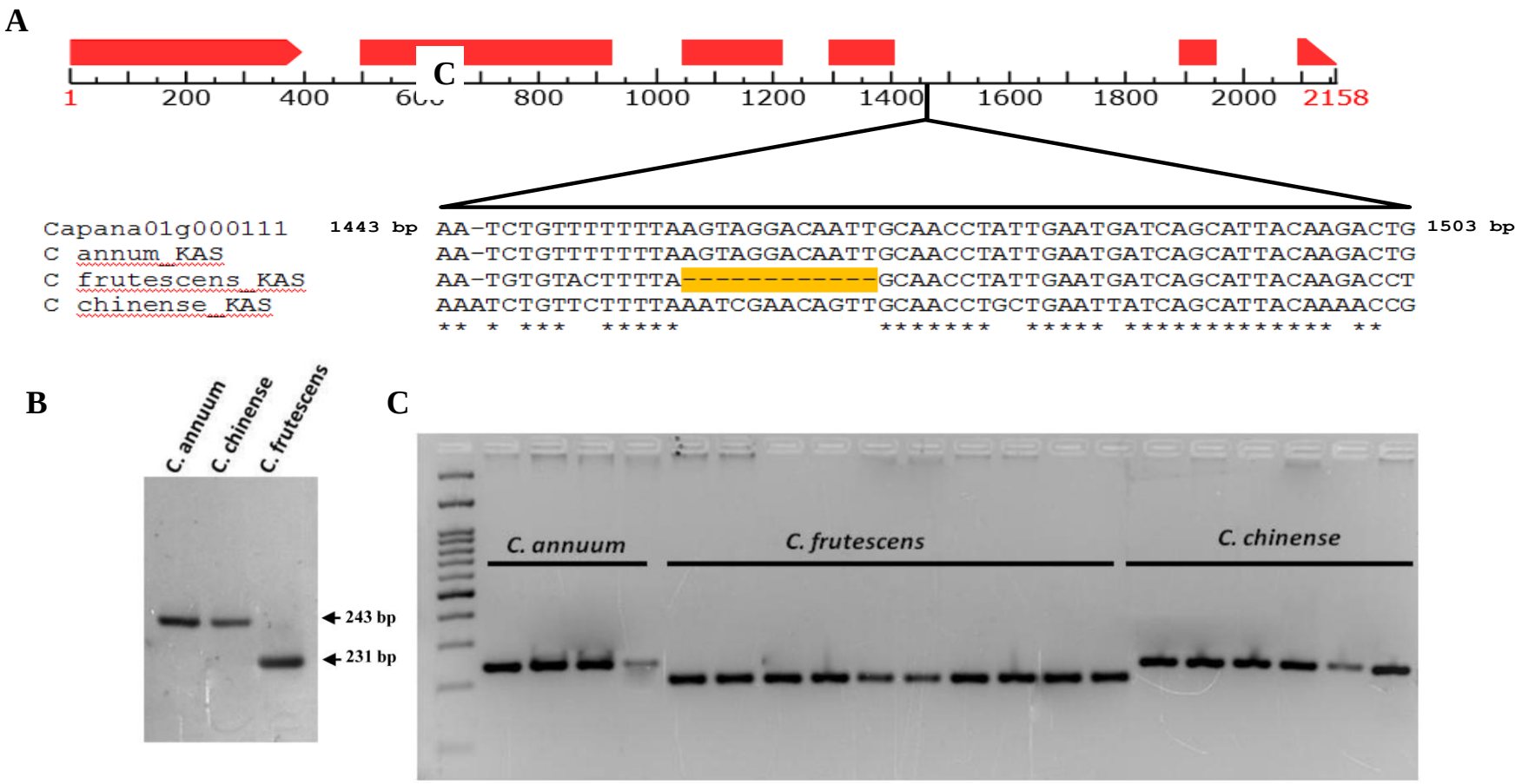


Figure. 3.4 Development of *KAS*-gene derived marker. A) Gene structure of *KAS*. Exons are displayed by red boxes. Multiple sequence alignment of *KAS* gene derived from three *Capsicum* species; *C. annuum*, *C. frutescens*, *C. chinense* and *KAS* gene from reference genome assembly. The highlighted region shows the deleted portion from *C. frutescens* accession. B) PCR product amplified with genomic DNA using specific primers from polymorphic sequence. The fragment size is mentioned on right side. C) MetaPhore agarose gel electrophoresis of *KAS*1 marker showing 231 bp fragment in *C. frutescens* and 243 bp fragmentin *C. annuum* and *C. chinense* accessions.

Table 3.6 The genotyping of contrasting *Capsicum* accessions with *KAS*1 marker showing distinction between *Capsicumchinense*/*C. annuum genotypes* from *C. frutescens* genotypes

Accession	Species	SHU value	Phenotype	Genotype
Acc. 16	*C. annuum*	3155.3 ± 323.1	LP	+
Acc. 42	*C. annuum*	5423.9±453.1	LP	+
Acc. 90	*C. annuum*	198055.6 ± 24968	LP	+
Acc. 83	*C. annuum*	8129.4±789	LP	+
Acc. 23	*C. frutescens*	200694.0± 18912	MP	-
Acc. 65	*C. frutescens*	452337.3±36183	MP	-
Acc. 32	*C. frutescens*	313296±23431	MP	-
Acc. 88	*C. frutescens*	228398±12341	MP	-
Acc. 96	*C. frutescens*	487619.6 ± 32359	MP	-
Acc. 71	*C. frutescens*	410382±22123	MP	-
Acc. 24	*C. frutescens*	200371±19847	MP	-
Acc. 87	*C. frutescens*	231935±14532	MP	-
Acc. 12	*C. frutescens*	423485±15567	MP	-
Acc. 26	*C. frutescens*	389261±16723	MP	-
Acc. 17	*C. chinense*	1031624±23065	HP	+
Acc. 22	*C. chinense*	913413.0 ± 40426	HP	+
Acc. 80	*C. chinense*	782716±38716	HP	+
Acc. 72	*C. chinense*	717383±36719	HP	+
Acc. 84	*C. chinense*	891726±38913	HP	+
Acc. 85	*C. chinense*	871829±32134	HP	+

Abbreviation- HP: High pungent, MP: Moderate pungent, LP: Low pungent, -:231 bp allele, +: 243 bp allele.

*The accession numbers are the serial numbered given at the lab during the time of germplasm collection

4. Discussion

The pungency (capsaicinoids) is one of the most studied economically important traits in *Capsicum* species due its use as spice and for the health-beneficial compounds (Meghvansi et al., 2010). Wide diversity of *Capsicum* species evolved in North East India which includes extreme fiery hot Bhut Jolokia' or 'Ghost chili', moderatly pungent *C. frutescens* and other *Capsicum species* (Bosland and Baral, 2007). The unique and special environmental conditions of North Eastern India favored the evolution of Bhut jolokia and other *Capsicum* species with varying degree of pungency content. Despite such huge diversity of *Capsicum* species especially in Bhut jolokia and *C. frutescens*, no systematic study at molecular level to understand the pungency biosynthesis have been done using these germplasm. In the current study, an attempt was made to correlate the pungency content and their gene expression in different accessions of three *Capsicum* species C. *chinense, C. frutescens* and *C. annuum*. The biochemical analysis of pungency by using GC-MS revealed that, the capsaicin and dihydrocapsaicin account for a major proportion of capsaicinoid complex as compared to other capsaicinoid constituents which is in agreement with the reports of earlier studies (Perucka et al., 2001; Topuz et al., 2007; Rodriguez-Uribe et al., 2012). Our analysis substantiates that capsaicin is found to be the primary capsaicinoid constituent in all of *Capsicum* accessions analyzed. Apart from capsaicin, other Capsaicinoids which we could detect and quantify are dihydrocapsaicin, nordihydrocapsaicin and nonivamide. As reported earlier, intra-specific and inter-specific variations of capsaicinoids content was found in the present analysis (Dyah et al., 1998; Contreras-Padilla et al., 1998, Mathur et al., 2000, Gnayfeed et al., 2001; Antonious et al., 2006; Sanatombi et al., 2008).

The comparative analysis of pungency, as expected, showed the highest level in Bhut jolokia followed by *C. frutescens* and *C. annuum* which is in consistent with earlier studies (Bosland et al., 2015; Dubey et al., 2015). The elevated level of pungency (in *C. chinense*) is due to higher accumulation of capsaicinoids in the placental septum of fruit as compared to other parts like pericarp (Aza-Gonzalez et al., 2011; Tanaka et al., 2017). Furthermore, in our another study it was observed that, although Bhut jolokia is reported to have highest pungency, variation was observed from 0.27 milllion SHU to 1.1 million, for *C. frutescens* 0.1-0.48 million SHU and for *C. annuum* 0-0.2 million SHU. This might be due the fact that during the course of evolution, cross hybridization with low pungency and accumulation of

mutations in pungency biosynthesis genes might have resulted in low pungency Bhut jolokia genotypes (Sarpras et al., 2016). This needs detail molecular characterization which is partly completed in the current study.

The phenotype is the manifestation of genotype with the interaction of environment. In the last decade, several studies attempted to investigate the molecular genetic basis of the capsaicinoid biosynthesis pathway and to isolate and characterise key regulatory genes of this pathway (Curry et al., 1999; Kim et al., 2001; Aluru et al., 2003; Stewart et al., 2005; Mazourek et al., 2009; Aza-Gonzalez et al., 2011). However, till now, direct correlation of pungency phenotypes with *Pun1* and *AMT* genes have only been found despite several genes are identified in the biosynthesis pathway (Stewart et al., 2005; Mazourek et al., 2009, Kim et al., 2014, Arce-Rodriguez and Ochoa-Alejo 2015). Furthermore, up to what extent these reported genes are responsible for determining capsaicinoid content in fruit are not known. Our aim is to understand and dissect the molecular basis of pungency biosynthesis. In the present work, we have selected 10 already reported genes for expression analysis using the qRT-PCR technique in different tissues of contrasting pugency genotypes of different *Capsicum* species.

The expression analysis showed that several capsaicnoid biosynthesis genes such as *Pun1, pAMT, KAS, BCAT, ACS* and *FAT* are highly expressed in one or more fruit developmental stages which is much higher in *C. chinense* accessions as compared to *C. frutescens* and *C. annuum,* indicating a high degree of correlation between gene expression and capsaicinoid concentration/accumulation. The *Pun1* gene that encodes for acyltransferase enzyme which catalyzes condensation of vannilylamine and fatty acid moiety to yield final capsaicin compound revealed significantly higher expression in the pungent genotypes of Bhut jolokia followed by *C. frutescens*. This observation is in agreement with earlier studies where they reported that expression of *Pun1* gene is observed in pungent genotypes compared to non pungency genotypes (Stewart et al., 2007, Ogawa et al., 2015). The expression of *Pun1* gene is significantly high in early (20 DPA) fruit stage as compared to the breaker stage and almost negligible in the matured stage of all the three *Capsicum* species analyzed which is in agreement with the observation of Iwai et al., (1979) where they have reported that capsaicinoids are found to be synthesized between 20 to 30 DPA of pungent cultivars of *Capsicum* species. The expression pattern of *Pun1* geneshowed significant correlation with the pungency (capsaicinoid) content i.e. extremely high in Bhut jolokia,

medium in *C. frutescens*and low in *C. annuum*. The analysis suggests that *Pun1* gene expression plays crucial roles in formulating capsaicinoid levels. Similarly, the *Capsicum* association mapping study showed that *Pun1* acts as a key regulator of chief metabolites and the accumulation of capsaicinoids are depended on the *Pun1* expression (Reddy et al., 2014).

The expression analysis showed that the *pAMT* (final enzyme of phenylpropanoid pathway) was found to be the second most important gene of in the capsaicinoid biosynthesis by displaying significant up-regulation of expression in highly pungent *C. chinense* genotypes as compared to moderately pungent *C. frutescens* and low pungent *C. annuum.* The expression analysis revealed that, the *pAMT* gene is predominantly expressed in early (20 DPA) and breaker stages of fruit development and highly co-related with the pungency content. The *pAMT* genes encodes for putative amino transferase enzyme which produces vanillylamine from vanillinwhich is an immediate precursor of capsaicinoids. It has been observed that, functional loss of *pAMT* due to single nucleotide (T) insertion at the position of 1291 bp leads to the formation of a sweat analog of capsaicinoids known as capsinoids in non-pungent *C. annuum* cv. CH-19 (Lang et al., 2009). This study confirms the vital role of the *pAMT* gene in determining capsaicinoid content in chilli peppers.

The expression analysis of other genes in phenylpropanoid pathway such as *PAL*, *C4H*, and *COMT* showed that their expression levels are not correlated with the capsaicinoid content (Fig. 3). The possible reason could be, since these genes are present in upstream part of the pathway they may contribute in the production of numerous secondary metabolites like flavonoids and lignins other than capsaicinoids (Vogt, 2010).The expression analyses of *BCAT*, *KAS*, *ACL*, *FAT-A* and *ACS* genes of fatty acid biosynthesis pathway demonstrated that, among them the expression levels of *BCAT* and *KAS* gene were significantly high in early (20 DPA) and breaker stage of fruit in all the analyzed *Capsicum* accessions. The relative expression of *BCAT* and *KAS* gene was considerably high in *C. chinense* accessions than that of *C. frutescens* and *C. annuum*. The *BCAT* gene product is vital for the methyl branched amino acids degradation, which is a part of the initial step of (methyl-branched ester and) capsaicinoid biosynthesis pathways (Mazourek et al., 2009). Differential expression analysis of fatty acid biosynthesis genes showed that *KAS* gene expression was positively associated with pungency content (Aluru et al., 2003). Further, silencing of *KAS* through virus-induced gene silencing leads to reduction of capsaicin content in *C. chinense* genotypes, which indicates a decisive role of *KAS* gene in determining pungency content

(Abraham-Juarez et al., 2008). Altogether, the qRT-PCR analysis demonstrated that the expression of four structural genes (*Pun1*, *pAMT*, *KAS*, and *BCAT*) were significantly high in capsaicinoid accumulating tissues such as fruit especially at early (20 DPA) stage. Our results suggest that these four genes (*Pun1*, *pAMT*, *KAS*, and *BCAT*) could be used for enhancement of capsaicinoid production in plants.

Molecular markers facilitate quick and easier identification of pungent/non-pungent accession in a larger population at early plant development such as at seedling or before fruit setting (Lee et al., 2005). Several markers based on *Pun1* gene and *pAMT* has been developed (Blum et al., 2002; Minamiyama et al., 2005; Lang et al., 2009; Rodiguez-Maza et al., 2012) while it has been observed that many of the pungency markers were genotypes specific and not reproducible in nature (Truong et al., 2009). Since much work has been done in development of *Pun1* and *pAMT* gene based marker which are among the key regulatory genes of capsaicinoid biosynthesis pathway, herein we attempted to develop molecular marker based on the *KAS* gene, considered as third most important gene of the pathway. Our definite aim was to develop *KAS*-based marker which could differentiate between pungent and non-pungent accessions of chilli peppers at the seedling stage of plant. The results showed that *KAS*1 marker could precisely detect the *KAS* allele of moderately pungent *C. frutescens* and high pungent *C. chinense*/low pungent *C. annuum*. Our analysis suggests that *KAS*1 marker is associated with moderate pungency. However, the *KAS*1 marker failed to differentiate between pungent and non-pungent genotypes of pepper, this could be due to targeted polymorphic sequence is not contributing in determining pungency content. This marker could be used in *Capsicum* breeding programme in which advancement of *C. frutescens* based hybrids are involved such as *C. frutescens* X *C. annuum*, *C. frutescens* X *C. chinense*, *C. frutescens* X *C. baccatum* and vice versa. Furthermore, pungency biosynthesis genes based molecular markers are being developed to for extremely high Bhut jolokia and other *Capsicum* specific. The findings of our study, which revealed very high level of pungency genes expression in Bhut jolokia suggests that this might be reason for extremely hot phenotypes in addition to nucleotide change in coding region. However, detail study of haplotypes of pungency biosynthesis genes in contrasting germplasm of different *Capsicum* species would reveal the molecular basis of pungency variation and this study is being done in the lab.

CHAPTER - IV

Development of genome-wide molecular markers in *C. chinense* and *C. frutescens* based on transcriptome data

1. Introduction

The discovery of the molecular markers has been indispensable in crop improvement and among the Solanaceae crop plants, the most extensive study of agronomic and quality traits using molecular markers have been done in tomato. Among different classes of molecular markers, simple sequence repeats (SSR) are useful for a variety of applications in plant genetics and breeding because of their codominant inheritance, reproducibility, multi-allelic nature, good genome coverage and relative abundance along with low running cost (Varshney et al., 2005).In *Capsicum* species, several SSR markers have been developed in *C. annuum* and used for genetic dissection of agronomically important and quality traits (Ramchiary et al., 2014, see Chapter II Review and Literature section 6.1). Recent studies of transcriptome sequencing using next generation sequencing technology accelerated the development of molecular markers in many crop plants including *Capsicum* species, as result, the development of the genic SSRs are becoming more and more feasible. Apart from dissecting *Capsicum* genome and mapping QTL/gene governing economically important traits, diversity study of *Capsicum* germplasm using molecular markers, especially SSR markers have been reported, mostly in *C. annuum* (Portis et al., 2007; Stágel et al., 2009; Pacheco-Olvera et al., 2012; Carvalho et al., 2015).

Bhut jolokia (*C. chinense*) and *C. frutescens* (moderate pungent) species from North East India are untapped *Capsicum* genetic resources which is still unexplored for breeding programme. Although increasing reports of genic SSR and other markers have been reported, there is limited information on marker development in Bhut jolokia (*C. Chinense*) and *C. frutescens*. Furthermore, identification of elite germplasms of Bhut jolokia and *C. frutescens* with desirable agronomically important traits could be used in study of *Capsicum* genetics and breeding. Therefore, the current investigation was undertaken with the objectives to generate a new set of genic SSR makers from transcriptome sequences of Bhut jolokia and *C. frutescens* and analyze genetic diversity of *Capsicum* germplasm collected from North East India belonging to *C. annuum*, *C. chinense*, and *C. frutescens* species. The identification of elite germplasm with distinct important economically important traits and their genetic relationship will help in utilizing these germplasm in breeding programme. Furthermore, the new set of gene based SSR markers would supplement the existing marker repertoire of *Capsicum* which could be used in genetics and genomics study of *Capsicum*s.

2. Materials and Methods

2.1 Plant material and growth condition

A total of 96 *Capsicum* genotypes belonging to *C. chinense*, *C. frutescens* and *C. annum* were used in the current study for the evaluation of morphological traits and for diversity study using SSR markers and the details are given in Table 4.1. The cultivation practice and growth condition of the *Capsicum* accessions, please see Chapter 1 section 2.1.

2.2 Development of Simple Sequence Repeat markers and annotation

The transcriptome sequences from the two highly contrasting genotypes of Bhut jolokia (*C. chinense*) and *C. frutescens* available in the lab was used for the development of SSR markers. The assembled unigenes set derived from transcriptomes of *C. chinense* (Acc. 7) and *C. frutescens* (Acc. 4) available in the lab was used identification of SSR motifs using microsatellite identification tool (MISA, MIcroSAtellite; http://pgrc.ipk-g atersleben.de/ misa). The transcriptome sequencing was done using total RNA from the leaf, flower and three developmental stages of fruit [20DPA, breaker (30-45DPA), and mature (45-60 DPA)] of *C. chinense* and *C. frutescens*. The criteria for selecting the SSRs was ten for mono-nucleotide repeats, five for di-nucleotide repeats, four for tri-nucleotide repeats, and three for tetra-, penta-, and hexa-nucleotide repeats. In the case of a compound microsatellite, a maximum difference between two SSRs was set as 100. The BatchPrimer3 v1.0 software (You et al., 2008) were used for primer designing with the following parameters: primer lengths range of 18–24 bases with an optimum 22 bases, GC content of 40–60% with an optimum 50%, annealing temperature of 40°C-60°C with an optimum 50^{o}C, and PCR product size of 100 to 400 bp with an optimum 200bp. The transcripts containing SSR motifs were compared against the sequences from reference genome assembly of *C. annuum* Zunla-1 using the stand alone BLAST program. Default parameters of the program were used and the expectation value (e-value) cut-off was set at 1^{e-10} for sequence similarity searches. The putative gene id was identified and their function was predicted. The structural annotation of SSR which reveals its distribution in the genic and inter-genic regions of the genome was investigated based on the genome sequence annotation information available at NCBI GenBank. The circos plot was used to depict the distribution of *C. chinense* and *C. frutescens* SSRs on the *C. annuum* genome using method described by Krzywinski et al. (2009).

2.3 Marker validation

A total of 50 genic-SSR markers were validated using 96 *Capsicum* accessions. Genomic DNA of each *Capsicum* accession was extracted from young leaves using the Hexadecyl trimethyl ammonium Bromide (CTAB) method. The quality and quantity of DNA was evaluated on a 1% agarose gel. The DNA concentration was adjusted to 50 ng/µl. All PCR reactions were performed in the Eppendorf (Germany) thermal cycler. Each PCR reaction mix contained 25 ng of DNA, 1X PCR buffer (50 mM KCl, 20 mM Tris–Cl pH 8.4), 1.5 mM MgCl2, 0.125 mM of each dNTPs, 0.5 lM of each primer, and 0.5U of Taq DNA polymerase (Life Technologies). The final reaction volume was made up to 20 ul with MQ water. Amplification was done using the touchdown profile. In this PCR, the initial denaturation was carried out at 95^{O}C for 2 min. This was followed by 10 cycles, each consisting of denaturation at 95^{O}C for 20 s, annealing at 65^{O}C for 50 s and extension at 72^{O}C for 50 s. During these cycles the annealing temperature was decreased at a uniform rate of 1^{O}C per cycle from 65 to 55^{O}C. This was followed by another 25 cycles, each having denaturation at 95^{O}C for 20 s, annealing at 55^{O}C for 50 ^{O}C. for 50 s, extension at 72^{O}C for 50 s and a final extension at 72^{O}C for 7 min.The PCR products were electrophoresed on 3% Metaphor agarose gels along with size markers and stained with ethidium bromide (EtBr).

2.4 Preparation of Metaphor agarose gel and running conditions

Three percent Metaphor (Cambrex, East Rutherford, N.J) agarose gels containing 0.15 µg ethidium bromide/ml were used to separate PCR amplification products. The gel was prepared according to manufacturer's instructions with slight modifications. Briefly, for 3% metaphor agarose gel, metaphor agarose and agarose were taken in 3:1 ratio and added to prechilled IX TBE buffer. Care was taken to avoid the formation of agarose clumps in thebuffer and mixed well. After the addition, the metaphor agarose was allowed to swell by incubating the mixture at 4^{O}C for 1-1.5 h. The resulting solution was weighed and boiled in a microwave for 2 min. The conical flask was swirled in order to dissolve the agarose properly. After complete dissolution the flask was weighed again and the distilled water was added to make up the weight loss. The solution was cooled down to 55^{O}C, and gel was cast after adding the EtBr (0.15 µg/ml). The PCR products were mixed with the tracking dye, loaded on gel and electrophoresed at 6V/cm for 2 h in 1X TBE.

2.5 SSR Data Analysis

All the *Capsicum* accessions were genotyped using set of 20 representative SSR markers. Genotyping data was recorded in the format of length variation. Diversity analysis including, total number of alleles (NA), allele frequency, major allele, accession-specific alleles, gene diversity (GD), and polymorphism information content (PIC) was performed. Genetic distances between each pair of accessions were calculated. A neighbor-joing (NJ) phylogram was constructed using distance matrix using MEGA4 package (Tamura et al. 2007; available in PowerMarker). All the above mentioned genetic diversity analysis were performed using PowerMarker version 3.23 (Liu and Muse, 2005).

2.6 *In silco* identification of polymorphic SSR between *C. chinense* and *C. frutescens*

In order to discover *in silico* polymorphic SSRs from the transcriptome data between *C chinense* and *C. frutescens*, we have extracted 250 bp transcript sequences flanking (5' and 3' end) the microsatellite repeat-motifs from *C. chinense* andcompared with flanking sequences of assembled contigs of *C. frutescens* using BLASTN (stand-alone). The precisely matched (about 95%) of 5´ and 3´ end flanking sequences of *C. chinense* displaying increase or decrease in number of identical microsatellite repeats in the assembled sequences of *C. frutescens* were believed as *in silico* polymorphic SSRs The criteria used for BLAST homology were, E-value <1e-10 and the sequence homology should be more than 95%. The rest of the reads of *C. chinense* with dissimilar alleles of identical microsatellite repeat motifs as compared to *C. frutescens* were not incorporated in analysis. In our analysis, if any insertion or deletions were found in flanking regions of the microsatellite repeat motifs from two species, the sequences were not described as polymorphic. The forward and reverse primers were designed for these polymorphic SSRs using BatchPrimer3 programme.

3. Results

3.1 Morpho-agronomic characterization of chilli pepper germplasm

The characterization 96 *Capsicum* germplasm, collected mainly from the North East India, belonging to *C. chinense, C. frutescence* and *C. annuum*, showed diverse morpho-agronomic traits. The morpho-agronomic characters were studied following the descriptors mentioned by International Plant Genetic Resources Institute (IPGRI, 1995). In our study, four different colors were observed for the mature fruit ranging from yellow to red. Among the 96 germplasm, the red color of the fruit (51%) was the most common color observed followed by orange (12.5%), yellow (5.2%) and chocolate or brown (2%) fruit colors. Elongated fruit shape (57%) was predominately observed in accessions followed by triangular (23%), round (11%) and blocky (5%) (Table 4.1).Among the other fruit descriptors, we observed varioustypes of fruit shape at blossom end, the majority of the accessions had a pointed shape at blossom end (72%), followed by blunt (20%) and sunken (7%). Significant variation was observed for other quantitative traits. Fruit weight varied from lowest 0.11 g (in Acc 23 of *C. frutescens*) to highest 8.6 g (Acc 95 of *C. annuum*), while fruit length ranged from lowest 0.87 cm (Acc 12 of *C. frutescens*) to highest 8.1 cm (Acc 19 of *C. chinense*) (Table 4.1). The pungency analysis of these 96 germplasm was done in another experiment by using GC-MS (Gas Chromatography-Mass Spectrometry) (Sarpras et al., 2016). The pungency analysis revealed that all of the *C. chinense* accessations are highly pungent (6,00,000 to 8,00,000 Scoville Heat Unit), while *C. frutescens* are of moderate pungent (3,00,000 to 6,00,000 SHU) and *C. annuum* accessions are low pungent (0 to 1,50,000 SHU). The sweet pepper/bell pepper (*C. annuum*) has negligible pungency (0 SHU value) level. The details of the morpho-agronomic data of all ninety-six accessions such as origin, the degree of pungency, fruit weight, fruit width, fruit position, fruit shape, fruit shape at blossom end have been listed in Table 4.1 The Figure 4.1 depicts the selected 96 chilli pepper accessions showing contrasting characters such as fruit color, size, shape, fruiting habit.

Table 4.1 Details of the chilli pepper accessions, origin and morpho-agronomic characteristics.

S No.	Accession number	Species	Source of origin	Degree of pungency	Fruit position	Fruit shape	Fruit colour	Fruit shape at blossom end	Fruit Length (cm)	Fruit Weight (g)	Seed Count (n)	Seed Weight (10)
1	Acc 1	*C. annuum*	NBPGR, India	Low pungent	Pendant	Almost round	Dark red	Blunt	6	3.5	40	0.322
2	Acc 2	*C. annuum*	NBPGR, India	Low pungent	Pendant	Elongated	Red	Blunt	5	2.5	78	0.382
3	Acc 3	*C. annuum*	NBPGR, India	Low pungent	Pendant	Almost round	Light red	Blunt	5.5	4	67	0.422
4	Acc 4	*C. annuum*	NBPGR, India	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	6.7	3.6	60	0.047
5	Acc 5	*C. chinense*	NBPGR, India	Highly pungent	Pendant	Elongated	Dark red at maturity	Pointed	5.8	2.9	66	0.036
6	Acc 6	*C. annuum*	NBPGR, India	Low pungent	Pendant	Elongated	Green in immature to dark red in mature	Pointed	6	1.8	42	0.055
7	Acc 7	Other *Capsicum* species	NBPGR, India	Low pungent	Pendant	Almost round	Dark red	Blunt	4.5	3.8	26	0.028
8	Acc 8	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Dark red	Pointed	4.5	1.3	53	0.037
9	Acc 9	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Dark red	Pointed	4.1	1.5	43	0.039
10	Acc 10	*C. annuum*	NBPGR, India	Low pungent	Pendant	Elongated	Light red	Blunt	6	4.5	41	0.0389
11	Acc 11	*C. chinense*	NBPGR, India	Highly pungent	Pendant	Triangular	Green in immature to orangish red in mature	Blunt	4.5	1.3	44	0.0283
12	Acc 12	*C. frutescens*	Assam, India	Moderately pungent	Pendant	Triangular	Green in immature to red in mature	Pointed	0.87	0.8	13	0.044
13	Acc 13	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	4.6	1.5	25	0.039
14	Acc 14	*C. annuum*	NBPGR, India	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	4.5	2.1	68	0.0283
15	Acc 15	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Dark red	Blunt	3.4	1.5	31	0.0425
16	Acc 16	*C. annuum*	Jammu & Kashmir, India	Low pungent	Pendant	Almost round	Dark red	Blunt	5	4	70	0.026
17	Acc 17	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Orange	Pointed	3.4	2.1	16	0.036

18	Acc 18	*C. chinense*	Assam, India	Highly pungent	Pendant	Campanulate	Green in immature turns to light red in mature	Pointed	6.1	1.3	14	0.025
19	Acc 19	*C. chinense*	Manipur, India	Highly pungent	Pendant	Triangular	Light green in immature turns to red in mature	Pointed	8.1	1.817	8	0.021
20	Acc 20	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Chocolate	Pointed	6	1.68	12	0.025
21	Acc 21	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Orangish red	Pointed	6.5	0.368	61	0.033
22	Acc 22	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Orangish red	Pointed	5.2	0.19	10	0.036
23	Acc 23	*C. frutescens*	Manipur, India	Moderately pungent	Errect/up ward	Elongated	Green in immature turns to red in mature	Pointed	0.925	0.115	8	0.044
24	Acc 24	*C. frutescens*	Manipur, India	Moderately pungent	Errect/up ward	Elongated	Green in immature turns to red in mature	Pointed	5	0.8	18	0.0288
25	Acc 25	*C. annuum*	Nagaland, India	Low pungent	Pendant	Elongated	Green in immature turns to red in mature	Pointed	4.5	1.3	53	0.037
26	Acc 26	*C. frutescens*	Nagaland, India	Moderately pungent	Pendant	Elongated	Yellowish red in mature	Blunt	3.4	1	28	0.044
27	Acc 27	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	3.2	1.6	59	0.035
28	Acc 28	*C. frutescens*	Assam, India	Moderately pungent	Pendant	Elongated	Dark red at maturity	Pointed	4.5	1.7	23	0.044
29	Acc 29	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	3.6	1	38	0.03
30	Acc 30	*C. annuum*	Assam, India	Low pungent	Pendant	Triangular	Red	Pointed	3	1.2	63	0.037
31	Acc 31	*C. annuum*	Nagaland, India	Low pungent	Pendant	Elongated	Red	Pointed	4.5	1.8	39	0.027
32	Acc 32	*C. frutescens*	Nagaland, India	Moderately pungent	Erect/upw ard	Elongated	Green in immature to orangish green in breaker to red in mature	Pointed	2.5	0.7	21	0.0418
33	Acc 33	*C. chinense*	Nagaland, India	Highly pungent	Pendant	Elongated	Red	Pointed	3.5	1.1	21	0.041
34	Acc 34	*C. chinense*	Assam, India	Highly pungent	Pendant	Elongated	Light green	Pointed	3.8	1.6	12	0.03
35	Acc 35	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Red	Pointed	3.8	1	37	0.038

36	Acc 36	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Light red	Pointed	3.8	1.2	46	0.028
37	Acc 37	*C. annuum*	Assam, India	Low pungent	Pendant	Almost round	Light to dark red	Pointed	4.8	1.9	11	0.023
38	Acc 38	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Dark red	Pointed	3.3	1.8	35	0.041
39	Acc 39	*C.annuum*	Canada	Low pungent	Pendant	Elongated	Yellow at maturity	Sunken	6.5	3.1	41	0.51
40	Acc 40	*C.annuum*	Canada	Low pungent	Pendant	Block shaped	Red at maturity	Sunken	4.5	3.5	37	0.0568
41	Acc 41	*C.annuum*	Canada	Low pungent	Pendant	Block shaped	Red at maturity	Sunken	2.9	4.6	25	0.0618
42	Acc 42	*C.annuum*	Canada	Low pungent	Pendant	Block shaped	Red	Sunken	3.8	3.6	28	0.041
43	Acc 43	*C. annuum*	Canada	Low pungent	Pendant	Block shaped	Yellow	Sunken	7.1	2.562	35	0.0768
44	Acc 44	*C. annuum*	New Delhi, India	Low pungent	Pendant	Elongated	Bright yellow in mature	Pointed	5.5	2.8	24	0.0768
45	Acc 45	*C. annuum*	Canada	Low pungent	Pendant	Elongated	Lemon yellow	Pointed	8	6.5	95	0.077
46	Acc 46	*C. chinense*	Canada	Low pungent	Pendant	Triangular	Red at maturity	Pointed	4.5	3.8	21	0.032
47	Acc 47	*C. annuum*	Manipur, India	Low pungent	Pendant	Campanulate	Light orange at maturity	Pointed	5.5	3.2	23	0.027
48	Acc 48	*C. annuum*	Manipur, India	Low pungent	Pendant	Campanulate	Light red at maturity	Pointed	7	22	66	0.047
49	Acc 49	*Other Capsicum species*	Assam, India	Low pungent	Erect	Elongated	Red	Blunt	1.5	0.9	40	0.04
50	Acc 50	*C. annuum*	Assam, India	Low pungent	Erect/ upward	Almost round	Red	Blunt	2.6	1	41	0.037
51	Acc 51	*Other Capsicum species*	Jammu & Kashmir, India	Low pungent	Erect/ upward	Almost round	Bright red	Blunt	2.1	1.6	12	0.028
52	Acc 52	*C. annuum*	Assam, India	Low pungent	Pendant	Almost round	Red	Blunt	3	1.3	46	0.035
53	Acc 53	*C. annuum*	Mizoram, India	Low pungent	Pendant	Elongated	Red	Pointed	3.5	1.6	34	0.012
54	Acc 54	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Green in immature turns tu purple greenish to red in mature	Blunt	1.7	1.5	57	0.04

55	Acc 55	*C. annuum*	Assam, India	Low pungent	Pendant	Almost round	Green in immature turns to orange in mature	Blunt	1.2	1.8	21	0.029
56	Acc 56	*C. chinense*	Nagaland, India	Highly pungent	Pendant	Elongated	Green in immature, to greenish orange in breaker,turns red when mature	Pointed	8	0.337	23	0.043
57	Acc 57	*C. annuum*	Assam, India	Low pungent	Pendant	Triangular	Dark red at maturity	Pointed	3.6	0.89	41	0.045
58	Acc 58	*C. annuum*	Nagaland, India	Low pungent	Pendant	Elongated	Green in immature turns to red in mature	Pointed	4.8	1.5	60	0.047
59	Acc 59	*C. annuum*	Nagaland, India	Low pungent	Pendant	Elongated	Red	Pointed	4	1.4	69	0.048
60	Acc 60	*C. chinense*	Assam, India	Highly pungent	Pendant	Elongated	Red	Pointed	3.2	1.6	34	0.024
61	Acc 61	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Orangish red	Sunken	4.1	3.2	15	0.048
62	Acc 62	*C. chinense*	Assam, India	Highly pungent	Errect/up ward	Triangular	Orange	Pointed	4	3.1	12	0.032
63	Acc 63	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Green in immature turns to red in mature	Pointed	6	2.4	36	0.042
64	Acc 64	*Other Capsicumm species*	Assam, India	Low pungent	Errect/up ward	Almost round	Purple colour fruit at immature turns to red at maturity	Blunt	1.4	2.1	18	0.032
65	Acc 65	*C. frutescens*	Assam, India	Moderately pungent	Errect/up ward	Elongated	Green in immature turns to orange in mature	Pointed	2.8	0.6	23	0.256
66	Acc 66	*Other Capsicum species*	Assam, India	Low pungent	Pendant	Elongated	Red	Pointed	2.1	1.1	16	0.032
67	Acc 67	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Red	Pointed	1.8	1.8	28	0.021
68	Acc 68	*C. chinense*	Assam, India	Highly pungent	Pendant	Elongated	Red	Pointed	2.2	0.5	18	0.0396
69	Acc 69	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Red	Pointed	2.8	1.8	37	0.025

70	Acc 70	*C. chinense*	Manipur, India	Highly pungent	Pendant	Elongated	Red	Pointed	2.8	1.2	15	0.028
71	Acc 71	*C. frutescens*	Assam, India	Moderately pungent	Errect/up ward	Elongated	Light red in mature	Pointed	2.4	0.5	23	0.0396
72	Acc 72	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Red	Pointed	2.8	2.1	18	0.03
73	Acc 73	*C. chinense*	Nagaland, India	Highly pungent	Pendant	Triangular	Red	Pointed	3	1.2	34	0.027
74	Acc 74	*C. chinense*	Manipur, India	Highly pungent	Pendant	Triangular	Green in immature turns to red in mature	Pointed	5.5	3.8	24	0.032
75	Acc 75	*C. chinense*	Manipur, India	Highly pungent	Pendant	Triangular	Green in immature, to greenish orange in breaker,turns red when mature	Pointed	4.8	3.2	19	0.0185
76	Acc 76	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Light green in immature turns to red	Pointed	4.5	2.6	21	0.0485
77	Acc 77	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Green in immature to turns to red in maturity	Pointed	4.2	2.9	23	0.0627
78	Acc 78	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Orange	Pointed	4.2	2.5	12	0.035
79	Acc 79	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Green in immature turns to red in mature	Pointed	4	2.6	14	0.035
80	Acc 80	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Light green in immature turns to red in mature	Pointed	5.8	6.5	19	0.044
81	Acc 81	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Green in immature to orange red in maturity	Blunt	4.5	4.4	16	0.046
82	Acc 82	*C. annuum*	Manipur, India	Low pungent	Pendant	Elongated	Purple colour fruit at immature turns to red at maturity	Pointed	3.1	1.6	40	0.037
83	Acc 83	*C. annuum*	Uttar pradesh, India	Low pungent	Pendant	Elongated	Dark red	Pointed	6.1	4.9	25	0.025
84	Acc 84	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Orange	Pointed	5.1	2.9	18	0.046
85	Acc 85	*C. chinense*	Assam, India	Highly pungent	Pendant	Triangular	Dark chocolate at maturity	Pointed	4.2	4.1	14	0.032
86	Acc 86	*C. frutescens*	Mizoram, India	Moderately pungent	Pendant	Elongated	Dark red	Pointed	2.5	1	38	0.025

87	Acc 87	*C. frutescens*	Assam, India	Moderately pungent	Pendant	Almost round	Red	Blunt	1.8	2.5	8	0.063
88	Acc 88	*C. frutescens*	Manipur, India	Moderately pungent	Pendant	Elongated	Green in immature to red in mature	Pointed	4.7	2.4	39	0.039
89	Acc 89	*C. annuum*	Uttarakhand, Assam	Low pungent	Pendant	Elongated	Green in immature turns red in mature	Pointed	5.5	2.1	59	0.043
90	Acc 90	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Chocolate at immature turns to dark red at maturity	Sunken	5	2.7	32	0.029
91	Acc 91	*C. annuum*	Jammu & Kashmir, India	Low pungent	Pendant	Elongated	Red	Pointed	5	3.9	28	0.069
92	Acc 92	*C. annuum*	Jammu & Kashmir, India	Low pungent	Pendant	Elongated	Green in immature turns to red in mature	Pointed	7.2	2.4	23	0.024
93	Acc 93	*C. frutescens*	Assam, India	Moderately pungent	Erect	Elongated	Red at maturity	Pointed	2.8	1.8	21	0.018
94	Acc 94	*C. annuum*	Assam, India	Low pungent	Pendant	Elongated	Green in immature,purplish green in breaker to red inmature	Pointed	5	1.4	39	0.027
95	Acc 95	*C. annuum*	Mizoram, India	Low pungent	Pendant	Elongated	Red	Blunt	4.8	8.6	48	0.0478
96	Acc 96	*C. frutescens*	Mizoram, India	Moderately pungent	Erect/upward	Elongated	Red	Pointed	2.1	1.4	9	0.031

Abbreviation:-NBPGR - National Bureau of Plant Genetic Resources, New Delhi, India

Figure 4.1 The charactetistics showing contrasting fruit color, size, shape, and fruiting habits of all the ninety six *Capsicum* germplasm. The details of germplasm with species has mentioned in Table 4.1

3.2 Identification, characterization and distribution of SSR markers in *C. chinense* and *C. frutescens*

For identification of microsatellite (SSR) motifs a total of 184,975 (165,766,957 bp) *C. chinense* and 179,780 (159,424,197 bp) *C. frutescens* sequences from the assembled transcripts were examined by using MISA programme. The numbers of SSR-containing transcripts were 49,136 in *C. chinense* and 46,771 in *C. frutescens*. In our study, we have excluded mono SSRs. We observed, an average of one SSR locus for every 7.52 kb of *C. chinense* and 7.42 kb of *C. frutescens*. After filtration a total of 12,473 and 11,835 SSR motifs containing sequences were found in *C. chinense* and *C. frutescens,* respectively. Of these, the total number of compound SSRs identified was 898 in *C. chinense* and 862 in *C. frutescens*. The distribution of different repeat classes in *C. chinense* was 5796 (46.46%) di-, 6312 (50.6%) tri-, 297 (2.4%) tetra-, 34 (0.27%) penta- and 30 (0.24%) hexa-nucleotides while for *C. frutescens* it was 5319 (45%) di-, 6182 (52.2%)tri-, 269 (2.26%) tetra-, 43 (0.36%) penta- and 22 (0.18%) hexanucleotides repeats (Table 4.2). Among the identified SSR motifs, di- and trinucleotide repeats comprised of 12,108 (97.06%) and 11,501 (97.2%) in *C. chinense* and *C. frutescens,* respectively. The number of reiterations of a given repeat unit varied from 5 to 20, and SSRs the minimum of five reiterations (cut off limit) was the most abundant. The inverse relationship was observed between the number of repeat units and the frequency of a given SSR structure. Motifs showing more than ten reiterations were rare with a frequency of <1%.Among the dinucleotide repeat motifs, AG/CT, AT/TA and AC/GT were the most abundantly both in C. chinesne (17.85%, 15.9% and 11.05%, respectively) and *C. frutescens* (19.83%, 14.3% and 10.66% respectively) (Figure 4.3A). Similarly, the trinucleotide repeats AAC/GTT, AAG/CTT and AAT/AAT were the most abundant both in *C. chinense* (with frequencies of 18.98%, 9.96% and 5.6%, respectively) and *C. frutescens* (18.5%, 9.9% and 5.18%, respectively).

Of these, for primer designing, about 250 bp flanking sequences of SSR motif`s were extracted both for *C. chinense* and *C. frutescens*. After considering all the criteria, primer pairs flanking for a total of 4989 and 4781 SSR motifs were generated for *C. chinense* and *C frutescens*, respectively. The distribution and mapping of different types of SSRs from *C. chinense* (4989 SSRs) and *C. frutescens* (4781 SSRs) on twelve chromosomes of *C. annuum* reference genome were depicted in the form of circos plot which showed the high density of trinucleotide SSRs followed by di-, tetra-, penta- and hexanuclotide (Fig. 4.2)

Table 4.2 Detail characteristics of SSR motif`s identified in the *C. chinense* and *C. frutescens*

	Repeat unit number											
SSR Motif Length	**Species**	**5**	**6**	**7**	**8**	**9**	**10**	**>10**	**Total**	**Percentage**	**Number of primers designed**	**Percentage**
Di	*C. chinense*	-	2535	1367	802	552	379	161	5796	46.46	2185	43.8
	C. frutescens	-	2266	1233	753	532	349	186	5319	45	1908	39.9
Tri	*C. chinense*	3801	1799	638	73	-	-	-	6312	50.6	2395	48
	C. frutescens	3699	1707	707	62	-	5	1	6182	52.2	2419	50.6
Tetra	*C. chinense*	242	50	-	2	2	-	1	298	2.4	285	5.7
	C. frutescens	222	38	7	2	-	-	-	269	2.26	291	6.1
Penta	*C. chinense*	18	16	-	-	-	-	-	34	0.27	64	1.3
	C. frutescens	41	2	1	-	-	-	-	44	0.36	375	7.8
Hexa	*C. chinense*	13	11	3	-	-	3	-	30	0.24	59	1.2
	C. frutescens	12	7	1	1	-	-	-	21	0.18	88	1.8
Total	*C. chinense*	4074	4411	2008	877	554	382	162	**12473**		**4989**	
	C. frutescens	3974	4020	1949	818	532	354	187	**11835**		**4781**	

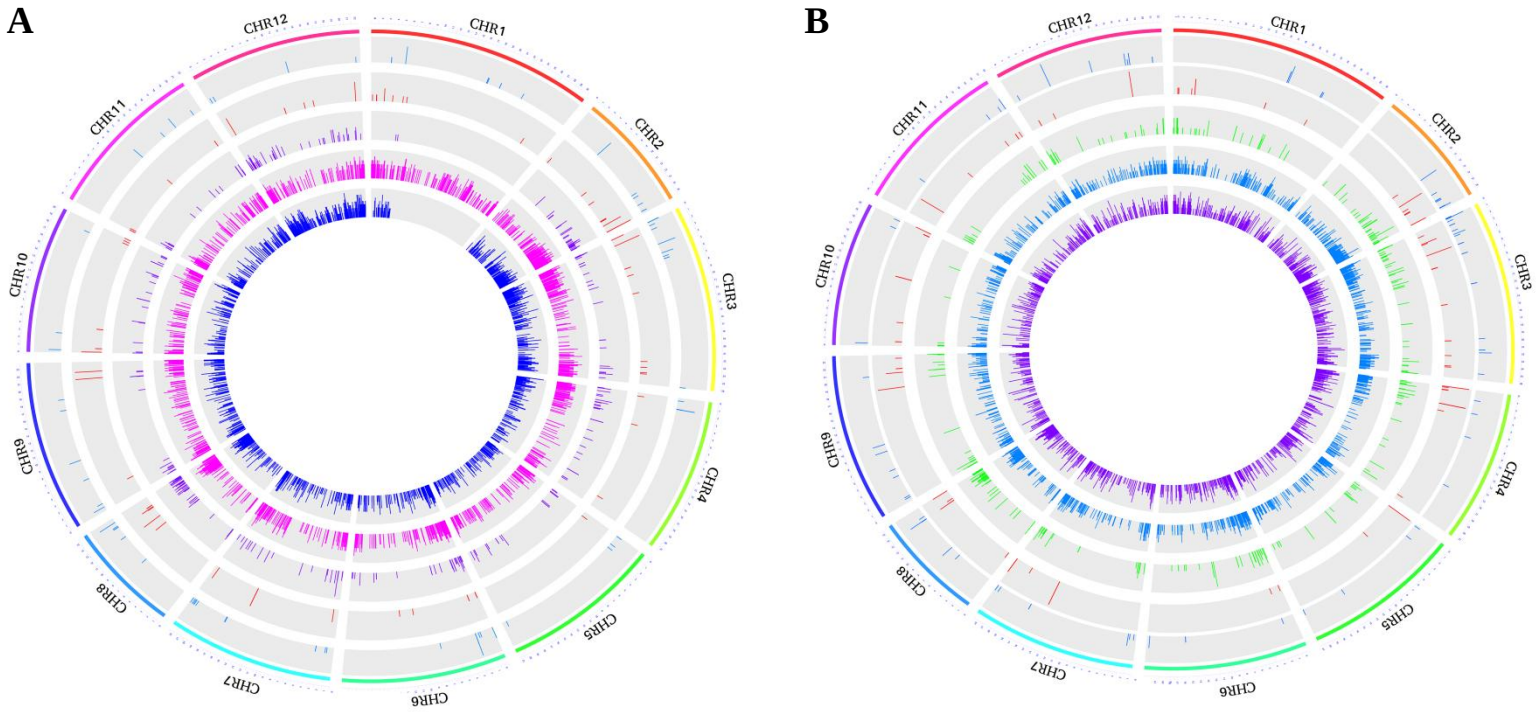


Figure 4.2 The distribution of SSR makers from *C. chinense* (4989) and *C. frutescens* in the *C. annuum* Zunla-1 reference genome. All the identified SSRs includingdi-, tri-, tetra-, penta-, hexa-, identified are shown in the form of circos plot in the in the reference genome. Each track from inside to outside refers to, di-, tri-, tetra-, penta-, hexanucleotide class of SSR repeat motifs.

Furthermore, we predicted the putative functions of each of the SSRs containing sequences from *C. chinense* and *C. frutescens*. Our analysis revealed that putative functions of about 88.27 % (of *C. chinense*) and 95.2 % (of *C. frutescens*) SSRs could be predicted based on the significant homology with reported proteins available in the *Capsicum annuum* Zunla-1 reference genome using BLAST (stand-alone). The remaining SSRs did not showed significant homology with proteins hence named as unknown protein or (N/A). The complete list of these 4989 *C. chinense* and 4781 *C. frutescens* SSR markers along with primer sequence, chromosome number, gene name and protein function are mentioned in Annexure 4&5 (The data is given separately in the form of CD Drive).

The distribution and position of all these identified SSRs on twelve chromosomes of *C. annuum* reference genome is given in Figure 4.3B. Furthermore, the analysis of origin of SSR motif`s identified different origin i.e. intergenic, coding, intronic, 5' and 3'-UTR of the genes. The SSR repeats found in the intergenic regions was more abundant in all types of repeats (2195 SSRs, 44% in *C. chinense* and 2221 SSRs, 46.4% in *C. frutescens*) (Fig. 4.3C). The SSRs motif`s within gene components showed highest proportion in the intronic regions (967 SSRs, 19.3% in *C. chinense* and 979 SSRs, 20.4% in *C. frutescens*) followed by CDS region (725, 14.5% in *C. chinense* and 804, 16.8% in *C. frutescens*) and 5'-UTRs (335, 6.7% in *C. chinense* and 350, 7.3% in *C. frutescens*). However, the number was the lowest in 3-UTR region of the genes (197, 4.1% in *C. chinense* and 804, 16.8% in *C. frutescens*) (Fig. 4.3C). In the coding region of genes, trinucleotide repeats were abundantly found (684 SSRs of 725, 94.3% in *C. chinense* and 732 SSRs of 804, 91% in *C. frutescens*). The di-nucleotide SSR repeats followed by tri-nucleotide were majorly found in the intronic regions of *C. chinense* and *C. frutescens*.

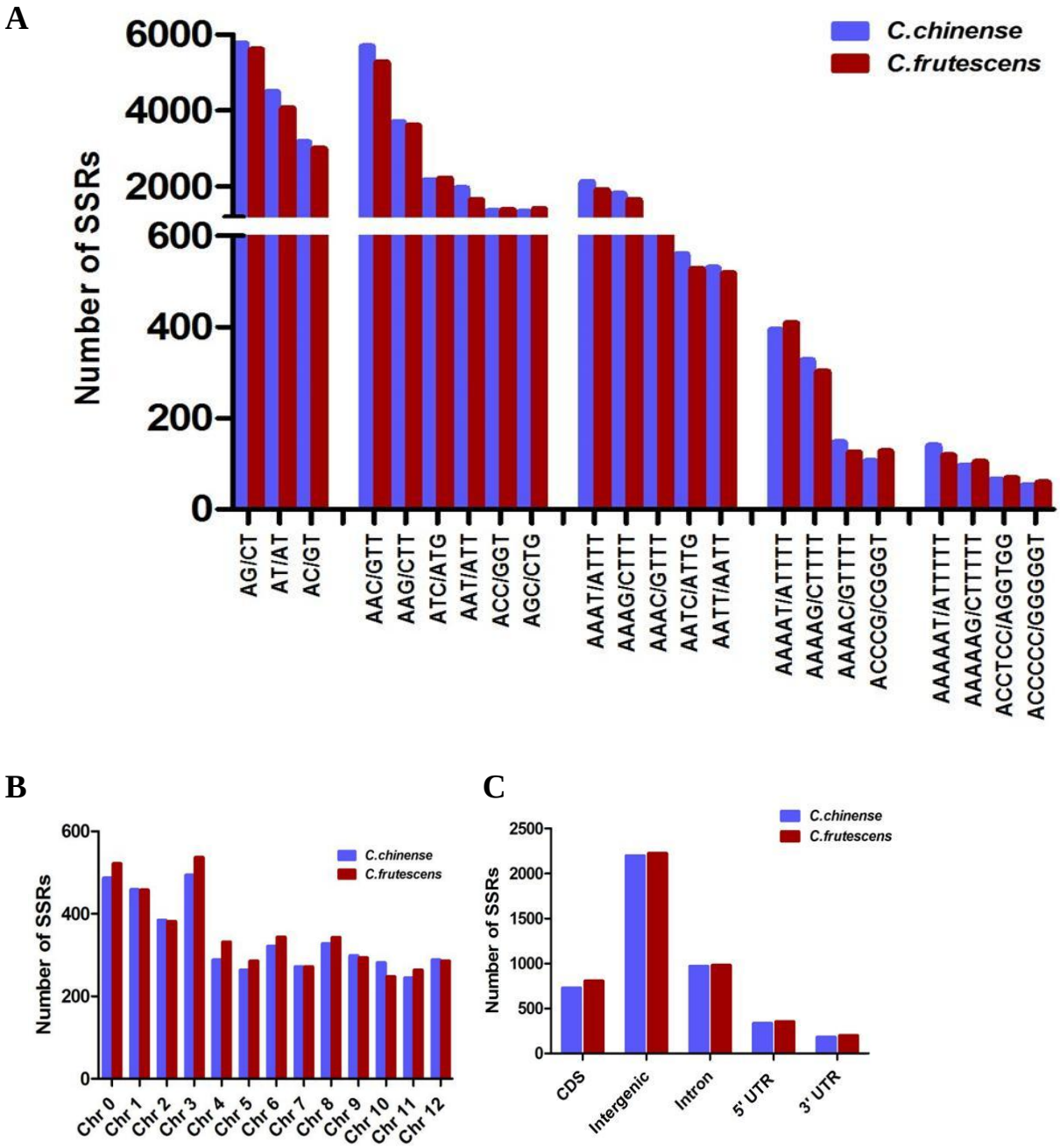


Figure 4.3 Detail characteristics of SSRs motif`s identified in *C. chinense* and *C. frutescens*. (A)Frequency of different SSR motifs identified in *C. chinense* and *C. frutescens* genomes, (B) Number of SSRs in distribution in different chromosomes of *Capsicumannuum* reference genome, (C) Characteristic features/origins of different SSRs in *C. chinense* and *C. frutescens.*

3.3 Validation of genic-SSR markers and diversity analysis of a set of *Capsicum* germplasm.

Of the total 4989 *C. chinense* and 4781 *C. frutescens* primer pairs designed flanking SSRs, 50 random SSR primer pairs representing both *C. chinense* and *C. frutescens* were custom synthesized. Initially these 50 primers pairs/SSR markers were PCR amplified in 10 *Capsicum* genotypes, 3 genotypes each from *C. chinense*, *C. frutescens* and *C. annuum*, respectively, along with one wild species, *Solanumpseudocapsicum*. The optimization of reaction components and conditions for efficient amplification of SSRs was done in the 10 genotypes. PCR amplification showed 100% amplification of SSR markers in at least one species (Fig.4.4). Of these 50 SSR primers/markers, 20 SSR primer pairs giving robust, clear and polymorphic bands amplification in all the three species were selected further for diversity analysis of 96 *Capsicum* germplasm belonging to *C. chinense*, *C. annuum* and *C. frutescens* (Fig. 4.5). Amplification of 96 genotypes using 20 SSR primer pair showed one to four polymorphic bands.

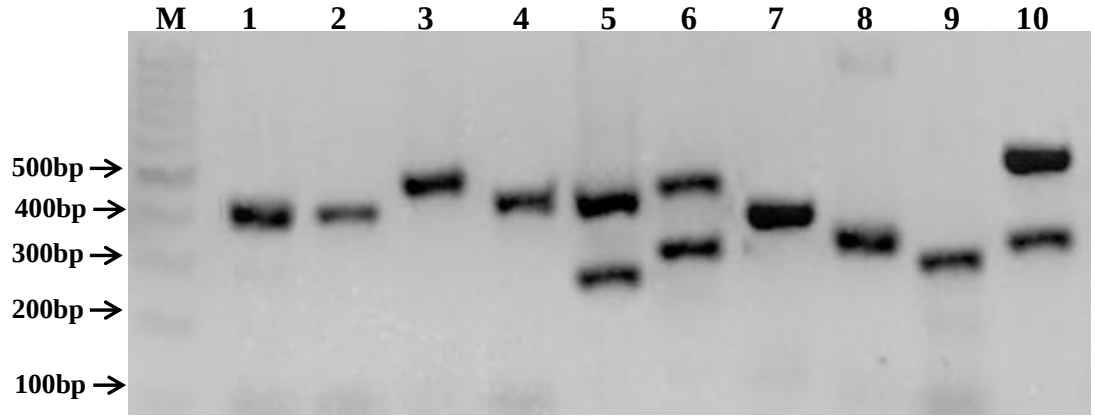


Figure 4.4 Representative gel picture of amplification and polymorphism confirmation using CAP_SSR1 primer. M: 100bp ladder; lane 1-3: *C. annuum* genotypes; lane 4-6: *C. chinense* genotypes; lane 6-9: *C. frutescens* genotypes; lane 12: wild species.

SSR polymorphisms were determined in terms of the heterozygosity, numbers of alleles, gene diversity, allele frequency and PIC, using the PowerMarker software. The 20 SSR markers revealed 84 alleles across the 96 accessions, with an average of 4.6 alleles per locus (Table 4.3). The allele size ranged from 150 (CAP_SSR1) to 450 bp (CAP_SSR15). The allelic richness per locus varied widely among the markers, ranging from 3 (CAP_SSR2, CAP_SSR3, CAP_SSR6, and CAP_SSR12) to 7 (CAP_SSR19) alleles. SSR markers are

highly informative and polymorphic as depicted from its PIC value. The polymorphism information content (PIC) value is a measure of polymorphism among genotypes for a marker locus. The PIC value of each SSR marker, which can be evaluated based on its alleles, varied greatly for all tested SSR loci - from 0.26 to 0.76 with an average of 0.53 (Table 6). The highest PIC value 0.76 was obtained for CAP_SSR1 followed by CAP_SSR20 (0.69), CAP_SSR14 (0.67) and lowest PIC value was recorded for CAP_SSR5 marker (0.26) (Table 4.3). A genetic distance-based analysis was carried out by calculating the shared allele frequencies among the ninety-six *Capsicum* accessions. The gene diversity ranged from 0.308 (CAP_SSR5) to 0.719 (CAP_SSR1) with averages of 0.513. The major alleles frequency for each locus varied from 0.8177 (CAP_SSR7) to 0.307 (CAP_SSR1) with the average of 0.567. The highest heterozygosity (0.667) was detected at loci CAP_SSR14 followed by CAP_SSR6 (0.625), CAP_SSR1 (0.583), CAP_SSR16 (0.567), while the lowest was found for two loci namely CAP_SSR18 and CAP_SSR8 (0.0289) followed by loci CAP_SSR4 (0.0313) and CAP_SSR11 (0.0423).

The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) cluster analysis was performed using Nei's genetic distance (1972) with PowerMarker (3.23) and the Mega 4 software (Tamura et al., 2007) and observed that 96 *Capsicum* germplasm formedgroups in a species specific manner, ie. Genotypes belonging to same species clustered together in one group. All the *Capsicum* accessions clustered into four main clusters (Cluster I, Cluster II and Cluster III and Cluster IV; Figure 4.6). The cluster I comprises *C. frutescens* accessions from various locations of North-Eastern part of India such as Assam, Meghalaya, Nagaland and rest of the Indian states. However, the *C. frutescens* genotype 'Acc 87' grouped under this cluster is morphologically dissimilar in relation to fruit shape and orientation (round shape, pendant fruit), in contrast to the slender shape and erect orientationof *C. frutescens* fruit. The cluster II consists of highly pungent *C. chinense* genotypes that has pendant fruit position with diverse fruit colors such as dark red, orange, chocolate, and of various fruit shape ranging from conical to ovate. Generally, *C. chinense* fruit position is pendant. The Acc 34 placed in separate subgroup of *C. chinense* has distinct characteristics such as light green fruit color (in contrast to red or orange color commonly found in *C. chinense*) with elongated/slander shape (Figure 4.6).

The cluster III represents *C. annuum* accessions with low pungency level compared to other two species. Diverse fruit color was observed in the germplasm analyzed which ranged from red to yellow, and fruit shape varied from round to elongate. This cluster consists of

genotypes collected from different geographical locations such as diverse states of India for example Jammu & Kashmir, Uttar Pradesh, Assam, Mizoram, Meghalaya and other countries such as Canada. Cluster IV contains five uncharacterized accessions separated from the three main clusters. Since we were not able to confirm to which species these genotypes i.e. Acc 51, Acc 7, Acc 49, Acc 64, Acc 66 belongs, we have categorize them as other *Capsicum* species. They have distinct characteristics like round or slender fruit shape, low pungency and less seed count (6-12).

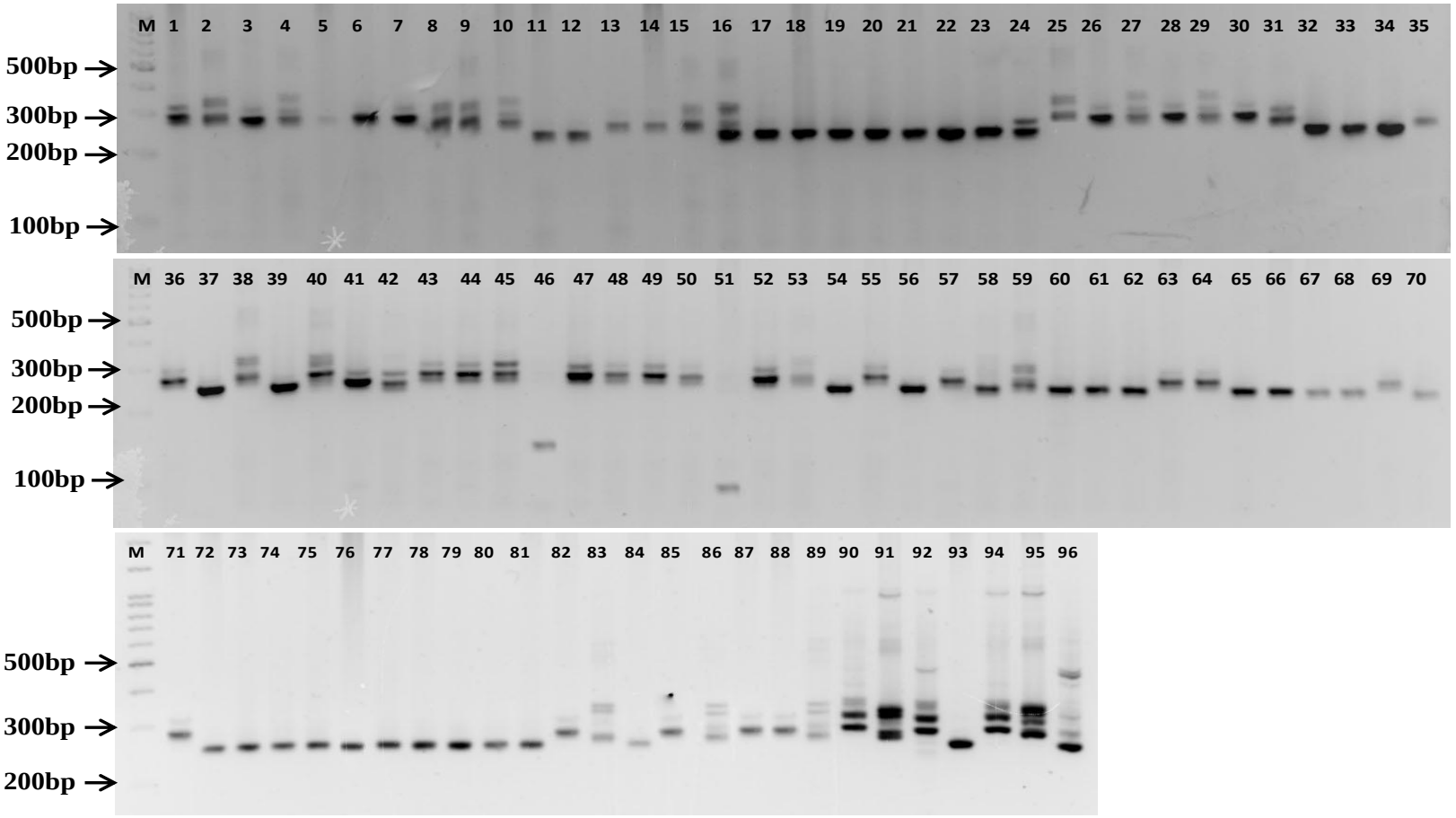


Figure 4.5 Representative gel picture showing allelic variation of CAP_SSR9 marker in 96 *Capsicum* genotypes in metaphor agarose gel electrophoresis. *C. annuum*: Lane 1-4, 6, 8-10, 13-16, 25, 27, 29-31, 35-42, 43-45, 47-48, 50, 52-55, 57-59, 63, 67, 69, 82-83, 89-92, 94-95; *C. chinense*: Lane 5, 11, 17-22, 33-34, 46, 56, 60-62, 68, 70, 72-81, 84-85; *C. frutescens*: Lane 12, 23-24, 26, 28, 32, 65, 71, 86-88, 93, 96; Other *Capsicum* sp.: Lane 7, 49, 51, 64 and 66; M: 100bp ladder.

Table 4.3 Characteristic of SSR primer pairs/markers used for amplification and diversity analysis of 96 *Capsicum* germplasm belonging to *C. chinense, C frutescens* and *C. annuum.*

Marker	Forward Primer Sequence	Reverse Primer Sequence	Motif	Size Range	Major Allele Frequency	Allele No	Gene Diversity	Heterozygosity	PIC
CAP_SSR 1	TGCTTGGAGTTTTCGATTGCAA	TATGCCGTCGGAACAGTAACAA	(TA)8	200-400	0.3073	5	0.7190	0.5833	0.7652
CAP_SSR 2	ACATGGCTGAGTTTCTTCTGTTC	ATCTAGCTCTCCGGTTTAGCTG	(TA)6	250-360	0.5052	3	0.6357	0.5208	0.5743
CAP_SSR 3	GGCAAGTTGTGGTATTGCACTT	ACTTATGTGAACTCCAACTCCA	(GAT)5	250-350	0.6979	3	0.4516	0.4375	0.3914
CAP_SSR 4	TACCCTCCCCTTGTATACTGCA	GAGAAGGAGATTGAGCTGGAGG	(CCCTGT)5	300-380	0.5000	5	0.5861	0.0313	0.6019
CAP_SSR 5	ATTCAGTGGTCCCTACCCTTTT	CGACTCTGATCCAAATATGCAC	(TC)10	300-400	0.8125	40	0.3084	0.3542	0.2668
CAP_SSR 6	TCGGGGTGTTCAGCTTTCTC	CACTACAAGAGCTGCCGGAA	(AGG)5	280-380	0.5729	3	0.5610	0.6250	0.6557
CAP_SSR 7	GGGAACTAGAAAGGGAGTTATCAA	GCTGTTAGTGATGTTGAGGAGA	(AGAA)5	150-250	0.8177	4	0.3161	0.2083	0.2957
CAP_SSR 8	GGATCTGGCTGCTTTCTTCTTA	GAGAACTTTGGTTCCATCTGCT	(GGA)6	200-300	0.6354	4	0.5171	0.0289	0.5242
CAP_SSR 9	AAACCGCACTGACTCACTGT	GGCCGGCGACATATAAGGAA	(ATA)5	280-380	0.5313	5	0.6228	0.4792	0.6367
CAP_SSR 10	AAGAGCTACAGATTCAACGGCA	CATCCACGTCAGCTTTCAGTTG	(ACG)7	280-320	0.7168	4	0.3412	0.2162	0.3144
CAP_SSR 11	ATACTCTGCCCAACTCAAACCG	CAGAATCATCCGCATCCACAAC	(GCC)7	380-420	0.4800	5	0.5272	0.0423	0.4915
CAP_SSR 12	TGACAAAGTGTGGCTTCATCAG	GCCATCGTCTATTCTTTCCAGA	(TA)7	330-360	0.6234	3	0.4245	0.4284	0.5725
CAP_SSR 13	TCCAAACCCAGTAATCCATACC	AGTAGGCATCCGAATAGGTTCA	(CCG)7	290-350	0.5212	6	0.6141	0.5017	0.5121
CAP_SSR 14	CGTTGATAAGGTCTCGATGAAA	AAAGAAAGGTCTCTCCGCATTA	(CTG)7	300-400	0.5216	5	0.4618	0.6670	0.6736
CAP_SSR 15	AAGACGAAGATGAGGGAGATGA	AAAAGCTAGAATCGACAATCGC	(TGA)6	350-450	0.6979	4	0.4516	0.4375	0.3914
CAP_SSR 16	GTTGTGTGTGTGTGCTGCTC	TTAAGCTCTCCGGGCACTTC	(AGC)5	280-400	0.4623	5	0.5135	0.5675	0.5676
CAP_SSR 17	AGCTTCTGGTGGGATGATGC	CCTCCATCACCACCATGACC	(AGC)6	360-420	0.4023	6	0.6142	0.4241	0.5815
CAP_SSR 18	TATAGTTTCGCTGGTGGGCA	TGGAACCAGAGGCAGATTCA	(AATA)5	250-350	0.6354	4	0.5171	0.0289	0.6524
CAP_SSR 19	ACCTTCACTTTGAGTCCATC	CAAAAGCCTTTCCAATACCG	(TGTT)7	250-330	0.3819	7	0.4679	0.4790	0.5824
CAP_SSR 20	ACACATTTGAGTCCATGTTCACA	GGGAGTTTGGTGTTGTAGTTGC	(CCCCA)5	280-320	0.5345	6	0.6127	0.4241	0.6943
Mean					0.5679	4.5500	0.5132	0.3743	0.5373

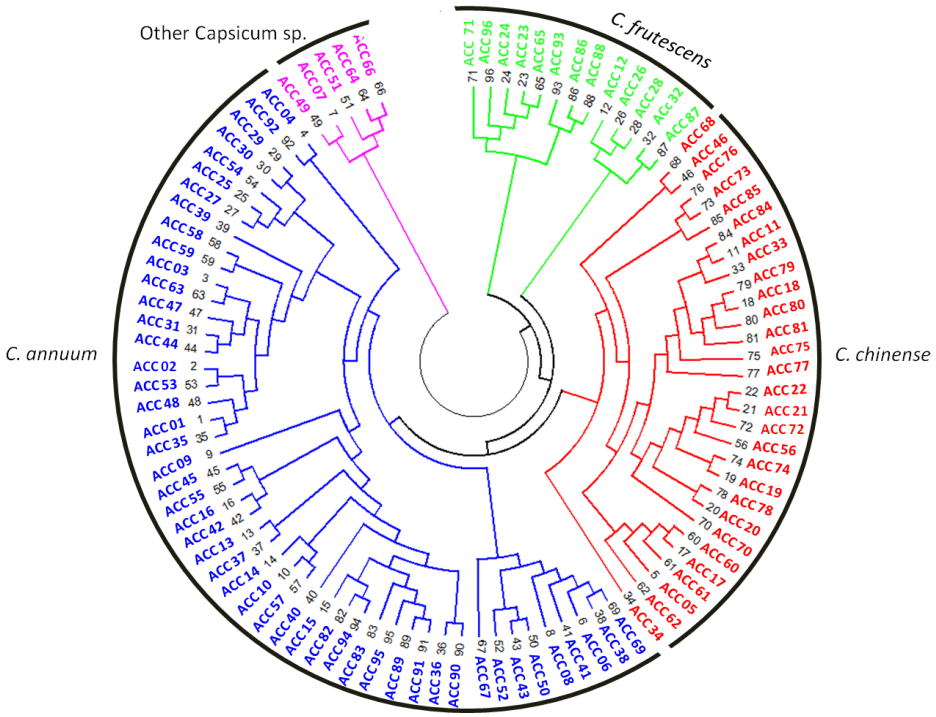


Figure 4.6 Dendrogram generated by UPGMA clustering analysis demonstrating genetic relatedness of 96 chilli pepper accessions. Green color represents cluster I of *Capsicum frutescens* group, Red color represents cluster II of *C. chinense* genotypes, Blue color represents Cluster III of *C. annuum* genotypes, and pink *color* signifies Cluster IV with other *Capsicum* species (species not confirm).

3.4 *In silico* polymorphism analysis between *C. chinense* and *C. frutescens* and validation of SSR markers

To identify the polymorphic SSRs (repeat motif number differences) between *C. chinense* and *C. frutescens* for the development SSR-based inter-specific genetic map, identical SSR markers (with same repeat sequence and genomic location both in *C. chinense* and *C. frutescens* in the *C. annuum* reference genome) were aligned using stand alone BLASTN. All the SSRs for which primers were designed i.e. 4989 *C. chinense* and 4781 *C. frutescens* primer pairs were used to identify polymorphic SSRs *in silico*. A total of 1,123 SSR motifs found to be common in both the species were tested for identification of polymorphism between the two species. The *in silico* analysis revealed a total of 337 polymorphic SSRs between the *C. chinense* and *C. frutescens*. The list of these 337 polymorphic SSRs along with primer sequences are given in Annexure 2. Additionally, in our *in silico* analysis we retained only such SSR motifswhichhas size differences of minimum of 2-12 bp. Out of 337 *in silico* polymorphic SSR markers, we selected 9 SSR marker showing

allelic size differences of more than 4 bp (between *C. chinense* and *C. frutescens*) so that polymorphism could be identified through electrophoresis analysis. The PCR amplification was detected for all these nine SSR markers however, some marker resulted in larger amplicon product size than expected, while some markers could not able to demonstrate polymorphism with gel electrophoresis possibly because of small difference in product size (average 5 bp variation)/genuine lack of polymorphism. For further validation, we have amplified and sequenced one SSR motif (CFpSSR3) which showed polymorphism between *C. chinense* and *C. frutescens* in *in silico* analysis (Fig. 4.7). The sequencing confirmed the InDel region of SSR. Further this polymorphic CFpSSR3 is validated in a panel of genotypes comprised of *C. chinense*, *C. annuum* and *C. frutescens* by PCR. The result of metaphor-agarose gel electrophoresis corroborates the expected 6 bp product size difference (allelic variation) between genotypes belonging to *C. chinense* and *C. frutescens* (Figure 4.8).

