Industrial Crops & Products 154 (2020) 112687

Contents lists available at ScienceDirect

Industrial Crops & Products

journal homepage: www.elsevier.com/locate/indcrop

Transcriptome profiling, simple sequence repeat markers development and genetic diversity analysis of potential industrial crops *Capsicum chinense* and *C. frutescens* of Northeast India

Sushil Satish Chhapekar[a,1,2], Vijaya Brahma[b,1], Abdul Rawoof[a,1], Nitin Kumar[a,c], Rashmi Gaur[a], Vandana Jaiswal[a,3], Ajay Kumar[a,d], Satish K. Yadava[e], Rajnish Kumar[a], Vineet Sharma[a], Salga S. Babu[a], Nirala Ramchiary[a,f,*]

[a] Translational and Evolutionary Genomics Lab, School of Life Sciences, Jawaharlal Nehru University, New Delhi, 110067, India
[b] School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, 110067, India
[c] Department of Bioengineering and Technology, Gauhati University, Gopinath Bordoloi Nagar, Guwahati 781014, Assam, India
[d] Department of Plant Science, School of Biological Sciences, Central University of Kerala, Kasaragod 671316, India
[e] Department of Genetics, Delhi University South Campus, Benito Juarez Road, New Delhi 110021, India
[f] Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi 110042, India

ARTICLE INFO

Keywords:
Capsicum
C. chinense
C. frutescens
Genic SSRs
Transcriptome
Genetic diversity

ABSTRACT

Capsaicinoids, a group of secondary metabolites (alkaloids) which imparts hot property, are unique to the *Capsicum* fruits and have high value in the pharmaceutical and food industries. In spite of having high capsaicinoids content with ethnopharmacological and economic importance, the *C. chinense* and *C. frutescens* of the Northeast India, are underexplored for the genetics and breeding programs. In this study, we developed transcriptome-based Simple sequence repeat (SSR) markers and analyzed the expression(s) of capsaicinoids and carotenoid biosynthesis pathway genes in flower and fruit tissues at three developmental stages (early, breaker and mature) of *C. chinense*, *C. frutescens* and *C. annuum*. A total of 4988 and 4781 genic SSRs were identified from 123,118 and 121,017 *de novo* assembled unigenes, in *C. chinense* and *C. frutescens*, respectively. Also, the reference-based transcriptome analysis identified 70.8–73.6 % commonly expressed transcripts in all tissues besides 2929, 1327, 1193 and 937 unique transcripts in flower, early, breaker and mature fruits, respectively. The gene expression profiling showed significantly high expression of the key capsaicinoids and carotenoid biosynthetic pathway genes at breaker and mature fruit stages. Furthermore, *in silico* analysis identified a total of 335 polymorphic SSRs between *C. chinense* and *C. frutescens* with allelic size difference of > 4 bases. The trinucleotide repeats was found to be predominant in both *C. chinense* (50.6 %) and *C. frutescens* (52.2 %). High genetic-diversity of SSRs was observed with mean gene diversity of 0.51 and polymorphism information content (PIC) of 0.53. Based on twenty polymorphic SSRs, the UPGMA cluster analysis differentiated 96 genotypes belonging to *C. chinense*, *C. frutescens* and *C. annuum* into distinct groups. The identified SSRs and diverse *Capsicum* genotypes will serve as important genetic resources for future applications in genetics study and breeding of *Capsicum* varieties with improved metabolites (pungency, carotenoids etc.) and agronomic traits (fruit shape, size etc.) for agricultural, food and pharmaceutical industries.

1. Introduction

The development of advanced high throughput Next Generation Sequencing (NGS) technologies with the ability to produce a large transcriptome and genome sequence data have substantially expedited the identification and application of genic simple sequence repeat markers (Zalapa et al., 2012) in many model and non-model plant species including soybean (Li et al., 2010), *Brassica rapa* (Ramchiary

* Corresponding author at: Translational and Evolutionary Genomics Lab, School of Life Sciences, Jawaharlal Nehru University, New Delhi, 110067, India.
E-mail address: nrudsc@gmail.com (N. Ramchiary).
[1] These authors contributed equally to this work.
[2] Present address: Department of Horticulture, Chungnam National University, Daejeon 34134, South Korea.
[3] Present address: Biotechnology Division, CSIR-Institute of Himalayan Bioresource and Technology, Palampur, Himachal Pradhesh, India

https://doi.org/10.1016/j.indcrop.2020.112687
Received 18 February 2020; Received in revised form 4 May 2020; Accepted 8 June 2020

et al., 2011), sweet potato (Wang et al., 2011), potato (Dutta et al., 2011), pineapple (Wöhrmann and Weising, 2011), sorghum (Reddy et al., 2011), faba bean (Gong et al., 2011; El-Rodeny et al., 2014), chickpea (Parida et al., 2015), ragweed (Meyer et al., 2017), *Rhododendron* L. (Zhang et al., 2017), common centaury (Banjanac et al., 2018), and *Parrotia subaequalis* (Zhang et al., 2019). The simple sequence repeats (SSRs) are the most preferred markers for genetic mapping as they are easy to genotype, are cost effective and co-dominant as well as widely abundant spanning both the coding as well as the non-coding regions of the genome. Because of their multi-allelic, highly polymorphic, and reproducible nature, SSRs have been extensively used in plant genetics and breeding studies (Varshney et al., 2005; Yang et al., 2011; Singh et al., 2013; Parida et al., 2015; Saha et al., 2017). Furthermore, genic SSRs designed from the candidate genes are potentially beneficial over genomic SSR as they promise higher cross-transferability across species as the coding regions are relatively better conserved across different species (Varshney et al., 2005).

Chili peppers (*Capsicum* spp.), belonging to Solanaceae family, are the world's most extensively cultivated vegetable and spice crops (Bosland and Votava, 2012). The pepper fruits show remarkable diversity and are consumed in every part of the world. Apart from that, *Capsicum* fruits are a worthy source of several nutritional and dietary antioxidant compounds including capsaicinoids, vitamins (A & C), pigments, minerals, and essential oils (Sarpras et al., 2016). An ethno medicinal survey in India reported the use of *C. frutescens* fruits in soothing waist pain (Rout and Panda, 2010). The fruit tissues of *Capsicum* have also been used in treating several skin disorders, dog/snake bite and wounds due to their antimicrobial activities (Meghvansi et al., 2010). The flavonoids, peptides, polyphenols, vitamins and alkaloids, possess various beneficial properties, such as anti-diabetic (Okumura et al., 2012) anti-inflammatory (Spiller et al., 2008), antioxidant (Hernández-Ortega et al., 2012), anticancer (Ullah et al., 2014; Singh et al., 2017), antimicrobial (Careaga et al., 2003) and cardiovascular properties (Arora et al., 2011). The fleshy *C. chinense* fruits are also being used as a source of coloring agent in the food industry (Ramchiary et al., 2013). The presence of high amount of capsaicinoids (pungency property) and coloring agents generated high demand for chili pepper fruits in the agricultural, pharmaceutical, food industries (Baenas et al., 2019). The indigenous people of Northeast India have been using *Capsicum* (Bhut jolokia or Ghost chilli belonging to *Capsicum chinense* and Kon jolokia belonging to *C. frutescens*) fruits as a traditional remedy for treatment of different human ailments including headache, fever, cold coughs, gastritis, ankylosing, spondylitis, rheumatism, indigestion etc.(Meghvansi et al., 2010; Umashanker and Shruti, 2011; Haanpää and Treede, 2012).

In *Capsicum*, the development of SSR markers have been reported mainly for the *C. annuum* (Ramchiary et al., 2013; Shirasawa et al., 2014; Ibarra-Torres et al., 2015; Cheng et al., 2016; Chhapekar, 2016; Zhang et al., 2016; Dubey et al., 2019). The developed SSR markers were used for genetic mapping, QTLs identification and gene discovery for disease resistance and other economically important traits (Portis et al., 2007; Stágel et al., 2009; Pacheco-Olvera et al., 2012; Carvalho et al., 2015). Furthermore, several genomics and transcriptomics based studies have been mostly performed in *C. annuum* compared with *C. chinense* and *C. frutescens* (Ashrafi et al., 2012; Kim et al., 2014; Martínez-López et al., 2014; Qin et al., 2014; Liu et al., 2017; Chhapekar et al., 2018; Kim et al., 2018). Martínez-López et al. (2014) investigated transcriptome of pepper fruits at different developmental stages and found that the genes related to the synthesis of capsaicin and vitamin C were significantly up-regulated at immature green stage. Recently, Liu et al. (2017) generated an extensive data set of transcriptomes for an elite pepper breeding line 6421 and developed a public data platform named "pepper- Hub". Kim et al. (2018) studied the global gene expression profiles and identified the transcriptome landscapes of fruit development in *C. annuum*. They also studied the expression of genes in leaves infected with pathogens (*Phytophthora infestans*, *pepper mottle virus* and *Tobacco mosaic virus*) and identified molecular networks of gene(s) expression in response to the pathogen infection. In a recent study, Guzmán et al. (2020) developed a set of 21 SSR markers to distinguish 42 *Capsicum* genotypes of 11 *Capsicum* species for efficient assessment of molecular variability in *Capsicum*. However, there is very limited information in *C. chinense* and *C. frutescens*. In *C. frutescens* putative SSR markers were identified but their validation and characterization was not reported (Liu et al., 2013). In the *C. chinense* genome, the development of trinucleotide SSRs markers were reported (Uncu, 2019). In our previous study we developed a total of 623 non coding RNA based SSRs which includes 119 microRNASSRs (miRNASSRs) and 504 long non-coding RNASSRs (lncRNASSRs) mainly from *C. chinense* and *C. frutescens* (Jaiswal et al., 2020). Thus, limited information in *C. chinense* and *C. frutescens* suggests the urgent need to explore gene/transcript expression during fruit development and for the development of genetic and genomic resources for the improvement of these two *Capsicum* crops.

The International Bureau of Plant Genetic Resources (IBPGR, 1983), identified Northeast India as the second center of origin for *Capsicum* species after South America (Islam et al., 2016). Bhut jolokia, with its unique fiery pungent phenotypes was reported as the naturally occurring highest pungency containing chili pepper in Guinness Book of World Records (2006). Bosland and Baral (2007), described it as a cultivar of *C. chinense* possibly resulting from natural hybridization between *C. chinense* and *C. frutescens* with more genome complement from the former. However, study by Purkayastha et al. (2012a, 2012b) using internal transcript spacer DNA sequences (ITS1 and ITS2) and 5.8 rRNA gene observed a distinct clade for all the accessions of 'Bhut jolokia' from the *C. annuum, C. chinense* and *C. frutescens* genotypes. They also observed distinct proteome and morpho-agronomic traits suggesting it to be a distinct species and re-named it as *C. assamicum*. However, in this study, we considered Bhut jolokia under *C. chinense* and hereafter it is referred to as *C. chinense*. Furthermore, Northeast India is also bestowed with other *Capsicum* species such as very small upright fruit with medium pungency belonging *C. frutescens* {a type of chili peppers commonly known as Kon jolokia (Kon means small), Mem jolokia, or bird eye chilli} and *C. annuum* with low pungent genotypes (Sarpras et al., 2016; Dutta et al., 2017). In our earlier study, we reported the presence of wide morphological and metabolite diversity in landraces belonging to these two *Capsicum* species from Northeast India (Therefore, Bhut jolokia and Kon jolokia are not individual cultivars, but names of type of chilli peppers consisting of many cultivars, Sarpras et al., 2016).

Bhut jolokia and Kon jolokia (Hereafter mentioned as *C. chinense* and *C. frutescens* respectively for uniformity) constitute relatively underexplored, yet highly useful *Capsicum* genetic resources of Northeast India. In this study we developed genic SSRs and used them in diversity analysis to identify the diverse genotypes belonging to *C. annuum, C. chinense* and *C. frutescens* which could be important genetic stocks with distinct fruit morphology and metabolite content for the future breeding programs. The new genomic resources such as transcriptome sequences and genic SSR markers would supplement the existing marker repertoire of *Capsicum*, which could be used in molecular breeding and for further *Capsicum* genetics studies. Moreover, the information generated on intra and inter-specific genetic relationships would aid germplasm conservation and utilization in the genetic improvement of *C. chinense* (Bhut jolokia) and *C. frutescens* (Kon jolokia), the two potential crops important for pharmaceutical and food industries.

2. Materials and methods

2.1. RNA extraction, transcriptome sequencing and assembly

Plants belonging to three *Capsicum* species i.e. *C. chinense* (Bhut jolokia Accession No. 17 in Table 1), *C. frutescens* (Kon jolokia

Table 1
Details of the chilli pepper accessions, their origin and morpho-agronomic characteristics.

S. No.	Accession No.	Species	Source of origin (Indian States)	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	Low pungent	Pendant	Elongated	Dark red	Pointed	6	3.5	40	0.322
2	Acc 2	*C. annuum*	NBPGR	Low pungent	Pendant	Elongated	Red	Pointed	5	2.5	78	0.382
3	Acc 3	*C. annuum*	NBPGR	Low pungent	Pendant	Elongated	Light red	Pointed	5.5	4	67	0.422
4	Acc 4	*C. annuum*	NBPGR	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	6.7	3.6	60	0.047
5	Acc 5	*C. chinense*	NBPGR	Highly pungent	Pendant	Elongated	Dark red at maturity	Blunt	5.8	2.9	66	0.036
6	Acc 6	*C. annuum*	NBPGR	Low pungent	Erect/ Upward	Block shaped	Green in immature to dark red in mature	Blunt	6	1.8	42	0.055
7	Acc 7	OCS	NBPGR	Low pungent	Pendant	Triangular	Dark red	Pointed	4.5	3.8	26	0.028
8	Acc 8	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Dark red	Blunt	4.5	1.3	53	0.037
9	Acc 9	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Dark red	Pointed	4.1	1.5	43	0.039
10	Acc 10	*C. annuum*	NBPGR	Low pungent	Pendant	Elongated	Light red	Blunt	6	4.5	41	0.0389
11	Acc 11	*C. chinense*	NBPGR	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	Moderately pungent	Erect	Elongated	Green in immature to red in mature	Pointed	0.87	0.8	13	0.044
13	Acc 13	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	4.6	1.5	25	0.039
14	Acc 14	*C. annuum*	NBPGR	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	4.5	2.1	68	0.0283
15	Acc 15	*C. annuum*	Assam	Low pungent	Pendant	Almost round	Dark red	Blunt	3.4	1.5	31	0.0425
16	Acc 16	*C. annuum*	J&K	Low pungent	Pendant	Elongated	Dark red	Pointed	5	4	70	0.026
17	Acc 17	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Orangish red	Pointed	3.4	2.1	16	0.036
18	Acc 18	*C. chinense*	Assam	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	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*	DrAlok_JNU-20191218T100220Z-002	Highly pungent	Pendant	Triangular	Chocolate	Pointed	6	1.68	12	0.025
21	Acc 21	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Orangish red	Pointed	6.5	0.368	61	0.033
22	Acc 22	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Orangish red	Pointed	5.2	0.19	10	0.036
23	Acc 23	*C. frutescens*	Manipur	Moderately pungent	Erect/ upward	Elongated	Green in immature turns to red in mature	Pointed	5	0.8	18	0.0288
24	Acc 24	*C. frutescens*	Manipur	Moderately pungent	Erect/ upward	Elongated	Green in immature turns to red in mature	Pointed	0.925	0.115	8	0.044
25	Acc 25	*C. annuum*	Nagaland	Low pungent	Pendant	Elongated	Green in immature turns to red in mature	Blunt	4.5	1.3	53	0.037
26	Acc 26	*C. frutescens*	Nagaland	Moderately pungent	Pendant	Elongated	Yellowish red in mature	Pointed	3.4	1	28	0.044
27	Acc 27	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	3.2	1.6	59	0.035
28	Acc 28	*C. frutescens*	Assam	Moderately pungent	Erect/ upward	Elongated	Dark red at maturity	Pointed	4.5	1.7	23	0.044
29	Acc 29	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Dark red at maturity	Pointed	3.6	1	38	0.03
30	Acc 30	*C. annuum*	Assam	Low pungent	Pendant	Triangular	Red	Pointed	3	1.2	63	0.037
31	Acc 31	*C. annuum*	Nagaland	Low pungent	Pendant	Block shaped	Red	Blunt	4.5	1.8	39	0.027
32	Acc 32	*C. frutescens*	Nagaland	Moderately pungent	Erect/ upward	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	Highly pungent	Pendant	Elongated	Red	Pointed	3.5	1.1	21	0.041
34	Acc 34	*C. chinense*	Assam	Highly pungent	Pendant	Elongated	Light green	Pointed	3.8	1.6	12	0.03
35	Acc 35	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Red	Pointed	3.8	1	37	0.038
36	Acc 36	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Light red	Pointed	3.8	1.2	46	0.028
37	Acc 37	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Light to dark red	Pointed	4.8	1.9	11	0.023
38	Acc 38	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Dark red	Pointed	3.3	1.8	35	0.041
39	Acc 39	*C. annuum*	New Delhi	Low pungent	Pendant	Block shaped	Red at maturity	Sunken	6.5	3.1	41	0.51
40	Acc 40	*C. annuum*	New Delhi	Low pungent	Pendant	Elongated	Yellow at maturity	Pointed	4.5	3.5	37	0.0568
41	Acc 41	*C. annuum*	New Delhi	Low pungent	Pendant	Elongated	Red at maturity	Pointed	2.9	4.6	25	0.0618

