PLOS ONE

RESEARCH ARTICLE

Development of a PCR-based, genetic marker resource for the tomato-like nightshade relative, *Solanum lycopersicoides* using whole genome sequence analysis

Puneet Kaur Mangat, Ritchel B. Gannaban, Joshua J. Singleton, Rosalyn B. Angeles-Shim*

Department of Plant and Soil Science, College of Agricultural Sciences and Natural Resources, Texas Tech University, Lubbock, Texas, United States of America

* rosalyn.shim@ttu.edu

Citation: Mangat PK, Gannaban RB, Singleton JJ, Angeles-Shim RB (2020) Development of a PCR-based, genetic marker resource for the tomato-like nightshade relative, *Solanum lycopersicoides* using whole genome sequence analysis. PLoS ONE 15(11): e0242882. https://doi.org/10.1371/journal.pone.0242882

Editor: David D. Fang, USDA-ARS Southern Regional Research Center, UNITED STATES

Received: June 30, 2020

Accepted: November 10, 2020

Published: November 23, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0242882

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Abstract

Solanum lycopersicoides is a wild nightshade relative of tomato with known resistance to a wide range of pests and pathogens, as well as tolerance to cold, drought and salt stress. To effectively utilize *S. lycopersicoides* as a genetic resource in breeding for tomato improvement, the underlying basis of observable traits in the species needs to be understood. Molecular markers are important tools that can unlock the genetic underpinnings of phenotypic variation in wild crop relatives. Unfortunately, DNA markers that are specific to *S. lycopersicoides* are limited in number, distribution and polymorphism rate. In this study, we developed a suite of *S. lycopersicoides*-specific SSR and indel markers by sequencing, building and analyzing a draft assembly of the wild nightshade genome. Mapping of a total of 1.45 Gb of *S. lycopersicoides* contigs against the tomato reference genome assembled a moderate number of contiguous reads into longer scaffolds. Interrogation of the obtained draft yielded SSR information for more than 55,000 loci in *S. lycopersicoides* for which more than 35,000 primers pairs were designed. Additionally, indel markers were developed based on sequence alignments between *S. lycopersicoides* and tomato. Synthesis and experimental validation of 345 primer sets resulted in the amplification of single and multilocus targets in *S. lycopersicoides* and polymorphic loci between *S. lycopersicoides* and tomato. Cross-species amplification of the 345 markers in tomato, eggplant, silverleaf nightshade and pepper resulted in varying degrees of transferability that ranged from 55 to 83%. The markers reported in this study significantly expands the genetic marker resource for *S. lycopersicoides*, as well as for related *Solanum spp*. for applications in genetics and breeding studies.

Introduction

Solanum lycopersicoides ($2n = 2x = 24$) from the *Lycopersicoides* section of the genus *Solanum* is a wild nightshade species that is distantly related to the cultivated tomato (*S. lycopersicum*)

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

[1]. It is endemic to the west Andes by the Chile-Peru border and thrives at high altitudes of up to 3800 m above sea level [2]. The species has known adaptation to cold, drought and salt stress [3–5], as well as resistance to phytophagous pests (i.e. leafminers) [6] and pathogenic fungi (e.g. *Botrytis cinerea* and *Phytophthora parasitica*) [7,8], bacteria (e.g. *Xanthomonas campestris*, *Clavibacter michiganensis* subsp. *michiganensis* and *Pseudomonas syringae* pv. *tomato*) [6,9] and viruses (e.g. tomato mosaic virus, tomato yellow leaf curl virus and tomato crinivirus) that commonly afflict the cultivated tomato [10–12].

Early efforts to introgress desirable traits from *S*. *lycopersicoides* to tomato have led to the successful generation of diploid F_1 hybrids via embryo rescue following wide hybridization. Unfortunately, the resulting F_1s are functionally male-sterile and unilaterally incompatible with tomato pollen and hence cannot be used directly for backcrossing [13]. Despite the initial challenges, introgression of *S*. *lycopersicoides* chromosomes in the genetic background of tomato was achieved through various strategies developed to overcome reproductive barriers associated with the interspecific crossing. These included the combined use of male-fertile amphidiploids from the interspecific F_1 hybrids and bridging lines of *S*. *pennellii* to circumvent the issue of unilateral incompatibility [14], modification of bud pollination to systematically avoid or suppress crossability barriers [8], and synthesis of a partially male-fertile F_1 hybrid by pollinating tomato with pollen pooled from several *S*. *lycopersicoides* plants [6]. Adoption of these techniques led to the successful generation of *S*. *lycopersicoides*-derived monosomic alien addition lines (MAALs) and chromosome segment substitution lines (CSSLs) [6,14,15]. MAALs are plants having the full chromosome complement of the cultivated species used as the recipient parent in an interspecific cross and an extra chromosome from the wild relative donor (2n + 1) [16]. CSSLs on the other hand, comprise a set of plants in the genetic background of an elite cultivar that represent the whole genome of the wild species in small, contiguous or overlapping chromosome segments [17–19]. These pre-breeding materials are unique genetic resources that capture novel genetic variations from the wild nightshade species in the tomato background. To date, these lines have been extensively evaluated for a range of agronomic characteristics [7,20] but the limited availability of DNA-based markers that can unlock the genetic basis of the observed phenotypes has restricted their efficient utilization in actual breeding programs to improve tomato.

Marker systems that have been used to characterize pre-breeding materials derived from *S*. *lycopersicoides* include morphological, biochemical and molecular markers. The DNA markers are in the form of restriction fragment length polymorphism (RFLP) and simple sequence repeats (SSRs) that are based solely on the tomato genome [6,14,15,21]. More recently, PCR-based, cleaved amplified polymorphic sequence (CAPS) markers developed based on existing RFLPs have also been used to map chromosome introgressions from *S*. *lycopersicoides* in the tomato background [22]. Despite the availability of DNA markers to characterize pre-breeding stocks developed from *S*. *lycopersicoides*, the marker resource available for the species remains limited in number, genome coverage and polymorphism rate. In case of the RFLP and CAP markers, digestion reactions that add to the cost, time and labor necessary to complete the genotyping make them less ideal for genetics and breeding studies.

Advances in molecular biology and instrumentation have facilitated the development of next generation, technological platforms for the rapid sequencing and assembly of whole genomes of several species [23,24]. With the availability of sophisticated but user-friendly computational tools, interrogation of new genome assemblies for sequence variations such as SSRs, insertions/deletions (indels) and single nucleotide polymorphism (SNPs) that can be used as targets for molecular marker development has become mainstream [25–27].