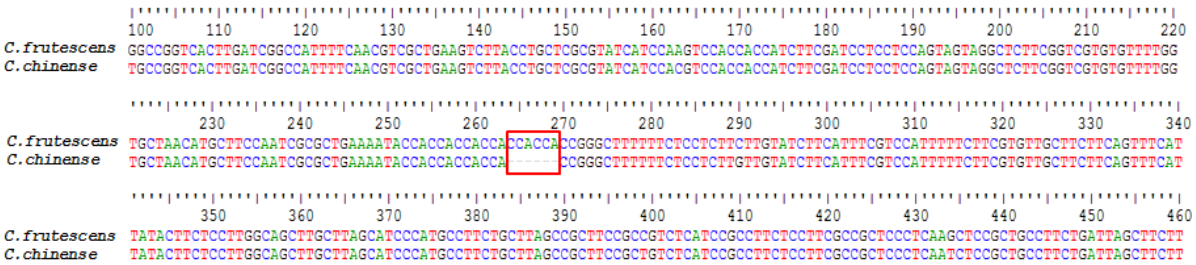


Figure 4.7 Sequence alignment of the CFpSSR3 microsatellite marker showing deletion ofrepeat motif in *C. chinense.* The highlighted red color box shows presence of repeat motif in *C. frutescens* while absence of motif in *C. chinense*

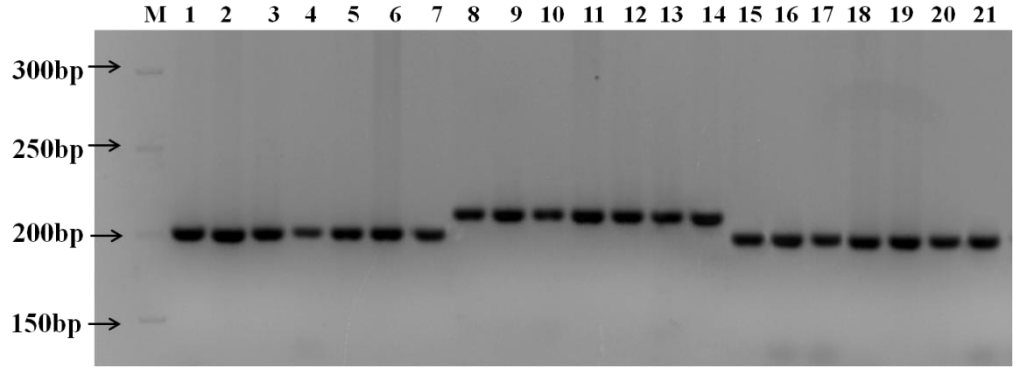


Figure 4.8 Allelic variation detected among 21 genotypes of *Capsicum* germplasm for microsatellite CFpSSR3 marker. M: 50 bp ladder; Lane 1-7: *C. annuum* genotypes Acc 16, Acc 42, Acc 92, Acc 83, Acc 3, Acc 4, Acc 38; Lane 8-14: *C. frutescens* genotypes Acc 23, Acc 65, Acc 32, Acc88, Acc 96, Acc 12, Acc 24; Lane 15-21: *C. chinense* genotypes Acc 17, Acc 22, Acc 80, Acc 72, Acc 84, Acc 85, Acc 61.

4. Discussion

The expanding transcriptome databases resulting out of functional genomics projects have led to the development of extensive gene based functional moelcularmarkers and are being exploited for mapping and identfication of genetic loci, candidate gene analysis, for enhancement of crop yield, quality and stress resistance traits (biotic and abiotic), and for and syntheny analysis among the crop genomes (Dutta et al., 2011, Khajuria et al., 2014, Zhang et al., 2014, Ahn et al., 2014; Tan et al., 2015; Buso et al., 2016). Among these, mining and development oflarge scale gene based microsatellite markers have become not only faster but also very cost-effective. The genic-SSRs are efficiently utilised and several studies have suggested their crucial involvement in gene regulation, DNA repair, chromatin organization, adaptation to various biotic and abiotic stress and hence they would greately helps in marker-trait association (Li et al., 2004, Varshney et al., 2005). Furthermore, these markers efficiently complements the genomic SSRs in their use in linkage mapping to enrich genetic maps to find the co-localization of QTLs and candidate genes in crop plants. However, among Solanaceae crop plants, only recently studies of discovery of transcriptome based SSR are being reported in addition to EST-derived SSRs in mostly *C. annuum* (Nicolai et al., 2012; Lu et al., 2012; Shirasawa et al., 2013; Ahn et al., 2014; Tan et al., 2015).Therefore the present study was designed to develop gene based SSR markers form transcriptome sequences of Bhut jolokia, the hottest naturally and the *C. frutescens* with medium pungency chilli peppers of North east India, as till now no reports are available of marker development and genetic mapping in utilizing these untapped genetic resources.

Of the total transcriptome sequences, in this study, we identified a total of 12,473 and 11,835 SSR motif`sin *C. chinense* and *C. frutescens* with the frequency of one SSRsper 7.52 kb in *C. chinense* and1 in 7.42 kb of *C. frutescens*. Although it is relative estimation and may differs from crop to crops, our finding is comparable with the frequencies observed in barley (1 per 6.3 kb; Thiel et al., 2003) and soybean (1 per 8.1 kb; Cardle et al., 2000). However, this frequency is higher than sugarcane (1 per 10.9 kb; Parida et al. 2010), maize (1 per 8.1 kb), Arabidopsis (1 per 13.83 kb), tomato (1 per 11.1 kb), poplar (1 per 14.0 kb), and cotton (1 per 20.0 kb) (Cardle et al. 2000); and smaller than wheat (1 per 5.4 kb; Pen et al. 2005) and pearl millet (1 per 1.75 kb; Senthilvel et al., 2008). The distribution, frequency and abundance of genic-SSRs may fluctuate due to several factors such as the number and size of the sequences taken for analysis, repeat length and SSR development tools (Varshney et al.,

2005; Poncet et al., 2006). Previeous studies also reported the observation of high frequency SSR in small genomes (Morgante et al., 2002), however, the frequency is found to be reduced in large genomes like *Capsicum*.

The SSR analysis revealed that the trinucleotide repeats are the most abundant (50.2% in *C. chinense* and 52.2% in *C. frutescens*) followed by di (46.46% in *C. chinense* and 45% in *C. frutescens)*, tetra-, penta- and hexa-nucleotide repeats in both the *Capsicum* species. Similar observations were made in earlier studies in pepper (Yi et al., 2006; Portis et al., 2007; Nicolai et al., 2012; Cheng et al., 2016). Furthermore, these observation was also made in previous studies in other crops, both monocot and dicot plants such as grape, barley, rice, wheat, citrus, cotton, soybean and flax (Scott et al., 2000; Thiel et al., 2003; Han et al., 2004; La et al., 2005; Chen et al., 2006; Hasona et al., 2007; Cloutier et al., 2009). In contrast to these observations, dinucleotide repeats were reported more in almond, spruce and cucurbit (Xu et al., 2004; Rungis et al., 2004; Gong et al., 2008) compared to trinucleotides repeats. The high abundance of trinucleotide repeats in the *Capsicum* transcriptome sequences suggests that their prevalence in coding region may prevent frame-shift mutations even when length variation occurs in trinucleotides (Metzgar et al. 2000).

Among trinucleotide motifs, the highest percentage (about 18%) was estimated for the AAC/TTG motif in both *C. chinense* and *C. frutescens* followed by AAG/TTC, ATC/TAG motifs. Similar obervations was reported in few studies of transcriptome derived SSR development in *C. annuum* where, AAC motif was detected as major fraction (Lu et al., 2011; Ashrafi et al., 2012; Xu et al., 2015). However, in another study of genome wide SSR identification in *C. annuum* showed that AAT motif is over-represented followed by AAC repeat motif (Chen et al., 2016), while the AAG motif was reported to be abuntantly found in other plants (Ueno et al., 2009; Siju et al., 2010; Durand et al., 2010; Joshi et al., 2011). On the other hand, among dimeric SSRs, AG/TC motifs were the most abundant (17.8% in *C. chinense* and 19.8% in *C. frutescens*), followed by AT/TA and AC/TG. Other studies also supported the predominance of AG/TC motif in plants such as coffee, cereals and forage crops (Temnykh et al. 2000; Thiel et al. 2003; Gao et al. 2003; Saha et al. 2004; Poncet et al. 2006) as well as perennial crops such as eucalyptus (Ceresini et al. 2005), apple (Newcomb et al. 2006), blackberry (Lewers et al. 2008), strawberry (Folta et al. 2005), citrus (Chen et al. 2006; Dong et al. 2006), oak (Durand et al. 2010) and cassava (Sraphet et al. 2011). AG/TC represents codons such as GAG, AGA, UCU, and CUC in mRNA population and translates into amino acids Arg, Glu, Ala, and Leu respectively. Since Ala and Leu are found in ample

amounts in proteins (Gao et al. 2003), this justifies the excess of AG/TC motifs in the genome (Joshi et al. 2011).

The strutural annotation analysis of all the identified microstellite repeats revealed that maximum number of SSRs belongs to intergenic regions of *C. chinense* and *C. frutescens* genome. The consistent observations are found in *C. annuum* (Cheng et al., 2016) and chickpea (Parida et al., 2015). The abundence of di-nucleotide (especially AT) microsatellite repeats in the intronic region of *C. chinense* and *C. frutescens* is also similar with the observations in the previous reports on *Arabidopsis* and rice (Lawson and Zhang, 2006; Parida et al., 2009). In our analysis, we found that, about 91 to 94% of the SSRs belongs to CDS region were trinucleotide repeat class. Our observation is in agreement with earlier reports on mungbean, cowpea, pigeonpea (Morgante et al., 2002; Gupta and Gopalakrishna 2010; Gupta et al., 2014; Dutta et al., 2014). The abundence of trinucleotide SSR repeat type in the CDS region as compared with other repeat types might be beacuse of the neccesity of coding region to preserve the reading frame of genes (Metzgar et al. 2000).

The identified SSRs from *C. chinense* and *C. frutescens* were validated and the genetic diversity among various chilli peppers accessions belonging to different *Capsicum* species mostly from North-Eastern India were studied. These chilli pepper accessions showed a wide range of diversity in the capsaicinoid content (Sarpras et al. 2016). We obsreved wide range of phenotypic variation in fruit color, weight and width, fruit position, fruit shape, fruit shape at blossom end. The observed phenotypic variability in these genotypes corroborates with the high level of genetic diversity detected in the present study with genic-SSRmarkers. Our result is in consistent with previous study where morphological (Gogoi and Gautam 2002; Bhagowati and Chngkija 2009) and biochemical diversity (Yatung et al., 2014) were reported among thepepper genotypes from North-Eastern India.

The 96 chilli peppers accessions belonging to different *Capsicum* species were subjected to genotyping with 20 polymorphic SSR markers to study the genetic diversity. The 20 SSR markers gave a total of 84 alleles in 96 genotypes, with an average of 4.6 alleles per SSR. The allelic richness per locus varied widely among the markers, ranging from 3 to 7 which is comparable with the earlier reports in *Capsicum* where on an average 2 to 6 alleles (Cheng et al., 2016) and 2 to 11 (Buso et al., 2016) was detected. However, in pigeonpea,on an average of 6.25 alleles per SSR locus was observed (Dutta et al., 2014). *Capsicum* SSR markers are highly informative and polymorphic in nature as depicted from the polymorphism information content (PIC) value. The PIC value is a measure of polymorphism

among genotypes for a marker locus used in linkage mapping analysis. The PIC value of each SSR marker, which can be evaluated based on its alleles, varied greatly for all tested SSR loci from 0.26 to 0.76 with an average of 0.53. This is slightly less than previously observed in *Capsicum* which were 0.60 (Buso et al., 2016) and 0.70 (Ibarra-Torres et al., 2015). However, our PIC result is comparable with recent study on chilli pepper diversity where, PIC value for each SSR marker was ranged from 0.05 to 0.64 (Cheng et al., 2016). According to the parameters recommended by Botstein et al., (1980), the marker with a PIC value above 0.5 is believed to be highly polymorphic and considered to be high useful. Hence, the set of *Capsicum* SSRs reported here will be useful for germplasm characterization, apart from molecular mapping of important traits, of both cultivated chilli peppers as well as their wild relatives and might contribute to the understanding of the genetic relationships among *Capsicum* accessions, wild taxa, or both.

The ability to use the same SSR primers in different plant species depends on the extent to which the primer binding sites flanking SSR motifs are conserved between related species. Due to high level sequence conservation in transcriptome/coding region, the genic-SSR markers are found higher transferable across species compared to genomic-SSRs (Cho et al., 2000; Eujayl et al., 2001; Chabane et al., 2005). The cross-species transferability of SSR markers in *Capsicum* was found to be high, with high percentages of loci producing amplicons in all tested species. This indicates that a high level of sequence conservation exists within the primer regions and also closely relatedness of these *Capsicum* species.

The UPGMA based cluster analysis using 20 highly polymorphic genic-SSRs grouped the 96 chilli pepper accessions into 4 major clusters, 3 major (many accessions) and 1 minor (few accessions) according to their gene pool/species i.e. Cluster I comprises of *C. frutescens* accessions, cluster II comprises of *C. chinense*, cluster III consists of *C. annuum* accessions, and cluster IV five unassigned accessations designated as (other) *Capsicum* species. The UPGMA analysis clearly classified the Bhut jolokia (*C. chinense*) genotypes from non-Bhut jolokia (*C. annuum* and *C. frutescens*). Except one accession (Acc. 49) of *C. chinense*, all accessions were grouped in cluster II of dendogram. The observation of several accessions from Assam, Manipur, Nagaland, and Meghalaya grouping in the same cluster (II) in dictates that they are genetically similar despite growing in different regions. Similarly, *C. annuum* accessions were collected from different parts of the country, they grouped into same cluster III suggesting genetic relatedness. The high genetic diversity observed in the Bhut jolokia and *C. frutescens*, may be attributed to a number of factors such as cross pollination, selection and

adaptation to varied microclimatic conditions together with its long history of cultivation in North-East India.

The developed SSRs in the present study from *C. chinense* and *C. frutescens* will be useful in mapping and identifying important QTLs/genes for economically important traits from *C. chinense* and *C. frutescens* in biparental QTL/gene mapping or association mapping. After, QTL/gene identification, linked SSRs may be utilized in marker-assisted selection to transfer desirable QTL/gene allele. Furthermore, these genome wide SSR markers can also be used for genomic selection to transfer desirable phenotypes from one genotype to another or form one species to another due to their high transferability. Furthermore, the genetically diverse *Capsicum* germplasm identified from the North East India could be used in future *Capsicum* improvement programme.

CHAPTER - V

Physiological and transcriptomic study of heterotic intra-specific and inter-specific *Capsicum* hybrids

1. Introduction

The world's annual production in 2014 for dry chilli was 3.446 million tons and for fresh green chilli was 31.11 million tons (FAOSTAT, 2016). Despite India is the world's largest producer, consumer and exporter of chilli peppers accounting for 36% share in total dry chilli production of the world (FAO 2015), the yield potential is very low due to poor yielding varieties and high incidence of pests and diseases. One of the ways to achieve quantum jump in yield and quality is by developing superior heterotic hybrids and identification of genes responsible for heterotic performance. Heterozygosity and inter-genomic interactions in diploid plants might result in phenotypic variation and increased growth vigor, a phenomenon referred to as heterosis (Ni et al., 2009). Heterosis/hybrid vigour is an event in which hybrids exhibit superior phenotypes, such as enhanced biomass production, development rate, grain yield, stress tolerance, relative to their parents. Although heterosis has been widely exploited in crop plants including *Capsicum* species, the study at molecular and genetic mechanisms underlying the superiority of F_1s over parental lines are still poorly understood, especially in *Capsicum* crops. Genetic models to explain the increased yield of hybrids consider the interaction of alleles at many loci generating altered expression levels and patterns of gene expression (Lipman and Zamir 2007; Charlesworth and Willis 2009). Whole-genome transcriptomes of parents and hybrids have been analyzed in a number of species (Hochholdinger and Hoecker, 2007; Birchler, 2010). In maize, different gene actions such as additive, dominance, and over dominance was collectively observed as basis of hybrid vigour in F_1 hybrids relative to both the parents (Swanson-Wagner et al., 2006). Similarly, the altered gene expression compared to parental genotypes in several tissues and various developmental stages in super hybrid rice LYP9 was found to be responsible for heterosis (Wei et al., 2009). Genes with altered expression are overrepresented in processes such as energy metabolism and transport, but no consistent pattern of altered gene expression has emerged, with many genes having expression levels up- or down-regulated in different tissues. In addition to whole genome transcriptome analyses, epigenetic states such as DNA methylation, small RNA production, and histone modifi cation have been found to be altered in hybrids (He et al., 2010; Groszman et al., 2011).

Furthermore, understanding the molecular basis of heterosis is important both for its application in crop improvement and for understanding plant evolution and adaptation. An

attempt has been made in the current study to understand the cause of early seedling heterosis at whole transcriptomes, physiological and metabolite level in intra (*C. chinense* X *C. chinense*) and inter-specific (*C. chinense* X *C. frutescens* and *C. annuum* X *C. chinense*) heterotic F_1 *Capsicum* hybrids. The F_1 hybrids were generated by crossing genetically distinct/contrasting genotypes belonging to different *Capsicum* species of North-East India. The most pungent chilli, Bhut Jolokia (*C. chinense*) was used to develop intra and inter-specific hybrids.

2. Materials and Methods

2.1 Plant materials, development of F_1 hybrids and growing conditions

The experimental material used in the present investigation is F_1 hybrid and its parents. A total of six F_1 hybrids were developed by crossing six contrasting genotypes of chilli pepper from different *Capsicum* species. Before crossing, the genotypes were self pollinated for 3-4 generations to make homozygous although *Capsicum*s are known to be self pollinated. The genotypes used are:- 1) *C. annuum* (Dudu), 2) *C. chinense* (Lota bhut), 3) *C. chinense* (Acc. 7), 4) *C. frutescens* (Acc. 4), 5) *C. chinense* (Chocolate bhut) 6) *C. chinense* (Umorok).All the genotypes belonging to three species were grown in glass house of School of Life Sciences, at Jawaharlal Nehru University, New Delhi during summer, 2014. The cultivation practice and growth conditions of these accessions and F_1 hybrids are mentioned in chapter 1 section 2.1.

Genotypes designated as female parents were crossed to each of the male parents by hand pollination method. The healthy buds from new flower flush, suitable to be open on the next day, were selected for emasculation and pollination. The selected buds were hand emasculated using forceps in the evening hours between 4 pm to 5 pm and covered with butter paper bags to prevent pollination with undesired pollens. Pollination of the emasculated flowers was carried out next day morning between 8 am to 10 am. Fully opened flowers with dehisced anthers were collected from the male parents and the stigma of female parent was dusted with the respective dehisced anthers of male flowers and covered with butter paper bag. The crossed flowers were tagged depicting name of the cross and date of pollination. After the fruit is seen forming, the butter paper bag is removed to allow the young fruit to grow. At maturity, red ripened fruits were harvested and sun dried. Seeds were extracted manually from the fruits, sun dried and stored. These seeds were obtained for evaluation in the next season. The self seeds of the parents were also collected during the same seasons. Following this method, a total of 6 hybrids, 5 interspecifc and one intraspecifi c hybrids were produced. The 5 inter-specific hybrids are :- A) *C. annuum* (Dudu) x *C. chinense* (Lota bhut), B) *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) C) *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4) D) *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) E) *C. chinense* (Umorok) x *C. frutescens* (Acc. 4),and one intra-specific hybrid is from *C. chinense* (Lota bhut) x *C. chinense* (Chocolate bhut) cross.

2.2 Measurement of quantitative characters

The measurements were recorded on randomly taken five competitive seedlingplants from each entry and replication followed by computing their means for the following traits.

A. Plant height (cm)

Plant height was measured in centimetres from the base to the top of the central apical shoot of the 30 day old seedlings of F_1 hybrid and parents in each entry over the replications.

B. Root length (cm)

Root length was measure in centimetres from the base to end of the primary roots of 30 day old seedlings of hybrid and parents and measured in 5 plants per replication. The average was taken as the final value.

C. Germination time (T_{50})

The time to 50% germination (T50) was calculated according to the following formula of Coolbear et al. (1984)

$$T50 = \frac{ti + [(N/2 - ni)\,(ti - tj)}{ni - nj}$$

Where, N is the final number of emergence and ni, nj cumulative number of seeds germinated by adjacent counts at times ti and tj, respectively when $ni < N / 2 < nj$.

D. Fruit length (cm)

We have randomly taken 10 fresh fruits from each F_1 hybrid and parents. The length was measured in centimetres using a scale and further average fruit length was recorded from total data of fruits.

E. Fruit weight (in gms)

The average fruit weight was estimated using the ratio between total weight of fruit for randomly selected 10 plants and number of fruits per plants.

F. Number of fruits per plant

The total number of fruits picked from 10 random plants at first harvest were measured.

2.3 Estimation of physiological and biochemical characters

A. Relative Water Content

Relative water content (RWC) of F_1 hybrid and parents was quantified according to Barrsand Weatherly (1968). Healthy and fully expanded fresh leaves were collected from seedlings of F_1 and parents seedlings (30 days old) and leaf discs of 1 cm diameter were prepared. Approximately, 120 leaf discs were made from each of the F_1 hybrid and respective parents and used for the experiment. The leaf discs were floated in 10 ml of water for 6 hours and allowed to become turgid. The time taken to turn into fully turgid was concluded by repeated weighing of leaf discs until there was no increase in weight. The fully turgid leaf discs were blotted in tissue paper to remove surface moisture and turgid weight was recorded. All samples were dried at 80°C for 24 hours and dry weight was recorded. The RWC (in percent) was calculated using the following equation. RWC was measured in three biological replicates for each sample.

$$\text{RWC (\%)} = 100 \text{ X } \frac{\text{Fresh weight - Dry weight}}{\text{Turgid weight - Dry weight}}$$

B. Total Chlorophyll Content

The total chlorophyll content was estimated in the seedling leaves of F_1 and parental lines according to the method of Arnon (1949). Fresh leaves were taken, washed, blotted in tissue paper to dry the external moisture and the leaf weight of the material was recorded by weighing in the balance. Approximately, 100 mg leaf samples were homogenized in a pre-chilled mortar using cold 80% acetone and tissue was pulverized completely. The homogenate was centrifuged at 3200 rpm for 30 min and the supernatant was collected. 80% acetone was taken as blank and the optical density of the acetone extract was measured at 645 and 663 nm in a spectrophotometer (Thermo Scientific). Total chlorophyll content was calculated from three biological replicates each replicates three times using the following formula:-

$$\text{Total chlorophyll} = 20.1 \text{ X O.D. (at 645 nm)} + 8.02 \text{ X O.D. (at 663 nm)}$$

C. Photosynthetic yield

For the measurement of net photosynthetic rate and transpiration rate, fully expanded matured leaves of F_1 and parents were taken, in infra-red gas analyzer (IRGA, WALZ, Germany) during the sunny days between 8 to 11 am in glass house. Assimilation chambers are used to enclose the plant or leaf, in the chamber CO_2 concentration was kept at 380 μLL^{-1} and air temperature was kept at 25°C. 15 minutes later, CO_2 concentration was controlled across a series of 1500, 1200, 800, 600, 400, and 200 μmol^{-1} to achieve steady state equilibrium. Later, leaves were pre exposed for 15 min at 1200 µmol photons m^{-2} s^{-1}. Data were measured after equilibration to a steady state (approx. 15 min), during which no signifi cant changes was observed on these parameters. The experiment was carried out three times in each of the three biological replicates.

D. Capsaicinoid content

Quantification of capsaicinoid content was performed by methods described earlier (Refer chapter 1 section 2.2)

2.4 Percent heterosis determination

Mid-parent heterosis (MPH) values was calculated to quantify the % increase of different F_1 hybrids in comparison to their respective parents for different traits following the formula of Geleta and Labuschagne (2004) as indicated below.

$$MPH = (F1 - MP)\ MP\ X\ 100$$

Better-parent heterosis (BPH) values was calculated in a similar manner as above using the value of the better-parent instead of the mid-parent

$$BPH = (F1 - BP)\ BP\ X\ 100$$

2.5 Global metabolite analysis

For non-targeted metabolic profiling, fresh leaves from each F_1 hybrid seedlings with respective parents were used. The GC-MS based metabolite profiling was performed by using the method described by Lisec et al., (2006). Metabolites present in all the 3 biological replicates were selected for further analysis.

2.6 RNA Extraction, Illumina Sequencing

The seedlings F_1 hybrids and parental lines were used for the RNA extraction. The method of total RNA extraction has been mentioned in Chapter 1 section 2.5b. The quality and quantity estimation was described in section 2.6 and 2.7. The cDNA libraries were prepared by using TruSeq cDNA sample preparation kit (Illumina, US). Briefly, the 3' and 5' adapters were ligated to total RNA followed by reverse transcription and amplification of the ligated product. The cDNA library was then purified and quality was checked by using gel electrophoresis (refer chapter 1 section 2.6) and bioanalyzer. After confirming the quantity and quality, the cDNA libraries were used for deep sequencing. Each cDNA library was sequenced using Illumina Genome Analyser II and preliminary analysis was conducted at Genotypic Technologies Pvt. Ltd, Bangalore, India.

2.7 Read Mapping and differential gene expression analysis

The quality of raw reads from each library was assessed by software FastQC (version 0.11.3). Further, the sequences were filtered and the adaptor sequences, low-quality tags, filter empty tags (reads with only 3' adaptor sequences without transcripts) sequences which are too short or too long are removed. The reference assembly of *C. annuum* Zunla-1 were used for the analysis (Qin et al., 2014). Reads were aligned to *Capsicum* genome using the default parameters for TopHat (version 2.09) (Trapnell et al., 2009; Kim et al., 2013), that provides sensitive and accurate alignment results for highly repetitive genomes. The expression level of each gene was anticipated by the occurrence of clean tags and afterward normalized to FPKM (Fragments Per Kilobase of transcript per Million mapped reads), which is a principal method and widely accepted in DGE analysis (Hoen et al., 2008; Morrissy et al., 2009). Gene ontology (GO), enrichment analysis was performed using AgriGo tool and significantly enriched terms were depicted by using REVIGO tool.

To study differential expression across samples (F_1 hybrids and parents), the high quality reads in every library was normalized to FPKM to attain normalized gene expression levels. Identification of different tags across samples was performed as described by Marioni et al., (2008). Differential expression analysis was performed using R packages of DESeq (Anders and Huber, 2010). Differential expression was assessed between the different F_1 hybrids relative to their parents, using Benjamini and Hochberg's statistical test. After multiple testing between pair-wise comparisons, FDR $\leq$ 0.001 and the definite value of

log_2Ratio $\geq$ 2 were used, as the threshold to determine the significance of gene expression variation. An additional stringent criterion with greater fold-change values and smaller FDR were used to discover different genes. In this present study, the same approach was followed in a linear-in-genotype contrast when F_1 genotype was compared to the two parental lines as described by Rapp et al., (2009) and Li et al., (2014). To further classify these genes, the high-parent dominant genes and the low-parent dominant genes were selected and categorized from the non-additive group following on the criteria, that "the F_1 genotype was significantly different from one parent and not significantly different from another parent". In the non-additive group, the expression of genes was recognized as under dominant or over dominant, when expression in the F_1 genotype was significantly lower or higher than both inbred parents, respectively.

2.8 Quantitative Real Time PCR

The qRT-PCR was performed with cDNA from F_1 hybrid seedling and its parent as describes earlier (Refer chapter 1 section 2.2.2)

3. Results

3.1 Development and confirmation of Intra- and Inter-specific chilli pepper hybrids

Several chilli pepper accessions of different *Capsicum* species such as *C. chinense,C. annuum* and *C. frutescens* collected from various part of the India were grown in glass house at controlled conditions. To develop intra- and inter-specific hybrids, we have studied a number of quantitative and qualitative traits of these genotypes. Among these, accessions showing contrasting characters such as capsaicinoid content, flowering time, fruit size & shape, average fruit weight, fruit color, plant height and chlorophyll content were selected for cross hybridization. We have developed about 22 different intra- and inter-specific *Capsicum* hybrids by cross hybridization of genotypes from three *Capsicum* species with multiple parental combinations. Among all these F_1 hybrids, based on desired agro-economic traits we have selected six hybrids for the study which has shown heterotic behaviour compared to their respective parents. The difference between parental genotypes and confirmation of F_1 hybrid was done by genotyping with one of the Conserved orthologous set (COS) of markers of Solanaceae C2At3g13700. This marker confirmed the F_1 hybrids two parents (Fig. 5.1), The intra-specific hybrid of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) was confirmed by phenotypic marker, color (*C. chinense* Lota bhut parent has red color fruit and *C. chinense* Chocolate has chocolate color, the hybridity of this cross was confirmed by observation of varying fruit colors i.e. chocolate to red, in the fruits of segregating F_2 plants, thereby confirming the hybrid). The selected F_1 hybrids i.e. 1) *C. annuum* (Dudu) x *C. chinense* (Lota bhut), 2) *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4), 3) *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia), 4) *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4), 5) *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4), and 6) *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) were used for analysis of agronomic, physiological, metabolite and transcriptomes to understand the molecular basis of heterosis (Fig. 5.2).

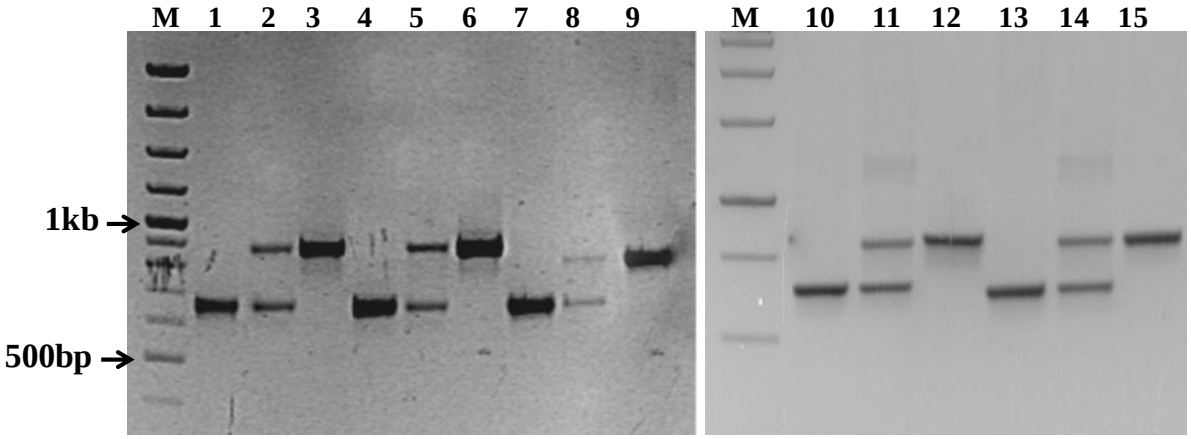


Figure 5.1 Agarose gel picture showing heterozygosity of F_1 hybrids representing both the parental genome with C2At3g13700 COS marker. M: 1kb ladder; Lane 1: *C. chinense* (Acc. 7), Lane 2: F_1 of *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4), Lane 3, Lane 6, Lane 12 and Lane 15: *C. frutescens* (Acc. 4),Lane 4: *C. chinense* (Chocolate jolokia), Lane 5: F_1 of *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4), Lane 7: *C. annuum* (Dudu), Lane 8: F_1 of *C. annuum* (Dudu) x *C. chinense* (Lota bhut), Lane 9 and Lane 10: *C. chinense* (Lota bhut), Lane 11: F_1 of *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4), Lane 13: *C. chinense* (Umorok), Lane 14: *C. chinense* (Umorok) x *C. frutescens* (Acc. 4).

3.2 Study of heterotic behaviour of *Capsicum*F_1 hybrids

The seeds of the six F_1 hybrids were germinated and kept for growing in glass house to test their heterotic performance over both of the parents. All the seedlings were grown in identical climatic conditions and after 30 days the heterotic behaviour were studied for several traits. We observed superiority of F_1 seedlings compared to both parents for time to 50 percent germination (T_{50}), plant height, plant vigour, root length, and leaf area (Fig. 5.3, 5.4 and 5.5). To estimate the heterosis in F_1 hybrids we used 5 plants for each trait.

Germination time (T_{50})

Seed germination is considered as one of the critical process in life cycle of plants. We observed remarkable heterosis in seed germination time of F_1plants over parents. In all of the intra- and inter-specific hybrid plants, the germination rate has found to be increased significantly as T_{50} (days to germination of 50% of all germinated seeds) of F_1 hybrids is significantly less as compared to parents (Fig. 5.5), except F_1 hybrid of (Chocolate jolokia) x *C. frutescens* (Acc. 4). We detected significant negative heterosis over mid parent and better parent in the case of F_1 such as *C. annuum* (Dudu) x *C. chinense* (Lota bhut), *C. chinense*

(Acc. 7) x *C. frutescens* (Acc. 4), *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) and *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) and *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) (Table 5.1). This negative heterosis is highly desirable for early plant growth.

Plant height

In the case of plant height, most of the F_1 seedlings showed positive mid parent and better parent heterosis except a cross of *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) (Fig. 5.3 and 5.5). Among F_1 hybrids, *C. annuum* (Dudu) x *C. chinense* (Lota bhut) resulted in highest mid parent heterosis (75 %), followed by *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) (50.6 %) and *C. chinense* (Chocolate Jolokia) x *C. frutescens* (Acc. 4) (35.6 %) (Table 5.1). While in one of the cross of *C. chinense* (Umorok) x *C. frutescens* (Acc. 4), negative heterosis has been observed over mid parent.

Leaf area

Large leaf area (in cm^2) was observed in all the F_1 hybrids over parents contributing towards vigourous growth / hybrid vigour of F_1 seedlings (Fig. 5.3). All the inter-specific F_1 hybrids exhibit positive mid parent heterosis for leaf area with highest percentage of heterosis (62%) observed in cross of *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) followed by49% in *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) (Table 5.1).

Root length

We detected superiority of F_1 seedlings over parents for root length trait in most of the crosses (Fig. 5.4). In all the intra- and inter-specific hybrids, root length has been recorded considerably high over the parents displaying positive mid parent and better parent heterosis(Table 5.1), except in the cross of *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) which showed negative mid parent heterosis. Over all, high-root vigour (such as length of primary root, number of lateral root, root diameter) has been observed in most inter-specific F_1 hybrids indicating the presence of heterosis behaviour in *Capsicum* hybrids. The highest mid parent heterosis (80.8 %) was observed in F_1 hybrid of *C. annuum* (Dudu) x *C. chinense* (Lota bhut) followed by *C. chinense* (Chocolate Jolokia) x *C. frutescens* (Acc. 4) cross (76.4 %). However, intra-specific hybrid *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) could not display high-root vigour compared to bothparents although, length of primary root

has been recorded high in F_1 plant (Fig. 5.4 and 5.5) resulting in positive mid parent heterosis (Table 5.1).

Seed number per fruit

Further, we analysed seed number per fruit of the F_1 plant. We observed negative mid parent and better parent heterosis in all the intra and inter-specific hybrids.

Fruit weight

In the case of fruit weight trait, negative heterosis was observed in all the inter-specific hybrids (Fig. 5.5). In general, it has been found that the fruits of the *C. annuum* and *C. chinense* have high fruit weight while *C. frutescens* has very low fruit weight. The crosses involving *C. frutescens* as one of the parent leads to F_1 plant with low fruit weight. Nevertheless, F_1 hybrid from intra-specific cross of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) showed positive mid-parent heterosis (Table 5.1).

Fruit length

Among all the hybrids, inter-specific cross of *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) showed highest mid parent heterosis (4.34 %), followed by 1.29 % in *C. annuum* (Dudu) x *C. chinense* (Lota bhut) cross. The intra-specific cross of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) also showed considerably positive mid parent heterosis (0.6 %) for fruit length (cm) (Table 5.1). In general, it has been observed that the fruits of the *C. annuum* and *C. chinense* have large fruit length while *C. frutescens* has very small fruit length. However, in case of *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) F_1 hybrid, negative heterosis was observed.

Fruits per plant

Fruit number per plant is an important quantitative trait for breeding purpose. We observed positive mid parent and better parent heterosis for all the hybrids of inter-specific cross (Fig. 5.5) (Table 5.1 and 5.2). The crosses involving *C. frutescens* (Acc. 4) as maternal parent, such as *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) showed highest percentage (107%) of better parent heterosis followed by *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4) (89%), *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) (71%) and *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) (65%) (Table 5.2). Negative mid parent heterosis has been

detected in intra-specific F_1 hybrid of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia).

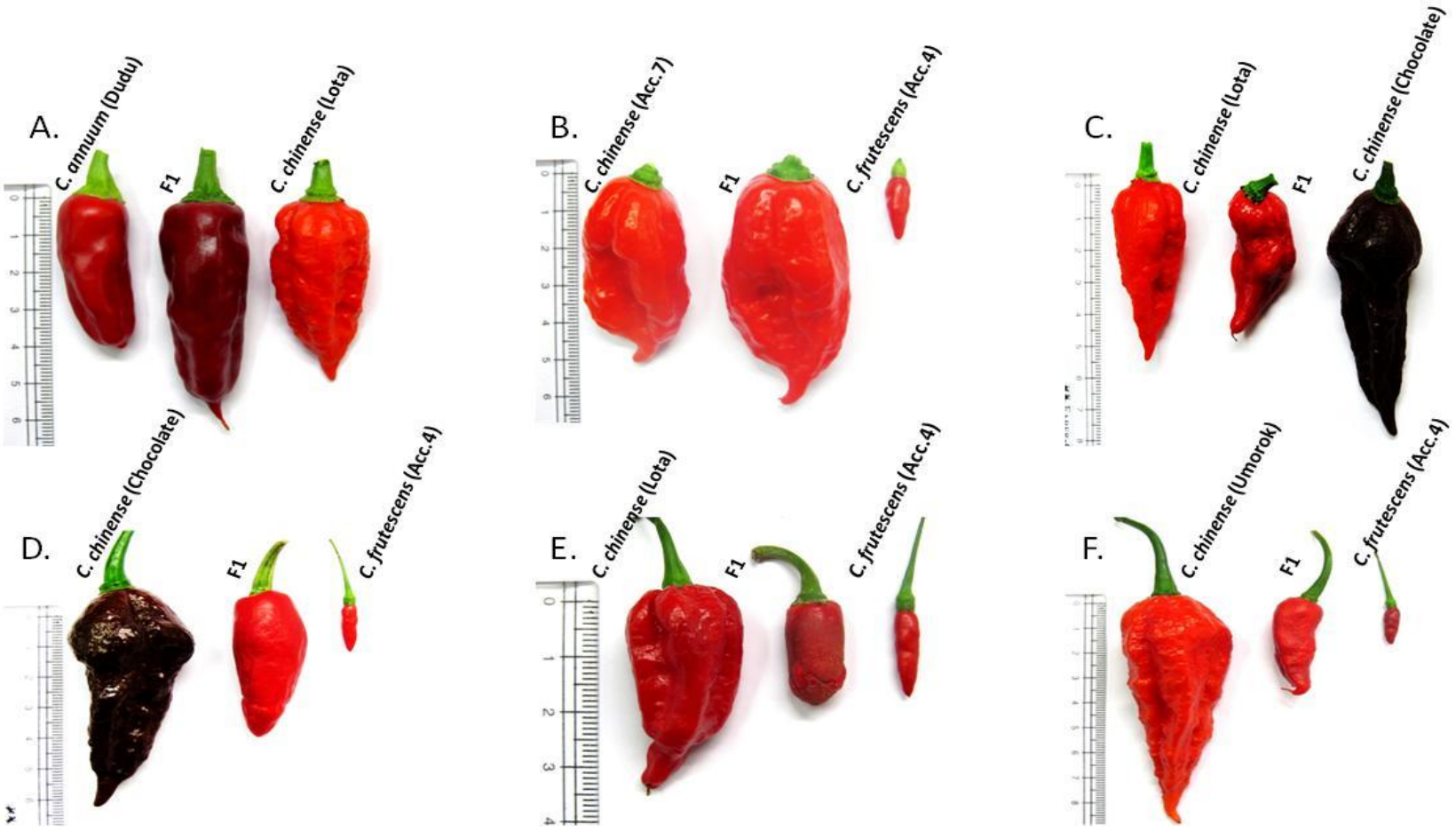


Figure 5.2 Fruit morphology of F_1 hybrids with parents in which A) *C. annuum* (Dudu) x *C. chinense* (Lota bhut), B) *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4),C) *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia), D) *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4), E) *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4), andF) *C. chinense* (Umorok) x *C. frutescens* (Acc. 4).

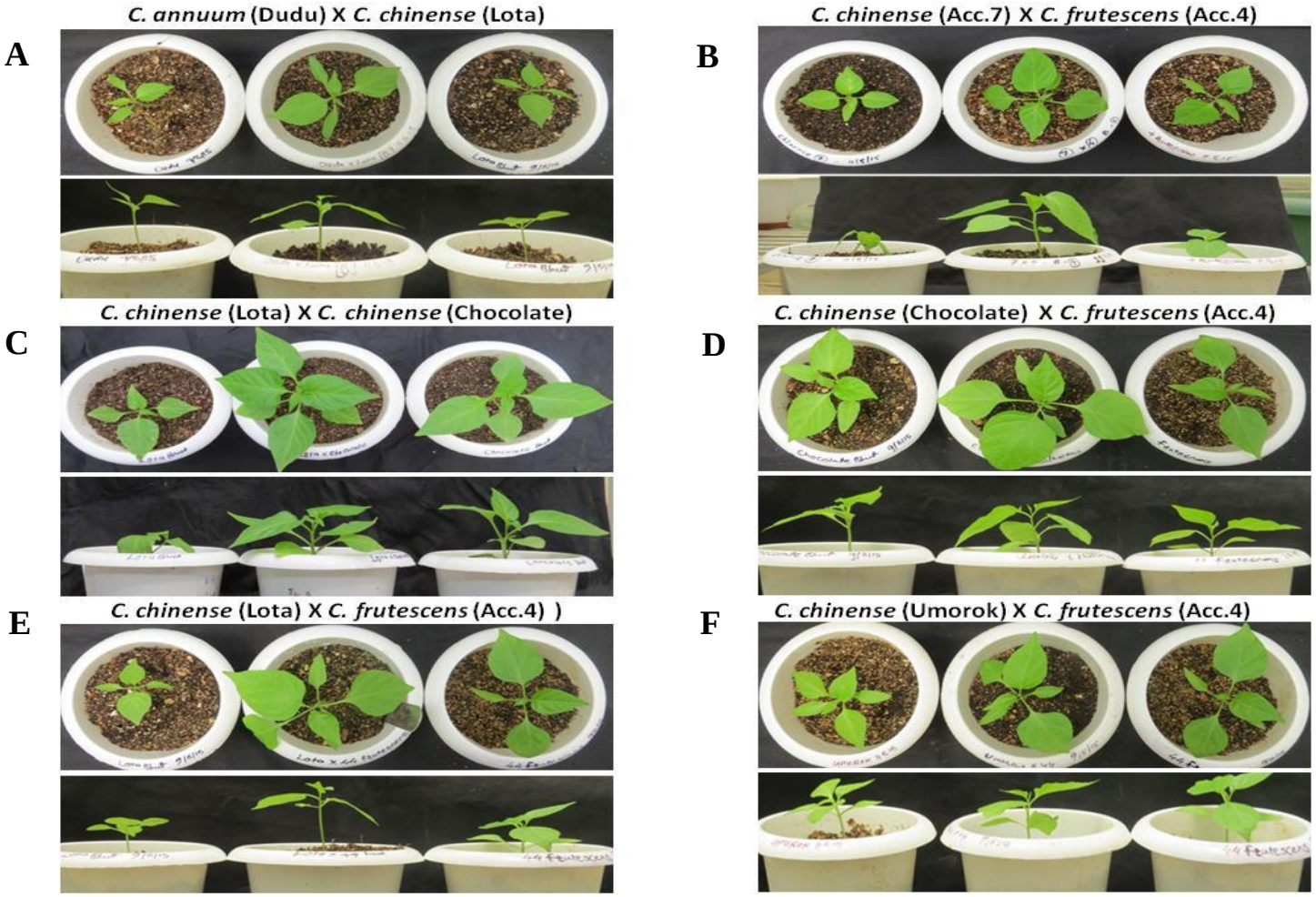


Figure 5.3 Heterotic behaviour of F_1 seedlings over parents :- A) *C. annuum* (Dudu) x *C. chinense* (Lota bhut), B) *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4), C) *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia), D) *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4), E) *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4), F) *C. chinense* (Umorok) x *C. frutescens* (Acc. 4).

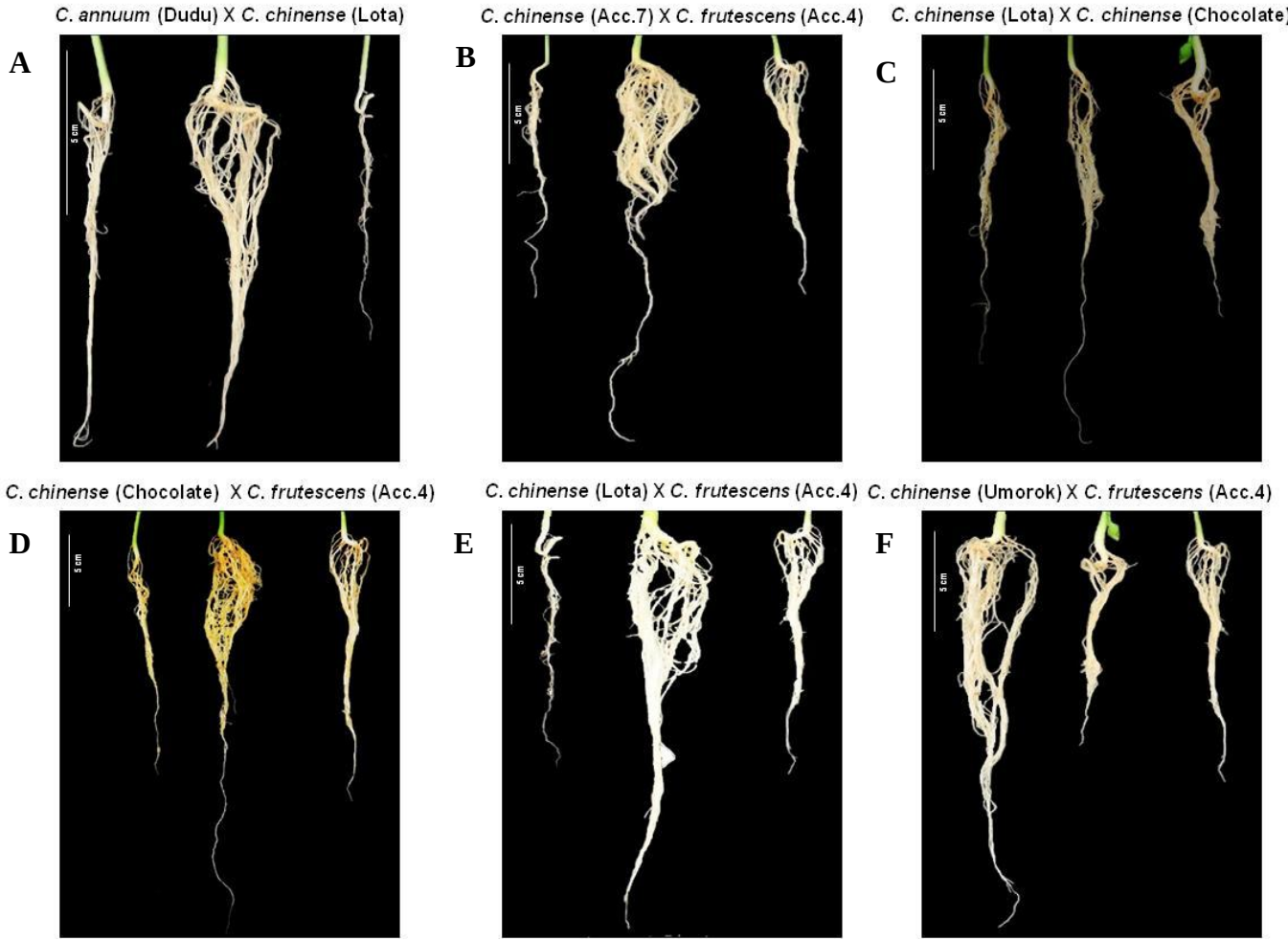


Figure 5.4 Heterotic behaviour of roots of the F_1 hybrids over parents :- A) *C. annuum* (Dudu) x *C. chinense* (Lota bhut)B) *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4),C) *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia)D) *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4)E) *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4)F) *C. chinense* (Umorok) x *C. frutescens* (Acc. 4).

Table 5.1 Percent heterosis of six F_1 hybrids over mid-parent and better parent for quantitative traits

	C. annuum **Acc. Dudu X** ***C. chinense*** **Acc. lota**					***C. chinenese*** **Acc. 7 X** ***C. frutescens*** **Acc. 4**					***C. chinense*** **Acc. Lota X** ***C. chinense*** **Acc. Choco Jol**				
Traits	**P1**	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**
Plant height	4.4	6.1	5.3	25.7	15.1**	5	7	3	75.0**	40.0*	3	6.25	5.3	50.6	17.9
Root length	20.5	32	15	80.2	56.1	13	14	11	16.6***	7.6**	11	16	15	23.1*	6.7**
Seed no. per fruit	17	9	4.25	-15.2	-47.1**	50	27	17	-19.4	-46.0**	17	10	4.25	-5.9*	-41.2**
Fruit weight	6.2	0.86	0.07	-72.5*	-86.1**	3.05	2.67	5.8	-39.6***	-53.9***	5.8	0.921	0.07	-68.6	-84.1***
Fruit length	4.9	3	0.85	4.3	-38.8	4.8	4.71	4.5	1.2	-1.8**	4.5	0.8	0.85	-70.1***	-82.2***
Fruit number	15	79	38	198.1**	107.9*	14	18	11	44.0***	28.5***	11	63	38	157.1***	65.8**
T50	8	5	9	-41.1***	-44.4***	10	3	14	-75.0**	-78.5**	14	3	9	-73.9***	-78.6***
Leaf area (cm2)	7.7	16	12	62.4	33.3	7.5	8.3	9	0.6**	-7.7*	9	15.7	12	49.5	30.8
	C. chinense **Acc. Choco Jol X** ***C. frutescens*** **Acc. 4**					***C. chinense*** **Acc. Lota X** ***C. frutescens*** **Acc. 4**					***C. chinense*** **Acc. Umorok X** ***C. frutescens*** **Acc. 4**				
Traits	**P1**	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**
Plant height	6.5	8	5.3	35.6**	23.1**	3	6.25	6.5	31.6**	-3.8*	5.43	5.05	5.3	-5.9**	-6.9
Root length	19	30	15	76.5**	57.9**	11	16	19	6.7	-15.8	22	12	15	-35.1	-45.4
Seed no. per fruit	30	6	4.25	-64.9**	-80.0***	17	20	30	-14.9	-33.3**	17	7	4.25	-34.1	-58.8*
Fruit weight	6.3	0.603	0.07	-81.1***	-90.4***	5.8	6.1	6.3	0.8***	-3.2***	5.11	2.2	0.07	-15.1	-56.9
Fruit length	5.5	3.2	0.85	0.8	-41.8**	4.5	5.03	5.5	0.6***	-8.5***	5.3	0.8	0.85	-73.9	-84.9
Fruit number	10	72	38	200.0***	89.5**	11	7	10	-33.3***	-36.4***	14	65	38	150.0*	71.1*
T50	6	8	9	6.7	-11.1	14	5	6	-50.0***	-64.3***	10	3	9	-68.4**	-70.0***
Leaf area (cm2)	15	17.25	12	27.8	15	9	11	15	-8.3**	-26.7*	11.5	13	12	10.6	8.3*

*Significant at 0.01level,
**Significant at 0.001 level
***Significant at 0.0001 level

Table 5.2 Percent heterosis of six F_1 hybrids over mid-parent and better parent for physiological and biochemical characters

	***C. annuum* Acc. Dudu X *C. chinense* Acc. lota**					***C. chinenese* Acc. 7 X *C. frutescens* Acc. 4**					***C. chinense* Acc. Lota X *C. chinense* Acc. Choco Jol**				
Traits	**P1**	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**
Chlorophyll content	10.9	16.6	14.1	32.1**	17.1**	14.7	18.9	15.6	24.5**	20.8*	14.1	17.5	16.3	15.0*	7.5
Transpiration rate	2.3	2.9	2.6	22.4*	15.6	2.5	3.3	2.7	25.4**	20.5**	2.6	2.9	2.3	19.5*	13.6*
Photosynthetic rate	6.5	10.0	7.9	39.1**	26.7**	8.1	11.0	8.9	29.0**	22.8*	7.9	10.8	9.9	21.8**	9.5
RWC	89.6	94.8	87.1	7.3*	5.8	87.2	95.3	88.1	8.8*	8.2*	86.1	93.1	89.7	5.9	3.8
Pungency (SHU)	7708	565903	953687	17.7	-40.7***	914271	598178	487501	-14.65	-34.6**	953687	752074	931948	-20.2**	-19.3**
	***C. chinense* Acc. Choco Jol X *C. frutescens* Acc. 4**					***C. chinense* Acc. Lota X *C. frutescens* Acc. 4**					***C. chinense* Acc. Umorok X *C. frutescens* Acc. 4**				
	P1	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**	**P1**	F_1	**P2**	**MP**	**BP**
Chlorophyll content	16.3	18.5	15.6	16.1*	13.7*	14.1	17.3	15.6	16.5*	11	12.32667	16.72	15.61	19.7*	7.1
Transpiration rate	2.3	3.0	2.7	19.9*	10.2	2.6	3.0	2.7	11.5	7.5	3.06	3.31	2.74	14.1***	8.2
Photosynthetic rate	9.9	10.8	8.9	15.1**	9.6*	7.9	9.8	8.9	16.3*	9.3	8.5	9.83	8.93	12.8*	10.07
RWC	89.7	94.6	88.1	6.3*	5.4	86.1	91.3	87.1	5.4	4.8	90.23	93.34	87.12	5.3***	3.4
Pungency (SHU)	931948	758767	487501	6.9	-18.6**	953687	574574	487501	-20.3*	-39.7**	944604	630888	487501	-11.9	-33.2***

Significance at level:- *- P<0.01, **- P< 0.001 and *** - P< 0.0001

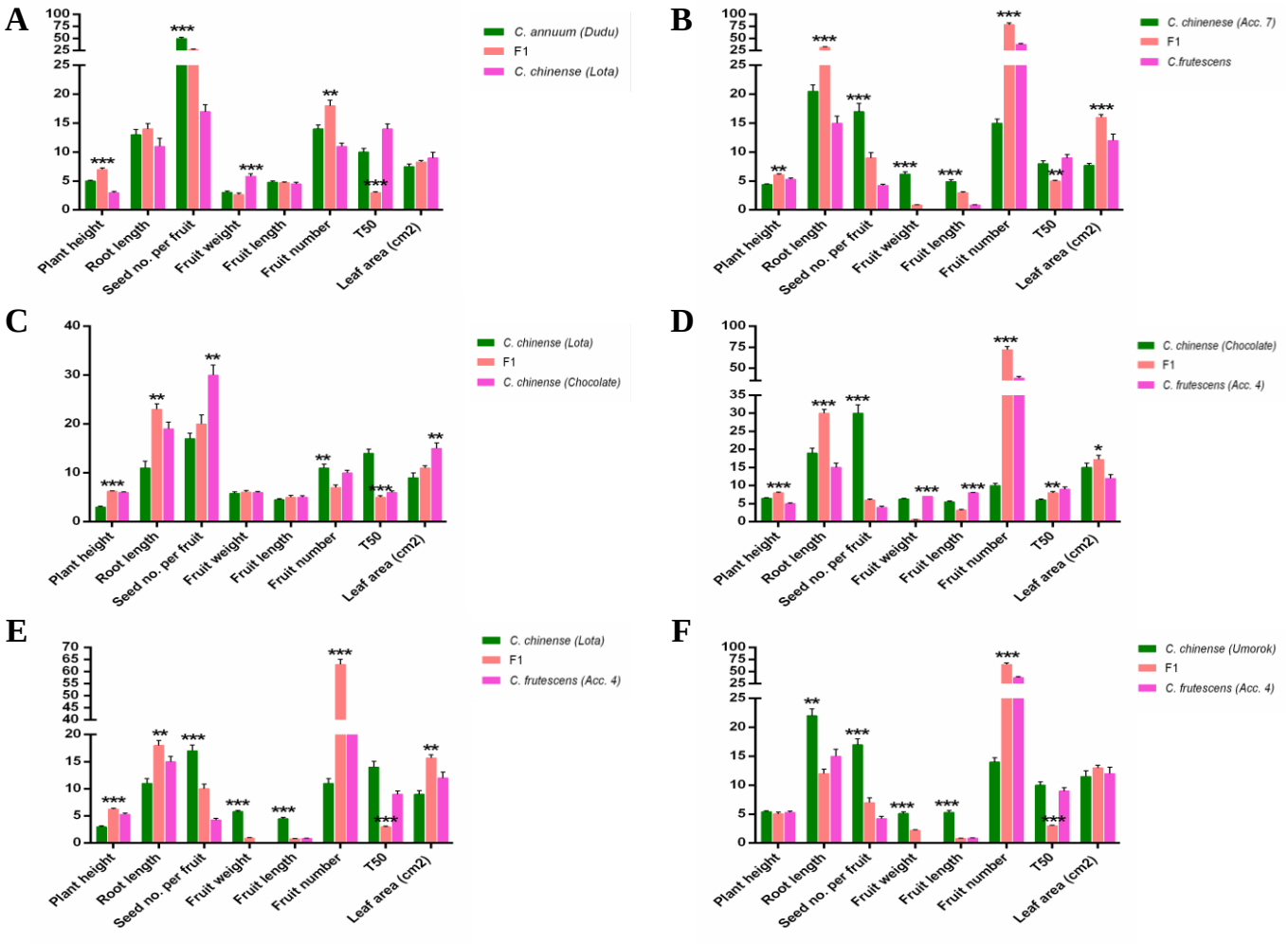


Figure 5.5 Quantitative traits observed in F_1 hybrids and parents

3.3 Physiological parameter analysis of F_1 hybrids and parents

To further understand the basis of heterosis, 30 day old seedlings of six F_1 hybrids along with respective parents were used for analysis of several physiological parameters. We analysed chlorophyll content, photosynthetic rate, transpiration rate, relative water content and capsaicinoid content of all the hybrids with their parents (Fig. 5.6) and (Table 5.2).

Chlorophyll content

To see whether chlorophyll content is one of the reasons for getting high vigour in F_1 seedlings over both parents, we measured chlorophyll content (mg/g^{-1} of FW) in leaves of F_1 seedlings and parents. In all of the intra- and inter-specific hybrids, significantly greater chlorophyll content was found compared to parents (Figure 6A). We observed positive mid parent and better parent heterosis in all the F_1 hybrids. Among them, inter-specific hybrids such as *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) showed highest BPH (20.8 %) followed by *C. annuum* (Dudu) x *C. chinense* (Lota bhut) (17 %) and *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4) (13.73 %) (Table 5.2).

Transpiration rate

Further, we have analysed the transpiration rate (mmol H_2O m^{-2} sec^{-1}) of all the F_1 seedlings and parents (Figure 6B). The transpiration rate has been found to be considerably high in almost all of the F_1 hybrids. The *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) showed highest BPH (20.53 %) followed by *C. annuum* (Dudu) x *C. chinense* (Lota bhut) (15.5 %) and *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) (13.59 %) (Table 5.2)

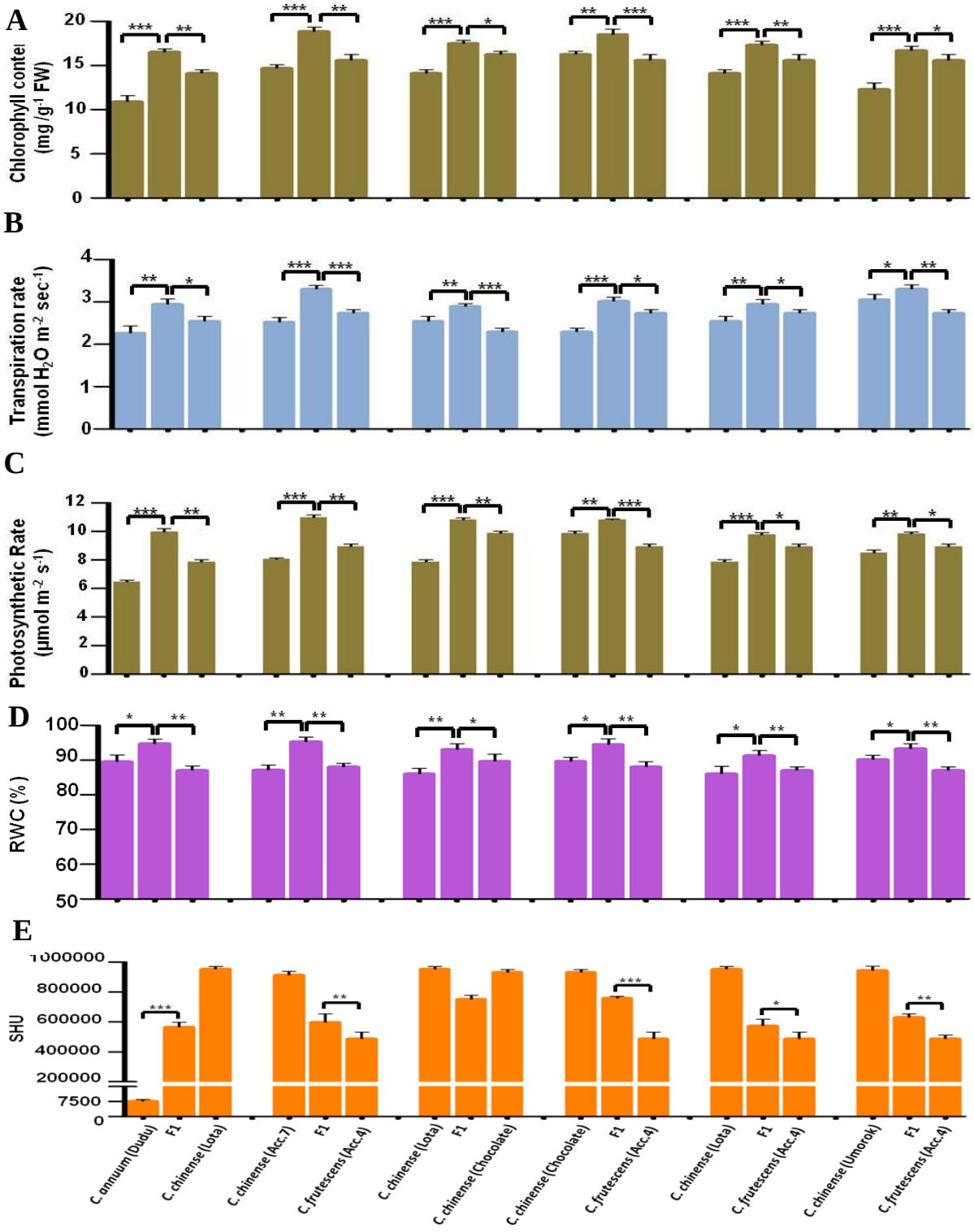


Figure 5.6. Physiological and biochemical analysis of six F_1 hybrids and parents. A) Chlorophyll content, B) transpiration rate, C) photosynthetic rate, D) relative water content, E) capsaicinoid content

Photosynthetic rate

The photosynthetic rate of F_1 seedlings were determined to investigate the photosynthetic activity of plants. We observed significantly increased levels of photosynthetic rate in F_1 hybrids (Fig. 5.6C). Among them, cross of *C. annuum* (Dudu) x *C. chinense* (Lota bhut) showed highest (26.6 %) BPH followed by *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) (22.76 %) and *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) (10 %) compared to other hybrid seedlings (Table 5.2).

Relative water content

Relative water content has been found to be increased in all the F_1 seedlings as compared with parents (Fig. 5.6D). The highest BPH was observed in cross of *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) (8.19 %) followed by *C. annuum* (Dudu) x *C. chinense* (Lota bhut) (5.8 %) (Table 5.2)

Capsaicinoid content

The capsaicinoid content of the fruits of F_1 plants were quantified to determine the pungency level of hybrids with respective parents (Fig. 5.6E). The capsaicinoid content has been converted in Scoville heat unit (SHU).We observed positive mid parent heterosis in only two inter-specific hybrids such as F_1 hybrid of *C. annuum* (Dudu) x *C. chinense* (Lota bhut) (17.7 %) and *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4) (6.9 %) (Table 2). However, hybrid vigour could not be detected for pungency content in remaining F_1 hybrids.

3.4 Global metabolite analysis of F_1 hybrid and parent using GC-MS

To study what metabolites are contributing in heterosis performance in F_1 hybrids, metabolite profiling was performed to investigate the expression/alterations of a variety of metabolites in F_1 hybrids with respective parents. We used Mass spectroscopy (MS) spectrum data to visualize various metabolites. Figure 5.7 A, B, C, D, E,and F indicates Heat Maps of the metabolites variation in the seedlings of the F_1 hybrids of *C. annuum* (Dudu) x *C. chinense* (Lota bhut), *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4), *C. chinense* (Lota bhut) x *C. chinense* (Chocolate bhut), *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4), *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) and *C. chinense* (Umorok) x *C. frutescens* (Acc. 4), respectively, with their respective parents . Our results showed that an increase in number

of metabolites which display higher concentration in F_1 compared with parents such as 64 metabolites in *C. annuum* (Dudu) x *C. chinense* (Lota bhut), 60 in *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4) 55 in *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia), 61 in *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4), 67 in *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) and 56 in *C. chinense* (Umorok) x *C. frutescens* (Acc. 4). The number and abundance of amino acids are found higher in F_1 plants than that of parents in almost all cross combinations (for example, alanine, threonine) (Table 1-6). Furthermore, sugar metabolites are observed in higher concentration in F_1's of all inter-specific hybrids (for example turanose, glucopyranoside, etc.) compared to parents, while overall reduction in the sugar metabolites level in intra-specific hybrid of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) is observed (Fig. 5.7). The analysis revealed that, all F_1 hybrids invariably showed considerably higher levels of fatty acid and sugar derivatives as compared to parents (Fig. 5.7). We observed some of the metabolites which are not detected in both of the parents but abundantly found in F_1 hybrids such as sugars (xylose, galactoside, melibiose, etc.), sugar derivatives (2, keto-glutaric acid, palatinose, glucuronic acid, etc.), fatty acids (octadecanoic acid/stearic acid, oleic acid, linolenic acid, propanoic acid, etc.), carboxylic acid (tetracoconoic acid, dioxoheptanoic acid, propenoic acid, etc.), and various other metabolites (Annexure 4-11) Overall, GC-MS profiling revealed considerable heterosis in F_1 hybrid over both parents at metabolite level.

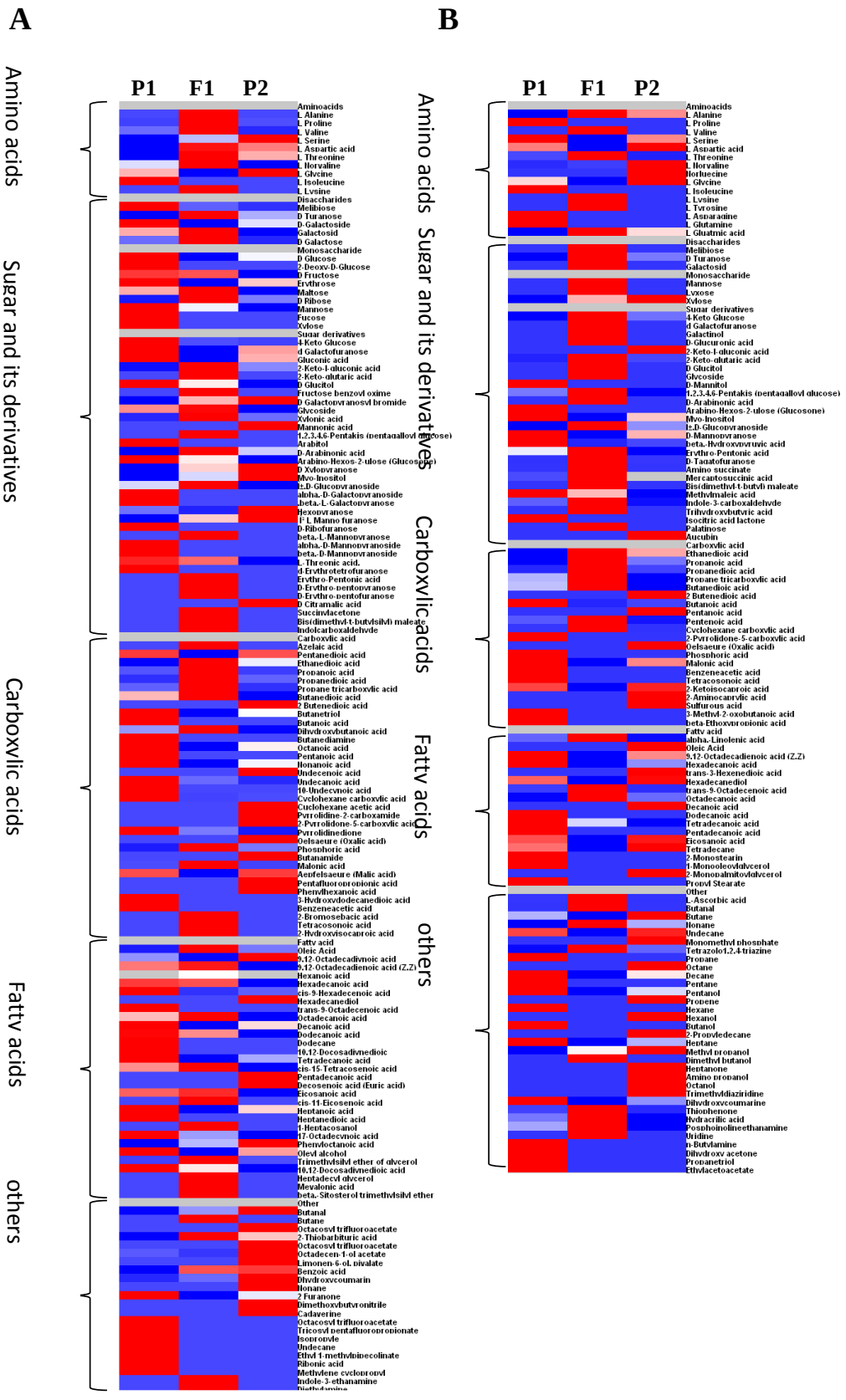


Figure 5.7 Heat map showing level of expression of various metabolites in F_1 hybrid and parental lines:- (A) *C. annuum* (Dudu) and *C. chinense* (Lota bhut), (B) *C. chinense* (Acc 7) and *C. frutescens* (Acc 4)

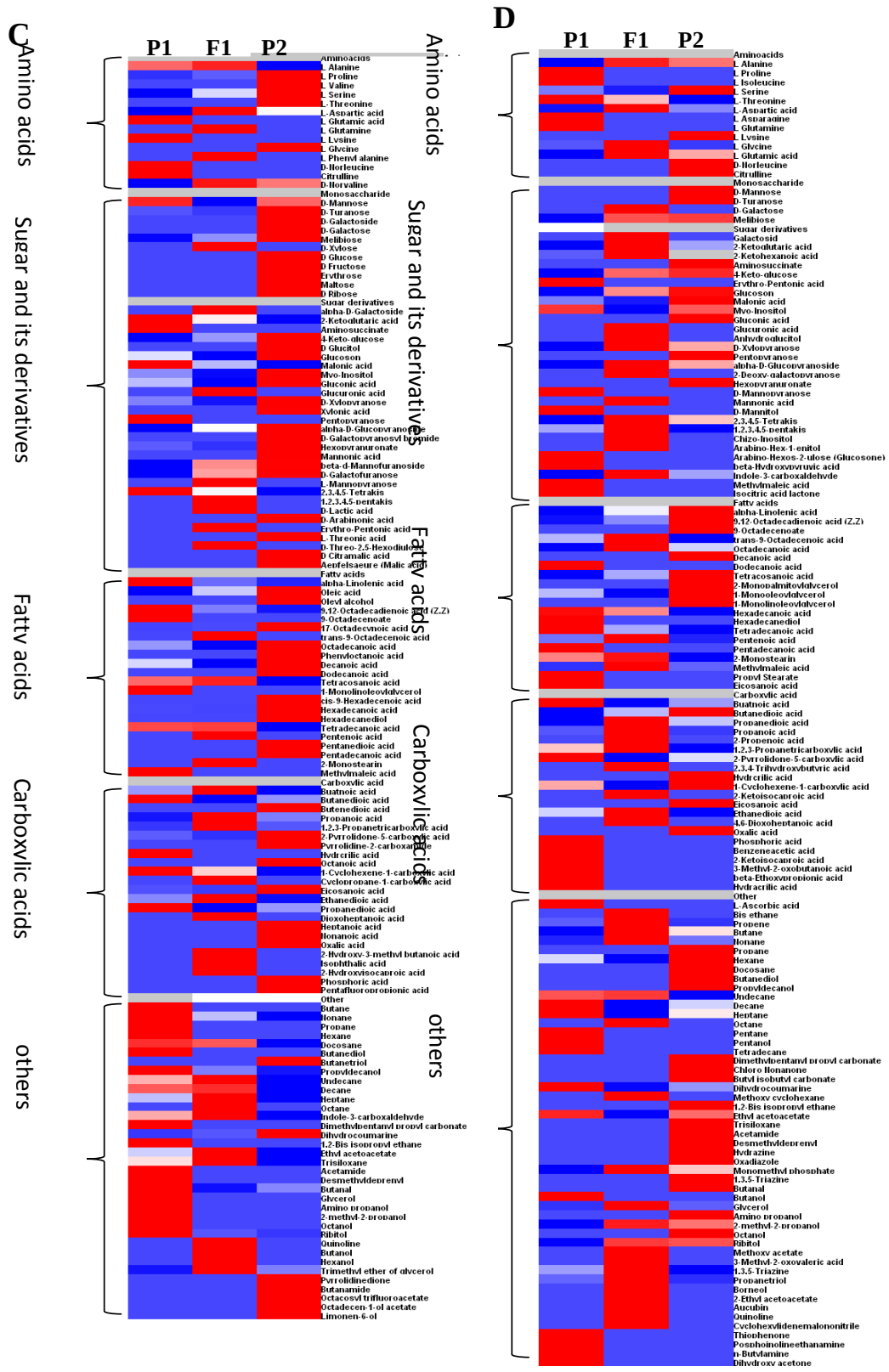


Figure 5.7 Heat map showing level of expression of various metabolites in F_1 hybrid and parental lines:- (C) *C. chinense* (Lota bhut) and *C. chinense* (Chocolate jolokia), (D) *C. chinense* (Chocolate jolokia) and *C. frutescens* (Acc 4).

