J Plant Biotechnol (2016) 43:311–316
DOI:http://dx.doi.org/10.5010/JPB.2016.43.3.311

ISSN 1229-2818 (Print)
ISSN 2384-1397 (Online)

Research Article

Transcriptome analysis, microsatellite marker information, and orthologous analysis of *Capsicum annuum* varieties

Yul-Kyun Ahn · Sandeep Karna · Jeong-Ho Kim · Hye-Eun Lee · Jin-Hee Kim · Do-Sun Kim

Received: 26 July 2016 / Revised: 17 August 2016 / Accepted: 17 August 2016

Abstract The efficacy of plant breeding has been enhanced by application of molecular markers in population screening and selection. Pepper (*Capsicum annuum* L.) is a major staple crop that is economically important with worldwide distribution. It is valued for its spicy taste and medicinal effect. The aim of this study was to discover single nucleotide polymorphisms (SNPs), microsatellite markers information, and percentage sharing through orthologous analysis of pepper-specific pungency-related genes. Here, we report the results of transcriptome analysis and microsatellite markers for four pepper varieties that possess a pungency-related gene. Orthologous analyses was performed to identify species-specific pungency-related genes in pepper, *Arabidopsis thaliana* L., potato (*Solanum tuberosum* L.), and tomato (*Solanum lycopersicum* L.). Advancements in next-generation sequencing technologies enabled us to quickly and cost-effectively assemble and characterize genes to select molecular markers in various organisms, including pepper. We identified a total of 9762, 7302, 8596, and 6886 SNPs for the four pepper cultivars Blackcluster, Mandarine, Saengryeg 211, and Saengryeg 213, respectively. We used 454 GS-FLX pyrosequencing to identify microsatellite markers and tri-nucleotide repeats (54.4%), the most common repeats, followed by di-, hexa-, tetra-, and penta-nucleotide repeats. A total of 5156 (15.9%) pepper-specific pungency-related genes were discovered as a result of orthologous analysis.

Keywords Molecular marker, Next-generation sequencing, Plant breeding, Screening

Y.-K. Ahn (✉) · S. Karna · J.-H. Kim · H.-E. Lee · J.-H. Kim · D.-S. Kim
Vegetable Research Division, National Institute of Horticultural & Herbal Science, Rural Development Administration, Jeonju 55365, Republic of Korea
e-mail: aykyun@korea.kr

Introduction

Pepper (*Capsicum annuum* L.) is an economically important vegetable crop that belongs to the *Capsicum* genus and Solanaceae family. A wild variety of peppers were domesticated as long ago as 6000 BC by Native Americans (Qin et al. 2014). Pepper fruits are very diverse for their color, shape, size, and appearance. Pepper is cultivated worldwide and has been widely used as a vegetable, condiment, spice, medicine, coloring agent, and a source of essential vitamins (Bosland and Votava 2000). Capsinoids, which are naturally present in pepper, are very important for the pharmaceutical industry. They have multiple pharmacological and physiological effects, including analgesic, anti-inflammatory, antioxidant, and antiobesity attributes (Luo et al. 2011).

There are multiple marker technologies available to increase the density of linkage maps. Single nucleotide polymorphisms (SNPs) have generated much interest for two reasons: (1) they are the most profuse form of genetic variation and appear at regular intervals in the genome (Studer et al. 2012), and (2) they are highly appropriate for multiplexed genotyping assays on microarray- or bead array-based platforms (Gupta et al. 2008). The main advantage of this technology is that it is high-throughput at low cost. DNA pyrosequencing, based on 454 Life Sciences technology (Margulies et al. 2005), has been successfully applied for large-scale sequencing of expressed sequence tags (ESTs) in maize (Ohtsu et al. 2007), *Medicago* (Cheung et al. 2006), *Arabidopsis* (Weber et al. 2007), and pepper (Kim et al. 2008a) to identify additional ESTs for these species (Novaes et al. 2008). Microsatellites, which are often referred to by plant geneticists as simple sequence repeats (SSR), are one of the most powerful genetic markers in plant science research. A general method to innovate SSR loci is to build genomic DNA libraries enriched for SSR sequences, followed by DNA sequencing (Edwards et al. 1996). Several computational devices are available to discover SSRs within sequence data and to design polymerase

chain reaction (PCR) primers appropriate for amplification of specific loci (Robinson et al. 2004). SSR markers enable the detection of multiple alleles per locus. SNP and SSR markers are used to study genetic variation among diverse genotypes. These markers are used for quantitative trait loci (QTL) mapping and other genomic applications (Liu et al. 2013).