(*continued on next page*)

Table 1 (*continued*)

S. No.	Accession No.	Species	Source of origin (Indian States)	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)
42	Acc 42	*C. annuum*	New Delhi	Low pungent	Pendant	Block shaped	Red	Sunken	3.8	3.6	28	0.041
43	Acc 43	*C. annuum*	New Delhi	Low pungent	Pendant	Block shaped	Yellow	Sunken	7.1	2.5	35	0.0768
44	Acc 44	*C. annuum*	New Delhi	Low pungent	Pendant	Block shaped	Bright yellow in mature	Sunken	5.5	2.8	24	0.0768
45	Acc 45	*C. annuum*	New Delhi	Low pungent	Pendant	Elongated	Lemon yellow	Pointed	8	6.5	95	0.077
46	Acc 46	*C. chinense*	New Delhi	Low pungent	Pendant	Triangular	Red at maturity	Pointed	4.5	3.8	21	0.032
47	Acc 47	*C. annuum*	Manipur	Low pungent	Pendant	Campanulate	Light orange at maturity	Pointed	5.5	3.2	23	0.027
48	Acc 48	*C. annuum*	Manipur	Low pungent	Pendant	Campanulate	Light red at maturity	Pointed	7	2.2	66	0.047
49	Acc 49	OCS	Assam	Low pungent	Erect	Elongated	Red	Blunt	1.5	0.9	40	0.04
50	Acc 50	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Red	Blunt	2.6	1	41	0.037
51	Acc 51	OCS	J&K	Low pungent	Pendant	Almost round	Bright red	Blunt	2.1	1.6	12	0.028
52	Acc 52	*C. annuum*	Assam	Low pungent	Erect	Elongated	Red	Blunt	3	1.3	46	0.035
53	Acc 53	*C. annuum*	Mizoram	Low pungent	Pendant	Elongated	Red	Pointed	3.5	1.6	34	0.012
54	Acc 54	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Green in immature turns to purple greenish to red in mature	Blunt	1.7	1.5	57	0.04
55	Acc 55	*C. annuum*	Assam	Low pungent	Pendant	Almost round	Green in immature turns to orange in mature	Sunken	1.2	1.8	21	0.029
56	Acc 56	*C. chinense*	Nagaland	Highly pungent	Pendant	Triangular	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	Low pungent	Erect	Elongated	Dark red at maturity	Pointed	3.6	0.89	41	0.045
58	Acc 58	*C. annuum*	Nagaland	Low pungent	Erect	Elongated	Green in immature turns to red in mature	Pointed	4.8	1.5	60	0.047
59	Acc 59	*C. annuum*	Nagaland	Low pungent	Erect	Triangular	Red	Pointed	4	1.4	69	0.048
60	Acc 60	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Red	Pointed	3.2	1.6	34	0.024
61	Acc 61	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Orangish red	Pointed	4.1	3.2	15	0.048
62	Acc 62	*C. chinense*	Assam	Highly pungent	Erect/ upward	Triangular	Orange	Pointed	4	3.1	12	0.032
63	Acc 63	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Green in immature turns to red in mature	Pointed	6	2.4	36	0.042
64	Acc 64	OCS	Assam	Low pungent	Erect/ upward	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	Moderately pungent	Erect/ upward	Elongated	Green in immature turns to orange in mature	Pointed	2.8	0.6	23	0.256
66	Acc 66	OCS	Assam	Low pungent	Pendant	Block shaped	Red	Sunken	2.1	1.1	16	0.032
67	Acc 67	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Red	Blunt	1.8	1.8	28	0.021
68	Acc 68	*C. chinense*	Assam	Highly pungent	Pendant	Elongated	Red	Pointed	2.2	0.5	18	0.0396
69	Acc 69	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Red	Pointed	2.8	1.8	37	0.025
70	Acc 70	*C. chinense*	Manipur	Highly pungent	Pendant	Triangular	Red	Pointed	2.8	1.2	15	0.028
71	Acc 71	*C. frutescens*	Assam	Moderately pungent	Erect/ upward	Elongated	Light red in mature	Pointed	2.4	0.5	23	0.0396
72	Acc 72	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Red	Pointed	2.8	2.1	18	0.03
73	Acc 73	*C. chinense*	Nagaland	Highly pungent	Pendant	Triangular	Red	Pointed	3	1.2	34	0.027
74	Acc 74	*C. chinense*	Manipur	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	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	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	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	Highly pungent	Pendant	Triangular	Orange	Pointed	4.2	2.5	12	0.035
79	Acc 79	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Green in immature turns to red in mature	Pointed	4	2.6	14	0.035

(*continued on next page*)

Table 1 (*continued*)

S. No.	Accession No.	Species	Source of origin (Indian States)	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)
80	Acc 80	*C. chinense*	Assam	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	Highly pungent	Pendant	Triangular	Green in immature to orange red in maturity	Pointed	4.5	4.4	16	0.046
82	Acc 82	*C. annuum*	Manipur	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	Low pungent	Pendant	Elongated	Dark red	Pointed	6.1	4.9	25	0.025
84	Acc 84	*C. chinense*	Assam	Highly pungent	Pendant	Elongated	Orange	Blunt	5.1	2.9	18	0.046
85	Acc 85	*C. chinense*	Assam	Highly pungent	Pendant	Triangular	Dark chocolate at maturity	Pointed	4.2	4.1	14	0.032
86	Acc 86	*C. frutescens*	Mizoram	Moderately pungent	Pendant	Elongated	Dark red	Blunt	2.5	1	38	0.025
87	Acc 87	*C. frutescens*	Assam	Moderately pungent	Pendant	Almost round	Red	Blunt	1.8	2.5	8	0.063
88	Acc 88	*C. frutescens*	Manipur	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	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*	J&K	Low pungent	Pendant	Elongated	Red	Pointed	5	3.9	28	0.069
92	Acc 92	*C. annuum*	J&K	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	Moderately pungent	Pendant	Elongated	Red at maturity	Pointed	2.8	1.8	21	0.018
94	Acc 94	*C. annuum*	Assam	Low pungent	Pendant	Elongated	Green in immature, purplish green in breaker to red immature	Pointed	5	1.4	39	0.027
95	Acc 95	*C. annuum*	Mizoram	Low pungent	Pendant	Elongated	Red	Sunken	4.8	8.6	48	0.0478
96	Acc 96	*C. frutescens*	Mizoram	Moderately pungent	Erect/ upward	Elongated	Red	Pointed	2.1	1.4	9	0.031

(Abbreviation: NBPGR - National Bureau of Plant Genetic Resources, New Delhi; OCS: Other Capsicum Species; J&K: Jammu & Kashmir).

Accession No. 24 in Table 1) and *C. annuum* (Accession No 16 in Table 1) were grown in greenhouse of School of Life Sciences, Jawaharlal Nehru University, New Delhi, India using standard growth conditions (16 h light and 27 °C/19 °C day/night temperature cycle). Tissue samples from flower and three developmental stages i.e. early (20–25 days post anthesis; DPA), breaker (30–40 DPA) and mature (50–60 DPA) fruit for each *Capsicum* species were collected in liquid nitrogen. Total RNA was extracted and transcriptome sequencing for four aforementioned samples of each individual species, was performed as described in our earlier study (Dubey et al., 2019). Three biological replicates were pooled for each individual stage prior to sequencing. The clean reads were obtained after removing adapter sequences and poor quality reads, with phred score < Q20 from raw reads using trimgalore (v0.4.4). The raw reads are deposited in National Center for Biotechnology Information (NCBI) and the accession numbers are PRJNA327797 (*C. chinense*), PRJNA327800 (*C. frutescens*) and PRJNA505972 (*C. annuum*). The clean reads from each sample were aligned to zunla-1 (Qin et al., 2014) reference genome using TopHat (Trapnell et al., 2009) with default parameters. The aligned reads were further assembled, transcripts abundance was calculated and their expressions were identified across tissue samples of three *Capsicum* species following the pipeline described by Trapnell et al. (2012). Transcripts with fragment per kilobase of transcript per million mapped reads (FPKM) > 0.5 across the tissues were considered as expressed (Lee et al., 2019). Trinity was utilized for *de novo* transcriptome assembly using *C. chinense* and *C. frutescens* samples with default parameters (Haas et al., 2013).

2.2. Gene annotation and ontology analysis

The identified transcript using reference based assembly was annotated based on tomato orthologs. Blastx was performed against tomato protein sequences (SL3.0; ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG3.0_release/) with e-value ≤1e-10 and > 95 % percent identity. Gene ontology (GO) information of transcripts/gene showing differential expression across tissue samples was retrieved using agriGO tool (Tian et al., 2017). The hypergeometric test and false discovery rate correction were applied and GO terms with *p*-value < 0.01, and FDR < 0.05 were considered as significant terms, and were visualized using R *ggplot2* package (Wickham, 2016).

2.3. Transcriptome data validation by quantitative real-time PCR

Transcriptome/RNAseq data for 14 different genes showing differential expression across flower and three fruit developmental stages was validated using quantitative real-time PCR. The sequences of primers in this study are listed in Supplementary Table 1. Total RNA was extracted from four different tissue samples belonging to *C. chinense, C. frutescens* and *C. annuum*. One µg of total RNA was converted into cDNA using SuperScript III first-strand cDNA synthesis kit (Clontech, USA) following manufacturer's instructions. The qRT-PCR analysis was performed as described by Dubey et al. (2019) which included initial denaturation at 95 °C for 2 min, followed by 40 cycles of amplification for 15 s at 95 °C and 1 min at 60 °C. Actin was used as internal control. Analysis of relative gene expression data was done using $2^{-[\Delta\Delta Ct]}$ method (Livak and Schmittgen, 2001).

2.4. Development of SSRs in C. chinense and C. frutescens

The unigenes obtained after *de novo* assembly of transcriptome data from each *Capsicum* species were utilized for SSR identification using MIcroSAtellite (MISA) identification tool. The nucleotide stretch with di, tri, tetra, penta and hexamers with minimum six repeats were considered as SSRs and mononucleotide repeats were excluded. The successive SSRs with distance intervals less than 100 bp were considered as overlapped and therefore removed from further analysis. BatchPrimer3 v1.0 software (You et al., 2008) was used for designing primers with the following parameters: primer length of 18–24 bases with an optimum of 22 bases, GC content of 40–60 % with an optimum of 50 %, annealing temperature of 40 − 60 °C with an optimum at 50 °C, and PCR amplicon size of 100–400 bp with an optimum size of 200 bp.

Transcripts containing SSR motifs were compared against the sequences from the reference genome assembly of the *C. annuum* (Zunla-1; Qin et al., 2014) using the standalone BLASTN program. Default parameters of the program were used and the expectation value (e-value) cut-off was set at 1e-10 for sequence similarity searches. The putative gene IDs were identified and their function, KEGG orthology (KO) as well pathways were predicted using BlastKOALA (Kanehisa et al., 2016). Also, SSRs were characterized as genic and inter-genic SSRs based on their genomic distribution and structural annotations as per the annotation of *C. annuum* reference genome available at NCBI. The chromosome-wise distribution of SSRs in *C. chinense* and *C. frutescens* were depicted using circos plot (Hu et al., 2014).

2.5. In silico identification of polymorphic SSRs between C. chinense and C. frutescens

In order to discover polymorphic SSRs *in silico* analysis was performed using the transcriptome data between *C. chinense* and *C. frutescens*. Around 250 bp transcript sequences flanking with 5′ and the 3′ ends of the microsatellite repeat-motifs from *C. chinense* were extracted and then compared with the flanking sequences of assembled contigs of *C. frutescens* using BLASTN (standalone). The identified sequence matches (with percent identity > 95 % and e-value < 1e-10) to the flanking sequences of *C. chinense* displaying increase or decrease in the number of identical microsatellite repeats in the *C. frutescens* were considered as *in silico* polymorphic SSRs. Furthermore, if any insertions or deletions were found in the flanking regions of the microsatellite repeat motifs from the two species, the sequences were not classified as polymorphic and therefore were not considered for the analysis.

2.6. Validation of SSR markers across three Capsicum species with distinct agronomic traits

A set of 96 *Capsicum* genotypes belonging to *C. chinense, C. frutescens* and *C. annuum* were used in the present study for the evaluation of agronomic traits and diversity analysis using SSR markers (Fig. 1, except few genotypes, these genotypes are different from the genotypes used in Jaiswal et al., 2020). The detailed information related to the source and fruit morpho-agronomic features of these 96 *Capsicum* genotypes is provided in Table 1. Seeds of all 96 genotypes were surface sterilized with 4% sodium hypochlorite and grown in a glass house at 24 − 26 °C with 16 h light period. One-month-old plants were transferred to the research field at Jawaharlal Nehru University campus, New Delhi, India for phenotypic evaluation. Phenotypic data were recorded for the following nine agronomic traits- seven of these related to the fruit: degree of pungency, position, shape, color, shape at blossom end, length and weight; and two related to the seed-number and weight, using standard protocols as prescribed by the International Plant Genetic Resources Institute (IPGRI, 1995).

Genomic DNA of each *Capsicum* accession was extracted from young leaves using the Hexadecyl trimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1990). The quality and quantity of DNA was checked on 1 % agarose gels. Primer pairs for a total of 50 genic-SSR markers were custom synthesized and were used initially for genotyping of 10 *Capsicum* genotypes. SSRs showing polymorphisms were selected and used for genotyping of 96 *Capsicum* accessions. In a 20 µL reaction volume, each PCR reaction mixture contained 25 ng of DNA, 1X PCR buffer [50 mM KCl, 20 mM Tris-Cl (pH 8.4)], 1.5 mM $MgCl_2$, 0.125 mM of each dNTPs, 0.5 µM of each primer, and 0.5 U of Taq DNA polymerase (Life Technologies). Amplifications were performed using a touchdown profile using Thermal Cycler (Eppendorf, Germany) with an

Fig. 1. The *Capsicum* germplasm showing diversity of fruit colour, size, shape, and fruiting habit.

initial denaturation at 95 °C for 2 min, followed by 10 cycles of denaturation at 95 °C for 20 s, annealing at 65 °C for 50 s and extension at 72 °C for 50 s. During this touchdown phase the annealing temperature was decreased at a uniform rate of 1 °C per cycle from 65 to 55 °C. This was followed by another 25 cycles, each having denaturation at 95 °C for 20 s, annealing at 55 °C to 50 °C for 50 s, extension at 72 °C for 50 s and a final extension at 72 °C for 7 min. PCR products were electrophoresed on 3 % metaphor agarose gels along with size markers and stained with ethidium bromide (EtBr). The genotyping data was recorded in the format of length variation (in bases).