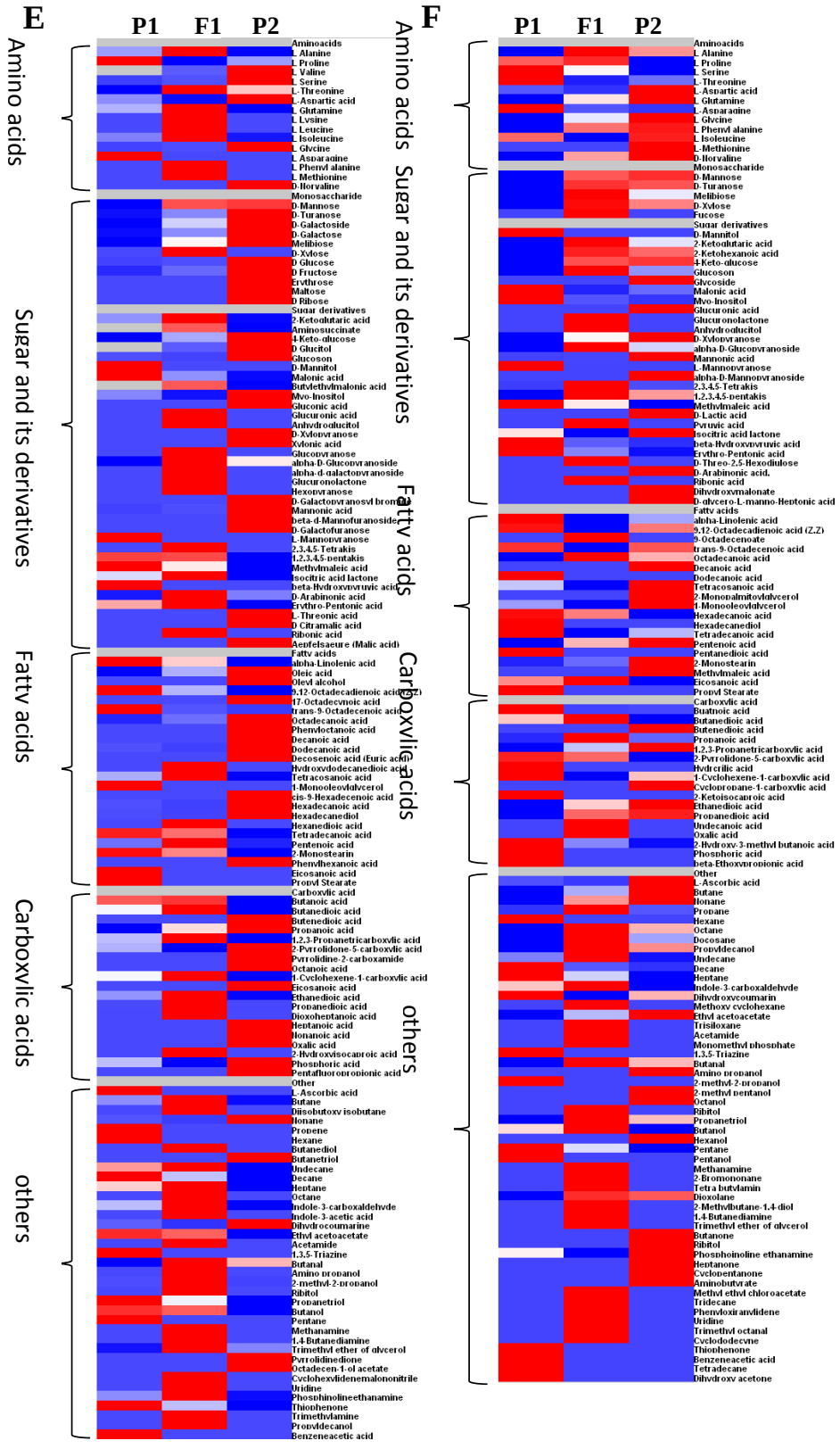


Figure 5.7 Heat map showing level of expression of various metabolites in F_1 hybrid and parental lines:- (E) *C. chinense* (Lota bhut) and, *C. frutescens* (Acc 4) (F) *C. chinense* (Umork) and *C. frutescens* (Acc 4).

3.5 Transcriptome sequencing and data processing

The paired end sequencing of cDNA libraries constructed from each of the F_1 hybrid and their respective parents was done with Illumina platform. The transcriptome sequencing analysis yielded a minimum of 22,817,552 and maximum of 26,137,550 raw reads from these twelve libraries (Table 5.3). Further, these raw reads were processed and filtered through various criteria such as removal of adaptor sequences, low-quality etc.to get high quality processed reads and mapped to *C. annuum* reference genome. Minimum of 72 % andmaximum of 86 % reads could be mapped in the reference genome (Table 5.3). The data analysis generated a minimum of 21,846 to 23,210 transcripts (Table 5.4). The maximum transcript length was found to be 13,179 bp while minimum length is of 300 bp. The average, median and total transcript length per tissue sample along with transcripts more than 500 bp, 1 kb and 10 kb has been mentioned in Table 5.4. The N50 (half the assembled bases found in contigs length) ranged from minimum 1,470 bases to maximum of 1,490 bases (Table 5.4).

3.6 Differential gene expression and expression level dominance analysis

The numbers of differentially expressed genes (DEG) in each genome of F_1 hybrid and their parents are slightly varied. A total of 24719, 24097, 23521, 24037, 23843, 23828 differentially expressed genes were obtained in F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota), *C. chinense* (Acc.7) X *C. frutescens* (Acc.4), *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia), *C. chinense* (Chocolate)X *C. frutescens* (Acc.4), *C. chinense* (Lota) X *C. frutescens* (Acc.4), *C. chinense* (Umorok) X *C. frutescens* (Acc.4) in comparison to respective parents. To obtain highly stringent significantly expressed genes we applied several criteria's such as genes with FDR <0.05, genes with FPKM>1, and log2 value >2. The following these criteria, DEGs were selected for further analysis (Table 5.5). To investigate expression level dominance, these DEGs were further classified into 12 possible differential expression categories (Fig. 5.8) as accordingly to the method described earlier (Rapp et al. 2009 and Li et al 2014). All these DEGs from each F_1 hybrids are listed in table 5.5.

Table 5.3 Statistics of sequence reads of *Capsicum* F_1 hybrids and their parents.

Sample name	Raw reads	High quality processed reads	High quality nucleotides (bp)	Mapped reads (%)
S1	23102776	22008674	6144702534	72.3
S2	26137550	24956127	6990359453	79.1
S3	23895852	23239500	6602393691	76
S4	24462351	23272320	6536067787	74.2
S5	23648377	22298386	6212543173	73.5
S6	25117123	24020188	6783400448	75.8
S7	22817552	21674325	6056359453	70.9
S8	23612156	22117334	6168560929	72.5
S9	23224962	21735292	6019602496	72
S10	23546220	22245105	6229111377	72.8
S11	24383279	23187505	6515013274	86.5
S12	25252656	23857206	6678136611	71.5

Abbrevation: S1- *C. chinense* (Acc. 7) X *C. frutescens* (Acc. 4), S2-*C. annuum* (Dudu) X *C. chinense* (Lota), S3-*C. chinense* (Umorok) X *C. frutescens* (Acc. 4), S4- *C. chinense* (Lota) X *C. frutescens* (Acc. 4), S5- *C.chinense* (Chocolate jolokia) X *C.frutescens* (Acc. 4), S6- *C. chinense* (Lota) X *C.chinense* (Chocolate jolokia), S7- *C.chinense* Chocolate jolokia, S8-*C. frutescens* (Acc. 4), S9- *C. chinense* (Umorock), S10- *C. chinense* (Lota Bhut), S11- *C. annuum* (Dudu), S12- *C. Chinense*(Acc. 7).

Table.5. 4 Statistics of reference based transcriptome assembly of *Capsicum* F_1 hybrids and their Parental lines

Sample Name	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Transcripts generated	22723	23074	22241	22410	22603	21848	22158	21960	22035	21846	23210	22256
Average transcripts length	1206.3	1194.8	1211.2	1207.1	1204.5	1217.7	1214.8	1217.1	1216.4	1218.5	1193.2	1211.4
Median transcripts length	636	709.5	324	1611	735	753	525	451.5	519	618	1899	640.5
Total transcripts length	27409713	27568848	26937600	27050748	27224376	26603505	26916915	26727714	26803590	26619324	27695175	26961708
Transcripts >=500 bp	18939	19125	18574	18696	18835	18310	18568	18408	18461	18336	19153	18607
Transcripts >=1 Kb	11293	11297	11132	11159	11217	11006	11149	11083	11087	11027	11343	11154
Transcripts >=10 Kb	7	7	7	7	7	7	7	7	7	7	7	7
N50 value	1482	1470	1485	1479	1476	1491	1485	1488	1488	1488	1473	1482

Abbrevation: S1- *C. chinense* (Acc. 7) X *C. frutescens* (Acc. 4), S2-*C. annuum* (Dudu) X *C. chinense* (Lota), S3-*C. chinense* (Umorok) X *C. frutescens* (Acc. 4), S4- *C. chinense* (Lota) X *C. frutescens* (Acc. 4), S5- *C.chinense* (Chocolate jolokia) X *C.frutescens* (Acc. 4), S6- *C. chinense* (Lota) X *C.chinense* (Chocolate jolokia), S7- *C.chinense* Chocolate jolokia, S8-*C. frutescens* (Acc. 4), S9- *C. chinense* (Umorock), S10- *C. chinense* (Lota Bhut), S11- *C. annuum* (Dudu), S12- *C. Chinense*(Acc. 7).

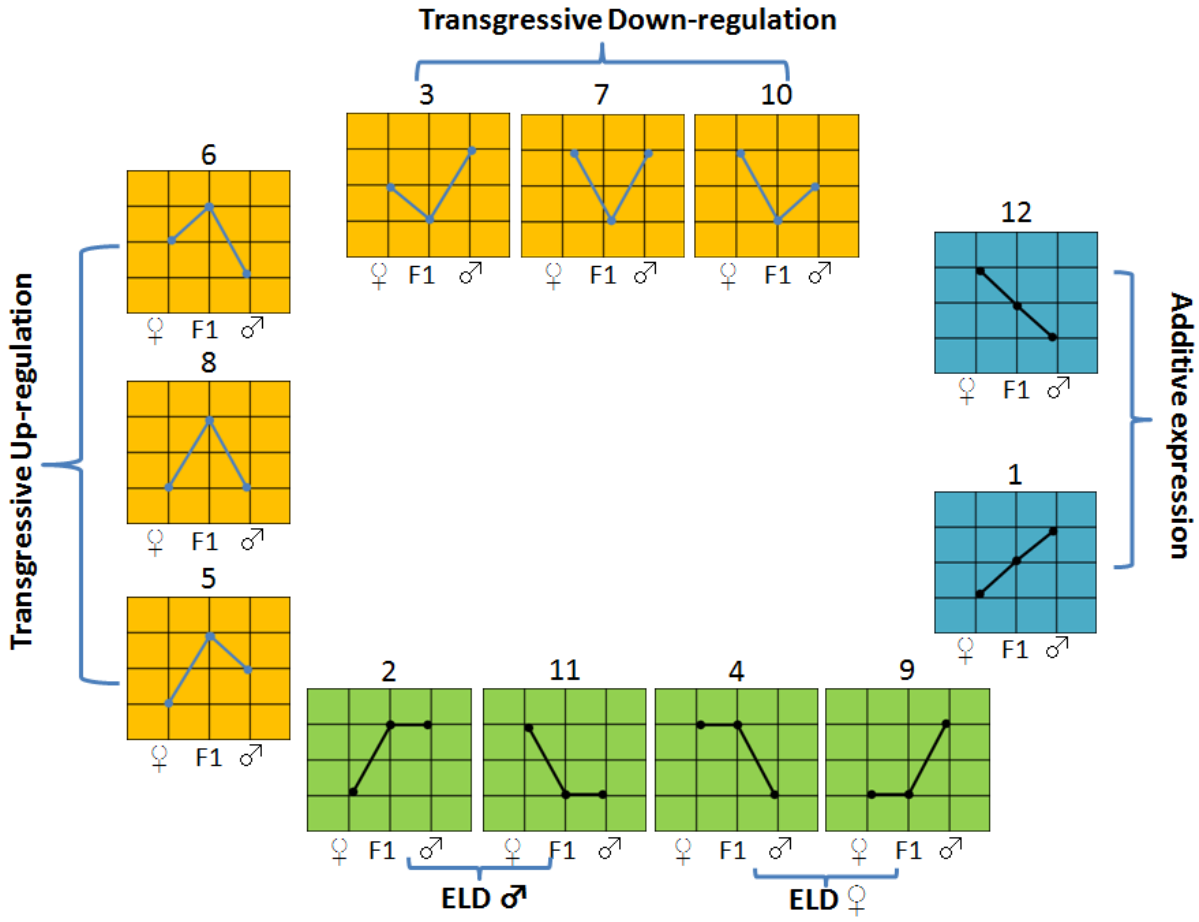


Figure 5.8 Twelve possible differentially expressed states between the F_1 hybrids relative to its parents. The categorization was done according to the method described by Rapp et al., (2009)

Furthermore, we have analysed the mode of gene action of these DEGs. About 60 % (7230 of 12284) of genes exhibited non-additive expression pattern while, 40 % (4964 of 12284) of genes resembles to additive gene expression in F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota) (Table 5.5). Similarly, another inter-specific F_1 hybrid of *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) showed 67.3 % (8832 of 13108) of genes with non-additive expression and 33.7 % (4276 of 13108) genes with additive expression. The intra-specific F_1 hybrid of *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia) showed 71.86 % (8397 of 11684) genes revealed non-additive gene expression and 28.1% (3287 of 11684) showed additive expression pattern. Similarly, 59.3%, 64%, 92.5% genes from F_1's of *C. chinense* (Chocolate) X *C. frutescens* (Acc.4), *C. chinense* (Lota) X*C. frutescens* (Acc.4), *C. chinense* (Umorok) X *C. frutescens* (Acc.4), respectively displayed non-additive gene expression (Table 5.5).These observations suggest that the non-additive gene expression is a major

contributor of gene action than additive expression. The non-additive expressed genes are further categorized into high parent dominance (HPD), low parent dominance (LPD), over dominance (ODO), under dominance (UDO). Out of 7320 non-additive genes, about 1336 HPD, 1151 LPD, 1389 ODO and 1576 UDO genes were obtained in F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota) (Table 5.5). The different categories of gene expression in all the F_1 hybrids are mentioned in table 5. The top twenty significant differentially expressed genes in F_1 hybrid and their parents are illustrated in figure 5.9.

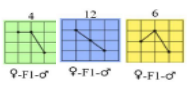
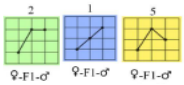
To understand the expression level dominance of either of the parents in F_1 hybrid, genome, differentially expressed genes were classified into 12 possible states (Table 5.5). We observed that, in the case of F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota), 6944 genes are significantly differentially expressed which are derived from *C. annuum* (Dudu) parent while, 7686 genes are significantly differentially expressed which are derived from another parent *C. chinense* (Lota) (Table 5.6). So, considerably larger numbers of genes from *C. chinense* (Lota) parent are present in F_1 genome. In another inter-specific cross of *C. chinense* (Acc.7) X *C. frutescens* (Acc.4), 4083 genes derived from *C. chinense* (Acc.7) parent are expressed and 9228 genes derived from other parent *C. frutescens* (Acc.4) are expressed (Table 5.6). Interestingly in all cross combination where, *C. frutescens* (Acc.4) was used as paternal parent, the significantly larger contributions of genes are observed as compared to that of maternal parent (Table 5.6). Overall the expression level dominance analyses from all the studied cross combination suggest that *C. frutescens* genotypes are dominant over *C. chinense* genotypes while; *C. chinense* genotypes are dominant over *C. annuum* genotypei.e. the dominant relationship of different *Capsicum* genomes are *C. frutescens* >*C. chinense* >*C. annuum*.

Table 5.5 Summary statistics of dominance patterns of differentially expressed genes.

Hybrid	Total genes	Total number of significantly DEG (FDR <0.05)	Number of genes with FPKM >1.0	Number of genes with Log2 >2.0	Number of genes used for analysis	Additive	Non-Ad ditive	HPD	LPD	ODO	UDO	Others
C. annuum **(Dudu) X** ***C. chinense*** **(Lota)**	24719	19207	14611	12284	12284	4964	7320	1336	1151	1389	1576	1868
C. chinense **(Acc.7) X** ***C. frutescens*** **(Acc.4)**	24097	15316	14472	13108	13108	4276	8832	1773	756	2402	1443	2458
C. chinense **(Lota) X** ***C. chinense*** **(Chocolate jolokia)**	23521	14998	13998	11684	11684	3287	8397	506	2720	527	4246	398
C. chinense **(Chocolate) X** ***C. frutescens*** **(Acc.4)**	24037	15110	14356	12962	12962	5273	7689	2104	987	1517	1960	1121
C. chinense **(Lota) X*C. frutescens*** **(Acc.4)**	23843	15162	14264	12719	12719	4573	8146	1649	774	2602	1479	1642
C. chinense **(Umorok) X** ***C. frutescens*** **(Acc.4)**	23828	12840	11816	11021	11021	825	10196	1948	1503	2145	2595	2005

Abbreviations: DEG- Differentially expressed genes, HPD- High parent dominance, LPD- Low parent dominance, ODO- Over dominance, UDO- Under dominance.

Table 5.6 Expression level dominance of genes from parents in F_1 hybrids

F1 Hybrids	 Parent 1	 Parent 2
C. annuum **(Dudu) X** ***C. chinense*** **(Lota)**	6944	7686
C. chinense **(Acc.7) X** ***C. frutescens*** **(Acc.4)**	4083	9228
C. chinense **(Lota) X** ***C. chinense*** **(Chocolate)**	3084	10021
C. chinense **(Chocolate) X** ***C. frutescens*** **(Acc.4)**	8124	8831
C. chinense **(Lota) X** ***C. frutescens*** **(Acc.4)**	6796	12332
C. chinense **(Umorok) X** ***C. frutescens*** **(Acc.4)**	3845	6571

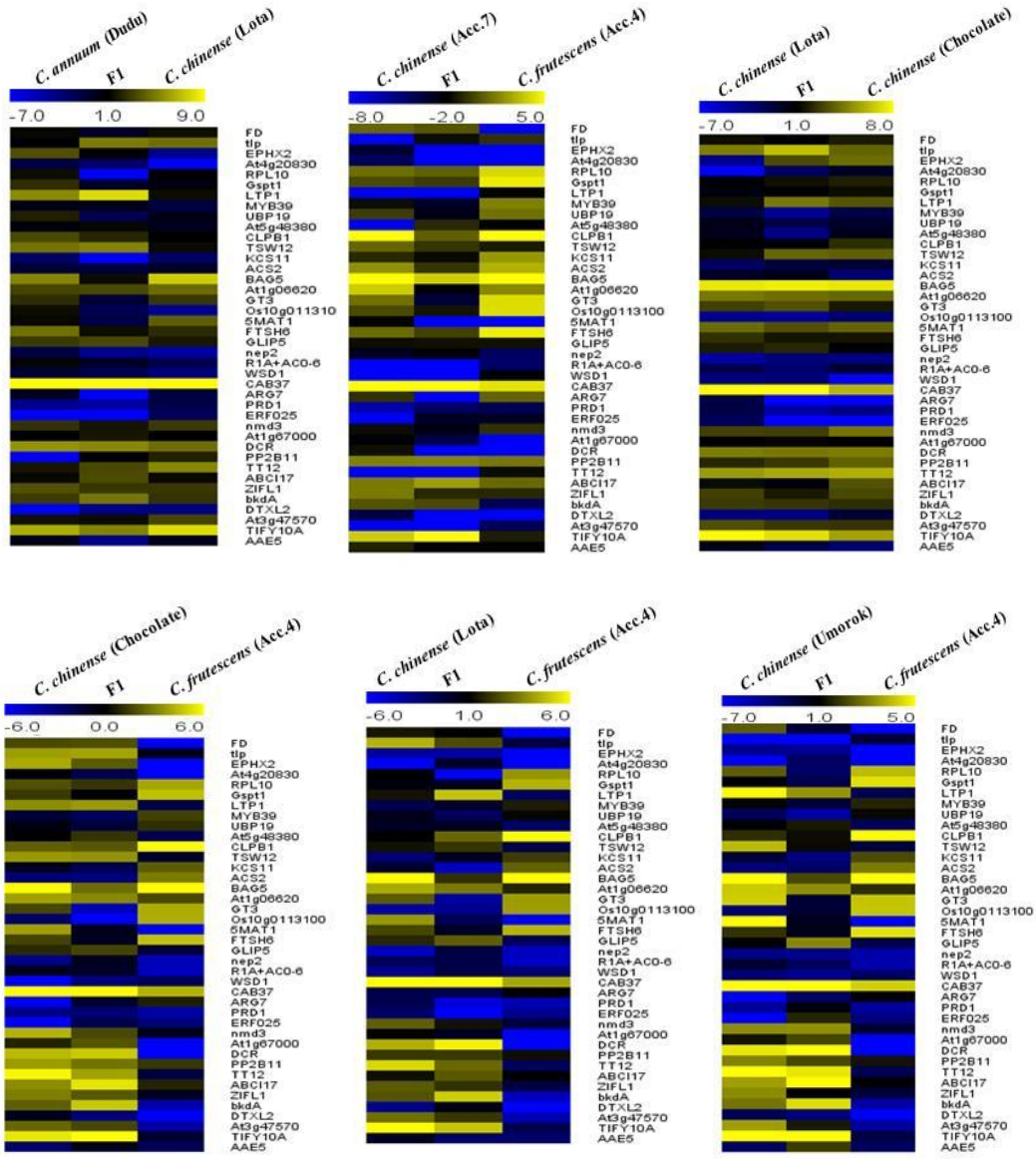


Figure 5.9 Heat maps showingtop twenty differentially expressed genes between the *Capsicum* F_1 hybrids and their respective parents.

3.7 Gene ontology enrichment analysis of differentially expressed genes

To investigate the functional role of differentially expressed genes observed between F_1 hybrids and their respective parents' gene ontology (GO) analysis was done. The GO enrichment analysis identifies the significantly over-represented class of gene. The analysis categorised all the genes into three broad classes such as genes involved in biological processes, cellular components and molecular functions. Among biological processes, transcription and its regulation was found to be enriched followed by protein ubiquitination and defense response (Fig. 5.10). In cellular component class, genes associated with nucleus and integral component of membranes were abundantly found followed by cytoplasm and plasma membrane. Among molecular function category, number of genes associated with ATP binding is copious followed by metal binding and DNA binding (Fig. 5.10).

The functional enrichment analysis was done using agriGO tool using Fisher's exact test with Benjamini-Yekutieli (FDR under dependency) multiple comparison corrections and FDR <0.05. In F_1 hybrid of *C. chinense* (Acc.7) X *C. frutescens* (Acc.4), the GO enrichment analysis revealed significantly higher abundance of expression of genes involved in biological processes such as cellular protein modification and metabolism, pollen-pistil interaction etc. (Fig. 5.11). In the cellular component class, plastid, thylakoid and plasma membrane organelles are significantly enriched (Fig. 5.11). Among the molecular function category, nucleotide binding, carbohydrate binding, protein kinase activity, protein binding and hydrolase activity were highly greatly represented. Similarly, in the case of inter-specific F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota), the GO enrichment analysis of biological process followed the same pattern except carbohydrate metabolism and response to biotic stimulus is highly enriched (Fig. 5.11). Also, similar representation from cellular components and molecular function category was observed in F_1 hybrid and no obvious difference could be found in the enrichment analysis of this F_1 hybrid with the earlier inter-specif cross of *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) (Fig. 5.12A). However, enrichment analysis of intra-specific F_1 hybrid of *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia), showed presence of some enriched process such as DNA metabolism in biological process class (Fig. 5.12B). In cellular processes in addition to thylakoid, plastid, and mitochondria associated genes, gollgi apparatus associated genes are also found to be enriched. No obvious difference has been noted in molecular function category of this F_1 hybrid with earlier inter-specific hybrids.

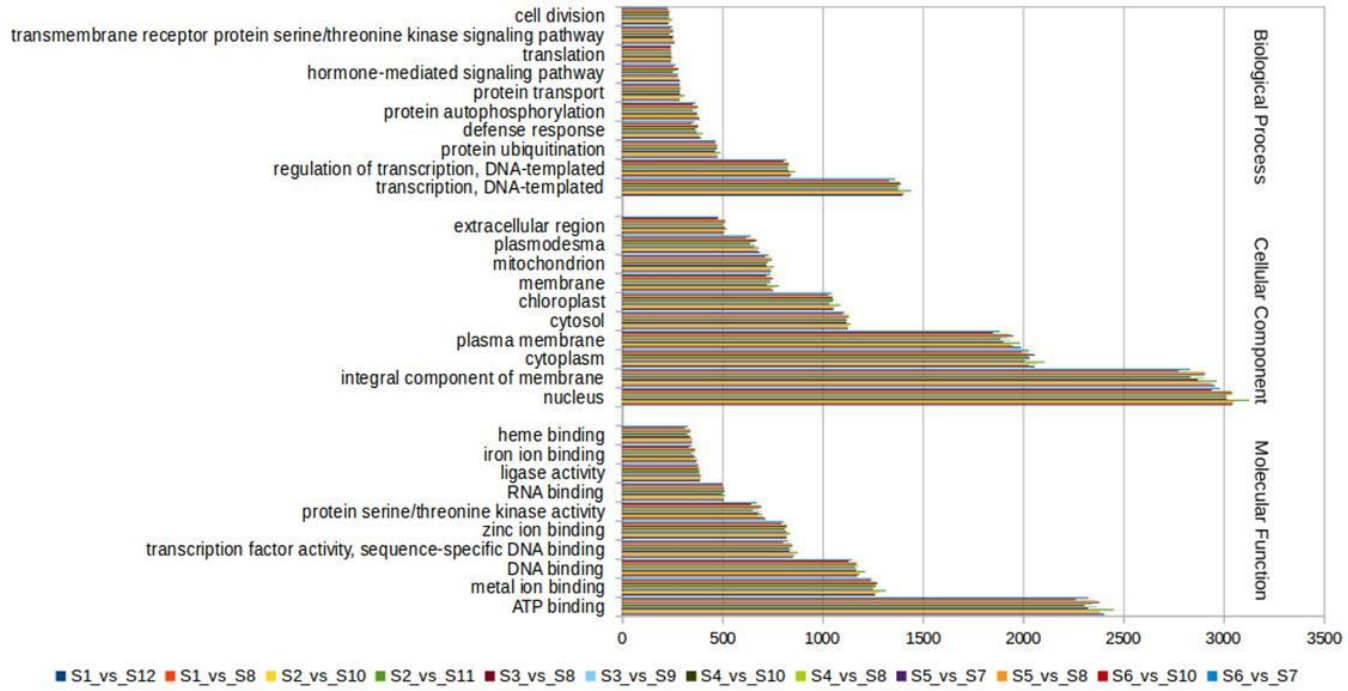


Figure. 5.10 Gene ontology analyses of F_1 hybrids with parent showing top ten GO terms of biological processes, cellular component and molecular function.

Abbrevation: S1- *C. chinense* (Acc. 7) X *C. frutescens* (Acc. 4), S2-*C. annuum* (Dudu) X *C. chinense* (Lota), S3-*C. chinense* (Umorok) X *C. frutescens* (Acc. 4), S4- *C. chinense* (Lota) X *C. frutescens* (Acc. 4), S5- *C.chinense* (Chocolate jolokia) X *C.frutescens* (Acc. 4), S6- *C. chinense* (Lota) X *C.chinense* (Chocolate jolokia), S7- *C.chinense* Chocolate jolokia, S8-*C. frutescens* (Acc. 4), S9- *C. chinense* (Umorok), S10- *C. chinense* (Lota Bhut), S11- *C. annuum*(Dudu), S12- *C. hinense*(Acc. 7).

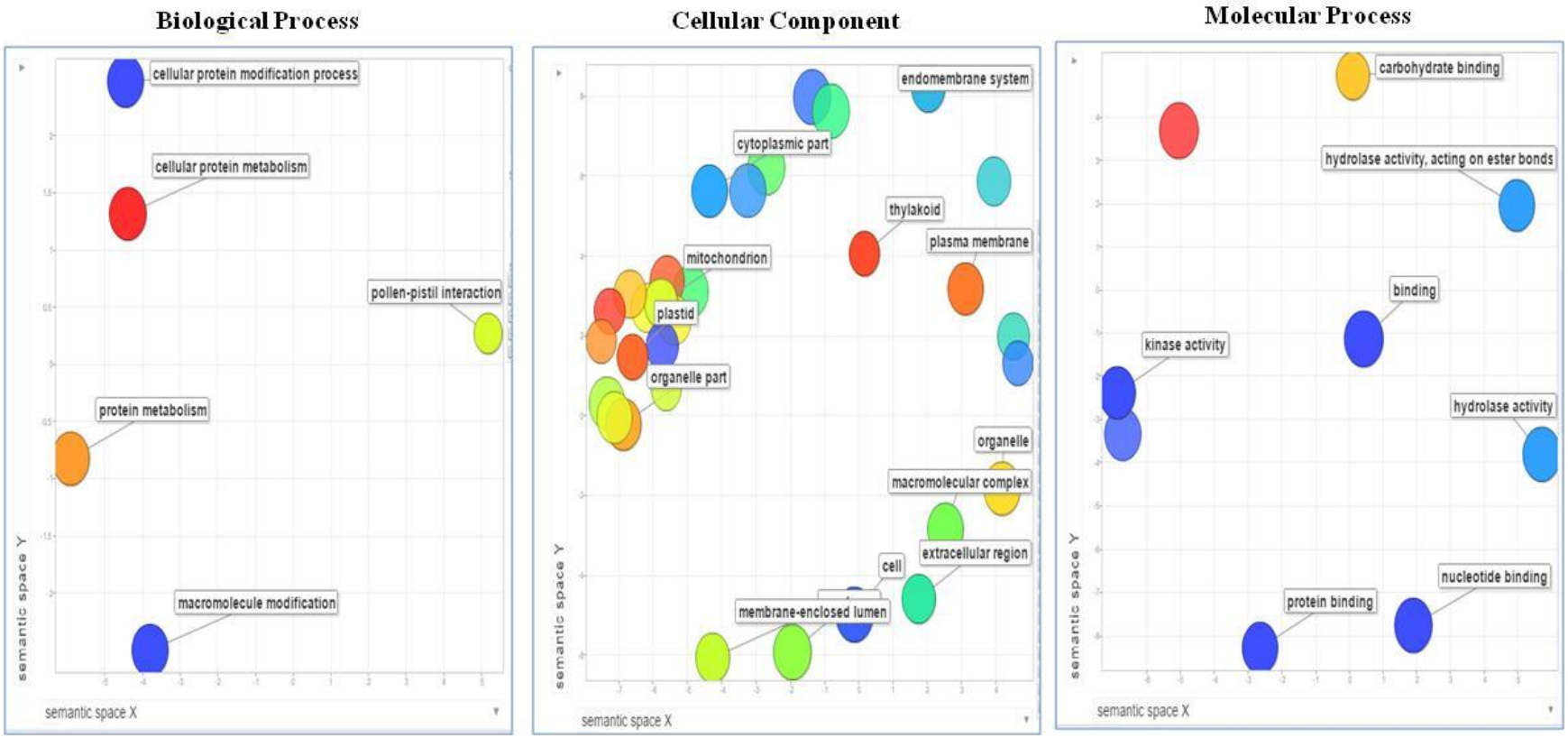


Figure 5.11 Gene ontology enrichment analysis of *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) F_1 hybrid with parent showing significantly enriched GO terms of biological processes, cellular component and molecular function.

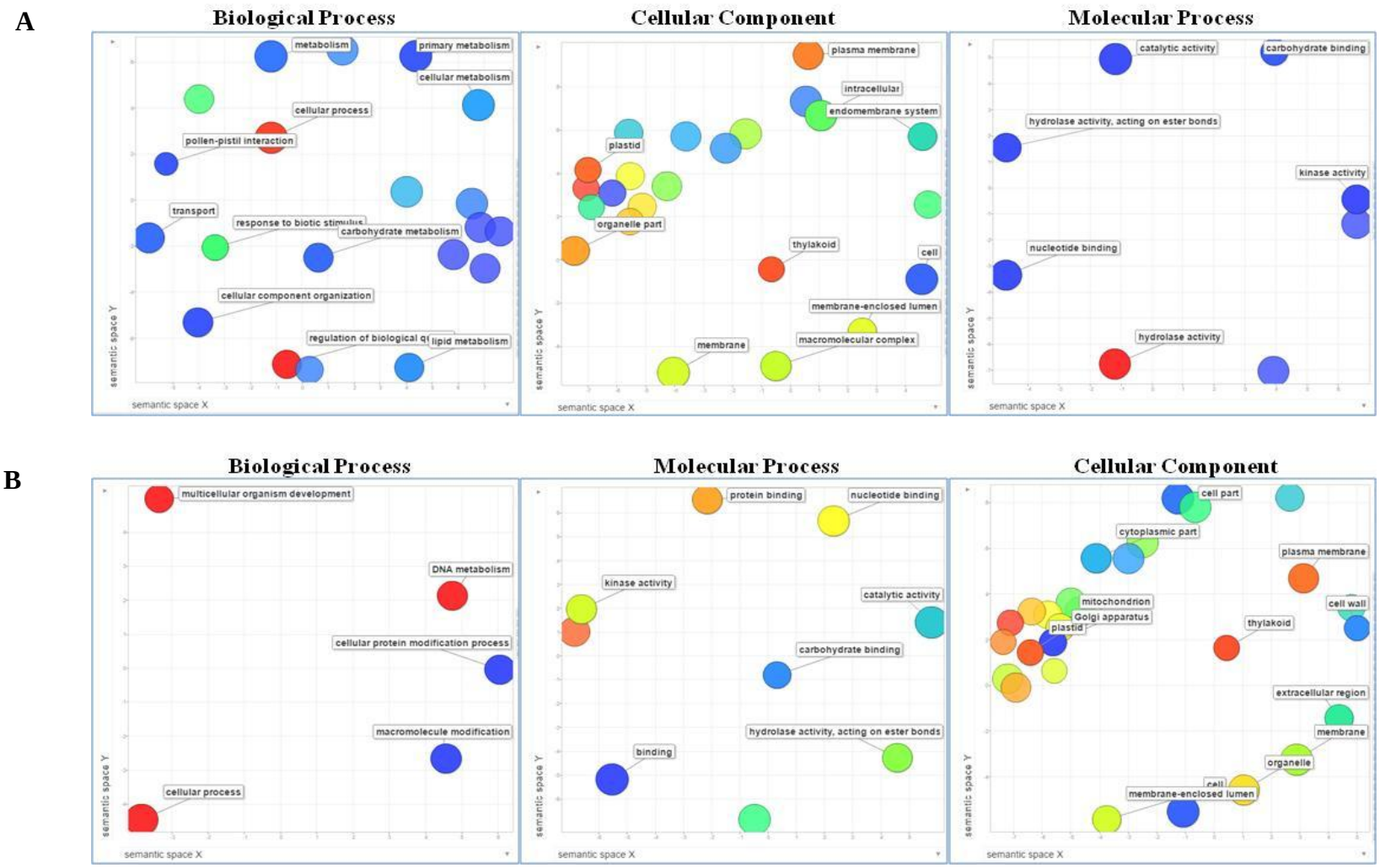


Figure 5.12 Gene ontology enrichment analyses of F_1 hybridsof A) *C. annuum* (Dudu) X *C. chinense* (Lota) andB) *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia) with parent showing significantly enriched GO terms of biological processes, cellular component and molecular function.

3.8 Analysis of differentially expressed carbohydrate metabolism genes in transcriptomes of *Capsicum* F_1 hybrids and their respective parents

The metabolite profilling revealed considerable increased levels of sugars and its deriavtives in F_1 hybrids as compared to parents. To confirm their dominance at genomic level, we have analysed the expression level of carbohydrate metabolism genes. The normalised FPKM values for each gene was transformed to Log2 value and indicated in the form of Heatmap (Fig.5.13). We observed that most of the carbohydrate metabolism genes are considerably present in higher level as compared to both the parents. The inter-specific F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota) and *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) showed higher expression of multiple genes for enzymes such as trehalose-phosphate synthase (TPS), Glucose-1-phosphate adenylyltransferase (AGPS), UDP-glucose 6-dehydrogenase (UGDH),UDP-glucose flavonoid 3-O-glucosyltransferase (GT6/GT7), GDP-mannose transporter (GONST3), Hexokinase (HXK2/1) than that of parents (Fig. 5.13). The remaining inter-specific hybrids such as *C. chinense* (Chocolate)X *C. frutescens* (Acc.4), *C. chinense* (Lota) X *C. frutescens* (Acc.4), *C. chinense* (Umorok) X *C. frutescens* (Acc.4) showed similar expression pattern in which majority of genes are highly expressed in F_1 as compared to both the parents (Fig. 5.13). Contrastingly, the intra-specific hybrid of *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia) revealed reduction in expression levels of these genes compared to parents.

3.9 Analysis of differentially expressed chlorophyll biosynthesis pathway genes in transcriptomes of *Capsicum* F_1 hybrids and their respective parent

We have observed high chlorophyll content and photosynthetic activity in F_1 hybrids compared to that of parents. To see the correlation of phenotypes with gene expression, the expression pattern of chlorophyll biosynthesis pathway genes in F_1 hybrids and parents was analysed. The results revealed that most important genes of the chlorophyll biosynthesis pathway such as *Aminoacyl tRNA synthase* (AIMP), *Glutamyl-tRNA reductase* (GluRS encoded by HEMA gene), *Magnesium-chelatase* (CHLI), *Protochlorophyllide reductase* (PORA) and *Chlorophyll synthase* (CHLG) were highly expressed in F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota) and *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) in contrast to parents (Fig. 5.14). The other inter-specific hybrids such as *C. chinense* (Chocolate)X *C. frutescens* (Acc.4), *C. chinense* (Lota) X *C. frutescens* (Acc.4), *C. chinense* (Umorok) X *C.*

frutescens (Acc.4) demonstrated identical expression pattern where majority of genes are highly expressing in F_1's compared to both parents (Fig. 5.14). Although, alone intra-specific hybrid of *C. chinense* (Lota) X *C. chinense*

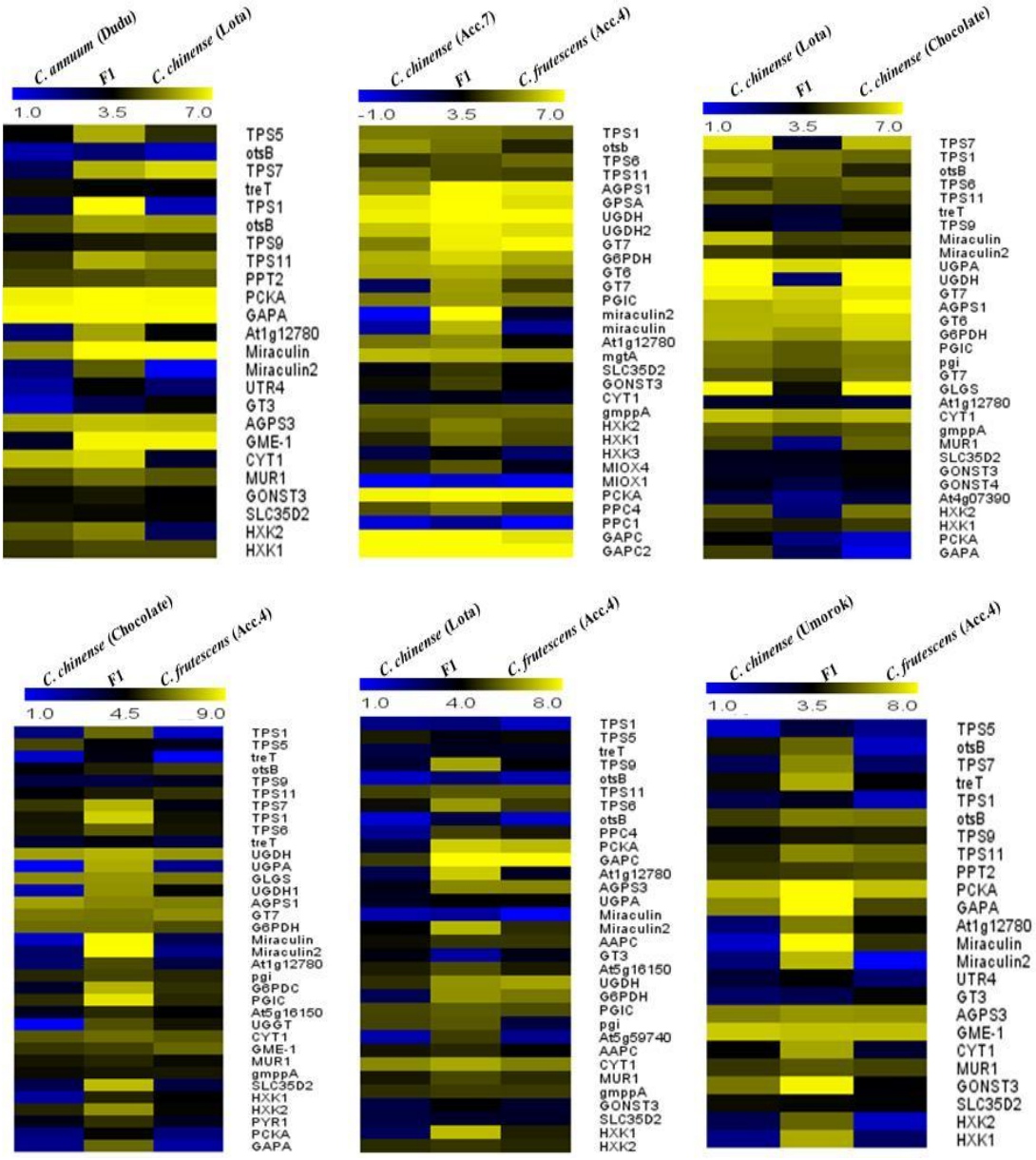


Fig. 5.13 Heat map showingexpression genes involved in carbohydrate metabolism in *Capsicum* F_1 hybrid and their respective parent

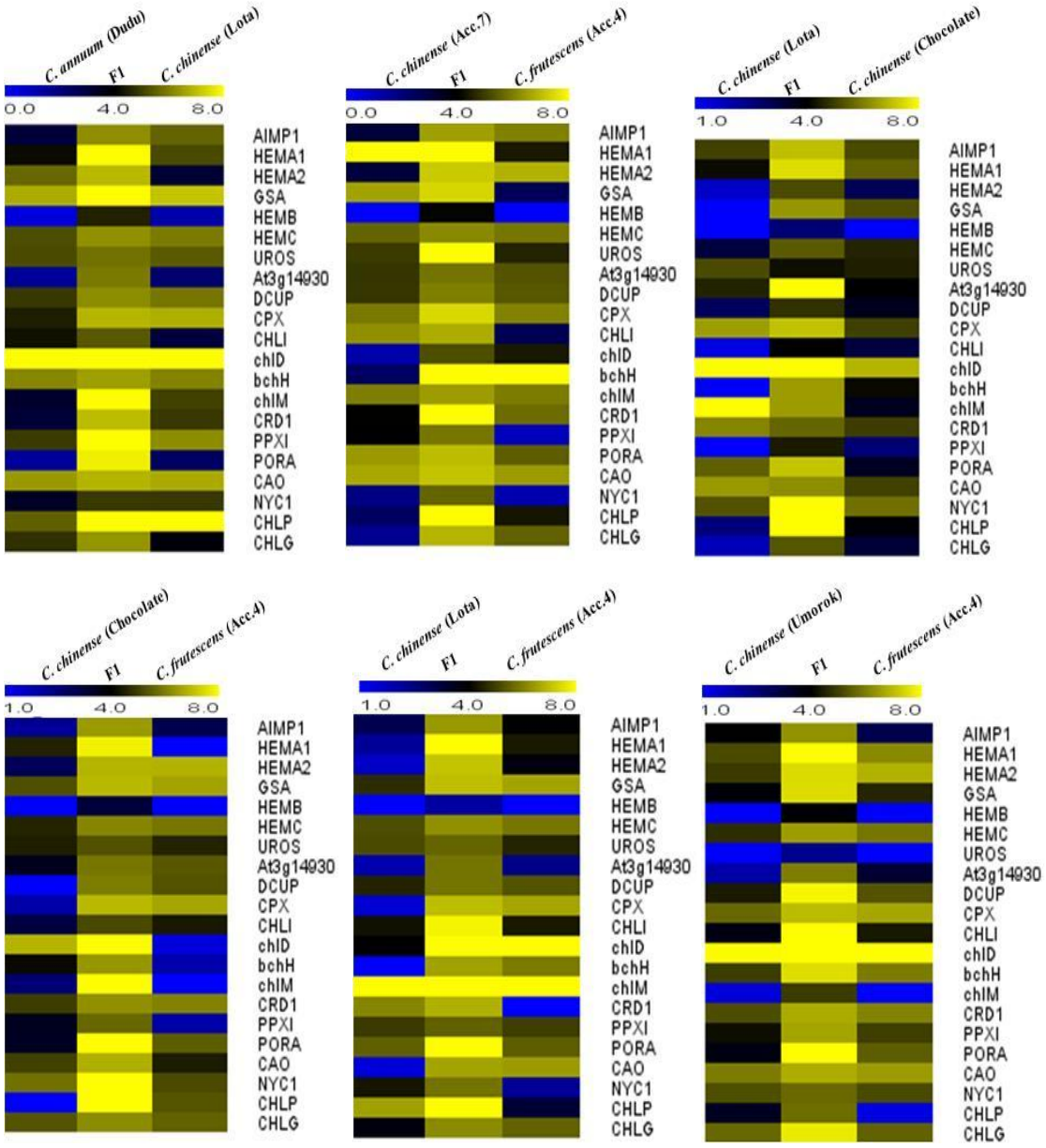


Figure 5.14 Heat map showingexpression genes involved in chlorophyll biosynthesis in *Capsicum* F_1 hybrid and their respective parent

(Chocolate jolokia) showed deviated expression profiling as contrast to others. This F_1 hybrid demonstrated reduction in gene expression level of important structural genes like *AIMP*, *HEMA*, and *CHLI* compared to parents (Fig. 5.14). However, *Geranylgeranyl diphosphate reductase* (CHLP) and PORA genes showed high expression in F_1 hybrid in contrast to parents.

3.10 Analysis of differentially expressed transcription factors in transcriptomes of *Capsicum*F_1 hybrids and their respective parent

The role of transcription factor in gene regulation is known. In the present study analysis of identify transcription factors (TFs) and their mode of action or role was done. In our F_1 dataset we identified a total of 46 different types of TFs which regulates the expression of several crucial pathways of plant development. Further, we determined their mode of gene expression and categorized into additive and non-additive pattern. Among these TFs, WRKY and bHLH gene family is abundantly found followed by ERF, NAC and MYB families (Fig.5. 15A). The WRKY and bHLH TF family majorly exhibited non-additive expression pattern relative to additive. Moreover, comparative analysis of all TFs reveals non-additive expression is major contributor of mode of gene action. Since WRKY gene family is majorly found compared to other TFs, we analyzed their expression in all the intra-specific and inter-specific F_1 hybrids and parents. The expression analysis revealed that majority of WRKY transcription family members are highly expressed in all the interspecifc hybrids compared to their respective parents (Fig.5.15B). However, in intra-specific F_1 hybrid of *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia), we could identify reduced number of WRKY TFs and the expression levels are also observed to be low in F_1 hybrid as compared to parents (Fig. 5.15B). Overall, the expression of WRKY gene family members are observed to be high in case of inter-specific hybrids than that of intra-specific.

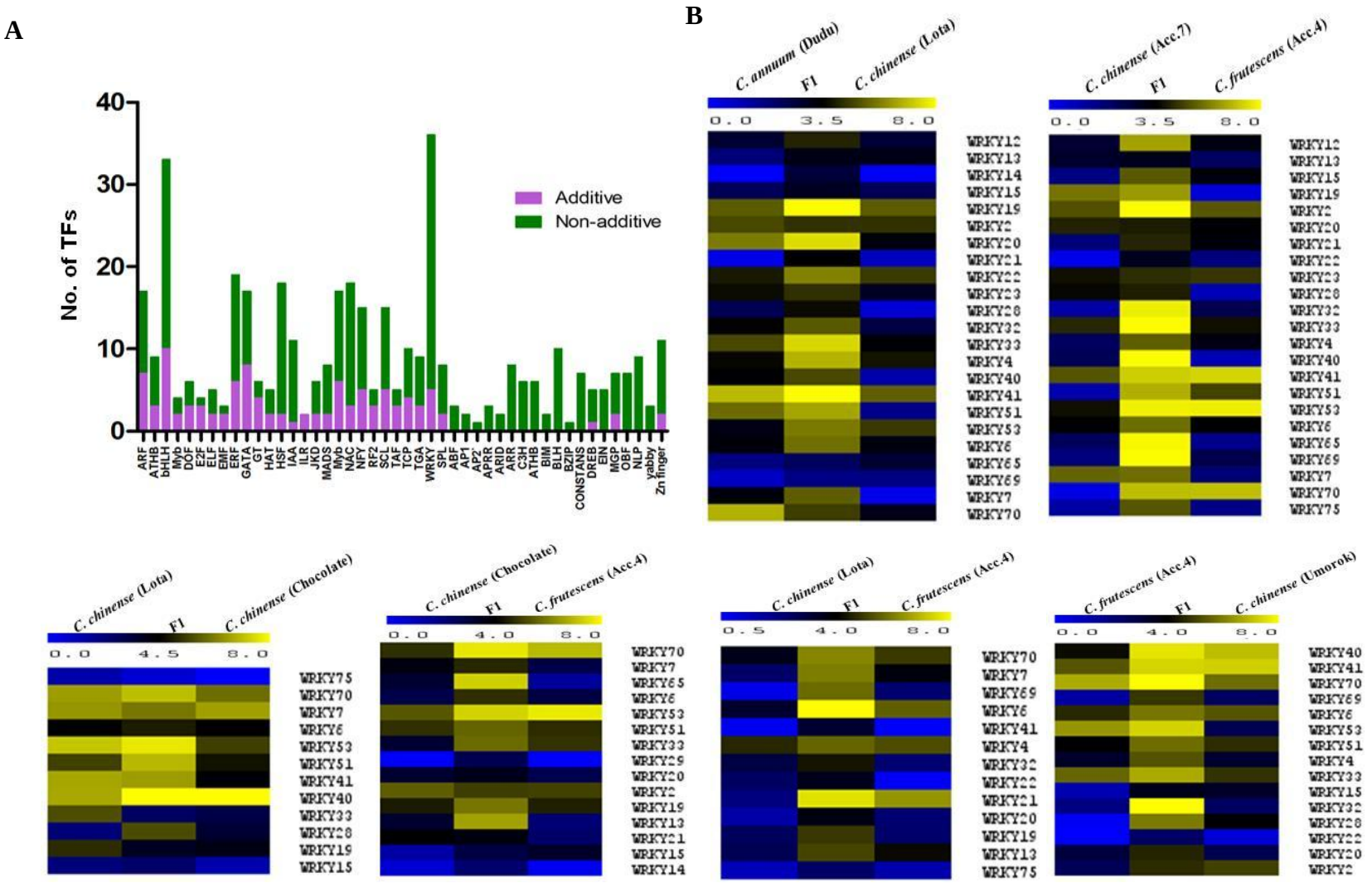


Figure 5.15 Expression analysis of TFs in hybrid. A) Estimation of total number of TFs in F_1 hybrid and their mode of gene action B) Heat map showing expression analysis of WRKY TFs *Capsicum* F_1 hybrids and their respective parents.

3.11 Experimental validation of gene expression observed in transcriptomes by Quantitative real time PCR (qRT-PCR)

To validate the expression profiling and mode of gene action of differentially expressed genes, we performed Quantitative real-time PCR. We randomly selected 12 DEGs from each cross of a unique parental combination such as *C. annuum* (Dudu) X *C. chinense* (Lota), *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) and *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia) and expression pattern was studied (Fig. 5.16A, B and C). These 12 genes are involved in variety of functions such as pathogenesis related protein, photosystem II protein, RuBiSCO small subunit, wound induced protease inhibitor, TFs (NAC and F-box), universal stress protein, etc. The list of these genes along with gene id and protein function for each set of hybrid has been mentioned in table 5.7. These genes belongs to overdominance, underdominance, and additive typesof gene action possibly imparting the hybrid vigour or heterosis in F_1 hybrids.The qRT-PCR data revealed that, for *C. annuum* (Dudu) X *C. chinense* (Lota) cross 10 of 12 genes, for *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) cross 9 of 12 genes and for *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia) cross 10 of 12 genes showed similar expression patterns observed in differential gene expression data from transcriptomes (Fig. 5.16).Overall, considerable resemblancein the expression patterns of genes derived through transcriptome sequencing and qRT-PCR was found indicating reliability of the differential gene expression analysis of transcriptomes data in F_1 hybrids and their respective parents.

Table 5.7 List of the selected genes from each F_1 hybrid used for the qRT-PCR experiment along with protein function(for primer details please refer to Annexure 3)

Hybrid	Gene Name	Protein	Hybrid	Gene Name	Protein
C. chinense x C. frutescens	Capana03g004339	Light-regulated protein	*C. annuum X C. chinense*	Capana08g002266	18.2 kDa class I heat shock protein
	Capana03g001469	Wound-induced proteinase inhibitor 2		Capana02g000066	Transketolase, chloroplastic (TK)
	Capana02g001912	Auxin-repressed 12.5 kDa protein		Capana02g000541	RuBisCO small subunit
	Capana02g002880	ADP-ribosylation factor GTPase-activating protein		Capana04g001287	Wound-induced proteinase inhibitor
	Capana03g002388	Protein DREB1C		Capana03g001487	21 kDa seed protein
	Capana03g000778	Chitin-binding lectin 1 (PL-I)		Capana02g002781	RuBisCO small subunit 2A
	Capana02g000764	Oxygen-evolving enhancer protein 1, chloroplastic	*C. Chinense X C. chinense*	Capana07g000107	Flower-specific defensin (NaD1)
	Capana08g002538	Glycine-rich RNA-binding protein		Capana09g001847	PR protein
	Capana03g004202	Extensin-3 (AtExt3)		Capana09g001520	Chlorophyll a-b binding protein
	Capana03g004449	Pathogenesis-related protein STH-2		Capana06g001739	NAC TFs
	Capana08g001896	Ferredoxin, chloroplastic (PFLP)		Capana03g001467	Wound-induced proteinase inhibitor 2
	Capana02g003498	Carbonic anhydrase, chloroplastic		Capana02g003359	Calmodulin-like protein 45
C. annuum X C. chinense	Capana08g002306	endochitinase		Capana05g002288	F-box
	Capana03g000778	GDSL esterase/lipase		Capana03g003392	Cytochrome P450 82C4
	Capana03g004339	Light-regulated protein		Capana08g000734	Photosystem I P700 chlorophyll a
	Capana12g000132	Photosystem II 5 kDa protein		Capana02g002192	Peroxidase 42
	Capana09g001750	Universal stress protein A-like protein		Capana08g002351	CCoAMT-6
	Capana01g000893	Phosphoribulokinase, chloroplastic		Capana07g000730	Sucrose synthase

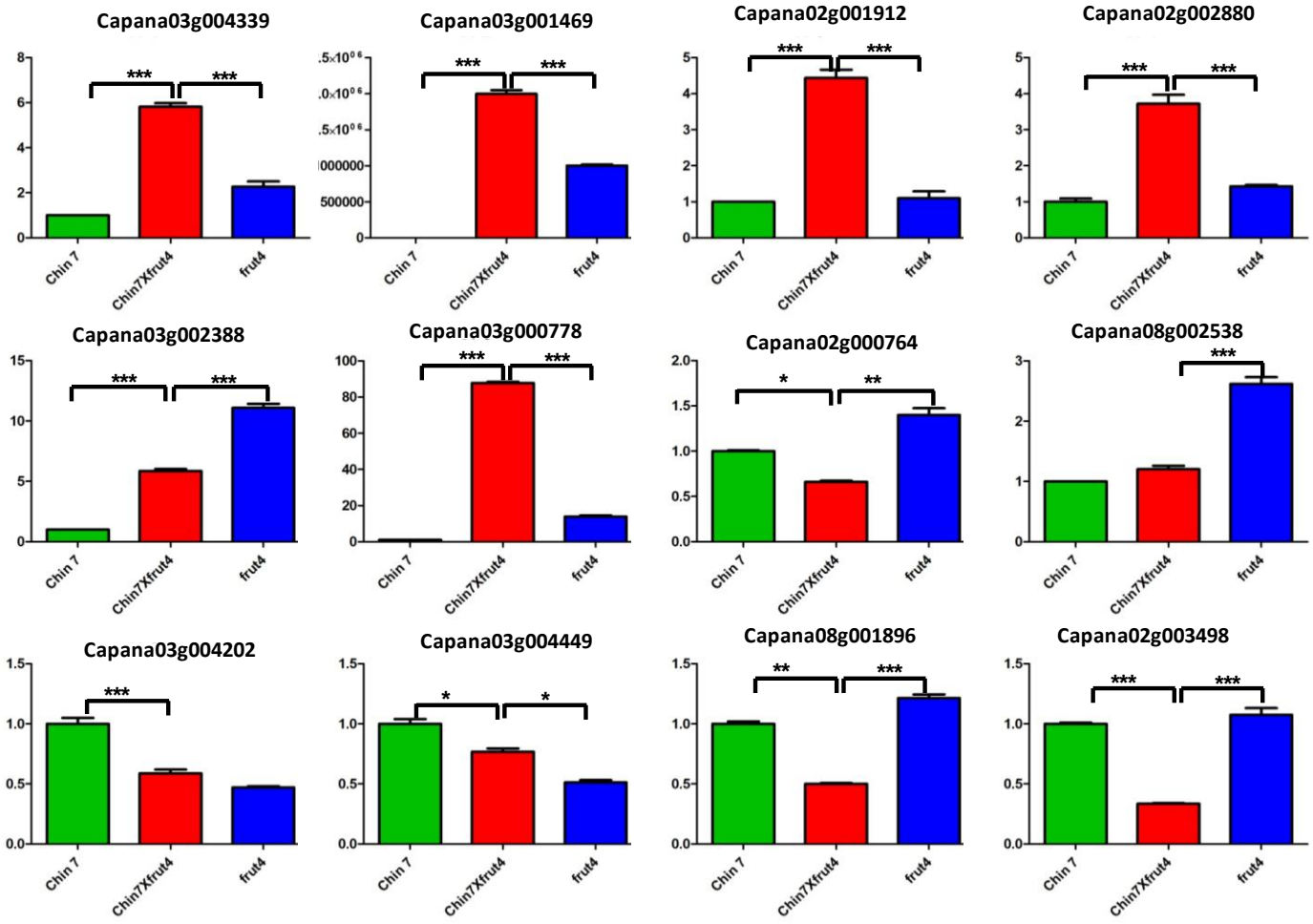


Figure 5.16A. Expression analysis of the twelve genes in F_1 hybrid of *C. chinense* (Acc.7) X *C. frutescens* (Acc.4) by qRT-PCR. The error bars signifies standard deviation among biological replicates of tissue samples. **P*<0.05, ***P*<0.01, ****P*<0.001.

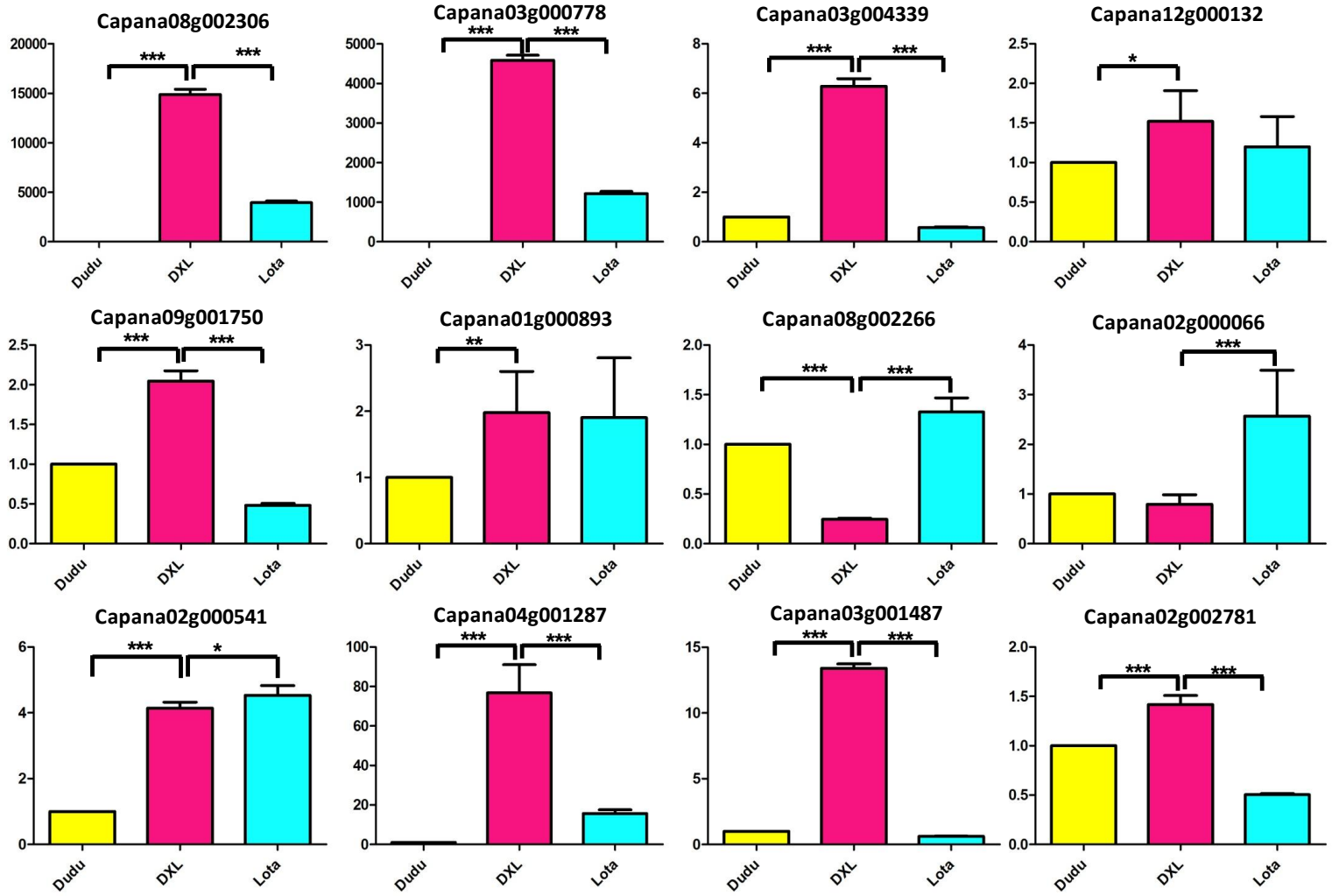


Figure 5.16B Expression analysis of the twelve genes in F_1 hybrid of *C. annuum* (Dudu) X *C. chinense* (Lota) by qRT-PCR. The error bars signifies standard deviation among biological replicates of tissue samples. **P*<0.05, ***P*<0.01, ****P*<0.001.

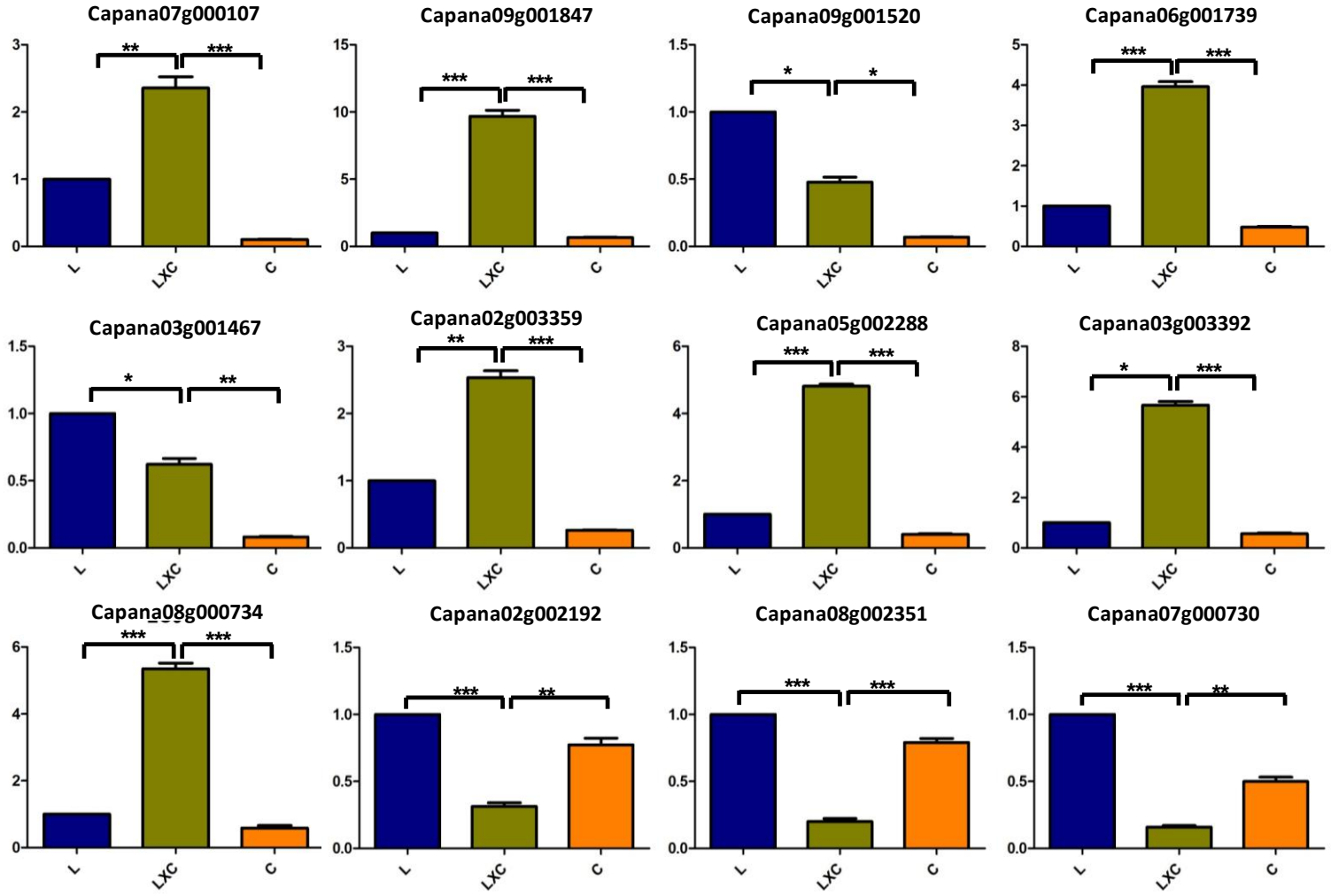


Figure 5.16C Expression analysis of the twelve genes in F_1 hybrid of *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia) by qRT-PCR. The error bars signifies standard deviation among biological replicates of tissue samples. **P*<0.05, ***P*<0.01, ****P*<0.001.

4. Discussion

Investigation of heterotic behaviour of F_1 hybrids in *Capsicum*

The presence of wide genetic variation with respect to pungency, fruit morphological and other agronomically important traits was observed in Bhut jolokia (*C chinense*) and *C. frutescens* from North East India was observed in our study (Sarpras et al., 2016). Bhut jolokia (*C.chinense*) being the naturally occurring highest pungency containing *Capsicum* species are currently in high demand around the world as the extreme high pungent property in the fruit which is due to the presence of capsaicnoids are used as spices in the form of dry powder orcommercial chilli sauce, for application in medicinal purposes, and for other commercial purposes. Apart from that, Defense Research and Development Organisation, Govt. Of India has successfully been using extreme fiery hot property (capsaicnoids) of Bhut jolokia to produce chilli grenades and sprays, and teargas carnisters since 2009, and successful used against terrorist and communal riots. Therefore, the market price of this chilli species is 4-5 times more than the normal chilli varieties (*C. annuum*). In addition to that, *C. frutescens* with medium pungency content evolved in North East India are also being used widely. Although very high in demand commercially, the yield potential are quite low and *Capsicum* genetic resources of these two species are unexploited in systematic breeding programme to produce high yielding varieties with desired traits. Therefore, the present study was undertaken to see if heterotic F_1 hybrids could be produced by crossing those genetically diverse germplasm, as development of F_1 hyrbid is quick compared to pedigree and other breeding methods.

Our study in five interpspecific i.e. *C. annuum* (Dudu) x *C. chinense* (Lota bhut), *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4), *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4),*C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4) and *C. chinense* (Umorok) x *C. frutescens* (Acc. 4) and one intra-specific hybrid of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia) showed heterosis for several traits. Significant heterosis was observed for plant height, days to 50% germination, fruit number per plant, root length, leaf area, photosynthetic rate, transpiration rate and chlorophyll content suggesting combinations of genes from genetically diverse parents are importantfor getting heterotic F_1s for all these traits. Seed germination event may have larger influence on determining early establishment of seedlings leading to hybrid vigour in plants and we observed that, the days to 50%

germination was considerably less in almost all the F_1 hybrids resulting in a significant negative heterosis over mid parent (-41 to -75%) and better parent (-44 to -78%).It is well known that, germination is a critical period in the plant life cycle which is sternly regulated by endogenous and environmental signals (Rajjou et al., 2012). The T_{50} trait has not been explored vigorously in plants although some reports are available in which, the positive heterosis was detected in early maize seedling development (Fu et al., 2011; Ding et al., 2012). Furthermore, significantly positive heterosis over mid parent (25 to 75%) and better parent (15.1 to 40%) for plant height. The maximum MPH and BPH was detected in inter-specific cross of *C. chinenese* Acc. 7 X *C. frutescens* Acc. 4. The positive heterosis for plant height was reported in several chilli hybrids but all these F_1 hybrids were developed in *C. annuum* (Gelata and Labuschagne, 2004; Shrestha et al., 2011; Kumar et al., 2014; Rao et al., 2017). While, we obtained a significant positive heterosis in almost all the parental combinations representing three diverse *Capsicum* species like *C. chinense*, *C. annuum* and *C. frutescens*.

The root length heterosis is very crucial for establishment of plants and determining the hybrid vigour in F_1 plants. Five of the six F_1 hybrids showed positive mid parent heterosis (6 to 80%), among all, a significant better parent heterosis was detected in F_1 cross of *C. chinense* (Acc. 7) X *C. frutescens* (Acc. 4). The healthier root organization in F_1 hybrids could potentially lead to enhanced nutritional uptake to sustain high growth rates in plants, ultimately leading to heterosis in hybrids. The heterosis for root length has not been reported in *Capsicum* species. The positive heterosis for root length in F_1 hybrids was observed during early seedling development in rice (Sasmal and Banerjee, 1986), maize (Hoecker et al., 2006; Paschold et al., 2010) and wheat (Tahira et al., 2011). Nellathambi and Kumari (2003) obtained positive heterosis for root length in most of the maize hybrids which varied from -2.89 to 30.62 over mid parent. A greater leaf area during early seedling development allows F_1 hybrids to absorb more light relative to their parents, potentially leads to enhanced photosynthetic efficiency of the plant. Our study observed significant positive heterosis (0.6 to 62%) for leaf area in most of intra- and inter-specific hybrids. The similar results have been reported in maize (Pavlikov and Rood, 1987), cotton (Wells et al., 1988) and tomato (Rao et al., 1992).

The number of fruits per plant is one of the most important traits which is directly related to the increase crop yield. In the present study, significantpositive heterosis over mid

parent (44 to 200%) and better parent (28 to 107%) was observed which is similar to the findings reported earlier in *Capsicum* species(Gelata and Labuschagne, 2004; Shrestha et al., 2011; Kumar et al., 2014). For fruit length a positive mid parent heterosis was detected in four of the five inter-specific hybrids (0.6 to 4.3) with maximum of 4.3% in F_1hybrid of *C. annuum* (Acc. Dudu) X *C. chinense* (Acc. Lota). The similar results were observed in other studies of *Capsicum* (Sood and Kumar, 2011; Shrestha et al., 2011; Kumar et al., 2014). In the case of fruit weight, only two inter-specific F_1 hybrids exhibited positive mid parent heterosis (0.6 to 0.8%) while significant negative heterosis was obtained in most of the F_1 hybrids over mid parent (-15 to -72%) and better parent (-3.2 to -86.1). This might be because the inter-specific hybrids were developed using *C. frutescens* (Acc. 4) accession with small fruit weight. During our study it was observed that in all the F_1s, the *C. frutescens* traits were mostly dominant when it was used as one of the parents including fruit weight. However, in several reports positive heterosis for fruit weight in *Capsicum* was observed (Gelata and Labuschagne, 2004; Kumar et al., 2014; Rao et al., 2017), the possible reason is most of these studies are performed with large fruit weighted bell pepper and other chilli pepper cultivars belonging to *C. annuum*. For seed number per fruit, negative heterosis was detected in almost all of the intra- and inter-specific hybrids which in contrast with other study (Kumar et al., 2014).

