Scientia Horticulturae 257 (2019) 108665

Contents lists available at ScienceDirect

Scientia Horticulturae

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

Refining of the egusi locus in watermelon using KASP assays

Lucky Paudel[a], Josh Clevenger[b], Cecilia McGregor[c,*]

[a] Institute for Plant Breeding, Genetics & Genomics, University of Georgia, 1111 Plant Sciences Bldg., Athens, GA, 30602, USA
[b] Institute for Plant Breeding, Genetics & Genomics, University of Georgia, 111 Riverbend Rd. CAGT, Athens, GA, 30602, USA
[c] Department of Horticulture and Institute for Plant Breeding, Genetics & Genomics, University of Georgia, 1111 Plant Sciences Bldg., Athens, GA, 30602, USA

ARTICLE INFO

Keywords:
Citrullus lanatus
Citrullus mucosospermus
Egusi watermelon
Seed coat type
QTL-seq
KASP assays

ABSTRACT

Egusi watermelon (*Citrullus mucosospermus*), a close relative of sweet watermelon, is an economically important crop grown in many West African countries for its protein and lipid rich edible seeds. Egusi watermelon seeds have a thick, fleshy mucilaginous seed coat layer surrounding the seed coat which is unique to egusi watermelon. The egusi seed trait is controlled by a single recessive Mendelian locus, *eg*, located on chromosome 6 from 6.75 Mb to 11.03 Mb. This region is 4.28 Mb wide and contains 241 candidate genes. The region lacks adequate markers for fine mapping and for marker-assisted selection (MAS) of the egusi trait. In this study, we used QTL-seq to validate the position of the *eg* locus and to identify SNP markers to refine the locus. A genomic region associated with the egusi trait was confirmed on chromosome 6 from 5.25 Mb to 7.85 Mb partially overlapping the previously mapped *eg* locus. SNPs identified from QTL-seq were used to design Kompetitive Allele Specific PCR (KASP™) assays for refining the *eg* locus. The *eg* locus was remapped at 86.3 cM on the genetic map, co-segregating with marker UGA6_7026576. The refined *eg* locus is 398.25 Kb long, extending from 6.95 Mb to 7.35 Mb, containing 30 candidate genes. Additional validation of the markers in a diversity panel and the validation population identified marker UGA6_7026576 as associated with the egusi phenotype and useful for MAS.

1. Introduction

Watermelon (*Citrullus lanatus* (Thumb.) Matsum. & Nakai) is an annual, vining, herbaceous crop which is cultivated throughout the world, predominantly for its sweet flesh. However, in several Asian and African countries different types of watermelons are cultivated for edible seeds and are collectively called edible seed watermelon (Zhang and Jiang, 1990). In Asia, most cultivated edible seed watermelon are *C. lanatus*, the same species grown for edible flesh, whereas in Africa, especially in West African countries, *C. mucosospermus* is cultivated. *Citrullus. mucosospermus* is a close relative of sweet watermelon (Guo et al., 2013) and is locally known as egusi watermelon. The term 'egusi' comes from the igbo and yoruba language spoken in Nigeria, meaning 'melon' (Adebayo and Yusuf, 2015). Egusi watermelons have round fruits, with light green rind and white, bland flesh. The large, flat seeds present within the fruit are used for human consumption and are a rich source of oil (approx. 35%), protein (approx. 28%), carbohydrate, vitamins and minerals (Oyolu, 1977a; Akobundu et al., 1982; Jarret and Levy, 2012; Prothro et al., 2012). The composition of egusi seed oil is similar to sunflower, soybean and safflower oil which makes it a potential feedstock for biodiesel production (Giwa et al., 2010; Bello and Makanju, 2011; Jarret and Levy, 2012).

The seed coat enveloping the seed in egusi watermelon has a typical fleshy outer mucilaginous layer when fresh. The mucilaginous tissue on the seed coat is considered a remnant of nucellar tissues and it appears during the second and the third week of seed development (Gusmini et al., 2004). Upon drying seeds, the mucilaginous layer becomes desiccated and seeds look like normal seeds however, rehydration causes the layer to reappear. This seed coat layer is unique to *C. mucosospermus*. However, not all accessions classified as *C. mucosospermus* have the egusi type seed coat. Accessions like plant introduction (PI) 189,317 do not possess the egusi type seed coat but are still classified as *C. mucosospermus* (*C. lanatus* subsp. *mucosospermus* in Guo et al. (2013)). Therefore, it is important to note that the egusi phenotype cannot be used to identify *C. mucosospermus*.

Oyolu (1977b) classified egusi seeds into five different classes based upon the morphological traits of seeds, including seed thickness, seed size and the thickness of seed edges, but does not discuss if all classes

Abbreviations: MAS, marker-assisted selection; KASP™, Kompetitive Allele Specific PCR; NGS, next generation sequencing; BSA, bulk segregant analysis; PI, plant introduction

* Corresponding author.
E-mail address: cmcgre1@uga.edu (C. McGregor).

https://doi.org/10.1016/j.scienta.2019.108665
Received 29 January 2019; Received in revised form 13 June 2019; Accepted 15 July 2019

possess egusi type seed coat. The type one class of miniature seeds with thin seed coat and flat edge looks morphologically similar to *C. lanatus* 'neri-type' in Achigan-Dako et al. (2015). This confusion is also complicated by the fact that in West Africa, where egusi watermelon is predominantly cultivated, several other cucurbits that do not have egusi type seed coat are collectively referred as egusi. In the current study, only seeds with a thick fleshy mucilaginous seed coat layer will be considered egusi type seed coat and watermelons with egusi type seed coat will be called egusi watermelon for the sake of simplicity.

The egusi seed coat is genetically controlled by a single recessive locus, *eg* (Gusmini et al., 2004; Prothro et al., 2012). The *eg* locus has been mapped to the region between markers NW0248325 and NW0250248 of the Strain II (PI 279461) x Egusi (PI 560023) genetic map (Prothro et al., 2012), which corresponds to the 6.75 Mb to 11.03 Mb region on chromosome 6 of the 97,103 reference genome (Guo et al., 2013). The region is 4.28 Mb long, contains 241 candidate genes and the closest marker, NW0248325, is 5.1 cM from the *eg* locus. This number of genes are still too large for candidate gene identification.