During the last 20 years, there has been massive progress in linking plant genomics through comparative genetic maps, particularly for species belonging to the same family (Wang et al. 2008). Genomic and genetic information may be shared among leguminous species (Menancio-Hautea et al. 1993) or among members of the Solanaceae family (Livingstone et al. 1999), and orthologous genes are genes that diverged after speciation events (Fitch 2000) or that appear in different organisms but share a common ancestor. This evolutionary connection implies that products of orthologous genes likely maintain their original functions. The identification of true orthologs in plants is further complicated by the fact that most plants are paleopolyploids, and widespread gene repetition events have occurred during their evolution (Wu et al. 2006). In pepper, numerous molecular markers have been reported and numerous studies have been based on these markers (Ashrafi et al. 2012; Mimura et al. 2012). In this study, we have collected microsatellite marker information, transcriptome assembly, and ortholog analysis of four pepper varieties, to create high-quality, DNA-based molecular markers. Molecular marker information is an important resource when attempting to determine functional genetic variation and it can be used in breeding programs to improve the horticultural, nutritional, and medicinal value of crops.

Materials and Methods

Plant Material and cDNA Library Construction

Blackcluster, Mandarin, Saengryeg 211, and Saengryeg 213 pepper varieties were used. These plants were cultivated in a greenhouse at the National Institute of Horticultural & Herbal Science of the Rural Development Administration (RDA), Suwon, Republic of Korea. Mature fruits were harvested 60 days after flowering and tissue was thoroughly pulverized using liquid N, and stored at -80°C. Total RNA was extracted with the RNeasy plant mini kit (Qiagen). Extracted RNA was purified according to the PolyATract® mRNA isolation system IV (Promega). RNA samples were quantified and examined for protein contamination (A260/A280 nm and A260/A230 mm ratio) by GE Healthcare Bio-Science (NanoVue). Full-length cDNA was synthesized with the ZAP-cDNA® synthesis kit (Stratagene). The cDNA was fragmented with the Agilent 2100 BioAnalyzer (Waldbronn, Germany) to construct the sequencing library. The fragmented cDNA was used for high-throughput sequencing of the 454 sequencing library, according to the manufacturer's protocol (GS-FLX Titanium General Library Preparation Kit/emPCR kit/sequencing Kit, http://www.roche.com). Single effective copies of template species to be sequenced were hybridized to DNA capture beads and the immobilized library was re-suspended in the amplification solution. The mixture was emulsified and subjected to PCR amplification.

Discovery of SNPs

Raw reads for each pepper variety were counted separately, using the 454 GS-FLX sequencer. Isotig and singleton transcriptome data were used to mine SNPs and indel markers. SNP markers were detected by aligning individual reads against contigs, from the assembly using the CLC Genomics Workbench ver. 4.6.1. A minimum of two individual reads aligned with the reference (NCBI database; http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=4072&lvl=3&lin=f&keep=1&srchmode=1&unlock) were required to place the variant alleles in order and to consider a sequence difference as a true polymorphism. High-confidence variants were screened, assuming that ≥ 3 non-duplicate reads verified the same variant in both forward and reverse reads. We set the screening parameters so that single-nucleotide indels had to be identified by at least 10% of the total unique sequencing reads.