The present study considered all the traits from seed germination to fruiting stage to understand at what stage(s) and traits manifestation of heterosis/hybrid vigour takes place and influence the overall performance of the F_1 hybrids towards heterosis. Our study observed that heterotic behaviour at all stages and several traits collectively influence the ultimate performance and yield of F_1 hybrids. Although previous studies observed considerable positive heterosis for yield associated traits such as number of fruits and fruit length including plant height Sweet pepper hybrids (*C. annuum*) (Gelata and Labuschagne, 2004; Sood and Kaul, 2006; Shrestha et al., 2011; Rao et al., 2016) and Bangladeshi chilli (*C. annuum*) (Hasanuzzaman et al., 2013), they did not study early traits of plant development such as days to 50% germination and root length. However, we found these early developmental traits of plants are associated with chilli pepper hybrid vigour. Moreover, majority of the studies are focused on *C. annuum* and other important *Capsicum* species such as Bhut jolokia and *C. frutescens* remained unexplored. Furthermore, our study is the first to develop and study the

hybrid vigour in several F_1 hybrids derived from crossing between the fiery hot Bhut jolokia(*C. chinense*) other *Capsicum* species (*C. annuum* and *C. frutescens*).

Photosynthesis and carbohydrate metabolism influences hybrid vigour

Recently, along with the study at phenotypic and molecular level (gene expression), several studies have been reported in which metabolites of parents and F_1s were estimated to understand the contribution of metabolites towards heterotic performance of F_1 hybrids (Schauer et al., 2008; Song et al., 2010; Zhao et al., 2015). In the current study, our analysis indicated in the six hybrids and their parents indicated that several metabolites showed positive mid parent and better parent heterotic expression. The most important metabolites, sugars and its derivatives showed significant positive mid parent and better parent heterosis in F_1 seedlings of the inter-specific hybrids. Sugars or carbohydrates are crucial for fundamental cellular and developmental processes of plants (Evland and Jackson 2011). Sugars act as the primary substrate for energy source and are one of the basic structural components for defense mechanism in plants, where they interact with various hormonal signalling networks controlling the plant disease resistance (Smeekens et al., 2010). Meyer et al., (2012) revealed that higher levels of sugars and fatty acid components are associated with early seedling heterosis in *Arabidopsis*. The disaccharide sugars and its derivatives are abundantly present in F_1 seedling of *Arabidopsis* than parents which is similar to our observation in the current study. Korn et al., (2010) found significant levels of glucose, raffinose, galactose, succinic acid, ketoglutaric acid in hybrid and were positively associated with heterosis with freezing tolerance in *Arabidopsis*. Similarly, in our investigation, we detected significantly higher concentration of these metabolites in inter-specific hybrids. In the intra-specific cross of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia), sugar and fatty acids concentration is reduced as compared to parents resulting in negative mid parent and better parent heterosis which might be potential reason behind low hybrid vigour in this hybrid. We observed considerable increase in fatty acids such as hexadecanoic acid, octadecanoic acid, linolenic acid in F_1 hybrids of inter-specific cross. This observation is similar to previously reported result in a study of heterosis in *Arabidopsis* (Meyer et al., 2012), and wheat (Zhao et al., 2015). In wheat, amino acids and sugars were found abundantly in F_1 hybrids compared to parents (Zhao et al., 2015) Similarly, in our current study, amino acids such as alanine, valine, threonine, and glutamic acid were found in significantly higher concentration in F_1 hyrbids compared to the parents.

Furthermore, to see whether increase in metabolites is the result of corresponding increase of gene expressions, we analyzed the expression of genes involved in carbohydrate metabolism and found that majority of the genes showed significantly higher amount of expression in F_1 hybrids compared to parents. The GO enrichment analysis revealed several genes associated with primary metabolism, such as carbohydrate, amino acids as well as energy metabolisms are significantly enriched.Lisec et al. (2008) suggested that primary metabolism was more tightly associated with plant growth and development. Our analysis revealed increase in amino acid content in F_1 hybrids showed prominence in F_1 hybrids and hence heterosis than that of other metabolic pathways (Ti fu et al., 2011). Moreover, we found several important metabolites that are strictly present only in F_1 hybrids with higher concentration than that of both parents. These metabolites are sugars, amino acids, fatty acids, carboxylic acids which indicate their vital involvement in determining hybrid vigour in F_1 plants. The GO enrichment analysis describes that the terms associated with energy and carbohydrate metabolism was significantly enriched in F_1 hybrids of intra- and inter-specific cross. The transcriptome analysis in rice hybrid demonstrates enrichment of energy metabolism genes in differentially expressed genes between super-hybrid and their parents (Wei et al., 2009). The pathways associated with energy metabolism are majorly photosynthesis and CO_2 fixation pathway.It is well known facts that increase in grain/fruit yield necessitates enhancement of photosynthetic efficiency (Horton 2000). We observed significantly high chlorophyll content, transpiration rate and photosynthetic rate in F_1 hybrids compared to their respective parents. These traits possibly might increase the photosynthetic yield which further directly confer hybrid vigour and enhanced crop yield. Further, the differentially expressed genes engaged in the enriched photosynthetic pathway may work together for hybrid vigour and yield enhancement. Moreover, to confirm this, we have analyzed the expression of photosynthetic pathway genes. We observed considerable increase in the expression level of major structural genes of this pathway in the F_1 hybrids of both intra- and inter-specific cross compared to their parents. Similar results were reported earlier on rice (Wei et al. 2009) and maize (Ti fu et al., 2011) in which not all metabolic pathways, but only specific enriched pathways such as carbohydrate and photosynthetic pathways, were significantly associated with hybrid vigour and yield. From the observation in our study, it is sensible to infer that the differentially expressed genes (DEGs) of carbohydrate and energy metabolism along with high expression of the corresponding metabolites in F_1s compared to

parental lines positively correlates with heterosis. Furthermore, these identified important genes could possibly use in future breeding programme to develop *Capsicum* varieties with increase carbohydrate and photosynthetic ability.

Non-additive gene expression is largely contributing to molecular basis of heterosis

Although exploitation of heterosis by conventional breeding led to significantly increase of production in many agriculturally important crop plants, no consensus mechanism has been established to identify the molecular mechanism of heterosis. The recent progress in genomics, especially ability to sequence global transcriptomes by using next generation sequencing technology (NGS) in a short span of time and cost effective manner, are being used at a large scale to understand the functions of genes governing developmental and economically important traits in many of the crop plants. As a result, study at molecular level to understand the basis of heterosis at transcriptomes level have been reported in heterotic F_1 hybrids and parental lines. Several hypothesis suggests that heterosis could be originated due to differential gene expression between F_1 hybrids and its parents (Song and Messing 2003; Hubner et al. 2005) and different genes expression models are constructed to decode heterosis. The inter-specific hybrids of rice (Ge et al., 2008; Wei et al., 2009; Song et al., 2010), *Arabidopsis thaliana* (Wang et al., 2006; Fujimoto et al., 2012) and maize (Hoecker et al., 2006; Thiemann et al., 2010) demonstrates that additive and non-additive pattern of gene expression are responsible for hybrid vigour in hybrids, and heterosis is principally regulated by genetic distinction between two parents. The transcriptome sequencing has been utilized to examine themolecular basis of heterosis in several intra- and inter-specific hybrids of rice, *A. thaliana*, Brassica and maize (Hoecker et al., 2006; Wang et al., 2006; Ge et al., 2008; Wei et al., 2009; Song et al., 2010; Thiemann et al., 2010; Fujimoto et al., 2012; Zhang et al., 2015).The possible role of additive gene regulation in maize heterosis was observed due to their significant enhancement in grain yield QTL (Thiemann et al., 2014). In maize hybrid, transcriptome analysis reveals concurrence of multiple gene action and transposable element coupled gene regulation as the basis of heterosis (Zhang et al., 2015). Our transcriptomes study to understand gene expression in F_1 hybrids and their respective parents from *C. chinense*, *C. frutescens* and *C. annuum* observed both additive and non-additive gene expression contributed to intra- and inter-specific heterosis in F_1 hybrids of different combinations. The differential gene expression analysis between different F_1 hybrids and

their respective parental genotypes revealed that the non-additive gene action largely contributed to heterosis compared to the additive gene action. We observed about minimum of 60 to maximum of 90% gene action in different inter-specific hybrids are controlled by non-additive gene expression, similarly the intra-specific hybrid of *C. chinense* (Lota) X *C. chinense* (Chocolate jolokia) revealed 72% non-additive gene expression. Previous studies also reported similar kind of observationheterotic rice (Swanson-Wagner et al., 2006), maize (Zhang et al., 2008; Ding et al., 2014), and wheat (Li et al., 2014) suggesting that the contribution of multiple types of non-additive gene expression such as dominance and over-dominance might be majorly influencing towards giving heterotic/ hybrid vigour phenotype in F_1 hybrids.

CHAPTER - VI

MicroRNA identification, target prediction and expression analysis

1. Introduction

Increasing evidence of small RNAs regulating the expressions of genes governing diverse developmental and biological processes are being reported in plants (Rubio-Somoza et al., 2011; Djami-Tchatchou et al., 2017). Therefore, the knowledge of entire repertoire of small RNAs (microRNA) is becoming essential to understand the convoluted gene regulatory pathways in plants. Solanaceous plants are the third most agro-economically valuable crops after Poaceae and Fabaceae. Among the Solanaceae plants, identification and characterization of several small RNAs/microRNAs have been reported in *Solanum lycopersicum*, *S. tuberosum* and *S. melongena*. (Zhang et al., 2008; Moxon et al., 2008; Zuo et al., 2011; Kim et al., 2011; Xie et al., 2011; Yang et al., 2013; Hwang et al., 2013; Lakhotia et al., 2014). Regulation of target genes governing economically important traits by few of those identified miRNAs have also been validated experimentally which is an indispensable footstep towards manipulation in crop improvement programme. In rice, over-expression ofOsmiR393 resulted more tillers and early flowering (Jiao et al., 2010), while over-expression of osamiR7695 leads to resistance against blast disease pathogen (Campo et al., 2013). Transgenic tomato with miR319 shows larger leaflets and promote leaf margins growth (Ori et al., 2007).

The recently completed whole genome sequencing and miRNA profiling in *C. annuum* also reported many miRNAs in hot pepper (Hwang et al., 2013; Kim et al., 2014; Qin et al.,2014). In comparison to other plants including tomato and potato from the same family, the identification of miRNAs and their functional characterization in *Capsicum* is very limited. In *C. annuum*, Hwang et al., (2013) reported identification of 29 conserved and 35 novel miRNA families from ten different tissue libraries, followed byidentification of 59 known and 310 novel miRNAs by Liu et al., 2017. However, the presence of more miRNAs is expected in *Capsicum* since the reported numbers of miRNAs are very small as compared to those reported in tomato, *Arabidopsis*, rice and other crops. Therefore, comprehensive profiling in different tissues and environmental conditions are required to identify more number of miRNAs to understandtheir regulatory roles in governing economically important traits. Furthermore, no study of miRNAs identification has been reported in another two most important species *C. chinense* and *C. frutescens*.

In this study, we performed deep sequencing ofsmall RNAs from four tissues/organs (leaf, flower, fruit and stem) in Bhut jolokia (*C. chinense*) and Acc 4 (*C. frutescens*). These two species is being cultivated in larger parts of the world after *C. annuum*. The *C. chinense* and *C. frutescens* were selected for deep sequencing because of its contrasting fruit

characteristics such as pungency, fruit size, and shape, color, metabolite content, antioxidant properties, flowering time, fruiting habit and yield. We report the identification and characterization of high-confidence miRNAs and predicted their potential targets that appertain to distinct biological and cellular processes. Our differential expression analysis showed spatio-temporal variation in miRNA expression, implying their multiferous roles in *Capsicum* development. Furthermore, one miRNAs was identified which possibly regulates *Yabby* gene controlling fruit size and validated in contrasting *Capsicum* germplasm differing in fruit size. The results will not only contribute to understand the miRNA-mediated developmental regulation in *C. chinense* and *C. frutescens* but also in other Solanaceous plants.

2. Materials and methods

2.1 Plant material and growth conditions

Two contrasting genotypes, one each from *C. chinense* (Lota bhut) and *C. frutescens* (Acc 4) were selected for the study. The leaf, flower, stem and fruit tissues were collected for miRNA sequencing. The fresh fully expanded leaf along with flower and stem were collected from healthy mature glass house grown plant. For fruit tissue, samples were collected at various fruit developmental stages such early, breaker and mature. All the fruit stages were pooled and considered as a fruit tissue sample. For growing the plants and growth condition refer to chapter 3 section 2.1.

2.2 RNA Extraction and small RNA sequencing

The method of total RNA extraction has been mentioned in Chapter 1 section 2.5b. The quality and quantity estimation was described in section 2.6 and 2.7. The small RNA libraries were prepared by using TruSeq Small RNA sample preparation kit (Illumina, US). Briefly, the 3' and 5' adaptors were ligated to total RNA followed by reverse transcription and amplification of the ligated product. The cDNA library was then purified and quality was checked by using gel electrophoresis (Refer chapter 1 section 2.6) and bioanalyzer. After confirming the quantity and quality, the cDNA libraries were used for deep sequencing. Each sRNA library was sequenced using Illumina Genome Analyser II and preliminary analysis was performed at Genotypic Technologies Pvt. Ltd, Bangalore, India.

2.3 Data pre-processing

After the sequencing of sRNAs, the quality of FASTQ fi les was assessed by SeqQCv2.2 software (http://genotypic.co.in/SeqQC.html). In this procedure, the low-quality sequence reads with <30 Phred scores and sequences shorter than 18 nucleotide long were eliminated. The remaining high-quality reads were trimmed for adapter and polyA tail containing sequences using UEA srna-Workbench programme (Stocks et al., 2012).

2.4 Identification of conserved and novel miRNAs

The analysis to identify potentialmiRNAs was performed following the method described by Hwang et al., (2013) with little modifications. To analyze the miRNAs expression and genomic distribution, the high-quality reads were mapped to the *C. annuum*

Zunla-1 genome (Qin et al., 2014) using Bowtie program allowing maximum of two base pair mismatches. Aligned reads were extracted and checked for the presence of ncRNAs such as tRNA, rRNA, snRNA and snoRNA contamination. The unaligned reads to above ncRNAs were used to predict conserved and novel miRNAs. Further, unique reads were retained and read count profile was generated. For the identification of conserved miRNAs from *Capsicum* sRNA libraries, the filtered reads populations from individual tissues were aligned against all reported plant miRNAs (including from *C. annuum*) sequences from miRBase 21.0 database (Griffiths-Jones et al., 2008) using ncbi-blast-2.2.30.

To identify novel miRNAs, the remaining reads were aligned against the *C. annuum* genome sequence and putative precursor sequences were extracted (Yang et al., 2011). These potential miRNA precursor sequences were used for the identification of novel miRNA using MIREAP 0.2. (http://sourceforge.net/projects/mireap) programme. Further, secondary structure, minimum free energy (MFE) and dicer cleavage site were predicted with the MIREAP 0.2 programme. MIREAP combines miRNA biogenesis, sequencing coverage and structural features for the identification of genuine miRNAs and their expression level from deep sequenced small RNA libraries. The MIREAP 0.2 programme was used with following parameters:- 1) eighteen and thirty-six nucleotide were the minimal and maximal length of miRNA sequence, 2) twenty and twenty-four were the minimum and maximum length of reference miRNA sequence, 3) twenty was the maximum miRNA copy number on the reference, 4) -18 kcal/mol were the maximum allowed free energy for a miRNA precursor, 5) minimum length of mature base pairs between miRNA and miRNA* were fourteen, 6) maximum bulge of miRNA and miRNA* were four, and 7) maximum asymmetry between miRNA/miRNA* duplex were five. The predicted small RNA sequences were recognized as potential candidate miRNA genes once they satisfied all the following standards:- 1) The miRNA and miRNA* are originated from distinct arms of the stem-loop structure of unique duplex with two nucleotides at 3' end, 2) The predicted secondary structures of hairpin should be stable and free energy of hybridization lower than − 18 kcal/mol., 3) The hairpin should be located in intra-genic regions or introns, 4) The mismatches of miRNAs and miRNA* sequence should be <2 with presence of asymmetric bulges restricted to one or none (Mayers et al., 2008). The RNAfold programme of Vienna RNA software package was used to predict the hairpin secondary structures (Hofacker, 2003).

2.5 Differential expression analysis of miRNA

To investigate the amount of expression of individual miRNAs the DESeq programme was used to normalize the expression in every tissue (Anders and Huber, 2010). Additionally, normalization factor was used to normalize read count of each miRNA in all the plant tissues. The identification of tissue-specific expression of miRNAs was done following the method described by Breakfield et al., (2012). The expression patterns of miRNAs shown in the form of heatmaps were produced using Multiple Experiment Viewer tool (v4.9) (Howe et al., 2010).

2.6 miRNA target prediction and functional annotation

The prediction of potential targets of the conserved and novel miRNAs was performed with psRNATarget tool (Dai and Zhao, 2011). The psRNATarget algorithm predicts the miRNA targets based on sequence complementarity scores and target site accessibility by estimating the thermodynamic stability of RNA duplex structure. The miRNA sequences were used as an input with *C. annuum* mRNA sequences using strict option [need strict alignment in the seed region (offset positions 2 to 8)] in the programme. This stringent option precludes the detection of target sites that lacks perfect complementarity in the seed region such as gaps or non-canonical base pairing. The miRNA hits having a minimum free energy <= -25 are assumed to be targets for reported miRNA.The *C. annum* genome annotation (Qin et al., 2014 and Kim et al., 2014) was used for the identification of putative functions of the predicted targets.Subsequently, thegene ontology (GO) terms were allotted to the target genes using Blast2GO. The significantly enriched GO terms were identified using agriGO tool (v2.0) with P-value ≤ 0.05.

2.7 Distribution miRNAs and comparative /synteny mapping of *Capsicum* miRNA genes with tomato and potato

The whole genome sequences of *Capsicum* (accession GCF_000710875.1), (Qin et al., 2014), tomato (accession GCF_000188115.3), (TGC, 2012) and potato (PGSC DM assembly version 3) (PGSC, 2011) were retrieved from respective resource databases. The ncbi-blast-2.5.0 was used to map the 22 high confidence *Capsicum* miRNA families in tomato and potato genome. The circos plot was used to depict the synteny relationships of *C. chinense* and *C. frutescens* miRNAs with that of tomato and potato genomes (Krzywinski et al., 2009).

2.8 Expression analysis by Quantitative Real-time PCR (qRT-PCR)

To evaluate and confirm the expression profiling of miRNAs, we performed the qRT-PCR analysis. Here, the aliquote of same RNA samples i.e. from leaf, flower and three developmental stages of the fruit (20DPA, 40 DPA, and 60DPA) of *C. chinense* and *C. frutescens* used for miRNA sequencing was used for expression analysis. The forward and reverse stem-loop primers for each miRNAs were designed and were synthesized from Sigma (Sigma-Aldrich, USA, Annexure table 5). The synthesis of first-strand cDNA was performed with 1 ug of template RNA by using SuperScript III first-strand synthesis system (Invitrogen, USA) as per manufacturer's instruction with some modifications. To increase reverse transcription efficiency, a pulsed RT reaction was carried out with following conditions: single step of 30 min at 16^{o}C, followed by pulsed RT of 60 cycles of 30s at 30°C, 30s at 42°C, 1s at 50°C (Vakonyi-Gasic et al.,2007). After this, for the inactivation of the reverse transcriptase, the reaction was incubated at 85°C for 5 min. The qRT-PCR was done with SYBR Premix Ex Taq (Clontech, USA) as per the manufacture's instruction. The qRT-PCR was conducted in ABI7500 Fast system (Applied Biosystems) with the following thermal protocol: 95°C for 2 min followed by 40 cycles of amplification of 15 s at 95°C and 1 min at 60°C. Immediately after final PCR cycle, a melt curve analysis was performed from 60 to 95°C in increments of 0.5°C to confirm the specificity of PCR products.The reactions were carried out in triplicate, and the experiment was repeated at least twice. Also, the control reaction with absence of template and reverse transcription were included for individual miRNA. In this study as an internal reference, U6 snRNA gene was used. All the stem loop primers used in the study are listed in Annexure 4. After the completion of reaction, the comparative Ct method $2^{-[\Delta\Delta Ct]}$ method was used to quantify the relative expressionof individual miRNA gene (Livak et al., 2001). For target gene expression analysis the qRT-PCR was performed as mentioned in chapter 3 section 2.9B.

3. Results

3.1 Sequencing and discovery of miRNAs in *Capsicum* species

The miRNA sequencing analysis yielded a total of 66,550,467 and 61,791,387 raw reads from the four tissues (leaf, flower, fruit and stem; Fig.6.1) of *C. chinense* and *C. frutescens*,respectively. After filtering the adaptor sequences and low quality reads and sequences of <18 and >30 nucleotide lengthfrom the raw reads,a total of 5,220,387 reads in *C. chinense* and 10,557,693 reads in *C. frutescens* clean high quality sequences were obtained (Table 6.1). Furthermore, after removing the redundant sequences,the remaining cleaned unique reads sequences of 3,025,801 in *C. chinense* and 6,614,813 in *C. chinense* were used to identify conserved and novel miRNAs.

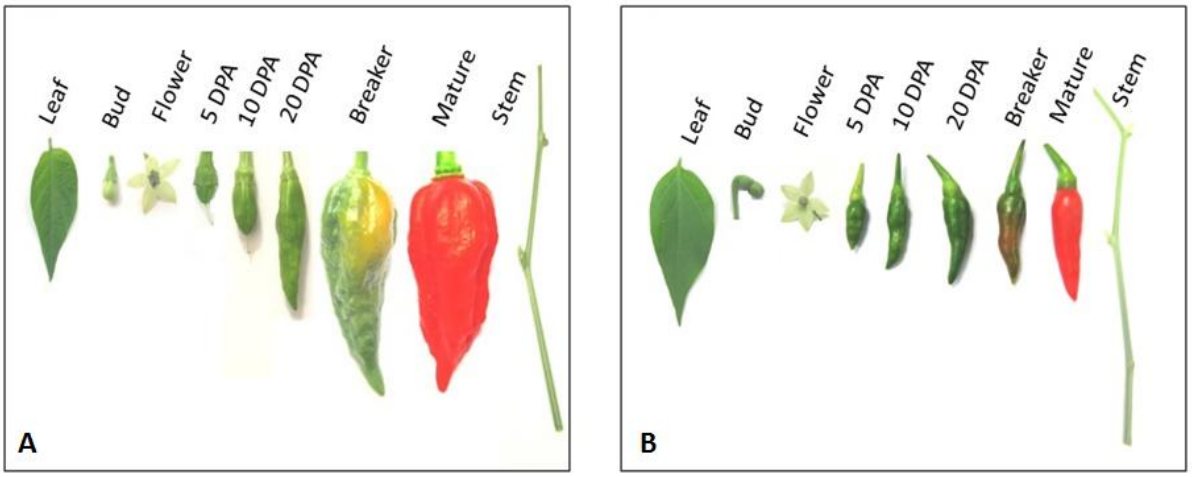


Figure 6.1 The leaf, flower, stem and different fruit developmental stages (A) *C. chinense* and (B) *C. frutescens* used for miRNA analysis.

The size distribution analysis of these miRNA displayed considerably identical pattern of length distribution in among all the tissue libraries (Figure. 6.2A and 6.2B**)**. Of the miRNA reads length between 18 to 30nt, the major fractions of reads were of 21 to 24nt indicating the characteristics of Dicer like protein (DCLs) processed miRNAs (Axtell, 2013) in all the libraries. The 24nt miRNA class was the most abundant(55% in *C. chinense* and 59% in *C. frutescens*) among all the analyzed miRNA libraries, followed by 23nt sRNAs (7.1% in *C. chinense* and 8.2% in *C. frutescens*), 22nt miRNAs (6.2% in *C. chinense* and 7.6% in *C. frutescens*), and 21nt miRNAs (6% in *C. chinense* and 6.6% in *C. frutescens*). The finding of majority of 21 to 24nt miRNAs both in *C. chinense*(74%) and *C. frutescens* (82%) confirms that they are exclusive cleavage products of DCL proteins.

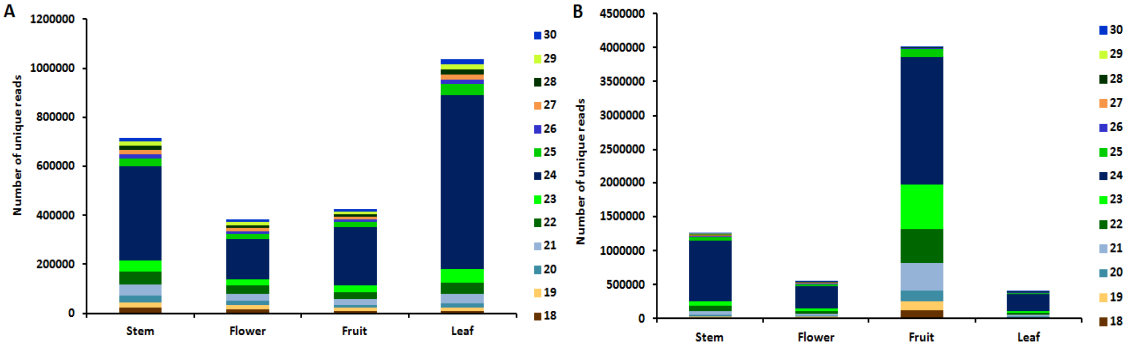


Figure 6.2 Size distribution (18-30nt) of total high quality unique miRNA reads in stem, flower, fruit and leaf tissues of (A) *C. chinense* and (B) *C. frutescens*.

3.2 Identification of conserved and novel miRNAs

These cleaned reads sequences of *C. chinense* and *C. frutescens* of 18-30 nucleotides length were further mapped onto the *C. annuum* reference gnome genome using UEA sRNA Workbench (http://srna-tools.cmp.uea.ac.uk/). About 100 bp flanking regions of each aligned miRNA reads were extracted from the *C.annuum* reference genome. For identification ofprecursor sequences of putative miRNAs criteria described by Meyers et al., (2008) was followed. To identify the potential miRNA precursor sequences, these extracted sequences were submitted to RNAfold programme of Vienna RNA package and the secondary hairpin structures were predicted. The minimum free energy (MFE) is one of the most important criteria to be considered during the identification of miRNAs which indicates the strength of hairpin structure (Llave et al., 2002; Adai et al., 2005). The miRNA with MFE less than -18 kcal mol^{-1} indicates the high confidence or genuine miRNA. In the current study, only miRNAs with MFE below –18 kcal mol^{-1} were retained. Using this criteria, the entire predicted miRNAs were aligned with *Viridiplantae* matured miRNA sequences retrieved from the miRbase-21(www.mirbase.org) using NCBI-BLAST (Griffiths-Jones et al., 2008) These sequences with 0 to 2 mismatches with known plant miRNAs were described as conserved miRNAs. Further, for the generation of high-confidence miRNA dataset, miRNAs with less than 10 read count in tissue samples were excluded and finally the number of miRNAs identified in each tissues were reduced in *C. chinense* and *C. frutescens* (Table 1).

Over all, we identified a total of 531 and 432 conserved miRNAs from four tissues of *C. chinense* and *C. frutescens*, respectively. Several of these miRNAs were found to expressin multiple tissues. Our analysis found that the largest number of the *C. chinense*

miRNAs were conserved with the reported miRNAs of soybean (80) followed by potato (68), rice (67), *Medicago* (53) and *Arabidopsis* (52), whereas more number of *C. frutescens* miRNAs were found conserved with that of potato (83) followed by soybean(82), rice (77),*Brachypodium* (61), and *Arabidopsis* (59) (Figure 6.3A and 6.3B).

The remaining unaligned distinct reads showing 3 or more mismatches (except in seed region of miRNA), andno homology with any of the earlier reported plant miRNAs are considered as novel and were further analyzed to predict secondary hairpin loop structure andMFE values were calculated using RNAfold as criteria mentioned above. After that the sequences fulfilling above criteria were submitted to MIREAP 0.2 programme for the prediction of novel miRNAs. Finally a total of 521 high confidence non-redundant novel miRNAs in *C. chinenese* and 159 novel miRNAs in *C. frutescens* were identified.These total miRNA populations high confidence putative miRNAs. The representative secondary structure of *C.chinense* and *C. frutescens* novel miRNAs are depicted in Figure 6.4A and 6.4B. respectively.

Table 6.1 Summary statistics ofsequencing and identification of potential miRNAs in *C. chinense* and *C. frutescens*

	C. chinense				***C. frutescens***			
	Stem	**Flower**	**Fruit**	**Leaf**	**Stem**	**Flower**	**Fruit**	**Leaf**
Total Reads	18987696	13325106	10943078	23294587	17654289	15504035	18952786	9680277
Trimmed Unique Reads	1340217	833591	846114	2200465	2092477	1242870	6419797	802549
Reads aligned to genome	1221789	765790	781783	1898614	1862097	1079187	5847905	727724
Reads aligned to mirBase	19345	14320	14374	22518	28666	14409	46347	11957
Conserved miRNA	**196**	**326**	**288**	**165**	**233**	**135**	**307**	**170**
Reads utilized for Novel miRNA	591006	337601	368108	828390	917586	413928	2769405	304876
Novel miRNA predicted	**92**	**353**	**119**	**16**	**39**	**53**	**41**	**47**
Total miRNAs predicted	**288**	**679**	**407**	**181**	**272**	**188**	**349**	**217**

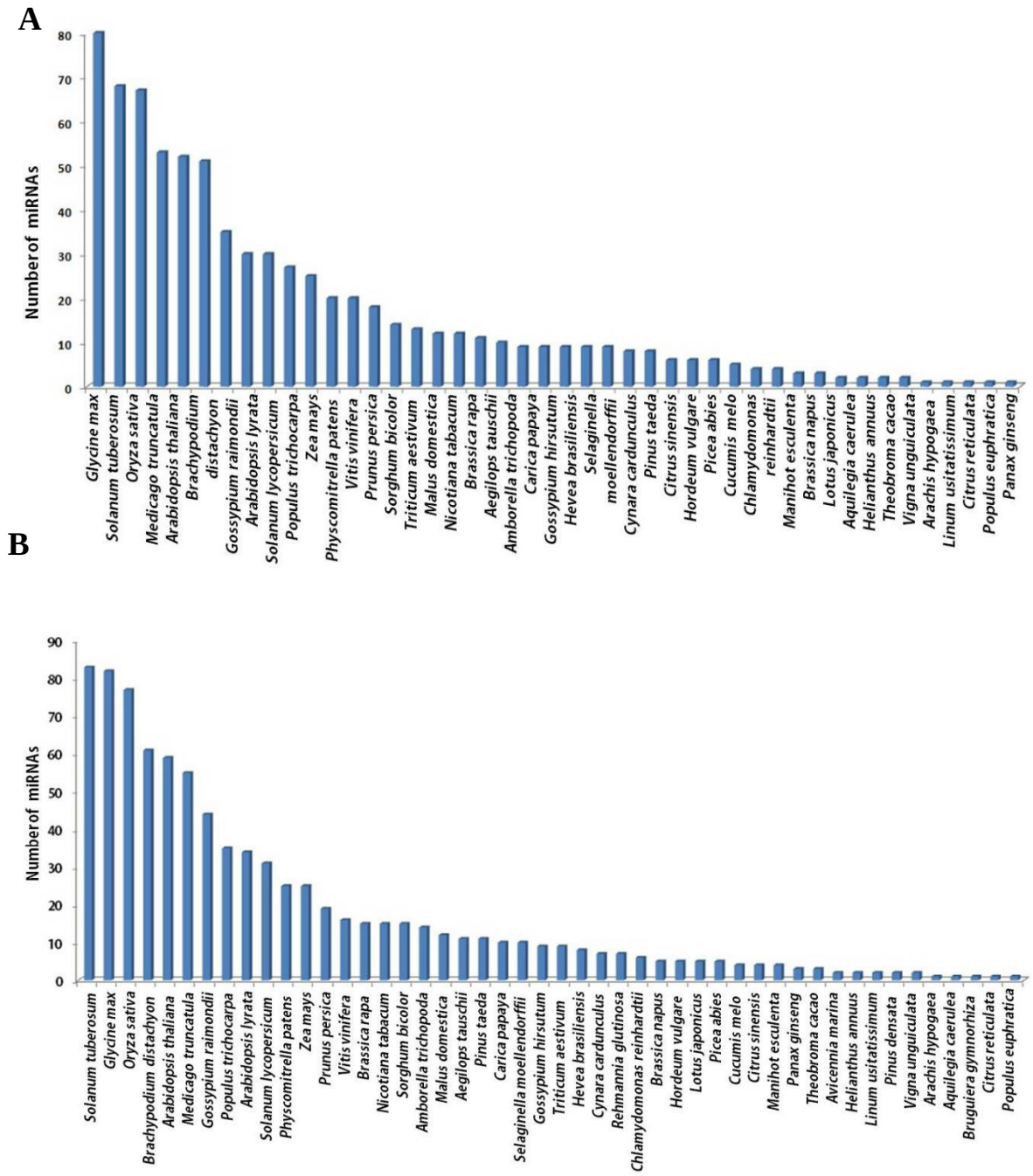


Figure 6.3 The number of conserved miRNAs from *C. chinense* (A) and *C. frutescens* (B) with known plant miRNAs of other plant species

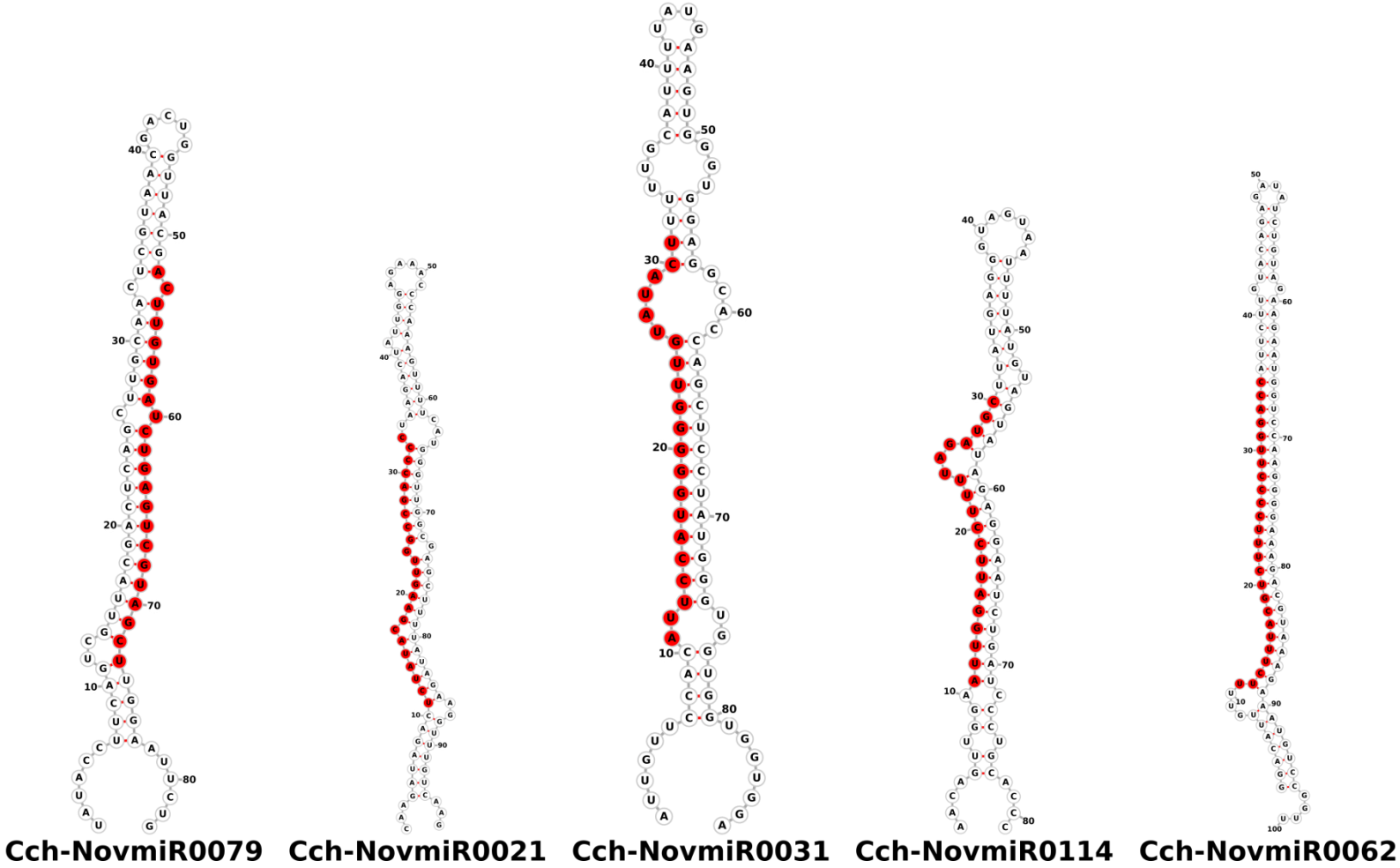


Figure. 6.4. Representative hairpin structures of novel miRNAs from (A) *C. chinense* and (B). *C. frutescens*

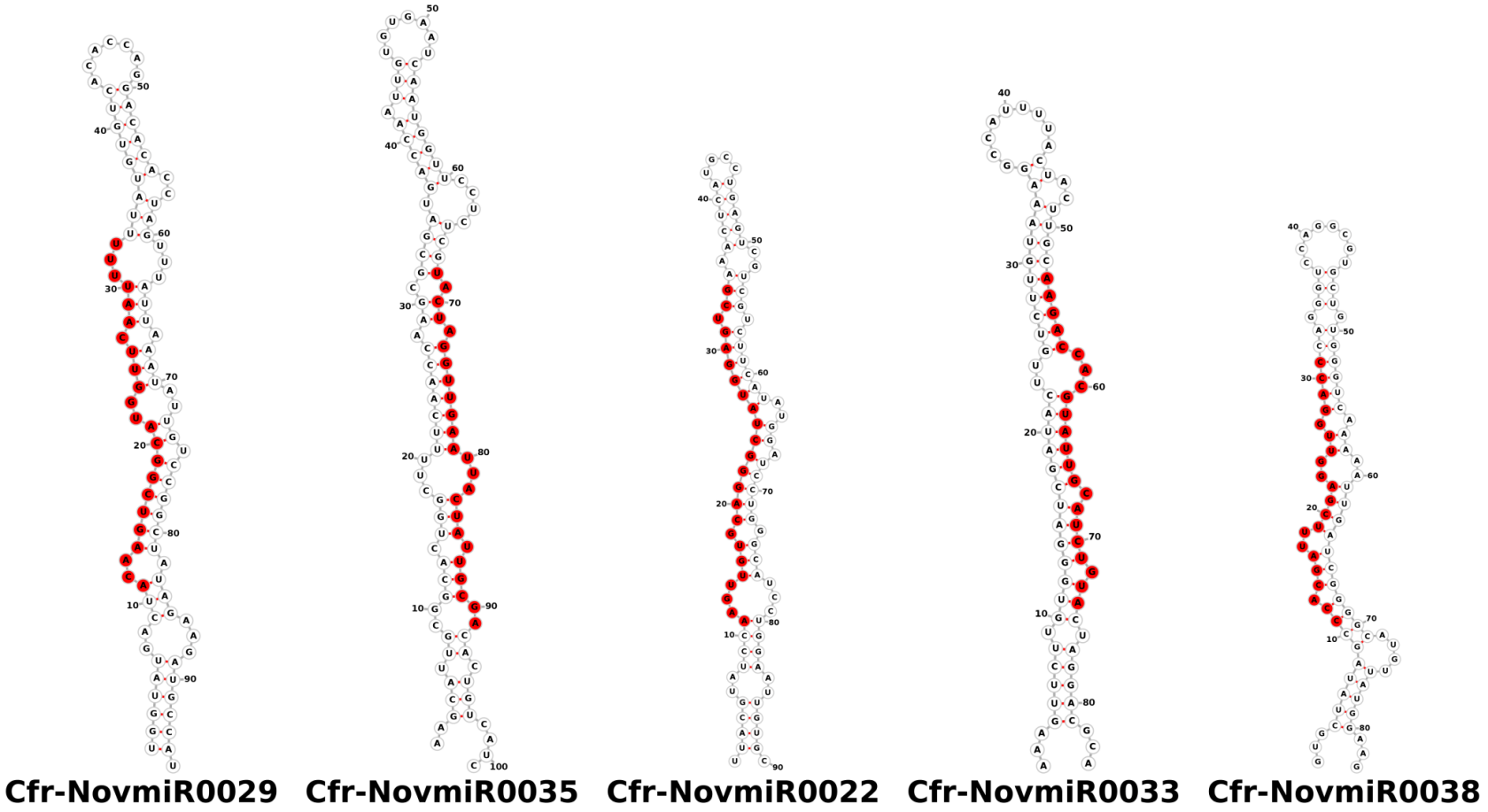


Figure 6.4B Continued…

3.3 Characteristics of miRNAs

Altogether, a total of 1054 in *C. chinense* and 591 in *C. frutescens* non-redundant miRNAs was identified across the tissues. This entire set of miRNAs from different tissues displayed broader distribution which varied from a minimum of 181 (in leaf) to maximum 680 (in flower) of *C. chinense*. However, in the case of *C. frutescens* the minimum 217 miRNAs in leaf tissue to a maximum of 340 miRNAs in fruit tissue were identified. The higher number of novel miRNAs in *C. chinense* was identified in flower (353) followed by fruit (119) and stem (92). Similarly, in *C. frutescens* the highest number was identified in flower tissue (53) followed by fruit (47) and leaf (41) (Table 5.1). The number of tissue specific, expressing in more than one tissue but not common to all tissues, and common common in all tissues are depicted in the figure 6.5A and 6.5B.

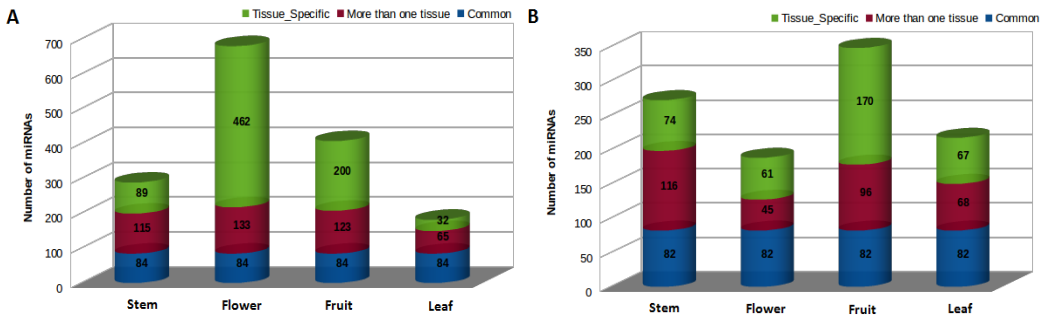


Figure 6.5 The number of tissue specific, found in more than one tissue but not all, and common miRNAs observed in *C. chinense* (A), and *C. frutescens* (B).

The miRNA size distribution analysis indicated that the major fraction of miRNAs are represented by 24 nucleotides in both *C. chinense* and *C. frutescens* (Fig. 6.6A and 6.6B). Furthermore, substantial variation in the proportion of 24nt class miRNAs in different tissues was observed. The distribution analysis revealed that in the stem (21%) and leaf (20%) tissue their proportion is similar,howeverin flower (29-31%) and fruit tissue (32-34%) the proportion was quite high for both *C. chinense* and *C. frutescens* The high level of 24nt miRNAs in reproductive tissues compared to vegetative tissues indicates that reproductive tissues require more distinct repression of this class. The base specificity analysis revealed the majority *Capsicum* small RNAs exhibited a characteristic U/A at 5'terminal, which are indication of genuine micro RNAs (Fig. 6.6A and 6.6B).

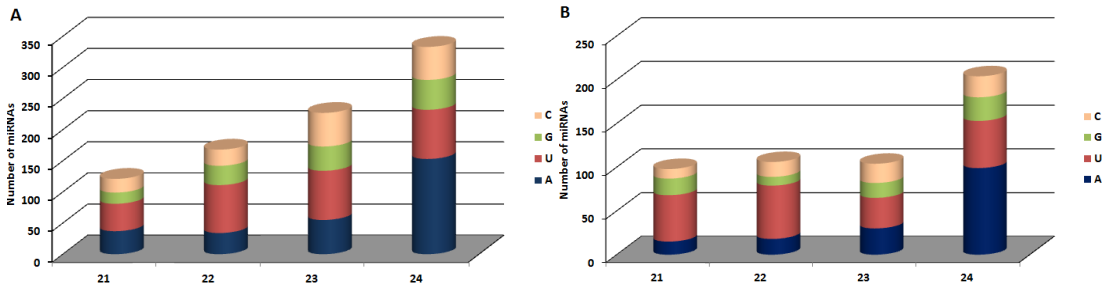


Figure 6.6 Size distribution and base specificity analysis of miRNAs in *C. chinense* (A) and *C. frutescens* (B)

Additionally, we have investigated nucleotide constitution of mature miRNAs in *C. chinense* and *C. frutescens*. As per Watson–Crick base pairing rule, GC composition is a crucial factor for the strength of RNA secondary structure, and they might have a significant control on sRNA biology, especially Dicer intervened dsRNA cleavage. The majority of the *Capsicum* miRNAs has GC content in the range of 35-70% which is similar to plants. The analysis revealed the average GC composition of miRNAs in *C. chinense* (44%) and *C. frutescens* (46%). Overall, our results indicated the prediction of high-confidence miRNAs in *Capsicum*. Detailed list of GC content of all the identified miRNAs in *C. chinense* and *C. frutescens* is given in Annexure Table 14 and 15 (compiled in CD optical drive for vetting).

In order to identify the common and specific miRNAs between *Capsicum* species, we performed comparative analysis of miRNAs. The *C. annuum* miRNA information were collected from the previous sRNA studies (Hwang et al., 2013 and Liu et al., 2017) while, our miRNA dataset were used for *C. chinense* and *C. frutescens.* The venn diagram showed 951 (50.6%), 484 (25.7%), and 334 (17.8%) miRNAs are specific to *C. chinense*, *C. frutescens*, and *C. annuum*, respectively (Figure 5.7). About 81 (4.3%) miRNAs are common in between *C. chinense* and *C. frutescens*, followed by 10 (0.5%) miRNAs in between *C. frutescens* and *C. annuum*, and 8 (0.4%) miRNAs between *C. annuum* and *C. chinense*. We observed about 12 miRNAs are common in all three species. The analysis revealed a major portion of miRNAs are species-specific while relatively small fraction is in common between *Capsicum* species.

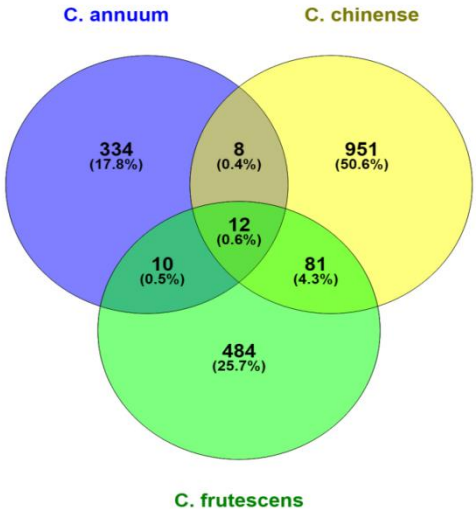


Figure 6.7 Venn diagram showing common and species specific miRNAs in *Capsicum* species.

The precursor sequences (pre-miRNAs) of miRNAs were investigated to identify miRNA distribution and location across the genic and inter-genic regions of the genome. Combining genomic annotation information with corresponding coordinates, the miRNA distributions were discovered. In *C. chinense*, we observed that about 863 miRNAs (81%) were located in intergenic regions and only 37 were located in genic regions (Fig. 6.8A). In the genic regions, 34 of the miRNAs were located in introns and 13 in exonic regions. Similarly, in *C. frutescens* 498 miRNAs (84%) were originated from intergenic regions whereas only 19 were originated from genic regions of which 13 are from the intronic region while 6 are from exonic regions (Fig. 6.8B). Our results suggested that major proportion of the miRNAs in *C. chinense* and *C. frutescens* are derived from non-coding regions of the genome.

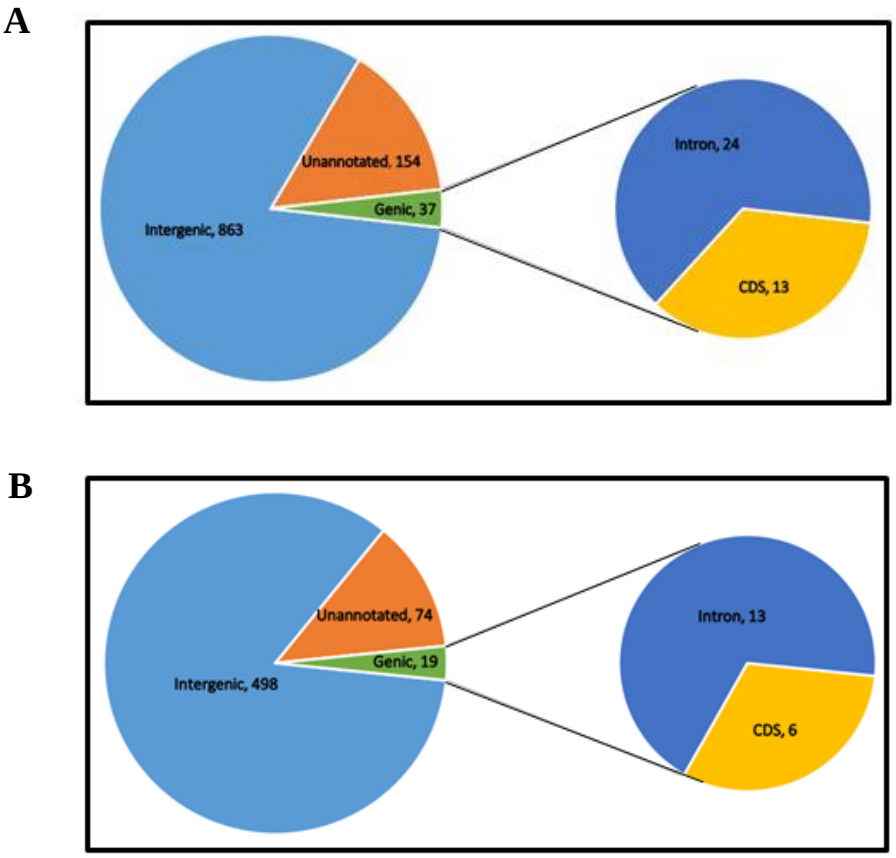


Figure 6.8 Genomic distribution of *Capsicum* miRNAs. The Venn diagram of miRNAs distributed in different genomic locations of the (A). *C. chinense* and (B). *C. frutescens* is shown.

To assemble novel miRNAs into families, we performed sequence clustering analysis with all identified miRNAs. We observed each conserved family comprised of variable number of miRNAs. The number varies from minimum two to maximum twenty nine miRNAs per family (Table 2 and 3). Due to recent evolutionary emanation and stringent similarity criteria, most of the identified novel miRNAs did not clustered with conserved families and described as unique novel miRNA. While, a few novel miRNAs was clustered with conserved miRNA families (Figure 6.9). The length of miRNA family and nucleotides at 3'-end varies within and between families. We found representation of 21 to 24nt variants of novel miRNAs in the conserved miRNA families. A major proportion of 22 and 24nt novel miRNAs were found among miRNA families such as thirteen members of miR168 family, eight members of miR390 family (with two novel miRNAs of *C. chinense* and *C. frutescens* clustered together) and fifteen members of miR169 family, etc. The variable size of miRNAs is responsible for distinct functions and role in plant system (Chen, 2009; Cuperus et al., 2011).

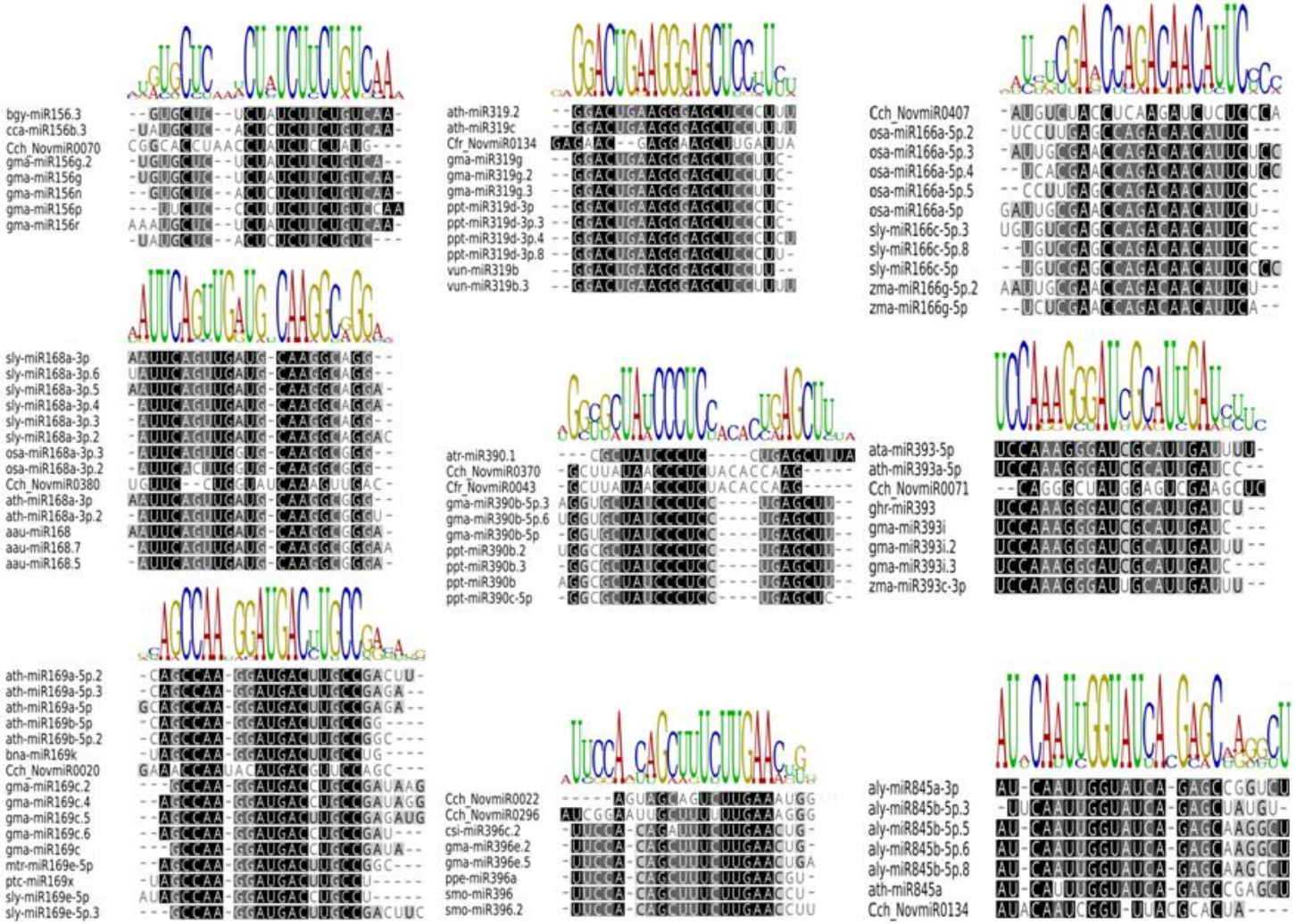


Figure. 6.9 Conserved miRNAs of *Capsicum* species with other known plant miRNAs.

3.4 Genomic distribution of miRNAs

To study genomic distribution of identified miRNAs in *Capsicum* species, the genomic coordinates of each miRNAs was identified in the *C. annuum* reference genome (Qin et al 2014). Our analysis revealed that 87% of the *C. chinense* and 89% of *C. frutescens* miRNA were successfully mapped to twelve *Capsicum* chromosomes while remaining miRNAs were mapped on scaffolds (unassigned contig sequences) in the *C. annuum* reference assembly (Qin et al.,2014). The analysis indicated distribution and clustering of conserved and novel miRNAs throughout the twelve chromosomes of *C.annuum* genome (Figure 6.10 & 6.11). The miRNA clusters are constituted by the members of same miRNA family. We observed that some miRNA family clusters are strictly present on same chromosomes across the two species, for example, miR398 and miR6025 family located on chromosome 12 and 5, respectively of *C. chinense* and *C. frutescens*. However, some clusters of the miRNA family from *C. chinense* and *C. frutescens* were located either on the same or different chromosomes, for example, miR159 family of both the species has two clusters and were detected on chromosomes 3 and 6. The analysis suggests a high level of conservation of miRNA family in *C. chinense* and *C. frutescens*. Furthermore, we observed five clusters of miR166 family members i.e., single cluster each on chromosome 1, 4 and 8 and two clusters on chromosome 3;one large cluster of the miR408 family on chromosome 8, and on chromosome 1 two extensive clusters of miR168 family were identified for both *C. chinense* and *C. frutescens*. Besides this, one each large cluster of miR393 family on chromosome 5 and for miRNA397 family on chromosome 7 was observed in *Capsicum* genome. In few cases the novel miRNAs were clustered together with the conserved miRNAs i.e. Cch-NovmiR251 and Cfr-NovmiR0056 were clustered with miR482 on chromosome 6 for *C. chinense,*andchromosome 4 for *C. frutescens,* respectively, indicating miRNA rich region at the clustered chromosomal region of *Capsicum* genome. The miR159, miR169 and miR172 families were also observed clustering on the same chromosome of *C. chinense, C. frutescens,* and *S. tuberosum* suggesting a high degree of conservation of miRNA families in Solanaceae crop plants (Lakhotia et al., 2014).

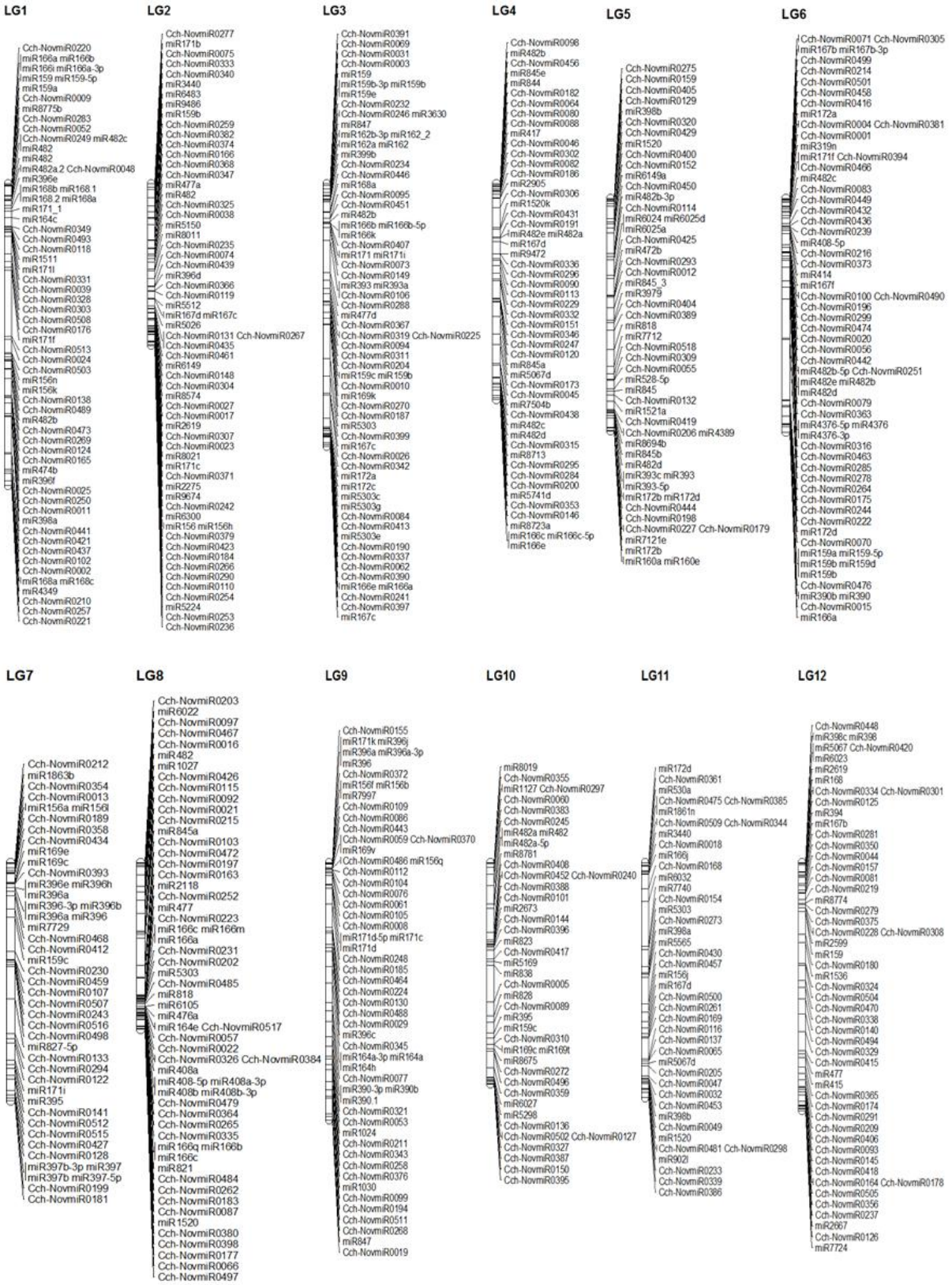


Figure 6.10 Distribution of miRNA from *C. chinense* on twelve chromosome of *Capsicum* genome. (Novel miRNAs from *C. chinense* are symbolized as Cch-NovmiRNA)

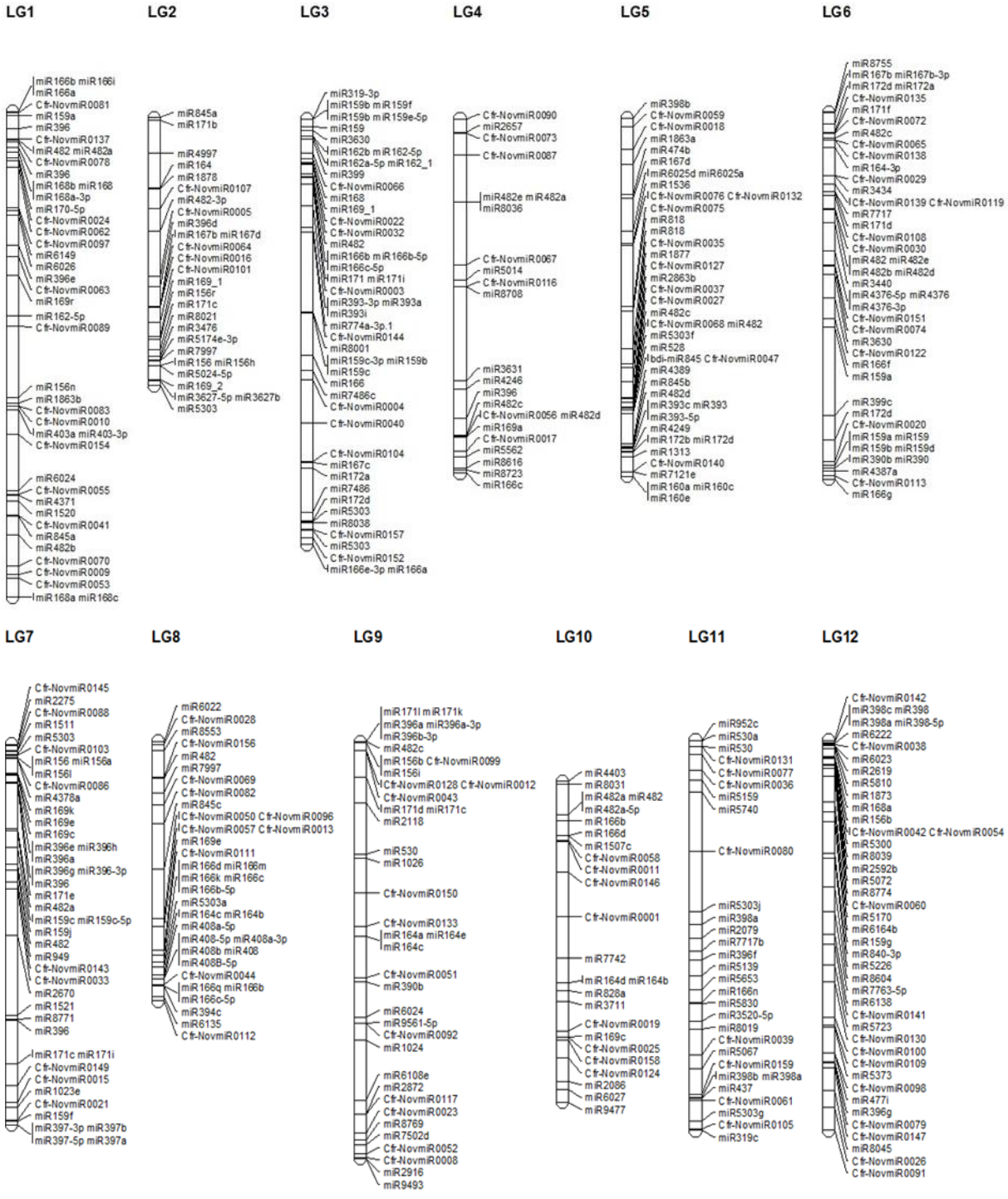


Figure 6.11 Distribution of identified novel and conserved miRNA families from *C. frutescens* on twelve chromosomes of *C. annuum* reference genome. (Novel miRNAs from *C. chinense* are symbolized as Cch-NovmiRNA).

3.5 Synteny analysis of *Capsicum* miRNAs with other Solanaceae plants

Further to confirm the conservation of miRNA families in Solanaceae plants, we performed comparative mapping of *Capsicum* miRNAs with tomato and potato genomes. The twenty two high-confidence miRNA families of *C. chinense* and *C. frutescens* were selected for the analysis. The precursor sequences of miRNAs were used to study the shared synteny among homologous *Capsicum*, tomato and potato miRNA genes (Fig. 6.12A and 6.12B). As shown in figures, conserved synteny between miRNAs was observed in majority of chromosomes of *C. chinense*, *C. frutescens*, tomato and potato genome. For example, the miR398 family found on chromosome 12 of both *C. chinense* and *C. frutescens* genome was also mapped to corresponding orthologous chromosome 12 of tomato and potato. The miR169 family located on chromosome 1, 2 and 7 of *C. frutescens* showed homologous sequence hit in the tomato and potato genome on the same chromosome. In our study, a considerable expansion of several miRNA families across chromosomes in the genome such as miR159 and miR172 was observed. These miRNA families were detected on more than one chromosomes i.e. miR159 found on chromosome 3, 5 and 6 of *C. chinense*, and chromosome 3 and 12 of *C. frutescens* and their corresponding orthologous tomato and potato chromosomes.

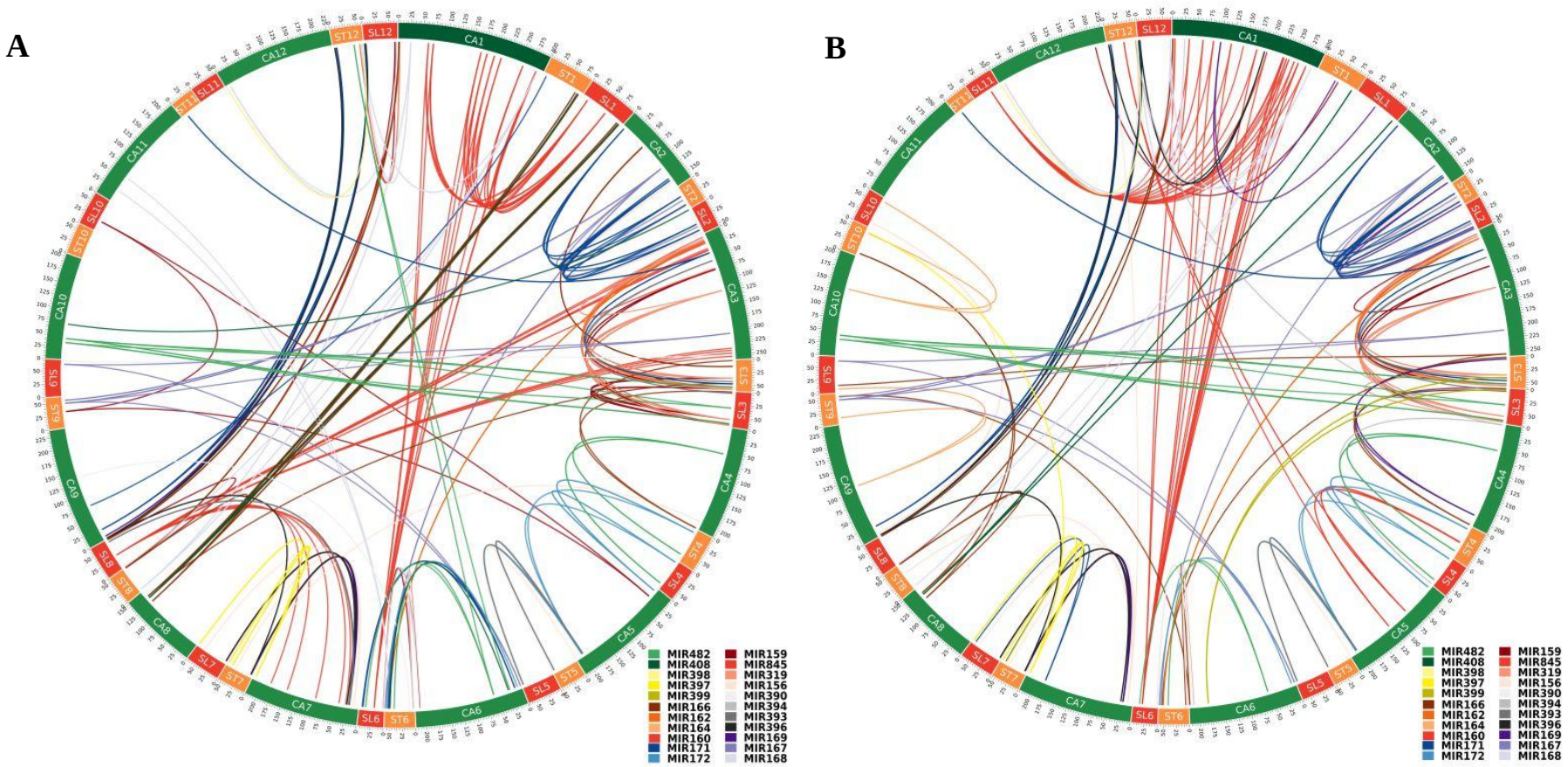


Figure 6.12 Comparative map of (A). *C. chinense* and (B). *C. frutescens* with tomato and potato genomes showing links between syntenic MIR genes. The links connect the locations of miRNA homeologs between genomes which is based on the comparison of sequence information of mapped miRNAs of *C. chinense* and *C. frutescens* with genome sequence of tomato and potato. Each miRNA family is colored uniquely.

3.6 Identification of potential targets of *Capsicum* miRNAs

To understand the role in multiple biological and cellular processes, the putative target mRNAs were identified for all the predicted miRNAs using psRNATarget tool with high stringent parameters (Mayers et al., 2008). In our study we were able to identify all the *C. chinense* and *C. frutescens* miRNAs target. The transcript sequences of *C. annuum* Zunla-1 were used as a reference genome (Qin et al., 2014). It was observed that approximately 80% of *Capsicum* miRNA targets were predicted to be governed by cleavage mechanism and the remaining through translational inhibition. The target prediction analysis showed variation in the number of identified targets for each miRNAs i.e. the targets differs from minimum one to maximum 30 (Annexure Table 16 and 17; Please refer to CD drive for further details). The analysis revealed that the highest fractions (25%) of targets are genes involved in the transcription.*Capsicum* miRNAs target mRNA/gene with diverse functions belonging for an extensive series of proteins, like those associated with inorganic ion, metal, and amino acid transport, response to diverse stress, RNA processing and modification, post-translational modification, signal transduction pathways, carbohydrate and secondary metabolite biosynthesis were identified.

The majority of the conserved *Capsicum* miRNAs indicated targeting protein with identical functions, as observed in other plants. For instance, we observed the miR156 family targets SBP transcription factors; miR160 targets ARF transcription factors, miR164 targets NAC, miR159 targets MYB, miR171 targets Scarecrow/GRAS, miR165/166 targets homeobox (HB), and miR172 targets AP2 in *Capsicum* (Annexure Table 16 and 17). The family members of miR167, miR168, miR169, miR172, and miR390 targets genes transcribing F-box proteins harboring distinct conserved domains, like, WD40 and LRR.These F-box proteins are involved in proteasome-guided degradation pathway (Lechner et al., 2006) and hormone signalling pathways (Wang et al., 2009). Moreover, we were able to identify novel targets of some conserved miRNA family members. For instance, besides SBP proteins, several other transcription factor genes such as MYB, ERF, bHLH, RF2a, MADS-box and WRKY were found to be targeted by miR156 family. Further we observed *Capsicum* miR171 and miR396 family members' targets GRAS transcription factors and sulphate transporter protein genes. The miR482 targets copper superoxide dismutase enzymes that are involved in assimilation of inorganic sulphate (Kliebenstein et al., 1998). Our target prediction analysis suggests that several potential targets of conserved miRNAs were believed to be a pepper-specific in nature. Like conserved targets, the novel pepper-specific

targets were also found to be enriched in transcription factors, but along with this the targets comprised of mRNAs encoding the helicase protein, gypsy/Ty-3 retroelement polyprotein, DNA methyltransferase, splicing factor, and copper/ferrous transporter, suggesting that the analogous novel targets of conserved miRNAs might be associated with particular biological/cellular processes in pepper. Intriguingly, we found the miR164 and miR396 novel miRNAs targets flowering time control protein indicating the vital role of miRNAs in flower development. We observed that miR396 targets DNA methyltransferase gene in both *C. chinense* and *C. frutescens*. The DRM methyltransferase gene has high sequence complementarity with miR396 in *Capsicum* and tomato as compared to tobacco and potato. The miR396 associated with methyltransferase may be involved in cytosine methylation process, thereby influencing transcriptional silencing of repetitive elements in complex *Capsicum* genome.The analysis confirms the vital role of conserved miRNAs in fundamental biological and cellular processes.