Recent advances in next generation sequencing (NGS) technology and the availability of reference genomes have facilitated the use of different tools to conduct a genome wide comparison to map loci and to rapidly detect large numbers of molecular markers throughout the genome. QTL-seq combines the principles of bulk segregant analysis (BSA) (Michelmore et al., 1991) and whole genome sequencing to map QTL and to identify markers simultaneously. QTL-seq has been used in mapping several major QTL in a wide variety of cereals (Takagi et al., 2013; Masumoto et al., 2016) and vegetables (Lu et al., 2014; Illa-Berenguer et al., 2015; Shu et al., 2018), including watermelon (Dong et al., 2018). The objective of this study is to refine the *eg* locus using the QTL-seq approach to generate abundant SNPs and to determine the utility of the linked SNPs for the MAS in different genetic backgrounds.

2. Materials and methods

2.1. Plant materials

An interspecific Strain II (*C. lanatus,* PI 279461) x Egusi (*C. mucosospermus*, PI 560023) F_2 mapping population (hereafter SII x Egusi) previously described by Sandlin et al. (2012) and used by Prothro et al. (2012) to map the *eg* locus was used for QTL-seq, fine mapping and identification of markers associated with the egusi phenotype. The F_2 plants were open pollinated among plants within the population, and the resulting F_3 plants were selfed in the greenhouse to produce F_4 seed.

A diversity panel consisting of 12 egusi seed coat type and 17 normal seed coat type PIs and cultivars was compiled to verify marker-phenotype association (Table 1). In addition, an F_2 population UGA147 (normal seed type, selection from PI 169233) x Egusi (hereafter 147 x Egusi) with 156 individuals was developed as a validation population to validate marker-phenotype association in an independent population (Fig. 1). Parental, F_1 and F_2 plants of the validation population were grown in the summer of 2017 at the Durham horticulture farm (Watkinsville, GA) and were visually phenotyped in the field.

2.2. DNA extraction for QTL-seq

Leaf tissue of F_3 plants originating from the open-pollinated F_2 progenies of the SII x Egusi population was used to extract DNA for preparation of bulks. Based on the phenotype of F_4 seeds obtained after selfing of F_3 plants, 7 F_3 plants, with normal type F_4 seeds, originating from 6 F_2 plants were selected to prepare the N-bulk and 7 F_3 plants, with egusi type F_4 seeds, originating from 6 F_2 plants were selected to prepare the E-bulk. DNA from all 14 samples was extracted using E. Z. N. A. Plant DNA kit (Omega Bio-Tek Inc., Norcross, GA) using the manufacturer's protocol. DNA concentrations were measured using an Infinite M200Pro plate reader (Tecan, Group Ltd., Mannerdorf, Switzerland). An equal amount of DNA was pooled from each individual to create the bulks. Both normal type and egusi type bulks were sent to the HudsonAlpha Institute for Biotechnology (Huntsville, AL) for library preparation and 151 base pair paired-end whole genome sequencing on the Illumina HiSeq X (Illumina, San Diego, CA).

2.3. DNA extraction for refining the egusi locus and marker validation

DNA from 139 freeze dried samples of the SII x Egusi population (Prothro et al., 2012; Sandlin et al., 2012) was extracted using the modified Dellaporta et al. (1983) protocol with a few modifications (CJ Tsai, personal communications). A TissueLyser II (QIAGEN, Hilden, Germany) was used to grind 35 mg of leaf tissue. For each sample, 525 µL of extraction buffer (50 mM of Tris – HCL pH 8, 10 mM EDTA pH 8, 100 mM NaCl, 1% SDS, and 10 mM β-mercaptoethanol) and 140 µg of proteinase K were added, followed by vortexing. Samples were incubated at 60 °C for 20 min, after which 140 µL of ice-cold 5 M potassium acetate (pH 4.8) was added and samples were incubated on ice for another 20 min. Samples were centrifuged for 25 min at 4000 rpm and 560 µL of supernatant was transferred into a new plate and mixed with 336 µL of isopropanol. The mixture was centrifuged at 4000 rpm for 10 min to precipitate the DNA. DNA pellets were washed twice with 70% ethanol, dried and resuspended in 100 µL diH_2O.

DNA from leaf tissue of 29 individuals of the diversity panel, 145 F_2 progenies of validation population and parents and F_1 of both mapping and validation population was extracted using the protocol described by King et al. (2014) with some modifications. Approximately 50 mg of frozen leaf tissue was ground using a TissueLyser II (QIAGEN) and 500 µL of extraction buffer mixture containing 40% (v/v) 5 M NaCl and 60% (v/v) Edward's extraction buffer (200 mM Tris/HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) was added to the ground samples. Samples were incubated at 65 °C for 20 min and centrifuged for 10 min at 3600 rpm. The supernatant was mixed with isopropanol (0.6 times the volume of supernatant) to precipitate the DNA. DNA pellets obtained after centrifuging at 3600 rpm for 10 min were washed with 70% ethanol, dried and resuspended in TE buffer.

2.4. Analysis of reads

A total of 79,551,582 and 155,774,971 reads were obtained from the next generation sequencing for the N-bulk and the E-bulk, respectively. The quality of raw reads was analyzed using FastQC (Andrews, 2010). The first two and last seven bases of the forward reads and the first two and last twenty bases of the reverse reads were trimmed to make sure that the average phred score for all base positions of all reads was higher than 20. A total of 99.02% of trimmed reads from the N-bulk and 98.90% of trimmed reads from the E-bulk were successfully aligned against the 97103 watermelon genome (Guo et al., 2013) using BWA and BWA MEM (Li and Durbin, 2009). The SAM files were converted to BAM files, sorted and indexed using SAM tools (Li et al., 2009). SAM tools was also used to calculate the genotype likelihood. SNP calling and filtering with a minimum depth of 10 were done using BCF tools and python tool. The SNP-index for all positions of the genome was calculated by counting the number of reads harboring the SNP as compared to the reference genome sequence and dividing it by the total number of reads. The SNP-indices between two bulks was subtracted to obtain the ΔSNP-index and a custom-made python script was used to generate a marker-specific threshold for $p < 0.01$ and $p < 0.001$ as described by Takagi et al. (2013). ΔSNP-index was plotted along with threshold $p < 0.001$ to identify the genomic region containing the egusi locus.