SSRs are tandemly arranged, repetitive sequences that make up a significant portion of eukaryote genomes [28], whereas indels are genomic insertions and deletions resulting from

replication slippage, simple sequence replications, unequal crossovers, retrotransposon insertions and segmental duplications [29–31]. Markers based on SSRs and indels are co-dominant, highly polymorphic and abundant in the genome. They are easily assayed by PCR and the amplicons directly resolved by agarose gel electrophoresis without any additional steps, making them more economical [32–34]. The robustness and technical simplicity of these markers for routine genotyping make them the marker of choice for genetics and breeding applications in many laboratories.

In this study, we aim to expand the limited genetic marker resource for *S. lycopersicoides* by developing SSR and indel markers based on whole genome sequence analysis. The molecular markers generated in this study are expected to accelerate basic research on the discovery and functional validation of genes/quantitative trait loci (QTL) conditioning traits of agronomic value in S. *lycopersicoides* towards their utilization in breeding for trait improvement in tomato.

Materials and methods

Plant materials, DNA extraction and whole genome sequencing

Seeds of the wild nightshade species, *S. lycopersicoides* (Acc LA1964) were provided by the Tomato Genetics Resource Center of the University of California, Davis (http://tgrc.ucdavis.edu). Seeds were surface-sterilized with 1% hypochlorite solution, plated in petri dishes lined with moist, sterile paper towels and germinated at an ambient temperature of 25˚C in the laboratory. The germinated seeds were transferred individually in 1-litre pots containing conventional potting media (composed of 45–50% composted pine bark, vermiculite, Canadian sphagnum, peat moss, perlite and dolomitic limestone) supplemented with slow-release NPK fertilizers and maintained in the greenhouse of the Horticultural Gardens of the Department of Plant and Soil Science (PSS) at Texas Tech University. Total genomic DNA was isolated from young leaves of *S. lycopersicoides* following a modified CTAB method [35]. The quality and quantity of purified genomic DNA were estimated using the NanoDrop™ One Microvolume UV-Vis Spectrophotometer (ThermoFisher, USA). Library preparation and whole genome sequencing using the Illumina HiSeq 3000 PE150 platform was outsourced to the Clinical Genomics Center of the Oklahoma Medical Research Foundation.

Whole genome analysis for repetitive sequence detection and primer design

Raw sequence data composed of 151-bp paired end reads were filtered using the Trimmomatic tool [36] to remove the Illumina adapter sequences and reads with poor quality. The trimmed reads were then assessed for quality using FastQC [37] before *de novo* assembling them into contigs using the short-read assembler, ABySS 2.0 [38]. Guided by the Build 3.0 of the reference genome for tomato (cv. Heinz), the contigs were used to generate longer scaffolds using the post-assembly genome improvement toolkit or PAGIT [39]. The quality of the newly built, draft assembly in comparison to the tomato reference genome was assessed using the quality assessment tool for genome assemblies or QUAST [40].

Genomic features in the draft assembly that can be used as basis for primer design were identified using various computational tools. Families of repetitive sequences were determined *de novo* based on existing Repbase libraries and collated into a species-specific database using RepeatModeler [41]. Repbase is maintained by the Genetic Information Resource Institute (GIRI) and is used as a reference for the annotation of eukaryotic repetitive DNA [42]. Repbase has libraries that are specifically available for the RepeatMasker software. The identified repeats were classified and annotated as retrotransposons, DNA transposons, small RNAs, satellites, simple repeats or low complexity DNA sequences using the RepeatMasker program

[43]. Mining the assembly for SSRs was carried out using the GMATA software [44]. SNPs and indels were identified based on sequence alignment between the *S. lycopersicoides* assembly and the tomato reference genome using the NUCmer (nucleotide MUMmer) package of the MUMmer program [45]. All programs used in the study ran on default settings.

The Primer3 program integrated into the GMATA software was used to generate primers for the SSRs that were identified in the draft assembly. To design the indel markers, comparative sequence alignment between the draft *S. lycopersicoides* and reference tomato genomes was generated using the Burrows-Wheeler aligner (BWA) [46]. The output file was then converted into a binary file that can be viewed using the Integrative Genomics Viewer [47]. Primers targeting the indels were manually designed using open source software and following specifications for standard primer design (i.e. 40–60% GC content, 20–25 bp in length). The GC content of the primer sequences were validated using the EndMemo-DNA/RNA GC content calculator [48]. The reverse primers were generated by reverse complementing DNA sequences through http://arep.med.harvard.edu/labgc/adnan/projects/Utilities/revcomp.html [49].

All SSRs and indel primers were designed at an average chromosome interval of 2–2.6 Mb. The specificity of the designed primers was validated *in silico* by BLAST searches [50] against the available tomato sequences curated at the NCBI database. Synthesis of all SSR and indel primers was outsourced to Sigma, USA.

Target amplification and cross-species transferability of *S. lycopersicoides* DNA markers

The ability of the newly designed markers to amplify targets in *S. lycopersicoides* was validated following a standard PCR protocol [51]. Adjustments in annealing temperature from 53˚C to 55˚C were carried out to optimize target amplification in *S. lycopersicoides*. PCR amplicons were resolved in 3% agarose gel in 1X Tris-Borate-EDTA buffer [51].

Additionally, the transferability of the *S. lycopersicoides*-specific markers to two other *S. lycopersicoides* accessions (LA2951 and LA2387) and other Solanaceous plants including tomato (Acc LA3122), eggplant (*S. melongena*) cv. Black Beauty, pepper (*Capsicum annuum*) cv. California Wonder and silverleaf nightshade (*S. elaeagnifolium*) was also determined. Young leaves from tomato, eggplant and pepper were sampled from seedlings germinated in the greenhouse as previously described. Leaf tissues of silverleaf nightshade were randomly sampled from populations growing at the Horticultural Gardens of PSS. Total genomic DNA for PCR was extracted from the young leaves of each of the Solanaceous species following a modified CTAB method [35].

Results and discussion

Whole-genome assembly and sequence repeats analysis

Illumina sequencing generated a total of 15.8 Gb of raw data containing 88,457,926 paired end reads that are 151 bp long (SRA accession SRX9292807). After trimming the adapters and removal of the poor-quality reads, the calculated average genome coverage [52] based on the tomato reference genome was 25X. *De novo* assembly generated a total of 6,874,225 contigs spanning a total length of 1,452,602,585 bp (Table 1).