To identify statistically significant over-represented GO terms we performed functional enrichment analysis with agriGO tool (Du et al., 2010). In *C. chinense*, the GO enrichment analysis revealed a significantly higher abundance of biological process terms such as cellular responses to stress and stimulus followed by transcription and their regulation from RNA polymerase II promoter (Fig. 6.13A). Moreover, in the cellular component category, an 'integral component of the membrane' term was over-represented, followed by nucleus and plasma membrane (Fig. 6.13A). While, in the case of molecular function category, serine/threonine and protein kinase activity, ATP binding, DNA binding and adenyl ribonucleotide binding activity, was abundantly represented. Similarly, in *C. frutescens* the GO enrichment analysis of biological process followed the same pattern except protein amino acid phosphorylation is highly enriched followed by transcription and their regulation and hormone-mediated signaling pathway (Fig. 6.13B). The identical representation of processes from cellular component class and molecular function category was observed in *C. frutescens*, and no obvious difference could be found in the enrichment analysis of these two *Capsicum* species except these terms are found to be highly enriched in *C. frutescens* compared to *C. chinense* (Fig. 6.13B). The top thirty most abundant gene ontology processes of *C. chinense* and *C. frutescens* is given in Figure 6.14A and 6.14B.

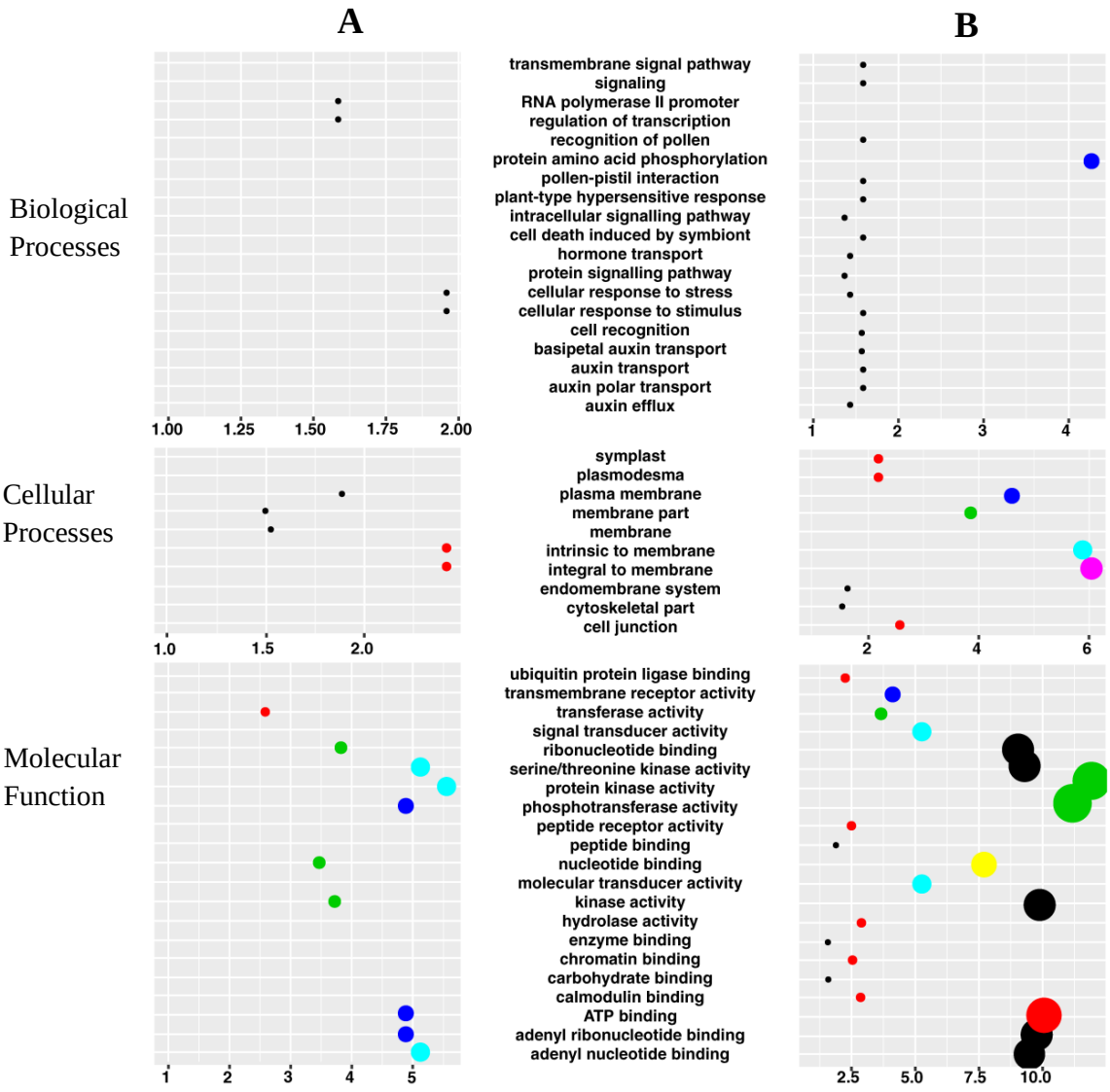


Figure 6.13 The Gene ontology enrichment analysis of the target genes of identified miRNAs in (A) *C. chinense*, and (B) *C. frutescens.* The significantly enriched terms obtained using agriGO were summarized and visualized as a scatter plot using R programme.

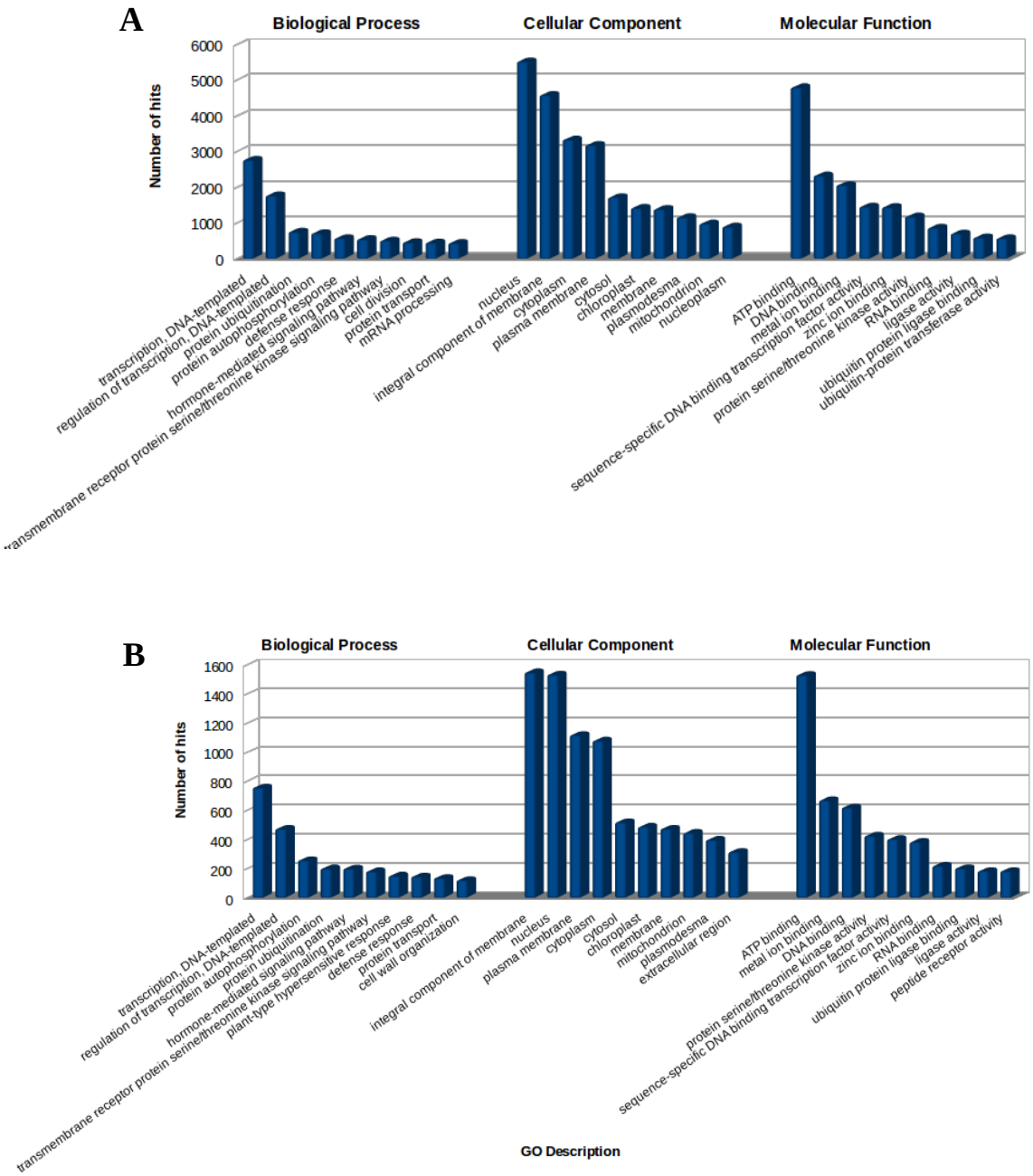


Figure 6.14 Most abundant (top 30) miRNA targets involved in biological process, molecular function, and cellular component GO terms in (A) *C. chinense*, and (B) *C. frutescens*.

About 41 diverse miRNA target transcription factor families were identified in *C. chinense* and *C. frutescens*. In *C. chinense*, it comprises majorly the members of MYB and NAC (both 14%) and followed by various other members like bHLH (8%), ARF (5%), C3H (5%), TCP (4%), ABI3 (4%), HB (4%) and others (Fig. 6.15A).In our analysis a minimum of seven members from about 25 transcription factor families were identified as putative targets of *C. chinense* and *C. frutescens* miRNAs. Similarly, in *C. frutescens* the identified major

targets were NAC (15%) and MYB (13%) transcription factors (Fig. 6.15B). The transcription factor abundance analysis signifies that MYB and NAC play very crucial role in overall development of *Capsicum* plant. The remaining target transcription factor family includes ARF, bHLH and TCP constitutes of 7% of the total transcription factor families. The AP2 (6%), C3H (5%) and HB (4%) were frequently targeted by miRNAs and contributes to the entire 37 TF families of *C. frutescens* miRNA families. We found 70 % (15 of 21) and 52 % (11 of 21) of total annotated ARF members in the *C. chinense* and *C. frutescens* genome which was predicted as the miRNAs targets.

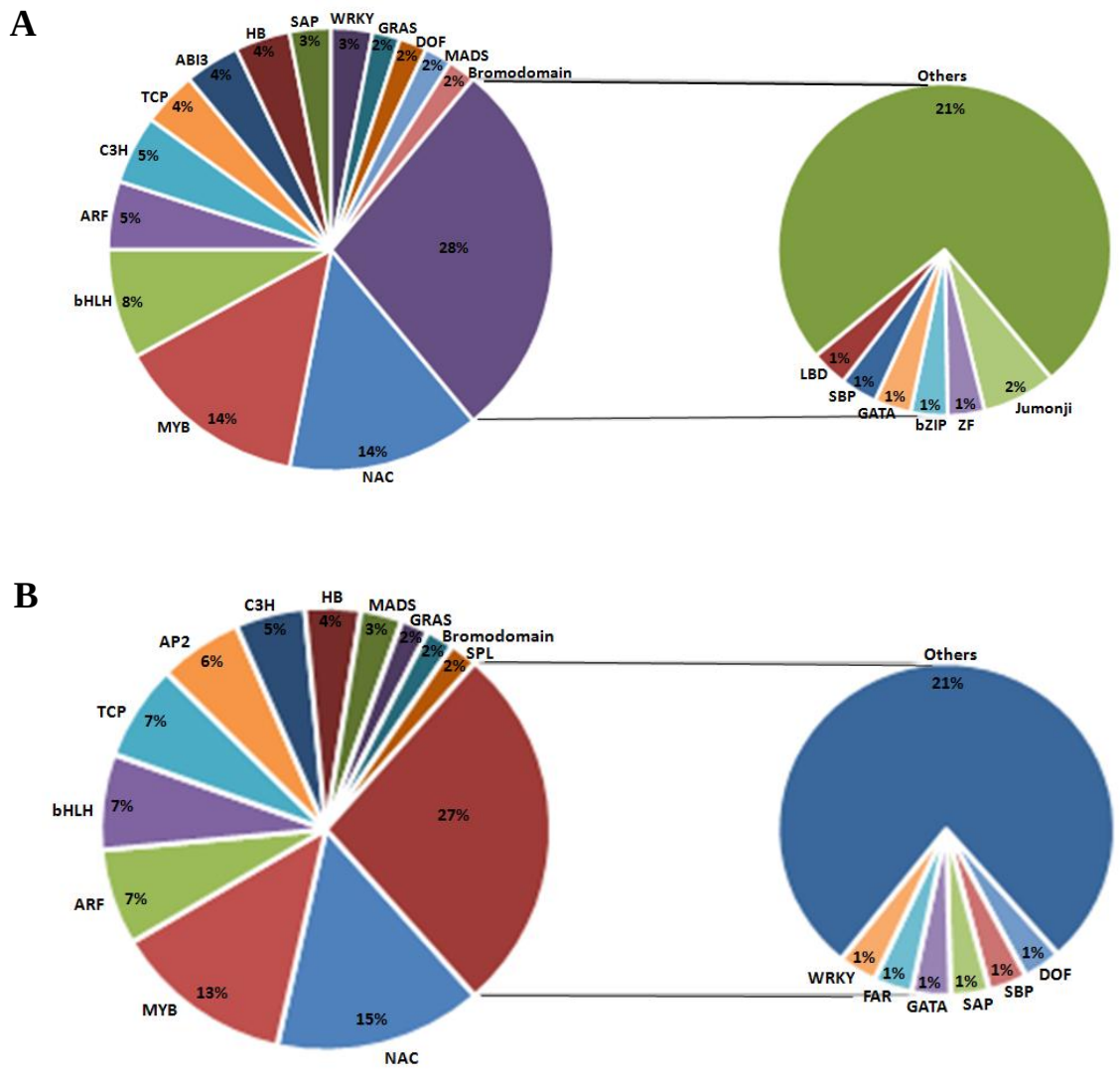


Figure 6.15 Proportions of different predicted miRNAs targets transcription factor family in (A) *C. chinense* and(B) *C. frutescens.*

3.7 Differential expression analysis of *Capsicum* miRNAs

The expression profile of all the identified miRNAs from *C. chinense* (1054) and *C. frutescens* (591) were investigated to study the potential roles of miRNAs in plant development. Each miRNAs expression/read count in tissue was normalized through DESeq approach (Anders and Huber, 2010). We found significant diversity in the expression profiles of *Capsicum* miRNAs. The expression abundance analysis categorize the gene expression levels in five classes:- 1) very low, 2) low, 3) moderate, 4) high, and 5) very high. Among this, the largest proportion (43–54%) of miRNAs displayed very low expression abundance throughout the tissues of both *Capsicum* species followed by low expression category (18–29%). Moreover, the moderate expression varies between 8-13% followed by high (7-13%) and very high expressed (5-11%) miRNAs (Fig. 6.16A).

Further, we observed broad differential expression profiling of miRNAs in *Capsicum* species. Variations in miRNA expression from ubiquitous to tissue or organ specific were observed. We detected variations in the expression level of novel miRNAs among the analyzed tissues as contrast to the conserved miRNAs. To identify tissue-specific miRNAs, the low expressed miRNAs were eliminated from the analysis leading to the selection of high-confidence miRNAs. The tissue preferential categories were assigned to individual miRNA as demonstrated by Breakfield et al., (2012). Our analysis showed an extensive population of miRNAs specifically expressed in one/several other analyzed tissue types of *Capsicum* species. In *C. chinense*, about 43% miRNAs were preferentially expressed in flower tissue followed by 19% in fruit, 9% in the stem and 8% in leaf (Table 2). In contrast, in *C. frutescens* 16% miRNAs were preferentially expressed in fruit tissue followed by 8% in the stem, 7% in leaf and 6% in flower. The tissue specificity analysis revealed that, In C chinense, the flower tissue was abundant in tissue-specific miRNAs followed by the fruit, and in the case of *C. frutescens;* fruit tissue was abundant with tissue-specific miRNAs followed by the stem (Fig. 16B). About 18% and 13% of miRNAs were expressed in more than one tissue both in *C. chinense* and *C. frutescens*, respectively. Besides this, 8% and 7.7% miRNAs are expressed in all the tissues of *C. chinense* and *C. frutescens*. This specificity analysis implies that the proportions of tissue-specific/developmental stage-specific expression of miRNAs are significantly higher compared to other types.Further, we detected differential expression of several miRNAs between flower and fruit tissues in both *Capsicum*

species, indicating a rigid control on the miRNA-mediated regulation of flower and fruit development in *Capsicum* species.

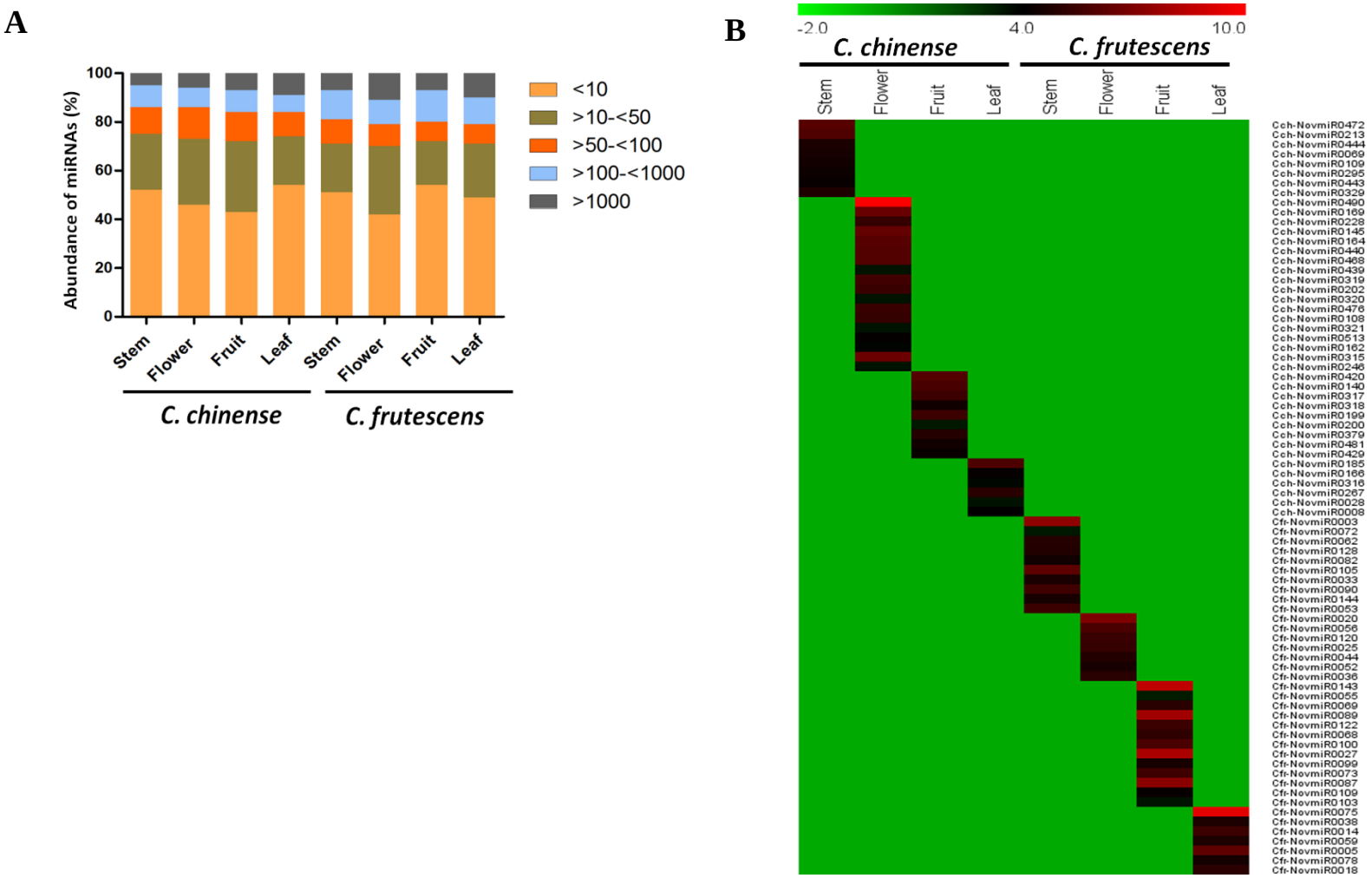


Figure 6.16. Expression analysis of *Capsicum* miRNAs. A). Percentage of miRNAs with different level of expression abundances among tissues of *C. chinense* and *C. frutescens.* B). Representative diagram of tissue specific novel miRNAs in *Capsicum* Species. Heat map showing expression profiling of tissue specific novel miRNAs in $\log_2$ transformed normalized expression values .The bar correspond to the degree of the $\log_2$ transformed expression levels for individual miRNA.

It was observed that members of miRNA from the same family do not always display identical expression profiling. The analysis of the expression profiling of miRNA families with four or more miRNAs showed this result. Further, miRNA families showing differential expression throughout the *Capsicum* tissues were also observed (Fig. 6.17) For example, miR156l majorly expressed in flower and stem of *C. chinense* compared to the *C. frutescens*; in contrast, miR156f was expressed in all the tissues apart from the flower of *C. chinense*. A significantly high expression of miR156b was detected in the stem and fruit of *C. chinense*, and leaf tissue of *C. frutescens*. The high expression of miR156i was observed in fruit and stem tissues of both *Capsicum* species but found low expression in leaf tissues. In the case of miR159 family members, miR159b found to be expressed at a high level throughout the tissues of *C. chinense* and *C. frutescens*, whereas expression of miR159j was limited to fruit and stem tissues of *C. frutescens*. miR159c was expressed only in the leaf of both *Capsicum* species, and stem of *C. frutescens*, while in remaining all tissues the expression could not be observed. A high expression of the miR482 family was observed in all the tissues of both *Capsicum* species suggesting their major role in *Capsicum* plant development. Overall, in stem tissue, the higher expression of miR482 family members was detected compared to other tissues.

3.8. Validation of miRNAs by qRT-PCR analysis

To validate the digital expression profile of *Capsicum* miRNAs, qRT-PCR analysis was performed. We have selected 21 miRNAs from all the tissue samples from both the *Capsicum* species. Five miRNAs showing differential expression across all the tissues of *C. chinense* and *C. frutescens*, and 16 tissue specific or developmental stage preferential miRNAs (revealed by digital expression/tissue specificity analysis) were selected from both the *Capsicum* species. The qRT-PCR analysis revealed that differentially expressed novel miRNAs are expressed at moderately high level in the flower and fruit tissues (Fig. 6.18A). This suggests that they might play a crucial role in flower and fruit development and differentiation. Apart from this, we detected significant fruit preferential expression of novel miRNAs such as Cch-NovmiR0140 and Cch-NovmiR0199 in *C. chinense* and Cfr-NovmiR0027 and Cfr-NovmiR0089 in *C. frutescens* compared to other tissues (Fig. 6.18B). This result indicates that these novel miRNAs and their corresponding target gene may play significant role in fruit development of *Capsicum* species.

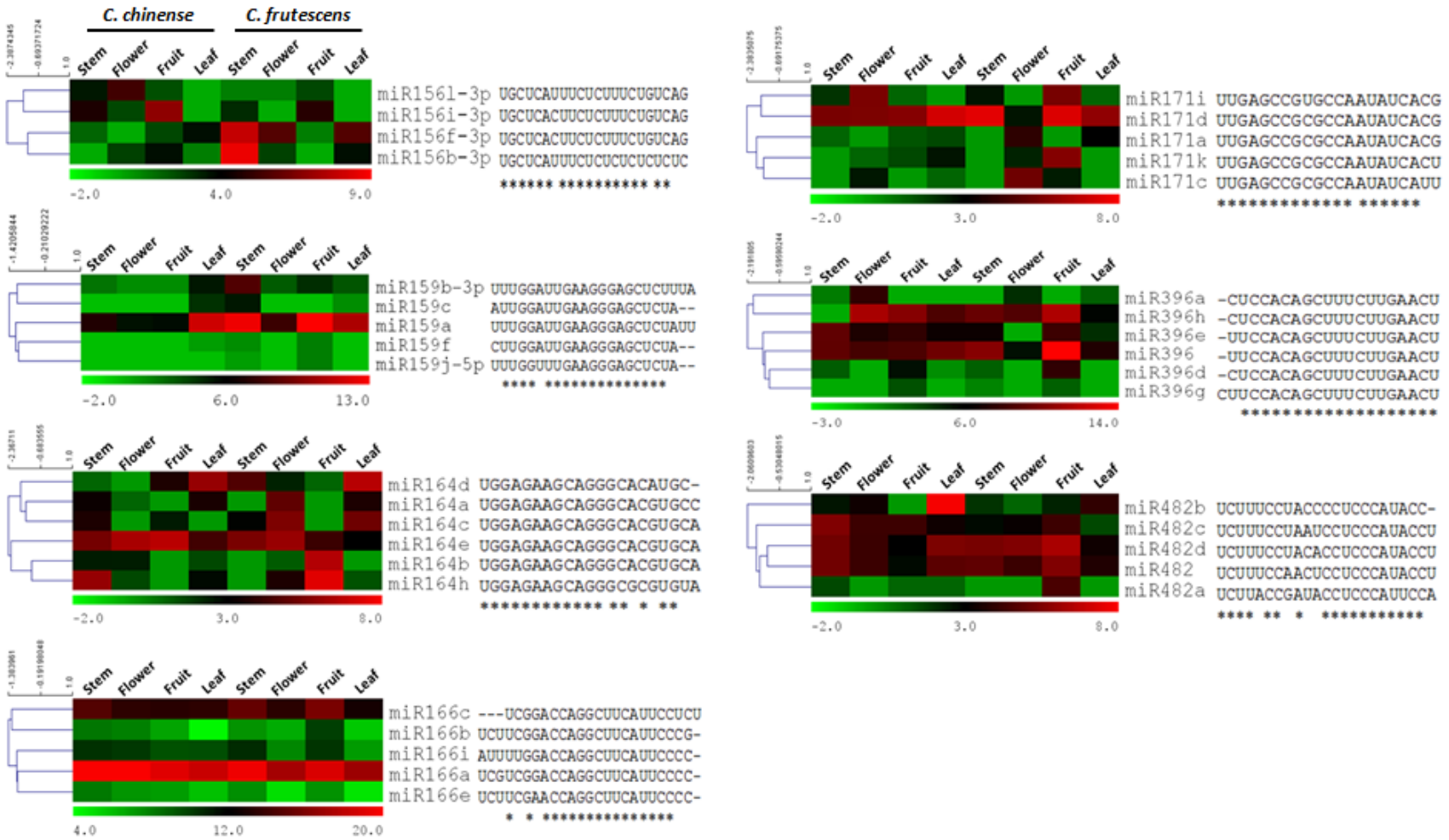


Fig. 6.17 Heat map showing differential expression of miRNA members from the same family. Heat map showingmiRNA expression data from leaf, flower, fruit and stem from *C. chinense* and *C. frutescens*. The name of each miRNA family is mentioned on the left side while its sequence is indicated on the right side. The scales represent Log_2 transformed values.

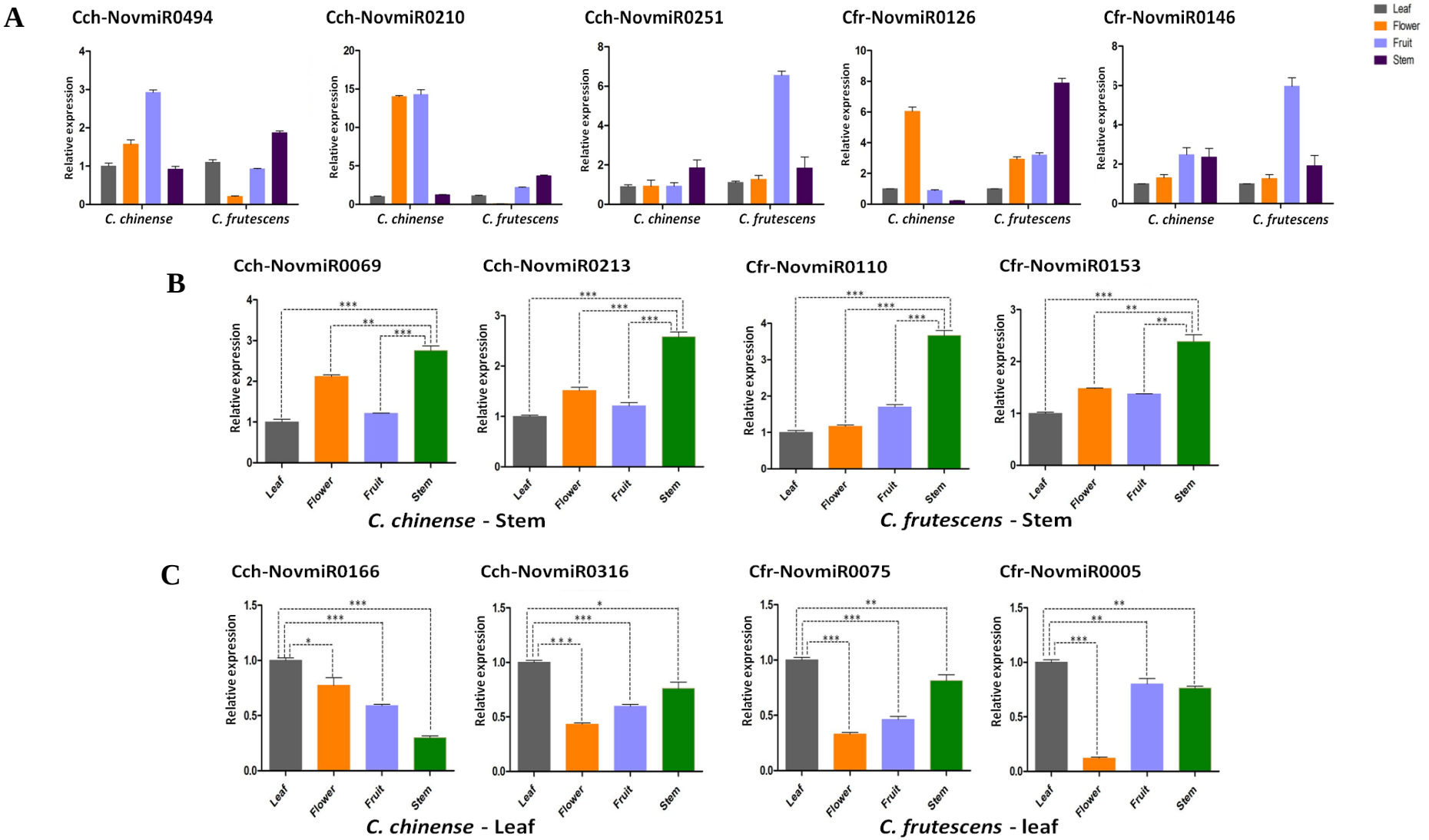


Figure 6.18. Continued....

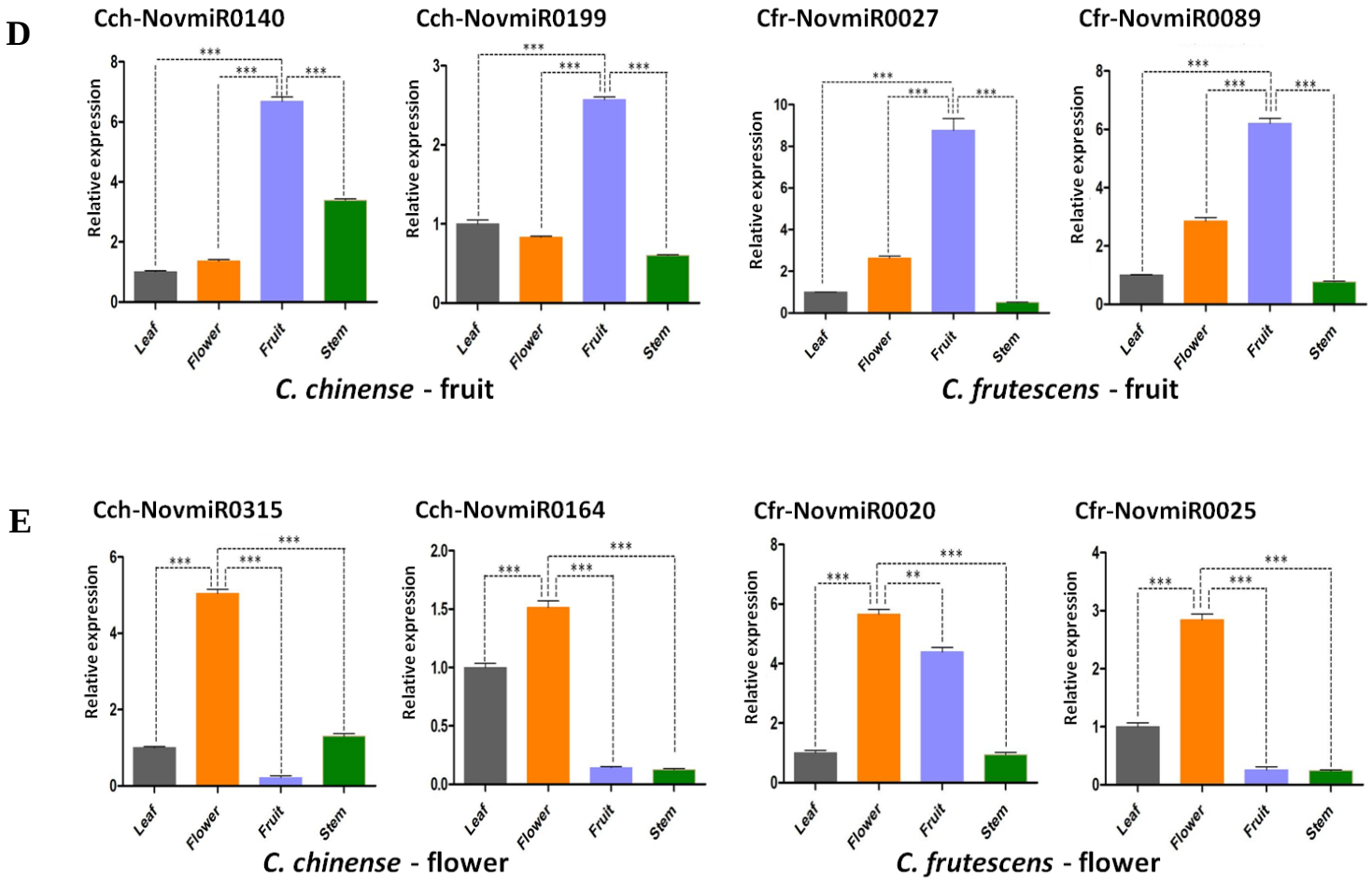


Figure 6.18 qRT-PCR analysis of differentially expressed and tissue-specific miRNAs in different tissues of *C. chinense* and *C. frutescens* (A) The histograms demonstrate relative expression levels of miRNAs in leaf, flower, fruit and stem tissues of both *Capsicum* species. The tissue specificity of miRNAs is shown in (B) stem, (C) leaf, (D) fruit, and (E) flower tissue. The expression level of each miRNA was normalized with U6 snRNA expression. The error bars signifies standard deviation among biological replicates of tissue samples. **P*<0.05, ***P*<0.01, ****P*<0.001.

Similarly, Cch-NovmiR0315 and Cch-NovmiR0164 (in *C. chinense*), and Cfr-NovmiR0020 and Cfr-NovmiR0025 (in *C. frutescens*) were significantly enriched in flower tissue in contrast to other tissue types (Fig. 6.18C). This implies a probable role of these miRNAs in flower development. In addition qRT-PCR analysis revealed tissue preferential expression of miRNAs in stem and leaf tissues of *Capsicum* species (Fig. 6.18D and 6.18E). Overall, we observed considerable consonance in the expression patterns of miRNAs derived through small RNA-sequencing and qRT-PCR. Several of these miRNAs displayed major differences in the expression levels throughout various tissues (Fig. 6.18 to 6.18E). In general, the differential and tissue-specific expression profile observed here will be useful to explore the specific function/biological roles of miRNAs in *Capsicum* development.

3.9 Identification of miRNA associated with fruit development

Tissue specific expression analysis revealed that few miRNAs are abundantly expressed (according to normalized sequencing reads and qRT-PCR analysis) in fruits compared to other reproductive and vegetative tissues. In particular, Cch-NovmiR0140 and Cch-NovmiR0199 of *C. chinense*, and Cfr-NovmiR0027 and Cfr-NovmiR0089 of *C. frutescens* significantly expressed in respective fruit tissuescompared to other (Fig. 6.18D). Further, the target prediction of these fruit preferential miRNAs revealed that Cfr-NovmiR0089 targets YABBY5 protein which regulates fruit size and shape in Solanaceae family especially in tomato (Cong et al., 2008; Huang et al., 2013; Han et al., 2015). To confirm the fruit preferentiality and species specificity of Cfr-NovmiR0089 in *C. frutescens* species, qRT-PCR analysis was done in few contrasting genotypes *i.e.* four genotypes of *C. frutescens*, andtwo genotypes each from *C. chinense* and *C. annuum.* We found significantly higher expression of Cfr-NovmiR0089 in fruit tissues of *C. frutescens* genotypes compared to genotypes of *C. chinense* and *C. annuum* thataffirms the species-specificity of miRNA (Fig. 6.19A).To investigate the possible regulatory roles of the miRNA in *Capsicum* fruit development, we have done a qRT-PCR analysis of the target genes i.e YABBY5a and 5b protein. We observed that relative expression of these two target genes is considerably high in large fruited *C. chinense* and *C. annuum* while very low level of expression in small-fruited *C. frutescens* genotypes was observed (Fig. 6.19B). However, the *YAB5b* gene has found to be highly expressed in vegetative tissues but in reproductive tissues negligible expression was detected. Overall, the qRT-PCR analysis illustrates the negative correlation of target gene expression with their analogous Cfr-NovmiR0089 miRNA in *C.*

frutescens, suggesting that this miRNA might have significant role in *Capsicum* fruit development.

3.10 Identification of Solanaceae-specific miRNAs

The advances in next generation small RNA sequencing technology enabled fast and accurate identification of miRNAs along with understanding their evolution and function. Several miRNA families are lineage-specific or family-specific or tissue-specific in nature (Allen etal., 2004). The proportion of miRNAs conserved between plant families are less than species or family-specific miRNAs indicating many of the known miRNA genes emerged in recent evolutionary period (Cuperus et al., 2011). These highly conserved miRNA families are the intrinsic component of various functional regulatory pathways of plant development, nutrition, stress and signaling response (Bartel et al., 2003, Rubio-Somoza et al., 2011). In this study, prediction of putative Solanaceae-specific miRNAs to explore their role in plant development was done. To achieve this, all plant miRNA (Viridiplantae) sequences from miRBase was retrieved. The tomato and potato miRNAs were further extracted from this miRNA list. Later, the BLASTN was performed with tomato, and potato miRNA sequences as a query against total Viridiplantae plant miRNA sequences, and 104 miRNA sequences (from each potato and tomato) did not show any hit and were inferred as tomato and potato specific miRNAs. Further, alignment with BLASTN by taking these sequences as a query sequences was done against in-house developed *Capsicum* miRNA sequences which resulted in 51 unique miRNA sequences. The *Capsicum* miRNA database were developed by collecting all the miRNAs from the previous studies (Hwang et al., 2013 and Liu et al., 2017) and our small RNA dataset of *C. chinense* and *C. frutescens.*

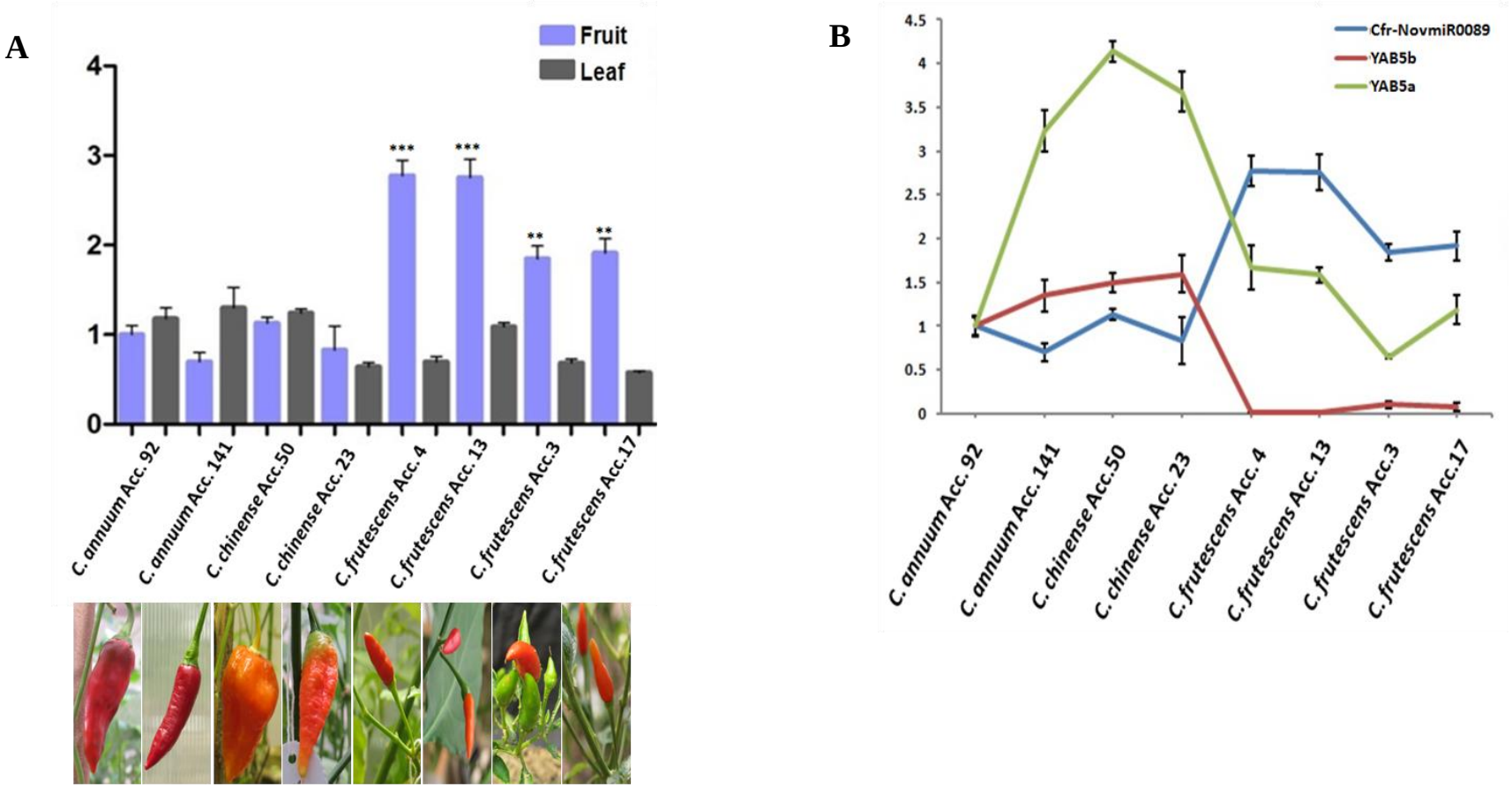


Fig. 19. Expression analysis for fruit specific miRNA and target gene in *Capsicum* species. (A). qRT-PCR analysis for confirmation of fruit specific expression of Cfr-NovmiR0089. The leaf and fruit tissue of two genotypes of *C. annuum* and *C. chinense* with four genotypes of *C. frutescens* were selected to confirm the fruit and species specific expression of Cfr-NovmiR0089. (B) Expression pattern of miRNA and target mRNA (*YAB5a*&*YAB5b*) from fruit tissues of *Capsicum* species shown by qRT-PCR analysis. At least three plants from each genotype were included and three technical replicates were used in analysis.$P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$.

Our analysis revealed that several miRNAs families such as miR166, miR167, miR171, miR172, miR408 and miR482 were Solanaceae-specific in nature i.e.:-, sly-miR166c, sly-miR167b, sly-miR171c,sly-miR482a, stu-miR167b, stu-miR167c, stu-miR172c, etc. These potential Solanaceae-specific miRNAs targets a diverse range of proteins such as protein kinase, ascorbate oxidase, GDP-mannose epimerase, calcium binding protein, LRR receptor-like serine/threonine-protein kinase, late blight resistance proteins which plays important role in plant development and defense signaling pathways. Several transcription factors (NAC, MYB, WRKY, bHLH) were revealed as targets of the miRNAs suggesting the pivotal role of conserved miRNAs in fundamental biological processes.

4. Discussion

As microRNAs are increasingly being studied and their role in regulation of diverse metabolic and developmental pathways including economically important traits in crop plants are established, the present study was designed for comprehensive profiling of small RNAs in Bhut jolokia (*C. chinense*) and *C. frutecens* since no such study is reported in these two *Capsicum* species. Furthermore, miRNA families identified in the most commonly studied *C. annuum* are also currentlyinadequate. Our aim was to identify and develop small non coding miRNA resources in *Capsicum* species which could be used in future to understand the regulation of genes governing economically important traits and their possible manipulations in *Capsicum* breeding programme. The raw reads after filtering with various criteria generated high quality small RNA tags. The sRNA reads length in individual library was selected in between 18 to 30nt and among these range, the major fractions of reads were of 21 to 24nt which indicating the characteristics of DCLs processed sRNAs (Axtell, 2013). The observation of major proportions of 24nt miRNA in our study is in agreement with earlier reports in other plant species like *Arabidopsis thaliana*, tomato, *Citrus trifoliata*, cucumber, peanut, maize, *C. annuum*, rice, potato and chickpea (Fahlgren et al., 2007, Moxon et al., 2008,Song et al., 2010, Martinez et al., 2011, Chi et al., 2011, Li et al., 2013, Hwang et al., 2013, Yang et al., 2013, Lakhotia et al., 2014, Shrivastava et al., 2015). Contrary to this observation, however, several previous studies reported findings of more fractions of 21nt or 23nt class miRNAs in wheat, chinese yew and grapevine (Yao et al., 2007; Qiu et al., 2009; Pantaleo et al., 2010). The MFE is one of the sternly followed criteria in the discovery of miRNAs (Reinhart et al., 2002; Zhang et al., 2006). A lower free energy indicates higher efficiency of miRNAs (Bonnet et al., 2004). Usually, miRNAs with MFE below –18 kcal mol^{-1} were indicative of canonical miRNAs (Hwang et al., 2013, Jain et al., 2014, Lakhotia et al., 2014), the similar parameter was followed to identify miRNAs in the present investigation.

The observation of predominancy of 24 nucleotide miRNA class in both *C. chinense* and *C. frutescens* was also reported in recent study of miRNAs in *C. annuum*(Liu et al., 2017). The richness of 24nt sRNAs might signify the intricacy of the *Capsicum* genome as mainly long-miRNAs are associated with repeats and heterochromatic modifications (Axtell et al., 2013). In plants usually, 21nt sRNA class are predominantly found when processed by DCL1 and AGO1 association, while the 24nt miRNAs are generated by DCL3 protein in association with AGO4 (Rajagopalan et al., 2006, Axtell, 2013). Among the analyzed tissues,

the representation of 24nt miRNAs was observed to be high in reproductive tissues (fruits and flower) relative to vegetative tissues (leaf and stem). This might be due to the fact thatthat the reproductive tissues requires more strict and distinct repression machinery of target genes by these miRNAs (Jeong et al., 2011). The similar findings was observed in *Arabidopsis* and rice in which higher proportion of 24nt sRNAs was observed in reproductive tissues like inflorescences and rice panicles, respectively as compared to vegetative tissues such as leaves and roots (Kasschau et al., 2007; Jeong et al., 2011). The larger fraction of *Capsicum* miRNAs showed a typical uracil/adenine (U/A) at 5'base of miRNA structure as consistent with previous studies in *Capsicum* and other plants (Mi et al., 2008; Czech et al., 2011; Hwang et al., 2013; Liu et al., 2017).

The GC content of miRNA plays a vital role in the prediction of the putative targets (Adai et al., 2005). Some studies suggested that low GC content (20%) miRNAs usually have few numbers of predicted binding sites and, high GC content (70%) miRNAs have several target sites (Ho et al., 2007). The GC content of majority of the *Capsicum* miRNAs in the study varies in between 35-70% which is similar to the observation made in other plants (Ho et al., 2007; Mishra et al., 2009; Jain et al., 2014). In *C. chinense* and *C. frutescens*, the average GC composition of miRNAs was found to be (44%) and (46%), respectively, and was identical with *Arabidopsis*,chickpea, *Medicago* (all have 44%) and rice, soybean (46%) but lower than grapevine (50%), sorghum and maize (both 52%) (Adai et al., 2005, Ho et al., 2007, Lelandais-Briere et al., 2009; Turner et al., 2012, Jain et al., 2014).

Each family comprised of variable number of miRNA members ranging from two to twenty nine. In *Arabidopsis* such a large expansion of miRNA families were observed due to several rounds of genome duplication events i.e. families like miR156, miR159, miR166, miR395 (Maher et al., 2006; Li and Mao, 2007). In *Capsicum* also, owing to a large genome size (3.6 GB), several duplication regions of the genome including the expansion of coding and non coding gene (miRNAs) families are expected. This might be one of the reasons of finding several members with varying length of miRNAs form one family. However, due to recent evolutionary emanation and stringent similarity criteria, most of the identified novel miRNAs did not clustered with conserved families, although a few novel miRNAs were clustered with conserved miRNA families. The length of miRNA family and nucleotides at 3'-end varies within and between families. We found representation of 21 to 24nt variants of novel miRNAs in the conserved miRNA families.

The mapping of miRNAs identified in the present study in the *C. annuum* reference genome, showed extensive genome wide distribution and clustering of conserved and novel miRNAs throughout the twelve linkage groups of *Capsicum* genome. It was observed that some miRNA family clusters are strictly present on the same chromosome of *C. chinense* and *C. frutescens*, while, some clusters of the miRNA family were located on either similar or different chromosomes of two species. These findings were consistent with chickpea in which four clusters of miR166 family detected on multiple linkage groups (Jain et al., 2014), and with potato in which the tandem arrays of miRNAs was observed on chromosome 1, 3, 6 and 8 of the genome (Lakhotia et al., 2014). The similar miRNA clustering pattern was detected in other studies like *Arabidopsis*, rice, *Medicago*, potato and chickpea (Jones-Rhoades and Bartel, 2004; Cui et al., 2009; Lelandais-Briere et al., 2009; Lakhotia et al., 2014; Jain et al., 2014). The miRNA distribution analysis suggested that major proportion of the miRNAs in *C. chinense* (81%) and *C. frutescens* (84%) are derived from non-coding regions of the genome and is consistent with other reports (Kim et al. 2005, Jain et al., 2014; Lakhotia et al., 2014).

The target prediction was performed to assign the possible functions to identified miRNAs in multifarious biological and cellular processes. We identified targets for all the *C. chinense* and *C. frutescens* miRNAs. However, in other sRNA studies in only about 85% to 90 % miRNAs target could be predicted (Hwang et al., 2013; Jain et al., 2014; Lakhotia et al., 2014). Furthermore, our finding is in accordance with the previous studies in plants that indicates the principal mechanism of miRNA-mediated gene regulation is mRNA cleavage (Schwab et al., 2005; Pasquinelli, 2012; Rogers and Chen, 2013; Jain et al., 2014). It has been found in previous sRNA studies that transcription factors (TFs) is one of the principal targets of miRNAs (Chen, 2009; Jeong et al., 2011; Hwang et al., 2013; Lakhotia et al., 2014; Jain et al., 2014). In early investigations, it was seen that the conserved miRNAs targeting transcription factor families were involved in various processes of plant development (Bartel, 2004; Mallory and Vaucheret, 2006). In our study, we observed novel miRNA targets multiple genes which encodes for a diverse series of proteins, such as transcription factors like MYB, NAC, AP2-EREBP, ARF, SBP, HB, GRAS, protein like PPR proteins, kinases, F and U-box, defense and signaling-related proteins. Overall, putative targets of novel *Capsicum* miRNAs are more distinct compared to conserved miRNAs, which is in consistent with previous studies (Jeong et al., 2011; Hwang et al., 2013). Majority of the plant transcription factors like SBP, AP2, ARF, GRAS, NAC, YA, MYB, etc. targeted by

conserved miRNAs were found to be analogous with the conserved miRNA targets identified in *Arabidopsi*s and several other plant species (Yanhui et al., 2006, Zhang et al., 2006, Moxon 2008, Hwang et al., 2013, Jain et al., 2014, Varshney et al., 2015), indicating the vital role of conserved miRNAs in fundamental biological and cellular processes.

Furthermore, our present study revealed that, members from the same miRNA family does not always display identical expression profile. The finding further confirmed by by investigating the expression patterns of the miRNA families with at least four miRNA members. In *Arabidopsis*, rice and chickpea, the differential expression of members of miRNA families such as miR156, 159, 164, 166, 169, 319, 171, and 172 have been observed (Jeong et al., 2011, Jain et al., 2014). As demonstrated in earlier studies (Sieber et al., 2007; Palatnik et al., 2007; Jeong et al., 2011) different isoforms of miRNAs are found to target diverse set of genes that control spatio-temporal gene regulation. The tissue specificity analysis implies that the proportions of tissue-specific/developmental stage-specific expression of miRNAs are significantly higher compared to other types. This observation is similar to the observations reported earlier in *C. annuum* (Hwang et al., 2013) and chickpea (Jain et al., 2014).

For the validation/confirmation of digital expression profile of *Caspicum* miRNAs observed from small RNA sequencing, we carried out qRT- PCR analysis following stem loop method (Vakonyi-Gasic et al., 2007). The expression analysis using qRT-PCR showed overall, similarity in the expression patterns of miRNAs derived through small RNA-sequencing and qRT-PCR. Among many miRNAs, the fruit specific expression of Cfr-NovmiR0089 was validated using qRT-PCR with diverse genotypes of *C. frutescens*, *C. chinense* and *C. annuum*. The result confirms the significantly higher expression of Cfr-NovmiR0089 in small fruited *C. frutescens*. The expression analysis of target gene (*YABBY5a* and *YABBY5b*) shows that their relative expression is considerably high in large fruited *C. chinense* and *C. annuum* while, very low level of expression was observed in small-fruited *C. frutescens* genotypes. The *YABBY* genes are reported to play a vital role in establishing abaxial cell fate in lateral organs such as fl oral parts, ovules, leaves and lamina expansion and maintenance of meristems (Golz and Hudson 1999; Bowman 2000). In rice spikelet, *OsYABBY*5 (TONGARI-BOUSHI1) is found to regulate lateral organ/tissue development and control the meristem organization (Tanaka et al. 2012). The expression analysis of *YABBY* gene family in tomato discovered that *YAB5a* gene was almost gradually

expressed in large size tomatoes, however in small size tomatoes; it was expressed at an early fruit development (5DAA) and but not seen in later stages (young, breaker and matured fruit; Han et al., 2015). Similarly, in our study the *YAB5b* gene was found to be highly expressed in vegetative tissues but in reproductive tissues negligible expression was detected. Our results in *Capsicum* sp. has been conceding with these previous studies in tomato and other plants in which *YAB5a* and *YAB5b* have relatively higher expression in vegetative parts and large-fruited species compared to small-fruited *C. frutescens*. Furthermore, the qRT-PCR analysis demonstrates the negative correlation of target gene expression with their analogous Cfr-NovmiR0089 miRNA in *C. frutescens*, thereby suggesting their crucial role in determining fruit size/shape. Overall, the validation via qRT-PCR indicates that this miRNA might have significant contribution in *Capsicum* fruit development especially in fruit size and shape.

The prediction of putative Solanaceae-specific miRNAs provided insights into the highly conserved miRNA families that are intrinsic component of various functional regulatory pathways of plant development, nutrition, stress and signaling response (Bartel et al., 2003, Rubio-Somoza et al., 2011). Several miRNA families are lineage-specific or family-specific or tissue-specific in nature (Allen et al., 2004). The proportion of miRNAs conserved between plant families are less than that of species or family-specific miRNAs indicating that many of the known miRNA genes emerged in recent evolutionary period (Cuperus et al., 2011). Synteny analysis illustrates that majority of high confidence *Capsicum* miRNA families originated from identical loci of the chromosomes of *Capsicum*, potato and tomato. This indicates higher level conservation (miRNA genes) between *C. chinense*, *C. frutescens*, potato and tomato genome. Our approach for the prediction of Solanaceae specific miRNAs resulted into identification of 51 putative unique miRNAs which are specific to Solanaceae species comprising *Capsicum*, potato and tomato. Similarly, in wheat small RNA analysis, two monocot-specific miRNAs such as tae-miR3075 and tae-miR3014b were found to be conserved in all the studied monocots (Sun et al., 2014). In legume plants, the sequence conservation and secondary structure analysis of small RNA revealed four novel small RNAs named miR1507, miR2118, miR2119 and miR2199 which were identified as legume-specific miRNAs (Jagadeeswaran et al., 2009).The expression analysis confirmed that two of the miRNAs showed expression in all of the tested legume plants such as *Medicago*, chickpea, pea, soybean, peanuts but not in *Arabidopsis* and rice. These predicted miRNAs belongs to

various different miRNA families and targets a diverse range of proteins such as protein kinase, ascorbate oxidase, GDP-mannose epimerase, calcium binding protein, LRR receptor-like serine/threonine-protein kinase, late blight resistance protein as well as multifarious TFs playing essential roles in plant development and defense signaling pathways. Similarly, the TIR-NBS-LRR genes are predicted as targets of legume-specific miRNAs that are responsible for miRNA-directed plant defense response particularly in *Medicago* and other legume plants (Jagadeeswaran et al., 2009).

SUMMARY

Summary of the Thesis

The chilli peppers are found to be important in agriculture, medicinal and nutritional biology due to its versatility. Over the years, *Capsicum* genetics and genomics have been studied for various agronomically important traits. However, these studies are focused only in *C. annuum* and other equally important species like *C. chinense* (Bhut jolokia/Ghost chilli) and *C. frutescens* remain ignored globally. In this regard, for the first time in Bhut jolokia along with *C. frutescens* genomes were explored for :-1) Comparative study of expression of genes involved in capsaicinoid biosynthesis pathway in different *Capsicum* species, 2) Development of genome-wide molecular markers in *C. chinense* and *C. frutescens* based on transcriptome data, 3) Physiological and transcriptomic study of heterotic intra-specific and inter-specific *Capsicum* hybrids, 4) MicroRNA identification, target prediction and expression analysis. The results obtained in the current study are summarized below.

1. Comparative study of genes involved in capsaicinoid biosynthesis pathway in different *Capsicum* species

a) The comparative analysis of pungency was investigated in 20 genotypes belonging to three different *Capsicum* species i.e. *C. chinense* (Bhut jolokia), *C. frutescens* and *C. annuum* using GC-MS. The analysis revealed that genotypes belonging to *C. chinense* genotypes are of extremely high pungent following the genotypes from *C. frutescens* with moderate pungent while *C. annuum* genotypes showed very low to mild pungency.

b) Furthermore, to see whether varied pungency phenotypes observed in different species are the result of differential expression of already reported pungency biosynthesis genes in *C. annuum* (Kim et al., 2014), the comparative expression profiling was performed. The qRT-PCR analysis with the ten genes i.e. *PAL, BCAT, C4H, ACL, KAS, COMT, ACS, FAT, pAMT* and *AT3* (*Pun1*) in two genotypes each from *C. chinense, C. frutescens* and *C. annuum* revealed manifold higher (few cases more than 1000 fold) expression of the majority of the biosynthesis genes in *C. chinense* compared to *C. frutescens* and *C. annuum* genotypes at early, breaker and matured stages of fruit development, thereby correlating with the pungency phenotypes.

c) Among the ten analyzed pungency biosynthesis genes, the expression of *Pun1*, *pAMT, KAS* and *BCAT* are found to be significantly high in *C. chinense* (Bhut jolokia) relative to

C. frutescens and *C. annuum* accessions indicating their positive correlation with pungency content. The higher expression of these four genes in Bhut jolokia (*C. chinense*) is the reason of extremely pungent (fiery hot) phenotypes compared to that of *C. frutescens* and *C. annuum*.Most of the structural genes showed higher expression at early (20 DPA) fruit development stage and further decreases as fruit turns to maturity suggesting that during the fruit development genes that are expressed in early stage convert to proteins in later stage and accumulates to give increased pungent property in matured *Capsicum* fruits.

d) The considerable low expression of *Pun1, pAMT, KAS and BCAT* genes in *C. frutescens* compared to *C. chinense* accessions correlates with their moderate capsaicinoid (pungency) content while, the expression of these genes was seen almost negligible in low capsaicinoid content *C. annuum* accessions. These observations suggest that the variation in pungency phenotype in different *Capsicum* species is the result of variation in the level of expression of pungency biosynthesis genes, although many unknown genes influencing pungency phenotypes cannot be ruled out.

e) Furthermore, to develop molecular markers, we sequenced and aligned *KAS* gene, one of the genes involved in pungency biosynthesis pathway, and observed 12 bp deletion in the fourth intron region of *KAS* gene from *C. frutescens* and such Indel was absent in *C. chinense* and *C. annuum*. Based on this sequence variation observed between different *Caspicum* species, we further design the *KAS*1 marker flanking the InDel region. The PCR analysis showed that *KAS*1 marker could be used in chilli pepper breeding programme to precisely differentiate the *KAS* allele between *C. frutescens* and high pungent *C. chinense*/low pungent *C. annuum*.*KAS*1 marker is found to associate with moderate pungency. Development of pungency biosynthesis gene(s) based markers is being done in the lab.

2. Development of genome-wide molecular markers in *C. chinense* and *C. frutescens* based on transcriptome data

a) The evaluation of 96 *Capsicum* genotypes (collected mainly from North East India) belonging to *C. chinense*, *C. annuum* and *C. frutescens* showed diverse morpho-agronomic characteristics such as fruit size, shape and color, fruit weight and length, and pungency content .

b) Since, no report of molecular marker development in Bhut jolokia/Ghost Chilli and very limited study in *C. frutescens* are available, we for the first time developed a total of 4,988 genic SSRs markers in Bhut jolokia and 4,781 genic-SSR markers in *C. frutescens* using a total of 184,975 *C. chinense* and 179,780 *C. frutescens* transcriptome sequences available in the lab. The characterization, frequency distribution and annotation were done for all these SSRs. We observed trinucleotide repeat is abundantly present followed by di-, tetra- and penta- in both the species. The SSR repeats found in the intergenic regions was more abundant in all types of repeats.

c) Among them 50 SSR markers were experimentally validated for their amplification in 10 genotypes i.e. 3 genotypes each form *C. chinense* (Bhut jolokia, *C. frutescens*, *C. annuum* and one wild species. We found that all of the SSR primer pairs showed amplification and of these, 20 SSRs showed robust amplification in all the species with few clear bands were further used for genotyping of 96 *Capsicum* germplasm belonging to different *Capsicum* species for diversity study. These 20 markers showed high polymorphisms in terms of the heterozygosity, numbers of alleles, gene diversity, allele frequency and PIC content, and genetic diversity analysis enabled clustering of *Capsicum* germplasm in species-specific manner. The analysis showed highly pungent *C. chinense*, moderately pungent *C. frutescens* accessions and low pungent *C. annuum* accessions clustered independently in separate groups.

d) Additionally, we have identified a total of 337 polymorphic SSRs between *C. chinense* and *C. frutescens* using *in-silico* approach with stringent criteria. We sequenced and validated CFpSSR3 polymorphic marker in using genotypes of *C. chinense*, *C. frutescens* and *C. annuum*.Further, these markers are being used for genetic map construction in our lab.

3. Physiological and transcriptomic study of heterotic intra-specific and inter-specific *Capsicum* hybrids

Heterosis breeding is the most preferred breeding method among breeders to increase the crop production in a short span of time. However, the identification of perfect combination of parental genomes giving heterotic/hybrid vigour in F_1 is required. In the current study we have explored the possibility of production of F_1 heterotic hybrids by using the diverse *Capsicum* germplasm from the North East India including Bhut

jolokia/Ghost chilli. We developed 5 inter-specific hybrids :- i) *C. annuum* (Dudu) x *C. chinense* (Lota bhut), ii) *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4), iii) *C. chinense* (Chocolate jolokia) x *C. frutescens* (Acc. 4), iv) *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4), v) *C. chinense* (Umorok) x *C. frutescens* (Acc. 4),and one intra-specific hybrid vi) *C. chinense* (Lota bhut) x *C. chinense* (Chocolate bhut). Study of various quantitative traits starting from seed germination to number of fruit per plants was done to see if heterotic phenomenon is manifested from beginning of plant development until maturity. Our analysis observed significant positive heterosis over mid parent for plant height, root length, leaf area, number of fruits per plant, germination time, chlorophyll content, relative water content, photosynthetic and transpiration rate (using Infra Red Gas Analyser).

a) Global metabolite analysis using GC-MS revealed abundance of sugar, amino acids, fatty acids, carboxylic acids and sugar derivates in F_1 hybrids of inter-specific cross relative to both parents indicating the observed heterosis may be due to these metabolites. While in intra-specific F_1 hybrid, the reduction in sugar metabolites was observed compared to their parental lines correlating with lower degree of heterosis found in F_1 hybrid. Furthermore, expression of few metabolites (sugars and fatty acids) are observed in F_1 hybrids but not in their respective parents suggesting that in addition to above mentioned primary metabolites, these metabolites may also influence in manifestation of heterosis phenomenon in F_1 hybrids.

   a) The differential gene expression analyses reveals that non-additive expression contributed about 60 to 90% gene action in all the inter-specific hybrids while 72% in intra-specific hybrid suggesting that non-additive gene expression major contributor of heterosis as compared to additive gene expression.Expression level dominance analysis showed that *C. frutescens* genome is most dominant followed by *C. chinense* and *C. annuum* in various F_1 hybrids.

   b) The DGE analysis revealed that majority of TFs exhibits non-additive expression pattern relative to additive. The expression analysis of predominant WRKY transcription factor family revealed considerably higher expression of its members in F_1 hybrids than parents suggesting their involvement in hybrid vigour of F_1.

c) Digital gene expression and qRT-PCR analyses demonstrated that, genes involved in photosynthesis and carbohydrate metabolism pathway showed considerably higher level

of expression in F_1 hybrid as compared to parents, which confirms the result obtained in physiological and global metabolite analysis. The finding indicates that, these genesare playing crucial role in determining heterotic behaviour of F_1 hybrids. For the first time in our study, the molecular mechanism of hybrid vigour in *Capsicum* species is understood.

d) The qRT-PCR analysis of randomly selected non-additive and additive genes from intra-and inter-specific hybrids showed substantial consonance in the expression patterns analysed through transcriptome sequencing.

4. MicroRNA identification, target prediction and expression analysis

a) Since microRNAs are being largely studied for their role in gene regulation of plant developmental processes including genes governing economically important traits, we did genome wide profiling for microRNAs in two highly contrasting accessions belonging to *C. chinense* (Bhut jolokia) and *C. frutescens* to identify conserved and novel miRNAs from stem, leaf, flower and fruit tissues. Our study revealed the identification of 531 conserved and 521 high-confidence novel miRNAs from 3,025,801 unique reads in *C. chinense*, and a total of 432 conserved and 159 novel miRNAs from 6,614,813 unique reads in *C. frutescens*. This is the first comprehensive small RNA study in both the *Capsicum* species.

b) Size distribution analysis showed that 21 to 24nt miRNAs are abundantly found in both species which is a characteristic of DCLs processed miRNAs, among them 24nt miRNAs are predominantly present which signifies the complexity of the *Capsicum* genome as mainly long-miRNAs are associated with repeats and heterochromatic modifications.

c) The mapping of miRNA in *C. annuum* reference genome showed extensive distribution and clustering of conserved and novel miRNAs throughout the twelve chromosomes of *Capsicum* genome.We observed that some miRNA family clusters are strictly present on same chromosomes of *C. chinense* and *C. frutescens*, whereas, some clusters of the miRNA family were located on either similar or different chromosomes of in the two species.

d) The genomic distribution analysis demonstrated that about 82% miRNAs are derived from inter-genic portion of genome.

e) The target prediction analysis reveals that approximately 80% of *Capsicum* miRNA targets were predicted to be governed by cleavage mechanism and the remaining through translational inhibition.

f) We have observed, *Capsicum* miRNAs may target diverse mRNA/genes coding for several proteins, like those associated with inorganic ion, metal, and amino acid transport, response to diverse stress, RNA processing and modification, post-translational modification, signal transduction pathways, carbohydrate and secondary metabolite biosynthesis indicating crucial role of miRNAs in fundamental biological and cellular processes.

g) The digital expression analysis discovered the differential and tissue-specific expression of miRNAs, indicating their involvement in tissue/organ development. Among them our findings of potential regulation of fruit size/shape via novel miRNA Cfr-NovmiR0089 provides insights into miRNA-mediated fruit development. We confirmed the involvement of Cfr-NovmiR0089 in determining fruit size/shape by digital expression and qRT-PCR analysis of both miRNA and its target (WABBY) gene in *Capsicum* germplasm belonging to *C. chinense, C. frutescens* and *C. annuum* with contrasting fruit sizes.

h) Overall, our combinatorial efforts of small RNA profiling and experimental validation offers high-confidence small RNA dataset that contributes to strenthen miRNA population in pepper and Solanaceae which can be used in future study to understand their regulation of gene(s) expression involved in diverse economically important traits.

Conclusion

&

Future Prospects

Conclusion of the thesis and future perspectives

Bhut jolokia or Ghost chilli of North East India, being the naturally occurring hottest chilli in the world is not only in high demand commercially but also for studying the pungency and other economically important traits. However, until recently, no systematic study was done to understand the level of genetic diversity for pungency and other economically important traits so that research findings could be used in applied breeding programme. In the present study, we tried to explore their applicability at phenotypic and genetic level with the final aim of developing genetics and genomics resources for future translational research. With the above aim, we have dissected the molecular mechanism of extremely high pungent Bhut jolokia; developed genomewide SSR markers; could dissect physiological, transcriptomic and metabolites based regulation of heterotic phenotype of inter and intraspecific F_1*Capsicum* hybrids; and develop genome wide microRNAs, of which one miRNA was found regulating target gene governing fruit size/shape.

The findings of up-regulation of four structural genes (*BCAT, KAS, pAMT* and *Pun1*) involved in pungency biosynthesis directly correlated with the elevated pungency in Bhut jolokia. These genes could be used in translational research to enhance the capsaicinoid content of chilli peppers. The development of a large number of gene based SSRs in *C. chinense* and *C. frutescens* opened up avenue for development of linkage map, mapping QTL/Gene governing economically important traits, identification of genes/linked markers and use in precision breeding of pungency and other economically important traits in *C. chinense,* and *C. frutescens*. The observation of diverse *Capsicum* germplasm gives scope to effectively use in future *Capsicum* breeding programme. Furthermore, to our knowledge this is the first work that systematically studied heterosis at molecular, physiological and metabolite level in *Capsicum* among Solanaceae crop plants. Our study revealed that significant positive correlated genes and metabolites expressions are responsible for heterotic F_1 phenotype in *Capsicum* hybrids compared to parents. Furthermore, it was observed that the more genetically distinct between parents, the more heterosis, suggesting genetically distinct parents with desirable traits could be used for development of *Capsicum* heterotic hybrid. Of the many genes and metabolites, we observed the major influence of up-regulation of photosynthesis and carbohydrate metabolism associated genes in deciding the degree of heterosis in *Capsicum*. Genes governing these traits, a few of which are validated in the present study could be used in breeding programme for development of high vigour *Capsicum* varieties. As observed earlier, in *Capsicum* species, we confirmed and

strengthened the quantitative genetic hypothesis through whole transcriptome study in seedlings by observing that the non additive expression of genes i.e. dominance and over-dominance gene action plays larger role in heterosis compared to additive gene action. Furthermore, since gene regulation by small RNAs are increasingly becoming important, the genome-wide profiling of miRNAs in *C. chinense* and *C. frutescens* provides ample opportunities for future miRNAs mediated gene regulation study of economically important traits in *Capsicum* species, especially in Bhut jolokia and *C. frutescens*. The high-confidence population of miRNAs identified in the present study with wide functional roles in several fundamental biological processes could be important genomics resources for future study and to our knowledge this is the first comprehensive profiling of miRNAs in *Capsicum* species, especially in Bhut jolokia and *C. frutescens*. The higher expression of miRNA Cfr-NovmiR0089 in *C. frutescens* indicates their crucial involvement in determining fruit size in *Capsicum*. Overall, the huge genetic and genomic resources developed in the present study would be used in accelerating the Bhut jolokia and other *Capsicum* improvement programme.

References

References

Abraham-Juarez MD, Del Carmen Rocha-Granados M, Lopez MG, Rivera-Bustamante RF, Ochoa-Alejo N. (2008) Virus-induced silencing of *Comt*, *pAmt* and *Kas* genes results in a reduction of capsaicinoid accumulation in chili pepper fruits. Planta. 227: 681–695.

Adai A, Johnson C, Mlotshwa S, Archer-Evans S, Manocha V, et al. (2005) Computational prediction of miRNAs in *Arabidopsis thaliana*. Genome Res. 15: 78–91.

Adetula OA, (2006) Genetic diversity of *Capsicum* using random amplified polymorphic DNAs. *Afr J Biotech* 5:120–122

Ahn Y K, Tripathi S, Kim J H, Cho Y I, Lee H E, et al. (2014) Transcriptome analysis of *Capsicum annuum* varieties Mandarin and Blackcluster: Assembly, annotation and molecular marker discovery. *Gene*, *533*(2), 494-499.

Ahn YK, Tripathi S, Cho YI, Kim JH, Lee HE, et al. (2013). De novo transcriptome assembly and novel microsatellite marker information in *Capsicum annuum* varieties Saengryeg 211 and Saengryeg 213. Bot Stud.*54*(1), 58.

Aktas H, Abak K, Sensoy S, (2009) Genetic diversity in some Turkish pepper (*Capsicum annuum* L.) genotypes revealed by AFLP analyses. *Afr J Biotech* 8:4378–4386

Akter A, Islam MM, Mondal SI, Mahmud Z, Jewel NA, et al. (2014) Computational identification of miRNA and targets from expressed sequence tags of coffee (*Coffea arabica*). Saudi J Biol Sci. 21: 3–12.