2.5. Kompetitive allele specific PCR (KASP™) assay design

Primers for KASP™ assays (LGC Genomics LLC, Teddington, UK) were designed (Table 2) for SNPs in the egusi region identified by QTL-seq and Prothro et al. (2012) using Primer3Plus (Untergasser et al.,

Table 1
Watermelon cultivars and Plant Introductions used to develop mapping population (MP), validation population (VP) and the diversity panel (DP) for mapping and marker validation of the egusi seed type locus.

Accession names	Species	Seed coat type	Origin	Uses	Genotype for KASP UGA6_7026576
Allsweet	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Black Diamond	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Blacktail Mountain	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Charleston Gray	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Cream of Saskatchewan	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Crimson Sweet	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Georgia Rattlesnake	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Mickylee	*C. lanatus*[b]	Normal	U.S.	DP	C:T
New Hampshire Midget	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Klondite Black seeded	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Orangeglo	*C. lanatus*[b]	Normal	U.S.	DP	T:T
PI 296341-FR	*C. amarus*[b]	Normal	South Africa	DP	T:T
ZWRM (PI 593359)	*C. lanatus*[b]	Normal	China	DP	C:C
Stars and Stripe	*C. lanatus*[b]	Normal	U.S.	DP	C:C
Strain II (PI 279461)	*C. lanatus*[b]	Normal	Japan	MP	T:T
Sugar Baby	*C. lanatus*[b]	Normal	U.S.	DP	T:T
Sugarlee	*C. lanatus*[b]	Normal	U.S.	DP	C:C
UGA147 (PI 169233)	*C. lanatus*[b]	Normal	Turkey	VP	T:T
PI 189317[a]	*C. mucosospermus*[c]	Normal	Nigeria	DP	T:T
PI 494532[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 559994[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 559997[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560006[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560010[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560011[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560014[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560017[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560018[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560020[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 560023 (Egusi)[a]	*C. mucosospermus*[b]	Egusi	Nigeria	MP, VP	C:C
PI 560024[a]	*C. mucosospermus*[b]	Egusi	Nigeria	DP	C:C
PI 595203[a]	*C. mucosospermus*[b]	Egusi	U.S.	DP	C:C

[a] Plant materials obtained from USDA-ARS Plant Genetic Resources Unit (Griffin, GA).
[b] Species classification based on USDA-ARS Plant Genetic Resources Unit (Griffin, GA).
[c] Species classification based on (Guo et al., 2013).

2007). PCR amplification was done using an S1000™ thermo cycler (Bio-Rad Laboratories, Inc., Hercules, CA) in a 4 µL reaction volume, comprised of 1.96 µL 2x low rox KASP™ master mix (LGC Genomics LLC), 0.06 µL of 0.81 mM primer mix, and 2 µL of 50–100 ng/µL genomic DNA. PCR conditions were set as 95 °C for 15 min, followed by 10 cycles of touch down PCR with 20 s of 95 °C, 25 s of primer annealing temperature +9 °C, with 1 °C decrease each cycle, and 15 s of 72 °C, then 35 cycles of 10 s at 95 °C, 1 min at primer annealing temp, and 15 s at 72 °C. KASP™ florescent end-point readings and genotyping calls were done using an Infinite M200Pro plate reader (Tecan, Group Ltd.) and KlusterCaller™ (LGC Genomics LLC), respectively. The genotype of a few individuals for some KASP™ markers could not be called because fluorescent signals were ambiguous. Such signals were considered missing data and were not included in calculation of phenotype prediction accuracy of markers. Missing data created differences in the total number of individuals with genotypic information available (n) for the different markers.

2.6. Mapping of the egusi locus and identification of candidate genes

One hundred and thirty-nine F_2 individuals from the SII x Egusi population were genotyped with 15 KASP™ assays. A genetic map was developed using JoinMap 5 software (Van Ooijen, 2006) and the maximum likelihood mapping function with a LOD score of 5. The final

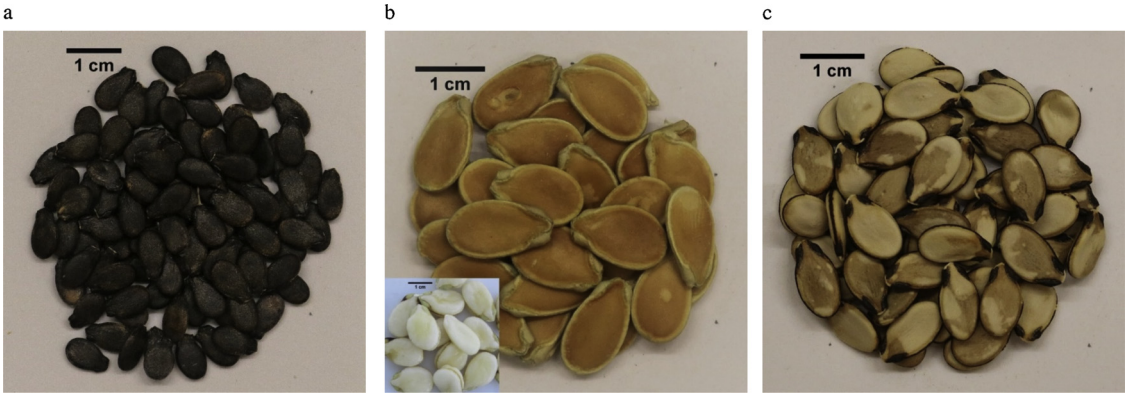

Fig. 1. Seeds of the parents used to develop mapping and validation populations. (a) Seeds of Strain II (PI 279461), (b) dried and fresh (lower left corner insert) seeds of PI 560023 (Egusi), (c) seeds of UGA147 (PI 169233).