SSR Detection and Primer Design

All singletons and isotigs were applied independently from transcriptome data to mine SSR motifs and obtain information on molecular markers. Repeat motifs were screened from non-redundant trimmed sequences, using Repeatmasker (www.repeatmasker.org). The Simple Sequence Repeat Identification (SSRI, www.gramene.org/db/markers/ssrtool) tool was used to find perfect SSRs within the sequence (Temnykh et al. 2001). At least six di-nucleotides and four tri-nucleotides, or larger repeat sequences, were selected as microsatellites. Motifs were classified into three groups: perfect, compound, and interrupted repeats. Continuous, uninterrupted repeats were considered perfect repeats. Two or more sets of successive perfect repeats were considered compound repeats and those with few base pair interruptions were considered interrupted repeats. Duplicate SSRs within the same group were filtered and excluded, as isotigs from the same isogroup might share

the same sequence. Forward and reverse primers adjacent to SSRs containing sequences were designed using the PRIMER 3 software (Rozen and Skaletsky 2000).

Identification of Tentative Orthologous Groups

We identified orthologs using DNA rather than protein sequences, as we wanted to be very conservative in developing criteria for ortholog identification. We believed that DNA should be well preserved and identified based on reflexive, high stringency, transitive sequence matches across three or more species. Tentative consensus (TC) and expressed transcripts (ETs) was compared pair wise between the species. Tentative ortholog genes (TOGs) were identified by requiring transitive, reflexive best hits across at least three species, with a maximum BLASTN E-value of 10^{-5}. TC sequences and singleton ETs from each of the TIGR Gene Indices (TGIs) were searched independently using WU-Blast and the best hit for every sequence was recorded (Altschul et al. 1990). Matches meeting or beyond the maximum BLASTN E-value of 10^{-5} were stored in TIGR Orthologous Gene Alignment (TOGA), a relational database, implemented in SYBASE (http://sybase.com), and designed to capture relationships between orthologous genes. The search results were used to discover reciprocal best hit pairs. TOGs were assembled by choosing reciprocal best hit pairs (Lee et al. 2002).

Results and Discussion

Pepper consumption continues to grow because of the high nutritional and medicinal value of peppers. *Capsicum* is the only genus that produces capsaicinoids, which consist of more than 20 related alkaloids that cause pungency (Kim et al. 2014). In this study, we conducted transcriptome and microsatellite marker analyses of four pepper varieties Blackcluster, Mandarin, Saengryeg 211, and Saengryeg 213. Further orthologous analyses were performed to identify the species-specific related genes in pepper, *Arabidopsis*, potato, and tomato. NGS technology has been employed in different plant species, including pepper, to generate large amounts of sequence data (Ashrafi et al. 2012). The 454 GS-FLX pyrosequencing of these four pepper varieties was performed with four single-stranded template DNA (sstDNA) libraries, which produced high-quality cDNA fragments. The SNPs were filtered to obtain cultivar-specific SNPs. A total of 12,741, 9,701, 11,584, 9,641, and 15,077 SNPs were identified for Blackcluster, Mandarine, Saengryeg 211, Saengryeg 213, and a source from a public database (NCBI), respectively. A total of 9,762 (76.6%) of the specific SNPs and 2,979 (23.4%) of the nonspecific SNPs were identified for Blackcluster. After data mining, 75.3% total specific SNPs were discovered for Mandarine, 74.2% for Saengryeg 211, 71.4% Saengryeg 213, and 77.9% from public database. Table 1 present detailed SNP statistics for four pepper varieties along with public database.

NGS technology has been used to discover microsatellite markers in plant species (Zalapa et al. 2012) including pepper (Kim et al. 2008b; Lee et al. 2004). The SSR polymorphism level can be influenced by many factors including nucleotide motifs and repeat numbers, which is an important factor for their effective application. Kayser *et al.* (Kayser et al. 2004) suggested that SSR polymorphisms are positively correlated with motif size and repeat number. SSRs with higher numbers of repeats tend to be more polymorphic in humans (Weber

Table 1 SNP statistics for the four pepper varieties used in this study along with public database