Reference-guided assembly of the contigs into longer scaffolds mapped 589,717,391 bp (40.59%) of the *S. lycopersicoides* sequence data against the tomato genome. *S. lycopersicoides* and tomato are distant relatives that belong to different sections under the genus *Solanum*. The former belongs to the section *Lycopersicoides* which also includes one other species, *S.*

Table 1. General statistics obtained for *S. lycopersicoides* genome assembly using ABySS.

Descriptive statistics	Value (bp)
Number of contigs	6,874,225
Total length	1,452,602,585
Largest contig size	46755
Number of contigs that are ≥500bp	394658
Reference length[a]	828,076,956
Total length of contigs aligned to the reference	589,717,391
N50[b]	2141

[a]length of the Build 3.0 of tomato cv. Heinz reference genome.

[b]N50 is the length for which the collection of all contigs of that length or longer covers at least 50% of the assembly.

https://doi.org/10.1371/journal.pone.0242882.t001

sitiens, whereas the latter belongs to the section *Lycopersicon* which includes twelve other wild relatives [53]. Throughout the course of evolution, the genomes of these plants have been subjected to mutations, chromosomal rearrangements, transposon amplifications, gene duplications and extensive genome expansion/contraction. The genetic differentiation of each species that resulted from such genomic events may explain the moderate alignment of the *S. lycopersicoides* contigs against the tomato reference genome. This observation is consistent with the high proportion (17–25%) of paired sequence reads generated for *S. arcanum*, *S. pennellii* and *S. habrochaites* that also did not map against Build 2.40 of the tomato cv. Heinz reference genome, despite the three species belonging to the same section as tomato [54]. Given the 1.45 Gb total contig length obtained for *S. lycopersicoides* in this study, assembly of a draft that is guided by a genome of a closer relative other than tomato has the potential to generate a longer consensus sequence for the species. Alternatively, the draft assembly can be improved by refining, gap filling and expanding the initial assembly with long reads generated by third-generation sequencing technology such as the PacBio SMRT.

Interrogation of the draft assembly for repetitive sequences using the RepeatModeler in conjunction with the RepeatMasker detected a total of 712,011 repeats, covering 164,390,084 bp (18.83%) of the draft. The interspersed repeats consisted of short interspersed nuclear elements (0.06%), short interspersed nuclear elements (0.85%), long terminal repeats (6.63%), DNA elements (1.24%) and unclassified repeats (9.63%). The proliferation and deletion of transposable elements (TEs) are key determinants of genome size variation in eukaryotes [55]. The loss of TEs in tomato during domestication may be one of the primary reasons behind the genome size difference between *S. lycopersicoides* and tomato. RepeatMasker also classified the repeats into small RNAs (0.05%), satellites (0.01%), simple repeats (0.39%) and low complexity repeats (0.08%) (Fig 1). With the draft assembly capturing less than 50% of the *S. lycopersicoides* genome, analysis of an improved assembly is expected to increase the proportion of these repeats in the genome.

SSR mining identified 56,901 SSRs with motifs ranging from 2 to 9 bp (Table 2). SSRs with di-nucleotide motifs were the most abundant (74.05%), whereas those having penta-, hexa-, hepta-, octo- and nona-nucleotide repeats comprise only approximately 1.26% of the total SSRs identified in the assembly. Among the di-nucleotide motifs, AT and TA make up more than 50% of the total SSRs with 2-bp repeats (Fig 2A). In terms of length, 10-bp SSRs were the most predominant (37.9%), whereas those that are 34-bp long were the scarcest (0.5%) (Fig 2B).

Primer design and target DNA amplification

The built-in Primer3 software in the GMATA tool was used to design primer pairs for 35,801 SSR loci with 34,198 unique markers (Table 3). To validate the ability of primer pairs to

