

# De Novo Transcriptome Analysis of *Cucumis melo* L. var. *makuwa*

Hyun A Kim<sup>1,5</sup>, Ah-Young Shin<sup>1,5</sup>, Min-Seon Lee<sup>2</sup>, Hee-Jeong Lee<sup>2</sup>, Heung-Ryul Lee<sup>2</sup>, Jongmoon Ahn<sup>2</sup>, Seokhyeon Nahm<sup>2</sup>, Sung-Hwan Jo<sup>3</sup>, Jeong Mee Park<sup>1,4</sup>, and Suk-Yoon Kwon<sup>1,4,\*</sup>

Oriental melon (*Cucumis melo* L. var. *makuwa*) is one of six subspecies of melon and is cultivated widely in East Asia, including China, Japan, and Korea. Although oriental melon is economically valuable in Asia and is genetically distinct from other subspecies, few reports of genome-scale research on oriental melon have been published. We generated 30.5 and 36.8 Gb of raw RNA sequence data from the female and male flowers, leaves, roots, and fruit of two oriental melon varieties, Korean landrace (KM) and Breeding line of NongWoo Bio Co. (NW), respectively. From the raw reads, 64,998 transcripts from KM and 100,234 transcripts from NW were *de novo* assembled. The assembled transcripts were used to identify molecular markers (e.g., single-nucleotide polymorphisms and simple sequence repeats), detect tissue-specific expressed genes, and construct a genetic linkage map. In total, 234 single-nucleotide polymorphisms and 25 simple sequence repeats were screened from 7,871 and 8,052 candidates, respectively, between the KM and NW varieties and used for construction of a genetic map with 94 F2 population specimens. The genetic linkage map consisted of 12 linkage groups, and 248 markers were assigned. These transcriptome and molecular marker data provide information useful for molecular breeding of oriental melon and further comparative studies of the Cucurbitaceae family.

## INTRODUCTION

Melon (*Cucumis melo* L.) is an important cultivated cucurbit and produces an aromatic, sweet fruit. Melons exhibit significant variation in terms of botanical phenotypes and fruit types, high-

lighting the genetic diversity of the Cucurbitaceae family (Mliki et al., 2001). Six melon subspecies are cultivated: *cantalupensis*, *reticulatus*, *inodorus*, *acidulous*, *saccharinus*, and *makuwa* (Liu et al., 2004).

Melon is a diploid species, with a basic number of chromosomes ( $x = 12$  [ $2x = 2n = 24$ ]) and an estimated genome size of 450 Mb, similar to that of rice (419 Mb). The melon genome is being sequenced as part of the Spanish Genome Initiative (MELOGENOMICS). Moreover, BAC libraries, high-resolution genetic maps, oligo-based microarrays, and a large number of transcriptome sequences (RNA-Seq and expressed sequence tag(EST)) for melon are also available as genetic and genomic tools.

Oriental melon (*C. melo* L. var. *makuwa*) is cultivated in large areas of Asia, especially throughout the temperate regions of China, Japan, and Korea. KM (Korean landrace; Gotgam) is one of the major landrace of oriental melon in Korea. This landrace contains more nutrients and has useful traits with greater disease resistance than other varieties. NW is high quality breeding line of NongWoo Bio Company with deep yellow, middle sized, oval type fruit and contained high sugar contents. Although oriental melon is economically valuable in Asia and exhibits genetic features distinct from the other subspecies, few reports of genome-scale studies of oriental melon have been published. Therefore, development of genomic resources and markers of these oriental melon will be important for the characterization of genetic markers linked to highly desirable traits and oriental melon breeding.