2.7. Diversity analysis using genic-SSR

The genotyping data of 20 selected polymorphic SSR markers was used for diversity analysis of the 96 *Capsicum* genotypes using PowerMarker version 3.23 (Liu and Muse 2005). The following diversity parameters were estimated - total number of alleles (NA), allelic frequency, major allele, accession-specific alleles, gene diversity (GD), polymorphism information content (PIC), and genetic distance between each pair of accessions. MEGA7 package (Kumar et al., 2016) was used to construct a neighbor-joining (NJ) phylogram using a distance matrix.

3. Results

3.1. Sequencing and assembly of transcripts

The dynamics of transcripts were studied at flower and three fruit developmental (early, breaker, mature) stages in *C. chinense*, *C. frutescens* and *C. annuum*. Overall a total of ~407 million clean reads of 100bp length were obtained from all samples and aligned against *C. annuum* (zunla-1) reference genome. The number of reads uniquely aligned to the reference genome ranges from approximately 23 million to ~42 million reads per sample library. An average of 88.8 %, 87.6 % and 86.5 % read mapping efficiency was obtained for *C. chinense*, *C. frutescens* and *C. annuum*, respectively (Table 2A). *De novo* assemblies using the clean reads from *C. chinense* and *C. frutescens* were performed using the Trinity software. A total of 184,975 and 179,780 transcripts including 123,118 and 121,017 unigenes were obtained from *C. chinense* and *C. frutescens* respectively. The minimum transcript length was 201 bp. The maximum length was 16,894 bp with N50 value of 1188 bp in *C. chinense,* while it was 15,670 bp with N50 value of 1148 in *C. frutescens.* The length of unigenes ranged from minimum of 201 bp and maximum of 16,894 bp with N50 value of 1569 bp in *C. chinense* and 15,670 bp with N50 value of 1573 bp in *C. frutescens,* respectively. Around 50.5–51.4 % transcripts and 63.82–65.18 % unigenes were of length < 500 bp, while ~4% transcripts and 2.6–2.8 % unigenes were of length with > 3000 bp (Table 2B).

3.2. Expression analysis and qRT-PCR validation of few genes

A total of 26,469 transcripts with FPKM > 0.5 in any one of the tissues across *Capsicum* samples were identified as expressed (Fig. 2A). The information related to these transcripts is supplied in Supplementary Table 2. Further, for each species, transcripts were grouped into 12 clusters based on their expression across four tissue samples using k-means clustering (Supplementary Fig. 1). The common and uniquely expressed transcript in each *Capsicum* species were analyzed and represented in a venn diagram. Around 70.8–73.6 % transcripts were commonly expressed in all tissue stages while a total of 2929, 1327, 1193 and 937 uniquely expressed transcripts in flower, early fruit, breaker fruit and mature fruit of three *Capsicum* species respectively was observed (Fig. 2B–D). Furthermore, the functional category of these uniquely expressed transcripts was analyzed by gene ontology (GO) enrichment analysis. The transcripts uniquely expressed in flower showed that the most significant GO terms were mainly from cellular and biological categories involved in cell periphery [GO:0071944], plasma membrane [GO:0005886] and transmembrane transport [GO:0055085], and ion transport [GO:0006811]. The GO term cysteine-type endopeptidase inhibitor activity [GO:0004869] was significantly enriched in the early fruit. The GO terms identified in the

Table 2
Summary of transcriptome sequence assembly. (A) Reference based read mapping of transcriptomes data from different tissue samples of three *Capsicum* species, (B) Quality parameters of *de novo* assembly and specific features of transcripts/unigenes in *C. chinense* and *C. frutescens*.

A

Group	Developmental stage	Left Reads	Mapped with Left	Right Reads	Mapped with Right	Overall Read Mapping	Average Read Mapping
C. chinense (Bhut jolokia)	flower	13903579	12273702	13903579	12238890	88.20 %	88.83 %
	early	12610302	11249381	12610302	11223597	89.10 %	
	breaker	15988530	14305578	15988530	14264616	89.30 %	
	mature	9764619	8675388	9764619	8650859	88.70 %	
C. frutescens	flower	8904084	7882298	8904084	7861672	88.40 %	87.58 %
	early	18672924	16559997	18672924	16519329	88.60 %	
	breaker	11708070	9694396	11708070	9669532	82.70%	
	mature	11163055	10127660	11163055	10092809	90.60 %	
C. annuum	flower	23705444	21161845	23705444	20379081	87.60 %	86.55 %
	early	28568535	25073140	28568535	24196770	86.20 %	
	breaker	23151739	20375620	23151739	19320667	86.20 %	
	mature	25637874	20375620	25637874	19541841	86.20 %	

B

	C. chinense		*C. frutescens*	
length (bp)	transcript	unigene	transcript	unigene
200 – 500	93496	78580	92364	78884
501 – 1000	37680	21984	35507	20796
1001 – 1500	19591	8187	18544	7698
1501 – 2000	13156	5319	12970	5078
2001 – 2500	8281	3452	8276	3362
2501 – 3000	5043	2147	4922	2064
> 3000	7728	3449	7197	3135
Total	184975	123118	179780	121017
N 50 (bp)	1188	1569	1148	1573
Average length (bp)	696.72	896.16	678.94	886.77
Median length (bp)	369	492	361	481
Min length (bp)	201	201	201	201
Max length (bp)	16894	16894	15670	15670
% GC	38.6	38.6	38.86	38.86

breaker specific transcripts were those involved in ADP binding [GO:0043531], lignin catabolic process [GO:0046274], phenylpropanoid catabolic process [GO:0046271], and nutrient reservoir activity [GO:0045735]. The transcripts specific to the mature fruit stage showed the molecular function related to monooxygenase activity [GO:0004497] (Fig. 3A–D; Supplementary Table 3). We investigated the expression profile of capsaicinoids and carotenoid biosynthesis pathway genes in all three *Capsicum* species. Significantly high expression of *pAMT* (putative aminotransferase), *AT3* (acyltransferase) and *Kas* (ketoacyl-ACP synthase) genes was observed in the breaker and mature fruit stages of *C. chinense* and *C. frutescens* compared with *C. annuum* (Fig. 4A). Other genes such as *BCAT* (branched-chain amino acid aminotransferase), *PAL* (phenylalanine ammonia-lyase) and *C4H* (cinnamate 4-hydroxylase) showed considerably high expression in *C. chinense* and *C. frutescens*. While in carotenoid biosynthesis pathway, the high expression of *Capana06g002492 (CrtR-b1; beta-carotene hydroxylase), Capana11g001999 (PSY2; phytoene synthase 2)*, *Capana08g001316 (ZDS; zeta-carotene desaturase) and Capana03g000054 (PDS; 15-cis-phytoene desaturase)* genes, was observed in all the three *Capsicum* species but were more abundant in *C. annuum* as compared with the other two species (Fig. 4B). Also some genes such as *Capana00g003114 (NCED1; nine-cis-epoxycarotenoid dioxygenase), Capana01g000984 (CYP707A2; abscisic acid 8′-hydroxylase), Capana08g001316 (ZDS), Capana04g002519 (PSY1; phytoene synthase 1),* and *Capana03g000054 (PDS)* were highly expressed in the breaker and mature stages of *C. annuum* followed by *C. chinense* and *C. frutescens*.

Furthermore, to validate the transcriptome data a set of 14 genes/transcripts showing differential expression in the flower and three developmental stages of the fruit tissues were selected. The correlation graph between $\log_2$ ratios from RNAseq and qRT-PCR data showed the consistency of gene expression from two methods (Fig. 5). The genes/transcripts such as *Capana08g000016* (*L-ascorbate oxidase homolog*), *Capana04g001678*, *Capana12g000154* (*pectin methylesterase*), *Capana06g000820* (*purple acid phosphatase 17-like*) and *Capana09g002437* (*LIN7; cell-wall invertase*) showed flower specific expression while the transcripts/genes such as *Capana04g000478* (*lysine histidine transporter-like 8*), *Capana06g000967* (*nucleobase-ascorbate transporter 6 isoform X1*), *Capana09g001131* (*patellin-6*) and *Capana06g001096* (*aspartyl protease family protein At5g10770-like*) showed early specific expression across *Capsicum* species. While the remaining transcripts/genes such as *Capana08g000223* (*MLO2; MLO family protein 2*), *Capana08g00245* (*MKS1b; methylketone synthase Ib*), *Capana07g001537* (*scarecrow-like protein 32*), *Capana04g000279* (*PAP1; Purple acid phosphatase*) and *Capana00g002265* (*alpha/beta-hydrolases superfamily protein precursor*) showed breaker and mature specific expression across the *Capsicum* samples. The gene ontology and tomato orthologs of these genes are represented in Supplementary Table 1.

3.3. SSR development and distribution in Capsicum genome

For identification of SSR motifs, a total of 184,975 (~166 Mb) sequences of *C. chinense* and 179,780 (~159 Mb) sequences of *C. frutescens* were examined in the MIcroSAtellite (MISA) identification tool. A total of 49,136 transcripts in *C. chinense* and 46,771 in *C. frutescens* harboring SSRs were identified. In this study, mono-nucleotide SSRs were not included. There was approximately one SSR locus at every 7.52 kb in *C. chinense* and 7.42 kb in the *C. frutescens* genomes. After filtration, a total of 12,473 and 11,835 SSR motifs containing sequences with a total of 898 and 862 compound SSRs were identified in *C. chinense* and *C. frutescens,* respectively.

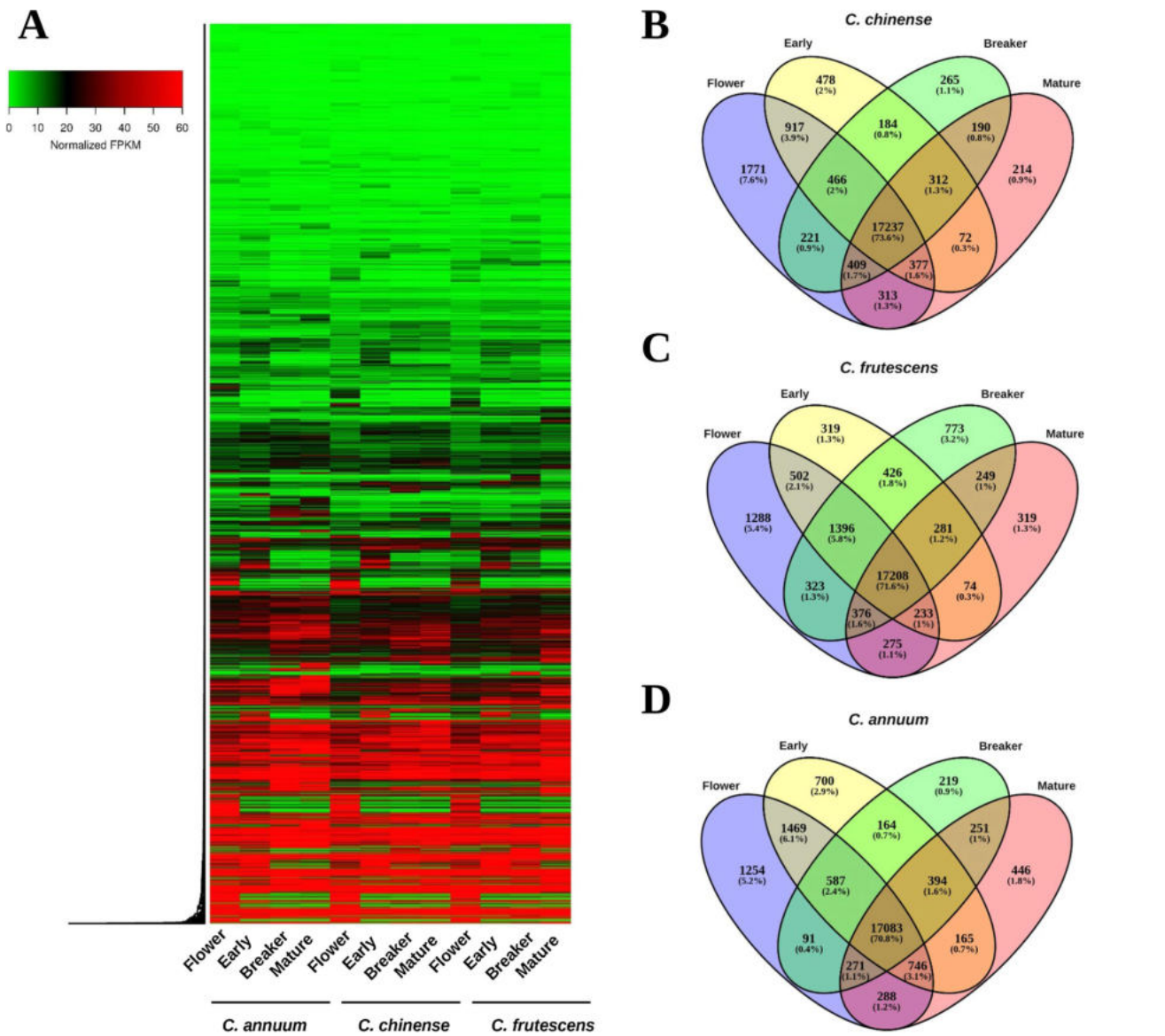

Fig. 2. Expression of genes/transcripts in different tissues belonging to three Capsicum species i.e. *C. annuum*, *C. chinense* and *C. frutescens*. (A) Heatmap of expressed genes across tissues of three *Capsicum* species with FPKM > 0.5, (B-D) Common and unique genes expressed across flower and fruit developmental stages in *C. chinense*, *C. frutescens* and *C. annuum,* respectively.

The distribution of different repeat classes of identified SSRs in the two *Capsicum* genomes is given in Table 3. Among the identified SSR motifs, the 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 with a minimum of five reiterations (cut off limit) were the most abundant. We also observed an inverse relationship between the number and frequency of repeat units of SSRs. Motifs showing more than ten reiterations were rare with a frequency of < 1 %.

Primers were designed using the 250 bp flanking bases of SSR motifs to amplify a total of 4988 and 4781 SSRs in *C. chinense* and *C. frutescens*, respectively (Supplementary Tables 4–5). The distribution of these SSRs on the twelve chromosomes of *C. annuum* reference genome is represented in the form of a circular plot showing the high density of trinucleotide SSRs followed by di-, tetra-, penta- and hexa nucleotide SSRs, respectively (Fig. 6A and 6B). The frequencies of abundant di- (AG/CT, AT/TA and AC/GT) and tri (AAC/GTT, AAG/CTT and AAT/AAT) nucleotide repeats in *C. chinense* and *C. frutescens* are shown in Fig. 6C, while the distribution of the identified SSRs in the *C. annuum* reference genome is given in Fig. 6D.

On the basis of significant homology with reported proteins available in the *C. annuum* Zunla-1 reference genome using BLAST (standalone), putative functions of 88.3 % (of *C. chinense*) and 95.2 % (of *C. frutescens*) SSRs were predicted (Supplementary Tables 4–5). The remaining SSRs did not show significant homology with any of the characterized proteins and hence were termed as unknown proteins (N/A).