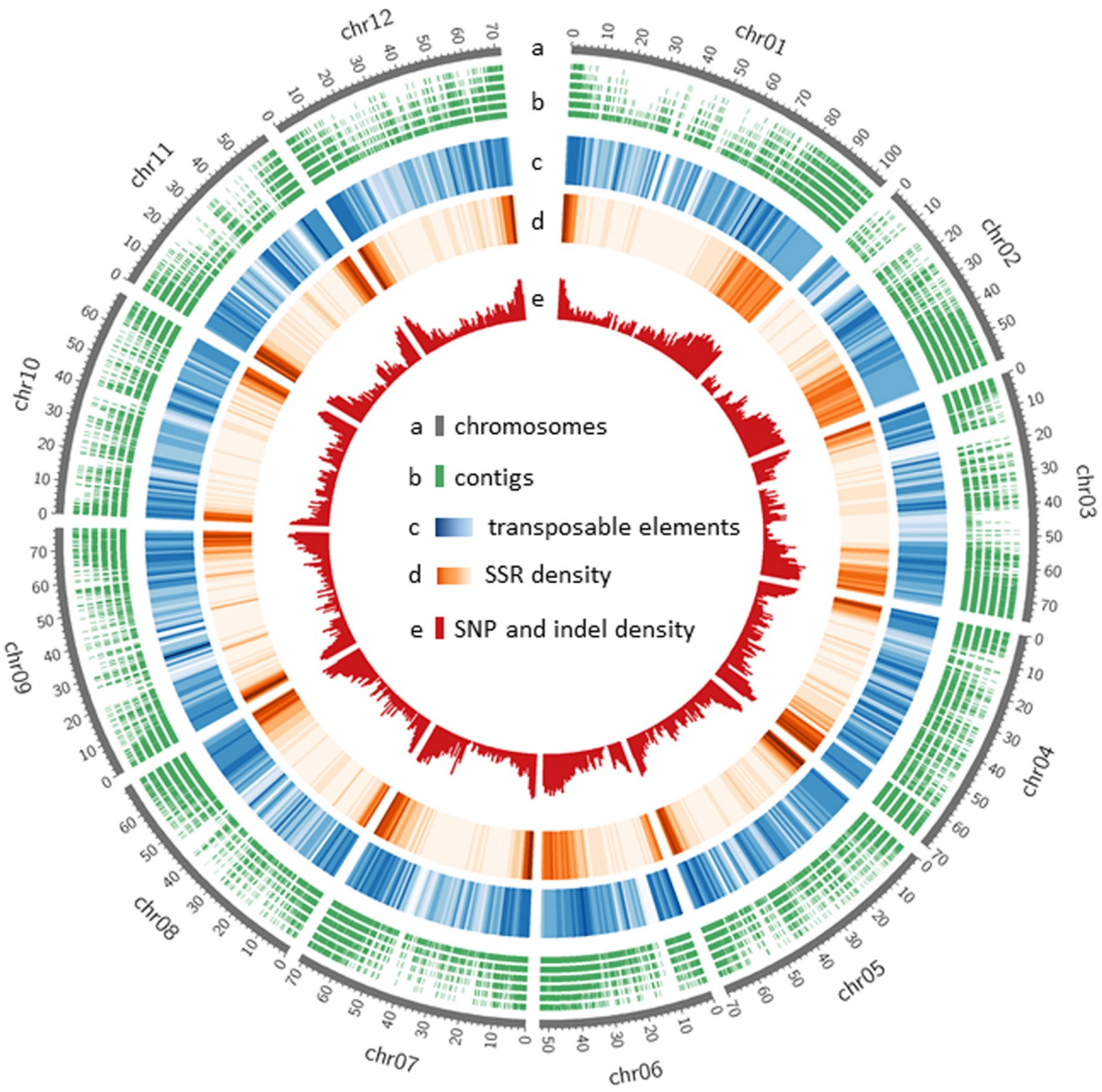

Fig 1. Circular view of the *S. lycopersicoides* genome assembly used for sequence variation mining. The outermost to the innermost rings represent the 12 representative pseudomolecules, contigs (≥500 bp), transposable elements, SSRs, and indels and SNPs. Color keys for the transposable elements and SSRs indicate the density of the repeats. The more intense the color, the more repetitive sequences in the pseudomolecule position. The indel and SNP density was determined based on sequence alignments between *S. lycopersicoides* and tomato. All tracks show binned data with a window size of 1 Mb.

https://doi.org/10.1371/journal.pone.0242882.g001

amplify targets in *S. lycopersicoides*, primers targeting 196 SSRs with di-nucleotide to hexa-nucleotide repeat motifs were selected (S1 Table). All primer pairs were 20–25 bp long and have an estimated amplicon size of 150–350 bp. Of the 196 SSRs, 182 successfully amplified targets in *S. lycopersicoides* Acc LA1964, with 148 annealing at 55˚C and two at 53˚C (Tables 4 and S2). Fifty-nine of the SSRs were multilocus, with 33 amplifying two bands and 26 amplifying more than two bands.

In addition, 149 indel markers (Tables 4 and S1) were also manually designed based on the alignment between the *S. lycopersicoides* draft assembly and the available reference genome for tomato. Of the 149 primer pairs, 143 successfully amplified targets in *S. lycopersicoides*

Table 2. SSRs mined from assembled *Solanum lycopersicoides* genome using GMATA software.

Motif(-mer) [a]	Total	Percentage (%)
2	42,135	74.05
3	12,691	22.30
4	1,353	2.38
6	350	0.62
5	303	0.53
7	64	0.11
9	3	0.01
8	2	0.00

[a]range of motif length was chosen while running the software.

https://doi.org/10.1371/journal.pone.0242882.t002

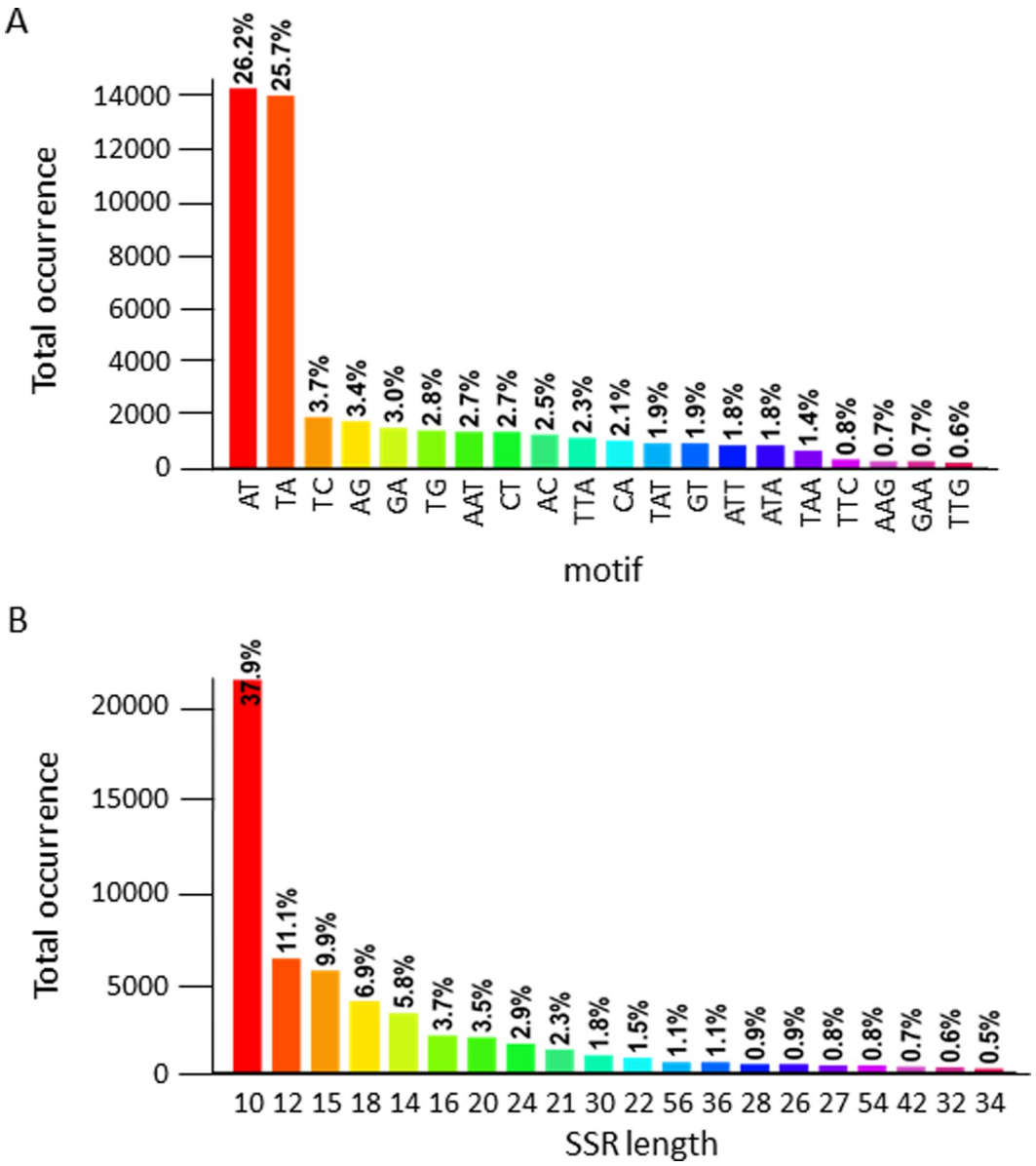

Fig 2. Distribution of the different (A) SSR motifs and (B) SSR lengths throughout the draft assembly of *S. lycopersicoides* based on GMATA data.

https://doi.org/10.1371/journal.pone.0242882.g002

Table 3. Summary of primer pairs designed by Primer3 of GMATA based on the SSRs mined.