Due to the high-throughput capacity of next-generation sequencing (NGS) technology, which was developed in the last decade, transcriptome analysis has become widely used for genome-scale studies. Transcriptome analysis can be used to profile gene expression and identify novel transcripts, splicing isoforms, and sequence variations, including single-nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs). In the present work, we generated a total of 67,440,566,178 raw sequence reads (67.4 Gb) from the female and male flowers, leaves, roots, and fruit of two oriental melon varieties (KM and NW). A total of 64,998 transcripts from KM and 100,234 from NW specimens were *de novo* assembled, with transcript N50 values of 939 and 1,138, respectively. A total of 259 SNP and SSR markers were developed, and genotype- and species-specific genes were identified. The transcripts generated in the present study will be a useful resource for the characterization of gene expression patterns and traits in oriental melon.

<sup>1</sup>Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea, <sup>2</sup>Nongwoo Bio Co., Ltd., Yeosu 469-885, Korea, <sup>3</sup>SEEDERS, Daeduk Industry Academic Cooperation Building, Daejeon 34016, Korea, <sup>4</sup>Biosystems and Bioengineering Program, University of Science and Technology, Daejeon 305-350, Korea, <sup>5</sup>These authors contributed equally to this work.  
\*Correspondence: sykwon@kribb.re.kr

Received 2 October, 2015; revised 30 October, 2015; accepted 2 November, 2015; published online 7 January, 2016

**Keywords:** genetic linkage map, Korean melon, simple sequence repeat, single-nucleotide polymorphism, transcriptome analysis

## MATERIALS AND METHODS

### Plant materials and mRNA sequencing

To characterize the oriental melon transcriptome and increase the sequence coverage of *C. melo*, sequence libraries for two oriental melon cultivars were generated. Specimens of the KM (Korean landrace) and NW (Breeding line of NongWoo Bio Co.) cultivars were provided by NongWoo Bio Company for use in the study (Supplementary Fig. S1). Thirty two seeds of each cultivar were sowed in tray and grown in a greenhouse until 3–4 true leaf stage. Ten healthy plants of each cultivar were transplanted into plastic pot and grown until flowering stage in greenhouse in which temperature was maintained 28°C in daytime and 24°C at night. Total RNA was isolated from the leaves, roots, female and male flowers, and fruit of both cultivars using TRIzol reagent according to the manufacturer's instructions. RNA from three biological replicates were pooled before cDNA synthesis. Purified total RNA was used to synthesize cDNAs, which were amplified according to the Illumina RNA-Seq protocol and sequenced using an Illumina HiSeq2000 system, producing 2.6 G 101-bp paired-end reads.

### De novo assembly and annotation

Sequence data with a quality score above 20 ( $Q \geq 20$ ) were extracted using SolexaQA (Cox et al., 2010; Kim et al., 2014). Sequence reads from different tissue samples were *de novo*

assembled using two software tools based on the de Bruijn graph algorithm: Velvet (v1.2.07)(Zerbino and Birney, 2008) and Oases (v0.2.08)(Schulz et al., 2012). A K-mer value of 59 was considered indicative of the optimum length for *de novo* assembly of oriental melon sequence reads. A schematic illustration of the process is shown in Fig. 1.

The quality-checked reads from each tissue were merged and used for transcript assembly. The assembled transcripts were validated by direct comparison with gene sequences in the SEEDERS plant annotation database using BLASTX (e-value  $\leq 1e^{-05}$ ) (Altschul et al., 1990). Protein sequences with the highest similarity were retrieved for further analysis. Short reads of KM transcripts were mapped to the MELONOMICS melon genome (<http://melonomics.net/>) (Garcia-Mas et al., 2012), and CDS regions were defined as in the melon genome.

### Functional enrichment analysis

For gene ontology (GO) term analysis, the assembled loci were annotated to the GO database (downloaded from <http://www.geneontology.org/>) using BLASTP (e-value =  $1e^{-06}$ ). GO term annotation was carried out using GO classification results from the Map2Slim.pl script. Functional enrichment analysis was carried out using DAVID, a web-accessible program that provides a comprehensive set of function annotation tools for inferring biological meaning from large lists of genes (Huang da et al., 2009a; 2009b). Fisher's exact test was used to analyze the gene lists annotated with the TAIR identifications of the transcripts with respect to GO terms, under the following criteria: counts  $\geq 10$ ; false discovery rate (FDR)  $\leq 0.05$ .