Table 2
Primer sequence of KASP™ assays used for identification and narrowing down of the egusi (*eg*) locus in watermelon. Numbers after "UGA" indicate chromosome number and physical position of the SNP on the 97103 watermelon genome (Guo et al., 2013).

KASP assay	Primer type	Primer Sequence (5'-3')	Tm (°C)
UGA6_4756466	FAM	GAAGGTGACCAAGTTCATGCTGTGTTTTGACCGCAGTGCATC	63.9
	VIC	GAAGGTCGGAGTCAACGGATTGTGTTTTGACCGCAGTGCATT	63.9
	Reverse	GCAGCTTCAAGGCATCTTGT	60.5
UGA6_5333406	FAM	GAAGGTGACCAAGTTCATGCTTGCCACTAGCCAACCTTTAAAACC	64.5
	VIC	GAAGGTCGGAGTCAACGGATTTGCCACTAGCCAACCTTTAAAACT	64.5
	Reverse	AATGCATTTGACAACTCCTTCC	60.4
UGA6_5596144	FAM	GAAGGTGACCAAGTTCATGCTTTTTACATCTTGGGATAATTACGGATA	61.2
	VIC	GAAGGTCGGAGTCAACGGATTTTTTACATCTTGGGATAATTACGGATT	61.2
	Reverse	CACTCGATTGTTAAGGGGCATT	62.4
UGA6_5706068	FAM	GAAGGTGACCAAGTTCATGCTCCCATGACCTCCCTATTTTCAC	61.9
	VIC	GAAGGTCGGAGTCAACGGATTCCCATGACCTCCCTATTTTCAG	61.9
	Reverse	TCTTCGATGTCGAATGAATGGA	62.8
UGA6_5912358	FAM	GAAGGTGACCAAGTTCATGCTAGGAAAATCAAAATATACAGGTCGAGA	61.6
	VIC	GAAGGTCGGAGTCAACGGATTAGGAAAATCAAAATATACAGGTCGAGT	61.6
	Reverse	AACACATCTACCCCGGAGCTT	62.1
UGA6_6396525	FAM	GAAGGTGACCAAGTTCATGCTGGGATCAAAGAGCAACCAGTG	62.0
	VIC	GAAGGTCGGAGTCAACGGATTGGGATCAAAGAGCAACCAGTA	62.0
	Reverse	TTGTATCGTAACATTTGTTGTGTGC	61.0
UGA6_6501374	FAM	GAAGGTGACCAAGTTCATGCTGCCATGTTGTTTTTGCAAGTC	60.5
	VIC	GAAGGTCGGAGTCAACGGATTGCCATGTTGTTTTTGCAAGTT	60.5
	Reverse	TGATCAATGGGCCAAGTTTTG	63
UGA6_6737954	FAM	GAAGGTGACCAAGTTCATGCTTGGTCGTGGCTTACACATAAAAG	61.3
	VIC	GAAGGTCGGAGTCAACGGATTTGGTCGTGGCTTACACATAAAAC	61.3
	Reverse	AGTGGAGCTAAGGATTCCAACA	60.1
UGA6_6829416	FAM	GAAGGTGACCAAGTTCATGCTACCCATCCTTGTTCCTTCCTTT	62.2
	VIC	GAAGGTCGGAGTCAACGGATTACCCATCCTTGTTCCTTCCTTC	62.2
	Reverse	TGCTAACTGGCGTTTCAAGATAGA	62.4
UGA6_6903757	FAM	GAAGGTGACCAAGTTCATGCTTCAACTGACATTAAGTTCATACAATCG	60.6
	VIC	GAAGGTCGGAGTCAACGGATTTCAACTGACATTAAGTTCATACAATCT	60.6
	Reverse	TGTGGGTGGAAGAATCAAACC	62.1
UGA6_6958189	FAM	GAAGGTGACCAAGTTCATGCTCGCTTCAGTCGGCAGCTAT	61.6
	VIC	GAAGGTCGGAGTCAACGGATTCGCTTCAGTCGGCAGCTAC	61.6
	Reverse	TTTGCTCACCTATACTCAGACCCATC	63.8
UGA6_7026576	FAM	GAAGGTGACCAAGTTCATGCTAATGTAAATGAGGTTCAAAGATGTGAT	60.4
	VIC	GAAGGTCGGAGTCAACGGATTAATGTAAATGAGGTTCAAAGATGTGAC	60.4
	Reverse	CCTTTTTGGGACCTCCAAATGT	63.4
UGA6_7356440	FAM	GAAGGTGACCAAGTTCATGCTTACAACAATGCCTAAATCCAACC	60.1
	VIC	GAAGGTCGGAGTCAACGGATTTACAACAATGCCTAAATCCAACA	60.1
	Reverse	TGGAAACCAAGCCCCTTATTG	63.2
UGA6_7533594	FAM	GAAGGTGACCAAGTTCATGCTGGTGCATGGAATTCAAACTGAC	61.7
	VIC	GAAGGTCGGAGTCAACGGATTGGTGCATGGAATTCAAACTGAT	61.7
	Reverse	GATGGATGTAACGACGGTCAA	60.8
UGA6_8289797	FAM	GAAGGTGACCAAGTTCATGCTTGTCATGTGATGTGTGCTAAACTAAAC	61.6
	VIC	GAAGGTCGGAGTCAACGGATTTGTCATGTGATGTGTGCTAAACTAAAG	61.6
	Reverse	CCATGATGGTTGGTTCATAATTCA	62.8
UGA8_17929262	FAM	GAAGGTGACCAAGTTCATGCTCACCAATAGTGCATGTAACCCTCA	63.2
	VIC	GAAGGTCGGAGTCAACGGATTCACCAATAGTGCATGTAACCCTCG	63.2
	Reverse	GGTTGTCGAAGGTGGTCGTC	62.9

genetic map with the refined position of the egusi locus was drawn in MapChart (Voorrips, 2002). KASP™ assays which mapped close to the egusi locus in the genetic map were tested on the diversity panel. The marker with highest phenotypic prediction accuracy on the mapping population and the diversity panel was tested on the validation population. For identification of candidate genes the physical position of the markers flanking the refined egusi locus were identified on the 97103 reference genome (Guo et al., 2013) and genes present in the region were examined for their annotated functions, homologs and orthologs in the Cucurbit Genomics Database (http://cucurbitgenomics.org/).