Alghanim HJ and Almirall JR. (2003) Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. Anal. Bioanal. Chem., 376:1225-33

Allen E, Xie Z, Gustafson AM, Sung GH, Spatafora JW, et al. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet. 36: 1282–1290.

Aluru MR, Mazourek M, Landry LG, Curry J, Jahn MM. et al. (2003) Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in *Capsicum* fruit. J. Exp. Bot. 54, 1655–1664.

Alvarez-Saavedra E, Horvitz HR. (2010) Many families of *C. elegans* microRNAs are not essential for development or viability. Curr Biol. 20(4):367-73.

Ambros V. The functions of animal microRNAs. (2004) Nature. 431(7006):350.

amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44(3): 397.

Anders S, Huber W. (2010) Differential expression analysis for sequence count data. Genome Biol.11: R106.

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME. (1993) Optimizing parental selection for genetic linkage maps. Genome. 36: 181–186.

Andrea P, Quave C L, Santoro F R. (2004) Folk pharmaceutical knowledge in the territory of the Dolomiti Lucane, inland southern Italy. Journal of nanopharmacology 95.2: 373-384.

Antonious GF, Jarret RL. (2006) Screening Capsicum accessions for capsaicinoids content. J Environ Sci Health B. 41: 717–29.

Arnon DI. (1949) Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris*. Plant Physiol. 24: 1-15.

Ashby E. (1937) The physiology of heterosis. The American Naturalist. 71(736):514-20.

Ashrafi H, Hill T, Stoffel K, Kozik A, Yao J, et al. (2012). De novo assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for in silico discovery of SNPs, SSRs and candidate genes. *BMC genomics*, *13*(1), 571.

Axtell JM. (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64: 137–159.

AZa-Gonzalez C, Nunez-Palenius HG, Ochoa-Alejo N. (2011) Molecular biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). Plant Cell Rep. 30: 695–706.

Baar K, Kavasch EB, (1999) American Indian Healing Arts: Herbs, Rituals, and Remedies for Every Season of Life. Bantam.

Baba VY, Rocha KR, Gomes GP, de Fátima Ruas C, Rua PM, Genetic diversity of *Capsicum chinense* accessions based on fruit morphological characterization and AFLP markers. *Genet. Resour. Crop Evol.*1-11.

Bao N, Lye KW, Barton MK. (2004) MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev cell. 7(5):653-62.

Barbosa AM, Geraldi IO, Benchimol LL, Garcia AA, Souza CL, et al. (2003) Relationship of intra- and interpopulation tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. Euphytica. 130(1):87-99.

Barrera-Figueroa BE, Gao L, Diop NN, Wu Z, Ehlers JD, et al. (2011) Identifi cation and comparative analysis of drought associated microRNAs in two cowpea genotypes. BMC Plant Biol. 11: 127.

Bartel B, Bartel DP. (2003) MicroRNAs: at the root of plant development? Plant Physiol. 132: 709–717.

Bartel DP. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 116: 281–297.

Bartel DP. (2009) MicroRNAs: target recognition and regulatory functions. Cell. 136(2):215-33.

Barth S, Busimi AK, Utz HF, Melchinger AE. (2003) Heterosis for biomass yield and related traits in five hybrids of *Arabidopsis thaliana* L. Heynh. Heredity. 91(1):36.

Ben-Chaim A, Brodsky Y, Falise M, Mazourek M, Kang BC, et al. (2006) QTL analysis for capsaicinoid content in *Capsicum*. Theor Appl Genet. 113: 1481–1490.

Benjamini Y, Hochberg Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc. 57: 289–300.

Bennett DJ, Kirby GW. (1968) Constitution and biosynthesis of capsaicin. J Chem Soc C. 442–446.

Bhagowati RR, Chngkija S. (2009) Genetic variability and traditional practice in Naga king chili landrace of Nagaland. Asian Agrihist. 13: 171–180.

Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA. (2010) Heterosis. Plant Cell. 22: 2105–2112.

Birchler JA, Yao H, Chudalayandi S. (2006) Unraveling the genetic basis of hybrid vigor. P Natl Acad Sci USA. 103(35):12957-8.

Birchler JA. (2013) Genetic rules of heterosis in plants. John Wiley & Sons, Inc.

Blum E, Liu K, Mazourek M, Yoo EY, Jahn M, Paran I, (2002) Molecular mapping of the C locus for presence of pungency in *Capsicum*. Genome, 45(4), 702-705.

Blum E, Mazourek M, O'Connell M, Curry J, Thorup T, et al. (2003) Molecular mapping of capsaicinoid biosynthesis genes and QTL analysis for capsaicinoid content in Capsicum. Theor Appl Genet. 108: 79–86.

Bonnet E, Wuyts J, Rouze P, Van de Peer Y. (2004) Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics. 20: 2911–2917.

Bosland PW, Baral JB. (2007) 'Bhut Jolokia'- the world's hottest known chile pepper is a putative naturally occurring interspecifi c hybrid. HortScience. 42: 222–224.

Bosland PW, Coon D, Cooke PH. (2015) Novel formation of ectopic (nonplacental) capsaicinoid secreting vesicles on fruit walls explains the morphological mechanism for super-hot chile peppers. J Am Soc Hort Sci. 140: 253–256.

Bosland PW, Walker SJ, (2010) Measuring chili pepper heat, Guide H-237. Department of Plant and Environmental Sciences and Department of Extension Plant Sciences, New Mexico State University, URL: http:/laces. nmsu. edu/pubs/_h H, 237.

Bosland PW. (1996) *Capsicums*: Innovative uses of an ancient crop." Progress in new crops. ASHS Press, Arlington. VA: 479-487.

Botstein D, White R L, Skolnick M, Davis R W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American journal of human genetics. 32(3): 3-14.

Bowman JL. (2000) The YABBY gene family and abaxial cell fate. Curr Opin Plant Biol. 3: 17–22.

Breakfield NW, Corcoran DL, Petricka JJ, Shen J, Sae-Seaw J, et al. (2012) High-resolutionexperimental and computational profiling of tissue-specific known andnovel miRNAs in *Arabidopsis.*Genome Res. 22: 163–176.

Bruce AB. (1910) The Mendelian theory of heredity and the augmentation of vigor. Science. 32(827):627-8.

Buso G S C, Reis A M M, Amaral Z P S, Ferreira M E. (2016) Novel and highly informative *Capsicum* SSR markers and their cross-species transferability. *Genet Mol Biol: GMR*, *15*(3).

C't Hoen PA, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RH, et al. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. 36(21): e141.

Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, et al. (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. P Natl Acad Sci USA. 101(9):2999-3004.

Campo S, Peris-Peris C, Sire C, Moreno A B, Donaire L, Zytnicki M, et al. (2013). Identifi cation of a novel microRNA (miRNA) from rice that targets an alternatively spliced transcript of theNramp6 (Natural resistance-associated macrophage protein 6) gene involved in pathogen resistance. New Phytol. 199, 212–227.

Canto-Flick A, Balam-Uc E, Bello-Bello JJ, Lecona-Guzman C, Solı ´s-Marroquin D, et al. (2008) Capsaicinoids content in Habanero pepper (*Capsicum chinense* Jacq.): Hottest known cultivars. HortScience. 43: 1344–1349.

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

Carnavale-Bottino M, Rosario S, Grativol C, Thiebaut F, Rojas CA, et al. (2013) High-throughput sequencing of small RNA transcriptome reveals salt stress regulated microRNAs in sugarcane. PLoS ONE. 8: e59423.

Carrington JC, Ambros V. (2003) Role of microRNAs in plant and animal development. Science. 301: 336–338.

Carvalho SIC, Ragassi CF, Oliveira IB, Amaral ZPS, Reifschneider FJB, et al. (2015) Transferability of microsatellite markers of *Capsicum annuum* L. to *C. frutescens* L. and *C. chinense* Jacq. Genet Mol Res. 14(3): 7937-7946.

Ceresini P C, Silva C L S P, Missio R F, Souza E C, Fischer C N, et al. (2005). Satellyptus: Analysis and database of microsatellites from ESTs of Eucalyptus., *28*(3), 589-600.

Chabane K, Ablett GA, Cordeiro GM, Valkounn J, Henry RJ. (2005) EST versus genomic derived microsatellite markers for genotyping wild cultivated barley. Genet Resour Crop Evol. 52: 903-909.

Chaitieng B, Kaga A, Tomooka N, Isemura T, Kuroda Y, et al. (2006) Development of a black gram [*Vigna mungo* (L.) Hepper] linkage map and its comparison with an azuki bean [*Vigna angularis* (Willd.) Ohwi and Ohashi] linkage map. *Theor Appl Genet* 113:1261–1269

Charlesworth D, Willis JH. (2009) The genetics of inbreeding depression. Nat Rev Genet. 10: 783–796.

Chaves SS, Fernandes-Brum CN, Silva GFF, Ferrara-Barbosa BC, Paiva LV, et al. (2015) New Insights on Coffea miRNAs: Features and Evolutionary Conservation. Appl Biochem Biotechnol. 177: 879–908.

Chen C, Zhou P, Choi Y A, Huang S, Gmitter F G. (2006) Mining and characterizing microsatellites from citrus ESTs. *Theor Appl Genet*, *112*(7), 1248-1257.

Chen X. (2009) Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol. 25: 21–44.

Cheng J, Zhao Z, Li B, Qin C, Wu Z, et al. (2016). A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in Capsicum. *Sci. Rep.*, *6*, 18919.

Chi X, Yang Q, Chen X, Wang J, Pan L, et al. (2011) Identification and characterization of microRNAs from peanut (*Arachis hypogaea* L.) by high-throughput sequencing. PLoS One. 6: e27530.

Chinn MS, Sharma-Shivappa RR, Cotter JL. (2011) Solvent extraction and quantification of capsaicinoids from *Capsicum chinense*. Food Bioprod Process. 89: 340–345.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, et al. (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet. 100: 713-722.

Cloutier S, Niu Z, Datla R, Duguid S. (2009) Development and analysis of EST SSRs for flax (*Linum usitatissimum* L.). Theor Appl Genet. 119: 53-63.

Coe FG, Anderson GJ. (1997) Ethnobotany of the Miskitu of eastern Nicaragua. J Ethnobiol 17, 171–214.

Cong B, Barrero LS, Tanksley SD. (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. Nature genet. 40: 800-804.

Contreras-Padilla MM, Yahia E. (1998) Changes in capsaicinoids during development, maturation, and senescence of chile peppers and relation with peroxidase activity. J Agric Food Chem. 46: 2075–2079.

Coolbear P, Francis P, Grierson D. (1984) The effect of low temperature pre-sowing treatment under the germination performance and membrane integrity of artificially aged tomato seeds. J. Exp. Bot. 35: 1609-1617.

Cui X, Xu SM, Mu DS, Yang ZM. (2009) Genomic analysis of rice microRNA promoters and clusters. Gene.431: 61–66.

Cuperus JT, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke RT, et al. (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in *Arabidopsis*. Nat Struct Mol Biol. 17: 997–1003.

Cuperus JT, Fahlgren N, Carrington JC. (2011) Evolution and functional diversification of MIRNA genes. Plant Cell. 23: 431–442.

Curaba J, Spriggs A, Taylor J, Li Z, Helliwell C. (2012) miRNA regulation in the early development of barley seed. BMC Plant Biol. 12: 120.

Curry J, Aluru M, Mendoza M, Nevarez J, Melendez M, et al. (1999) Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp. Plant Sci. 148, 47–57.

Czech B, Hannon GJ. (2011) Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet. 12: 19–31.

Dai X, Zhao PX. (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 39: W155–159.

Darwin C. (1876) The effects of cross and self fertilisation in the vegetable kingdom. J. Murray

Davenport CB. (1908) Degeneration, albinism, and inbreeding. Science. 28(718):454-5.

Del Rosario Abraham-Juarez M, Del Carmen Rocha-Granados M, Lopez MG, Rivera-Bustamante RF, Ochoa-Alejo N. (2008) Virus-induced silencing of *Comt*, *pAmt* and *Kas* genes results in a reduction of capsaicinoid accumulation in chili pepper fruits. Planta. 227: 681–695.

Deng P, Wang L, Cui L, Feng K, Liu F, et al. (2015) Global identifi cation of microRNAS and their targets in barley under salinity stress. PLoS ONE. 10: e0137990.

Din M, Barozai MYK, Baloch IA. (2014) Identifi cation and functional analysis of new conserved microRNAs and their targets in potato (*Solanum tuberosum* L.). Turk J Bot. 38: 1199–1213.

Ding H, Gao J, Luo M, Peng H, Lin H, et al. (2013). Identifi cation and functional analysis of miRNAs in developing kernels of a viviparous mutant in maize kernel. Crop J. 1: 115–126.

Ding H, Qin C, Luo X, Li L, Chen Z, et al. (2014) Heterosis in early maize ear inflorescence development: a genome-wide transcription analysis for two maize inbred lines and their hybrid. Int J Mol Sci. 15(8): 13892-915.

Dobzhansky T. (1950) Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. Genetics. 35(3):288.

Du JF, Wu YJ, Fang XF, Xia CJ, Liang Z, et al. (2010) Prediction of sorghum miRNAs and their targets with computational methods. Chin Sci Bull. 55: 1263–1270.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. (2010). agriGO: a GO analysis toolkit for the agricultural community.Nucleic Acids Res, 38(suppl_2), W64-W70.

Dubey RK, Singh V, Upadhyay G, Pandey AK, Prakash D. (2015) Assessment of phytochemical composition and antioxidant potential in some indigenous chilli genotypes from North East India. Food Chem. 188: 119–125.

Durand J, Bodénès C, Chancerel E, Frigerio J M, Vendramin G, et al. (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC genomics*, *11*(1), 570.

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea (*Cajanus cajan* (L.) Millspaugh). BMC plant biology. 11(1): 1.

Dyah J, Oen S, Azizahwati LH, Winarno FG. (1997) Capsaicin content of various varieties of Indonesian chillies. Asia Pacific J Clin Nutr. 6: 99-101.

East EM. (1908) Inbreeding in corn. Rep. Conn. Agric. Exp. 419-28.

El-Rodeny W, Kimura M, Hirakawa H, Sabah A, Shirasawa K, et al. (2014) Development of EST-SSR markers and construction of a linkage map in faba bean (*Vicia faba*). Breeding Sci. *64*(3), 252-263.

Eshbaugh WH (1993) Peppers: history and exploitation of a serendipitous new crop discovery. New crops. Wiley, New York. 132-139.

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

Eveland, Andrea L., and David P. Jackson. (2011) "Sugars, signalling, and plant development." J Exp Bot. 63.9: 3367-3377.

Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, et al. (2007) High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One. 2: e219.

FAO, (2016) FAOSTAT database collections. Food and Agriculture Organization of the United Nations. Rome. Access date: 2017-06-22. (http://faostat.fao.org)

Ferreira TH, Gentile A, Vilela RD, Costa GG, Dias LI, et al. (2012) microRNAs associated with drought response in the bioenergy crop sugarcane (*Saccharum* spp.). PLoS ONE. 7: e46703.

Filipowicz W, Bhattacharyya SN, Sonenberg N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?. Nat Rev Genet. 9(2):102.

Fire A, Xu S, Montgomery MK, Kostas SA. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 391(6669):806.

Folta, K. M., Staton, M., Stewart, P. J., Jung, S., Bies, D. H., et al. (2005). Expressed sequence tags (ESTs) and simple sequence repeat (SSR) markers from octoploid strawberry (*Fragaria*× *ananassa*). BMC Plant Biol, *5*(1), 12.

Frazier TP, Xie F, Freistaedter A, Burklew CE, Zhang B. (2010) Identi fi cation and characterization of microRNAs and their target genes in tobacco (*Nicotiana tabacum*). Planta. 232: 1289–1308.

Fu C, Sunkar R, Zhou C, Shen H, Zhang JY, et al. (2012). Overexpression of miR156 in switchgrass (*Panicum virgatum* L.) results in various morphological alterations and leads to improved biomass production. Plant Biotechnol J. 10: 443–452.

Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES. (2012) Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. Proc Natl Acad Sci USA. 109: 7109–7114.

Gao L, Tang J, Li H, Jia J. (2003) Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol Breed, *12*(3), 245-261.

Ge X, Chen W, Song S, Wang W, Hu S, et al. (2008) Transcriptomic profiling of mature embryo from an elite super-hybrid rice LYP9 and its parental lines. BMC Plant Biol. 8: 114.

Geleta LF, Labuschagne MT. (2004) Hybrid performance for yield and other characteristics in peppers (*Capsicum annuum* L.). J Agri Sci. 142: 411–419.

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ. (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu Rev Genet. 44: 445-477.

Gnayfeed MH, Daood HG, Biacs PA, Alcaraz CF. (2001) Content of bioactive compounds in pungent spice red pepper (paprika) as affected by ripening and genotype. J Sci Food Agric. 81: 1580–1585.

Gogoi D, Gautam BP. (2002) Variability, heritability and genetic advance in chilli (*Capsicum* spp.). Agric Sci Dig. 22: 102–104.

Golz JF, Hudson A. (1999) Plant development: YABBYs claw to the fore. Curr Biol. 9: R861–R863.

Gong L, Stift G, Kofler R, Pachner M, Lelley T. (2008) Microsatellites for the genus Cucurbita and an SSR-based genetic linkage map of *Cucurbita pepo* L. Theor Appl Genet. 117: 37-48.

GONG YM, XU SC, MAO WH, LI ZY, HU QZ, et al. (2011) Genetic diversity analysis of faba bean (*Vicia faba* L.) based on EST SSR markers. Agri Sci China. 10(6): 838-844.

Govindarajan VS, Rajalakshmi D, Chand N, Salzer UJ. (1987) *Capsicum*—Production, technology, chemistry, and quality. Part IV. Evaluation of quality. Crc Cr Rev Food Sci, 25(3), 185-282.

Govindarajan VS, Salzer UJ (1985) Capsicum- production, technology, chemistry, and quality part 1: History, botany, cultivation, and primary processing, CRC Critical Reviews in Food Science and Nutrition, 22:2, 109-176

Griffing B, Zsiros E. (1971) Heterosis associated with genotype-environment interactions. Genetics. 68(3):443.

Griffing B. (1990) Use of a controlled-nutrient experiment to test heterosis hypotheses. Genetics. 126(3):753-67.

Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ. (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res. 36: D154–D158.

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, et al. (2011) Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. Proc Natl Acad Sci USA 108:2617–2622.

Gu Y, Liu Y, Zhang J, Liu H, Hu Y, et al. (2013) Identifi cation and characterization of microRNAs in the developing maize endosperm. Genomics. 102: 472–478.

Guo H, Kan Y, Liu W. (2011) Differential expression of miRNAs in response to topping in fl ue-cured tobacco (*Nicotiana tabacum*) roots. PLoS ONE. 6: e28565.

Gupta PK, Varshney RK, Prasad M. (2002) Molecular markers: principles and methodology. Molecular techniques in crop improvement. 9-54

Gupta SK, Bansal R, Gopalakrishna T. (2014) Development and characterization of genic SSR markers for mungbean (*Vigna radiata* (L.) Wilczek). *Euphytica*, *195*(2), 245-258.

Gupta SK, Gopalakrishna T. (2010) Development of unigenederived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other Vigna species. Genome. 53: 508–523.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. (2013) *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 8(8): 1494-1512.

Han HQ, Liu Y, Jiang MM, Ge HY, Chen HY. (2015) Identification and expression analysis of YABBY family genes associated with fruit shape in tomato (*Solanum lycopersicum* L.). Genet Mol Res. 14: 7079-91.

Han J, Fang J, Wang C, Yin Y, Sun X, et al. (2014) Grapevine microRNAs responsive to exogenous Gibberellin. BMC Genomics. 15: 111.

Han J, Kong ML, Xie H, Sun QP, Nan ZJ, et al. (2013). Identifi cation of miRNAs and their targets in wheat (*Triticum aestivum* L.) by EST analysis. Genet Mol Res. 12: 3793–3805.

Han ZG, Guo WZ, Song XL, Zhang TZ. (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboretum* in allotetraploid cotton. Mol Genet Genomics. 272: 308-327.

Hasanuzzaman M, Hakim MA, Hanafi MM, Shukor-Juraimi A, Islam MM, et al. (2013) Study of heterosis in Bangladeshi chilli (*Capsicum annuum* L.) landraces. Agrociencia. 47(7): 683-90.

He G, Zhu X, Elling AA, Chen L, Wang X, et al. (2010) Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell. 22(1): 17–33.

Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, et al. (2007) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res. 14: 271-281.

Ho T, Wang H, Pallett D, Dalmay T. (2007) Evidence for targeting common siRNA hotspots and GC preference by plant Dicer-like proteins. FEBS Lett. 581: 3267–3272.

Hochholdinger F, Hoecker N. (2007) Towards the molecular basis of heterosis. Trends Plant Sci. 12: 427–432.

Hoecker N, Keller B, Piepho HP, Hochholdinger F. (2006) Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theor Appl Genet. 112(3): 421–429.

Hofacker IL. (2003) Vienna RNA secondary structure server. Nucleic Acids Res. 31: 3429–3431.

Horton P. (2000) Prospects for crop improvement through the genetic manipulation of photosynthesis: morphological and biochemical aspects of light capture. J Exp Bot. 51: 475-485.

Howe E, Holton K, Nair S, Schlauch D, Sinha R, et al. (2010) Mev: multiexperiment viewer. In Biomedical informatics for cancer research (pp. 267-277).

Huang HH, Zhang ZH, Zhang ZH, Mao SL, Wang LH, et al. (2011) Analysis of SSRs information in *Capsicum* spp. from EST database. *Agri Sci China.10*(10), 1532-1536.

Huang Z, Van Houten J, Gonzalez G, Xiao H, van der Knaap E. (2013) Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato. Mol Genet Genomics. 288: 111-129.

Hubner N, Wallace C A, Zimdahl H, Petretto E, Schulz H, et al.(2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature genet*, *37*(3), 243.

Hwang DG, Park JH, Lim JY, Kim D, Choi Y, et al. (2013) The hot pepper (*Capsicum annuum*) microRNA transcriptome reveals novel and conserved targets: a foundation for understanding MicroRNA functional roles in hot pepper. PLoS One. 8: e64238.

Ibarra-Torres P, Valadez-Moctezuma E, Pérez-Grajales M, Rodríguez-Campos J, Jaramillo-Flores ME. (2015) Inter-and intraspecific differentiation of *Capsicum annuum* and *Capsicum pubescens* using ISSR and SSR markers. Sci Hortic-Amsterdum. *181*, 137-146.

Ibarra-Torres, P, Valadez-Moctezuma E, Pérez-Grajales M, Rodríguez-Campos J, Jaramillo-Flores M E. (2015) Inter-and intraspecific differentiation of *Capsicum annuum* and *Capsicum pubescens* using ISSR and SSR markers. *Scientia Horticulturae*, *181*, 137-146.

Ince AG, Karaca M, Onus AN. (2009) Development and utilization of diagnostic DAMD-PCR markers for *Capsicum* accessions. Genet Resour Crop Evol. 56: 211–221.

Iwai K, Suzuki T, Fujiwake H. (1979) Formation and accumulation of pungent principle of hot pepper fruits, capsaicin and its analogues, in *Capsicum annuun* var. annuun cv. karayatsubusa at different growth stages after floweringAgric Biol Chem, 43(12), 2493-2498.

Jagadeeswaran G, Zheng Y, Li YF, Shukla LI, Matts J, et al. (2009) Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume specific microRNA families. New Phytol. 184: 85-98.

Jain M, Chevala VN, Garg R. (2014) Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. J Exp Bot. 65: 5945–5958.

Jansson MD, Lund AH. (2012) MicroRNA and cancer. Mol Oncol. 6(6):590-610.

Jeong DH, Park S, Zhai J, Gurazada SGR, De Paoli E, et al. (2011) Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. Plant Cell. 23: 4185-4207.

Jiang D, Zhong GY, Qi-Bing hong. (2006) Analysis of microsatellites in citrus unigenes. *Acta Genetica Sinica*, *33*(4), 345-353.

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of OsSPL14 by OsmiR156 defi nes ideal plant architecture in rice. Nat. Genet. 42, 541–544

Jones DF. (1917) Dominance of linked factors as a means of accounting for heterosis. P Natl Acad Sci USA. 3(4):310-2.

Jones-Rhoades MW, Bartel DP, Bartel B. (2006) MicroRNAs and their regulatory roles in plants. Annu Rev Plant Biol. 57: 19–53.

Jones-Rhoades MW, Bartel DP. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell.14: 787–799.

Joshi R K, Kar B, Nayak S. (2011) Exploiting EST databases for the mining and characterization of short sequence repeat (SSR) markers in *Catharanthus roseus* L. *Bioinformation*, *5*(9), 378.

Kadel C, Jain AK, (2008). Folklore claims on snakebite among some tribal communities of central India. Indian J Tradit Know. 7, 296–299.

Kaja E, Szcze´sniak MW, Jensen PJ, Axtell MJ, McNellis T, et al. (2015) Identifi cation of apple miRNAs and their potential role in fi re blight resistance. Tree Genet Genomes. 11: 812.

Kamatenesi-Mugisha M, Oryem-Origa H, (2005). Traditional herbal remedies used in the management of sexual impotence and erectile dysfunction in western Uganda. Afr Health Sci. 5(1), 40-49.

Kapoor M, Arora R, Lama T, Nijhawan A, Khurana JP, et al. (2008) Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA- dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. BMC Genomics 9: 451.

Karlova R, Haarst JC, Maliepaard C, Geest HVD, BovyAG, et al. (2013). Identifi cation of microRNA targets in tomato fruit development using high-throughput sequencing and degradome analysis. J Exp Bot. 64: 1863–1878.

Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, et al. (2007) Genome-wide profiling and analysis of *Arabidopsis* siRNAs. PLoS Biol. 5: e57.

Katiyar A, Smita S, Chinnusamy V, Pandey DM, Bansal KC. (2012) Identifi cation of miRNAs in sorghum by using bioinformatics approach. Plant Signal Behav. 7: 246–259.

Kavasch BE, Baar K. (1999) American Indian Healing Arts. Bantam Books, Random House Inc., New York.

Keeble F, Pellew C. (1910) The mode of inheritance of stature and of time of flowering in peas (*Pisum sativum*). J Genet. 1(1):47-56.

Kim HJ, Baek KH, Lee BW, Choi D, Hur CG. (2011) In silico identification and characterization of microRNAs and their putative target genes in Solanaceae plants. Genome. 54: 91-98.

Kim HS, Kwon HJ, Kim GE, Cho MH, Yoon et al. (2014) Attenuation of natural killer cell functions by capsaicin through a direct and TRPV1-independent mechanism. Carcinogenesis, 35(7), 1652-1660

Kim S, Park M, Yeom SI, Kim YM, Lee JM, et al. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat genet.46: 270-278.

Kim VN. (2005) MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol. 6: 376–385.

Kim, Daehwan, et al. (2013) "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biol. 14.4: R36.

Kim, Minwoo, et al. (2001) "Isolation of cDNA clones differentially accumulated in the placenta of pungent pepper by suppression subtractive hybridization." Mol cells (Springer Science & Business Media BV) 11.2

Kliebenstein DJ, Monde RA, Last RL. (1998) Superoxide dismutase in *Arabidopsis*: an eclectic enzyme family with disparate regulation and protein localization. Plant Physiol.118: 637–650.

Koc I, Filiz E, Tombuloglu H. (2015) Assessment of miRNA expression profi le and differential expression pattern of target genes in cold-tolerant and cold-sensitive tomato cultivars. Biotechnol Biotechnol Equip. 29: 851–860.

Koek MM, Bakels F, Engel W, Van Den Maagdenberg A, Ferrari MD, et al. (2010) Metabolic profiling of ultra small sample volumes with GC/MS: From microliter to nanoliter samples. Anal Chem. 82: 156–162.

Kong Q, Zhang G, Chen W, Zhang Z, Zou X. (2012) Identification and development of polymorphic EST-SSR markers by sequence alignment in pepper, *Capsicum annuum* (Solanaceae). Am J Bot. *99*(2), e59-e61.

Korn, Marina, et al. (2010) "Predicting Arabidopsis freezing tolerance and heterosis in freezing tolerance from metabolite composition." Mol Plant. 3.1: 224-235.

Kosuge S, Inagaki Y, Okumura H. (1965) Pungent principles in *Capsicum.* Nippon Nogei Kagaki Kaishi. 35: 540.

Krol J, Loedige I, Filipowicz W. (2010) The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet. 11(9):597.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an information aesthetic for comparative genomics. Genome Res. 19: 1639-1645.

Kwon YS, Lee JM, Yi GB, Yi SI, Kim KM, et al. (2005) Use of SSR markers to complement tests of distinctiveness, uniformity, and stability (DUS) of pepper (*Capsicum annuum* L.) varieties. Mol Cells. *(Springer Science & Business Media BV), 19*(3).

La Rota M, Kantety RV, Yu JK, Sorrells ME. (2005) Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics. 6: 23.

Lakhotia N, Joshi G, Bhardwaj AR, Katiyar-Agarwal S, Agarwal M, et al. (2014) Identification and characterization of miRNAome in root, stem, leaf and tuber developmental stages of potato (*Solanum tuberosum* L.) by high-throughput sequencing. BMC plant biology. 14: 6.

Lamkey KR, Edwards JW. (1998) Heterosis: theory and estimation. Proceedings of the 34th Illinois Corn Breeders' School, Urbana. 62-72.

Lang, Yaqin, et al. (2009) Functional loss of *pAMT* results in biosynthesis of capsinoids, capsaicinoid analogs, in *Capsicum annuum* cv. CH- 19 Sweet. The Plant J 59.6: 953-961.

Lapitan VC, Brar DS, Abe T, Redona ED, (2007) Assessment of genetic diversity of Philippine rice cultivars carrying good quality traits using SSR markers. *Breed Sci* 57:263–270

Lawson MJ, Zhang L. (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome biology*, *7*(2), R14.

Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P. (2006) Fbox proteins everywhere. Curr Opin Plant Biol. 9: 631–638.

Lee RC, Feinbaum RL, Ambros V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell, *75*(5), 843-854.

Lefebvre V, Palloix A, Max R. (1993) Nuclear RFLP between pepper cultivars (*Capsicum annuum* L.). Euphytica. 71: 189–199.

Lelandais-Briere C, Naya L, Sallet E, Calenge F, Frugier F, et al. (2009) Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. Plant Cell. 21: 2780–2796.

Lewers K S, Saski C A, Cuthbertson B J, Henry D C, Staton, et al. (2008). A blackberry (Rubus L.) expressed sequence tag library for the development of simple sequence repeat markers. *BMC plant biology*, *8*(1), 69.

Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. (2003) Prediction of mammalian microRNA targets. Cell. 115(7):787-98.

Li A, Liu D, Wu J, Zhao X, Hao M, et al. (2014) mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. The Plant Cell. 26(5): 1878-900.

Li A, Mao L. (2007) Evolution of plant microRNA gene families. Cell Res. 17: 212–218.

Li D, Wang L, Liu X, Cui D, Chen T, et al. (2013) Deep sequencing of maize small RNAs reveals a diverse set of MicroRNA in dry and imbibed seeds. PLoS One. 8: e55107.

Li J, Guo G, Guo W, Guo G, Tong D, et al. (2012) miRNA164-directed cleavage of *ZmNAC1* confers lateral root development in maize (*Zea mays* L.). BMC Plant Biol. 12: 220.

Li YC, Korol AB, Fahima T, Nevo E. (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol. 21(6): 991-1007.

Li YH, Li W, Zhang C, Yang L, Chang RZ, et al. (2010) Genetic diversity in domesticated soybean (Glycine max) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single- nucleotide polymorphism loci. New phytol. *188*(1), 242-253.

Lippman ZB, Zamir D. (2007) Heterosis: Revisiting the magic. Trends Genet. 23(2): 60–66.

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. (2006) Gas chromatography mass spectrometry–based metabolite profiling in plants. Nat Protoc. 1: 387–396.

Litt M. and Luty JA. (1989) A hypervariable microsatellite revealed by in vitro

Liu F, Wang YS, Tian XL, Mao ZC, Zou XX, et al. (2012) SSR mining in pepper (*Capsicum annuum* L.) transcriptome and the polymorphism analysis [J]. Acta Hortic Sinica. *39*(1), 168-174.

Liu K, Muse SV. (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 21(9): 2128-2129.

Liu R, Qian W, Meng J. (2002) Association of RFLP markers and biomass heterosis in trigenomic hybrids of oilseed rape (*Brassica napus*× *B. Campestris*). Theor Appl Genet. 105(6):1050-7.

Liu Z, Zhang Y, Ou L, Kang L, Liu Y, et al. (2017) Identification and characterization of novel microRNAs for fruit development and quality in hot pepper (*Capsicum annuum* L.). Gene. 608: 66-72.

Liu, Shaoqun, et al. (2013) De novo transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. PloS one 8.1: e48156.

Livak KJ, Schmittgen TD. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 25: 402-408.

Llave C, Xie Z, Kasschau KD, Carrington JC. (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science. 297: 2053–2056.

Loss-Morais G, Ferreira DCR, Margis R, Alves-Ferreira M, Corrêa RL. (2014) Identification of novel and conserved microRNAs in *Coffea canephora* and *Coffea arabica*. Genet Mol Biol. 37: 671–682.

Lu F H, Yoon M Y, Cho Y I, Chung J W, Kim K T, (2011). Transcriptome analysis and SNP/SSR marker information of red pepper variety YCM334 and Taean. Scientia Horticulturae, 129(1), 38-45.

Lu YB, Qi YP, Yang LT, Guo P, Li Y, et al. (2015) Boron deficiency-responsive microRNAs and their targets in *Citrus sinensis* leaves. BMC Plant Biol. 15: 271.

Lu YB, Yang LT, Qi YP, Li Y, Li Z, et al. (2014) Identification of boron-deficiency-responsive microRNAs in Citrus sinensis roots by Illumina sequencing. BMC Plant Biol. 14: 123.

M S,Gaur R,Sharma V,Chhapekar SS, Das J. (2016) et al. Comparative Analysis of Fruit Metabolites and Pungency Candidate Genes Expression between Bhut Jolokia and Other Capsicum Species. PloS one 11.12 : e0167791.

Maga JA, Todd PH. (1975) Capsicum. Crit Rev Food Sci Nutr. 6(2): 177-199.

Maher C, Stein L, Ware D. (2006) Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res. 16: 510–519.

Mallory AC, Vaucheret H. (2006) Functions of microRNAs and related small RNAs in plants. Nature Genet.38: S31–S36.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 18: 1509–1517.

Martinez G, Forment J, Llave C, Pallas V, Gomez G. (2011) High-throughput sequencing, characterization and detection of new and conserved cucumber miRNAs. PLoS One 6: e19523.

Mathur R, Dangi R, Dass S, Malhotra R. (2000) The hottest chilli variety in India. Curr Sci. 79: 287–288.

Mazourek M, Pujar A, Borovsky Y, Paran I, Mueller L, et al. (2009) A dynamic interface for capsaicinoid systems biology. Plant physiol, 150(4), 1806-1821.

Meghvansi MK, Siddiqui S, Khan MH, Gupta VK, Vairale MG, et al. (2010) Naga chilli: a potential source of capsaicinoids with broad-spectrum ethnopharmacological applications. J Ethnopharmacol. 132(1): 1-14.

Melchinger AE. (1999) Genetic diversity and heterosis. The genetics and exploitation of heterosis in crops. Thegeneticsande :99-118.

Mesfin F, Demissew S, Teklehaymanot T. (2009) An ethnobotanical study of medicinal plants in Wonago Woreda, SNNPR, Ethiopia. J Ethnobiol and Ethnomed. 5(1), 28.

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

Meyer S, Pospisil H, Scholten S. (2007) Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. Plant Mol Biol. 63(3):381-91.

Meyer, Rhonda C., et al. (2004) Heterosis of biomass production in *Arabidopsis*. Establishment during early development. Plant Physiol. 134.4: 1813-1823.

Meyers B C, Axtell M J, Bartel B, Bartel D P, Baulcombe D, et al. (2008) Criteria for annotation of plant microRNAs. Plant Cell. 20: 3186–3190.

Mi S, Cai T, Hu Y, Chen Y, Hodges E, et al. (2008) Sorting of small RNAs into *Arabidopsis* argonaute complexes is directed by the 5' terminal nucleotide. Cell. 133: 116–127.

Minamiyama Y, Kinoshita S, Inaba K, Inoue M. (2005) Development of a cleaved amplifi ed sequence (CAPS) marker linked to pungency in pepper. Plant Breed. 124:288–291.

Mishra AK, Agarwal S, Jain CK, Rani V. (2009) High GC content: critical parameter for predicting stress regulated miRNAs in *Arabidopsis thaliana*. Bioinformation. 4: 151–154.

Morgante M, Hanafey M, Powell W. (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet. 30: 194–200.

Morin RD, Aksay G, Dolgosheina E, Ebhardt HA, Magrini V, et al. (2008) Comparative analysis of the small RNA transcriptomes of *Pinus contorta* and *Oryza sativa*. Genome Res. 18: 571–584.

Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, et al. (2009) Next-generation tag sequencing for cancer gene expression profiling. Genome Res. 19: 1825–1835.

Moxon S, Jing RC, Szittya G, Schwach F, Pilcher RLR, et al. (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Res. 18: 1602–1609.

Mutum RD, Kumar S, Balyan S, Kansal S, Mathur S, et al. (2016) Identification of novel miRNAs from drought tolerant rice variety Nagina 22. Sci Rep. 6.

Naito Y, Suzuki S, Iwata Y, Kuboyama T (2008) Genetic diversity and relationship analysis of peanut germplasm using SSR markers. *Breed Sci* 58:293–300

Nallathambi G, Kumari B M. (2003) Heterosis for rooting characters in maize (*Zea mays* L.). *Madras Agric. J*. 90.1/3: 154-156.

Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, et al. (2006). A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science. 312(5772):436-9.\

Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EH, Allan AC, et al. (2006). Analyses of expressed sequence tags from apple. *Plant Physiol*, *141*(1), 147-166.

Ng DW, Lu J, Chen ZJ. (2012) Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. Curr Opin Plant Biol. 15(2):154-61.

Ni Z, Kim ED, Ha M, Lackey E, Liu J, et al. (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature. 457: 327–331.

Nicolai M, Pisani C, Bouchet J P, Vuylsteke M , Palloix A. (2012) Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*). *Genet Mol Res*, *11*(3), 2295-2300.

Ochoa-Alejo N, Gómez-Peralta JE. (1993) Activity of enzymes involved in capsaicin biosynthesis in callus tissue and fruits of chili pepper (*Capsicum annuum* L.). J Plant Physiol. 141: 147–152.

Odeny DA, Jayashree B, Gebhardt C, Crouch J. (2009) New microsatellite markers for pigeonpea (*Cajanus cajan* (L.) millsp.). BMC Res Notes. 2: 35.

Odugbemi TO, Akinsulire OR, Aibinu E, Fabeku PO. (2007) Medicinal plants useful for malaria therapy in Okeigbo, ondo state, southwest Nigeria. Afr J Tradit Complem**.** 4: 191–198.

Ogawa K, Murota K, Shimura H, Furuya M, Togawa Y, et al. (2015) Evidence of capsaicin synthase activity of the *Pun1*-encoded protein and its role as a determinant of capsaicinoid accumulation in pepper. BMC Plant Biol. 15: 93.

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira ML. (2006) Origin, evolution and genome distribution of microsatellites. Genet Mol Biol. *29*(2), 294-307.

Ori N, Cohen AR, Etzioni A, Brand A, Yanai,O, Shleizer S, et al. (2007). Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. Nature Genet. 39, 787–791.

Pacheco-Olvera A, Hernández-Verdugo S, Rocha-Ramírez V, González-Rodríguez A, Oyama K. (2012) Genetic diversity and structure of pepper (*Capsicum annuum* L.) from Northwestern Mexico analyzed by microsatellite markers. Crop Sci. 52: 231–241.

Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, et al. (2007) Sequence and expression differences underlie functional specialization of *Arabidopsis* microRNAs miR159 and miR319. Dev Cell.13: 115–125.

Pant BD, Buhtz A, Kehr J, Scheible WR. (2008) MicroRNA399 is a long distance signal for the regulation of plant phosphate homeostasis. Plant J. 53: 731–738.

Pantaleo V, Szittya G, Moxon S, Miozzi L, Moulton V, et al. (2010) Identification of grapevine microRNAs and their targets using highthroughput sequencing and degradome analysis. Plant J. 62: 960–976.

Parida S K, Verma M, Yadav S K, Ambawat S, Das S, et al. (2015). Development of genome-wide informative simple sequence repeat markers for large-scale genotyping applications in chickpea and development of web resource. *Front Plant Sci*, 6: 645

Parida S, Pandit A, Gaikwad K, Sharma TR, Srivastava PM, et al. (2010) Functionally relevant microsatellites in sugarcane unigenes. BMC Plant Biol., 10:251

Parida SK, Kalia SK, Kaul S, Dalal V, Hemaprabha G, et al. (2009) Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. Theor. Appl. Genet., 118:327-338

Paschold, Anja, et al. (2010) Molecular dissection of heterosis manifestation during early maize root development.Theor Appl Genet. 120.2: 383-388.

Pasquinelli AE. (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nature Rev Genet. 13: 271-282.

Patanun O, Lertpanyasampatha M, Sojikul P, Viboonjun U, Narangajavana J. (2013) Computational identifi cation of microRNAs and their targets in Cassava (*Manihot esculenta* Crantz.). Mol Biotechnol. 53: 257–269.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, et al. (2009) The Sorghum bi color genome and the diversifi cation of grasses. Nature. 457: 551–556.

Pavlrkova E, Roor SB. (1987) Cellular basis of heterosis for leaf area in maize. Can. J. Plant Sci. 67: 99-104.

Peng JH, Lapitan NL. (2005) Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct Integr Genomics. 5: 8-96.

Peng Zhang,Xiangdong Liu, Hanhua Tong, Yonggen Lu, Jinquan Li. 2014 "Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers." *PloS one* 9.10 (2014): e111508.

Dickau R, Zarrill S Hols I Perry L Piperno DR, et al. (2007) Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. Science. 315.5814: 986-988.

Perucka I, Materska M. (2001) Phenylalanine ammonia-lyase and antioxidant activities of lipophilic fraction of fresh pepper fruits *Capsicum annum* L. Innov Food Sci Emerg Technol. 2: 189–192.

Pickersgill, B. A. R. B. A. R. A. (1991) "Cytogenetics and evolution of Capsicum L." Chromosome engineering in plants: genetics, breeding, evolution, part B. Elsevier, Amsterdam : 139-160.

Pieroni A, Quave CL. (2005) Traditional pharmacopoeias and medicines among Albanians and Italians in southern Italy: a comparison. J Ethnopharmacol. 101(1), 258-270.

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, et al. (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the Coffea genus. Mol. Genet. Genomics, 276:436-449

Porebski S, Bailey LG, Baurn BR. (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep. 15: 8-15.

Portis E, Nagy I, Sasvari Z, Stágel A, Barchi L, et al. (2007) The design of *Capsicum* spp. SSR assays via analysis of in silico DNA sequence, and their potential utility for genetic mapping. *Plant Sci* 172:640–648

Potato Genome Sequencing Consortium. (2011). Genome sequence and analysis of the tuber crop potato. Nature, 475(7355), 189-195.

Powers L. (1945) Relative yields of inbred lines and F1 hybrids of tomato. Botanical Gazette. 106(3):247-68.

Pruthi JS. (1980) Spices and Condiments-Chemistry, Microbiology and Technology. New York: Academic Press Inc, USA. 1-450.

Purkayastha J, et al. (2012) Molecular characterization of'Bhut Jolokia'the hottest chilli. J biosciences. 37.4: 757-768.

Qin C, Yu C, Shen Y, Fang X, Chen L, et al. (2014) Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci USA.111: 5135-5140.

Qin YF, Fang HM, Tian QN, Bao ZX, Lu P, et al. (2011) Transcriptome profiling and digital gene expression by deep-sequencing in normal/regenerative tissues of planarian *Dugesia japonica*. Genomics. 97: 364–371.

Qiu D, Pan X, Wilson IW, Ketchum REB, Li F, et al. (2009) High-throughput sequencing technology reveals that the taxoid elicitormethyl jasmonate regulates microRNA expression in Chinese yew (*Taxus chinensis*). Gene. 436: 37–44.

Kumar RL, Sridevi , Kage U PM, Salimath, Madalageri, et al. (2014) Heterosis Studies in Chilli (*Capsicum annuum* L.). International Journal of Horticulture. 4(8), 40-43.

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev. 20: 3407–3425.

Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, et al. (2012) Seed germination and vigor. Annu Rev Plant Biol. 63: 507–533.

Ramchiary N, Kehie M, Brahma V, Kumaria S, Tandon P. (2014) Application of genetics and genomics towards *Capsicum* translational research. Plant Biotechnol Rep. 8(2): 101-123.

Ramchiary N, Nguyen VD, Li X, Hong CP, Dhandapani V, et al. (2011) Genic microsatellite markers in *Brassica rapa*: development, characterization, mapping, and their utility in other cultivated and wild Brassica relatives. DNA Res. 18(5): 305-320.

Rao GP, Reddy MK, Naresh P, Chalapathi VV, Indiresh KM. (2016) Combining ability and standard heterosis in bell pepper (*Capsicum annuum* L.) for yield and yield attributing traits. Green Farming Int J. 7(5): 1024-1028.

Rao NKS, Bhatt RM, Anand N. (1992) Leaf area, growth and photosynthesis in relation to heterosis in tomato. Photosynthetica. 26: 449–454.

Rao, P. Gangadhara, et al. (2017) "Heterosis in bell pepper (*Capsicum annuum* L.) for yield and yield attributing traits." BANGLADESH JOURNAL OF BOTANY 46.2: 745-750.

Rapp Ryan A, Joshua A Udall, Jonathan F Wendel. (2009) "Genomic expression dominance in allopolyploids." BMC Biol. 7.1: 18.

Reddy RN, Madhusudhana R, Mohan SM, Chakravarthi DV, Seetharama N. (2012) Characterization, development and mapping of Unigene derived microsatellite markers in sorghum [*Sorghum bicolor* (L.) Moench]. Mol Breeding. 29(3): 543-564.

Reddy UK, Almeida A, Abburi VL, Alaparthi SB, Unselt D, et al. (2014) Identification of gene-specific polymorphisms and association with capsaicin pathway metabolites in Capsicum annuum L. collections. PLoS One. 9: 1–10.

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. (2002) MicroRNAs in plants. Genes Dev. 16: 1616–1626.

Rodriguez-Uribe L, Guzman I, Rajapakse W, Richins RD, O'Connell MA. (2012) Carotenoid accumulation in orange-pigmented *Capsicum annuum* fruit, regulated at multiple levels. J Exp Bot. 63: 517–526.

Rogers K, Chen X. (2013) Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell. 25: 2383–2399.

Rubio-Somoza I, Weigel D. (2011) MicroRNA networks and developmental plasticity in plants. Trends Plant Sci. 16: 258–264.

Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, et al. (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor Appl Genet. 109: 1283-1294.

Saha M C, Mian M R, Eujayl I, Zwonitzer J C, Wang L, et al. (2004). Tall fescue EST-SSR markers with transferability across several grass species. *Theoretical and Applied Genetics*, *109*(4), 783-791.

Samal PK, Dhyani PP. (2006). Gender in the management of indigenous knowledge: reflections from Indian Central Himalaya. Curr Sci. 104-108.

Sanatombi K, Sharma GJ. (2008) Capsaicin Content and Pungency of Different *Capsicum* spp. Cultivars. Not Bot Horti Agrobot Cluj-Napoca. 36: 89–90.

Sasmal B, S P. Banerjee. (1986) "Combining ability for grain yield and other agronomic characters in rice." J Agron Crop Sci. 156.1: 18-23.

Schnable PS, Springer NM. (2013) Progress toward understanding heterosis in crop plants. Annu Rev Plant Biol. 64:71-88.

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, et al. (2005) Specific effects of microRNAs on the plant transcriptome.Dev Cell.8: 517–527.

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, et al. (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet. 100: 723-726.

Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, et al. (2006) Overdominant quantitative trait loci for yield and fitness in tomato. P Natl Acad Sci USA. 103(35):12981-6.

Senthilvel S, Jayashree B, Mahalakshmi V, Sathish Kumar P, Nakka S, et al. (2008) Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. BMC Plant Biol., 8:119

Sharma N, Panchal S, Sanan-mishra N. (2015) Protocol for artifi cial microRNA mediated over-expression of miR820 in indica rice. Am J Plant Sci. 6: 1951–1961.

Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, et al. (2012) Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of ligno cellulosic feed stocks. New Phytol. 193: 121–136.

Sheng L, Chai W, Gong X, Zhou L, Cai R, et al. (2015) Identifi cation and characterization of novel maize miRNAs involved in different genetic background. Int J Biol Sci. 11: 781–793.

Shirasawa K, Ishii K, Kim C, Ban T, Suzuki M, et al. (2013). Development of *Capsicum* EST–SSR markers for species identification and in silico mapping onto the tomato genome sequence. *Mol Breed*, *31*(1), 101-110.

Shirasawa K, Isobe S, Tabata S, Hirakawa H. (2014) Kazusa Marker DataBase: a database for genomics, genetics, and molecular breeding in plants. Breeding Sci. *64*(3), 264-271.

Shrestha SL, Luitel BP, Kang WH. (2011) Heterosis and heterobeltiosis studies in sweet pepper (*Capsicum annuum* L.). Hortic Environ Biotechnol. 52(3): 278-283.

Shui XR, Chen ZW, Li JX. (2013) MicroRNA prediction and its function in regulating drought-related genes in cowpea. Plant Sci. 210: 25–35.

Shull GH. (1908) The composition of a field of maize.

Shuzuo LV, Nie X, Le W, Du X, Biradar SS, et al. (2012) Identifi cation and characterization of microRNAs from barley (*Hordeum vulgare* L.) by high-throughput sequencing. Int J Biol Sci. 13: 2973–2984.

Sieber P, Wellmer F, Gheyselinck J, Riechmann JL, Meyerowitz EM. (2007) Redundancy and specialization among plant microRNAs: roleof the MIR164 family in developmental robustness. Development. 134: 1051–1060.

Siju S, Dhanya K, Syamkumar S, Sasikumar B, Sheeja T E, et al. (2010). Development, characterization and cross species amplification of polymorphic microsatellite markers from expressed sequence tags of turmeric (*Curcuma longa* L.). *Mol. Biotechnol*, *44*(2), 140-147.

Simonne AH, Simonne EH, Eitenmiller RR, Mills HA, Green NR. (1997) Ascorbic acid and provitamin A contents in unusually colored bell peppers (*Capsicum annuum* L.). J Food Comp Anal. 10(4): 299-311.

Smeekens S, Ma J, Hanson J, Rolland F. (2010) Sugar signals and molecular networks controlling plant growth. Curr Opin Plant Biol. 13: 274–279.

Smeekens, Sjef, et al. (2010) "Sugar signals and molecular networks controlling plant growth." Curr Opin Plant Biol. 13.3: 273-278.

Somta P, Seehalak W, Srinives P. (2009) Molecular diversity assessment of AVRDC-World vegetable center elite-parental mungbeans. Breed Sci. 59: 149–157.

Song C, Wang C, Zhang C, Korir NK, Yu H, et al. (2010) Deep sequencing discovery of novel and conserved microRNAs in trifoliate orange (*Citrus trifoliata*). BMC Genomics. 11: 431.

Song C, Yu M, Han J, Wang C, Liu H, et al. (2012) Validation and characterization of *Citrus sinensis* microRNAs and their target genes. BMC Res Notes. 5: 235.

Song GS, Zhai HL, Peng YG, Zhang L, Wei G, et al. (2010) Comparative transcriptional profiling and preliminary study on heterosis mechanismofsuper-hybridrice. Mol Plant. 3(6): 1012–1025.

Song QX, Liu YF, Hu XY, Zhang WK, Ma B, et al. (2011) Identifi cation of miRNAs and their target genes in developing soybean seeds by deep sequencing. BMC Plant Biol. 11: 5.

Sood S, Kaul S. (2006) Heterosis in intraspecific hybrids for quantitative traits of bellpepper. Veg Sci. 33(2): 178-179.

Springer NM, Stupar RM. (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. The Plant Cell. 19(8):2391-402.

Sraphet S, Boonchanawiwat A, Thanyasiriwat T, Boonseng O, Tabata S, et al. (2011). SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor Appl Genet*, *122*(6), 1161-1170.

Srivastava HK. (1983) Heterosis and intergenomic complementation: mitochondria, chloroplast, and nucleus. InHeterosis (pp. 260-286). Springer Berlin Heidelberg.

Srivastava S, Zheng Y, Kudapa H, Jagadeeswaran G, Hivrale V, et al. (2015) High throughput sequencing of small RNA component of leaves and inflorescence revealed conserved and novel miRNAs as well as phasiRNA loci in chickpea. Plant Science. 235: 46-57.

Stágel A, Gyurjan I, Sasvari Z, Lanteri S, Ganal M, et al. (2009) Patterns of molecular evolution of microsatellite loci in pepper (*Capsicum* spp.) revealed by allele sequencing. Plant System Evol. 281: 251–254.

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, et al. (2005) The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. Plant J. 42: 675–688.

Stewart Jr C, Mazourek M, Stellari GM, O'connell M, Jahn M. (2007) Genetic control of pungency in C. chinense via the *Pun1* locus. J Exp Bot. 58(5), 979-991.

Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, et al. (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics. 28: 2059-2061.

Storey JD, Tibshirani R. (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci USA. 100: 9440–9445.

Stuber CW. (1999) Biochemistry, molecular biology, and physiology of heterosis. The Genetics and Exploitation of Heterosis in Crops. (thegeneticsande):173-83.

Sun F, Guo G, Du J, Guo W, Peng H, et al. (2014) Whole-genome discovery of miRNAs and their targets in wheat (*Triticum aestivum* L.). BMC plant biology. 14: 142.

Sun G, Stewart CN Jr, Xiao P, Zhang B. (2012) MicroRNA expression analysis in the cellulosic biofuel crop Switchgrass (*Panicum virgatum*) under abiotic stress. PLoS ONE 7: e32017.

Sun G. (2012) MicroRNAs and their diverse functions in plants. Plant Mol. Biol. 80: 17–36.

Sun R, Guo T, Cobb J, Wang Q, Zhang B. (2015) Role of microRNAs during fl ower and storage root development in sweet potato. Plant Mol Biol Rep. 33: 1731–1739.

Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS. (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc Natl Acad Sci USA. 103: 6805–6810.

Tahira W, M Buutta, A. Salam. (2011) "Estimation of heterosis in root length of wheat under different saline environment." Cereal Res Commun. 39.2: 306-314.

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, et al. (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. Theor Appl Genet.*110*(5), 819-831.

Tamura K, Dudley J, Nei M, Kumar S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol. 24(8): 1596-1599.

Tan Z, Fang X, Tang S, Zhang J, Liu D, et al. (2015) Genetic map and QTL controlling fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Euphytica*, *203*(3), 615-628.

Tanaka W, Toriba T, Ohmori Y, Yoshida A, Kawai A, et al. (2012) The YABBY gene *TONGARI-BOUSHI1* is involved in lateral organ development and maintenance of meristem organization in the rice spikelet. Plant Cell. 24: 80–95.

Tanaka Y, Nakashima F, Kirii E, Goto T, Yoshida Y, et al. (2017) Difference in capsaicinoid biosynthesis gene expression in the pericarp reveals elevation of capsaicinoid contents in chili peppers (*Capsicum chinense*). Plant Cell Rep. 36(2): 267-279.

Tang S, Wang Y, Li Z, Gui Y, Xiao B, et al. (2012) Identifi cation of wounding and topping responsive small RNAs in tobacco (*Nicotiana tabacum*). BMC Plant Biol. 12: 28.

Tautz D. (1989) Hypervariabflity of simple sequences as a general source for polymorphic DNA markers. *Nuc Acids Res*, *17*(16), 6463-6471.

Temnykh S, Park W D, Ayres N, Cartinhour S, Hauck N, et al. (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theor Appl Genet*, *100*(5), 697-712.

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

Thiemann A, Fu J, Schrag TA, Melchinger AE, Frisch M, et al. (2010) Correlation between parental transcriptome and field data for the characterization of heterosis in *Zea mays* L. Theor Appl Genet. 120(2): 401–413.

Todd PH, Bensinger MG, Biftu T. (1977) Determination of pungency due to Capsicum by Gas-Liquid Chromatography. J Food Sci. 42: 660–665.

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature, 485(7400), 635.

Topuz A, Ozdemir F. (2007) Assessment of carotenoids, capsaicinoids and ascorbic acid composition of some selected pepper cultivars (*Capsicum annuum* L.) grown in Turkey. J Food Compos Anal. 20: 596–602.

Trapnell, C., L. Pachter, and S.L. Salzberg. (2009) TopHat: Discovering splice junctions with RNA-Sequence. Bioinformatics. 25:1105–1111.

Truong HTH, Kim KT, Kim S, Kim HR, Cho MC, et al. (2009) Development of gene-based markers for the *Pun1* pungency gene in pepper (*Capsicum* spp.) for marker-assisted selection. Hortic Environ Biotechnol. 50(4): 358-365.

Tsaballa A, Ganopoulos I, Timplalexi A, Aliki X, Bosmali I, et al. (2015) Molecular characterization of Greek pepper (*Capsicum annuum* L) landraces with neutral (ISSR) and gene-based (SCoT and EST-SSR) molecular markers. Biochem Syst Ecol. *59*, 256-263.

Turner M, Yu O, Subramanian S. (2012) Genome organization and characteristics of soybean microRNAs. BMC Genomics. 13: 169.

Ueno D, Kono I, Yokosho K, Ando T, Yano M, et al. (2009). A major quantitative trait locus controlling cadmium translocation in rice (*Oryza sativa*). *New Phytol.*, *182*(3), 644-653.

Upadhyaya HD, Dwivedi SL, Baum M, Varshney RK, Udupa SM, et al. (2008) Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). BMC Plant Biol. 8: 106.

USDA-ARS, (2011) https://www.ars.usda.gov/

Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP. (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods. 3: 12.

Varkonyi-Gasicet E, Gould N, Sandanayaka M, Sutherland P, MacDiarmid RM. (2010) Characterisation of microRNAs from apple (*Malus domestica* 'Royal Gala') vascular tissue and phloem sap. BMC Plant Biol. 10: 159.

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

Varshney R, Grosse I, Hähnel U, Siefken U, Prasad M, et al.(2006) Genetic mapping and BAC assignment of EST derived SSR markers shows non-uniform distribution of genes in the barley genome. Theor Appl Genet. 113(2): 239-250.

Varshney RK, Graner A, Sorrells ME. (2005). Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23(1): 48-55.

Vilanova S, Manzur JP, Prohens J. (2012) Development and characterization of genomic simple sequence repeat markers in eggplant and their application to the study of diversity and relationships in a collection of different cultivar types and origins. Mol Breed. 30: 647–660.

Vogt T. (2010) Phenylpropanoid biosynthesis. Mol Plant. 3: 2–20.

Wang C, Shangguan L, Kibet KN, Wang X, Han J, et al. (2011) Characterization of microRNAs identifi ed in a table grapevine cultivar with validation of computationally predicted grapevine miRNAs by miR-RACE. PLoS ONE. 6: e21259.

Wang J, Lydiate DJ, Parkin IA, Falentin C, Delourme R, et al. (2011) Integration of linkage maps for the amphidiploid Brassica napus and comparative mapping with Arabidopsis and Brassica rapa. BMC genomics 12(1): 101.

Wang J, Tian L, Lee HS, Wei NE, Jiang H, et al. (2006) Genome-wide non-additive gene regulation in Arabidopsis allotetraploids. Genetics. 172(1): 507–517.

Wang M, Wang QL, Zhang BH. (2013) Response of miRNAs and their targets to salt and drought stresses in cotton (*Gossypium hirsutum* L.). Gene. 530: 26–32.

Wang Q, Zhang B. (2015) MicroRNAs in cotton: an open world needs more exploration. Planta. 241: 1303–1312.

Weber JL, May PE, (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet. 44(3): 388.

Wei G, Tao Y, Liu G Z, Chen C, Luo R Y, et al. (2009) A transcriptomic analysis of superhybrid rice LYP9 and its parents. Proc Natl Acad Sci USA. 106: 7695-7701.

Wells R, Meredith WRJ, Williford JR. (1988) Heterosis in upland cotton II. relationship of leaf area to plant photosynthesis. Crop Sci. 28: 522–525.

Wightman B, Ha I, Ruvkun G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell. 75(5):855-62.

Williams C. (1995) Jungle Medicines. Australian Well Being Magazine 59, 18

Wöhrmann T, Guicking D, Khoshbakht K, Weising K. (2011). Genetic variability in wild populations of Prunus divaricata Ledeb. in northern Iran evaluated by EST-SSR and genomic SSR marker analysis. Genet Resour Crop Ev. *58*(8), 1157-1167.

Wyatt L E, Eannetta N T, Stellari G M, Mazourek M. (2012). Development and application of a suite of non-pungency markers for the *Pun1* gene in pepper (*Capsicum* spp.). Mol. breed, 30(3), 1525-1529.

Xia K, Wang R, Ou X, Fang Z, Tian C, et al. (2012) *OsTIR1* and *OsAFB2* down regulation via OsmiR393 overexpression leads to more tillers, early fl owering and less tolerance to salt and drought in rice. PLoS ONE. 7: e30039.

Xie F, Frazier TP, Zhang B. (2011) Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (*Solanum tuberosum*). Gene. 473: 8-22.

Xie F, Stewart CN Jr, Taki FA, He Q, Liu H, et al. (2014) High throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. Plant Biotechnol J. 12: 354–366.

Xie KB, Shen JQ, Hou X, Yao JL, Li XH, et al. (2012) Gradual increase of miR156 regulates temporal expression changes of numerous genes during leaf development in rice. PlantPhysiol. 158: 1382–1394.

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, et al. (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biology. 2: E104.

Xu X, Li T, Li T, Wang H, Xu X. (2015) Denovo assembly and characterization of transcriptome and molecular discovery in *Capsicum annuum* L. R597. Int J Control Autom. 8(2): 105-116.

Xu Y, Ma RC, Xie H, Liu JT, Cao MQ. (2004) Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. Genome. 47: 1091-1104.

Yang L, Jue D, Li W, Zhang R, Chen M, et al. (2013) Identification of MiRNA from eggplant (*Solanum melongena* L.) by small RNA deep sequencing and their response to Verticillium dahliae infection. PloS one. 8: e72840.

Yang X, Zhang H, Li L. (2011) Global analysis of gene-level microRNA expression in *Arabidopsis* using deep sequencing data. Genomics. 98: 40–46.

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, et al. (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol. 60: 107–124.

Yao YY, Guo GG, Ni ZF, Sunkar R, Du JK, et al. (2007) Cloning and characterization of microRNAs from wheat (*Triticum aestivum* L). Genome Biol. 8: R96.

Yatung T, Dubey RK, Singh V, Upadhyay G. (2014) Genetic diversity of chilli (*Capsicum annuum* L.) genotypes of India based on morphochemical traits. Aust J Crop Sci. 8: 97–102.

Yi F, Xie S, Liu Y, Qi X, Yu J. (2013) Genome-wide characterization of microRNA in foxtail millet (*Setaria italica*). BMC plant biology. 13: 212.

Yi G, Lee JM, Lee S, Choi D, Kim BD. (2006). Exploitation of pepper EST–SSRs and an SSR-based linkage map. Theor Appl Genet. *114*(1), 113-130.

Yin ZJ, Li Y, Yu JW, Liu YD, Li CH, et al. (2012) Difference in miRNA expression profi les between two cotton cultivars with distinct salt sensitivity. Mol Biol Rep. 39: 4961–4970.

You FM, Huo N, Gu YQ, Luo MC, Ma Y, et al. (2008) BatchPrimer3: A high throughput web application for PCR and sequencing primer design. BMC Bioinformatics. 9: 253.

Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, et al. (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot. 99(2): 193-208.

Zanca AS, Vicentini R, Ortiz-Morea FA, Bem LED, da Silva MJ, et al. (2010) Identication and expression analysis of microRNAs and targets in the biofuel crop sugarcane. BMC Plant Biol. 10: 260.

Zewdie Y, Bosland PW. (2001) Capsaicinoid profiles are not good chemotaxonomic indicators for *Capsicum* species. Biochemical Systematics and Ecology, 29(2), 161-169.

Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, et al. (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev. 25: 2540–2553.

Zhang BH, Pan XP, Cox SB, Cobb GP, Anderson TA. (2006) Evidence that miRNAs are different from other RNAs. Cell Mol Life Sci. 63: 246–254.

Zhang BH, Wang QL, Wang KB, Pan XP, Liu F, et al. (2007) Identification of cotton microRNAs and their targets. Gene. 397: 26–37.

Zhang HY, He H, Chen LB, Li L, Liang MZ, et al. (2008) A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. Mol Plant.1: 720–731.

Zhang JG, Zeng R, Chen JS, Liu X, Liao QS. (2008) Identification of conserved microRNAs and their targets from *Solanum lycopersicum* Mill. Gene. 423: 1–7.

Zhang L, Chia JM, Kumari S, Stein JC, Liu Z, et al. (2009) A genome-wide characterization of microRNA genes in maize. PLoS Genet. 5: e1000716.

Zhang X, Zou Z, Gong PJ, Zhang JH, Ziaf K, et al. (2011) Overexpression of microRNA169 confers enhanced drought tolerance to tomato. Biotechnol Lett. 33: 403–409.

Zhang Y, Zhu X, Chen X, Song C, Zou Z, et al. (2014) Identi fi cation and characterization of cold-responsive microRNAs in tea plant (*Camellia sinensis*) and their targets using high-throughput sequencing and degradome analysis. BMC Plant Biol. 14: 271.

Zhang, Jinpeng, et al. (2015) De novo transcriptome sequencing of Agropyron cristatum to identify available gene resources for the enhancement of wheat." Genomics. 106.2: 129-136.

Zhang, Ti-Fu, et al. (2012) Genome-wide transcriptional analysis of yield and heterosis-associated genes in maize (*Zea mays* L.). J Integr Agr. 11.8: 1245-1256.

Zhao CZ, Xia H, Frazier TP, Yao YY, Bi YP, et al. (2010) Deep sequencing identifi es novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). BMC Plant Biol. 10: 3.

Zhao M, Meyers BC, Cai C, Xu W, Ma J. (2015) Evolutionary patterns and coevolutionary consequences of MIRNA genes and MicroRNA targets triggered by multiple mechanisms of genomic duplications in soybean. Plant Cell. 27: 546-562.

Zhao, Yusheng, et al. (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. Proc. Natl. Acad. Sci. U.S.A. 112.51: 15624-15629.

Zhou L, Liu Y, Liu Z, Kong D, Duan M, et al. (2010) Genome-wide identifi cation and analysis of drought-responsive microRNAs in *Oryza sativa*. J Exp Bot. 61: 4157–4168.

Zhu QW, Luo YP. (2013) Identifi cation of miRNAs and their targets in tea (*Camellia sinensis*). Biomed Biotechnol. 14: 916–923.

Zuo JH, Wang YX, Liu HP, Ma YZ, Ju Z, et al. (2011) MicroRNAs in tomato plants. Sci China Life Sci. 54: 599-605.