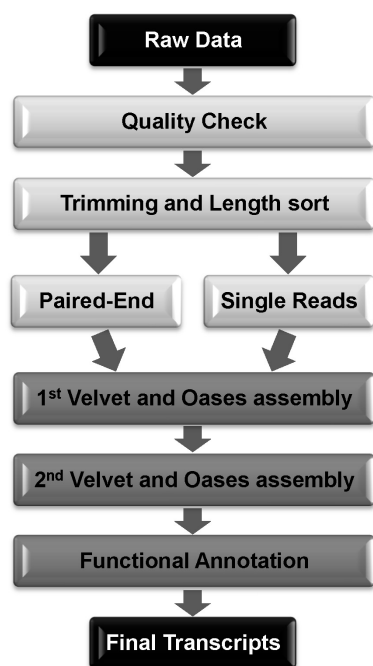
### Short-read counting and tissue-specificity scoring

Illumina sequencing was used to generate mRNA libraries for the various oriental melon tissues examined. Reads for each sequence tag were mapped to the assembled loci using Bowtie (mismatch  $\leq 2$  bp); the number of clean mapped reads for each locus was determined, and the data were normalized using the DESeq library in R. Only transcripts with a tag count  $\geq 50$  were retained for further analysis. Genes differentially expressed between samples were identified based on the fold-change in expression, with the results analyzed by *t*-test. The FDR determined from multiple tests and analyses was applied to calculate the p-value threshold via DESeq. For reliable determination of tissue specificity, Fisher's exact test was used to compare the proportion of a given transcript among all transcripts in each tissue examined. Tissue-specific genes were selected based on read counts of  $> 50$  in the target tissue and  $< 10$  in the other tissues. The degree of homology between the sequences of KM, NW, and melon transcripts was determined using BLASTn/tBLASTn (e-value  $\leq 1e^{-100}$ ; identity  $\geq 90$ ).

### Identification of SNPs

To identify SNPs, a quality check of the KM and NW raw reads was performed using the Solexa QA package. The raw reads were aligned against melon mRNA sequences using TopHat, with modified default parameters (mismatches [-N] = 1; maximum insertion length [-I] = 50; minimum intron length [-i] = 50; maximum intron length [-l] = 14,018; mate inner distance [-r] = 350; segment mismatches = 1; maximum segment intron = 100), and the results were saved as a BAM file for further analysis using SAMTools (Kim et al., 2014; Li et al., 2009).

Using the varFilter command in SAMTools, SNPs were called only for variable positions with a minimum mapping quality (-Q) of 30. The minimum and maximum read depths were set at 3 and 1,000, respectively. Significant SNP sites among the sequences of transcripts from KM, NW, and melon were identified



**Fig. 1.** Schematic illustrating transcriptome assembly and the analysis of oriental melon sequence data. Transcriptome assembly and the transcript sequence data analysis proceeded as a workflow. Sequence data quality analysis, data trimming, and read-length sorting were carried out using the SolexaQA, DynamicTrim, and LengthSort programs, respectively. Transcripts were *de novo* assembled using Velvet and Oases, with the optimum K-mer value set at 59. The assembled transcripts were annotated using BLASTX (e-value  $\leq 1e^{-05}$ ) with the SEEDERS plant annotation database.