	Total	Percentage (%)
total no. of loci detected	56,901	-
no. of SSR loci with designed primer pairs	35,801	62.90
no. of SSR loci without designed primer pairs	21,100	37.00
no. of unique markers	34,198	-

https://doi.org/10.1371/journal.pone.0242882.t003

genome, with 127 annealing at 55˚C and 16 annealing at 53˚C (S2 Table). Ten primer pairs amplified two bands, whereas two amplified more than two bands, indicating multilocus targets for the designed primers.

In summary, a total of 345 markers composed of 196 SSRs and 149 indels that are distributed across the 12 draft pseudomolecules of *S. lycopersicoides* were designed at an average map interval of 2.6 bp (Fig 3). Of these, 326 (94.50%) amplified targets in *S. lycopersicoides* Acc 1964. A slightly lower transferability of the markers was observed for *S. lycopersicoides* accessions LA2951 (68.00%) and LA2387 (70.00%), although multilocus amplifications were also observed. Of the total number of indels and SSRs tested, 11 and 24% amplified multiple targets in LA2951, and 27 and 49% in LA2387, respectively.

Previous studies on *S. lycopersicoides* have relied heavily on the use of RFLPs, SSRs and CAPS designed based on the tomato genome [15,21,56,57]. While these markers have been useful for genetic diversity studies and for monitoring wild chromosome introgressions, their distribution and number are not sufficient for mapping and cloning useful genes/QTLs in *S. lycopersicoides*. The newly designed markers in this study, 94% of which amplified their target loci, offers a much broader marker resource that can be used in genetics and breeding studies on *S. lycopersicoides*.

Cross-species transferability of *S. lycopersicoides*-specific markers

Cross-species amplification of all 345 *S. lycopersicoides*-specific markers in tomato, eggplant, silverleaf nightshade and pepper resulted in varying degrees of transferability ranging from 55% to 83% (Table 4 and Fig 4). In tomato, 148 SSRs and 138 indels amplified, with 20 markers showing multilocus targets. A total of 206 (59.71%) SSRs and indels amplified polymorphic targets between *S. lycopersicoides* and tomato, indicating the potential of these markers in mapping genes/QTLs in pre-breeding materials derived from crosses between the two species.

After tomato, silverleaf nightshade recorded the most number of target loci for the markers followed by eggplant and pepper. Section *Lycopersicoides* to which *S. lycopersicoides* belongs is an immediate outgroup of the tomato clade [53], making *S. lycopersicoides* closest to tomato. In contrast, pepper, which belongs to a separate genus, is the most distant to *S. lycopersicoides*.

Table 4. Target amplification and cross-species transferability of *S. lycopersicoides*-specific SSR and indel markers.

Plant species	No. of markers tested		No. of markers that amplified		Transferability rate (%)
	SSR	indel	SSR	indels	
S. lycopersicoides (Acc LA1964)	196	149	182	143	94.20
S. lycopersicoides (Acc LA2951)	196	149	123	114	68.69
S. lycopersicoides (Acc LA2387)	196	149	124	119	70.43
S. lycopersicum	196	149	148	138	82.90
S. melongena	196	149	111	90	58.26
S. elaeagnifolium	196	149	148	115	76.23
C. annum	196	149	105	83	54.49

https://doi.org/10.1371/journal.pone.0242882.t004

Fig 3. Chromosome distribution of *S. lycopersicoides*-specific SSR and indel markers. Map position of all the markers is based on their position in tomato chromosomes. Red markers are SSRs and blue markers are indels. Red triangle = centromere.

https://doi.org/10.1371/journal.pone.0242882.g003

The degree of the genetic relatedness of *S. lycopersicoides* to either species is consistent with the highest and lowest rate of marker cross-transferability recorded for tomato and pepper, respectively. In a similar manner, the phylogenetic relationship of *S. lycopersicoides* to eggplant and silverleaf nightshade reflects the observed rate of marker transferability to the latter two species. Silverleaf nightshade and eggplant belong to the subgenus *Leptostemonum* of the genus *Solanum*. Compared to tomato, silverleaf nightshade and eggplant are more distantly related