3.4. *In silico* polymorphism analysis between *C. chinense* and *C. frutescens*

Identical SSR markers (with same repeat sequence and genomic location both in *C. chinense* and *C. frutescens*) were aligned using standalone BLASTN in order to identify the polymorphic SSRs with differences in the repeat motif numbers between *C. chinense* and *C. frutescens* which can be used for the development of a SSR-based inter-specific genetic map. All the SSRs for which the primers were designed i.e. 4988 of *C. chinense* and 4781 of *C. frutescens* primer pairs, were used to identify polymorphic SSRs *in silico* (Supplementary Tables 4–5). A total of 1123 SSR motifs that were common in both the species were screened for the identification of polymorphisms. *In silico* analysis

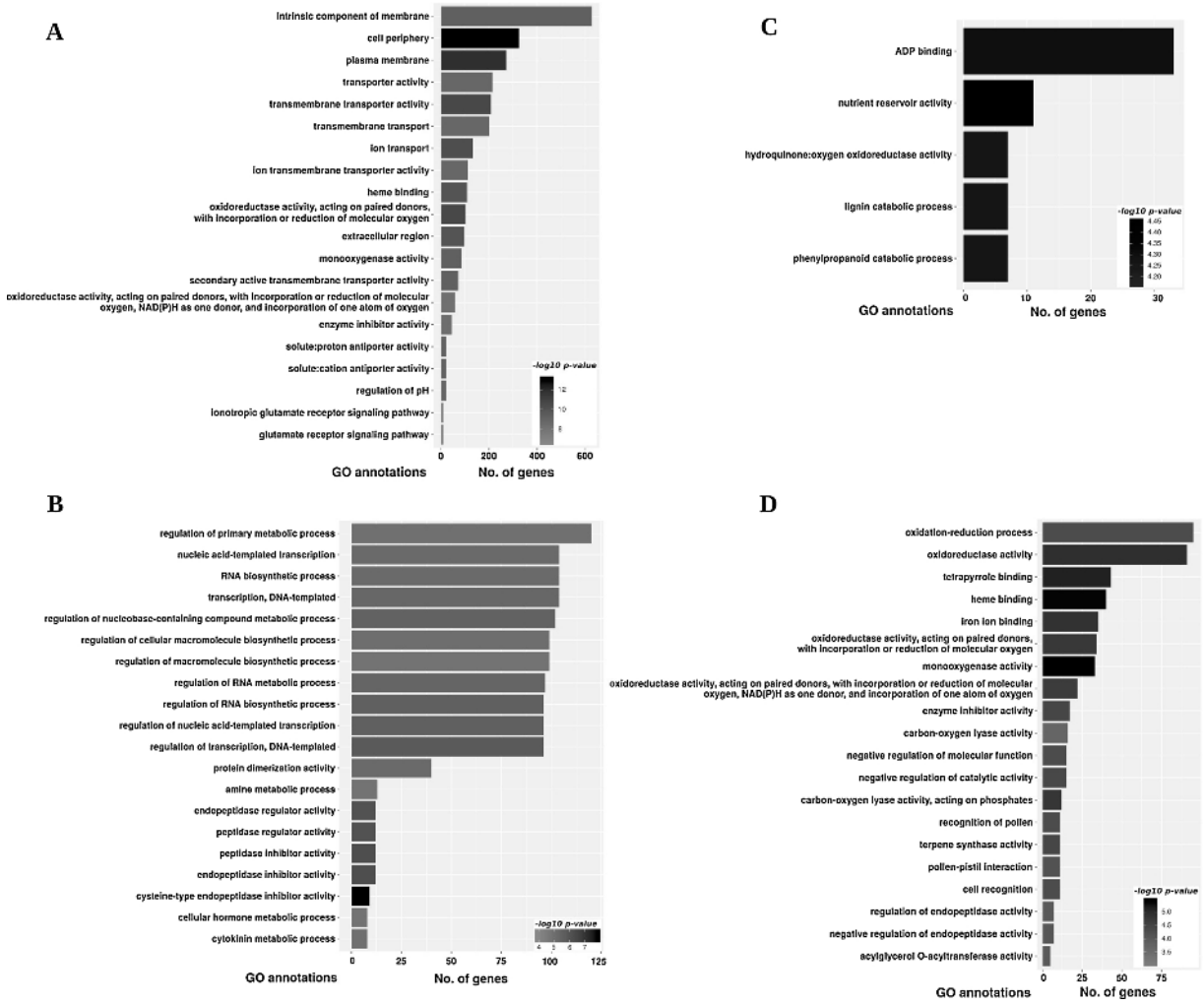

Fig. 3. Gene ontology (GO) analysis of transcripts uniquely expressed in a tissue across *Capsicum* species. Top 20 GO terms related to genes uniquely expressed in (A) flower, (B) early fruit, (C) breaker fruit and (D) mature fruit at p-value < 0.01 and FDR < 0.05.

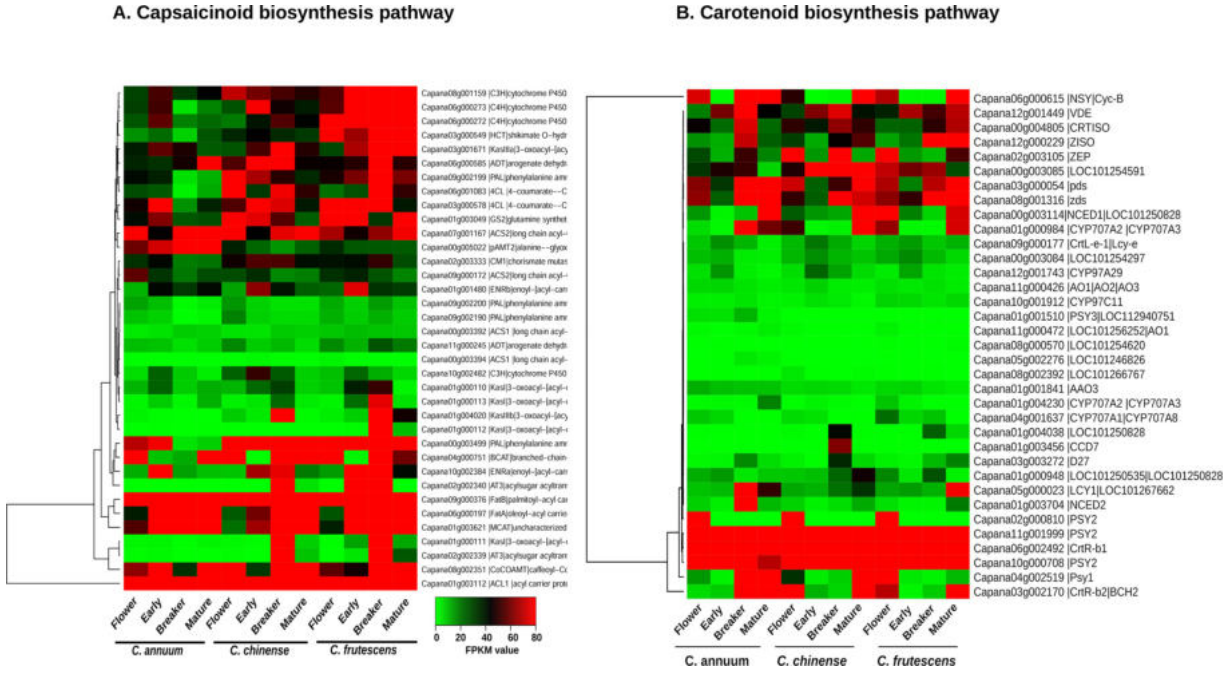

Fig. 4. Expression profile of (A) capsaicinoids and (B) carotenoid biosynthesis genes in the form of heatmap.

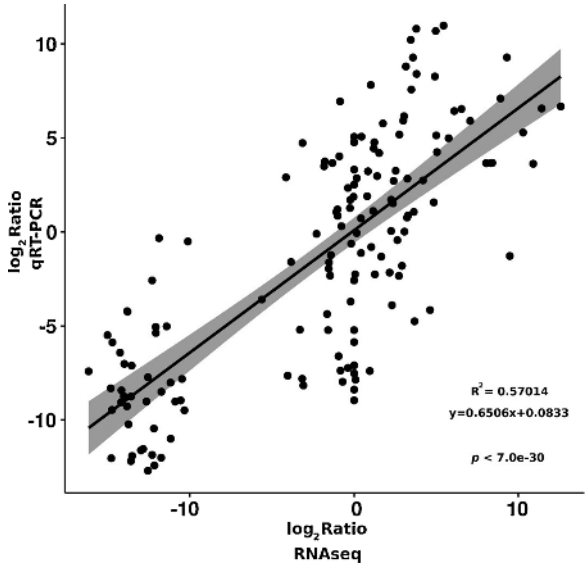

Fig. 5. Correlation of expression of genes observed in RNAseq (Transcriptome) data and qRT-PCR analysis.

revealed 335 polymorphic SSRs between *C. chinense* and *C. frutescens*. The details of these polymorphic SSRs along with the primer sequences are given in Supplementary Table 6. We however retained only those SSR motifs that showed a minimum size difference of 2 nucleotides. Moreover, to identify and differentiate the polymorphism through gel electrophoresis, 9 SSRs out of 335 polymorphic SSRs with an allelic size difference of more than 4 bases (between *C. chinense* and *C. frutescens*) were selected. Subsequently, the PCR amplifications were performed for all the nine SSR markers. However, four markers resulted in larger amplicon product sizes than expected (possibly due to the presence of introns), while three markers did not demonstrate polymorphism on the gel electrophoresis (because of the small differences in the product size or due to absence of polymorphism). For further validation, we then amplified and sequenced another *in silico* identified SSR motif (CFpSSR3), which showed polymorphism between *C. chinense* and *C. frutescens* (Fig. 7A). The sequencing of the amplified product confirmed the presence of an indel in the SSR. Furthermore, CFpSSR3 was also validated in a panel that comprised of *C. chinense*, *C. annuum* and *C. frutescens* genotypes. Metaphor-agarose gel electrophoresis confirmed the allelic variation in the form of a size difference of 6 bp between the genotypes belonging to these *Capsicum* species (Fig. 7B).

3.5. SSR validation and diversity analysis

Of the total 4988 *C. chinense* and 4781 *C. frutescens* primer pairs designed from sequences flanking SSRs, 50 SSR markers (25 SSRs from each *C. chinense* and *C. frutescens*) were randomly selected and the PCR amplification was done initially on 10 *Capsicum* genotypes (including 3 genotypes each from *C. chinense*, *C. frutescens* and *C. annuum* and one wild species, *Solanum pseudocapsicum*). Optimization of the reaction components and conditions for efficient amplification of SSRs was also done using these 10 genotypes. All of the PCR amplifications showed the presence of SSR bands in at least one of the three *Capsicum* species. Of these 50 SSR markers, 20 SSRs that highlighted robust, clear and polymorphic bands in all the three species were further selected for diversity analysis of 96 *Capsicum* accessions (Fig. 8)

Genotyping with the 20 SSRs revealed 84 alleles among the 96 accessions, ranging from 3 (CAP_SSR2, CAP_SSR3, CAP_SSR6, and CAP_SSR12) to 7 alleles (CAP_SSR19), with an average of 4.6 alleles per primer pair (Supplementary Table 7). The allele sizes ranged from 150 (CAP_SSR1) to 450 bp (CAP_SSR15). The polymorphism information content (PIC) value of the SSR markers ranged from 0.26 (CAP_SSR5) to 0.76 (CAP_SSR1) with an average of 0.53, (Supplementary Table 7) while the gene diversity ranged from 0.308 (CAP_SSR5) to 0.719 (CAP_SSR1) with a mean value of 0.513. Major allele 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 by CAP_SSR14 followed by CAP_SSR6 (0.625), CAP_SSR1 (0.583) and CAP_SSR16 (0.567). The lowest was found for two loci namely CAP_SSR18 and CAP_SSR8 (0.028 each) followed by loci CAP_SSR4 (0.031) and CAP_SSR11 (0.042).

The unweighted pair group method with arithmetic mean (UPGMA) cluster analysis revealed that 96 *Capsicum* accessions clustered in a species-specific manner. All the *Capsicum* accessions were grouped into four main clusters (Cluster I, II, III and IV; Fig. 9). Cluster I consisted of *C. frutescens* accessions from various locations of Northeast India – Assam, Meghalaya, Nagaland and other Indian states. However, the genotype 'Acc 87' in this cluster is morphologically dissimilar with respect to fruit shape and orientation (round shape, pendant fruit), in contrast to the slender shape and erect orientation of fruits of accessions belonging to *C. frutescens*. Cluster II consists of highly pungent *C. chinense* genotypes that have pendant fruit orientation with diverse fruit colors (dark red, orange and chocolate), and various fruit shapes ranging from conical to ovate. Acc 34 placed in a 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/slender shape; Fig. 9). Cluster III represents *C. annuum* accessions with low pungency levels compared with the other two species. In the analyzed germplasms, diverse fruit colors, ranged from red to

Table 3
Detail characteristics of SSR motif's identified in the *C. chinense* and *C. frutescens*.

SSR Motif Length	Repeat unit number									Percentage of SSR motif length	Number of primers designed	Percentage of designed primers
	Species	5	6	7	8	9	10	> 10	Total			
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	75	1.5
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		4988	
	C. frutescens	3974	4020	1949	818	532	354	187	11835		4781	

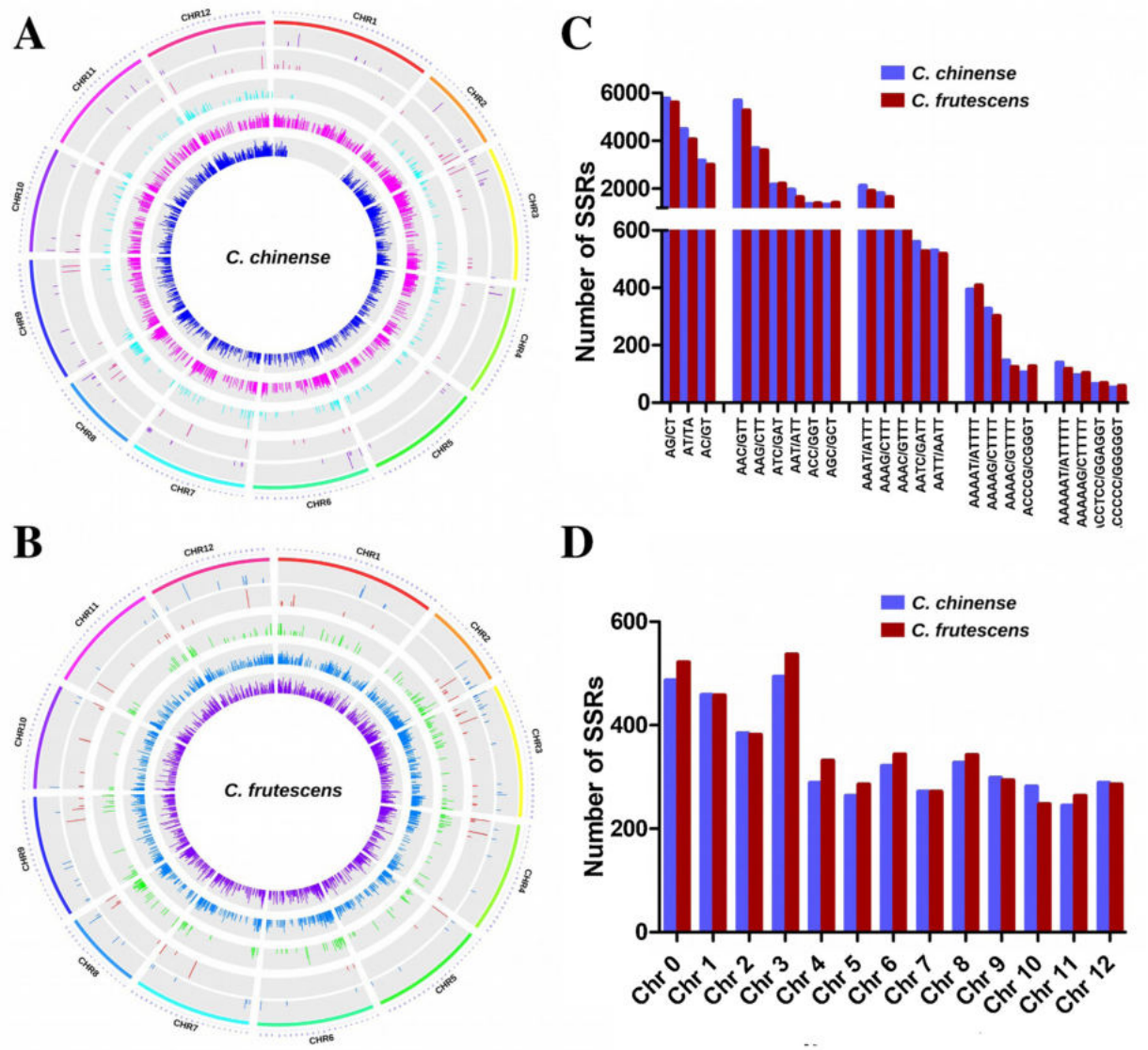

Fig. 6. The distribution and characteristics of SSR motifs. Genome-wide distribution of identified SSRs in (A) *C. chinense* (*Bhut jolokia*) and (B) in *C. frutescens* (Kon jolokia) on the *C. annuum* Zunla-1 reference genome (Qin et al., 2014). Each track (from inside to outside) shows di-, tri-, tetra-, penta-, hexanucleotide class of SSR repeat motif; (C, D) frequency of different SSR motifs and SSR distribution across 12 *Capsicum* chromosomes.

yellow whereas the fruit shape varied from round to elongate. This cluster consists of genotypes collected from diverse geographical locations of India such as Jammu & Kashmir, Uttar Pradesh, Assam, Mizoram, Meghalaya, and from Canada. Cluster IV contains five uncharacterized accessions (Acc 51, Acc 7, Acc 49, Acc 64, Acc 66) separated from the three main clusters. Since, these genotypes could not be assigned to any of the three *Capsicum* species; they were categorized as 'other *Capsicum* species'. These genotypes showed distinct characteristics like round or slender fruit shape, low pungency and fewer seed count (6–12) per fruit.