Annexure

Annexure-I

List of the primer used for qRT PCR analysis of capsaicinoid biosynthesis genes

Gene name	Sequence (5'-3')	Number of bases
ACL-F	CTGCATCTTCCTTCGCTATCT	21
ACL-R	CGAGTAGCTGGCTTCATTCT	20
ACS-F	ACGCCGTGAGATTGTAGATG	20
ACS-R	CTCCGGATGGAATTTCCTACTT	22
AMT-F	GCTGTCCTTGTAAGCCAGAA	20
AMT-R	CAGGGTGTCCGGAATAAGTAAA	22
BCAT-F	GCAAATTGGTGCAGAGAGAATG	22
BCAT-R	GGAATCCAGCGCTTGTTAGA	20
COMT-F	TCTGCTGATGAGGGACAATTC	21
COMT-R	CTGGAAGAGCAAGAGCAGTAG	21
C4H-F	TATCCTAGCGCTGCCAATTC	20
C4H-R	GACTGAACTGTCCACCTTTCTC	22
FatA-F	GTTTCGTATGAGCCGAGTCTT	21
FatA-R	CTCCAACCTCGTAACACCTAAC	22
Kas-F	GGACTAATGGGACCTTGTTACTC	23
Kas-R	CGTTCCACCTGCTACCATAAT	21
PAL-F	CAGATTGAGGCTGCTGCTATTA	22
PAL-R	GGAGATGTTCGGAGAGCATAAC	22
AT3-F	GGTTCTCTCATTACGCCACAA	21
AT3-R	CCTGATTCTTCTGCCACCTTAG	22
Actin -F	GTCCTCTTCCAACCATCCAT	20
Actin -R	TACTTTCTCTCTGGTGGTGC	20

List of the primer used for validation of *KAS*1 marker

*KAS*1-F	CGCGTCTCTTAAGGGAAAGAATC	23
*KAS*1-R	CGAGTCCATGACCAATCATCGAC	23

Primer sequences of YAB5a and YAB5b gene

YAB5a_FP	CAGCGAGTACCTTCTGCATATAA	23
YAB5a_RP	CAGTACTAAATGCCTCCCTGTG	22
YAB5b1_FP	TTGTTCTTGCGGTGAGTGT	19
YAB5b1_RP	AGCAGCAGCAGCCATATT	18

Annexure-II

List of the *insilico* polymorphic sequences between *C. chinense* and *C.frutescens*

Index	Primer name	*C. frutescens* Motif	*C. chinense* Motif	Forward Sequence	Reverse Seq	Avg. Tm	Prod size
1	CFpSSR1	TA(7)	TA(8)	GGTCGGATTGGACAAATTAT	CACGAACTTCTCCATCTTGTA	55.91	160
2	CFpSSR2	TA(8)	TA(7)	ACACTTGAACTTGCGAACAT	AATGGGACAGAGGGAGTAATA	55.545	149
3	CFpSSR3	TA(7)	TA(9)	TTGAGGATGGCTACAGTAGAA	TGTATCCTTCTCAGCATTCAC	55.265	140
4	CFpSSR4	CAA(6)	CAA(5)	ACCAACATACAAAGGTTTCG	CATGACTAGGCCCAATATCTA	54.83	128
5	CFpSSR5	TTC(5)	TTC(6)	CCGTTCTTTCTTTCCCTATT	GATGCTGTTGGAAGTTGTAGT	54.8	125
6	CFpSSR6	GCT(6)	GCT(5)	GCAGGACTGATAATGAAGTGA	CTTGTTGTTTCGATGAGGAG	55.64	143
7	CFpSSR7	ACA(6)	ACA(5)	CACTCCAATAGCAACCATAAC	GCTGATCTTGCTTCTGTAAAC	54.645	158
8	CFpSSR8	CTG(5)	CTG(6)	GGATGACATGAAGAAGAGGTT	AAGACAGATCGAGAATCCACT	55.58	159
9	CFpSSR9	AT(9)	AT(7)	TAACACGAACACCACGATTA	GCAATGAACTATGAACCCTCT	55.49	166
10	CFpSSR10	TG(10)	TG(6)	TATCCCCTTTTTACCCCTAA	CTAGCCAAATGTGTCATATCC	55.075	203
11	CFpSSR11	CTT(5)	CTT(6)	TCAAAGCCTCTTTAGTCTTCC	AAAGTGCTAGCGTTGTTGTT	55.41	147
12	CFpSSR12	GTT(6)	GTT(7)	TTAGCTGATACTACCGACGAC	GTTCGTCGAAATGTGAAGAT	54.98	157
13	CFpSSR13	TA(7)	TA(8)	ATACATGCCATACACACAAGC	ACAGACCTACATGATAACACGA	55.425	149
14	CFpSSR14	AT(7)	TA(6)	ATGATCTTAGGAGCTGAGGAT	CAAAGGTGAGGGTAGAGAAAT	54.86	143
15	CFpSSR15	CA(7)	CA(6)	CTCAGACACAAAATCCCAAT	GTAAACTGTTACAACGCCACT	54.82	127
16	CFpSSR16	TTC(5)	TTC(4)	TCACTTACTATCTCCCATGTCA	GTACCGAATGTAGATGAACGA	55.045	137
17	CFpSSR17	CT(8)	CT(6)	ACACCTTCCTCAATCTCTCAC	GTCGGAGAATTGAAAAACAC	55.27	141
18	CFpSSR18	TA(6)	TA(7)	GAAATTATACCGAGCTTCACC	AAACCACTCTGCCTCTTTTAC	55.44	153
19	CFpSSR19	ACA(7)	ACA(5)	GATGATGAATGTGGTGAAGAC	GTTGTTGCCGTGAACATAG	55.165	164
20	CFpSSR20	AT(5)	TA(7)	TATGGGAGAGCAAAACCTATC	GCTGTCACTTGTCTTCAACTT	55.31	150
21	CFpSSR21	CT(9)	CT(10)	CCCTTTCACAAACCAATCTA	ATATCAGCTCTCCCACTTTTC	55.185	141
22	CFpSSR22	AG(8)	AG(9)	CGTAAAATATCACCGGAAAC	CTCTCTCTCACGCTTTTTCTT	55.27	165
23	CFpSSR23	AC(7)	AC(9)	AAGAATTGGGAGTAGTGAGGT	CTCTCAAAGACTCAAACAGGA	54.6	136
24	CFpSSR24	CAA(4)	CAA(5)	TCCTGTTTGAGTCTTTGAGAG	GGAGAAAGAGATTCTTGGACT	54.475	156
25	CFpSSR25	TGG(5)	TGG(6)	TGGTACATCAATCAACTGGAC	TGATTCATCCACTAGTTCCAC	55.455	143
26	CFpSSR26	GCT(5)	GCT(6)	CACTTGACTCTATGCCCATTA	CAACCTGGAAAGAGTTATCTG	54.785	153
27	CFpSSR27	CA(8)	CA(6)	GTATTAGCCCAAAGTTGGAAG	CTGTTCTTGTCCTCTCTAGTCC	55.26	148
28	CFpSSR28	CT(7)	CT(6)	CTTTCTCTGATTGTCTTGCTG	AAAGAAAGGACAGAGATACGG	55.29	137
29	CFpSSR29	TC(7)	TC(6)	CCAGCTGGAATTAGTTAGTCA	GTTGGATGAAGAAGGAAGAAC	55.085	151
30	CFpSSR30	TCT(6)	TCT(5)	TCTGTTCAGATGTCATTCTCC	CTAATCCAGGCAGGTTAGTCT	55.21	149
31	CFpSSR31	ATG(6)	ATG(5)	GATGATGATAATGACGACGAC	CTTCTTTGGAGGTTGGATAGT	55.245	156
32	CFpSSR32	GA(8)	AG(6)	GTCTGATTCCTCCTGCATATT	ACGATTGAGCAAAAGAGAGA	55.535	151
33	CFpSSR33	CT(7)	CT(6)	TTCCTATTTCCACTAGGACAAG	CTATTTCCATTGTTCGGTTC	54.995	136
34	CFpSSR34	AG(8)	AG(6)	AATCTTGTGCCCAATGTAAG	CTTAGCATGAGCAACTCAAAG	55.4	175
35	CFpSSR35	AGAA(5)	AGAA(6)	CTAAACGAAGACCTTCAGGAT	ATTGGTCTTAAACCTGTGGAG	55.53	137
36	CFpSSR36	ACC(7)	ACC(5)	TCCTCCAGTAGTAGGCTCTTC	CTGCCAAGGAGAAGTATAATG	54.835	173
37	CFpSSR37	GT(5)	GT(6)	GTTTCCATTTTCCAAGAGC	GAAGCCATTTGCATCATC	54.84	144
38	CFpSSR38	AT(6)	AT(7)	GGATTGAAAGGAGAACTCAAC	TCCAAACTGCAGCATCTATAC	55.505	146
39	CFpSSR39	GA(7)	GA(8)	CCTGGTTTATAGCAATCGAA	GCACATCAACTTTTGTACCTC	55.195	150
40	CFpSSR40	AAC(5)	AAC(6)	GTCCAAGAAAACGGAGATAAC	ATCTCTAAAACTCCTCGTGGT	55.055	137

41	CFpSSR41	AT(6)	AT(7)	TTTATACTGTGTAGGCGCTCA	TACATTTGGTCAATCGGTCA	56.825	152
42	CFpSSR42	AC(6)	AC(7)	AAAGTCCTCACACCTCAAACT	ATTCTTGTCACTGACTCTGCT	54.775	147
43	CFpSSR43	TAG(6)	TAG(5)	CAGTAGTGAAAAGGAGTGCTG	CCTTCGTTCGTTCAAATCTA	55.425	143
44	CFpSSR44	TGCT(3)	TGCT(4)	AAGAAGAAGAAGACCAGTGCT	TCCTTCAGTAACACAACAAGC	55.275	145
45	CFpSSR45	CT(8)	CT(5)	ACACTCACAAACCACAAGAAC	CTTTCTTGCTCTCTCTCTTCC	55.16	109
46	CFpSSR46	AG(5)	AG(7)	ACACTCACAAACCACAAGAAC	CTCCCATCATTACTTCCATTAC	55.2	130
47	CFpSSR47	TA(9)	TA(8)	CAGAGGTGGATTCAGTATTTG	ACCACTTCAATGTCAACTCAC	54.92	181
48	CFpSSR48	TGG(6)	TGG(7)	TGGTAGATATGGTAGCGACTG	CCACCACTAACCACCTCTG	55.955	138
49	CFpSSR49	GCC(7)	GCC(8)	CCTAAGTCCAAACCCAGTAAT	CTCCACAAAATGGGTGTTAT	54.945	135
50	CFpSSR50	CAT(5)	CAT(7)	TATTCCCAGTTTCCCACGA	GCCGTTGAATTTAGGGTTTACA	59.62	125
51	CFpSSR51	TAG(6)	TAG(5)	CACCCTTGATCAAATTTCTG	TGTACGGCTTGAACCTTATT	55.33	158
52	CFpSSR52	AT(7)	AT(6)	CAAGCTGGTAATCCTCACATA	GGAAGAACCACCTACTCAGTT	55.2	160
53	CFpSSR53	AC(6)	AC(7)	CACTGAAACCACAAGAAGAAC	AGAAGAGCTCACCAAACTAGC	55.26	185
54	CFpSSR54	TC(7)	TC(6)	TGTCATCTAAGTACCGATTCC	ATTCAGCTTGATTCCACAAC	54.755	150
55	CFpSSR55	TTC(5)	TTC(6)	CCTCCTTCCCTCTTATTATTCT	GAGGGGGCTATTTACTATTCA	55.055	116
56	CFpSSR56	TC(7)	TC(6)	AACTACATGAGCTGAATGTCC	TAGAGAAGAGAGACCGAAACC	54.85	149
57	CFpSSR57	AC(10)	AC(9)	ACCTTCCAACAATAACTCTCC	GAAACCTTTACACGATTTGG	54.86	149
58	CFpSSR58	TC(8)	TC(7)	CAGCATCAAGAACACCTAGTC	ACTATCTCTACAGCCAAATGC	54.44	156
59	CFpSSR59	AT(8)	AT(7)	AGTGGAAAGTTGCTTGGTTA	AGCGTAGTGCTTATCAGATTG	54.985	144
60	CFpSSR60	TG(8)	TG(9)	GGGTAAAGATTTATGCTTGC	GTTTGACCAAAAAGAAGTGG	54.52	168
61	CFpSSR61	AGA(5)	AGA(6)	GCTTTTGACTTGAAGGAGAGT	GATGCTTCCTCTGTTTCTTCT	55.37	141
62	CFpSSR62	TG(7)	TG(8)	GGGTTCTCGAATCTGAATACT	CATGATATCTTCCACAACCTC	54.765	150
63	CFpSSR63	AG(6)	AG(7)	GTGCAAATACTAGCATCCATC	GTTGCTTTTCTTTGCTCTGT	54.915	150
64	CFpSSR64	ATTT(5)	ATTT(6)	CAAGTAGTCGTGGACATTGTT	ACAAAGCCAGAACATCAGTAG	54.94	145
65	CFpSSR65	GAA(5)	AGA(7)	CAAGGGGTAATATTGGAAGAG	CCCACAATAGTGTTTATCACC	54.995	159
66	CFpSSR66	ATC(6)	ATC(5)	ACTCTAACTCCGATGCTTTCT	ACACTCCAATGTTGGAACTC	54.94	147
67	CFpSSR67	TC(7)	TC(6)	ATAACCCCAAAAGGCATC	CTTCTTTAGCCCTAGTTTCCA	55.215	148
68	CFpSSR68	AG(8)	AG(6)	GTGTCCAGATTTGCAAGAAT	TCTCCCTCTTCTCTCTCTTTC	54.85	157
69	CFpSSR69	AT(6)	AT(7)	AAAAGTGAAGACACGTGTGAG	CTTGAGTGGCCTGATTCTT	55.405	150
70	CFpSSR70	CTG(6)	CTG(5)	GGAAGTCTCTATTTTCGAACG	GATCCATAGTCTCACAAAATCC	55.51	143
71	CFpSSR71	ATT(7)	ATT(5)	CATAGTCTTCCAGATTGATGG	CACCATTCGTCTTCATTCTAC	54.765	259
72	CFpSSR72	TTC(6)	TTC(5)	GACCATTCTCCTCTTTCGTAT	CTGAGATAAAGCCCTTGAAA	54.88	141
73	CFpSSR73	TA(8)	TA(7)	AAGAGAGGAGGCGGTTAC	GAGTGTTACCGATCAATTCAG	54.58	180
74	CFpSSR74	AG(8)	AG(7)	TCTGAGTAATCCCAGTATTTCC	CCCAAGTTAGTGAGTGGATAA	54.725	161
75	CFpSSR75	TA(7)	TA(10)	GCTGCAGATACGAACATACAT	GGTGTTCCCAATAGTGTATCA	55.19	154
76	CFpSSR76	GGT(6)	GGT(4)	GTTGGGTTTTGTATGGAGAA	AACCATAGTGGTAAGTCCACA	54.82	151
77	CFpSSR77	TGT(5)	TGT(5)	AGAGGTTACCAGAAACAATCC	CCATCTATGGATACAACACATC	54.68	133
78	CFpSSR78	AAC(5)	AAC(4)	AGACATGTTCATCAGCGATAG	TGTGAGTACACTGAATACGTTG	54.645	167
79	CFpSSR79	AG(6)	AG(5)	CAATGAAACAACCCCAAC	GATCCAACCATATTGAACCTAC	54.83	182
80	CFpSSR80	ATG(6)	ATG(5)	GACCTGATATTTCCCTCAGTC	CGAAATCTTTCTCTCATCGT	54.895	133
81	CFpSSR81	GT(5)	GT(6)	ATGATGTGTGTATGTGTGGTG	TTCTTCTCTACCAGTTCAACG	54.84	161
82	CFpSSR82	GT(5)	GT(6)	AGGAGATCATAGCAGTGTGAA	GAAAGCAATACATCCACTCAG	54.96	126
83	CFpSSR83	TC(6)	TC(5)	AGGAGATCATAGCAGTGTGAA	GAAAGCAATACATCCACTCAG	54.96	126
84	CFpSSR84	CA(6)	CA(8)	ATAAATCCCTGTCCCTTTTC	CTCCCATTTTCACTACACCTA	54.59	161
85	CFpSSR85	TA(7)	TA(6)	GGAATATGAATGGACAGGTG	AGTGAATCCAATCCTTCTAGC	55.195	178

87	CFpSSR86	AT(7)	AT(10)	CTGCATCTTTGTAAGCTTCTC	CTAACCCTGTAATGATGATGC	54.695	138
88	CFpSSR87	TC(7)	TC(6)	CAGCCCCAATAGTTAAGAGTA	GCCAATATCAGAGAAAAGTCC	54.735	147
89	CFpSSR88	GAA(5)	GAA(4)	GCCTCTCTCTCTCTACACACTT	CAACGACGTTTTCGTATTCT	55.04	169
90	CFpSSR89	GTG(6)	GTG(5)	TCCTTTAAGTAATGGCCGTA	GTCCAGTCCTTCCACAGTTAT	55.39	149
91	CFpSSR90	GGT(4)	GGT(5)	ATGGTTATAGGAGGAATGGTG	ACAATCCCTAGCTAAATGTCC	55.29	158
92	CFpSSR91	AG(9)	AG(6)	AGTTGCACATAGAGTTGCTGT	TATCATTACTACGGTGGTTC	52.63	127
93	CFpSSR92	AG(8)	AG(7)	GTCAGGCTCCAGGTTATATTT	AAACTGTTGGTAGTGGCAAA	55.43	133
94	CFpSSR93	AC(9)	AC(6)	TGTATTACTGGCAAAGGGAAC	TGATGGGTGCTATGCTATACT	55.89	138
95	CFpSSR94	AGTC(4)	AGTC(3)	CACTTATCATCATCACCTTCC	CTCAAGAAGACAGCTACCAAA	54.745	124
96	CFpSSR95	AGA(5)	AGA(6)	CTTCAGCTATAAAATCCACAGG	GGTTGGTCAAACTACTCATCA	55.48	146
97	CFpSSR96	GA(8)	GA(10)	GGGTAACAAGACAACACACTT	GTGTGGTCTAACTGAAGGTTG	54.49	137
98	CFpSSR97	AT(10)	AT(8)	GTCAAATGAACTGAACAACG	GCATGAGCTTCACTCCTTAAT	55.15	161
99	CFpSSR98	TTC(5)	TTC(7)	GTGTAGTACAAGGTCGAGGTG	GTTGTTATTGTCACCGGAGT	54.895	131
100	CFpSSR99	TA(7)	TA(8)	GGTTGGTGAGTTTTGTTCAT	ACCTTTTGTTACACACACCAC	54.965	168
101	CFpSSR100	GAA(5)	GAA(4)	AATGAACTTCTCGATCCTACC	AAACAACTACTGCACACCACT	54.935	127
102	CFpSSR101	TCA(7)	TCA(6)	AACTCTTGCCACAACATTAG	GTAGACATGTAACGAGGCATT	53.69	144
103	CFpSSR102	AT(7)	AT(6)	AGCTAACCTTTTGCATGTTC	ACCATACATCACTTGTTGTCAC	54.81	156
104	CFpSSR103	GA(7)	GA(6)	CTCTGTTCCAGCAATTTTTC	CTAAATTAATTCCCGCACCAC	56.8	133
105	CFpSSR104	AT(6)	AT(8)	GCACAAACCTAGACAGGATTA	ACTATTTCCGCCTATCAACA	54.755	172
106	CFpSSR105	AC(8)	AC(6)	TTCACATCTCCTAGTCTGGAA	CATGAAATCAGACAAGTAGCC	54.945	131
107	CFpSSR106	AT(8)	AT(7)	TCTTGCAGGAGATAGTGTCAT	TAAGGGACAGACACAAAACAC	55.085	171
108	CFpSSR107	CTG(6)	CTG(7)	AGCTCGATGAGGATGAACTA	GAGGATTCGTTCTCTTGTGA	55.175	151
109	CFpSSR108	GGT(5)	GGT(5)	TGATCACGTACTAGCTCGATT	CCACCAGTAATAGGAGTTTCA	54.8	179
110	CFpSSR109	CAG(6)	CAG(5)	ACCTTCAACCCTTCATTTCT	AAGCCTCTTCCTCTTCTGTT	55.095	152
111	CFpSSR110	GAA(4)	GAA(5)	TAGAGGATTACCACCTGTTGA	CTCAAAAAGCAAGACCAACT	54.785	134
112	CFpSSR111	AC(6)	AC(7)	GTACAGCAGCTACCATCACTT	AAAACAGTCTTTCTACCTCGAC	54.575	148
113	CFpSSR112	GA(7)	GA(6)	ATGGAGGGAGATTATTGAGAG	GCAAGAAACTCACCAGAATAG	54.595	156
114	CFpSSR113	TCT(5)	TCT(6)	CCACCTTCGTCTAATTCATC	CAGCAGGAGATAAATCTATGG	54.455	145
115	CFpSSR114	AG(8)	AG(7)	CACCAAAAGCCTGTTATGTT	GAGAACAAGCTGGTTAAGCTA	54.785	160
116	CFpSSR115	TCA(6)	TCA(5)	TGACTCTGGAAAAGTTTTGG	CACTTCGACCTGAGAAAACTA	55.1	152
117	CFpSSR116	AC(8)	AC(9)	CCAATAGGCAATAATTCCAC	CGCTTCTAACTTCTTCCTTCT	54.795	163
118	CFpSSR117	AT(7)	AT(6)	ACCTTTCTTCCTCACTACACC	CAACCCACTCTTTTCTCTTCT	55.11	187
119	CFpSSR118	CT(6)	CT(7)	TGGCTCAAACTAACCATTTC	GTTCATTGCTGGTGGTAGTT	55.26	149
120	CFpSSR119	GTG(7)	GTG(5)	AACGTCACACATGCATCTACT	CAAGAAGGATGCAAAGAAAG	55.48	152
121	CFpSSR120	AT(6)	AT(5)	CTCTTGTTGGAGTGTGTGTG	ACCTGCTTAAGATGGATTGA	54.925	165
122	CFpSSR121	AAG(6)	AAG(5)	ATAGGCAAGAGGAAATCTCTG	CACAGGTGTTGAAATCAGTCT	55.255	155
123	CFpSSR122	TC(8)	TC(7)	CGTTGTCACCATTACCAAC	AAAACTCACTGGGAACTTGA	54.75	141
124	CFpSSR123	AT(8)	AT(6)	CGCATTCCAAACTAGTGTAAC	GAAGATTCAAAGGCTCCATA	54.815	136
125	CFpSSR124	ATA(7)	ATA(6)	GTAGCACCAGCTTTATATCCA	ACTGGTACAAAGGAAATAGGC	54.915	147
126	CFpSSR125	ATC(6)	ATC(5)	CTCATTAGTGGCAACATTCA	ATGATGGTGATTTAGCGTCT	54.705	176
127	CFpSSR126	GCG(7)	GCG(5)	GAATGTGGTGGATGAATTG	CATCAAACTCCCATCAATCT	54.695	142
128	CFpSSR127	CT(5)	CT(6)	TAGTAAGGGGTCATTTTCCTC	GAATAACCGGAGAGAGAGAGA	55.135	153
129	CFpSSR128	CT(6)	CT(8)	GCCAAGTTACGAAACTGCTA	AACAAAGAACACCCACAGAG	55.5	116
130	CFpSSR129	GAA(6)	GAA(5)	CATCCTCACGTCTCCTTAATA	ATAGCAGTGAGTTTTCGTGTG	55.04	130
131	CFpSSR130	CA(7)	CA(8)	TATCAGTTCGGAGTCACACTT	GGGAGATGAATCAGAGAACTT	54.895	171

132	CFpSSR131	TC(7)	TC(6)	CACAGCTTATTCCTTCATGTC	AACTGGAGCTTTAGCCTTTAG	54.955	151
133	CFpSSR132	CA(7)	CA(6)	GTCATAGCTGTGAAAGTGAATG	TAACTCAATATTGGGCAAGC	55.285	146
134	CFpSSR133	AAC(7)	AAC(5)	AGTCTTGGATCAGAAACTGGT	GGTGTCTAGTTCTTGTTGTGC	55.155	137
135	CFpSSR134	AT(8)	AT(9)	TTGAAGAAGCTAACACTCTGG	CCACAAGAAATTCCTTCTCAC	55.655	161
136	CFpSSR135	AT(5)	AT(5)	GCAGTGGGATTTTACTTTTG	TGAAGAAGAGTCCGTAGTTCA	55.08	152
137	CFpSSR136	CT(9)	CT(7)	TGACACAGGAGAGATAGTTGC	GGTGAGATGGGAATCTAGAAC	55.34	166
138	CFpSSR137	GA(8)	GA(9)	AGTGTGCCACATAAAATGG	GTGTTCTCAGAAGGCAAAAC	54.655	149
139	CFpSSR138	TC(7)	TC(8)	CTTTCACGCAATCTCTAACC	TTCAGCCACCAAACTAAAAG	55.375	150
140	CFpSSR139	TA(6)	TA(7)	TTAGGAATGTCTCACTGGATG	CCTCTCCGATATGAAAATACC	55.295	142
141	CFpSSR140	GCTGCG(4)	GCTGCG(3)	AGAGAGCCACTGGTTAGAGTT	CTTCTCCGTCAACAACACTT	55.115	147
142	CFpSSR141	TC(5)	TC(5)	AGGAGGAGGAGGAGGAGT	GTGAGCTTCTTGGACATGAT	55.16	156
143	CFpSSR142	AGG(6)	AGG(5)	GTTTTAAGAGCAAGGAAGCTC	GCATGTTCATCCTTAGAGAGA	54.92	114
144	CFpSSR143	AGG(6)	AGG(5)	CCACACTGAGAACTATCGTTTA	GAAGAGGCAGAGAAAGACTGT	55.045	152
145	CFpSSR144	TG(7)	TG(6)	CCCAATTGCAATAACTGACT	AGCCCAATTTACAAGGAAGT	55.365	173
146	CFpSSR145	TCA(6)	TCA(4)	GCCTCCAAAATCAGTAGAATC	TGACTATGAAGAAGGTGAGGA	55.23	157
147	CFpSSR146	AC(6)	AC(5)	GACCACCCTTGAAGAGAAAT	GTGTTTTTCCCGTTTTGTTG	56.865	150
148	CFpSSR147	GAT(7)	GAT(6)	CGGTTGAAGAGAAAAATGAG	GTATACGGAGGTTGTGTGATG	55.285	169
149	CFpSSR148	CT(8)	CT(7)	AATCTCTAAACCCCTTCCCTA	GACCAGGTTTAATTGGGATT	55.765	143
150	CFpSSR149	TG(9)	TG(8)	GAACCAGAAACCAATGCTAA	GTATCTTCACGGAGCATGAC	55.52	142
151	CFpSSR150	GT(7)	GT(9)	TGAGAGAAGGGATGAGTATTG	ACCTCTCTATTTCACCCACTC	54.655	151
152	CFpSSR151	GAA(6)	GAA(5)	AGTGTAAGAGATGGGATAAGGA	CTAAGGTCCCATCAATACTCC	54.945	166
153	CFpSSR152	AG(9)	AG(10)	GGTCAGAAAAAGAAGAGGAAG	GGATTCACATGGCTGTTATAC	54.76	145
154	CFpSSR153	TC(5)	TC(9)	GAGAGAGAGAGAGGACCCATA	GAAAGAGGAGAGAGACAGGAA	55.19	115
155	CFpSSR154	ATT(5)	ATT(7)	CTCCGCAATTATTCTTGGA	ATTCACACATTGGAGTTTGG	56.095	159
156	CFpSSR155	AG(6)	AG(7)	CACAACTACAAAACGGGTAGA	GGAAGATCTCCTCAAGATTGT	55.175	145
157	CFpSSR156	AC(8)	AC(7)	ATTGGTGAAGGAGTGACAAG	GCGCCCTTTTAGTTTTACTAC	55.15	167
158	CFpSSR157	CT(10)	CT(6)	GGAGCATGGTAGTCAGTAGAG	ACAAGGGAACTGTACAATCTG	54.2	153
159	CFpSSR158	AT(6)	AT(7)	CATCGTAATCTCTCTCAATGG	GTAGCAAAAGGAATGCTAGTG	54.71	178
160	CFpSSR159	CT(10)	CT(6)	ACCAAGAATAGTGTTGTGGTG	ACAACTGAGAACTGATGATGG	55.095	155
161	CFpSSR160	TA(9)	TA(6)	ACCTGATTACCCATTCAAGTC	ATCATCATGCATGCCTTC	55.38	156
162	CFpSSR161	GTG(4)	GTG(6)	TCTTATCATCCCAATTACCG	CTCATTCTCATTCATCACCTC	54.88	210
163	CFpSSR162	AG(7)	AG(8)	GACAGCAGCATAAGAGCTAAA	ACCACTTTATTGGATGGATG	55.03	131
164	CFpSSR163	TA(7)	TA(6)	GGTTAAATTAGGTCAGGGAGA	GCTACCTTAAGTTGTCTGACG	54.71	131
165	CFpSSR164	TA(9)	TA(8)	AGTTTCTCAAGTGTCTGCTCA	GACGTGAACACAAGAAGAAGA	55.375	183
167	CFpSSR165	AGA(6)	AGA(7)	CCAGAGTCAAAGATCCATACA	TCCAGCAGTCTAATCTTGATG	55.395	157
168	CFpSSR166	ACC(6)	ACC(5)	CAGCTTCTAAACAGGGATGA	TCCTTCACTTACTTCAACCAG	55.08	156
169	CFpSSR167	AT(8)	AT(9)	AGGATCACACACTGAACAAAC	CGCCACTGCTACACTTATATT	54.925	155
170	CFpSSR168	AG(8)	AG(7)	GATCCAAGACTTCTCCACAGT	CAACTTTACCCTCCACTACAA	55.155	138
171	CFpSSR169	TGA(6)	TGA(5)	AATTATATGCCTCGTTGCTC	CCTATGCCTCAACTTTAACTCT	54.63	155
172	CFpSSR170	CT(7)	CT(8)	ATCGCTTTATCTCTCTGCTCT	TAAGAATCACTGCCTCGAAC	55.375	158
173	CFpSSR171	TA(10)	TA(6)	GAGGCCAAAAGATAGATGAAC	GAATATGATGCTCCTCGTACC	55.9	233
174	CFpSSR172	GGCGGA(4)	GGCGGA(3)	GAGCCTTAGAGGATTTGTAGC	GGAGAAAAGACAACGAAGATAC	54.94	171
175	CFpSSR173	TA(9)	TA(8)	AAGGACTTGAGAACCATCACT	TCCTTCAAATCCTGTATGCT	55.12	159
176	CFpSSR174	TC(6)	CT(9)	GTCAACTTCATAAACCCCTCT	TTATCACAAAGAGCTCACACTC	54.835	130
177	CFpSSR175	TA(6)	TA(7)	ACCCTTAAGAAGAATCAGGTC	AATTAGTCGAGGTGCTTCAA	54.64	141

178	CFpSSR176	TC(5)	TC(6)	AAATTCTCTCCCTCTCTATCTC	GATCATGGAAAGACCAAAAC	53.6	149
179	CFpSSR177	TC(7)	TC(8)	TGTTAAAGCTGCTCTCTGTTC	GAGAGATGTTGTGGGGTTTAT	55.335	169
180	CFpSSR178	GAA(7)	GAA(5)	CCCTTATTCTTCCTATTCGTC	TGTGGGTTCTGTTATGTTACC	54.87	156
181	CFpSSR179	CA(6)	CA(5)	CTGATACAGAGTGGAAACACA	TTCTGTTTTCTGCTACTCCAC	54.21	139
182	CFpSSR180	AT(7)	AT(6)	GACTTGATCGTGACATTCTTG	TCACTGACCTCGTGTAACTTC	55.55	149
183	CFpSSR181	AG(6)	AG(10)	AGCTAGGGTTTCCTCTTCATA	CCGCTGTTTACTCTCTTTCTT	55.265	193
184	CFpSSR182	TGA(5)	TGA(6)	GATGAGGGAGATGATGAAGAT	GATTCCTCAACGTTCAACAC	55.545	158
185	CFpSSR183	CA(7)	CA(6)	CAAAGATGAGCAGTAACCTTG	GCCATTAATTATGCACGTCT	55.3	131
186	CFpSSR184	AC(7)	AC(8)	CTCTGGCATAAAACAACAGAG	GAACACCATACCAGAACAAGA	55.18	158
187	CFpSSR185	TC(9)	TC(10)	TGTGTCAGACAGTATGCTCAA	GGATTTCAATGGTGCTTTAC	55.005	140
188	CFpSSR186	CT(7)	CT(8)	TTACTCAGGAGCTTGATGAAC	CCCCTGTATTGATCTTTCTCT	54.83	148
189	CFpSSR187	AT(9)	AT(6)	TCCAAACCTCATTTTCTCC	AGATTCTTTACCACCTCCATC	54.945	114
190	CFpSSR188	TCA(5)	TCA(6)	TCCTCATCGACATTACTACCA	CAGGAAAGGGCTTATATGATG	56.05	152
191	CFpSSR189	AT(7)	AT(6)	CTTCTCTTTCTTCACTCACCA	AGTACTTCCTCCACCCATATT	54.515	130
192	CFpSSR190	TTA(7)	TTA(5)	TGCTTGTCATGAAGAGATTG	AGCGTGTTTACTAGCATCATC	54.775	142
193	CFpSSR191	AC(9)	AC(7)	CACAATACACTCCAATTCTCC	TTCCCTCTCTACTCTCTCGAT	54.74	139
194	CFpSSR192	CAG(7)	CAG(5)	ACTTGGGAGGAACATACTCAT	GGAAGACATCTGAACCATGTA	55.11	141
195	CFpSSR193	TCC(5)	TCC(6)	CAACTTCAGTACCATGGAAGA	CTTCGGAATACGATGTAAGGT	55.62	147
196	CFpSSR194	AT(8)	AT(7)	AAAGCACTTTCTTCTCTCTCC	TTCATTCAACTTCCATAGCC	54.975	150
197	CFpSSR195	CT(8)	CT(7)	TGTCTCTCTTCCACTCTCAAA	GGATTAGTAGACCCGAGTGAG	55.315	167
198	CFpSSR196	AT(7)	AT(8)	ACAAGGAGGAGACTAACAACC	GTCAAGACTTCTCATTTGCAG	55.065	153
199	CFpSSR197	AT(9)	AT(6)	GAAAGAGAAAGTGGTGATGTG	GTAAGATGTACGCAAACTCCA	55.17	144
200	CFpSSR198	TA(6)	TA(7)	GGAGAGAAGAAGTTGCAGAGT	GCTATATGAAGGCGAATGTT	55.005	183
201	CFpSSR199	AT(6)	AT(5)	ACTAGCCCTTGTGACAAAAA	GGTTATCGCAAAATCGAGTA	55.27	150
202	CFpSSR200	TGT(5)	TGT(6)	GGGCTAAGGTCTTTGTACACT	CACCTAGAAACTAAAACCCAAG	55.07	140
203	CFpSSR201	CACCAG(4)	CACCAG(3)	CCACCACTCTCTCTCTCTTT	AAGTAGCTGGTACTGATGCTG	55.035	194
204	CFpSSR202	TTA(6)	TTA(5)	AGTAACTTCCCTTTGCACTTC	TGATACACTTCCAGCTAGGTT	54.73	166
205	CFpSSR203	AG(9)	AG(7)	GAGTTGGATTTTGAACTGTCC	CCAACCTTTCTAATCAACCA	55.715	179
206	CFpSSR204	CAG(4)	CAG(6)	GCTATTGCCAACCATATCAT	ATCAGTGGCCATGTATAGAGA	54.94	161
207	CFpSSR205	AG(8)	AG(6)	GCAAAAGCCAAAAAGGTC	AGCTAACAGTTGTGTTGCAGT	55.58	131
208	CFpSSR206	AT(10)	AT(9)	CACAATTAGCCTTTCAGCTC	CACAAATGGTGATAGGCTGTA	55.755	146
209	CFpSSR207	AG(7)	AG(9)	AATCTCAGCTTCCTTGTTCC	TTAGCTGCTCTTCTCACTCTC	55.15	168
210	CFpSSR208	TCT(6)	TCT(5)	CCAAATTGTCGTCAGTATCA	GTCCTCTGAACGAACAAAAG	54.75	156
211	CFpSSR209	TCT(6)	TCT(5)	CCAAATTGTCGTCAGTATCA	GGATTAGGGCTGTTTTGTTA	54.51	153
212	CFpSSR210	AG(11)	AG(6)	TGACTGTGTGTGTTAGCTGAG	AACACCCAAGATCAACTCTTAC	54.96	207
213	CFpSSR211	GAT(5)	GAT(6)	ACTAATTGGCTGATGAGAGAGT	ATAGCATCGTCTTCCTCCATA	55.62	158
214	CFpSSR212	GCA(8)	GCA(7)	ACTCCTCCATAGCACATTTCT	GCAAATCCATCCTCTAAATG	55.15	172
215	CFpSSR213	TATT(4)	TATT(3)	CTGTGATAAGGAACGTGACAT	CTTTGCAGCTTTTGTCTGTA	54.79	113
216	CFpSSR214	ATC(6)	ATC(5)	AGGATGATACACTACCACTCG	GATAGCAAAGTCGCAACAA	54.625	151
217	CFpSSR215	TTA(5)	TTA(6)	CTCTCTCCACACTTTTCCTTT	GGAAGCTTCGGTTCTTCTTA	55.795	153
218	CFpSSR216	TG(9)	TG(7)	CACGGTGCTCATAAGTAGTTG	AACACTAGTAACAACCGTAGGA	54.71	202
219	CFpSSR217	TG(9)	TG(7)	CACGGTGCTCATAAGTAGTTG	GAATGCTAGCGTGTAACTCAG	55.725	143
220	CFpSSR218	TA(9)	AT(7)	GGTCGCCTTTATTATTAGGA	AAAAGGAGGAGGAGGAAAC	54.34	145
221	CFpSSR219	AC(7)	AC(6)	TACGACAGTTGTGGCTAGAAA	TCAATAATAGTGCCAGAGAGC	55.495	156
222	CFpSSR220	GGA(5)	GGA(6)	TATTGTAGATGGAACCTGCTG	GGCAGCAATTCTAGTCAGTTT	55.86	159

223	CFpSSR221	TCT(6)	TCT(7)	CACACCCAATATCAAACCTAC	CATATGGAATTCAGTGAGGAG	54.635	138
224	CFpSSR222	TTG(5)	TTG(6)	CTTTCTCCTTTTCCTTCTCTG	AAACACAACATCATCAGCAG	54.815	168
225	CFpSSR223	AT(8)	AT(7)	CATTGCAGATGAAGAACCAT	CAAGCCCAATTTAACTATGC	55.61	141
226	CFpSSR224	GTA(5)	GTA(6)	AGGTCTTTCTCTTAGTTTGTCG	ACTTAGAACTCAAAGGCACAC	54.22	158
227	CFpSSR225	AT(6)	AT(7)	CAAAACCCTAACCCTAACTTG	GGACAGATTCGGGTTTTATT	55.62	151
228	CFpSSR226	GGT(5)	GGT(6)	GGTTTGACAGGAGATCATCA	ACGGTCATAGAGTGTGTCTTG	55.58	165
229	CFpSSR227	TCT(6)	TCT(5)	GGCTGGAAACCCTAGTAATAA	CCATAGAAATGTTGAGTCGTC	55.005	144
230	CFpSSR228	AAG(6)	AAG(5)	GATTCCATCTGCTGAATAGAG	TCACACTTTCTACACACTTCG	54.02	149
231	CFpSSR229	AAG(6)	AAG(5)	GCTTTTCTGCTCAGAGACTTA	CTGGTGGAGACATTGTTATTG	54.965	170
232	CFpSSR230	ATG(8)	ATG(7)	GCTATGGTATACCTTCCCATC	TGTGAGAATCTCCACTCTCC	55.18	151
233	CFpSSR231	ATG(8)	ATG(7)	TACCTTCCCATCTTCTCCTAC	CCACTACTCCATTTTCTTCCT	55.03	147
234	CFpSSR232	AT(7)	AT(8)	TCAGTGTATGCTCTTGGATTC	AAATGTAGTTGGTGCAGAGC	55.43	145
235	CFpSSR233	CT(6)	CT(7)	GCTGAACAAGTAGCCATAGAA	CTACCCCTTTCTTCTTTTGTC	54.86	148
236	CFpSSR234	TC(9)	TC(8)	TAATGTTGGTGACTCAAGAGC	CAGCGTCTCTGCGTATATAAT	55.19	148
237	CFpSSR235	TG(8)	TG(7)	AATCTAATGGTCAACCCACA	TTTATTAACCCCACCCCTAC	55.44	167
238	CFpSSR236	TTC(5)	TTC(4)	GGTTTCTGTTTCATCCTCATC	ATCTGTTGAGACAGAAGATGG	55.21	163
239	CFpSSR237	AC(8)	AC(7)	CCGACTCATCTCTTTTCTCTT	CTAGGGAGAAGGTGAGAATTG	55.49	128
240	CFpSSR238	GT(8)	GT(10)	ACAGTGGACATACTGTGGAAG	TTAAGCAGCAAGCAATGTAG	54.815	137
241	CFpSSR239	CTT(5)	TTC(6)	CACTTTCGAGTTCACTAGCAG	GGCAAGAATGAAGAGAAGAA	55.085	152
242	CFpSSR240	TGG(7)	TGG(5)	TAGAGTGGCTTGACAAGAATC	CCCGAGCTGCCAGTAGTA	56.09	187
243	CFpSSR241	TCT(5)	TCT(6)	CATGGGGTTGTGATTTCTA	CTGCACACAATAGACGTATGA	54.52	183
244	CFpSSR242	ATG(9)	ATG(5)	AGAACAATAAGCGCTGAAAC	ATATCCTGATTGACTCGTTCC	55.275	139
245	CFpSSR243	AG(7)	AG(6)	GTGGGCACCAACCATTAT	ATCCAACTTACACTCATACGC	55.485	199
246	CFpSSR244	GTT(5)	GTT(6)	GTTGGAGTAGGTTTGTTCTTG	GATAATGCTGCCAACTCATAC	54.485	156
247	CFpSSR245	AT(8)	AT(7)	GAGGAAAAGATAGGGAGGAC	CAGCCTAAGTTCCCATTACTT	54.85	170
248	CFpSSR246	TCT(4)	TCT(5)	CGTGACATTACTCATTCCTCT	CCCTAACCCTAGAAATTGAGA	54.775	103
249	CFpSSR247	AT(6)	AT(7)	GTGGCAACAAATATACAGAAGC	GCCATTGCTGACAGAGTAAT	55.97	157
250	CFpSSR248	GT(8)	GT(7)	GGTTTCTCTCCTTTTCAGTTG	AACAACAGCTTCCCTTCATA	55.315	151
251	CFpSSR249	CA(6)	CA(7)	AACTTCCCCCTAAAACAAAG	CTCTACCGTTATTGTTTGTGC	55.095	202
252	CFpSSR250	AT(10)	AT(8)	GGGTTTGGTTTTTATGTTCC	GGGTTCATCCTATGAAAACA	55.47	157
253	CFpSSR251	TA(6)	TA(7)	TAGAACCCATCAACCTTCAC	CGGACGAACAAGTAAGTTTT	54.84	145
254	CFpSSR252	GTA(5)	GTA(6)	CTCCCTTGTTTATCCACTTTC	ACTGATGCTTCCTCCTTGT	55.09	160
255	CFpSSR253	GT(6)	GT(7)	ATTGGATAATGCTGATAGGG	TTTACCGTGTGTGATCTTTG	54.405	128
256	CFpSSR254	AT(8)	AT(5)	GGAGATCCACTATGGTCTTTT	GAACTCAGGTAAAAGCTAGCAG	55.105	143
257	CFpSSR255	CTCCGT(6)	CTCCGT(4)	GACGTAATGGGTGCTTCTT	GAGAGGTATGCAGTGGTTATG	55.03	161
258	CFpSSR256	TA(8)	TA(7)	CACCTTTAATGAGATGGTGAC	AAAGTCGCTAGGAAGAGAAAC	54.68	219
259	CFpSSR257	ATAC(6)	ATAC(5)	TGTGTCAGTACAACTGGGATT	TATCACCACGAGTTTAGTTGC	55.51	148
260	CFpSSR258	AG(6)	AG(7)	AAAAACGTCTGCGTACAGAT	ATCTCTGTCTCACTACCGTGT	54.425	155
261	CFpSSR259	AT(9)	AT(6)	GAATACAAAAGGGAAGAGTGC	AAGGAGAGGTAGCAATTGTGT	55.595	160
262	CFpSSR260	AT(7)	AT(6)	CTACAACAACTGGAAGACCAA	CTTAGCTAATACGCCACTTGA	55.255	142
263	CFpSSR261	TAG(6)	TAG(5)	TGGTATATATATCGCGTGTCC	CGGGTTTTCAGGATAACAA	55.645	157
264	CFpSSR262	AGC(5)	AGC(6)	GGGGTAGAGGAGGTAGAAAAT	CCATTCTTCTCAACAACTGAC	55.165	148
265	CFpSSR263	TA(7)	TA(8)	CTACCGGAGGAAGAGAGATT	ACCTGGAATAGAGCAAAAGTC	55.17	155
266	CFpSSR264	AC(7)	AC(6)	ACATAATTTCAGGCTCTACCG	CCTGATTAGCACAGCATAAAC	55.63	148
267	CFpSSR265	CAA(5)	CAA(6)	GTAAGCAGGTATCTGATGGAA	CTCTCTCTACCGCTGATGTC	54.79	153

268	CFpSSR266	ATA(5)	ATA(6)	CTGGATAGATAGTGCTGGATG	GGCTGGAGTTTAACAAAATG	54.955	144
269	CFpSSR267	TGT(5)	TGT(6)	ACCTCTAAGGTAGTGGTACGG	CTGCAAACCCCTTACATAAA	55.245	126
270	CFpSSR268	AG(6)	AG(9)	CATATCGTCATTGCTGCTAA	CATGTCCAGGTCGATAAACTA	55.14	235
271	CFpSSR269	TA(8)	TA(9)	GGAGTTGTTCGATGTTGATAG	GCAAATTGAGTTTCTCCTCA	55.15	156
272	CFpSSR270	AT(6)	AT(7)	ACGAGATAACACATGTTCGAC	GAGACCTCAAGACCAAAACA	55.2	161
273	CFpSSR271	TC(8)	TC(7)	GAGCGAAATGTGAATCAAAG	GCAAGAAGATATCCATCAGC	55.45	141
274	CFpSSR272	TA(7)	TA(8)	ATCACAGATGTAGCGTTCTTG	CTTGAGTTGGTGACAATGAGT	55.35	169
275	CFpSSR273	AG(6)	AG(7)	CAATGAATGAAGGAGACAGAG	CGAGCTAGACAACCAAACTAC	54.665	160
276	CFpSSR274	CA(8)	CA(6)	CCATTTCTACCAAGACAAAC	TCTTCTTCTTGTTGTCGTACTG	53.565	129
277	CFpSSR275	AC(8)	AC(7)	GCCCTACTTCCCATAAAAGA	ACCAGTGGTCATTATGTCATC	55.345	137
278	CFpSSR276	TC(9)	TC(8)	GATGACGATAAACAAGTGTGC	TTAGGCAGCCTGTAAGATATG	55.32	142
279	CFpSSR277	GCA(5)	GCA(6)	ACACCATAGATCCACACCATA	TGAACTTCACTGGTGAGAAAC	55.28	143
280	CFpSSR278	TA(6)	TA(8)	CTCTTAGCGGAAGCTTTATG	TCATCTCCATCATCCTCATAC	54.81	185
281	CFpSSR279	AG(10)	AG(6)	CATCATGTTTGCAGGAGAA	GTTTTGCATAGGAGGGATACT	55.29	146
282	CFpSSR280	TC(8)	TC(6)	CTACCACCGGACAACAAC	GTAGCTTGGTCCGTCAGAT	54.915	137
283	CFpSSR281	GAG(6)	GAG(5)	AAGTGACTGCATCCAACAGT	CCGTACTGTACTCGTCAAATC	55.295	170
284	CFpSSR282	CT(9)	CT(8)	TCATTTGATTCTGTGGGTCT	GAGGCAGATATCCTTTCATTC	55.465	153
285	CFpSSR283	AT(7)	AT(6)	TAAACTTCCGCTGGATACAC	CCTTTAGGAGAAGTGTCAGGT	55.275	171
286	CFpSSR284	TTC(6)	TTC(7)	GACAATGGCACAGAGAAAAT	TTCGACAGCTTTACCAGAAT	55.155	145
287	CFpSSR285	GA(9)	GA(10)	TCGGAGACAAAGAATGAGAC	TTCCTCTCTTAACTCCCTCAT	55.02	151
288	CFpSSR286	CTT(5)	CTT(6)	CAACGAAATTACTCCTTCTC	GTGCTCTGATACCATGTCATT	53.375	113
289	CFpSSR287	AC(7)	AC(8)	AGCAACTCATCCACCAATAA	TCAGCAACTCTTGGACTTTT	55.415	203
290	CFpSSR288	TG(7)	TG(10)	GCGAGGACATTTCTATTGAG	CCTGGAGACACTAACAAATGA	55.175	157
291	CFpSSR289	AT(6)	AT(7)	CCAGCACGGATCTTAAAATAG	GTTGATATTGCCCATTGAGT	55.805	147
292	CFpSSR290	AG(6)	AG(5)	TAAGAACGGAGATTGAGAGTG	ATCTTTTTCCCTTCACCTTC	54.81	162
293	CFpSSR291	AT(7)	AT(5)	ACCCTGATCTTCCATAGTTGT	TACCGACTGTGCTTATTTACC	54.905	162
294	CFpSSR292	AT(10)	AT(5)	GGTAAATAAGCACAGTCGGTA	CAAGAAGGCAGATGCTTAAT	54.725	232
295	CFpSSR293	GAA(6)	GAA(5)	GTGTATCTGGTGCTGGTACAT	CGAATCAAAGCATCTTCTTC	55.095	151
296	CFpSSR294	AG(6)	GA(7)	CCAACAAATCAACTTGTCCT	GTTCACTTGCCCTATAAAAGC	55.46	154
297	CFpSSR295	GAA(5)	GAA(6)	TCTGGAGCAAAATAAAGAGG	AGTCACTCTACCCCACCAC	54.755	146
298	CFpSSR296	AG(8)	AG(6)	CCCGTGAAGAGGATTTAATA	CTTGTTTGGGGTCTCTACAAT	55.135	138
299	CFpSSR297	AC(7)	AC(6)	TCATGTTACAGTGTCTGTGGA	CAATGCCTATTGTTGTACCTG	55.365	149
300	CFpSSR298	AT(10)	AT(8)	CTATGTCTAGAGGGAGGTTGC	ATACAGGAGCAATGCCTTC	55.48	192
301	CFpSSR299	ACA(5)	ACA(4)	GAAAGGCCTTCTTCTCATAAC	AGTCTTGATGTCCTTTCTCCT	54.735	126
302	CFpSSR300	TA(7)	TA(6)	ATCCTAGCTATGTTCCGAATG	GGATCTAGAGCGATTCAATG	55.48	143
303	CFpSSR301	AT(6)	AT(7)	GCTGATTACTGAAAGCAGAAG	CTCACAGATCATTTCCGTAAC	54.705	163
304	CFpSSR302	TA(10)	TA(7)	GTGGTATACTCCAGCAGATGA	CTTGAGAAATGACGGATGAT	55.185	149
305	CFpSSR303	TTG(5)	TTG(6)	TAGGCATGTCAAGACTCATTC	GTTTTTGAGGCTGCAAAGTA	55.58	136
306	CFpSSR304	AGT(5)	AGT(6)	GCCTCAAAAACATCAATCTC	CCATTACTGGAATACATCCTG	54.725	106
307	CFpSSR305	AC(8)	AC(7)	ACCACCAATCTACAATCACAG	GTTCTACCTAATGCACATGCT	54.8	173
308	CFpSSR306	TAT(6)	TAT(5)	TAGCTCTGGTGTTTTGCTAAC	ACCGCCAAGACTTTACTCTAT	54.925	154
309	CFpSSR307	AGG(6)	AGG(5)	ATATGGTGGAGGACGATACAT	CTACCTCTCTTTTCTCCCTTG	55.395	180
310	CFpSSR308	TA(9)	TA(8)	ACACCTAGCCCATTTGTATCT	CTTTGCTTTCCACTTTCTCA	55.565	191
311	CFpSSR309	AT(8)	AT(5)	GCACTAGCATGTAACTCAAGG	AAGCCTTATTCCCTATTTCG	55.26	132
312	CFpSSR310	TC(9)	TC(10)	CACGATTAGATTTCGTGCAT	CACCTTGGGATCCTGTATATT	55.92	160

313	CFpSSR311	TG(7)	TG(8)	CATACTCTGGAAGTTTGGATG	CCAAATTCCACTCCACATAA	55.38	170
314	CFpSSR312	CT(7)	CT(6)	GTGTGGTCGGTTCGTATAAT	CACCGAAAAATCCTAAATGC	56.105	184
315	CFpSSR313	TG(9)	TG(6)	CGGAACAAGGAAAATGTAAG	AGATAACACCCTGATTGATCC	55.25	159
316	CFpSSR314	TA(10)	TA(7)	CTTGATCCGGATGATTTTAC	GTTTTGATGGTCGTGTGTG	55.165	152
317	CFpSSR315	AGT(5)	AGT(4)	GATTAGCTTTGTAACCACCAC	ATTGCAAGTATGTCCACTCAC	54.59	160
318	CFpSSR316	TGT(5)	TGT(6)	AAGGATGGGTTAGTTTTCCT	AGAACCACTTGTTCCTCAGAT	54.9	169
319	CFpSSR317	CA(8)	CA(6)	TGAGTGAAAAGCTACAAGTCTG	ATAGGGGTATCCTCTTCCTCT	55.03	158
320	CFpSSR318	AT(7)	AT(8)	GCTTATTGCTTTGTCTAGCTG	CCTACTAACCACCTTCTACCC	54.7	142
321	CFpSSR319	AC(7)	AC(6)	CAAGATAATACTAGAGCCACAGG	GGATTAGGGTAGAGGGTTAGG	55.5	156
322	CFpSSR320	ATC(7)	ATC(6)	GCTTGAAATCATGCTCACTA	GATGATGAAGTTGAGAAGGTG	54.385	145
323	CFpSSR321	GT(6)	GT(7)	CGACCCAGATATACAAACAGT	TCCAATTCTCCTCTCTTCTTC	54.755	152
324	CFpSSR322	AT(6)	AT(8)	CTTTCAACTTTCCTGCCTAA	CAACTACCTCAACATTGCAT	54.255	154
325	CFpSSR323	GA(6)	GA(7)	CGATTAGTGGCATATATCCTG	ATCACCCATCCTCTTGTCTC	55.695	148
326	CFpSSR324	TGA(5)	TGA(6)	TGACATCCAGGCTCACTATTA	CTTCTGAAGTTGTTTGTCCAC	55.39	145
327	CFpSSR325	AT(6)	AT(8)	TCTCAAGTCTCAACTTCCTCA	CTCCTGTGAAATGAGATCAAC	54.935	143
328	CFpSSR326	TGG(7)	TGG(6)	CTGTAGAACTCCAAACCCTCT	ATTTCGCTTAGCTTGTCGT	55.22	151
329	CFpSSR327	AT(6)	AT(7)	CGATATAAGGTTGTGCCTATC	GACTGAGTGTCTTCGATTCAT	54.195	169
330	CFpSSR328	AT(8)	AT(7)	CCATGAATGTAACTGCAGAA	GAGATGTGCTAAAACCTGTTG	54.865	146
331	CFpSSR329	TGA(6)	TGA(5)	GGAACAACAGGTTGAAATGT	CTTGTCTTTAGCCAATTTCG	55.125	162
332	CFpSSR330	CTT(4)	CTT(5)	CAACGGTGGCTAACATACTAA	CTATTTCTTCTCCTTCCGAGT	55.02	204
333	CFpSSR331	CGG(5)	CGG(4)	CGGCATAATTAAAGGAGATG	CTCTTTCTCTCTCCCTTTTTC	54.605	159
334	CFpSSR332	AG(10)	AG(7)	TAACTAGTTGGCTGTTCTTGG	AGTTCAGATCCCACAGCTTAC	55.41	166
335	CFpSSR333	GC(7)	GC(9)	CCACTTCCTTTTATCCTGAAC	TACCGCTGTTGTTTCTCTTT	55.34	146
336	CFpSSR334	TA(6)	TA(8)	CCAAGAAAATGAAGGGTGT	ATTTCCCAACATGGATCAC	55.255	165
337	CFpSSR335	TA(8)	TA(6)	GTGAATACCGTCCCTAACAAT	ATCCGGGTAGTAAAAGTGAAC	55.29	126

Annexure-III

List of the primer for genes used in qRT-PCR analysis of F_1 hybrid and parents

F1 Cross name	Primer Name	Forward primer	Reverse primer	F1 Cross name		Forward primer	Reverse primer
C.annuum x *C.chinense*	DL1	CACGGACTCAGTCATCTCATTC	GGTTGCCATCTTCCGGTTAT	*C.chinense* X *C. frutescens*	ChiFru7	AAGGGAACTGGAACTGCTAAC	CGTGAATGATGTGGGCTCTAA
	DL2	CCACCATCATATCCAGAAGGAC	TTCCACACCAACCAGCTATAC		ChiFru8	GCGAAGTTACGGAATCCAAGA	GGCCGTTCATCCCTTCAATA
	DL3	ATAGGAGGTGAAGCTTGTGATG	CTCTGTCCACACTCTCTGAAAC		ChiFru9	CTCCACCACCTACTCCTTACTA	GTGCTTCTTGTCATGGGATTTC
	DL4	CCTAGGGTTGTCATGGTTAAGG	CACCGACGTTGTTGATCTCT		ChiFru10	GTGGAAGGTGGTCCAATGAA	GCAAGAGCATCACCTTCGATA
	DL5	AGCAGCAAAGCAGCTCTAA	GACTCAGCCCACAGCTTATT		ChiFru11	CACACCTGACGGACCAATAG	CCCTGCACGAATAAGGAAGAT
	DL6	CTCATGGATTCCTTGTGGTAGG	CGTCCATCTCCAAGACAGATAC		ChiFru12	TGGCATTGAAGGGAGGTTAC	CGAGCTTCCAGTGCAGTATAG
	DL7	TTTAGGCTGCCGGAGAATAC	CCTCAGGTTTCTTCTCCTCTTC	*C.chinense* X *C.chinense*	LC1	GATTATGCTGGACCGACTCTTC	CTTAGTGCATAGGCAGTTCCTT
	DL8	TGCGATGCCCAACATTCT	GACTGATGGTGTCTTCCTCTTG		LC2	GGCGAGTGCAGACACTATAC	CCAGTTGCCTACAGGATCATAG
	DL9	GGTGTTGAATGAGGTCCAAGAG	CTGGCTTGTAGGCGATGAAA		LC3	CACACCTTTCCAGCCTTACA	GCACCAAGAGTAGCCAACATA
	DL10	GAGAAGCACAAGGCAAAGAAAG	GGTGCTCATGGAATGCAAATC		LC4	TACCTGCCTTCAGATTCAGTTC	GTTAGCTTTGGCTCGCTTTG
	DL11	CGTAAGGCATCGAAGAAAGTAGA	TAACCTTCCACCAACCACAC		LC5	CTTGCCCTCGGTATTGTGAT	CCGTCAGCGCTGTAATAGTT
	DL12	GAAGTCCAGGAGGCAAAGAA	TGTAGGCGATGAAGCTGATG		LC6	TGGCAGCACCTGGTTATTT	TACGTGCGCAGTCAAAGATTA
C.chinense X *C. frutescens*	ChiFru1	ATAGGAGGTGAAGCTTGTGATG	CTCTGTCCACACTCTCTGAAAC		LC7	GAACAGGTAAAGAGGGCTACAC	ATTTCCATCCATCCATCACTTCT
	ChiFru2	CGAAGCAAGTGCAGAACAAC	GTCAGACTCTCCCTCACAAATG		LC8	TCAATGCCAAGGGAGGAATC	CAGCCAGTCTGGGAGAAATAAG
	ChiFru3	GAGTGTGTTTCACCCAGGAA	CCACTGTACAGCCAGTCATAAA		LC9	CTAGGACTTGGGTCTCTCTCTT	CATGAGGAAGTGGTATCTCTTTAGG
	ChiFru4	GAAGACTCGAGTGGTGAAGAAG	TGGAAGGATTGGTTCGGTAATAG		LC10	GAGTGTCCTGGTGTTGTTTCT	CTGCTCTTCCTTCCATCTCTTC
	ChiFru5	TGCATGCCGGGATTACTTAC	TATCCGCATCAGCAGTTTCC		LC11	GGATGCTGACAAGGACAACTA	ACCATTCCACAGGGTGTTATC
	ChiFru6	CCACCATCATATCCAGAAGGAC	TTCCACACCAACCAGCTATAC		LC12	GGCTAGATCGCGTGAAGAAT	CCTTCTGTCTCCACCAACAA

Annexure IV

List of the step loop and forward primer for qRT-PCR of miRNAs

miRNA ID	Stem loop-primer	Forward primer
Cfr-NovmiR0027	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAAGGCC	CGGCTTGGGTCGAAGTTTTGGA
Cfr-NovmiR0089	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACACATAA	GCCGGGTCCTCTCTTTGTCTTG
Cch-NovmiR0140	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGGGAT	GCGCGGAAGTCCTCGTGTTAC
Cch-NovmiR0199	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGGGCC	CGGTGAAAAGGACATGGGCC
Cfr-NovmiR0020	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGGCAC	CGGCCGAAACGCTCGGAATT
Cfr-NovmiR0025	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAGTTG	GCACACTCCAGGCAGACG
Cch-NovmiR0315	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCCAAG	CGGCTCTGGGATCCCCTTTG
Cch-NovmiR0164	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACCAGGTA	CGCCCCGGAGTCATAACCAG
Cfr-NovmiR0075	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGGCAG	GGCCATTTGTGGGATACCGGG
Cfr-NovmiR0005	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGCGGT	GGCTCCTGTTGCAGCTGCTA
Cch-NovmiR0166	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGAACA	GCTCCGGACGCTGGCCTG
Cch-NovmiR0316	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGGTGG	CGGCGGATCAGACTTGTAGGGATA
Cch-NovmiR0069	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGGGTAAAA	GCCCCGAAACGCTCGGAATT
Cch-NovmiR0213	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGTGGCAGC	CGGCCTGGGATCCTCTTTGT
Cfr-NovmiR0110	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCCAAG	GGCCTCTGGGATCCCCTTTG
Cfr-NovmiR0153	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACGGGCGT	CGGGCTAATGATATGCCTGGGC
Cfr-NovmiR0146	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACACAATG	CGCCCGCCATGATAGATGC
Cch-NovmiR0251	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTGGGGG	CGGCGGAAATCTTGCCGATTCC
Cch-NovmiR0494	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTCCACC	CGGCGTTGTTGTCACTACTGCT
Cch-NovmiR0210	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAGCCTA	CGGGCTTAGGAGACCCCATTGA
Cfr-NovmiR0126	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACTAATTC	CGGCGCTCTCGTACTAGGTT
U6	GTGCAGGGTCCGAGGTTTGGACCATTTCTCGAT	GGAACGATACAGAGAAGATTAG
Universal reverse primer	GTGCAGGGTCCGAGGT	

Annexure –V & VI are separately given in CD Drive

Annexure VII-XII

Annexure 7. Estimation of percent heterosis over mid parent and better parent for various metabolitesin F_1 hybrid of *C. annuum* (Dudu) x *C. chinense* (Lota bhut).

Metabolites	***C. annuum* Acc. Dudu**	**F_1**	***C. chinense* Acc. lota**	**MP**	**BP**
Aminoacids					
L Proline	1330107	3170659	1281494	142.8	147.4
L Valine	428263	549366	442255	26.2	24.2
L Serine	3022094	1522459	744648	-19.2	-49.6
L Aspartic acid	3705844	4353911	693382	97.9	17.5
L Norvaline	9379047	38206403	21523465	147.3	77.5
Sugar & its derivatives					
Melibiose	1884006	2538539	9977864	-57.2	-74.6
D Turanose	7563620	11762926	5703727	77.3	55.5
D Galactose	6887739	14707761	7820022	100.0	88.1
D Glucose	11584383	5861286	17750229	-60.0	-67.0
D Fructose	8300838	14874435	15199975	26.6	-2.1
Maltose	2503157	15182551	11391738	118.5	33.3
D Ribose	411393	1906452	98733	647.4	363.4
Mannose	643080	3377045	6480153	-5.2	-47.9
2-Keto-l-gluconic acid	810982	3464986	139800	628.9	327.3
D Glucitol	1339612	4769032	7677823	5.8	-37.9
Xylonic acid	475902	1364290	352249	229.5	186.7
Carboxylic acid					
Octanoic acid	8738138	7680062	9654241	-16.5	-20.4
Nonanoic acid	3336311	2594926	4151547	-30.7	-37.5
Undecanoic acid	657159	867403	2449083	-44.2	-64.6
Pyrrolidinedione	4030351	4359884	6021505	-13.3	-27.6
Fatty acid					
9,12-Octadecadiynoic acid	2441236	1374959	1622139	-32.3	-43.7
Hexadecanoic acid	54458707	61250914	61538496	5.6	-0.5
cis-9-Hexadecenoic acid	2156092	2132462	2439770	-7.2	-12.6
Octadecanoic acid	14633826	23607761	20576394	34.1	14.7

Annexure8. Estimation of percent heterosis over mid parent and better parent for various metabolitesin F_1 hybrid of *C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4).

Metabolites	***C. chinense* (Acc. 7)**	**F_1**	***C. frutescens* (Acc. 4)**	**MP**	**BP**
Amino acids					
L Alanine	420144	513792	134569	85.2	22.3
L Proline	0	0	1680027	-100.0	-100.0
L Serine	542034	222577	641694	-62.4	-65.3
L Aspartic acid	1478664	158078	1206751	-88.2	-89.3
L Threonine	486815	3278857	659488	472.1	397.2
L Norvaline	14371292	444726	0	-93.8	-96.9
L Glutamic acid	3511968	6062505	0	245.2	72.6
Sugar & its derivates					
D Turanose	236564	1189893	0	906.0	403.0
Xylose	159972	106906	0	33.7	-33.2
4-Keto Glucose	547945	2750898	84736	769.6	402.0
2-Keto-glutaric acid	503916	1672911	433877	256.8	232.0
1,2,3,4,6-Pentakis (pentagalloyl glucose)	75041	5753053	1195551	805.6	381.2
α ,D-Glucopyranoside	19506845	36210597	13754201	117.7	85.6
beta.-Hydroxypyruvic acid	143847	118392	581642	-67.4	-79.6
Erythro-Pentonic acid	0	274531	127059	332.1	116.1
D-Tagatofuranose	20093	2889114	0	28657.4	
Methylmaleic acid	90614	134760	158545	8.2	-15.0
Indole-3-carboxaldehyde	4331175	33937667	8118317	445.2	318.0
Carboxylic acid					
Ethanedioic acid	16296090	17772417	12967662	21.5	9.1
Propanoic acid	2090409	5027408	1404507	187.7	140.5
Propanedioic acid	2798473	4153517	2735601	50.1	48.4
Propane tricarboxylic acid	1261264	17593804	6845805	334.0	157.0
Butanedioic acid	5190217	17001471	9528159	131.0	78.4
Pentenoic acid	54339	708612	125366	688.6	465.2
Cyclohexane carboxylic acid	102160	42792722	386004	17432.1	10986.1
Fatty acid					
alpha.-Linolenic acid	973146	6098772	1674016	360.8	264.3
Hexadecanoic acid	7622347	7108675	9142708	-15.2	-22.2
trans-9-Octadecenoic acid	5378686	18068094	5747022	224.8	214.4
Octadecanoic acid	3959668	10072151	2810526	197.5	154.4

Annexure 9. Estimation of percent heterosis over mid parent and better parent for various metabolitesin F_1 hybrid of *C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia).

Metabolites	***C. chinense* (Lota bhut)**	**F_1**	***C. chinense* (Chocolate jolokia)**	**MP**	**BP**
Aminoacids					
L Serine	3022094	2408282	1984258	-3.8	-20.3
L-Threonine	3184120	568509	571134	-69.7	-82.1
L-Aspartic acid	3705844	5988023	1304107	139.0	61.6
Sugar & its derivatives					
D-Mannose	643080	151514	704512	-77.5	-78.5
Melibiose	1884006	708764	363763	-36.9	-62.4
4-Keto-glucose	5371015	2264427	1232328	-31.4	-57.8
Myo-Inositol	4812918	76686	1103653	-97.4	-98.4
alpha-D-Glucopyranoside	27613252	25068074	22421047	0.2	-9.2
Hexopyranuronate	5950555	105563	529175	-96.7	-98.2
Carboxylic acid					
Butanedioic acid	3016908	198405	12135002	-97.4	-98.4
Propanoic acid	4661573	20547526	1297757	589.6	340.8
1,2,3-Propanetricarboxylic acid	3506057	7461549	3273091	120.1	112.8
2-Pyrrolidone-5-carboxylic acid	5647473	1188850	1545876	-66.9	-78.9
1-Cyclohexene-1-carboxylic acid	212396	412863	543883	9.2	-24.1
Ethanedioic acid	6235916	11929097	7392342	75.1	61.4
Propanedioic acid	2725075	2247704	4156250	-34.7	-45.9
Oxalic acid	63089477	51204	128167	-99.8	-99.9
Fatty acids					
Octadecanoic acid	14633826	3895691	6533057	-63.2	-73.4
Decanoic acid	3224494	214601	1444141	-90.8	-93.3
Hexadecanoic acid	54458707	8015427	8351893	-74.5	-85.3
Tetradecanoic acid	614919	4047484	3989921	75.8	1.4

Annexure 10. Estimation of percent heterosis over mid parent and better parent for various metabolitesin F_1 hybrid of *C. chinense* (Chocolate Jolokia) x *C. frutescens* (Acc. 4).

Metabolites	***C. chinense* (Chocolate jolokia)**	**F1**	***C. frutescens* (Acc. 4)**	**MP**	**BP**
Aminoacids					
L Alanine	225425	240742	134569	33.7	6.8
L Serine	1984258	383873	641694	-70.8	-80.7
L-Threonine	571134	629281	659488	2.3	-4.6
Sugar and its derivatives					
2-Ketoglutaric acid	563339	906167	433877	81.7	60.9
4-Keto-glucose	1232328	1113499	84736	69.1	-9.6
Malonic acid	396263	206075	238947	-35.1	-48.0
alpha-D-Glucopyranoside	22421047	25853175	13754201	42.9	0.0
Indole-3-carboxaldehyde	10966754	18461509	8118317	93.5	68.3
Carboxylic acid					
Butanedioic acid	12135002	10388578	9528159	-4.1	-14.4
Propanedioic acid	4156250	6633287	2735601	92.5	59.6
Propanoic acid	1297757	3406166	1404507	152.1	142.5
1,2,3-Propanetricarboxylic acid	3273091	8889095	6845805	75.7	29.8
2-Pyrrolidone-5-carboxylic acid	1545876	1385653	1763763	-16.3	-21.4
Ethanedioic acid	7392342	21591506	12967662	112.1	66.5
Fatty acids					
alpha-Linolenic acid	8497566	4868057	1674016	-4.3	-42.7
9,12-Octadecadienoic acid	2755876	1370929	1016032	-27.3	-50.3
Octadecanoic acid	6533057	12216294	2810526	161.5	87.0
Tetracosanoic acid	80829	45720	31210	-18.4	-43.4
Hexadecanoic acid	8351893	8978730	9142708	2.6	-1.8
Tetradecanoic acid	3989921	4569699	5975184	-8.3	-23.5

Annexure 11. Estimation of percent heterosis over mid parent and better parent for various metabolitesin F_1 hybrid of *C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4).

Metabolites	***C. chinense* (Lota bhut)**	**F1**	***C. frutescens* (Acc. 4)**	**MP**	**BP**
Amino acids					
L Serine	3022094	723048	641694	-60.5	-76.1
L-Threonine	3184120	4608806	659488	139.8	44.7
L-Aspartic acid	3705844	526681	1206751	-78.6	-85.8
L Glycine	9379047	185557	20485	-96.1	-98.0
Sugar derivatives					
4-Keto-glucose	5371015	1614740	84736	-40.8	-69.9
Myo-Inositol	4812918	298256	1179692	-90.0	-93.8
alpha-D-Glucopyranoside	27613252	39015383	13754201	88.6	41.3
Carboxylic acid					
Butanedioic acid	3016908	16652968	9528159	165.5	74.8
Propanoic acid	4661573	3299228	1404507	8.8	-29.2
1,2,3-Propanetricarboxylic acid	3506057	13284193	6845805	156.7	94.0
1-Cyclohexene-1-carboxylic acid	212396	571630	386004	91.1	48.1
Ethanedioic acid	6235916	34545747	12967662	259.8	166.4
Propanedioic acid	2725075	6116381	2735601	124.0	123.6
Fatty acids					
Hexadecanoic acid	54458707	7867556	9142708	-75.3	-85.6
Tetradecanoic acid	614919	5255489	5975184	59.5	-12.0

Annexure 12. Estimation of percent heterosis over mid parent and better parent for various metabolitesin F_1 hybrid of *C. chinense* (Umorok) x *C. frutescens* (Acc. 4).

Metabolites	***C. chinense* (Umorok)**	**F1**	***C. frutescens* (Acc. 4)**	**MP**	**BP**
Aminoacids					
L Alanine	442197	540065	134569	87.3	22.1
L Proline	209147	1819135	1680027	92.6	8.3
L Serine	438903	544445	641694	0.8	-15.2
L-Threonine	378262	333065	659488	-35.8	-49.5
L-Aspartic acid	2945521	1085641	1206751	-47.7	-63.1
L Glutamine	4650697	3678855	2408232	4.2	-20.9
L Glycine	224646	111691	20485	-8.9	-50.3
L Isoleucine	58715	41783	56647	-27.6	-28.8
Sugar & its derivatives					
2-Ketoglutaric acid	695210	1031693	433877	82.7	48.4
4-Keto-glucose	926105	886642	84736	75.4	-4.3
Glucoson	100765	280600	44526	286.3	178.5
Myo-Inositol	46489	102983	1179692	-83.2	-91.3
alpha-D-Glucopyranoside	17464137	22981642	13754201	47.2	31.6
Carboxylic acid					
Buatnoic acid	76445	73591	315005	-62.4	-76.6
Butanedioic acid	91196	14972127	9528159	211.3	57.1
Propanoic acid	1552038	2864163	1404507	93.8	84.5
Ethanedioic acid	25590612	20753643	12967662	7.6	-18.9
Fatty acids					
alpha-Linolenic acid	1116124	892110	1674016	-36.1	-46.7
9,12-Octadecadienoic acid	896862	298289	1016032	-68.8	-70.6
Octadecanoic acid	3835768	4307998	2810526	29.6	12.3
Hexadecanoic acid	7806828	8898091	9142708	5.0	-2.7
Tetradecanoic acid	5456007	5199288	5975184	-9.0	-13.0
Pentenoic acid	229281	195194	125366	10.1	-14.9

Annexure 13. List of the metabolites only found in F_1 hybrids of *Capsicum.*

Metabolites	*C. annuum* (Dudu) x *C. chinense* (Lota bhut)	*C. chinense* (Acc. 7) x *C. frutescens* (Acc. 4)	*C. chinense* (Chocolate Jolokia) x *C. frutescens* (Acc. 4)	*C. chinense* (Lota bhut) x *C. chinense* (Chocolate jolokia)	*C. chinense* (Lota bhut) x *C. frutescens* (Acc. 4)	*C. chinense* (Umorok) x *C. frutescens* (Acc. 4)
Sugar & its derivatives						
1,2,3,4,5 pentakis	-	-	-	1313880	-	-
1,2,3,4,6-pentakis (pentagalloyl glucose)	15075857	-	-	-	-	-
2,3,4,5 tetrakis	-	-	-	-	1425204	-
2-deoxy-glactopyranose	-	-	240188	-	-	-
2-keto-glutaric acid	1016482	-	-	-	-	-
2-ketohexanoic acid	-	-	518696	-	-	-
Alpha d galactoside	-	-	-	153517	-	-
Alpha-d-galactopyranoside	-	-	-	-	86750	-
Amino succinate	-	3237204	-	-	4490479	-
Anhydroglucitol	-	-	1505022	-	161842	76298
Arabino-hex-1-enitol	-	-	53123	-	-	-
Beta -hydroxypyruvic acid	-	-	-	143055	-	-
Beta-l-manopyranose	216172	-	-	-	-	-
Bis-(dimethyl-t-butyl)maleate	-	547921	-	-	-	-
Bis-(dimethyl-t-butylsilyl)maleate	5413190	-	-	-	-	-
Butylethylmelonic acid	-	-	-	-	138692	-
Chizo-ionositol	-	-	169220	-	-	-
D galctosefuranose	-	253304	-	-	-	-
D-arabinoic acid	-	1023202	-	-	-	-
D-erythro-pentofuranose	116352	-	-	-	-	-
D-erythro-pentopyrynose	128796	-	-	-	-	-
D-galactose	-	-	642198	-	-	-
D-glucitol	-	148789	-	-	-	-
D-glucohexodialdose	-	-	-	587141	-	-
D-glucuronic acid	-	2173472	-	-	-	-
D-lactic acid	-	-	-	54695	-	-
D-threo--2,5,hexodiulose	-	-	-	113202	-	15792
D-xylose	-	-	-	29206	551941	-
Erythro-pentonic acid	146191	-	-	454057	-	-
Fructose benzoyl oxime	68280	-	-	-	-	-
Fucose	-	-	-	-	-	151039
Galactinol	-	1261827	-	-	-	-
Galactoside	-	449255	664268	-	-	-

Gluconsaeure	-	-	-	174730	-	-
Glucopyranose	-	-	-	-	159628	-
Glucuronic acid	-	-	1313257	186747	2766710	-
Glucuronolactone	-	-	-	-	265707	180654
Glycoside	-	496125	-	-	-	-
Hexopyranose	-	-	-	-	249446	-
Indolecarboxaldehyde	27528118	-	-	-	-	-
Isocitric acid lactone	-	-	-	2412256	-	-
L-mannopyranose	-	-	-	29956	-	-
Lyxose	-	366193	-	-	-	-
Mannonic acid	-	-	165159	-	-	-
Mannose	-	1726780	-	-	-	-
Melibiose	-	1077966	-	-	-	-
Methylmaleic acid	-	-	-	23607	-	-
Palatinose	-	218066	-	-	-	-
Pyruvic acid	-	-	-	-	-	169814
Ribonic acid	-	-	-	-	399783	399714
Succinylacetone	600301	-	-	-	-	-
Trihydroxybutyric acid	-	177629	-	-	-	-
Amino acid						
l- valine	-	149489	-	-	-	-
Citrulline	-	-	-	-	417829	-
L-glutamin	-	-	-	3837702	-	-
L-leucine	-	-	-	-	74409	-
L-lysine	-	196549	-	-	108907	-
L-methanonine	-	-	-	-	132341	-
l-phenyl alanine	-	-	-	25850	28426	-
l-tyrosine	-	290490	-	-	-	-
Carboxylic Acid						
2-hydroxy-3- methyl butanoic acid	-	-	-	81925	-	-
2,3,4, trihydroxybutyric acid	-	-	518239	68158	-	-
2-bromosebacic acid	234608	-	-	-	-	-
2-hydroxydodecanedioic acid	-	-	-	-	106702	-
2-hydroxyisocarproic acid	53778	-	-	57223	-	-
2-ketoisocarpoic acid	-	-	212700	-	-	-
2-propanoic acid	-	-	76866	-	-	-
4,6-dioxoheptanoic acid	-	-	1069555	-	-	-
Azelaic acid	1796823	-	-	-	-	-
Cyclopropane-1-carboxylic acid	-	-	-	15079	-	-

Dioxoheptanoic acid	-	-	-	162918	248554	-
Isophthalic acid	-	-	-	16168	-	-
Oxalic acid	-	-	-	-	-	39433
Tetracossonoic acid	574087	-	-	-	-	-
Undecanoic acid	-	-	-	-	-	76172
Fatty acid						
1-Heptacosanol	681372	-	-	-	-	-
2-monostrearin	-	-	-	82457	-	-
9-octadecenoate	-	-	-	-	-	4600125
Beta-sitasteroltrimethylsilyl ether	687516	-	-	-	-	-
Cis-11-Eicosenoic ACID	10372209	-	-	-	-	-
Heptadecyl glycerol	840235	-	-	-	-	-
Hexanedioic acid	-	-	-	-	171301	-
Hydroxydodecacanedioic acid	-	-	-	-	113009	-
Mevalonic acid	1551242	-	-	-	-	-
Pentenoic acid	-	-	-	121930	-	-
Trans-9-octadecenoic acid	-	-	-	236381	-	-
Others						
1,4-butanediamine	-	-	-	-	252607	241979
1,4-butanediamine -1,4-diol	-	-	-	77158	-	
2-bromononane	-	-	-	26681	-	30086
2-ethyl acetoacetate	-	-	663902	-	-	
2-methylbutane	-	-	-	60205	-	
2-methylbutane 1,4,-diol	-	-	-	-	-	46859
3-methy-2-oxovaleric acid	-	-	79631	-	-	
Acetamide	-	-	-	-	38047	24081
Amino propanol	-	-	-	-	54985	
Aucubin	-	-	240031	-	-	
Bis ethane	-	-	82686	-	-	
Borneol	-	-	275495	-	-	
Butanal	-	1138771	-	-	-	
Butane	1096975	-	-	-	-	
Butanediol	-	-	-	-	93118	
Butanol	-	-	-	43694	-	
Butenedioate	-	-	-	225839	-	
Cyclododecyne	-	-	-	-	-	201918
Cyclohexylidenemalononitril	-	-	179027	-	145585	
Diacetylethylenediamine	-	-	-	18244	-	
Diethylamine	136772	-	-	-	-	

Diisobutoxyisobutane	-	-	-	-	135790	
Dimethyl butanol	-	460660	-	-	-	
Dioxolane	-	-	-	20891	-	
Hexanol	-	-	-	89056	-	
Indole-3-acetic acid	-	-	-	-	7899839	
Indole-3-etnanamine	1376020	-	-	-	-	
Methanamine	-	-	-	87831	103163	90857
Metheoxy cyclohexane	-	-	-	-	-	27590
methyl ethyl Choloroactetate	-	-	-	-	-	115616
Methyoxy acetate	-	-	623261	-	-	
Monomethyl phosphate	-	-	-	-	-	163706
Octane	-	-	189271	80957	126397	
P-allyl-o-methoxy benzene	-	-	-	21474	-	
Penicillamine	-	-	-	29413	-	
Pentane	-	-	-	38479	-	
Phenyloxianylidene	-	-	-	-	-	6129
Propanetriol	-	-	-	309312	-	
Propyldecanol	-	-	-	-	113608	
Quinoline	-	-	-	119279	-	
Quinolinelononitril	-	-	123896	-	-	
Ribitol	-	-	-	-	237279	62563
Tetra butylamin	-	-	-	400946	-	272892
Tridecane	-	-	-	-	-	77221
Trimethyl ether	-	-	-	-	-	161367
TrimethylOctanal	-	-	-	-	-	33264
Trimethylamine	-	-	-	-	112397	
Trisiloxane	-	-	-	-	334204	242159
Uridin	-	165138	-	-	-	90568
Uridine anhydride	-	-	-	-	136609	

Annexure XIV & XV are separately given in CD Drive

PUBLICATIONS

PUBLICATIONS

1. Sarpras M, Rashmi Gaur, Vineet Sharma, **Sushil Satish Chhapekar**, Jharna Das, Ajay Kumar, Satish Kumar Yadava, Mukesh Nitin, Vijaya Brahma, Suresh K. Abraham, Nirala Ramchiary. (2016). Comparative Analysis of Fruit Metabolites and Pungency Candidate Genes Expression between Bhut Jolokia and Other *Capsicum* Species. PLoS ONE 11(12): e0167791. doi:10.1371/journal.pone.0167791.