**Table 1.** Metrics of oriental melon *de novo* assembly using Velvet and Oases

	KM	NW
Number of assembled transcripts ( <i>K</i> -mer = 57)	64,998	100,234
Minimum length (bp)	200	117
Maximum length (bp)	13,444	11,659
Mean length (bp)	706	739
N50	939	1,138
Number of assembled loci	49,409	51,557

**Table 2.** Functional annotation statistics of assembled oriental melon transcripts

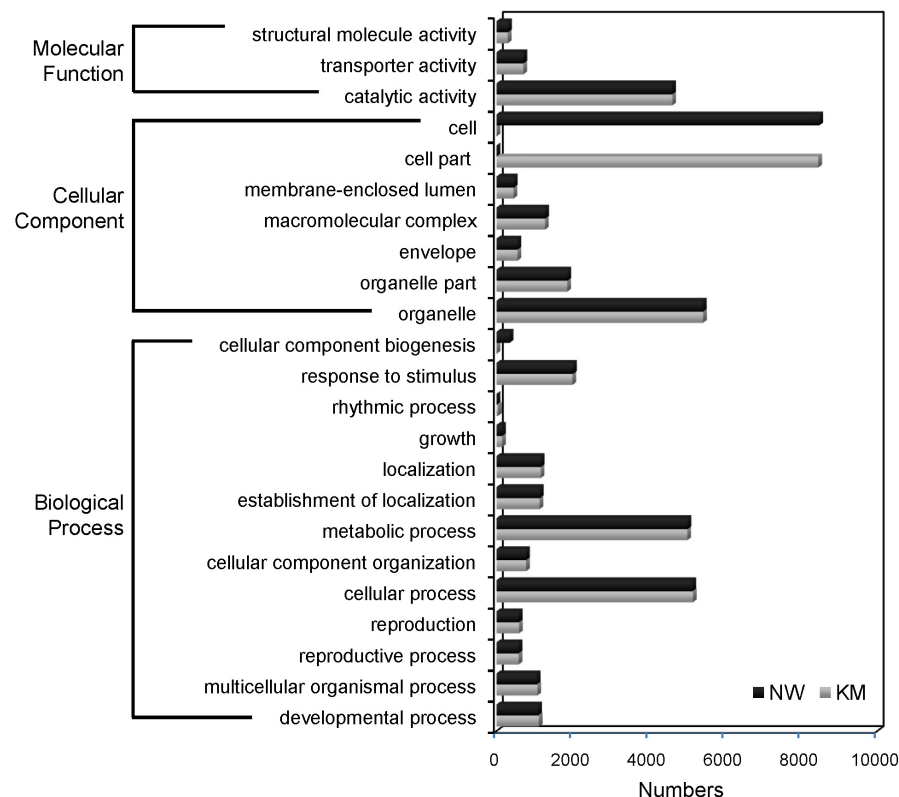
	Number of total transcripts	Number of annotated transcripts ( <i>e</i> -value $\leq 1e^{-05}$ )	Number of unigenes
KM	64,998	36,871	21,363
NW	100,234	64,149	21,914

using a Perl script developed in-house (Supplementary Fig. S2).

#### Identification of SSRs

To identify SSRs, assembled transcripts of NW specimens were formatted according to the SSR Locator's protocol (da Maia et al., 2008). Perfect SSRs (designated 'P-type' SSRs) forming dimer to hexamer motifs with more than five repeat units and located more than 100 bp from other SSRs were selected. Imperfect SSRs (designated 'I-type' SSRs) were selected by allowing for 5 bp of erroneous sequence. Previously

reported criteria (Garg et al., 2011; Kong et al., 2007) were used to select SSRs. Primers were designed using Primer 3 in the SSR Locator. Using the designed primer sets, virtual PCR was performed with SSRs from NW specimens according to the SSR Locator's protocol. Transcripts containing one P-type SSR were selected and used for the development of KM, NW, and melon markers. The following selection criteria were used for the primers: (i) the expected amplicon size should be the same as that of the virtual PCR; and (ii) the primer sets for the different SSRs should not overlap. The selected primer sets



**Fig. 2.** GO classification of the assembled unigenes. Oriental melon unigenes were classified into three functional categories: 'biological process', 'cellular component', and 'molecular function'. Bars indicate the number of genes in each GO term category.