3.6. Agronomic traits characterization in 96 Capsicum genotypes

The 96 *Capsicum* genotypes used in the present study showed diversity for nine different agronomic traits (Fig. 1). Among the 96 genotypes, red fruit color was observed in a maximum of 51 % genotypes, followed by orange (12.5 %), yellow (5.2 %) and chocolate or brown (2%) fruits. Elongated fruit shape (57 %) was predominantly observed in accessions followed by triangular (26 %), blocky (7.2 %), round (5.2 %) and Campanulate (3%) fruits (Table 1). Also, we have observed other fruit descriptors, which mainly include various types of fruit shapes at the blossom end. It was noted that the majority of the accessions had pointed shapes at the blossom end (73 %), followed by blunt (18 %) and sunken (8%) fruit shapes.

Distinct variations were also observed in other quantitative traits. Fruit weight varied from the smallest of 0.11 g (in Acc 23 of *C. frutescens*) to the largest of 8.6 g (Acc 95 of *C. annuum*), while fruit length ranged from a lowest of 0.87 cm (Acc 12 of *C. frutescens*) to the highest of 8.1 cm (Acc 19 of *C. chinense*; Table 1). In our earlier study, pungency analysis of these 96 accessions showed that the *C. chinense* accessions were highly pungent (600,000–8,00,000 Scoville Heat Unit; SHU), while *C. frutescens* accessions were moderately pungent (300,000–6,00,000 SHU), and the *C. annuum* accessions were less pungent (0–1,50,000 SHU). The sweet pepper or bell pepper (*C. annuum*) had negligible pungency (0 SHU) level (Sarpras et al., 2016).

4. Discussion

4.1. Development of genomic resources and gene expression study

Continuously expanding NGS based transcriptome sequencing data from different functional genomics projects have led to the discovery of

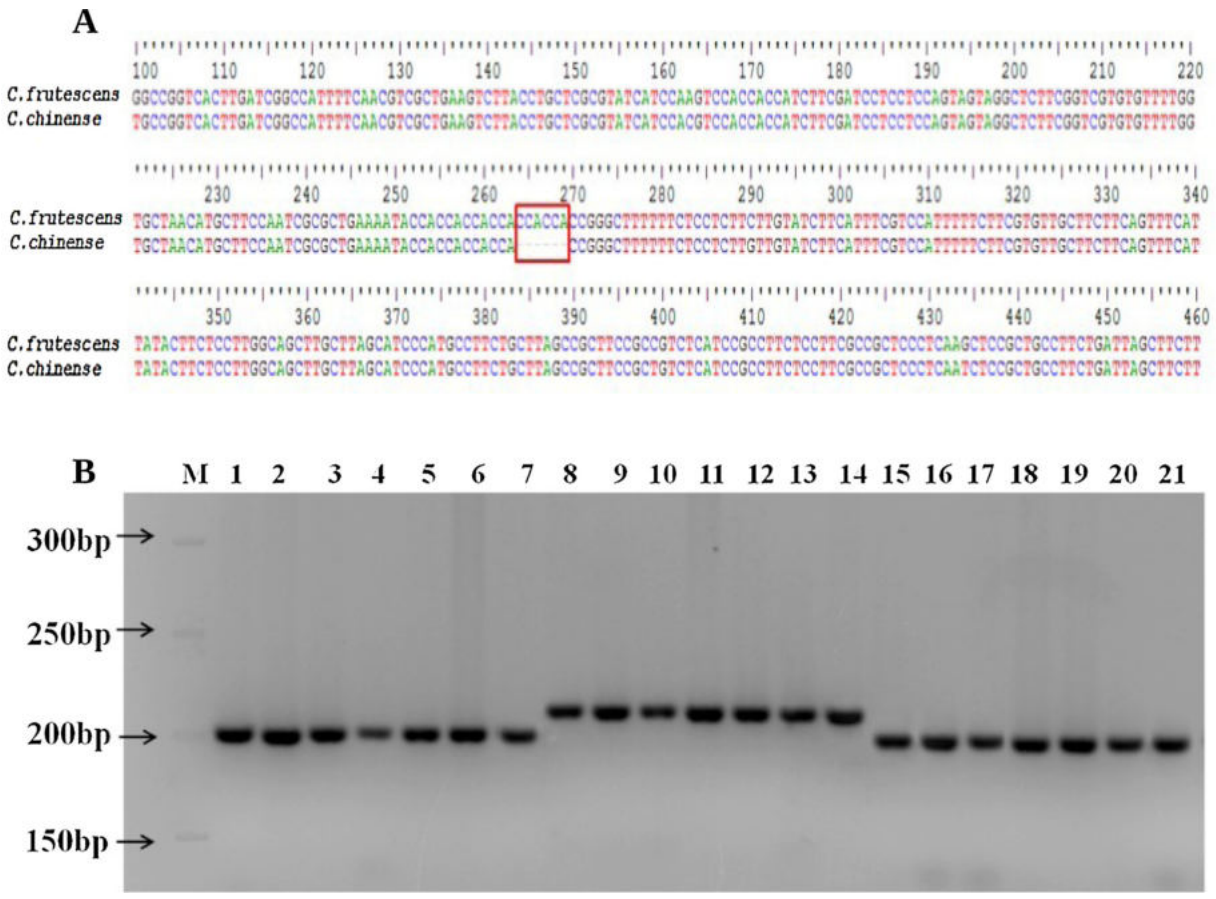

Fig. 7. Sequence comparison and allelic variation observed with CFpSSR3 marker in *Capsicum*. (A) Deletion of the repeat motif (highlighted in red box) in *C. chinense*; (B) Allelic variation among 21 genotypes of *Capsicum* germplasm. M: 50 bp ladder; Lanes 1-7: *C. annuum* genotypes (Accessions 16, 42, 92, 83, 3, 4 and 38); Lanes 8-14: *C. frutescens* genotypes (Accessions 23, 65, 32, 88, 96, 12 and 24); Lanes 15-21: *C. chinense* genotypes (Accessions 17, 22, 80, 72, 84, 85 and 61).

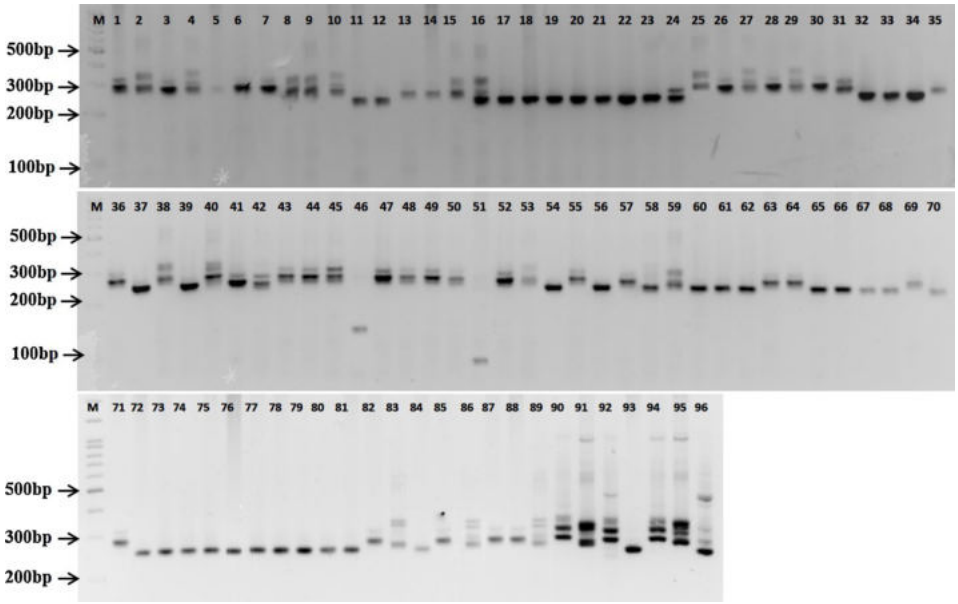

Fig. 8. Representative gel picture showing allelic variation detected with CAP_SSR9 marker in 96 *Capsicum* genotypes in metaphor agarose gel electrophoresis. *C. annuum*: Lanes 1-4, 6, 8-10, 13- 16, 25, 27, 29-31, 35-45, 47-48, 50, 52-55, 57-59, 63, 67, 69, 82-83, 89-92, 94-95; *C. chinense*: Lanes 5, 11, 17-22, 33-34, 46, 56, 60-62, 68, 70, 72-81, 84-85; *C. frutescens* : Lanes 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.

new genes, functional molecular markers, and expression profiling of global transcriptomes. These led to the successful identification of gene (s)/QTLs governing economically important traits and marker assisted transfer from one genetic background to another in many crop species. However, the lack of genetic and genomics information on potential industrial crops like *C. chinense* and *C. frutescens* native to Northeast India limited their utilization in breeding program using molecular tools. Furthermore, detailed genetic diversity study using molecular markers is required for the identification of potential genetic stocks to be used in breeding program. Therefore, the aim of the present study was to generate genomics resources using *C. chinense* and *C. frutescens,* and to identify diverse genotypes that can be used in future breeding

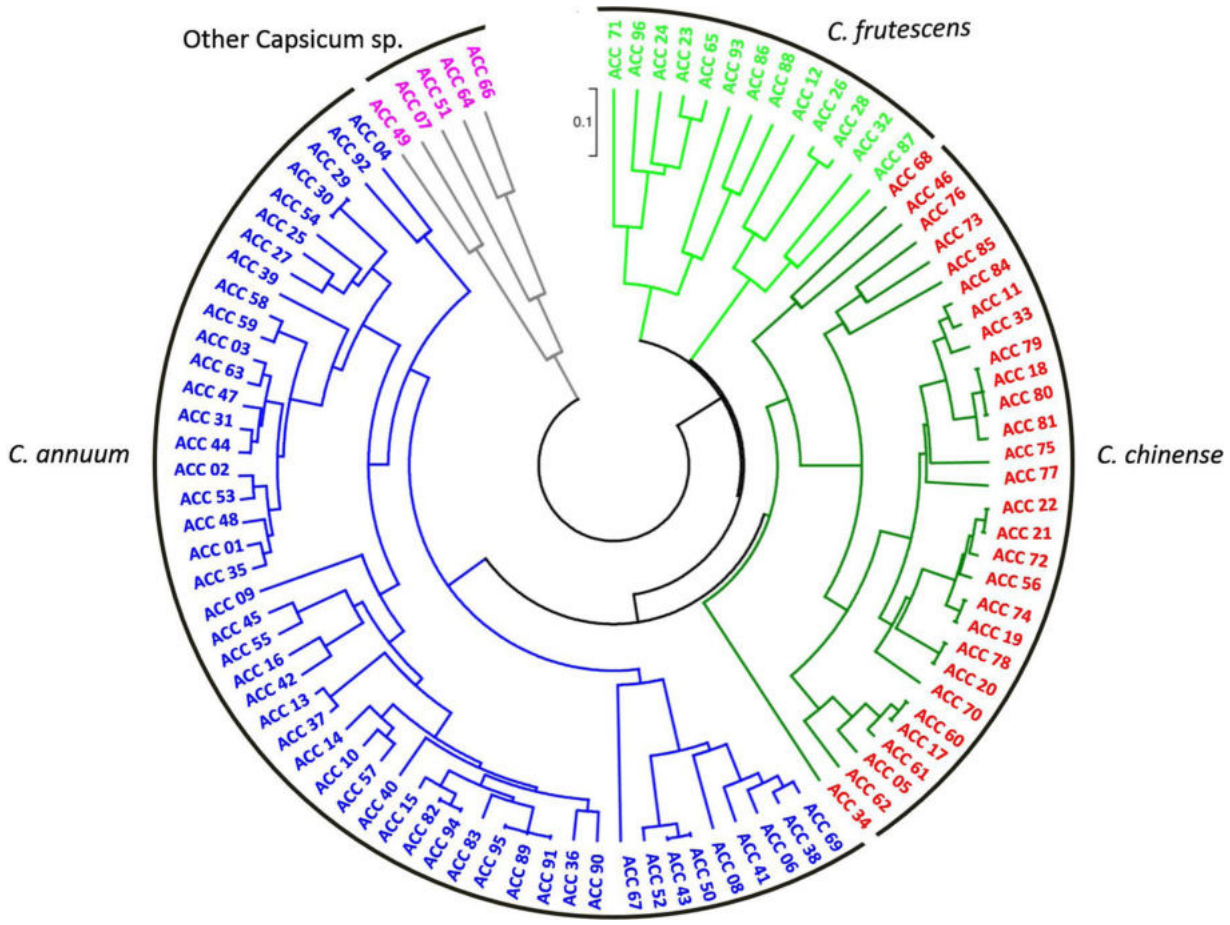

Fig. 9. UPGMA dendrogram demonstrating genetic relatedness of 96 chili pepper accessions. Accessions in green font represent cluster I of *C. frutescens* (Kon jolokia) group, accessions in red font represent cluster II of *C. chinense* (Bhut jolokia) genotypes and accessions in blue font represent Cluster III of *C. annuum* genotypes. Accessions in pink font (Cluster IV) belong to other *Capsicum* species.

programs.

We analyzed RNAseq of flower and different stages of fruit development for three *Capsicum* species i.e. *C. annuum*, *C. chinense* and *C. frutescens*. Overall > 87 % mapping efficiency of RNAseq reads was observed per sample (Table 2A). *De novo* assembly of RNAseq samples belonging to *C. chinense* and *C. frutescens* was performed to decipher their repetitive sequence information. A total of 244,135 unigenes with an average of N50 value 1569 and 1573 in *C. chinense* and *C. frutescens,* respectively, were found (Table 2B) which is comparable to the N50 value reported for *de novo* transcriptome assembly of *C. annuum* (N50 − 1647 bp; Ashrafi et al., 2012), *C. frutescens* (N50 − 1113 bp; Liu et al., 2013) and black pepper (N50 − 1496 bp; Hao et al., 2016). We further analyzed the expression patterns of transcripts across flower, fruit tissues at the three developmental stages (Fig. 2A) between three *Capsicum* species. Similar to chili pepper (Martínez-López et al., 2014), we also observed that the majority (70.8–73.6 %) of the transcripts were commonly expressed across all the tissues (Fig. 2B–D and Supplementary Fig. 1). The GO annotations revealed that the GO term cell periphery [GO:0071944] was most enriched for transcripts expressed only in the flower tissue (Fig. 3A) which were also reported as over-represented in the early developmental stages of floral organs (Cohen, 2016). The phenylpropanoid and lignin catabolic processes were most enriched in the breaker stages (Fig. 3C). Previous studies reported that several phenolic compounds produced from phenylpropanoid pathways and regulations of lignin biosynthesis are responsible for fruit pigmentation during fruit ripening (Seymour et al., 2008; Singh et al., 2010). In the present study our finding is similar to the observation reported by them.