Sample	Total SNP	TS SNP (%)	Genotypes		Variation Types			SFTSSNP (%)	SF Type SNP		NS SNP (%)
			HO SNP (%)	HE SNP (%)	SU (%)	IN (%)	DE (%)		SA SNP (%)	NSA SNP (%)	
Blackclus-ter	12,741	9,762 (76.6)	9,730 (99.7)	32 (0.3)	6,259 (64.1)	987 (10.1)	2,516 (25.8)	1,059 (8.3)	961 (7.5)	96 (0.8)	2,979 (23.4)
Mandarine	9,701	7,302 (75.3)	7,269 (99.5)	33 (0.5)	3,888 (53.2)	903 (12.4)	2,511 (34.4)	1,025 (10.6)	1009 (10.4)	15 (0.2)	2,399 (24.7)
Saengryeg 211	11,584	8,596 (74.2)	8,563 (99.6)	33 (0.4)	5,084 (59.1)	1,207 (14.0)	2,305 (26.8)	665 (5.7)	616 (5.5)	30 (0.3)	2,998 25.9
Saengryeg 213	9,641	6,886 (71.4)	6,873 (99.8)	13 (0.2)	3,850 (55.9)	936 (13.6)	2,100 (30.5)	632 (6.6)	634 (6.4)	16 (0.2)	2,755 (28.6)
Public Database	15,077	11,743 (77.9)	11,560 (98.4)	183 (1.6)	7,545 (64.3)	2,600 (22.1)	1,598 (13.6)	419 (2.8)	337 (2.2)	81 (0.5)	3,334 (22.1)

TS SNP = Total Specific SNP; HO SNP = Homozygous SNP; HE SNP = Heterozygous SNP; SU = Substitution; IN = Insertion; DE = Deletion; SFTS SNP = SF Type Specific SNP; SA SNP = Same Allele SNP; NSA SNP = Not Same Allele SNP; NS SNP = Not Specific SNP.

Table 2 SSR detection from raw reads and public database

Sample	TSR	SR	NS	Motif length					Total (%)	Frequency (bp/SSR)
				Di	Tri	Tetra	Penta	Hexa		
Black-cluster	3	0	3	246	1,920	66	72	190	2,494 (8.0)	10,308
Mandarine	9	3	3	161	1,567	96	99	144	2,067 (6.7)	12,437
Saengryeg 211	6	3	9	393	2,827	182	137	227	3,766 (12.1)	6,826
Saengryeg 213	9	0	6	234	1,847	120	80	150	2,431 (7.8)	10,575
Public Database	15	13	2	9,448	8,694	562	547	989	20,240 (65.3)	1,270
Total				10,482 (33.8%)	16,855 (54.4%)	1,026 (3.3%)	935 (3.0%)	1,700 (5.5%)	30,998 (100%)	829

TSR = Total Specific SSR, SR = Specific SSR, NS = No SSR.

Table 3 Ortholog analysis of orthologous gene families

Genes	Pepper Transcripts	Arabidopsis	Potato	Tomato
Total	32,325	27,416	52,925	34,727
Ortholog group number Total: 13,369	11,109	11,338	12,551	13,194
Ratio (%)	83.1	84.8	93.9	98.7
Ortholog group Genes	12,296	15,462	21,077	16,660
Ratio (%)	47.3	56.4	39.8	48.0
Specific Genes	5,126	2,562	7,136	3,131
Ratio (%)	15.9	9.3	13.5	9.0

1990), rice (Ellegren 2004), and *Medicago truncatula* (Mun et al. 2006). RepeatMasker (ver 3.2.7) was run to detect SSRs from sequencing reads for each cultivar. The tri-nucleotide repeats were most frequently detected in the coding regions (Yu et al. 2011), and most frequent repeats was tri-nucleotide (54.4%) in this study. The di-nucleotide repeats represented 33.8% of the SSRs, followed by 5.5% hexa-nucleotides 3.3% tetra-nucleotides, and 3.0% penta-nucleotides (Table 2) and similar results was previously reported (Sonah et al. 2011). Based on the SSRs identified in this study, further SSR optimization should be focused on tri-nucleotide repeats. Mononucleotide SSRs were excluded due to the frequent errors found in Roche 454 pyrosequencing.