3. Results

3.1. Validation of inheritance of the egusi phenotype

The validation population, 147 x Egusi, was phenotyped for seed coat type. The F_1 plants had a normal seed coat, confirming that the normal type is dominant over the egusi type and the F_2 individuals segregated at the ratio of 108 normal: 48 egusi ($\chi^2_{(0.05,1)}$, $P = 0.96$). These results confirm the conclusion made by Gusmini et al. (2004) and Prothro et al. (2012) that the trait is controlled by a single locus, with the egusi type being recessive.

3.2. QTL-Seq

From QTL-seq, 699,060 SNPs were detected between the N-bulk and E-bulk. When plotting the ΔSNP-index against the genome position, two statistically significant ($p < 0.001$) peaks were identified. The first peak was located from 5.25 Mb to 7.85 Mb on chromosome 6 and the second was located at 17.65 Mb to 19.63 Mb on chromosome 8 (Fig. 2). The peak on chromosome 6 partially overlaps the previously mapped egusi locus (Prothro et al., 2012). SNP markers within both peak regions (UGA6_7026576 and UGA8_17929262) were utilized to design KASP™assays to test association of the peaks with the phenotype in both mapping and validation populations. KASP™ assay UGA6_7026576 on chromosome 6 was highly associated to the egusi phenotype in both mapping and validation populations. However, the peak on the chromosome 8 did not associate with the phenotype. The recombination

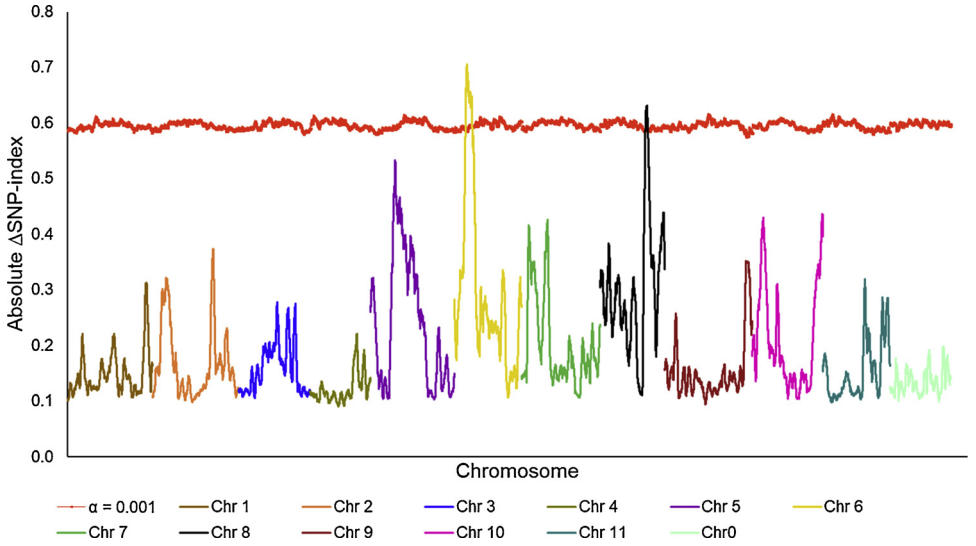

Fig. 2. Absolute ΔSNP-index graph plot of all chromosomes obtained from QTL-seq analysis. X-axis indicates chromosomes of watermelon placed in sequential order from 1 to 11 and 0 denoted by different colors. Y-axis indicates absolute ΔSNP-index. Absolute ΔSNP-index for each chromosome were calculated using 1Mb sliding window with a 10Kb increment increase and plotted with statistical threshold ($P < 0.001$) (horizontal red line) with criteria that no QTL below threshold. Two significant peaks on chromosome 6 (yellow) and chromosome 8 (black) were identified (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

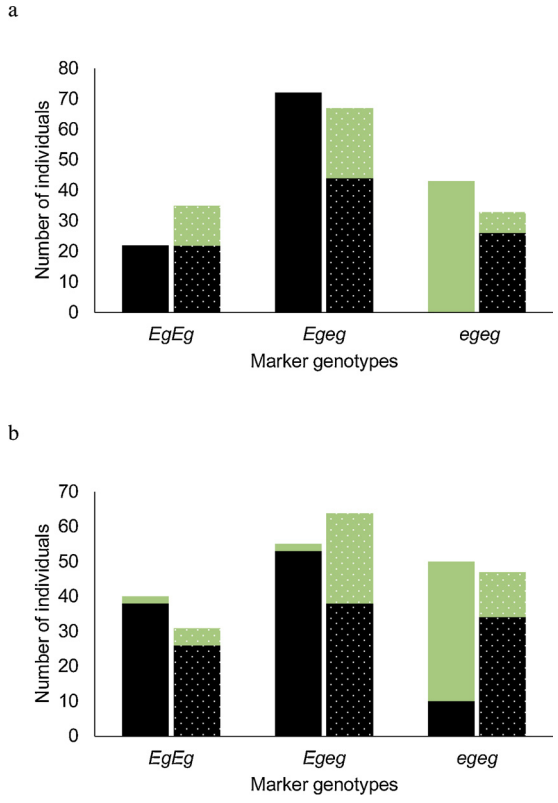

Fig. 3. Bar graphs showing the marker phenotype association between seed coat phenotype and KASP™ markers UGA6_7026576 (solid) and UGA8_17929262 (dotted) in the (a) Strain II (PI 279461) x Egusi (PI 560023) and (b) UGA147 (PI 169233) x Egusi population. Black and green sections denote the number of individuals with normal and egusi type seed coat, respectively. The x-axis genotypes represent the marker genotypes coded as the normal (*Eg*) or egusi (*eg*) type allele (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

frequency between KASP™ marker UGA8_17929262 and the egusi phenotype was 62.06 and 37.52 mapping units in the SII x Egusi and 147 x Egusi populations, respectively, confirming that the genomic region identified on chromosome 8 was a false discovery (Fig. 3a, b).