The gene expression analysis showed differential expression patterns of capsaicinoids and carotenoid biosynthesis genes during different fruit development stages among *C. chinense, C. frutescens* and *C. annuum* (Fig. 4). The expression patterns of genes involved in Capsaicinoids biosynthesis pathway such as of *PAL* (*Capana09g002199*), *C4H* (*Capana06g000273*, *Capana06g000274*) and *HCT* (*Capana03g000549*) gene in *C. annuum* was similar to that reported in hot pepper (Kim et al., 2014; Qin et al., 2014) i.e. increase with fruit development but gradually decrease during breaker and post breaker stages. However, in both the *C. chinense* and *C. frutescens* genotypes, the *C4H* and *HCT* genes comparatively showed higher expression than that of *C. annuum* (Fig. 4A). Furthermore, we observed that genes such as *ZDS* (*Capana08g001316*) and *PDS* (*Capana03g000054*) involved in carotenoid biosynthesis, were highly expressed in the breaker and mature stage of *Capsicum* fruit (Fig. 4B). Our results corroborate with those from Hou et al. (2018), and we also observe the higher expression of *ZDS* and *PDS* genes in fruit of *C. frutescens* during ripening. Furthermore, during fruit ripening, genes such as *NCED1* (*Capana00g003114*) and *CYP707A2* (*Capana01g000984*) showed increased expression levels from breaker to mature fruit of *Capsicum* (Fig. 4B). In tomato, the increased expressions of *NCED1* and *CYP707A2* were also observed during fruit ripening, thereby suggesting their potential role in fruit ripening and abscisic acid signaling (Ji et al., 2014). In *C. frutescens*, the fruit coloration is found to be directly affected by the expression of *NCED1/3* (Hou et al., 2018). Overall, we observed a significant positive correlation between RNAseq and qRT-PCR analysis of gene expression data validating our results (Fig. 5).

4.2. Characterization of genic SSRs

Due to novel fast-paced advances in the genome sequencing technologies, mining and development of large-scale gene based microsatellite markers, has not only become faster but also cost-effective. The genic-SSRs have been effectively utilized in several studies which have highlighted their varied roles in gene regulation, DNA repair, chromatin organization, adaptation to various biotic and abiotic stresses, and in marker-trait association studies (Li et al., 2004; Varshney et al., 2005; Chen et al., 2017; Hou et al., 2017). Among the crops belonging to the Solanaceae, various studies on discovery of transcriptome based SSRs have earlier been reported (Zhou et al., 2015; Tan et al., 2014; Gramazio et al., 2016). However, in *Capsicum,* such studies have only been undertaken mostly in *C. annuum* (Lu et al., 2011; Nicolaï et al., 2012; Shirasawa et al., 2013; Ahn et al., 2014). In the present study, we identified a total of 4988 and 4781 SSRs in *C. chinense* and *C. frutescens,* using the unexploited genotypes from Northeast India, which could be efficiently used in development of genetic map, QTL mapping, map based cloning of genes and marker assisted transfer of important traits in *Capsicum*. Moreover, the identification of *in silico* polymorphic SSRs between *C. chinense* and *C. frutescens,* could directly be used for the screening and mapping of important traits QTL in interspecific mapping population developed from crosses between those two Capsicum species. The developed gene based SSR markers from transcriptome sequences will greatly supplement the existing marker repertoire of Capsicum species. Furthermore, these markers could be specifically useful in the improvement of economically important traits utilizing untapped chili genetic resources of *C. chinense* (Bhut jolokia) and *C. frutescens* (Kon jolokia) of Northeast India.

4.3. Frequency and distribution of SSRs in Capsicum genome

SSR frequency varies among different crops and in different regions of the same genome (Li et al., 2002). In this study, the observed frequency of SSRs in the *Capsicum* genome (one SSR per 7.52 kb and 7.42 Kb in *C. chinense and C. frutescens*, respectively) is comparable with the frequencies observed in other crops like 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), *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) but lower than wheat (1 per 5.4 kb; Peng and Lapitan, 2005) and pearl millet (1 per 1.75 kb; Senthilvel et al., 2008). The distribution, frequency and abundance of genic-SSRs could fluctuate due to several factors such as the number and size of the sequences analyzed, repeat lengths and also due to the SSR development tools being used (Varshney et al., 2005; Poncet et al., 2006). Previous studies have also reported the observation of high frequencies of SSRs in small genomes (Morgante et al., 2002), however, the frequency is found to be reduced in large genomes like that of *Capsicum*.

In this study, we observed that the trinucleotide repeats were the most abundant type of repeats in the *Capsicum* transcriptome and more than 50 % of the SSRs were of the trinucleotide repeats class (Supplementary Tables 4–5). Our observation is in agreement with the earlier reports on mungbean, cowpea and pigeonpea (Morgante et al., 2002; Gupta and Gopalakrishna, 2010; Dutta et al., 2011; Gupta et al., 2014). The abundance of trinucleotide SSR repeat type in the CDS region as compared with the other repeat types could be attributed to their triplet nature which may prevent frameshift mutations in the coding regions of genes (Metzgar et al., 2000). Abundance of trinucleotide repeats in pepper has also been reported earlier in different studies (Yi et al., 2006; Portis et al., 2007; Nicolaï et al., 2012; Cheng et al., 2016). Such observations on their abundance have also been made in other, both monocot and dicot crops including grape, barley, rice, wheat, citrus, cotton, soybean and flax (Scott et al., 2000; Thiel et al., 2003; Han et al., 2004; La Rota et al., 2005; Chen et al., 2006; Cloutier et al., 2009). Other studies on genome wide SSR identification in *C. annuum* (Cheng et al., 2016) and *C. chinense* (Uncu, 2019) showed that the AAT motif is overrepresented followed by AAC repeat motif, while the AAG motif was reported to be abundantly found in other plants (Ueno et al., 2009; Durand et al., 2010; Siju et al., 2010; Joshi et al., 2011).

In contrast, the presence of dinucleotide repeats was reported to be higher in almond, spruce and cucurbits (Xu et al., 2004; Rungis et al., 2004; Gong et al., 2008) compared with the trinucleotides repeats. Among the dinucleotide SSRs, AG/CT motifs were found to be most abundant in the present study (Fig. 6C). Genes related to these motifs are found to be involved in different metabolic pathways such as amino acid metabolism (ko01000, ko00230, ko00240,ko00280), secondary metabolites, amino acid biosynthesis pathways (ko01110, ko01230) etc. (Supplementary Tables 4–5). Also, it has been suggested that AG/CT represents codons such as GAG, AGA, UCU, and CUC in mRNA populations, which translate into the amino acids Arginine, Glutamic acid, Alanine, and Leucine, respectively. Since Alanine and Leucine are found in ample amounts in proteins (Gao et al., 2003), this observation justifies the excess of AG/TC motifs in the genome (Joshi et al., 2011). Other studies also support the predominance of AG/CT motif in coffee, cereals, forage crops (Temnykh et al., 2000; Thiel et al., 2003; Gao et al., 2003; Saha et al., 2004; Poncet et al., 2006) and perennials 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).

4.4. SSRs effectively differentiated 96 diverse Capsicum genotypes

The SSRs designed in this study were highly polymorphic as revealed by the allelic richness and PIC and could effectively differentiate the 96 diverse *Capsicum* genotypes (Supplementary Table 7). Average alleles per locus observed in the present study were found to be comparable with other studies reported in *Capsicum* (Cheng et al., 2016; Buso et al., 2016). Although Buso et al. (2016) reported up to 11 alleles per locus in a relatively smaller population (48 accessions), we did not identify such a large number of alleles per locus in our study. This could be due to the different genetic background of *Capsicum* genotypes.

The PIC, which quantitatively measures the informativeness of a genetic marker for the diversity and linkage studies, also suggested the efficacy of SSRs for the future studies. According to the parameters recommended by Botstein et al. (1980), a marker with a PIC value above 0.5 is believed to be highly polymorphic and is considered to be very useful. A high PIC value (mean = 0.53) suggested that SSRs developed during the present study are highly suitable for the genetic and diversity studies in *Capsicum*, like germplasm characterization, QTL mapping and marker assisted selection in both cultivated chili peppers and their wild relatives; and could also contribute to the understanding of the interrelationships among *Capsicum* accessions, species and their wild relatives.

SSR primers could only be used in different plant species when the primer binding sites flanking the SSR motifs are conserved. Due to high levels of sequence conservation in transcriptome/coding regions, the genic SSRs are highly transferable across species, compared with the genomic SSRs (Cho et al., 2000; Eujayl et al., 2001; Chabane et al., 2003). The cross-species transferability of SSR markers designed in this study was also found to be high, with a large percentage of loci producing amplicons in all the tested *Capsicum* species. This indicates a high level of sequence conservation within the primer binding regions and that these *Capsicum* species shared close genetic identities.

4.5. UPGMA based cluster analysis distinguished Bhut jolokia from other Capsicum species

Further, the UPGMA analysis could clearly distinguish the Bhut

jolokia group of genotypes belonging to *C. chinense* from the other (*C. annuum* and *C. frutescens*) genotypes. With the exception of one accession of *C. chinense* (Acc 49), all the accessions were grouped in cluster II of the dendrogram. Several accessions from Assam, Manipur, Nagaland, and Meghalaya grouping in the same cluster (II) suggest that they are genetically similar, despite growing in different regions. Similarly, the *C. annuum* accessions collected from different parts of the country also grouped into the same cluster (III) suggesting their genetic relatedness (Fig. 9). Our earlier study of diversity analysis of different genotypes belonging to *C. annuum*, *C. chinense* and *C. frutescens* using non coding RNA based SSRs also showed similar results (Jaiswal et al., 2020). The high genetic diversity observed in the Bhut jolokia of *C. chinense* and *C. frutescens* accessions could be attributed to a number of factors such as cross pollination, selection and adaptation to varied micro-climatic conditions together with their long history of cultivation in the Northeastern India. The diversity analysis revealed identification of important genetic stocks with desired traits such as pungency (extreme high, moderate and low), fruit size and shape, fruit color and weight, early flowering, short life cycle. The Bhut jolokia of *C. chinense* has extremely high pungent red and chocolate color triangular fruit while, *C. frutescens* has moderate pungency with small size fruit. These could be used for intra-specific and inter-specific hybridization studies to understand the genetic regulation of a variety of traits as aforementioned that have high commercial value in agricultural, pharmaceutical and food industries.

5. Conclusion

In summary, the transcripts and SSRs identified in this study will i) serve as important genetic and genomic resources, and ii) useful in the genetic mapping and identification of key QTLs/genes for economically important traits in the potential industrial crops *C. chinense* and *C. frutescens*. The candidate genes linked to SSRs could also be utilized in marker-assisted selection and transfer of the desirable alleles. To our knowledge, this is the first study conducted on *Capsicum species* (*C. chinense*, Bhut jolokia and *C. frutescens,* Kon jolokia) of Northeast India origin, to generate genic SSRs using transcriptome data. The genic SSRs based markers developed in this study are highly polymorphic and showed considerable transferability among genotypes of different *Capsicum* species (*C. chinense*, *C. frutescens* and *C. annuum*). Lastly, this study will help to utilize the genetically diverse *Capsicum* germplasm of Northeast India to further augment *Capsicum* improvement programs.

Accession number

RNA sequencing data developed and used in this study are submitted to the National Center for Biotechnology Information (NCBI). The transcriptomes data of *C. chinense*, *C. frutescens* and *C. annuum* are available under bioproject of NCBI with accession numbers PRJNA327797, PRJNA327800, and PRJNA505972, respectively.

Author contribution

NR conceived and designed the experiment. SSC, RG, VS, RK, AK collected the samples. SSC, SSB, RG performed the experiments. VB developed SSRs. AR, VB, NK performed computational analysis. SSC, VB, AR, RG, VJ, SKY performed data analysis. SSC and AR drafted the manuscript. VB, AR, SKY, VJ provided inputs in manuscript. NR wrote, corrected and modified the manuscript. All authors contributed in drafting and revision and approved the final manuscript.

CRediT authorship contribution statement

Sushil Satish Chhapekar: Methodology, Investigation, Formal analysis, Validation, Resources, Data curation, Visualization, Writing - original draft. **Vijaya Brahma:** Investigation, Software, Data curation, Formal analysis, Writing - review & editing. **Abdul Rawoof:** Methodology, Formal analysis, Investigation, Data curation, Visualization, Writing - original draft. **Nitin Kumar:** Visualization, Data curation, Writing - review & editing. **Rashmi Gaur:** Methodology, Investigation, Validation, Writing - review & editing. **Vandana Jaiswal:** Data curation, Writing - review & editing. **Ajay Kumar:** Resources, Formal analysis. **Satish K. Yadava:** Formal analysis, Data curation, Writing - review & editing. **Rajnish Kumar:** Resources, Investigation. **Vineet Sharma:** Resources, Investigation. **Salga S. Babu:** Methodology, Validation. **Nirala Ramchiary:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

None declared.

Acknowledgements

This study was supported by DST-PURSE grant from the Department of Science and Technology, Ministry of Science and Technology, Government of India to Jawaharlal Nehru University, Science and Engineering Research Board, Ministry of Science and Technololgy, and Department of Biotechnology, Ministry of Science and Technology, Government of India for providing Ramalingaswami Re-Entry Fellowship cum Research Grant (BT/RFL/Re-entry/46/2011) to Nirala Ramchiary. AR acknowledges direct Senior Research Fellowship received from Council of Scientific & Industrial Research. VB acknowledge the receive of Postdoctoral Fellowship from the University Grants Commission, New Delhi, India. Authors acknowledge and are thankful to Dr. Simran Bhullar for her critical comments and help extended in revision of this manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.indcrop.2020.112687.

References

Ahn, Y.-K., Tripathi, S., Kim, J.-H., Cho, Y.-I., Lee, H.-E., Kim, D.-S., Woo, J.-G., Cho, M.-C., 2014. Transcriptome analysis of Capsicum annuum varieties Mandarin and Blackcluster: assembly, annotation and molecular marker discovery. Gene 533, 494–499. https://doi.org/10.1016/j.gene.2013.09.095.

Arora, R., Gill, N.S., Chauhan, G., Rana, A.C., 2011. An overview about versatile molecule capsaicin. Int. J. Pharm. Sci. Drug Res. 3.

Ashrafi, H., Hill, T., Stoffel, K., Kozik, A., Yao, J., Chin-Wo, S.R., Van Deynze, A., 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, 571. https://doi.org/10.1186/1471-2164-13-571.

Baenas, N., Belović, M., Ilic, N., Moreno, D.A., García-Viguera, C., 2019. Industrial use of pepper (*Capsicum annum* L.) derived products: technological benefits and biological advantages. Food Chem. 274, 872–885.

Banjanac, T., Skorić, M., Belamarić, M., Nestorović Živković, J., Mišić, D., Mihailo, J., Dmitrović, S., Šiler, B., 2018. Pursuit for EST microsatellites in a tetraploid model from de novo transcriptome sequencing. Genetika 50, 687–703. https://doi.org/10.2298/GENSR1802687B.

Bosland, P.W., Baral, J.B., 2007. 'Bhut Jolokia'—the world's hottest known Chile pepper is a putative naturally occurring interspecific hybrid. HortScience 42, 222–224. https://doi.org/10.21273/HORTSCI.42.2.222.

Bosland, P.W., Votava, E.J., 2012. Pepper Vegetable and Spice Capsicums. CABI, Cambridge.

Botstein, D., White, R.L., Skolnick, M., Davis, R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32, 314–331.