Identification of overlapping orthologous clusters across multiple species enables the elucidation of the function and evolution of proteins (Wang et al. 2015). In this study, we analyzed orthologous genome sequences in plants, and compared our results to previous studies on tomato and potato (Wu et al. 2006). Table 3 presents data from the analysis of orthologous gene families associated with pungency-related genes. A total number of 32,325, 27,416, 52,925 and 34,727 orthologous gene families were identified in pepper, *Arabidopsis*, potato, and tomato, respectively. Among these species, a total of 13,369 ortholog group numbers were identified, in which pepper transcripts shared 83.1% of genes, *Arabidopsis* shared 84.8%, potato shared 93.9% and tomato shared 98.7% of genes respectively. Figure 1 presents a Venn diagram showing species-specific genes related to pungency. In all, 5,126 (15.9%) pepper transcripts, 2,562 (9.3%) *Arabidopsis* transcripts,

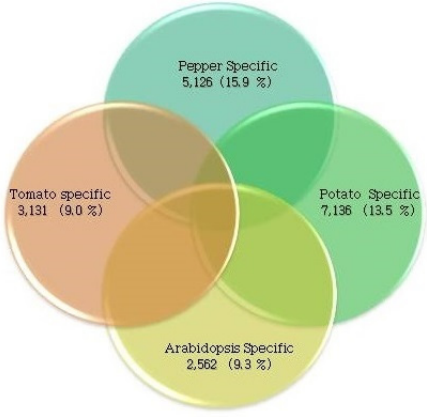

Fig. 1 Venn diagram showing pepper specific pungency related genes

7,136 (13.5%) potato transcripts, and 9% of the tomato transcripts shared species-specific related genes.

In summary, the transcriptome assembly of these pepper varieties provided high-quality gene-based molecular markers, which are an important resource for establishing functional genetic variation as applied to pepper breeding programs (Barbazuk et al. 2007). Our analyses included *de novo* transcriptome assembly, a classification of transcriptome, and the identification of large sets of candidate markers for population-level genetic analyses of pepper.

Acknowledgement

This study was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project no. PJ01106801), RDA, Republic of Korea".

References

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

Ashrafi H, Hill T, Stoffel K, Kozik A, Yao J, Chin-Wo SR, 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

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51: 910-918

Bosland PW, Votava EJ (2000) Peppers: Vegetable and Spice Capsicums. Crop Production Science in Horticulture Series Series No. 12, CABI Publishing, United Kingdom

Cheung F, Haas BJ, Goldberg SM, May GD, Xiao Y, Town CD (2006) Sequencing Medicago truncatula expressed sequenced tags using 454 Life Sciences technology. BMC Genomics 7: 272

Edwards KJ, Barker JH, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758-760

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435-445

Fitch DH (2000) Evolution of "rhabditidae" and the male tail. J Nematol 32:235-244

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity (Edinb) 101: 5-18

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C (2004) A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74: 1183-1197

Kim HJ, Baek KH, Lee SW, Kim J, Lee BW, Cho HS, Kim WT, Choi D, Hur CG (2008a) Pepper EST database: comprehensive in silico tool for analyzing the chili pepper (Capsicum annuum) transcriptome. BMC Plant Biol 8:101

Kim KS, Ratcliffe ST, French BW, Liu L, Sappington TW (2008b) Utility of EST-derived SSRs as population genetics markers in a beetle. J Hered 99:112-124

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E, Choi J, Cheong K, Kim KT, Jung K, Lee GW, Oh SK, Bae C, Kim SB, Lee HY, Kim SY, Kim MS, Kang BC, Jo YD, Yang HB, Jeong HJ, Kang WH, Kwon JK, Shin C, Lim JY, Park JH, Huh JH, Kim JS, Kim BD, Cohen O, Paran I, Suh MC, Lee SB, Kim YK, Shin Y, Noh SJ, Park J, Seo YS, Kwon SY, Kim HA, Park JM, Kim HJ, Choi SB, Bosland PW, Reeves G, Jo SH, Lee BW, Cho HT, Choi HS, Lee MS, Yu Y, Do Choi Y, Park BS, van Deynze A, Ashrafi H, Hill T, Kim WT, Pai HS, Ahn HK, Yeam I, Giovannoni JJ, Rose JK, Sorensen I, Lee SJ, Kim RW, Choi IY, Choi BS, Lim JS, Lee YH, Choi D (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species. Nat Genet 46: 270-278