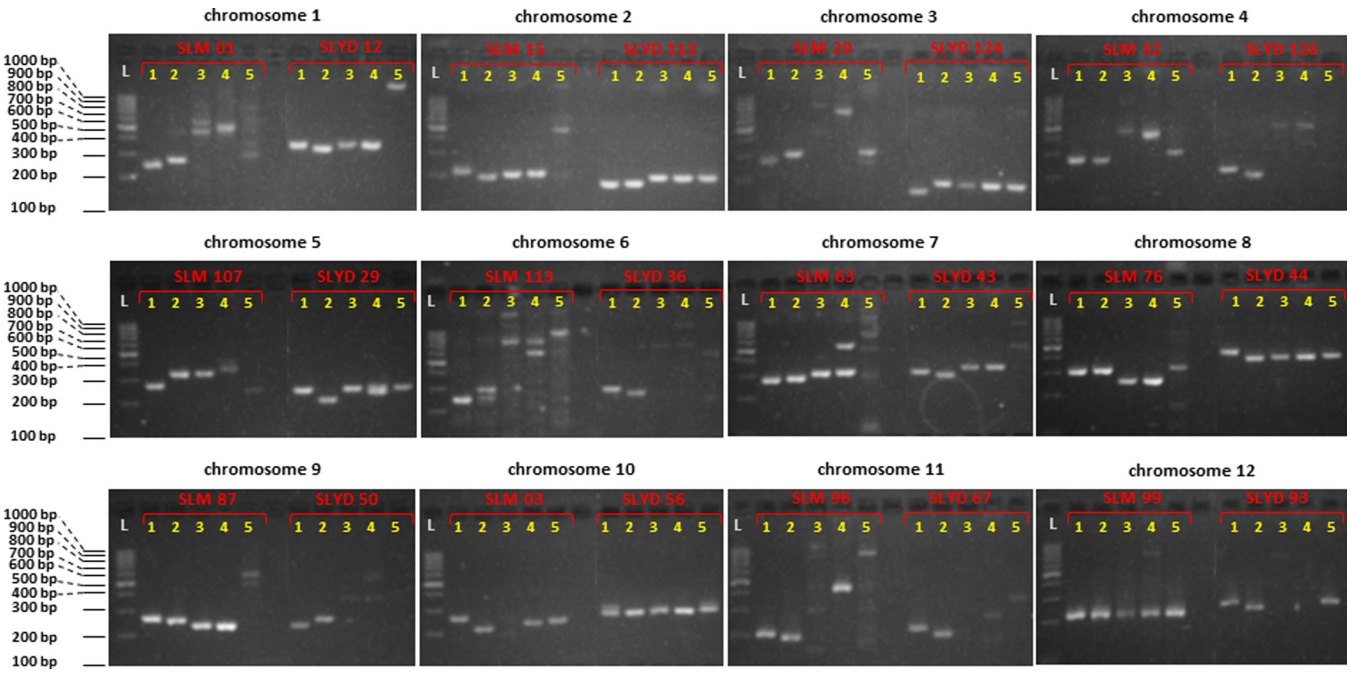

Fig 4. Cross-species amplification of *S. lycopersicoides*-specific markers in other members of Solanaceae. One SSR and one indel marker for each chromosome were used to amplify targets in tomato, silverleaf nightshade, eggplant and pepper. SLM = SSR marker, SLYD = indel marker, 1 = *S. lycopersicoides*, 2 = tomato, 3 = silverleaf nightshade, 4 = eggplant, 5 = pepper, L = 100 bp-ladder.

https://doi.org/10.1371/journal.pone.0242882.g004

to *S. lycopersicoides* [58] hence the lower rate of marker transferability in these two species compared to tomato. Comparative genomic studies in eggplant, potato, pepper and tomato indicate the highly conserved linkage order for markers despite the occurrence of major inversion events that drove the evolution of these related genomes [59–61]. This further support the relatively high transferability of *S. lycopersicoides* markers in the closely related species.

The generally high rates of cross-species amplification of *S. lycopersicoides* markers indicate their potential use in genetics and breeding applications in related Solanaceous plants. In fact, a subset of 54 *S. lycopersicoides*-specific, indel markers have been successfully used to assess the genetic diversity in silverleaf nightshade populations from different localities in Texas, USA. Genetic profiling using the indels, along with other DNA markers from related species, established the genetic differentiation of silverleaf nightshade populations in response to variations in selection pressures that are unique to the ecological habitats selected in the study [62].

Conclusions

Tomato production amidst worsening agro-environments can be sustained by harnessing natural genetic variation from wild tomato relatives that can provide durable forms of adaptation to the crop against both biotic and abiotic stresses. *S. lycopersicoides* is a distant tomato relative with known adaptation to marginal environments. Exploiting the genetic potential of *S. lycopersicoides* for tomato breeding will require understanding of the genetic basis of the adaptability of this wild species.

We designed and validated a total of 345 SSR and indel markers that are specific to *S. lycopersicoides* using whole genome sequence analysis. These markers, together with the more than 30,000 SSRs that are available for validation significantly expands the genetic marker resource that can be used for QTL analysis, mapping and positional cloning of genes in *S.*

lycopersicoides that can be utilized towards value-added trait improvements in tomato. The transferability of the *S. lycopersicoides* markers to tomato, eggplant, pepper and silverleaf nightshade indicate their applicability in similar genetics and breeding studies in these Solanaceous species.

Supporting information

S1 Table. *S. lycopersicoides*-specific DNA markers developed using whole genome sequence analysis.
(DOCX)

S2 Table. Forward and reverse primer sequences of *S. lycopersicoides*-specific markers.
(DOCX)

S1 Raw images. Amplification of *S. lycopersicoides*-specific markers in other Solanaceous species. One SSR and one indel marker for each chromosome were used to amplify targets in tomato, silverleaf nightshade, eggplant and pepper. SLM = SSR marker, SLYD = indel marker, 1 = *S. lycopersicoides*, 2 = tomato, 3 = silverleaf nightshade, 4 = eggplant, 5 = pepper, L = 100 bp-ladder. Lanes marked in X were not used to generate Fig 4.
(PDF)

Acknowledgments

The authors would like to thank the Tomato Genetics Resource Center (TGRC) of the University of California, Davis for generously providing us with the seeds of tomato Acc LA3122, and *S. lycopersicoides* accessions LA1964, LA2951 and LA2387.

Author Contributions

Conceptualization: Rosalyn B. Angeles-Shim.

Data curation: Puneet Kaur Mangat, Joshua J. Singleton.

Formal analysis: Puneet Kaur Mangat, Ritchel B. Gannaban.

Writing – original draft: Puneet Kaur Mangat, Rosalyn B. Angeles-Shim.

Writing – review & editing: Puneet Kaur Mangat, Rosalyn B. Angeles-Shim.

References

1. Bai Y, Lindhout P. Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? Ann Bot. 2007 Nov; 100(5):1085–94. https://doi.org/10.1093/aob/mcm150 PMID: 17717024.
2. Rick CM. Tomato-like nightshades: affinities, autoecology, and breeders' opportunities. Econ Bot. 1988 Apr; 42(2):145–54.
3. Li J, Liu L, Bai Y, Zhang P, Finkers R, Du Y, et al. Seedling salt tolerance in tomato. Euphytica. 2011; 178(3):403–14.
4. Wolf S, Yakir D, Stevens MA, Rudich J. 1986. Cold temperature tolerance of wild tomato species. J Amer Soc Hort Sci. 1986; 111:960–4.
5. Zhang L, Li Z, Li J, Wang A. Ectopic overexpression of SsCBF1, a CRT/DRE-binding factor from the nightshade plant *Solanum* lycopersicoides, confers freezing and salt tolerance in transgenic *Arabidopsis*. PLoS One. 2013 Jun 3; 8(6):e61810. https://doi.org/10.1371/journal.pone.0061810 PMID: 23755095.
6. Chetelat RT, Cisneros P, Stamova L, Rick CM. A male-fertile *Lycopersicon esculentum* x *Solanum lycopersicoides* hybrid enables direct backcrossing to tomato at the diploid level. Euphytica. 1997; 95(1): 99–108.