3.3. Refining the egusi locus

We designed 21 KASP™assays in the egusi region identified by QTL-seq and Prothro et al. (2012). Fifteen assays that gave the expected genotypes in the parental and F_1 plants were utilized to genotype the population for fine mapping the egusi locus (Table 2). The egusi locus was remapped to the 86.3 cM position, co-segregating with marker UGA6_7026576 on chromosome 6 of the SII x Egusi genetic map (Fig. 4). Marker UGA6_6958189, the closest flanking marker, was 1.3 cM away from the *eg* locus, whereas the other flanking marker UGA6_7356440 was 3.1 cM away from the *eg* locus. In the 97103

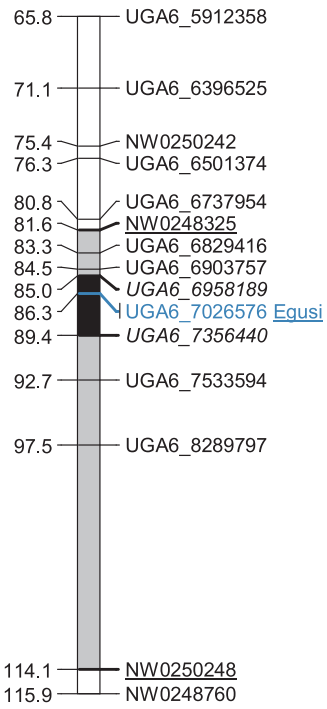

Fig. 4. Partial genetic map of chromosome 6 of the Strain II (PI 279461) x Egusi (PI 560023) F_2 watermelon population showing the refined position of the egusi locus. The grey region (flanking markers underlined) is the egusi locus described by Prothro et al. (2012), the black region (flanking markers italic) is the refined region with the egusi locus (blue) and UGA6_7026576 (blue). Map positions are in cM (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 3
Candidate genes present in the refine egusi locus from 6.95 Mb to 7.35 Mb based on the 97103 watermelon genome (Guo et al., 2013) along with their annotated function.

Gene ID	Annotated function
Cla007521	Homocysteine s-methyltransferase
Cla007522	Inositol 1 4 5-trisphosphate 5-phosphatase
Cla007523	Stress-induced-phosphoprotein 1
Cla007524	IQ calmodulin-binding motif family protein
Cla007525	Uncharacterized protein
Cla007526	Protein forked1
Cla007527	Protein forked1
Cla007528	Transmembrane BAX inhibitor motif-containing protein 4
Cla007529	DNA-directed RNA polymerase III subunit RPC9, transcription of DNA
Cla007530	Epidermal patterning factor-like protein 5, Stomata patterning
Cla007531	Protein TIFY 4A
Cla007532	Acireductone dioxygenase
Cla007533	Aldo/keto reductase
Cla007534	Aldo/keto reductase
Cla007535	Sulfate adenylyltransferase
Cla007536	Calmodulin
Cla007537	Os11g0586300 protein
Cla007538	G-protein gamma-subunit 1
Cla007539	Pyruvate kinase
Cla007540	Uncharacterized protein
Cla007541	Plant-specific domain TIGR01615 family protein, Uncharacterized protein
Cla007542	CBL-interacting protein kinase 20
Cla007543	Acylamino-acid-releasing enzyme, prolyl oligopeptidase active site region
Cla007544	Auxin responsive protein
Cla007545	Auxin responsive protein
Cla007546	Transport protein sec23
Cla007547	Magnesium transporter MRS2-2
Cla007548	Voltage-gated potassium channel beta subunit
Cla007549	Sieve element occlusion
Cla007550	Sieve element occlusion

genome physical map (Guo et al., 2013), the refined egusi locus is 398.25 Kb and extends from 6.95 Mb to 7.35 Mb and contains 30 candidate genes (Table 3).

3.4. Marker-phenotype association

In the SII x Egusi population, three KASP™ markers: UGA6_7026576, UGA6_6958189 and UGA6_6903757, were able to accurately predict 100% (137/137), 100% (139/139) and 99.26% (135/136) of the F_2 individuals (Fig. 5a). We further tested these markers on the diversity panel. Marker UGA6_7026576 was fixed in all the egusi genotypes but segregating in the normal type watermelon (Table 1). Cultivars "Allsweet", "Blacktail Mountain", "Charleston Gray", "Crimson Sweet", "Georgia Rattlesnake", "Stars and Stripes", "Sugarlee" and genotype ZWRM (PI 593359) had the same genotype as the egusi seed type watermelons. Among these cultivars, "Allsweet", "Charleston Gray", "Crimson Sweet" and "Sugarlee" share ancestry and are genetically related. The performance of marker UGA6_7026576 was assessed on the independent validation population, 147 x Egusi population, to evaluate its utility in predicting phenotype in a segregating population. UGA6_7026576 was able to accurately predict 90.34% (131/145) of phenotypes (Fig. 5b) indicating that the marker will be useful in MAS of egusi phenotype provided parents are segregating for the marker.

4. Discussion

The QTL-seq approach is a powerful tool which has been successfully utilized in QTL mapping and SNP discovery. The power of QTL-seq to reliably identify genomic regions associated with a trait depends on the nature and size of the population, the number of individuals in each bulk, the heritability of the QTL and the phenotypic variation explained by the QTL. Studies deploying QTL-seq have shown that an F_2 population with 200–300 individuals utilizing 15–20% of the individuals for each bulk is sufficient to identify both major and minor QTL (Magwene et al., 2011; Takagi et al., 2013; Tiwari et al., 2016). An F_2 population with 100 individuals has been shown to be sufficient in identifying a single qualitative locus or a QTL (Takagi et al., 2013; Illa-Berenguer et al., 2015).