Lee JM, Nahm SH, Kim YM, Kim BD (2004) Characterization and molecular genetic mapping of microsatellite loci in pepper. Theor Appl Genet 108:619-627

Lee Y, Sultana R, Pertea G, Cho J, Karamycheva S, Tsai J, Parvizi B, Cheung F, Antonescu V, White J, Holt I, Liang F, Quackenbush J (2002) Cross-referencing eukaryotic genomes: TIGR Orthologous Gene Alignments (TOGA). Genome Res 12:493-502

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 8:e48156

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK (1999) Genome mapping in capsicum and the evolution of genome structure in the solanaceae. Genetics 152:1183-1202

Luo XJ, Peng J, Li YJ (2011) Recent advances in the study on capsaicinoids and capsinoids. Eur J Pharmacol 650:1-7

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380

Menancio-Hautea D, Fatokun CA, Kumar L, Danesh D, Young ND (1993) Comparative genome analysis of mungbean (Vigna radiata L. Wilczek) and cowpea (V. unguiculata L. Walpers) using RFLP mapping data. Theor Appl Genet 86:797-810

Mimura Y, Inoue T, Minamiyama Y, Kubo N (2012) An SSR-based genetic map of pepper (Capsicum annuum L.) serves as an anchor for the alignment of major pepper maps. Breed Sci 62:

93-98

Mun JH, Kim DJ, Choi HK, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez AM, Kiss GB, Roe B, Young ND, Cook DR (2006) Distribution of microsatellites in the genome of Medicago truncatula: a resource of genetic markers that integrate genetic and physical maps. Genetics 172:2541-2555

Novaes E, Drost DR, Farmerie WG, Pappas GJ, Jr., Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. BMC Genomics 9:312

Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou R, Chen T, Zhang X, Timmermans MC, Beck J, Buckner B, Janick-Buckner D, Nettleton D, Scanlon MJ, Schnable PS (2007) Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). Plant J 52:391-404

Qin C, Yu C, Shen Y, Fang X, Chen L, Min J, Cheng J, Zhao S, Xu M, Luo Y, Yang Y, Wu Z, Mao L, Wu H, Ling-Hu C, Zhou H, Lin H, Gonzalez-Morales S, Trejo-Saavedra DL, Tian H, Tang X, Zhao M, Huang Z, Zhou A, Yao X, Cui J, Li W, Chen Z, Feng Y, Niu Y, Bi S, Yang X, Li W, Cai H, Luo X, Montes-Hernandez S, Leyva-Gonzalez MA, Xiong Z, He X, Bai L, Tan S, Tang X, Liu D, Liu J, Zhang S, Chen M, Zhang L, Zhang L, Zhang Y, Liao W, Zhang Y, Wang M, Lv X, Wen B, Liu H, Luan H, Zhang Y, Yang S, Wang X, Xu J, Li X, Li S, Wang J, Palloix A, Bosland PW, Li Y, Krogh A, Rivera-Bustamante RF, Herrera-Estrella L, Yin Y, Yu J, Hu K, Zhang Z (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

Robinson AJ, Love CG, Batley J, Barker G, Edwards D (2004) Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics 20:1475-1476

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386

Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, Rana JC, Singh NK, Sharma TR (2011) Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in Brachypodium. PLoS One 6:e21298

Studer B, Byrne S, Nielsen RO, Panitz F, Bendixen C, Islam MS, Pfeifer M, Lubberstedt T, Asp T (2012) A transcriptome map of perennial ryegrass (Lolium perenne L.). BMC Genomics 13:140

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

Wang Y, Coleman-Derr D, Chen G, Gu YQ (2015) OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 43:W78-84

Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD (2008) Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. Genetics 180:391-408

Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007) Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiol 144:32-42

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

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174:1407-1420

Yu JN, Won C, Jun J, Lim Y, Kwak M (2011) Fast and cost-effective mining of microsatellite markers using NGS technology: an example of a Korean water deer Hydropotes inermis argyropus. PLoS One 6:e26933

Zalapa JE, 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