BOOK CHAPTERS

1. **Chhapekar SS,** Gaur R, Kumar A, Ramchiary N (2016) Reaping the Benefits of Next-generation Sequencing Technologies for Crop Improvement - Solanaceae. In: Kulski JK (eds) Next Generation Sequencing - Advances, Applications and Challenges, InTech Publishers, Croatia, pp- 247-285 DOI: 10.5772/61656.
2. **Chhapekar S,** Kehie M, Ramchiary N (2016) Advances in Molecular Breeding of Capsicum Species. In Deka PC (eds) Biotechnological Tools for Genetic Resources, Daya Publishing House, New Delhi, pp- 233-274.

CONFERENCES

1. Presented a poster in an International Plant Physiology Congress, 11th - 14th Dec 2015, JNU, New Delhi pp-328.
2. Delivered an oral talk on ‘Comparative profiling of microRNAs between C. chinense and C. frutescens reveals potential miRNAs involved in fruit development’ in Biospark symposium held at Jawaharlal Nehru University, New Delhi, during 30th - 31st March 2017.

PLOS ONE

RESEARCH ARTICLE

Comparative Analysis of Fruit Metabolites and Pungency Candidate Genes Expression between Bhut Jolokia and Other *Capsicum* Species

Sarpras M[1], Rashmi Gaur[1], Vineet Sharma[1], Sushil Satish Chhapekar[1], Jharna Das[2], Ajay Kumar[1,3], Satish Kumar Yadava[4], Mukesh Nitin[1], Vijaya Brahma[5], Suresh K. Abraham[6], Nirala Ramchiary[1]*

1 Translational and Evolutionary Genomics Lab, School of Life Sciences, Jawaharlal Nehru University, New Delhi, India, **2** Department of Biological Science, Gauhati University, Guwahati, Assam, India, **3** Department of Plant Science, School of Biological Sciences, Central University of Kerala, Periya, Kasaragod, Kerala, India, **4** Centre for Genetic Manipulation of Crop Plants, University of Delhi South Campus, Benito Juarez Road, New Delhi, India, **5** School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India, **6** School of Life Sciences, Jawaharlal Nehru University, New Delhi, India

* nrudsc@gmail.com

Citation: M S, Gaur R, Sharma V, Chhapekar SS, Das J, Kumar A, et al. (2016) Comparative Analysis of Fruit Metabolites and Pungency Candidate Genes Expression between Bhut Jolokia and Other *Capsicum* Species. PLoS ONE 11(12): e0167791. doi:10.1371/journal.pone.0167791

Editor: Guangyuan He, Huazhong University of Science and Technology, CHINA

Received: July 30, 2016

Accepted: November 20, 2016

Published: December 9, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Department of Biotechnology, Ministry of Science and Technology, Government of India, in the form of prestigious Ramalingaswami Fellowship cum Project Grant (No. BT/RLF/Re-entry/46/2011) to Nirala Ramchiary. The authors are highly thankful to Jawaharlal Nehru University (JNU) for providing financial assistance for publication.

Abstract

Bhut jolokia, commonly known as Ghost chili, a native *Capsicum* species found in North East India was recorded as the naturally occurring hottest chili in the world by the Guinness Book of World Records in 2006. Although few studies have reported variation in pungency content of this particular species, no study till date has reported detailed expression analysis of candidate genes involved in capsaicinoids (pungency) biosynthesis pathway and other fruit metabolites. Therefore, the present study was designed to evaluate the diversity of fruit morphology, fruiting habit, capsaicinoids and other metabolite contents in 136 different genotypes mainly collected from North East India. Significant intra and inter-specific variations for fruit morphological traits, fruiting habits and 65 fruit metabolites were observed in the collected *Capsicum* germplasm belonging to three *Capsicum* species i.e., *Capsicum chinense* (Bhut jolokia, 63 accessions), *C. frutescens* (17 accessions) and *C. annuum* (56 accessions). The pungency level, measured in Scoville Heat Unit (SHU) and antioxidant activity measured by 2, 2-diphenyl-1-picrylhydrazyl (DPPH) free radical scavenging assay showed maximum levels in *C. chinense* accessions followed by *C. frutescens* accessions, while *C. annuum* accessions showed the lowest value for both the traits. The number of different fruit metabolites detected did not vary significantly among the different species but the metabolite such as benzoic acid hydroxyl esters identified in large percentage in majority of *C. annuum* genotypes was totally absent in the *C. chinense* genotypes and sparingly present in few genotypes of *C. frutescens*. Significant correlations were observed between fruit metabolites capsaicin, dihydrocapsaicin, hexadecanoic acid, cyclopentane, α-tocopherol and antioxidant activity. Furthermore, comparative expression analysis (through qRT-PCR) of candidate genes involved in capsaicinoid biosynthesis pathway revealed many fold higher expression of majority of the genes in *C. chinense* compared to *C. frutescens* and

Competing Interests: The authors have declared that no competing interests exist.

C. annuum suggesting that the possible reason for extremely high pungency might be due to the higher level of candidate gene(s) expression although nucleotide variation in pungency related genes may also be involved in imparting variations in level of pungency.

Introduction

The Chili peppers belonging to the family Solanaceae and genus *Capsicum* shows an incredible diversity and are consumed by a large section of population throughout the world because of its impressive health beneficial chemical compounds such as capsaicinoids, carotenoids (provitamin A), flavonoids, vitamins (Vitamins C and E), minerals, essential oils and aroma of the fruits [1,2,3,4,5]. These compounds have shown to possess anticancer [6,7,8,9] anti-inflammatory [10], antimicrobial [11] and antioxidant [12] properties.

Capsaicinoid contents, a group of alkaloids, specifically present only in the members of the genus *Capsicum*, is responsible for giving pungency or heat to the fruit. The capsaicinoid biosynthesis involves convergence of two pathways i.e. the phenylpropanoid pathway which provides the precursor phenylalanine for the formation of vanillylamine, and the branched chain fatty acid pathway which provides the precursors valine or leucine for 8-methyl-6-nonenoyl-CoA formation. Capsaicinoids accumulation occurs specifically in the epidermal layer, called dissepiment of the placental tissue, mostly after 20 to 30 days of pollination and continues till the fruit ripening stage [13,14]. Till date, 23 capsaicinoid analogues have been reported, among which, capsaicin (trans-8-methyl-N-vanillyl-6-nonenamide) and dihydrocapsaicin (8-methyl-N-vanillylnonanamide) constitute about 77–98% of the capsaicinoids content in capsicum [15,16,17]. Apart from these two major capsaicinoids, other capsaicinoids such as nordihydrocapsaicin, homocapsaicin, homodihydrocapsaicin, and nonivamide are also found in small quantities in capsicum fruits [18,19].

Some of the genes involved in capsaicinoid biosynthesis have been characterized and their sequence analysis and expression profiles are studied extensively in different pungent and non-pungent varieties, mostly in *C. annuum* [14,20,21,22]. Stewart et al. (2005) [14] reported that the presence of capsaicinoids is controlled by the *Pun1* locus, and confirmed their presence in the interlocular septa of pungent fruit by using HPLC analysis. Later, Stewart et al. (2007) [23] identified a 2.5 kb deletion in *C. annuum* sequence that constituted 1.8 kb of the promoter and 0.7 kb of the first exon of SB2-66 clone and named as *Pun1* or *AT3* as it contains acyltransferase domains. Recently, two separate groups i.e. Kim et al. (2014) [24] and Qin et al. (2014) [25] independently published the whole genome sequence and reported capsaicinoid biosynthesis genes in *C. annuum*. Reddy et al. (2014) [26] based on candidate gene association mapping studies suggested that the *Pun1* acts as a key regulator in the capsaicinoid pathway and only the expression of this gene decides the accumulation of capsaicinoids. Their analysis also revealed that the *CCR* (Cinnamoyl CoA reductase) and *KAS*I (β-ketoacyl carrier protein synthase I) are the two important enzymes involved in pathways for the regulation of capsaicinoid biosynthesis in capsicum.

Of the total 38 *Capsicum* species reported, *C. annuum* is the most extensively grown worldwide among the 6 cultivated species. The other cultivated species are *C. baccatum*, *C. chinense*, *C. frutescens*, *C. pubescens*, and *C. assamicum* [27,28]. It is believed that the unique climatic condition of North East India have made this region one of the biodiversity hotspots of the world and Bhut jolokia or "Ghost chili" (Assamese word) with its fiery hot pungent characteristics is one of them. It is also known as the Naga King chili or Naga morich in Nagaland and

"Umorok" in Manipur State of North east India and is considered as the world's naturally originated hottest chili (Guinness Book of World records, 2006) [29]. This particular species of pepper which is grouped into *C. chinense* is grown mostly in the backyard of North East India household since time immemorial, although recently it is being cultivated commercially because of its unique aroma, nutritive and medicinal properties. Apart from this species, wide variation observed in capsicum germplasm belonging to *C. annuum* and *C. frutescens* makes North Eastern India one of the important sources of genetic resources of chili peppers. However, only fragmented studies and reports on diversity in capsicum germplasm, which is based on capsaicinoids are currently available [30,31]. Recently, Islam et al. (2015) [32] evaluated the levels of variation in capsaicinoid content in 139 diverse accessions using high performance liquid chromatography (HPLC) method. However, the detailed characterization and documentation of capsicum germplasm with respect to morphological traits, pungency, other metabolites, vitamins and their contribution towards antioxidant activities have not been reported till date. Furthermore, extensive comparative studies on expression of candidate genes involved in capsaicinoid biosynthesis using germplasm belonging to different capsicum species of North Eastern India are lacking.

Therefore, in the present study, our main objectives were to i) characterize different genotypes of the three species—*C. chinense*, *C. frutescens* and *C. annuum* for fruit morphology and metabolites including pungency, vitamins, and antioxidant activity; ii) to understand the overall correlation between different metabolites and antioxidant activities; and iii) to compare the pungency related candidate gene expression in contrasting capsicum germplasm belonging to different capsicum species and their correlations with pungent phenotypes.

Materials and Methods

Plants materials

Majority of the 136 genotypes belonging to the three capsicum species (*C. chinense*, *C. frutescens*, and *C. annuum*) were collected from different regions of North East India i.e. Assam, Nagaland, Manipur and Meghalaya and grown in an experimental plot of School of Life Sciences, Jawaharlal Nehru University, New Delhi following standard cultivation practices. Few samples of *C. annuum* were collected from the states of Kerala, Jammu and Kashmir, and Delhi. Since the collections of germplasm were done from traditional market places, no permission was required. Furthermore, no restricted or endangered materials were damaged during sample collection and research activities. The geographical coordinates are provided in S1 Table. These 136 genotypes included 63 (Acc 1–63) genotypes from Bhut jolokia (*C. chinense*), 17 (Acc 64–80) genotypes from *C. frutescens* and 56 (Acc 81–136) genotypes from *C. annuum*. The Capsicum plants were grown during May to December, 2014 in sunny days in experimental research field with well drained loamy soils rich in nutrients. The seeds were treated with Bavistin and Sodium hypochlorite to prevent seed-borne diseases and sown in germination tray. The field is prepared with repeated plowing. Before sowing the field was sprayed with copper fungicide to prevent damping off and to control thrips. A 35 kg P (phosphorus) per hectare and 35 kg K (potash) per hectare was applied. The healthy seedlings of 1 months old were transplanted with spacing of 45 cm X 50 cM (plant to plant and row to row). A 70 kg of N (nitrogen) per hectare was applied at 30, 60, 90 days after transplanting for flowering and proper vegetative growth. The field is irrigated once in 4–5 days. The plants were grown in three rows, each of 3 meter length and 6–10 fruits (depending on the size) from middle plants and second flush of fruit settings were harvested carefully at ripening (mature) stage and kept for drying for further analysis.

Reagents & chemicals

The entire chemicals used in this study were HPLC grade and purchased from Himedia (India) and Sigma Aldrich Co. (USA). The standards of capsaicin and dihydrocapsaicin for estimation of capsaicinoid content and 2, 2-diphenyl-1-picrylhydrazyl (DPPH) for antioxidant assays were purchased from Sigma Aldrich Co. (USA).

Capsaicinoid and other metabolite extractions

The ripened fruits (deseeded) were homogenized in methanol (1:10, w/v) and filtered through Whatman paper No. 1 over anhydrous Sodium sulphate (Na_2SO_4). The filtered extract was evaporated to dryness in vacuum and the residue was suspended with 10 ml acetonitrile as reported earlier [33]. The samples were then centrifuged at 14,000 rpm for 10 minutes and filtered through 0.45 μm Polytetrafluoroethylene (PTFE) membrane filter (Millipore) before injecting to GC-MS. Three independent replicates of samples were used for extraction and GC-MS analysis.

GC-MS analysis

Detection and quantification of capsaicinoids and the presence of other metabolites was carried out by gas chromatography coupled with mass spectrometry (Shimadzu QP2010 Plus) equipped with a Rtx- 5 MS capillary column (0.25 mm film thickness, 0.25 mm internal diameter, and 30 m in length). The oven temperature was set at 100˚C for 2 min, then increased to 250˚C at a rate of 5˚C per minute, and finally to 280˚C at a rate of 10˚C per minute. One μl of each sample was injected to the column in split mode (split ratio 10) with helium as the carrier gas with a flow rate of 1.21 ml per minute. The presence of distinctive peak fragmentation patterns for various metabolites was detected by an MS detector in full scan mode. Capsaicin and dihydrocapsaicin were determined using external reference standards injected under the same conditions. Their identification was based on the retention times and mass measured under identical GC-MS conditions, while their quantitative determinations in the different samples were carried out using the peak areas. Identification of metabolites was confirmed by comparing the spectral data of peaks with the corresponding standard mass spectra from the library database [National Institute of Standards and Technology library (NIST05) and Wiley 8]. Capsaicinoid contents from all the genotypes were expressed in μg/g of fruits and final value was expressed as Scoville heat unit (SHU) by multiplying with the conversion factor of 16.0×10^6 for capsaicin and dihydrocapsaicin, 9.3×10^6 for nordihydrocapsaicin and 9.2×10^6 for nonivamide [34].

Antioxidant assay

Antioxidant activity of different capsicum species was evaluated by 2, 2 diphenyl-1-picrylhydrazyl (DPPH) free radical scavenging assay. The DPPH solution (100 μM) was freshly prepared in 100% methanol. Sample solutions (concentrations: 100mg/ml, 50mg/ml and 25mg/ml) were prepared in acetonitrile and 25 μl aliquots were then added to a 96 well micro plate containing a 225 μl DPPH (0.1 mM). The reaction mixtures were incubated in the dark, at room temperature for 15 minutes and the absorbance was measured at 517 nm in a Multiplate reader (Thermo Fisher Scientific). The free radical scavenging capability was evaluated by comparing to a blank, which contained only methanol. For obtaining the calibration curve, five concentrations of ascorbic acid (100μg– 6.25μg) and capsaicin (1000μg– 62.5μg) in acetonitrile were used. Percentage of free radical scavenging activity (AA) was determined by using

the following equation-

$$\%AA = \frac{(Acontrol - Asample)}{Asample} \times 100$$

Where $A_{Control}$ is the absorbance of the reaction mixture excluding test sample (DPPH solution) and A_{Sample} is the absorbance of the reaction mixture of the test sample (DPPH solution with sample). All the tests were conducted in triplicates and the values were expressed as means ± SD [35].

Quantitative Real-Time PCR

Total RNA was extracted using Lithium chloride (LiCl) precipitation method from fruit tissue at green, breaker and mature stages of fruit development. Complementary DNA (cDNA) was synthesized from 1 μg of RNA using Verso cDNA synthesis kit (Thermo Fisher Scientific) according to manufacturer's instructions. To perform expression analysis, genes from the capsaicinoid pathway were selected and primers were designed using Primer Express version 3.0 software (Applied Biosystems) and custom synthesized from Sigma Genosys (Sigma Aldrich). Using these primer pairs, qRT-PCR was performed in 10μl reaction volumes that contained 1μl cDNA, 5 μl SYBR green master mix (Agilent Technologies), 0.2 μl of 10 μM of each primer, 0.2 μl of the reference dye (Agilent Technologies) and 3.4 μl of nuclease free water. qRT-PCR was performed in a ABI7500 Fast Real-Time PCR system (Applied Biosystems) with the following thermal profile: initial denaturation at 95˚C for 2 min followed by 40 cycles of amplification of 15 sec at 95˚C and 1 min at 60˚C. Finally, a melting curve analysis was performed from 60 to 95˚C in increments of 0.5˚C to confirm the presence of a single product and absence of primer dimers. Each sample was assayed in triplicates, and each experiment was repeated at least twice. For expression analysis, comparative threshold cycle (Ct) method was used which also called as $2^{-[\Delta\Delta Ct]}$ method [36]. For data normalization a house keeping gene actin was used as an internal control.

Statistical analysis

Summary statistics and principal component analysis (PCA) of the metabolites obtained from GC-MS analysis of the *Capsicum* genotypes were performed using the mixOmics package in R environment for statistical computing (version 3.2.3). Summary statistics comprised of mean, standard deviation and analysis of variance (ANOVA) at 95% confidence limit, F-value ($P \leq 0.001$) significance. Correlation analysis using Pearson correlation method and adjusted for multiple testing by using Bonferroni correction were implemented in R (S2 Table). Student's t-test was used for analyzing qRT-PCR data.

Results

Morphological variations

The 136 different accessions collected mainly from North Eastern India were characterized for fruiting habits, fruit morphology and colors (Fig 1 and Table 1). The highest variations of fruit morphology, especially fruit shape, size and length were observed in *C. annuum* accessions followed by *C. chinense*, while *C. frutescens* showed mostly one type of fruit shape among the collected accessions. The fruit shapes observed were long, elongate, ovate, round, pumpkin shape and varied from small to large fruits in *C. annuum*; ovate and elongated type in *C. chinense*; and very small elongated fruits in *C. frutescens*. The contrast in fruit color varied from orange, red, yellow and chocolate colors. Fruiting habits were observed to be upright and pendant in

Fig 1. Morphological diversity of *Capsicum* species. Selected *Capsicum* germplasm from North East India showing contrasting phenotypes for fruit morphology, color, and fruiting habits. Accessions in 1-3rd rows are contrasting Bhut jolokia genotypes (*C. chinense*), 4th and 5th row contains *C. chinense*, *C. frutescens* and *C. annuum* accessions.

doi:10.1371/journal.pone.0167791.g001

Table 1. Morphological characteristics of *Capsicum* fruits.

Species	Fruit length (cm)			Fruit weight (g)			Seed count number			Seed weight (g)	Fruit characteristics			
	Mini-mum	Maxi-mum	Ave-rage	Mini-mum	Maxi-mum	Ave-rage	Mini-mum	Maxi-mum	Ave-rage	Ave-rage	Fruiting habit	Fruit shape	Fruit color at maturity	Fruit shape at blossom end
C. annuum	1.3	10.22	5.59	0.22	7	2.89	14	100	46.75	0.071	Mostly pendant	Elongated, almost round or block shaped	Light red, yellow, dark red	Pointed, blunt or sunken
C. chinense	2.7	8	4.73	0.7	10.58	4.98	8	60	26.11	0.035	Mostly pendant	Triangular, ovate	Red, Orange or chocolate	Pointed, blunt or sunken
C. frutescens	0.7	2.56	1.4	0.05	0.42	0.28	3	14	7.07	0.053	Erect upward	Short slender	Red or Orange	Pointed or blunt

doi:10.1371/journal.pone.0167791.t001

C. annuum, only upright in *C. frutescens* and only pendant in *C. chinense* (Bhut jolokia). Variation from single to bunch type fruiting habits were observed in *C. annuum* and *C. chinense*, whereas in *C. frutescens* only single fruiting habit was observed.

Determination of capsaicinoid contents

Pungency, a unique and important property of *Capsicum* species, attributed to the capsaicinoid contents was analyzed in all the 136 different germplasm collected using Gas Chromatography coupled with Mass Spectrometry (GC-MS). The extraction was carried out by using acetonitrile and the capsaicinoid contents was separated by using GC-MS. The quantity of the complex was calculated by means of calibration curves. The correlation coefficients for capsaicin and dihydrocapsaicin were >0.998 and >0.995, respectively (S1 Fig). Of the two major capsaicinoids, i.e. capsaicin and dihydrocapsaicin, the former was found to be more abundant in the collected *Capsicum* germplasm belonging to *C. annuum*, *C. frutescens* and *C. chinense* of North East India. Other two capsaicinoids, nordihydrocapsaicin and nonivamide were also present in many of the accessions but in small quantities. Capsaicinoid contents were measured both in Scoville Heat Unit and amount in μg/g of fruit for all the genotypes (S3 Table).

The pungency, as expected was observed to be high in *C. chinense* accessions compared to accessions belonging to the other two *Capsicum* species. The Scoville Heat Unit (SHU) value, a unit of heat/pungency measurement, ranged from 272897 (0.27 million) to 1037305 (1.0 million), 109508 (0.1 million) to 487619 (0.48 million) and 0 (bell pepper) to 203731 (0.2 million) in *C. chinense*, *C. frutescens* and *C. annuum* accessions, respectively (S3 Table and Fig 2). The highest pungency of more than 1 million SHU value was obtained for three *C. chinense* genotypes with accession numbers 8, 23 and 42. SHU values between 0.9 to 1.0 million were observed in 14 genotypes of *C. chinense* (Accessions 7, 11, 19, 20, 22, 24, 25, 29, 32, 43, 45, 48, 50 and 54), between 0.8 to 0.9 million in 10 genotypes (Acc 2, 6, 10, 18, 31, 40, 41, 49, 53 and 56), 0.7 to 0.8 million in 8 genotypes (Acc 4, 12, 16, 17, 34, 37, 53 and 55), 0.6 to 0.7 million in 10 genotypes (Acc 1, 14, 15, 28, 30, 38, 44, 46, 47 and 51), 0.5–0.6 million in 10 genotypes (Acc 5, 9, 13, 21, 26, 33, 35, 36, 59 and 63) and only 8 genotypes of *C. chinense* (Acc 3, 27, 39, 57, 58, 60, 61 and 62) showed pungency below 0.5 million with varying capsaicin and

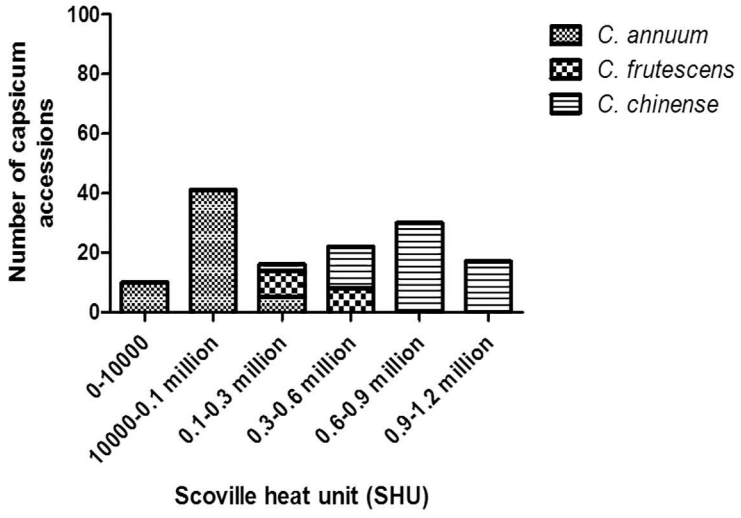


Fig 2. SHU range of different *Capsicum* species. Total capsaicinoids content observed in *C. chinense* (63 accessions), *C. frutescens* (17 accessions) and *C. annuum* (56 accessions) accessions in Scoville Heat Unit (SHU).

doi:10.1371/journal.pone.0167791.g002

dihydrocapsaicin levels. Most of the genotypes from *C. frutescens* showed moderate pungency with a SHU value ranging between 0.3–0.5—million, but 9 genotypes expressed a pungency level below 0.3 million. The *C. annuum* accessions 95, 98, 116, and 126 exhibited the lowest pungency level (<5000 SHU). Nordihydrocapsaicin and nonivamide peaks were absent in many of the analyzed genotypes of *Capsicum*. As expected, the accession 87 (bell pepper) showed zero pungency.

Analysis of other metabolites

In the current study, apart from the capsaicinoid contents, other metabolites were also analyzed using GC-MS. These 61 different metabolites were identified after acetonitrile extraction of dried capsicum fruit. These metabolites comprises carboxylic acids (such as Propanoic acid, butanoic acid, hexanoic acid etc.), fatty acid and esters (such as Decanoic acid, Palmitic acid etc.), hydrocarbons (Cyclopentane, Naphthalene etc.), aldehydes (Tetradecanoic acid, Pentadecanoic acid, Eicosanoic acid), terpenoids (2,7-Octadiene, Geranyl linalool isomer B), Alcohol (hexanol, isopropanol) and Vitamin E (α-tocopherol). However, the metabolites concentration varied with genotypes (S4 Table). Many of the compounds were found only in specific genotypes. *C. chinense* and *C. annuum* exhibited a slightly higher number of metabolites compared to *C. frutescens*. An average of 17, 14 and 17 metabolites was identified in the genotypes belonging to *C. chinense*, *C. frutescens* and *C. annuum*, respectively (Fig 3). The number of metabolites identified ranged from 7–32 for *C. chinense*, 5–31 for *C. annuum* and 9–32 for *C. frutescens*, respectively. The metabolites like benzoic acid hydroxyl esters, which are identified in large percentage in majority of *C. annuum* genotypes, were totally absent in the *C. chinense* genotypes and present sparingly in few genotypes of *C. frutescens*. Other metabolites like fatty acids and corresponding esters, hydrocarbons, aldehydes, alcohols and terpenoids were randomly distributed in all the genotypes of *C. chinense*, *C. frutescens* and *C. annuum*.

Antioxidant activity of different *Capsicum* genotypes

The antioxidant activity of different *Capsicum* varieties were analyzed by determining the DPPH scavenging capability. Significant differences in antioxidant activity were observed between *C. chinense*, *C. frutescens* and *C. annuum* accessions. The highest antioxidant (free radical scavenging) activity was observed in *C. chinense* accessions compared to *C. frutescens* and *C. annuum* accessions. The antioxidant activity determined by DPPH assay ranged from

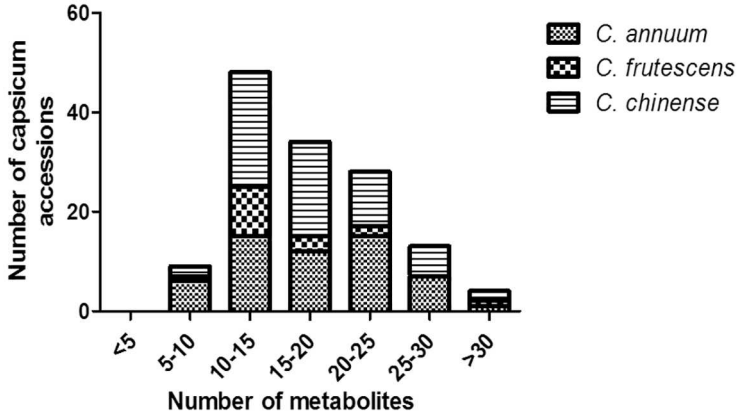


Fig 3. Metabolite range of different *Capsicum* species. Metabolite range of *C. chinense*, *C. frutescens* and *C. annuum* varieties.

doi:10.1371/journal.pone.0167791.g003

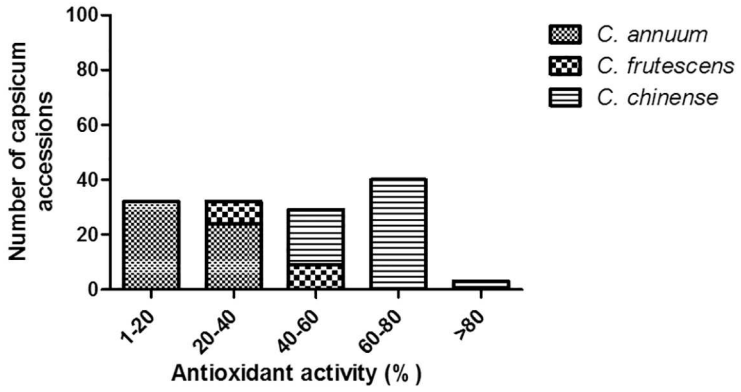


Fig 4. Range of antioxidant activity of different *Capsicum* species. Anti-oxidant activity using DPPH assay obtained for *C. chinense*, *C. frutescens* and *C. annuum* varieties and represented in 25mg/ml dilutions.

doi:10.1371/journal.pone.0167791.g004

40% to 83% in *C. chinense* accessions followed by 31% to 50% in *C. frutescens* and 3% to 38% in *C. annuum* accessions. The average free radical scavenging activity from all the genotypes of *C. chinense*, *C. frutescens* and *C. annuum* were 63.83 ± 2.21, 40.76 ± 3.72 and 18.63 ± 4.52 respectively. Three accessions of *C. chinense* showed more than 80% antioxidant activity, while 39 accessions exhibited antioxidant activity between 60 to 80%, and 21 accessions had antioxidant activity between 40 to 60%. The genotypes from *C. frutescens* showed antioxidant activity ranging from 20 to 50%, 8 accessions showed 20 to 40%, while 9 accessions showed 40–50% antioxidant activities. The majority of *C. annuum* genotypes exhibited antioxidant activity less than 20% (Fig 4 and S3 Table).

Principal component analysis

Metabolite profiling of capsicum fruits identified a total of 65 metabolites by GC-MS. Of these, the metabolites which were present in almost all 136 genotypes were selected for further analysis. These include various bioactive fatty acids like palmitic acid (hexadecanoic acid), octadeccanoic acid (stearic acid), 9(Z)-octadecenoic (oleic acid), cyclopentane and n-octacetylamide and alkaloids like capsaicin, dihydrocapsaicin, nordihydrocapsaicin and nonivamide and α-tocopherol (vitamin E). These metabolites play different roles in capsaicinoid biosynthesis pathway, maintaining cell membrane integrity, signaling and defense mechanism etc. The Principal component analysis (PCA) revealed that genotypes could be differentiated based on their metabolite profiles and the correlation variances explained by the two principal components (PC 1 and PC 2) were observed to be 51% and 11%, respectively (Fig 5). Even though majority of accessions from *C. chinense* and *C. annuum* fall in to separate clusters, the patterns of metabolite expression across the 136 genotypes were not completely differentiated based on the type of species (*C. chinense*, *C. frutescens* and *C. annuum*).

Correlation analysis of these 10 metabolites along with the antioxidant activity showed high correlation among the metabolites with their antioxidant activity. The correlation analysis showed significant correlations between many of these metabolites (Table 2 and S2 Table). Correlation circle plot of PCA analysis clearly illustrated that, there exists correspondence between capsaicin, dihydrocapsaicin, hexadecanoic acid, cyclopentane, α-tocopherol and antioxidant activity (S2A Fig). Metabolites forming a cluster were projected in the same direction with significant distances from the origin highlighting the strength of correlation. In addition PCA also revealed a similar pattern of metabolite correlation across the *Capsicum* genotypes from *C. chinense*, *C. frutescens* and *C. annuum* (S2B Fig).

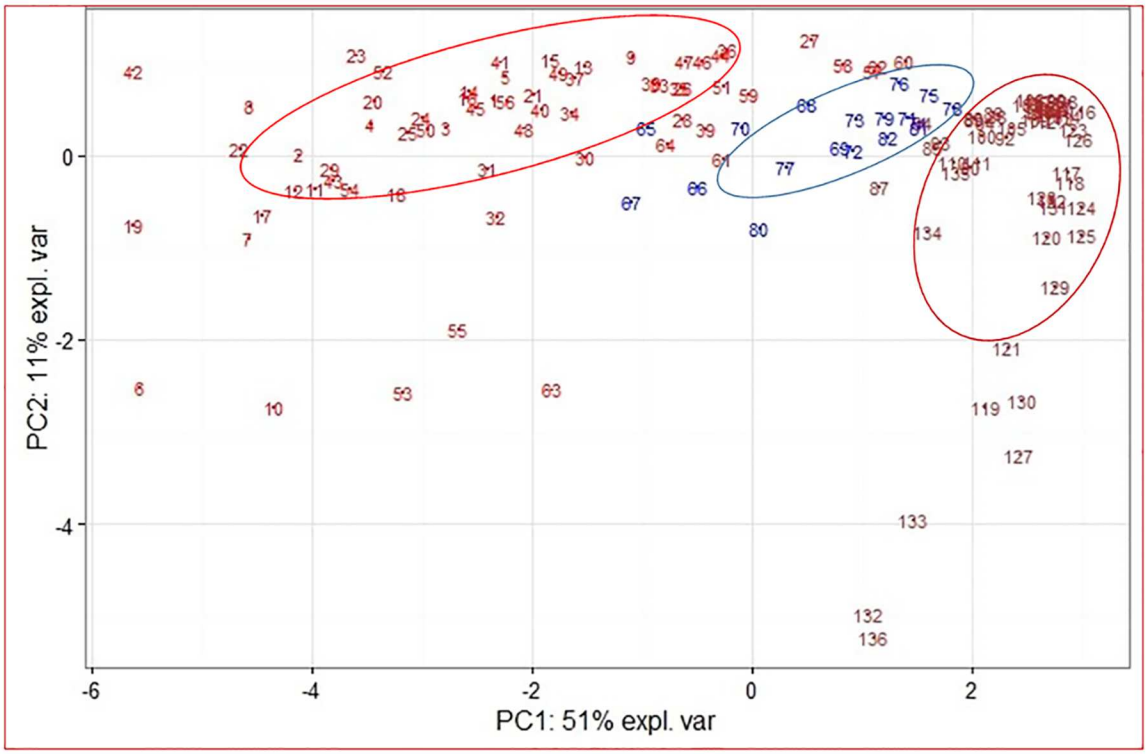


Fig 5. Principal component analysis of metabolites identified using gas chromatography–mass spectrometry (GC–MS) analysis. For GC–MS, different genotypes of *C. chinense* [Acc 1–63 (major accession formed red circle)], *C. frutescens* [Acc 64–80 (major accession formed blue circle)] and *C. annuum* [Acc 81–136 (major accession formed brown circle)] were analysed and the correlation variances explained by the PC1 and PC2 components are 51% and 11%, respectively.

doi:10.1371/journal.pone.0167791.g005

Expression analysis of candidate genes

The different *Capsicum* accessions collected from North East India showed wide variation in pungency contents as evidenced from biochemical analysis, and since the genes involved in capsaicinoid biosynthesis pathway (S3 Fig) have been reported in *C. annuum*, we made attempts to identify variations in the expression of candidate genes in accessions with contrasting pungency levels. In the present study, 10 candidate genes of the capsaicinoid biosynthesis pathway were selected to analyze their expression patterns in leaf, flower and three stages of fruit development (green, breaker and mature stages of the fruit) in highly pungent (*C. chinense* accessions 23 and 50), moderately pungent (*C. frutescens* accession 65) and low pungent genotypes (*C. annuum* accessions 93 and 95) (Fig 6A). The candidate genes selected were *PAL* (phenylalanine ammonia-lyase), *C4H* (cinnamate 4-hydroxylase), *COMT* (caffeic acid 3-O-methyltransferase), *ACL* (acyl-CoA synthetase), *pAMT* (putative aminotransferase), *BCAT* (branched-chain amino acid aminotransferase), *KAS* (ketoacyl-ACP synthase), *ACL* (malonyl-acyl carrier protein), *FAT* (acyl-ACP thioesterase) and *AT3* (*Pun1* or acyltransferase). The sequences of forward and reverse primer pairs are listed in (S5 Table). We observed that the expression levels of these genes varied with the genotypes having different pungency levels. The majority of genes showed significantly higher expression in *C. chinense* genotypes followed by moderately high pungent *C. frutescens*, whereas a very low level of expression was observed in *C. annuum* genotypes (low-pungent). Amongst these candidate genes, *pAMT*, *Pun1/AT3*, *PAL* from Phenylpropanoid pathway and *BCAT*, *KAS* and *ACL* from Fatty acid

Table 2. Pearson correlation co-efficient values observed between different metabolites on the basis of their peak intensities.

	Stearic acid	Oleic acid	Cyclopentane	n-octacetylamide	Capsaicin	Dihydrocapsaicin	Nordihydrocapsaicin	Nonivamide	Vitamin E	Antioxidant activity
Palmitic acid	-0.045	0.64***	0.41***	0.449***	0.782***	0.746***	0.327**	0.341**	0.715***	0.800***
Stearic acid		-0.044	0.07	0.065	-0.069	-0.08	-0.068	0.122	0.131	-0.054
Oleic acid			0.311*	0.561***	0.703***	0.601***	0.421***	0.087	0.426***	0.699***
Cyclo pentane				0.601***	0.351**	0.354**	0.066	0.093	0.307*	0.412***
n-octacetylamide					0.604***	0.449***	0.101	0.162	0.298	0.421***
Capsaicin						0.741***	0.261	0.212	0.647***	0.913***
Dihydrocapsaicin							0.442***	0.264	0.577***	0.832***
Nordihydrocapsaicin								0.224	0.293*	0.345**
Nonivamide									0.268	0.302*
Vitamin E										0.683***

* represents significant at 0.05 level

** represents significant at 0.01 level

*** represents significant at 0.001 level

doi:10.1371/journal.pone.0167791.t002

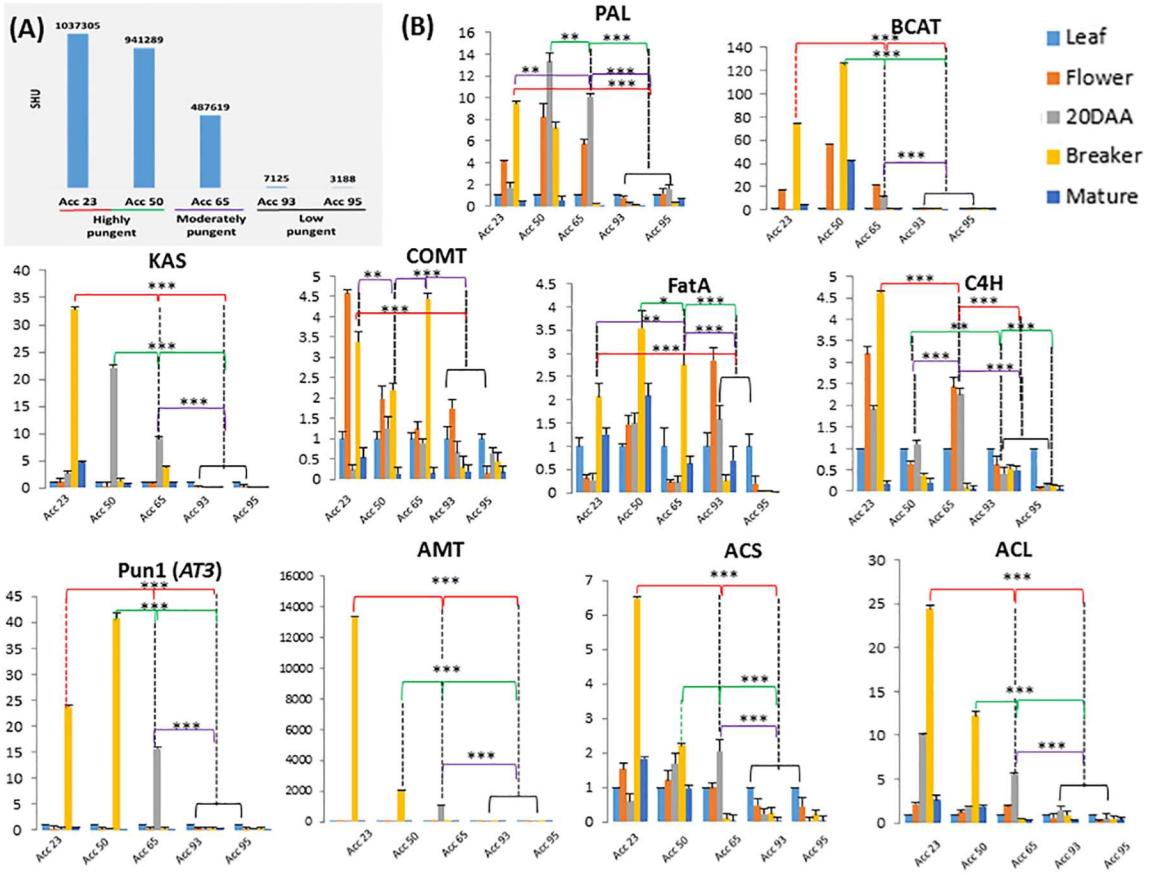


Fig 6. Pungency and capsaicinoid biosynthesis gene expression analysis. (A) Pungency analysis of selected *Capsicum* genotypes **(B)** Quantitative real time PCR analysis to analyze the expression of candidate genes involved in capsaicinoid biosynthesis pathway in highly pungent Bhut jolokia (Acc 23 and Acc 50), moderately pungent *C. frutescens* (Acc 65) and low pungent *C. annuum* (Acc 93 and Acc 95) accessions. The expression analysis was done in leaf, flowers, and three different stages of fruit developmental i.e. green (20 days after anthesis), Breaker (30–45 days after anthesis) and Mature stage of each genotype. The majority of the important genes involved in the capsaicinoid biosynthesis pathway (*Pun 1*, *AMT*, *ACS*, *ACL*, *KAS* and *BCAT*) were expressed very high in *C. chinense* accessions followed by *C. frutescens*. The low pungent *C. annuum* accessions showed very low expression of these genes. The other genes (*PAL*, *COMT*, *FatA* and *C4H*) were expressed variably among the three species. ***$P<0.001$

doi:10.1371/journal.pone.0167791.g006

biosynthetic pathway were found to be highly expressed in pungent genotypes especially in breaker stage of the fruit development (Fig 6B).

Discussion

The North Eastern region of India, one of the biodiversity hotspot of the world, harbors many of the endangered and endemic species of plants and animals. The unique climatic conditions of this region have also favored the evolution of *Capsicum* species, thereby producing landraces and traditional cultivars with diverse morphology, fruiting habits, metabolite contents with varied levels of biotic and abiotic resistances. Bhut jolokia or Ghost chili, reported by the Guinness Book of World Records as the naturally occurring world's most pungent chili pepper has also evolved in this region. Although this particular *Capsicum* species have been cultivated from time immemorial in the kitchen gardens of North East India, until recently, no systematic commercial cultivation was practiced. Having enormous commercial potential of this crop

particularly in producing spices, natural color and for its potential use in medicine, the systematic characterization and identification of potential germplasm for future use in breeding program is pre-requisite. The fragmented studies that have been reported so far are based on the analysis of pungency contents of only few genotypes. Importantly, detailed analysis of fruit metabolites and antioxidant activity and their correlations have not been reported. Furthermore, experimental designs understanding the fiery hot property of this species have also not been reported. We collected 136 different capsicum landraces and traditional cultivars mainly from different places of North East India where chili peppers are grown. As expected we observed a variation in fruit morphology reflected in fruit size, shape, length and color. The different fruit colors (orange, yellow, red etc.) are reported to be determined by both the amount and composition of carotenoids and pigments. The yellow-orange color of peppers is formed by α- and β-carotene, zeaxanthin, lutein, and β-cryptoxanthin and the red color of peppers is due to the presence of carotenoid pigments of capsanthin, capsorubin, and capsanthin 5,6-epoxide [37]. Of the many genes reported, only phytoene synthase (*Psy*) and capsanthin–capsorubin synthase (*Ccs*) are directly shown to be correlating with red, yellow and orange color in different allelic combinations, and studies to identify more genes imparting color are underway. However, no concrete evidence of gene(s) imparting chocolate color is reported in *Capsicum* species. The identification of chocolate color Bhut jolokia in the present study provides new opportunity to use this genotype in future study in identifying genes imparting chocolate color.

GC-MS data demonstrated the different capsaicinoid levels in 136 C*apsicum* genotypes. Our analysis showed that capsaicin and dihydrocapsaicin represented the major fractions compared to other capsaicinoid components, which is consistent with the results obtained in several published reports [38,39,40]. In addition, the present study confirms that capsaicin is the primary capsaicinoid component in almost all of the analyzed *Capsicum* genotypes. However, the total capsaicinoids concentration showed intra-specific as well as inter-specific variations as reported earlier [41,42,43,44,45,46]. The present study identified 17 genotypes from *C. chinense* group with more than 0.9 million SHU which is distinctly higher than the previously reported pungency of Habanero [47,48,49]. Although, Bhut jolokia (*C. chinense*) is known as the natural hottest chili pepper, low pungency Bhut jolokia genotypes were also observed (with pungency as low as 272897.1 ± 38759 SHU), suggesting that during the course of evolution, low pungency alleles were formed and accumulated in those genotypes. Another plausible reason for the low pungency in these accessions could be attributed to the crossability of Bhut jolokia with other cultivated species and therefore, cross pollinations in nature with low pungent *C. annuum* followed by selection might have developed low pungent Bhut jolokia genotypes. A recent study by Dubey et al. (2015) [30] also reports quantification of capsaicin content of 25 *Capsicum* genotypes from North Eastern states of India by using spectrophotometry in which they have also reported variations in pungency content. However, in our study we have listed and quantified all the components of capsaicinoids (capsaicin, dihydrocapsaicin, nor dihydrocapsaicin and nonivamide). The present study also shows the high antioxidant activity of Bhut jolokia accessions compared to the other two species suggesting a strong correlation between capsaicinoids and antioxidant activity. This was further supported by the fact that *Capsicum* accessions possessing lower capsaicin and dihydrocapsaicin also exhibited the lowest antioxidant activity thereby indicating that capsaicinoids also contributes in reduction of free radicals, a property which is deemed beneficial for human health. This finding is also supported by recently published data of Sora et al. (2015) [50].

GC-MS is one of the established and highly suitable techniques for metabolite profiling, as it combines a highly efficient separation technique with versatile and sensitive mass detection methods [51,52]. We have used GC-MS to determine the level and composition of fruit

metabolites in Bhut jolokia, in which except for the capsaicinoid contents other health beneficial fruit metabolites are largely uncharacterized. We identified a total of 65 metabolites including the four capsaicinoids components, fatty acids and esters, aliphatic esters and aldehydes, alcohols, hydrocarbons and vitamin E. These metabolites have been previously reported to be found in various ripened *Capsicum* varieties mostly in *C. annuum* [53,54], and are the primary components of the essential oils, which give aroma to chili peppers [1,4], but have not been characterized in Bhut jolokia. These metabolites were found in varied concentrations in the different genotypes and species. These health promoting compounds of Bhut jolokia can be studied in detail and manipulated in future breeding programs.

Among the total of 65 metabolites, 10 were detected in majority of the *Capsicum* genotypes. Principal component analysis showed five metabolites to be highly correlated among them along and to the antioxidant activity. These metabolites are palmitic acid, cyclopentane, capsaicin, dihydrocapsaicin and α-tocopherol. This appears to be the first study showing the positive correlations in fatty acid, capsaicinoid and vitamin E pathway. The present data is also consistent with the previous study, which demonstrates the role of long chain fatty acids in capsaicin biosynthesis pathway in *C. annuum* [55,56]. There are other efforts which demonstrate the correlation of vitamin E and capsaicinoid synthesis [57]. None of the studies so far have been reported underlining the correlation between fatty acids, capsaicinoids and vitamin E along with antioxidant activity in a large number of genotypes which comprise of *C. chinense*, *C. frutescens* and *C. annuum* groups. Although, pungency and other metabolites found in *C. annuum* genotypes varied, it was not as wide as in the case of Bhut jolokia (*C. chinense*) genotypes. These might be due to the fact that most of the *C. annuum* varieties are derived from more related genotypes showing homogeneity.

Our qRT-PCR results showed the relationship between the expression of candidate genes and the level of pungency in the *Capsicum* genotypes analyzed in this study. Further, among the different stages of fruit development used in this study, the maximum expression of these genes was obtained at 20 DAA (green in *C. frutescens*) and breaker stage (35–45 DAA in *C. chinense*). We found many candidate genes i.e. *pAMT*, *Pun1*, *KAS*, *ACS*, *BCAT*, *ACL* and *FAT* to be highly expressed in different fruit developmental stages which were many fold higher in *C. chinense* compared to the *C. frutescens* and *C. annuum* genotypes suggesting a high correlation of gene expression with higher pungency content. Iwai et al. (1979) [13] also observed that capsaicinoids are synthesized in the placenta in between 20 to 30 DAA in pungent varieties of *Capsicum*. Very recently, Ogawa et al. (2015) [58] verified the involvement of *Pun1* genes in capsaicin biosynthesis. They also studied the expression profiles of *Pun1* and *pAMT* genes and concluded that the accumulation of capsaicin content is highly correlated with the expression levels of these genes in different varieties of *Capsicum*. The expression analysis revealed that along with *Pun1*, the *pAMT* gene also significantly expressed very high in 20 DPA and breaker stages of fruit development in *C. frutescens* and *C. chinense* respectively, compared to low pungent *C. annuum*. Lang et al. (2009) [59] observed that functional loss of *pAMT* gene leads to formation of capsinoids (a sweat analog of capsaicinoid) in non-pungent *C. annuum* cv. CH-19. A single nucleotide (T) insertion at 1291 bp of *pAMT* resulted in formation of stop codon that prevented gene translation and protein accumulation. The study confirms the crucial role of *pAMT* gene in capsaicinoid biosynthesis pathway. Our result shows that *pAMT* gene is mainly expressed in 20 DPA and breaker stages of fruit and co-related with the amount of pungency. The expression of *pAMT* gene is significantly high in *C. chinense* (highly pungent) accessions compared to *C. frutescens* (moderately pungent) and *C. annuum* (low pungent). An association mapping study of Reddy et al. (2014) [26] revealed that *Pun1* acts as a key regulator of major metabolites and that the capsaicinoids accumulation depends on the expression of *Pun1*. Further, the evidences available support *KAS* as an important player in altering the

pungency in *Capsicum* varieties. For e.g. Aluru et al. (2003) [21] reported that *KAS* expression was positively correlated with the level of pungency. Later, this was confirmed by Abraham-Juarez et al. (2008) [60] by a virus induced silencing of *KAS* leading to very low levels of mRNA and thus low capsaicinoids in the pungent variety of *C. chinense*. Our study is the first comprehensive study in Bhut jolokia which shows correlation of expression of candidate genes of pungency biosynthesis pathway with the pungency content.

Conclusions

Diversity is a prerequisite for breeding improved varieties in any crop plant. Our findings from the present study, which observes large morphological and fruit metabolites diversity among the *Capsicum* genotypes found in the North East India, would constitute a valuable resource for future improvements on capsicum breeding. Bhut jolokia, although known as naturally occurring highest pungent chili pepper, also showed to have low pungent genotypes. Our results suggest that the many fold higher expression of candidate genes involved in capsaicinoid biosynthesis pathway is the most plausible reason for finding very high pungent phenotypes of Bhut jolokia compared to *C. frutescens* and *C. annuum*. Furthermore, the variability found in the nutritionally valuable metabolites including capsaicinoids (pungency), vitamins; and a positive correlation with antioxidant activities suggested that these genotypes would be potential genetic stocks towards improving health promoting *Capsicum* varieties through a combined genetics and genomics approach in future capsicum breeding programs.

Supporting Information

S1 Fig. Calibration curve for capsaicin and dihydrocapsaicin.
(TIF)

S2 Fig. Correlation circle plot of analysed metabolite in *Capsicum* genotypes. (A) Correlation circle plot shows that there is a similar correlation pattern between certain metabolites and antioxidant activity in majority of the *Capsicum* species from (*C. annuum*, *C. frutescens and C. chinense*). These metabolites are identified and represented as M1 (hexadecanoic or palmitic acid), M3 cyclopentane, M6 (capsaicin), M7 dihydrocapsaicin, M10 (α-tocopherol or vitamin E) and AA (antioxidant activity). The strongly correlated metabolites were projected in the same direction from the origin of the circle. The distance from the origin indicates the strong association of the metabolites. (**B)** Biplot analysis showing the association between the *Capsicum* accessions and metabolites. Majority of the accessions from the three species of *Capsicum* exhibited similar pattern of metabolites correlation. The angle between the arrows (vectors) showed inversely proportional to the correlation of metabolites. Highly correlated metabolites point in the same direction; uncorrelated metabolites are at right angles to each other.
(TIF)

S3 Fig. Flow chart of capsaicinoid biosynthesis pathway.
(TIF)

S1 Table. Geographical locations and coordinates of sampling sites.
(DOCX)

S2 Table. Bonferroni and Benjamini correction table.
(XLSX)

S3 Table. Concentration of different capsaicinoid components (in μg/g dry weight of fruit) and antioxidant activities observed in *Capsicum* accessions.
(DOCX)

S4 Table. Average area of different metabolites observed in capsicum accessions belonging to Bhut jolokia (*C. chinense*), *C. frutescens* and *C. annuum*.
(XLSX)

S5 Table. List of primer sequences used for expression studies of pungency candidate genes.
(DOCX)

Acknowledgments

This work was supported by the Department of Biotechnology, Ministry of Science and Technology, Government of India, in the form of prestigious Ramalingaswami Fellowship cum Project Grant (No. BT/RLF/Re-entry/46/2011) to Nirala Ramchiary. Financial help in the form of Junior and Senior Research Fellowships to Mr. Sarpras M by University Grants Commission, India, is highly acknowledged. We also acknowledge the kind help of Ms Deepa Bisht in correcting typographical mistakes and references.

Author Contributions

Conceptualization: NR.

Formal analysis: SM RG VS JD VB MN JD AK.

Funding acquisition: NR.

Investigation: SM RG NR SSC VS VB JD.

Methodology: SM SKY MN SSC VB VS JD.

Project administration: NR.

Software: SKY VB SA AK.

Supervision: NR.

Validation: SM RG JD SSC.

Writing – original draft: SM RG SSC VS VB SA.

Writing – review & editing: NR.

References

1. Gbolade AA, Omobuwajo OR, Soremekun RO. Evaluation of the quality of Nigerian chillies for pharmaceutical formulations. J Pharm Biomed Anal. 1997; 15: 545–548. PMID: 8953498
2. Kumar OA, Tata SS. Ascorbic Acid Contents in Chili Peppers (*Capsicum* L.). Not Sci Biol. 2009; 1: 50–52.
3. Howard LR. Antioxidant vitamin and phytochemical content of fresh and processed prepper fruit (*Capsicum annuum*). In: Wildman REC, editor. Hand book of Nutraceuticals and Functional Foods. CRC Press 2000.
4. Forero MD, Quijano CE, Pino JA. Volatile compounds of chile pepper (*Capsicum annuum* L. var. glabriusculum) at two ripening stages. Flavour Fragr J. 2009; 24: 25–30.
5. Purkayastha J, Alam SI, Gogoi HK, Singh L, Veer V. Molecular characterization of "Bhut Jolokia" the hottest chilli. J Biosci. 2012; 37: 757–768. PMID: 22922200
6. Surh Y, Lee RC, Park K, Mayne ST, Liem A, Miller JA. Chemoprotective effects of capsaicin and diallyl sulfide against mutagenesis or tumorigenesis by vinyl carbamate and. 1995; 16: 2467–2471. PMID: 7586153

7. Chanda S, Erexson G, Riach C, Innes D, Stevenson F, Murli H, et al. Genotoxicity studies with pure trans-capsaicin. Mutat Res—Genet Toxicol Environ Mutagen. 2004; 557: 85–97.
8. Anandakumar P, Kamaraj S, Jagan S, Ramakrishnan G, Devaki T. Capsaicin provokes apoptosis and restricts benzo(a)pyrene induced lung tumorigenesis in Swiss albino mice. Int Immunopharmacol. Elsevier B.V.; 2013; 17: 254–259.
9. Oyagbemi AA, Saba AB, Azeez OL. Capsaicin: a novel chemopreventive molecule and its underlying mechanism of action. Indian J Cancer. 2010; 47:53–8.
10. Spiller F, Alves MK, Vieira SM, Carvalho TA, Leite CE, Lunardelli A, et al. Anti-inflammatory effects of red pepper (*Capsicum baccatum*) on carrageenan- and antigen-induced inflammation. J Pharm Pharmacol. 2008; 60: 473–478. doi: 10.1211/jpp.60.4.0010 PMID: 18380920
11. Careaga Monica, Fernandez E, Dorantes L, Mota L, Jaramillo ME, Hernandez-Sanchez H. Antibacterial activity of *Capsicum* extract against *Salmonella typhimurium* and Pseudomonas aeruginosa inoculated in raw beef meat. Int J Food Microbiol. 2003; 83: 331–335. PMID: 12745238
12. Alvarez-Parrilla E, De La Rosa LA, Amarowicz R, Shahidi F. Antioxidant activity of fresh and processed Jalapeno and Serrano peppers. J Agric Food Chem. 2011; 59: 163–173. doi: 10.1021/jf103434u PMID: 21126003
13. Iwai K, Suzuki T, Fujiwake H. Formation and metabolism of pungent principle of Capsicum fruits. IV. Formation and accumulation of pungent principle of hot pepper fruits, capsaicin and its analogues, in *Capsicum annuum* var. *annuum* cv. Karayatsubusa at different growth stages after flow. Agric Biol Chem. 1979; 43: 2493–2498.
14. Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Eun YY, et al. The *Pun1* gene for pungency in pepper encodes a putative acyltransferase. Plant J. 2005; 42: 675–688. doi: 10.1111/j.1365-313X.2005.02410.x PMID: 15918882
15. Govindarajan VS, Rajalakshmi D, Chand N, Salzer UJ. Capsicum—Production, technology, chemistry, and quality. Part IV. Evaluation of quality. C R C Crit Rev Food Sci Nutr. 1987; 25: 185–282.
16. Collins M, Wasmund L, Bosland P. Improved method for quantifying capsaicinoids in *Capsicum* using high-performance liquid chromatography. HortScience. 1995; 30: 137–139. Available: http://hortsci.ashspublications.org/content/30/1/137.short
17. Zewdie Y, Bosland PW. Capsaicinoid profiles are not good chemotaxonomic indicators for *Capsicum* species. Biochem Syst Ecol. 2001; 29: 161–169. PMID: 11106844
18. Constant HL, Cordell GA, West DP. Nonivamide, a constituent of Capsicum oleoresin. J Nat Prod. 1996; 59: 425–426.
19. Huang XF, Xue JY, Jiang AQ, Zhu HL. Capsaicin and its analogues: structure-activity relationship study. Curr. Med. Chem. 2013; 20: 2661–2672. PMID: 23627937
20. Curry J, Aluru M, Mendoza M, Nevarez J, Melendrez M, O'Connell MA. Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent Capsicum spp. Plant Sci. 1999; 148: 47–57.
21. Aluru MR, Mazourek M, Landry LG, Curry J, Jahn M, O'Connell MA. Differential expression of fatty acid synthase genes, *Acl*, *Fat* and *Kas*, in Capsicum fruit. J Exp Bot. 2003; 54: 1655–1664. doi: 10.1093/jxb/erg176 PMID: 12810854
22. Blum E, Mazourek M, O'Connell M, Curry J, Thorup T, Liu K, et al. Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum*. Theor Appl Genet. 2003; 108: 79–86. doi: 10.1007/s00122-003-1405-y PMID: 13679988
23. Stewart C, Mazourek M, Stellari GM, O'Connell M, Jahn M. Genetic control of pungency in *C. chinense* via the *Pun1* locus. J Exp Bot. 2007; 58: 979–991. doi: 10.1093/jxb/erl243 PMID: 17339653
24. Kim S, Park M, Yeom S-I, Kim Y-M, Lee JM, Lee H-A, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat Genet. Nature Publishing Group; 2014; 46: 270–278.
25. Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, et al. Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. Proc Natl Acad Sci U S A. 2014; 111: 5135–5140. doi: 10.1073/pnas.1400975111 PMID: 24591624
26. Reddy UK, Almeida A, Abburi VL, Alaparthi SB, Unselt D, Hankins G, et al. Identification of gene-specific polymorphisms and association with capsaicin pathway metabolites in *Capsicum annuum* L. collections. PLoS One. 2014; 9: 1–10.
27. Taylor P, Maga JA, Todd PH. Capsicum. C R C Critical Reviews in Food Science and Nutrition. 2009; 37–41.
28. Ramchiary N, Kehie M, Brahma V, Kumaria S, Tandon P. Application of genetics and genomics towards *Capsicum* translational research. Plant Biotechnol Rep. 2013; 1–23.

29. Guinness Book of World Records. 2006. Hottest Spice. www.guinnessworldrecords.com.
30. Dubey RK, Singh V, Upadhyay G, Pandey AK, Prakash D. Assessment of phytochemical composition and antioxidant potential in some indigenous chilli genotypes from North East India. Food Chem. Elsevier Ltd; 2015; 188: 119–125.
31. Kehie M, Kumaria S, Devi KS, Tandon P. Genetic diversity and molecular evolution of Naga King Chili inferred from internal transcribed spacer sequence of nuclear ribosomal DNA. Meta Gene. The Authors; 2016; 7: 56–63.
32. Islam MA, Sharma SS, Sinha P, Negi MS, Neog B, Tripathi SB. Variability in capsaicinoid content in different landraces of *Capsicum* cultivated in north-eastern India. Sci Hortic (Amsterdam). Elsevier B.V.; 2015; 183: 66–71.
33. Chinn MS, Sharma-Shivappa RR, Cotter JL. Solvent extraction and quantification of capsaicinoids from *Capsicum chinense*. Food Bioprod Process. Institution of Chemical Engineers; 2011; 89: 340–345.
34. Todd PH, Bensinger MG, Biftu T. Determination of Pungency Due to Capsicum By Gas-Liquid Chromatography. J Food Sci. 1977; 42: 660–665.
35. Abraham SK, Khandelwal N. Ascorbic acid and dietary polyphenol combinations protect against genotoxic damage induced in mice by endogenous nitrosation. Mutat Res—Genet Toxicol Environ Mutagen. Elsevier B.V.; 2013; 757: 167–172.
36. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and $2^{-\Delta\Delta c}$T Methods. 2001; 25: 402–408.
37. Perucka I, Oleszek W. Extraction and determination of capsaicinoids in fruit of hot pepper *Capsicum annuum* L. by spectrophotometry and high-performance liquid chromatography. Food Chem. 2000; 71: 287–291.
38. Rodriguez-Uribe L, Guzman I, Rajapakse W, Richins RD, O'Connell MA. Carotenoid accumulation in orange-pigmented *Capsicum annuum* fruit, regulated at multiple levels. J Exp Bot. 2012; 63: 517–526. doi: 10.1093/jxb/err302 PMID: 21948863
39. Perucka I, Materska M. Phenylalanine ammonia-lyase and antioxidant activities of lipophilic fraction of fresh pepper fruits *Capsicum annum* L. Innov Food Sci Emerg Technol. 2001; 2: 189–192.
40. Topuz A, Ozdemir F. Assessment of carotenoids, capsaicinoids and ascorbic acid composition of some selected pepper cultivars (*Capsicum annuum* L.) grown in Turkey. J Food Compos Anal. 2007; 20: 596–602.
41. Dyah J, Oen S, Azizahwati LH, Winarno FG. Capsaicin content of various varieties of Indonesian chillies. Asia Pacific. J. Clin.Nutr. 1997; 6: 99–101.
42. Contreras-Padilla M, Yahia E. M.. Changes in capsaicinoids during development, maturation, and senescence of chile peppers and relation with peroxidase activity. J Agric Food Chem. 1998; 46: 2075–2079.
43. Mathur R, Dangi R., Dass S., Malhotra R. The hottest chilli variety in India. Curr Sci. 2000; 79: 287–288.
44. Gnayfeed MH, Daood HG, Biacs PA, Alcaraz CF. Content of bioactive compounds in pungent spice red pepper (paprika) as affected by ripening and genotype. J Sci Food Agric. 2001; 81: 1580–1585.
45. Antonious GF, Jarret RL. Screening *Capsicum* accessions for capsaicinoids content. J Environ Sci Health B. 2006; 41: 717–29. doi: 10.1080/03601230600701908 PMID: 16785178
46. Sanatombi K, Sharma GJ. Capsaicin Content and Pungency of Different *Capsicum* spp. Cultivars. Not Bot Horti Agrobot Cluj-Napoca. 2008; 36: 89–90.
47. Canto-Flick A, Balam-Uc E, Bello-Bello JJ, Lecona-Guzman C, Solis-Marroquin D, Aviles-Vinas S, et al. Capsaicinoids content in Habanero pepper (*Capsicum chinense* Jacq.): Hottest known cultivars. HortScience. 2008; 43: 1344–1349.
48. Nwokem CO, Agbaji EB, Kagbu J a, Ekanem EJ. Determination of Capsaicin Content and Pungency Level of Five Different Peppers Grown in Nigeria. New York Sci J. 2010; 3: 17–21.
49. Giuffrida D, Dugo P, Torre G, Bignardi C, Cavazza A, Corradini C, et al. Characterization of 12 *Capsicum* varieties by evaluation of their carotenoid profile and pungency determination. Food Chem. Elsevier Ltd; 2013; 140: 794–802.
50. Sora GTS, Haminiuk CWI, da Silva MV, Zielinski AAF, Gonçalves GA, Bracht A, et al. A comparative study of the capsaicinoid and phenolic contents and in vitro antioxidant activities of the peppers of the genus *Capsicum*: an application of chemometrics. J Food Sci Technol. 2015; 52: 8086–8094. doi: 10.1007/s13197-015-1935-8 PMID: 26604381
51. Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR. Gas chromatography mass spectrometry–based metabolite profiling in plants. Nat Protoc. 2006; 1: 387–396. doi: 10.1038/nprot.2006.59 PMID: 17406261

52. Koek MM, Bakels F, Engel W, Van Den Maagdenberg A, Ferrari MD, Coulier L, et al. Metabolic profiling of ultra small sample volumes with GC/MS: From microliter to nanoliter samples. Anal Chem. 2010; 82: 156–162. doi: 10.1021/ac9015787 PMID: 19947586

53. Pino J, Sauri-Duch E, Marbot R. Changes in volatile compounds of Habanero chile pepper (*Capsicum chinense* Jack. cv. Habanero) at two ripening stages. Food Chem. 2006; 94: 394–398.

54. Ziino M, Condurso C, Romeo V, Tripodi G, Verzera A. Volatile compounds and capsaicinoid content of fresh hot peppers (*Capsicum annuum* L.) of different Calabrian varieties. J Sci Food Agric. 2009; 89: 774–780.

55. Kobata K, Saito K, Tate H, Nashimoto AKI, Okuda H, Takemura I, et al. Long-chain N-vanillyl-acylamides from capsicum oleoresin. J Agric Food Chem. 2010; 58: 3627–3631. doi: 10.1021/jf904280z PMID: 20192218

56. Keum YS, Park HW, Song H-H, Kim B-D, Kang B-C, Kim J-H. Metabolite analysis of long chain branched fatty acids and capsaicin biosynthesis in *Capsicum annuum* placenta. J Korean Soc Appl Biol Chem. 2012; 55: 189–195.

57. Wahyuni Y, Ballester AR, Sudarmonowati E, Bino RJ, Bovy AG. Metabolite biodiversity in pepper (Capsicum) fruits of thirty-two diverse accessions: Variation in health-related compounds and implications for breeding. Phytochemistry. Elsevier Ltd; 2011; 72: 1358–1370.

58. Ogawa K, Murota K, Shimura H, Furuya M, Togawa Y, Matsumura T, et al. Evidence of capsaicin synthase activity of the Pun1-encoded protein and its role as a determinant of capsaicinoid accumulation in pepper. BMC Plant Biol. 2015; 15: 93. doi: 10.1186/s12870-015-0476-7 PMID: 25884984

59. Lang Y, Kisaka H, Sugiyama R, Nomura K, Morita A, Watanabe T, et al. Functional loss of pAMT results in biosynthesis of capsinoids, capsaicinoid analogs, in *Capsicum annuum* cv. CH-19 Sweet. The Plant Journal. 2009 Sep 1; 59:953–961. doi: 10.1111/j.1365-313X.2009.03921.x PMID: 19473323

60. Del Rosario Abraham-Juarez M, Del Carmen Rocha-Granados M, Lopez MG, Rivera-Bustamante RF, Ochoa-Alejo N. Virus-induced silencing of *Comt*, *pAmt* and *Kas* genes results in a reduction of capsaicinoid accumulation in chili pepper fruits. Planta. 2008; 227: 681–695. doi: 10.1007/s00425-007-0651-7 PMID: 17999078

Chapter 9

Reaping the Benefits of Next-generation Sequencing Technologies for Crop Improvement — Solanaceae

Sushil Satish Chhapekar, Rashmi Gaur, Ajay Kumar and Nirala Ramchiary

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/61656

Abstract

Next-generation sequencing (NGS) technologies make possible the sequencing of the whole genome of a species decoding a complete gene catalogue and transcriptome to allow the study of expression pattern of entire genes. The huge data generated through whole genome and transcriptome sequencing not only provide a basis to study variation at gene sequence (such as single-nucleotide polymorphism and InDels) and expression level but also help to understand the evolutionary relationship between different crop species. Furthermore, NGS technologies have made possible the quick correlations of phenotypes with genotypes in different crop species, thereby increasing the precision of crop improvement. The Solanaceae family represents the third most economically important family after grasses and legumes due to high nutritional components. The current advances in NGS technology and their application in Solanaceae crops made several progresses in the identification of genes responsible for economically important traits, development of molecular markers, and understanding the genome organization and evolution in Solanaceae crops. The combination of high-throughput NGS technologies with conventional crop breeding has been shown to be promising in the Solanaceae translational genomics research. As a result, NGS technologies has been seen to be adopted in a large scale to study the molecular basis of fruit and tuber development, disease resistance, and increasing quantity and quality of crop production.

Keywords: Solanaceae, NGS, capsicum, eggplant, tomato, potato

1. Introduction

In developing countries, "population" and "food security" are the two major issues. These problems get worse with the sudden climate changes that hamper production, yield, and quality of food crops. Therefore, to keep in mind the food security for billions of peoples, an initiative is required for improving the quality and yield of important crops. Several traditional

2016, Biotechnological Tools for Genetic Resources *Pages* **233–274**
Editor: **Pradip Chandra Deka**
Published by: **Daya Publishing House, NEW DELHI**

Chapter 8

Advances in Molecular Breeding of Capsicum Species

Sushil Chhapekar, Mechuselie Kehie and Nirala Ramchiary*

School of Life Sciences, Jawaharlal Nehru University, New Delhi – 110 067
****E-mail: nrudsc@gmail.com***

ABSTRACT

Capsicum species commonly known as Chilli pepper is one of the world's most important vegetable and spice crop. India is the world's largest producer, consumer and exporter of chilli pepper. Despite comprising 38 species, only six species of Capsicum namely, Capsicum annuum, C. assamicum, C. baccatum, C. frutescence, C. chinense and C. pubescens are cultivated. They are used as vegetables, spices, coloring agent and folk remedies (for diseases), proving their diverse role in human's life. Chilli pepper in mainly known for its pungency characteristics. The presence of capsaicinoids, a pungency character classifies pepper into sweet peppers and hot peppers. The molecular markers, quantitative trait loci developed by conventional breeding and classical genetic analysis have revealed number of genes for major and important traits. While, the newly developed next generation sequencing and genotyping technologies have been able to generate large scale genomic resources and to find the molecular basis of economically important traits in Capsicum genome. This chapter summarizes the advances in genomics and their applications in the form of generation of resources, comparative mapping, and identification of candidate genes for fruit trait, pungency, male sterility, disease resistance, viral resistance, nematode resistance and abiotic stress tolerance. In summary, we propose the integrated use of genomics and breeding approach in Capsicum translational research to enhance the crop productivity.

Keywords: *Capsicum, Biotic stress, Abiotic stress, Genomics, Transcriptome, Molecular markers.*

Introduction

Chilli or Capsicum is one of the world's most important vegetable and spice