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. Res. 15https://doi.org/10.4238/gmr.15038689. 10.4238/gmr.15038689.

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

Careaga, M., Fernández, E., Dorantes, L., Mota, L., Jaramillo, M.E., Hernandez-Sanchez, H., 2003. Antibacterial activity of Capsicum extract against Salmonella typhimurium

and Pseudomonas aeruginosa inoculated in raw beef meat. Int. J. Food Microbiol. 83, 331–335. https://doi.org/10.1016/s0168-1605(02)00382-3.
Carvalho, S.I.C., Ragassi, C.F., Oliveira, I.B., Amaral, Z.P.S., Reifschneider, F.J.B., Faleiro, F.G., Buso, G.S.C., 2015. Transferability of microsatellite markers of Capsicum annuum L. to C. frutescens L. and C. chinense Jacq. Genet. Mol. Res. 14, 7937–7946. https://doi.org/10.4238/2015.July.17.1.
Ceresini, P.C., Silva, C.L.S.P., Missio, R.F., Souza, E.C., Fischer, C.N., Guillherme, I.R., Gregorio, I., Silva, E.H.Tda, Cicarelli, R.M.B., Silva, M.T.Ada, Garcia, J.F., Avelar, G.A., Porto Neto, L.R., MarÃ\Son, A.R., Bacci Junior, M., Marini, D.C., 2005. Satellyptus: analysis and database of microsatellites from ESTs of Eucalyptus. Genet. Mol. Biol. 28, 589–600.
Chabane, K., Ablett, G.A., Cordeiro, G.M., Valkoun, J., Henry, R.J., 2003. EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. Genet. Resour. Crop Evol. 52, 903–909.
Chen, C., Zhou, P., Choi, Y.A., Huang, S., Gmitter Jr, F.G., 2006. Mining and characterizing microsatellites from citrus ESTs. Theor. Appl. Genet. 112, 1248–1257. https://doi.org/10.1007/s00122-006-0226-1.
Chen, H., Wang, L., Liu, X., Hu, L., Wang, S., Cheng, X., 2017. De novo transcriptomic analysis of cowpea (Vigna unguiculata L. Walp.) for genic SSR marker development. BMC Genet. 18, 65. https://doi.org/10.1186/s12863-017-0531-5.
Cheng, J., Zhao, Z., Li, B., Qin, C., Wu, Z., Trejo-Saavedra, D.L., Luo, X., Cui, J., Rivera-Bustamante, R.F., Li, S., Hu, K., 2016. A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in Capsicum. Sci. Rep. 6, 18919. https://doi.org/10.1038/srep18919.
Chhapekar, S.S., 2016. Reaping the benefits of next-generation sequencing technologies for crop improvement — solanaceae. In: Gaur, R. (Ed.), IntechOpen, Rijeka, pp. Ch. 9. https://doi.org/10.5772/61656.
Chhapekar, S.S., Jaiswal, V., Ahmad, I., Gaur, R., Ramchiary, N., 2018. Progress and prospects in capsicum breeding for biotic and abiotic stresses. In: Vats, S. (Ed.), Biotic and Abiotic Stress Tolerance in Plants. Springer, Singapore, pp. 279–322.
Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N., Cartinhour, S., 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (Oryza sativa L.). Theor. Appl. Genet. 100, 713–722. https://doi.org/10.1007/s001220051343.
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. https://doi.org/10.1007/s00122-009-1016-3.
Cohen, J.I., 2016. De novo sequencing and comparative transcriptomics of floral development of the distylous species Lithospermum multiflorum. Front. Plant Sci. 7, 1934. https://doi.org/10.3389/fpls.2016.01934.
Doyle, J.J., Doyle, J.E., 1990. Isolation of plant DNA from fresh tissue. Focus (Madison) 12, 13–15.
Dubey, M., Jaiswal, V., Rawoof, A., Kumar, A., Nitin, M., Chhapekar, S.S., Kumar, N., Ahmad, I., Islam, K., Brahma, V., Ramchiary, N., 2019. Identification of genes involved in fruit development/ripening in Capsicum and development of functional markers. Genomics 111, 1913–1922. https://doi.org/10.1016/j.ygeno.2019.01.002.
Durand, J., Bodénès, C., Chancerel, E., Frigerio, J.-M., Vendramin, G., Sebastiani, F., Buonamici, A., Gailing, O., Koelewijn, H.-P., Villani, F., Mattioni, C., Cherubini, M., Goicoechea, P.G., Herrán, A., Ikaran, Z., Cabané, C., Ueno, S., Alberto, F., Dumoulin, P.-Y., Guichoux, E., de Daruvar, A., Kremer, A., Plomion, C., 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11, 570. https://doi.org/10.1186/1471-2164-11-570.
Dutta, S., Kumawat, G., Singh, B.P., Gupta, D.K., Singh, S., Dogra, V., Gaikwad, K., Sharma, T.R., Raje, R.S., Bandhopadhya, T.K., Datta, S., Singh, M.N., Bashasab, F., Kulwal, P., Wanjari, K.B., K Varshney, R., Cook, D.R., Singh, N.K., 2011. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. BMC Plant Biol. 11, 17. https://doi.org/10.1186/1471-2229-11-17.
Dutta, S.K., Singh, S.B., Saha, S., Akoijam, R.S., Boopathi, T., Banerjee, A., Lungmuana, Vanlalhmangaiha, Somnath Roy, S., 2017. Diversity in bird's eye chilli (*Capsicum frutescens* L.) landraces of north-east India in terms of antioxidant activities. Proc. Natl. Acad. Sci. India. Sect. B. Biol. Sci. 87, 1317.
El-Rodeny, W., Kimura, M., Hirakawa, H., Sabah, A., Shirasawa, K., Sato, S., Tabata, S., Sasamoto, S., Watanabe, A., Kawashima, K., Kato, M., Wada, T., Tsuruoka, H., Takahashi, C., Minami, C., Nanri, K., Nakayama, S., Kohara, M., Yamada, M., Kishida, Y., Fujishiro, T., Isobe, S., 2014. Development of EST-SSR markers and construction of a linkage map in faba bean (Vicia faba). Breed. Sci. 64, 252–263. https://doi.org/10.1270/jsbbs.64.252.
Eujayl, I., Sorrells, M., Baum, M., Wolters, P., Powell, W., 2001. Progress and prospects in capsicum breeding for biotic and abiotic stresses. In: Bedö, Z., Láng, L. (Eds.), Wheat in a Global Environment. Developments in Plant Breeding. Springer, Dordr. 9, pp. 597–602.
Folta, K.M., Staton, M., Stewart, P.J., Jung, S., Bies, D.H., Jesdurai, C., Main, D., 2005. Expressed sequence tags (ESTs) and simple sequence repeat (SSR) markers from octoploid strawberry (Fragaria × ananassa). BMC Plant Biol. 5, 12. https://doi.org/10.1186/1471-2229-5-12.
Gao, L., Tang, J., Li, H., Jia, J., 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol. Breed. 12, 245–261.
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. https://doi.org/10.1007/s00122-008-0750-2.
Gong, Y., Xu, S., Mao, W., Li, Z., Hu, Q., Zhang, G., Ding, J., 2011. Genetic diversity analysis of faba bean (Vicia Faba L.) based on EST-SSR markers. Agric. Sci. China 10, 838–844. https://doi.org/10.1016/S1671-2927(11)60069-2.
Gramazio, P., Blanca, J., Ziarsolo, P., Herraiz, F.J., Plazas, M., Prohens, J., Vilanova, S., 2016. Transcriptome analysis and molecular marker discovery in Solanum incanum and S. aethiopicum, two close relatives of the common eggplant (Solanum melongena) with interest for breeding. BMC Genomics 17, 300. https://doi.org/10.1186/s12864-016-2631-4.
Guinness Book of World Records, 2006. Hottest Spice. www.guinnessworldrecords.com.
Gupta, S.K., Gopalakrishna, T., 2010. Development of unigene-derived SSR markers in cowpea (Vigna unguiculata) and their transferability to other Vigna species. Genome 53, 508–523. https://doi.org/10.1139/g10-028.
Gupta, S.K., Bansal, R., Gopalakrishna, T., 2014. Development and characterization of genic SSR markers for mungbean (Vigna radiata (L.) Wilczek). Euphytica v. 195 (2), 245–258. https://doi.org/10.1007/s10681-013-0993-0.
Guzmán, F.A., Moore, S., de Vicente, M.C., Jahn, M.M., 2020. Microsatellites to enhance characterization, conservation and breeding value of Capsicum germplasm. Genet. Resour. Crop Evol. 67, 569–585. https://doi.org/10.1007/s10722-019-00801-w.
Haanpää, M., Treede, R.-D., 2012. Capsaicin for neuropathic pain: linking traditional medicine and molecular biology. Eur. Neurol. 68, 264–275. https://doi.org/10.1159/000339944.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. https://doi.org/10.1038/nprot.2013.084.
Han, Z.-G., Guo, W.-Z., Song, X.-L., Zhang, T.-Z., 2004. Genetic mapping of EST-derived microsatellites from the diploid Gossypium arboreum in allotetraploid cotton. Mol. Genet. Genomics 272, 308–327. https://doi.org/10.1007/s00438-004-1059-8.
Hao, C., Xia, Z., Fan, R., Tan, L., Hu, L., Wu, B., Wu, H., 2016. De novo transcriptome sequencing of black pepper (Piper nigrum L.) and an analysis of genes involved in phenylpropanoid metabolism in response to Phytophthora capsici. BMC Genomics 17, 822. https://doi.org/10.1186/s12864-016-3155-7.
Hou, S., Sun, Z., Li, Y., Wang, Y., Ling, H., Xing, G., Han, Y., Li, H., 2017. Transcriptomic analysis, genic SSR development, and genetic diversity of proso millet (Panicum miliaceum; Poaceae). Appl. Plant Sci. 5https://doi.org/10.3732/apps.1600137. apps. 1600137.
Hou, B.-Z., Li, C.-L., Han, Y.-Y., Shen, Y.-Y., 2018. Characterization of the hot pepper (Capsicum frutescens) fruit ripening regulated by ethylene and ABA. BMC Plant Biol. 18, 162. https://doi.org/10.1186/s12870-018-1377-3.
Hu, Y., Yan, C., Hsu, C.-H., Chen, Q.-R., Niu, K., Komatsoulis, G.A., Meerzaman, D., 2014. OmicCircos: a simple-to-use R package for the circular visualization of multidimensional omics data. Cancer Inform. 13, 13–20. https://doi.org/10.4137/CIN.S13495.
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. Sci. Hortic. (Amsterdam) 181, 137–146. https://doi.org/10.1016/j.scienta.2014.10.054.
International Board for Plant Genetic Resources (IBPGR), 1983. Genetic Resources of Capsicum: a Global Plan of Action. International Board for Plant Genetic Resources, Rome, IBPGR Secretariat.
International Plant Genetic Resources Institute (IPGRI) Rome Italy, 1995. Descriptors for Capsicum (Capsicum spp.). https://www.bioversityinternational.org/fileadmin/user_upload/online_library/publications/pdfs/345.pdf.
Islam, M.A., Sinha, P., Sharma, S.S., Negi, M.S., Neog, B., Tripathi, S.B., 2016. Analysis of genetic diversity and population structure in Capsicum landraces from North Eastern India using TE-AFLP markers. Plant Mol. Biol. Rep. 34, 869–875. https://doi.org/10.1007/s11105-015-0968-5.
Jaiswal, V., Rawoof, A., Dubey, M., Chhapekar, S.S., Sharma, V., Ramchiary, N., 2020. Development and characterization of non-coding RNA based simple sequence repeat markers in Capsicum species. Genomics 112, 1554–1564. https://doi.org/10.1016/j.ygeno.2019.09.005.
Ji, K., Kai, W., Zhao, B., Sun, Y., Yuan, B., Dai, S., Li, Q., Chen, P., Wang, Y., Pei, Y., Wang, H., Guo, Y., Leng, P., 2014. SlNCED1 and SlCYP707A2: key genes involved in ABA metabolism during tomato fruit ripening. J. Exp. Bot. 65, 5243–5255. https://doi.org/10.1093/jxb/eru288.
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, 378–381. https://doi.org/10.6026/97320630005378.
Kanehisa, M., Sato, Y., Morishima, K., 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. 428, 726–731.
Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J.M., Lee, H.-A., Seo, E., Choi, J., Cheong, K., Kim, K.-T., Jung, K., Lee, G.-W., Oh, S.-K., Bae, C., Kim, S.-B., Lee, H.-Y., Kim, S.-Y., Kim, M.-S., Kang, B.-C., Jo, Y.D., Yang, H.-B., Jeong, H.-J., Kang, W.-H., Kwon, J.-K., Shin, C., Lim, J.-Y., Park, J.H., Huh, J.H., Kim, J.-S., Kim, B.-D., Cohen, O., Paran, I., Suh, M.C., Lee, S.-B., Kim, Y.-K., Shin, Y., Noh, S.-J., Park, J., Seo, Y.S., Kwon, S.-Y., Kim, H.-A., Park, J., Kim, H.-J., Choi, S.-B., Bosland, P.W., Reeves, G., Jo, S.-H., Lee, B.-W., Cho, H.-T., Choi, H.-S., Lee, M.-S., Yu, Y., Do Choi, Y., Park, B.-S., van Deynze, A., Ashrafi, H., Hill, T., Kim, W.T., Pai, H.-S., Ahn, H.K., Yeam, I., Giovannoni, J.J., Rose, J.K.C.C., Sørensen, I., Lee, S.-J., Kim, R.W., Choi, I.-Y., Choi, B.-S., Lim, J.-S., Lee, Y.-H., Choi, D., 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat. Genet. 46 (3), 270–278. https://doi.org/10.1038/ng.2877.
Kim, M.-S., Kim, S., Jeon, J., Kim, K.-T., Lee, H.-A., Lee, H.-Y., Park, J., Seo, E., Kim, S.-B., Yeom, S.-I., Lee, Y.-H., Choi, D., 2018. Global gene expression profiling for fruit organs and pathogen infections in the pepper, *Capsicum annuum* L. Sci. Data 5, 180103. https://doi.org/10.1038/sdata.2018.103.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874. https://doi.org/10.1093/molbev/msw054.

La Rota, M., Kantety, R.V., Yu, J.-K., Sorrells, M.E., 2005. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. BMC Genomics 6, 23. https://doi.org/10.1186/1471-2164-6-23.

Lee, J., Heath, L.S., Grene, R., Li, S., 2019. Comparing time series transcriptome data between plants using a network module finding algorithm. Plant Methods 15, 61. https://doi.org/10.1186/s13007-019-0440-x.

Lewers, K.S., Saski, C.A., Cuthbertson, B.J., Henry, D.C., Staton, M.E., Main, D.S., Dhanaraj, A.L., Rowland, L.J., Tomkins, J.P., 2008. A blackberry (Rubus L.) expressed sequence tag library for the development of simple sequence repeat markers. BMC Plant Biol. 8, 69. https://doi.org/10.1186/1471-2229-8-69.

Li, Y.-C., Korol, A.B., Fahima, T., Beiles, A., Nevo, E., 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol. Ecol. 11, 2453–2465. https://doi.org/10.1046/j.1365-294x.2002.01643.x.

Li, Y.-C., Korol, A.B., Fahima, T., Nevo, E., 2004. Microsatellites within genes: structure, function, and evolution. Mol. Biol. Evol. 21, 991–1007. https://doi.org/10.1093/molbev/msh073.

Li, Y.-H., Li, W., Zhang, C., Yang, L., Chang, R.-Z., Gaut, B.S., Qiu, L.-J., 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, 242–253. https://doi.org/10.1111/j.1469-8137.2010.03344.x.

Liu, K., Muse, S.V., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21, 2128–2129. https://doi.org/10.1093/bioinformatics/bti282.