In our study, we used QTL-seq to map a qualitative trait, *eg*, in watermelon. Each bulk was comprised of DNA from seven F_3 individuals originating from six open pollinated F_2 individuals. Lack of good quality leaf tissue or selfed seeds from the F_2 individuals prevented us from increasing bulk size. Nonetheless, we were able to map a significant peak partially overlapping the previously mapped egusi locus. We believe that the qualitative nature of the egusi trait made it possible to map the *eg* locus even with such non-ideal bulks. However, the use of smaller sized bulks increases the possibility of having bulks that are different, not only for alleles linked to the trait of interest but also for other alleles. This increases the probability of mapping false positives (Giovannoni et al., 1991; Michelmore et al., 1991; Tiwari et al., 2016). In this study, we also detected a significant peak on chromosome 8. However, the results from the chi-square test and an association test between the phenotype and the genotype of UGA8_17929262 in the peak region in both mapping and validation populations confirmed that this peak was a false positive and that the egusi phenotype is controlled by a single locus on chromosome 6 (Fig. 3a, b) as previously found by Gusmini et al. (2004) and Prothro et al. (2012).

QTL-seq identified 3112 SNPs in the egusi region, providing plenty of resources for developing KASP™ assays in the region. Fifteen KASP™ markers were added in the region of the egusi locus identified by QTL-

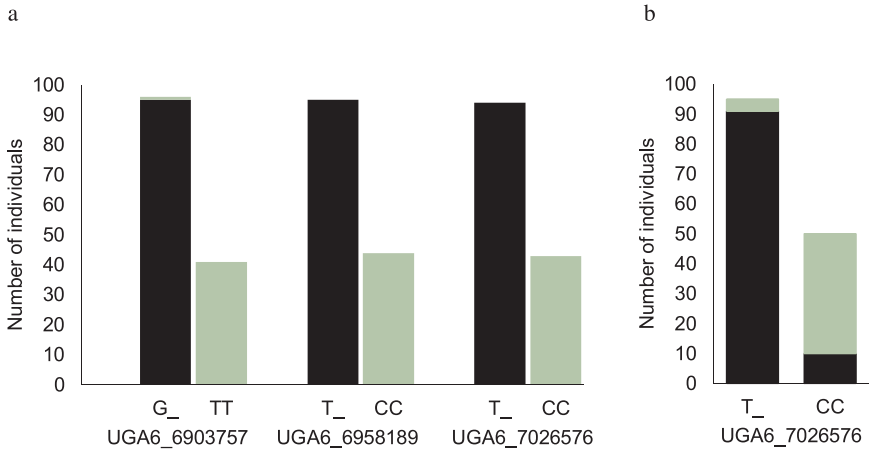

Fig. 5. (a) Genotypic and phenotypic data for KASP™ markers UGA6_6903757 (n = 136), UGA6_6958189 (n = 139) and UGA6_7026576 (n = 137), in the mapping population, Strain II (PI 279461) x Egusi (PI 560023), and (b) KASPTM™ marker UGA6_7026576 in the validation population UGA147 (PI 169233) x Egusi, (n = 145). X-axis indicates genotypes (G_ = GG or GT and T_ = TT or TC) of respective KASP™ markers and y-axis indicates number of F_2 individuals. Black and green sections denote the number of individuals with normal and egusi type seed coat, respectively (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

seq and Prothro et al. (2012). As expected, the closer the marker mapped to the egusi locus (less recombination), the better the marker was at predicting the phenotype. Marker UGA6_7026576 co-segregated with the egusi phenotype in the mapping population (Figs. 4 and 5) and had a 100% prediction efficiency in this population. In the larger validation population recombination between UGA6_7026576 and the egusi locus were observed, leading to lower prediction accuracy. The historical recombination in the diversity panel led to poor performance of UGA6_6903757 and UGA6_6958189 that were mapped further away from the egusi locus.

In the 97,103 genome physical map, the *eg* locus was narrowed from 4.29 Mb to 398.25 Kb and the number of candidate genes between the flanking markers was decreased from 249 to 30. We examined the annotated function of each gene present in the region of interest and their homologs and orthologs in *Arabidopsis* and other cucurbits, however, none were previously associated with seed coat development. Three uncharacterized genes (Cla007525, Cla007540 and Cla007541) were also found in this region. It is possible that none of these 30 genes are candidate genes. The 97103 reference genome we used to align reads and to identify candidate genes may lack the gene conferring the egusi phenotype since cultivar 97103 belongs to *C. lanatus*. The candidate gene may be present only in the genomic region specific to the *C. mucosospermus* species because the egusi phenotype is present only in *C. mucosospermus* species. However, the unavailability of a public reference genome of *C. mucosospermus* species hinders confirming this hypothesis.

The flesh of egusi watermelons is white, hard, bland and inedible and harvesting seeds from the hard flesh is a very time consuming and unsanitary process since fruit is cracked and then left to rot for several days (Oyolu, 1977b), but recently, breeders have started breeding egusi watermelons with red, sweet, edible flesh (Orji et al., 2016). The availability of egusi watermelon with both edible flesh and seed will increase food security in many West African countries. However, both favorable traits, red flesh and the egusi trait are recessive to white flesh and normal seed phenotype, respectively. Therefore, breeding for watermelon with these traits without MAS will be difficult and time consuming. The KASP™ marker UGA6_7026576 we developed can be used for MAS for the egusi trait in populations segregating for this marker.

Funding

This research was supported by The Nepal Fulbright Commission, Fulbright Foreign Student Program fellowship (grant number G-1-00001), awarded to LP. This research was partially supported by a U.S. Department of Agriculture Research and Education grant (GEO-2009-04819) and a U.S. Department of Agriculture Specialty Crop Research Initiative grant (2014-51181-22471).

Declaration of Competing Interest

None

Acknowledgement

We would like to acknowledge Reeve Daniel Legendre, Leigh Ann Fall, Winne Gimode, Jorge Reyes, Dr. Yihua Chen and Jesse Kuzy for their help in this research and critically reading the manuscript.

References

Achigan-Dako, E.G., Avohou, E.S., Linsoussi, C., Ahanchede, A., Vodouhe, R.S., Blattner, F.R., 2015. Phenetic characterization of *Citrullus* spp. (*Cucurbitaceae*) and differentiation of egusi-type (*C. mucosospermus*). Genet. Resour. Crop Evol. 62, 1159–1179. https://doi.org/10.1007/s10722-015-0220-z.

Adebayo, A., Yusuf, K., 2015. Performance evaluation of an egusi-melon washing machine. Analele Universitatii' Eftimie Murgu' 22, 11–22.