7. Davis J, Yu D, Evans W, Gokirmak T, Chetelat RT, Stotz HU. Mapping of loci from *Solanum lycopersicoides* conferring resistance or susceptibility to *Botrytis cinerea* in tomato. Theor Appl Genet. 2009; 119(2):305–14. https://doi.org/10.1007/s00122-009-1039-9 PMID: 19399472

8. Gradziel TM, Robinson RW. *Solanum lycopersicoides* gene introgression to tomato, *Lycopersicon esculentum*, through the systematic avoidance and suppression of breeding barriers. Sex Plant Reprod. 1989 Mar; 2:43–52.

9. Mazo-Molina C, Mainiero S, Hind SR, Kraus CM, Vachev M, Maviane-Macia F, et al. The *Ptr1* locus of *Solanum lycopersicoides* confers resistance to race 1 strains of *Pseudomonas syringae* pv. *tomato* and to *Ralstonia pseudosolanacearum* by recognizing the type III effectors AvrRpt2 and RipBN. Mol Plant Microbe Interact. 2019 Aug; 32(8), 949–60. https://doi.org/10.1094/MPMI-01-19-0018-R PMID: 30785360.

10. Pérez de Castro A, Díez MJ, Nuez F. Evaluation of a subset of Solanum lycopersicoides introgression lines for resistance to Tomato yellow leaf curl disease. In: Proceedings of the XVII Eucarpia Meeting Group—Tomato. Málaga, Spain, 2011, p. 17.

11. Zong YY, Liu L, Li T, Sayed-Rashid AS, Zhou LX, Sun YY, et al. Mapping of QTLs conferring the resistance to Tomato yellow leaf curl virus (TYLCV) in *Solanum lycopersicoides*. Acta Hortic Sin. 2012; 39 (5):915–22.

12. Soler S, Belmonte I, Aramburu J, Galipienso L, López C, Sifres A, et al. Identificación de una fuente de tolerancia al ToMV en una colección de líneas de introgresión derivada de Solanum lycopersicoides LA2951. In: Proceedings of the XVI Congreso de la Sociedad Española de Fitopatología, Málaga, Spain, 2012, p. 221.

13. Rick CM. Hybrids between Lycopersicon esculentum Mill. and Solanum lycopersicoides Dun. Proct Natl Acad Sci U S A. 1951 Nov; 37(11):741–4. https://doi.org/10.1073/pnas.37.11.741 PMID: 16578413.

14. Chetelat RT, Rick CM, DeVerna JW. Isozyme analysis, chromosome pairing, and fertility of *Lycopersicon esculentum* × *Solanum lycopersicoides* diploid backcross hybrids. Genome. 1989; 32(5):783–90.

15. Canady MA, Meglic V, Chetelat RT, Belzile F. A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. Genome. 2005 Aug; 48(4):685–97. https://doi.org/10.1139/g05-032 PMID: 16094436.

16. Brar DS, Khush GS. Alien introgression in rice. Plant Mol Biol. 1997; 35:35–47. PMID: 9291958

17. Shim RA, Ashikari M, Angeles ER, Takashi T. Development and evaluation of Oryza glaberrima Steud chromosome segment substitution lines in the background of O. sativa subsp. japonica cv. Koshihikari. Breeding Sci. 2010; 60:613–9. https://doi.org/10.1270/jsbbs.60.613

18. Furuta T, Uehara K, Angeles-Shim RB, Shim J, Nagai K, Ashikari M, et al. Development of chromosome segment substitution lines (CSSLs) harboring O. nivara segments and evaluation of yield-related traits. Breeding Sci. 2016; 66(5):845–50. https://doi.org/10.1270/jsbbs.16131 PMID: 28163601

19. Besho-Uehara K, Furuta T, Masuda K, Yamada S, Angeles-Shim RB, Ashikari M, et al. Construction of rice chromosome segment substitution lines harboring *Oryza barthii* genome and evaluation of yield-related traits. Breeding Sci. 2017; 67(4):408–415. https://doi.org/10.1270/jsbbs.17022 PMID: 29085251

20. Zhao L, Qiu C, Li J, Chai Y, Kai G, Li Z, et al. Investigation of disease resistance and cold tolerance of *Solanum lycopersicoides* for tomato improvement. HortSci. 2005; 40(1):43–6.

21. Albrecht E, Escobar M, Chetelat RT. Genetic diversity and population structure in the tomato-like nightshades *Solanum lycopersicoides* and *S. sitiens*. Ann Bot. 2010 Apr; 105(4):535–54. https://doi.org/10.1093/aob/mcq009 PMID: 20154348.

22. Peiró R, Díez MJ, Pùrez de Castro A. A set of PCR-based markers for management of a library of *Solanum lycopersicoides* introgression lines. J Hortic Sci Biotech. 2015 Nov; 90(3):279–84.

23. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012 May 30; 485(7400):635–41. https://doi.org/10.1038/nature11119 PMID: 22660326.

24. Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, et al. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol. 2015 May; 33(5):524–30. https://doi.org/10.1038/nbt.3208 PMID: 25893780.

25. Wu K, Yang M, Liu H, Tao Y, Mei J, Zhao Y. Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using insertion-deletion (InDel) and simple sequence repeat (SSR) markers. BMC Genet. 2014 Mar 19; 15:35. https://doi.org/10.1186/1471-2156-15-35 PMID: 24641723.

26. Ollitrault F, Terol J, Martin AA, Pina JA, Navarro L, Talon M, et al. Development of indel markers from Citrus *clementina* (Rutaceae) BAC-end sequences and interspecific transferability in Citrus. Am J Bot. 2012 Jul; 99(7);e268–73. https://doi.org/10.3732/ajb.1100569 PMID: 22733984.

27. Bus A., Hecht J., Huettel B., Reinhardt R., & Stich B. (2012). High-throughput polymorphism detection and genotyping in Brassica napus using next-generation RAD sequencing. BMC Genomics, 13(1), 281. https://doi.org/10.1186/1471-2164-13-281 PMID: 22726880

28. 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. 2001 Aug; 11(8):1441–52. https://doi.org/10.1101/gr.184001 PMID: 11483586.

29. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet. 2010 Oct 15; 19(R2):R131–6. https://doi.org/10.1093/hmg/ddq400 PMID: 20858594.

30. Terakami S, Matsumura Y, Kurita K, Kanamori H, Katayose Y, Yamamoto T, et al. Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. Tree Genet Genomes. 2012 Feb; 8:841–54.

31. Rockah-Shmuel L, Tóth-Petróczy Á, Sela A, Wurtzel O, Sorek R, Tawfik DS. Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. PLoS Genet. 2013 Oct; 9(10):e1003882. https://doi.org/10.1371/journal.pgen.1003882 PMID: 24204297.

32. Wu K, Yang M, Liu H, Tao Y, Mei J, Zhao Y. Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using insertion-deletion (InDel) and simple sequence repeat (SSR) markers. BMC Genet. 2014 Mar 19; 15:35. https://doi.org/10.1186/1471-2156-15-35 PMID: 24641723.

33. Angeles-Shim RB, Vinarao RB, Marathi B, Jena KK. Molecular analysis of *Oryza latifolia* Desv. (CCDD genome) derived introgression lines and identification of value-added traits for rice (O. sativa L.) improvement. J Hered. 2014 Sep-Oct; 105(5):676–89. https://doi.org/10.1093/jhered/esu032 PMID: 24939891.

34. Angeles-Shim RB, Shim J, Vinarao RB, Lapis RS, Singleton J. A novel locus from the wild allotetraploid rice, *Oryza latiolia* Desv. confers bacterial blight (*Xanthomonas oryzae* pv. oryzae) resistance in rice (O. sativa L.). PLoS One. 2020 Feb 21; 15(2):e0229155. https://doi.org/10.1371/journal.pone.0229155 PMID: 32084193.

35. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 1980 Oct; 8(19):4321–5. https://doi.org/10.1093/nar/8.19.4321 PMID: 7433111.

36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404.

37. Andrews S. FastQC: a quality control tool for high throughput sequence data. Version 0.11.9 [software]. 2010 [cited 2020 Mar 28]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

38. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009 Jun; 19(6):1117–23. https://doi.org/10.1101/gr.089532.108 PMID: 19251739.

39. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc. 2012 Jun 7; 7(7):1260–84. https://doi.org/10.1038/nprot.2012.068 PMID: 22678431.

40. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15; 29(8):1072–5. https://doi.org/10.1093/bioinformatics/btt086 PMID: 23422339.

41. Smit AFA, Hubley R. RepeatModeler Open-1.0 [software]. 2008–2015 [cited 2020 Mar 28]. Available from: http://www.repeatmasker.org.

42. www.girinst.org [Internet]. c2001-2020 [cited 2020 Nov 01]. RepeatMasker libraries; [about 3 screens]. Available from: https://www.girinst.org/downloads/.

43. Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0 [software]. 2013–2015 [cited 2020 Mar 28]. Available from: http://www.repeatmasker.org.

44. Wang X, Wang L. GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. Front Plant Sci. 2016 Sep 13; 7:1350. https://doi.org/10.3389/fpls.2016.01350 PMID: 27679641.

45. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004; 5(2):R12. https://doi.org/10.1186/gb-2004-5-2-r12 PMID: 14759262.

46. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [software]. 2013 [cited 2020 Mar 28]. Available from https://arxiv.org/abs/1303.3997.

47. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29(1):24–26. https://doi.org/10.1038/nbt.1754 PMID: 21221095

48. www.endmemo.com [Internet]. c2020 [cited 2020 Nov 01]. GC%; [about 2 screen]. Available from: http://www.endmemo.com/bio/gc.php.

49. Arep.med.harvard.edu [Internet]. c2020 [cited 2020 Nov 01]. Reverse and/or complement DNA sequences [about 5 screens]. Available from: http://arep.med.harvard.edu/labgc/adnan/projects/Utilities/revcomp.html.

50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5; 215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712.

51. Shim J, Torollo G, Angeles-Shim RB, Cabunagan RC, Choi IR, Yeo US, et al. Rice tungro spherical virus resistance into photoperiod-insensitive japonica rice by marker-assisted selection. Breed Sci. 2015 Sep; 65(4):345–51. https://doi.org/10.1270/jsbbs.65.345 PMID: 26366118.

52. Lander ES, Waterman MS. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics. 1988 Apr; 2(3):231–9. https://doi.org/10.1016/0888-7543(88)90007-9 PMID: 3294162

53. Peralta IE, Spooner DM, Knapp S. Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; *Solanaceae*). Syst Botany Monogr. 2008; 84:186.

54. Aflitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, et al. Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing. Plant J. 2014 Oct; 80 (1):136–48. https://doi.org/10.1111/tpj.12616 PMID: 25039268.

55. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 2007; 41:331–68. https://doi.org/10.1146/annurev.genet.40.110405.090448 PMID: 18076328.

56. Pertuzé RA, Ji Y, Chetelat RT. Comparative linkage map of the *Solanum lycopersicoides* and *S. sitiens* genomes and their differentiation from tomato. Genome. 2002 Dec; 45(6):1003–12. https://doi.org/10.1139/g02-066 PMID: 12502244.

57. Canady MA, Ji Y, Chetelat RT. Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics. 2006 Dec; 174(4):1775–88. https://doi.org/10.1534/genetics.106.065144 PMID: 17057228.

58. Stern S, Agra M, Bohs L. Molecular delimitation of clades within New World species of the "spiny solanums" (*Solanum* subg. *Leptostemonum*). Taxon. 2011 Oct; 60(5):1429–41. https://doi.org/10.1002/tax.605018 PMID: 32327851.

59. Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD. A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics. 2002 Aug; 161(4):1697–711. PMID: 12196412.

60. Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, et al. High density molecular linkage maps of the tomato and potato genomes. Genetics. 1992 Dec; 132(4):1141–60. PMID: 1360934.

61. Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK. Genome mapping in capsicum and the evolution of genome structure in the solanaceae. Genetics. 1999 Jul; 152(3):1183–202. PMID: 10388833.

62. Singleton JJ, Mangat PK, Shim J., Vavra C, Coldren C, ngeles-Shim RB. Cross-species transferability of Solanum spp. DNA markers and their application in assessing genetic variation in silverleaf nightshade (*Solanum elaeagnifolium*) populations from Texas, USA. Weed Sci. 2020:1–25. https://doi.org/10.1017/wsc.2020.25