Liu, S., Li, W., Wu, Y., Chen, C., Lei, J., 2013. De novo transcriptome assembly in chili pepper (Capsicum frutescens) to identify genes involved in the biosynthesis of capsaicinoids. PLoS One 8https://doi.org/10.1371/journal.pone.0048156. e48156–e48156.

Liu, F., Yu, H., Deng, Y., Zheng, J., Liu, Minglei, Ou, L., Yang, B., Dai, X., Ma, Y., Feng, S., He, S., Li, X., Zhang, Z., Chen, W., Zhou, S., Chen, R., Liu, Minmin, Yang, S., Wei, R., Li, H., Li, F., Ouyang, B., Zou, X., 2017. PepperHub, an informatics hub for the Chili pepper research Community. Mol. Plant 10, 1129–1132. https://doi.org/10.1016/j.molp.2017.03.005.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25, 402–408. https://doi.org/10.1006/meth.2001.1262.

Lu, F.H., Yoon, M.Y., Cho, Y.I., Chung, J.W., Kim, K.T., Cho, M.C., Cheong, S.R., Park, Y.J., 2011. Transcriptome analysis and SNP/SSR marker information of red pepper variety YCM334 and Taean. Sci. Hortic. (Amsterdam) 129, 38–45. https://doi.org/10.1016/j.scienta.2011.03.003.

Martínez-López, L.A., Ochoa-Alejo, N., Martínez, O., 2014. Dynamics of the chili pepper transcriptome during fruit development. BMC Genomics 15, 143. https://doi.org/10.1186/1471-2164-15-143.

Meghvansi, M.K., Siddiqui, S., Khan, M.H., Gupta, V.K., Vairale, M.G., Gogoi, H.K., Singh, L., 2010. Naga chilli: a potential source of capsaicinoids with broad-spectrum ethnopharmacological applications. J. Ethnopharmacol. 132, 1–14. https://doi.org/10.1016/j.jep.2010.08.034.

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

Meyer, L., Causse, R., Pernin, F., Scalone, R., Bailly, G., Chauvel, B., Délye, C., Le Corre, V., 2017. New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant Ambrosia artemisiifolia L. and can be transferred to other invasive Ambrosia species. PLoS One 12, e0176197.

Morgante, M., Hanafey, M., Powell, W., 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat. Genet. 30, 194–200. https://doi.org/10.1038/ng822.

Newcomb, R.D., Crowhurst, R.N., Gleave, A.P., Rikkerink, E.H.A., Allan, A.C., Beuning, L.L., Bowen, J.H., Gera, E., Jamieson, K.R., Janssen, B.J., Laing, W.A., McArtney, S., Nain, B., Ross, G.S., Snowden, K.C., Souleyre, E.J.F., Walton, E.F., Yauk, Y.-K., 2006. Analyses of expressed sequence tags from apple. Plant Physiol. 141, 147–166. https://doi.org/10.1104/pp.105.076208.

Nicolaï, 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, 2295–2300. https://doi.org/10.4238/2012.August.13.3.

Okumura, T., Tsukui, T., Hosokawa, M., Miyashita, K., 2012. Effect of caffeine and capsaicin on the blood glucose levels of obese/diabetic KK-A(y) mice. J. Oleo Sci. 61, 515–523. https://doi.org/10.5650/jos.61.515.

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. https://doi.org/10.2135/cropsci2011.06.0319.

Parida, S.K., Pandit, A., Gaikwad, K., Sharma, T.R., Srivastava, P.S., Singh, N.K., Mohapatra, T., 2010. Functionally relevant microsatellites in sugarcane unigenes. BMC Plant Biol. 10, 251. https://doi.org/10.1186/1471-2229-10-251.

Parida, S.K., Verma, M., Yadav, S.K., Ambawat, S., Das, S., Garg, R., Jain, M., 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. https://doi.org/10.3389/fpls.2015.00645.

Peng, J.H., Lapitan, N.L.V., 2005. Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. Funct. Integr. Genomics 5, 80–96. https://doi.org/10.1007/s10142-004-0128-8.

Poncet, V., Rondeau, M., Tranchant, C., Cayrel, A., Hamon, S., de Kochko, A., Hamon, P., 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. https://doi.org/10.1007/s00438-006-0153-5.

Portis, E., Nagy, I., Sasvári, Z., Stágel, A., Barchi, L., Lanteri, S., 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. https://doi.org/10.1016/j.plantsci.2006.11.016.

Purkayastha, J., Alam, S.I., Gogoi, H.K., Singh, L., Veer, V., 2012a. Molecular characterization of' Bhut Jolokia' the hottest chili. J. Biosci. 37, 757–768. https://doi.org/10.1007/s12038-012-9249-8.

Purkayastha, J., Alam, S.I., Gogoi, H.K., Singh, L., 2012b. *Capsicum assamicum* sp. nov. (Solanaceae), from Assam, northeastern India. Ozean J. Appl. Sci. 5, 1.

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al., 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proc. Natl. Acad. Sci. U. S. A. 111, 5135–5140. https://doi.org/10.1073/pnas.1400975111.

Ramchiary, N., Nguyen, V.D., Li, X., Hong, C.P., Dhandapani, V., Choi, S.R., Yu, G., Piao, Z.Y., Lim, Y.P., 2011. Genic microsatellite markers in Brassica rapa: development, characterization, mapping, and their utility in other cultivated and wild Brassica relatives. DNA Res. 18, 305–320. https://doi.org/10.1093/dnares/dsr017.

Ramchiary, N., Kehie, M., Brahma, V., Kumaria, S., Tandon, P., 2013. Application of genetics and genomics towards Capsicum translational research. Plant Biotechnol. Rep. https://doi.org/10.1007/s11816-013-0306-z.

Reddy, R.N., Madhusudhana, R., Mohan, S.M., Chakravarthi, D.V.N., Seetharama, N., 2011. Characterization, development and mapping of Unigene-derived microsatellite markers in Sorghum [Sorghum bicolor (L.) Moench]. Mol. Breed. 29, 543–564.

Rout, S.D., Panda, S.K., 2010. Ethnomedicinal plant resources of Mayurbhanj district, Orissa. Indian J. Tradit. Knowl. 9, 68–72.

Rungis, D., Bérubé, Y., Zhang, J., Ralph, S., Ritland, C.E., Ellis, B.E., Douglas, C., Bohlmann, J., Ritland, K., 2004. Robust simple sequence repeat markers for spruce (Picea spp.) from expressed sequence tags. Theor. Appl. Genet. 109, 1283–1294. https://doi.org/10.1007/s00122-004-1742-5.

Saha, M.C., Mian, M.A.R., Eujayl, I., Zwonitzer, J.C., Wang, L., May, G.D., 2004. Tall fescue EST-SSR markers with transferability across several grass species. Theor. Appl. Genet. 109, 783–791. https://doi.org/10.1007/s00122-004-1681-1.

Saha, D., Rana, R.S., Chakraborty, S., Datta, S., Kumar, A.A., Chakraborty, A.K., Karmakar, P.G., 2017. Development of a set of SSR markers for genetic polymorphism detection and interspecific hybrid jute breeding. Crop J. 5, 416–429. https://doi.org/10.1016/j.cj.2017.02.006.

Sarpras, M., Gaur, R., Sharma, V., Chhapekar, S.S., Das, J., Kumar, A., Yadava, S.K., Nitin, M., Brahma, V., Abraham, S.K., Ramchiary, N., 2016. Comparative analysis of fruit metabolites and pungency candidate genes expression between Bhut Jolokia and other Capsicum species. PLoS One 11, e0167791.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., Henry, R.J., 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. 100, 723–726. https://doi.org/10.1007/s001220051344.

Senthilvel, S., Jayashree, B., Mahalakshmi, V., Kumar, P.S., Nakka, S., Nepolean, T., Hash, C.T., 2008. Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. BMC Plant Biol. 8, 119. https://doi.org/10.1186/1471-2229-8-119.

Seymour, G., Poole, M., Manning, K., King, G.J., 2008. Genetics and epigenetics of fruit development and ripening. Curr. Opin. Plant Biol. 11, 58–63. https://doi.org/10.1016/j.pbi.2007.09.003.

Shirasawa, K., Ishii, K., Kim, C., Ban, T., Suzuki, M., Ito, T., Muranaka, T., Kobayashi, M., Nagata, N., Isobe, S., Tabata, S., 2013. Development of Capsicum EST-SSR markers for species identification and in silico mapping onto the tomato genome sequence. Mol. Breed. 31, 101–110. https://doi.org/10.1007/s11032-012-9774-z.

Shirasawa, K., Isobe, S., Tabata, S., Hirakawa, H., 2014. Kazusa Marker DataBase: a database for genomics, genetics, and molecular breeding in plants. Breed. Sci. 64, 264–271. https://doi.org/10.1270/jsbbs.64.264.

Siju, S.J., Dhanya, K.R., Syamkumar, S.A., Sasikumar, B., Sheeja, T.E., Bhat, A.I., Parthasarathy, V.A., 2010. Development, characterization and Cross species amplification of polymorphic microsatellite markers from expressed sequence tags of turmeric (Curcuma longa L.). Mol. Biotechnol. 44, 140–147.

Singh, R., Rastogi, S., Dwivedi, U.N., 2010. Phenylpropanoid metabolism in ripening fruits. Compr. Rev. Food Sci. Food Saf. 9, 398–416. https://doi.org/10.1111/j.1541-4337.2010.00116.x.

Singh, N., Choudhury, D.R., Singh, A.K., Kumar, S., Srinivasan, K., Tyagi, R.K., Singh, N.K., Singh, R., 2013. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. PLoS One 8, e84136.

Singh, A.K., Sharma, N., Ghosh, M., Park, Y.H., Jeong, D.K., 2017. Emerging importance of dietary phytochemicals in fight against cancer: role in targeting cancer stem cells. Crit. Rev. Food Sci. Nutr. 57, 3449–3463. https://doi.org/10.1080/10408398.2015.1129310.

Spiller, F., Alves, M.K., Vieira, S.M., Carvalho, T.A., Leite, C.E., Lunardelli, A., Poloni, J.A., Cunha, F.Q., de Oliveira, J.R., 2008. Anti-inflammatory effects of red pepper (*Capsicum baccatum*) on carrageenan- and antigen-induced inflammation. J. Pharm. Pharmacol. 60, 473–478. https://doi.org/10.1211/jpp.60.4.0010.

Sraphet, S., Boonchanawiwat, A., Thanyasiriwat, T., Boonseng, O., Tabata, S., Sasamoto, S., Shirasawa, K., Isobe, S., Lightfoot, D.A., Tangphatsornruang, S., Triwitayakorn, K., 2011. SSR and EST-SSR-based genetic linkage map of cassava (Manihot esculenta Crantz). Theor. Appl. Genet. 122, 1161–1170. https://doi.org/10.1007/s00122-010-1520-5.

Stágel, A., Gyurján, I., Sasvári, Z., Lanteri, S., Ganal, M.W., Nagy, I., 2009. Patterns of molecular evolution of microsatellite loci in pepper (*Capsicum* spp.) revealed by allele sequencing. Plant Syst. Evol. 281, 251–254.

Tan, Z., Fang, X., Tang, S., Zhang, J., Liu, Dajun, Teng, Z., Li, L., Ni, H., Zheng, F., Liu, Dexin, Zhang, T., Paterson, A.H., Zhang, Z., 2014. Genetic map and QTL controlling fiber quality traits in upland cotton (Gossypium hirsutum L.). Euphytica 203, 615–628.

Temnykh, S.V., Park, W.D., Ayres, N.M., Cartinhour, S., Hauck, N., Lipovich, L., Cho, Y.G., Ishii, T., McCouch, S.R., 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100, 697–712.

Thiel, T., Michalek, W., Varshney, R.K., 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. https://doi.org/10.1007/s00122-002-1031-0.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., Su, Z., 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 45, W122–W129. https://doi.org/10.1093/nar/gkx382.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562.

Ueno, D., Kono, I., Yokosho, K., Ando, T., Yano, M., Ma, J.F., 2009. A major quantitative trait locus controlling cadmium translocation in rice (*Oryza sativa*). New Phytol. 182, 644–653. https://doi.org/10.1111/j.1469-8137.2009.02784.x.

Ullah, M.F., Bhat, S.H., Husain, E., Abu-Duhier, F., Hadi, S.M., Sarkar, F.H., Ahmad, A., 2014. Cancer chemopreventive pharmacology of phytochemicals derived from plants of dietary and non-dietary origin: implication for alternative and complementary approaches. Phytochem. Rev. 13, 811–833. https://doi.org/10.1007/s11101-014-9341-9.

Umashanker, M., Shruti, S., 2011. Traditional Indian herbal medicine used as antipyretic, antiulcer, anti-diabetic and anticancer: a review. IJRPC 1 (4) ISSN: 2231-2781.

Uncu, A.T., 2019. Genome-wide identification of simple sequence repeat (SSR) markers in Capsicum chinense Jacq. with high potential for use in pepper introgression breeding. Biologia (Bratisl). 74, 119–126. https://doi.org/10.2478/s11756-018-0155-x.

Varshney, R.K., Graner, A., Sorrells, M.E., 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23, 48–55. https://doi.org/10.1016/j.tibtech.2004.11.005.

Wang, Z., Li, J., Luo, Z., Huang, L., Chen, X., Fang, B., Li, Y., Chen, J., Zhang, X., 2011. Characterization and development of EST-derived SSR markers in cultivated sweetpotato (Ipomoea batatas). BMC Plant Biol. 11, 139. https://doi.org/10.1186/1471-2229-11-139.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Available at:. Springer-Verlag, New York. http://ggplot2.org.

Wöhrmann, T., Weising, K., 2011. In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. Theor. Appl. Genet. 123, 635–647. https://doi.org/10.1007/s00122-011-1613-9.

Xu, Y., Ma, R.-C., Xie, H., Liu, J.-T., Cao, M.-Q., 2004. Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. Genome 47, 1091–1104. https://doi.org/10.1139/g04-058.

Yang, X., Xu, Y., Shah, T., Li, H., Han, Z., Li, J., Yan, J., 2011. Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica 139, 1045–1054. https://doi.org/10.1007/s10709-011-9606-9.

Yi, G., Lee, J.M., Lee, S., Choi, D., Kim, B.-D., 2006. Exploitation of pepper EST-SSRs and an SSR-based linkage map. Theor. Appl. Genet. 114, 113–130. https://doi.org/10.1007/s00122-006-0415-y.

You, F.M., Huo, N., Gu, Y.Q., Luo, M.-C., Ma, Y., Hane, D., Lazo, G.R., Dvorak, J., Anderson, O.D., 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9, 253. https://doi.org/10.1186/1471-2105-9-253.

Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., Simon, P., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J. Bot. 99, 193–208. https://doi.org/10.3732/ajb.1100394.

Zhang, X.-F., Sun, H., Xu, Y.C., Chen, B., Yu, S., Geng, S., Wang, Q., 2016. Development of a large number of SSR and InDel markers and construction of a high-density genetic map based on a RIL population of pepper (*Capsicum annuum* L.). Mol. Breed. 36, 1–10.

Zhang, Y., Zhang, X., Wang, Y.-H., Shen, S.K., 2017. *De Novo* assembly of transcriptome and development of novel EST-SSR markers in *Rhododendron rex* Lévl. through Illumina Sequencing. Front. Plant Sci. 8, 1664. https://doi.org/10.3389/fpls.2017.01664.

Zhang, Y., Zhang, M., Hu, Y., Zhuang, X., Xu, W., Li, P., Wang, Z., 2019. Mining and characterization of novel EST-SSR markers of *Parrotia subaequalis* (Hamamelidaceae) from the first Illumina-based transcriptome datasets. PLoS One 14, e0215874.

Zhou, R., Wu, Z., Jiang, F.L., Liang, M., 2015. Comparison of gSSR and EST-SSR markers for analyzing genetic variability among tomato cultivars (*Solanum lycopersicum* L.). Genet. Mol. Res. 14, 13184–13194. https://doi.org/10.4238/2015.October.26.14.