Akobundu, E., Cherry, J., Simmons, J., 1982. Chemical, functional, and nutritional properties of egusi (*Colocynthis citrullus* L.) seed protein products. J. Food Sci. 47, 829–835. https://doi.org/10.1111/j.1365-2621.1982.tb12725.x.

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.

Bello, E., Makanju, A., 2011. Performance evaluation of Egusi melon (*Citrullus colocynthis* L.) seeds oil biodiesel. J. Emerg. Trends Eng. Applied Sci. 2, 741–745.

Dellaporta, S.L., Wood, J., Hicks, J.B., 1983. A plant DNA minipreparation: version II. Plant Mol. Biol. Rep. 1, 19–21.

Dong, W., Wu, D., Li, G., Wu, D., Wang, Z., 2018. Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. Sci. Rep. 8, 1–7. https://doi.org/10.1038/s41598-018-21293-1.

Giovannoni, J.J., Wing, R.A., Ganal, M.W., Tanksley, S.D., 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res. 19, 6553–6568. https://doi.org/10.1093/nar/19.23.6553.

Giwa, S., Abdullah, L.C., Adam, N.M., 2010. Investigating "Egusi"(*Citrullus colocynthis* L.) seed oil as potential biodiesel feedstock. Energies 3, 607–618. https://doi.org/10.3390/en3040607.

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W.J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., Wang, Z., 2013. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat. Genet. 45, 51–58. https://doi.org/10.1038/ng.2470.

Gusmini, G., Wehner, T., Jarret, R., 2004. Inheritance of egusi seed type in watermelon. J. Hered. 95, 268–270. https://doi.org/10.1093/jhered/esh031.

Illa-Berenguer, E., Van Houten, J., Huang, Z., van der Knaap, E., 2015. Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. Theor. Appl. Genet. 128, 1329–1342. https://doi.org/10.1007/s00122-015-2509-x.

Jarret, R.L., Levy, I.J., 2012. Oil and fatty acid contents in seed of *Citrullus lanatus* Schrad. J. Agric. Food Chem. 60, 5199–5204.

King, Z., Serrano, J., Boerma, H.R., Li, Z., 2014. Non-toxic and efficient DNA extractions for soybean leaf and seed chips for high-throughput and large-scale genotyping. Biotechnol. Lett. 36, 1875–1879. https://doi.org/10.1007/s10529-014-1548-8.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., Sun, J., Zhang, Z., Weng, Y., Huang, S., 2014. QTL-seq identifies an early flowering QTL located near flowering locus T in cucumber. Theor. Appl. Genet. 127, 1491–1499. https://doi.org/10.1007/s00122-014-2313-z.

Magwene, P.M., Willis, J.H., Kelly, J.K., 2011. The statistics of bulk segregant analysis using next generation sequencing. Journal 7, e1002255. https://doi.org/10.1371/journal.pcbi.1002255.

Masumoto, H., Takagi, H., Mukainari, Y., Terauchi, R., Fukunaga, K., 2016. Genetic analysis of NEKODE1 gene involved in panicle branching of foxtail millet, Setaria italica (L.) P. Beauv., and mapping by using QTL-seq. Journal 36, 59. https://doi.org/10.1007/s11032-016-0481-z.

Michelmore, R.W., Paran, I., Kesseli, R., 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. 88, 9828–9832. https://doi.org/10.1073/pnas.88.21.9828.

Orji, V.O., Ogbonna, P.E., Onyia, V.N., Okechukwu, E.C., 2016. Improvement of fruit taste quality of 'egusi'melon (*Colocynthis citrullus* L.) through hybridization with watermelon (*Citrullus lanatus* Thumb.). J. Tropical Agri. 54, 1–6.

Oyolu, C., 1977a. Extraction rates and chemical composition of seed types in egusi (*Colocynthis citrullus* L.). Acta Hort. 53, 287–290.

Oyolu, C., 1977b. A quantitative and qualitative study of seed types in egusi (*Colocynthis citrullus* L.). Tropical Sci. 19, 287–290.

Prothro, J., Sandlin, K., Gill, R., Bachlava, E., White, V., Knapp, S.J., McGregor, C., 2012. Mapping of the egusi seed trait locus (*eg*) and quantitative trait loci associated with seed oil percentage in watermelon. J. Am. Soc. Hortic. Sci. 137, 311–315.

Sandlin, K., Prothro, J., Heesacker, A., Khalilian, N., Okashah, R., Xiang, W., Bachlava, E., Caldwell, D.G., Taylor, C.A., Seymour, D.K., 2012. Comparative mapping in watermelon [*Citrullus lanatus* (Thunb.) Matsum. et Nakai]. Theor. Appl. Genet. 125, 1603–1618. https://doi.org/10.1007/s00122-012-1938-z.

Shu, J., Liu, Y., Zhang, L., Li, Z., Fang, Z., Yang, L., Zhuang, M., Zhang, Y., Lv, H., 2018. QTL-seq for rapid identification of candidate genes for flowering time in broccoli × cabbage. Theor. Appl. Genet. 131, 917–928. https://doi.org/10.1007/s00122-017-3047-5.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., Uemura, A., Utsushi, H., Tamiru, M., Takuno, S., 2013. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74, 174–183. https://doi.org/10.1111/tpj.12105.

Tiwari, S., Krishnamurthy, S., Kumar, V., Singh, B., Rao, A., Sv, A.M., Rai, V., Singh, A.K., Singh, N.K., 2016. Mapping QTLs for salt tolerance in rice (*Oryza sativa* L.) by bulked segregant analysis of recombinant inbred lines using 50K SNP chip. Journal 11, e0153610. https://doi.org/10.1371/journal.pone.0153610.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A., 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 35, W71–W74. https://doi.org/10.1093/nar/gkm306.

Van Ooijen, J., 2006. JoinMap 4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma B.V, Wageningen.

Voorrips, R., 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. J. Hered. 93, 77–78. https://doi.org/10.1093/jhered/93.1.77.

Zhang, X., Jiang, Y., 1990. Edible seed watermelons (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) in Northwest China. Report-Cucurbit Genetics Cooperative 13, 40–42.