**Table 3.** Number of SNPs among KM, NW and Melon

	KM/NW	KM/Melon
Number of SNPs detected	7,871	3,730
Number of SNPs at CDS	4,752	2,297
Number of Loci with SNPs	2,156	1,063
Number of Transcripts with SNPs	3,110	1,547

were used for virtual PCR analysis of the KM and melon sequence data to distinguish NW-specific marker candidates (Supplementary Fig. S3).

### Genetic linkage map construction

An oriental melon genetic map was constructed using MAPMAKER 3.0/EXP (Lander et al., 1987) with 234 dCAPS (Neff et al., 1998) and 25 SSR markers. F2 population NW and KM specimens were used for mapping. Recombination fractions were converted to map distances in centimorgans (cM) using the Kosambi mapping function (Kosambi, 1943).

## RESULTS AND DISCUSSION

### De novo assembly of the oriental melon transcriptome

To perform transcriptome analysis, RNA-Seq data were generated from five different tissues (female and male flowers, fruit, leaves, and roots) of two oriental melon cultivars (KM and NW). In total, 30.5 Gb (251,752,490 raw reads) and 36.8 Gb (287,233,170 raw reads) of KM and NW sequence data, respectively, were generated using an Illumina HighSeq 2000 (Supplementary Table S1). The quality of the sequence data ( $Q \geq 20$ ) was assessed using SolexaQA, and the reads were trimmed and sorted by length using the DynamicTrim and LengthSort programs, respectively.

Transcripts of each oriental melon cultivar were assembled using Velvet (v1.2.07) and Oases (v0.2.08) ( $k$ -mer = 59), based on de Bruijn graphs. These transcripts were used to construct extended transcripts using Velvet, followed by Oases (Fig. 1). A total of 64,998 KM transcripts were generated, with a mean length of 706 bp and N50 length of 939 bp; the transcript length ranged from 200 to 13,444 bp. For the NW cultivar, 100,234 transcripts were assembled, with a mean length of 739 bp and N50 length of 1,138 bp; the transcript length ranged from 117 to 11,659 bp. The transcripts from each cultivar were then clustered, resulting in 49,409 predicted loci for KM and 51,557 loci for NW (Table 1).

### Functional annotation and classification of oriental melon transcripts

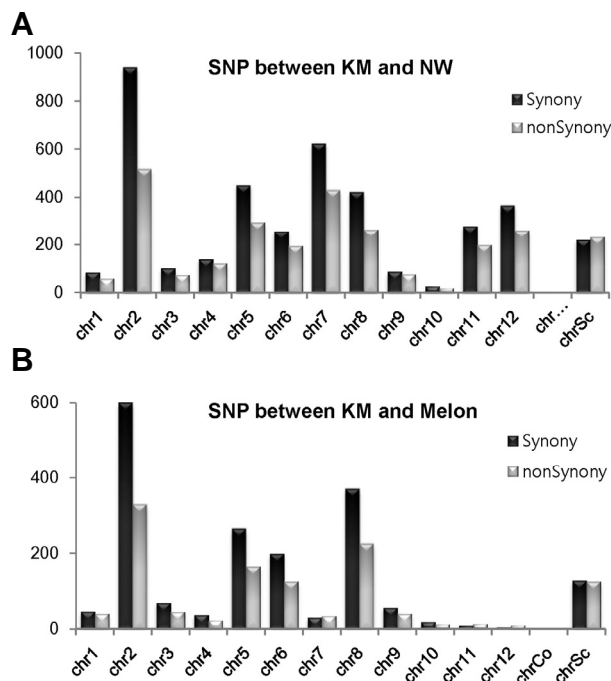
Putative functions of the assembled transcripts were annotated using BLASTP ( $e$ -value  $\leq 1e^{-06}$ ) with the SEEDERS non-redundant protein database. Of the 64,998 KM transcripts, 36,871 were assigned to 21,363 reference proteins, and 64,149 of 100,234 NW transcripts were assigned to 21,914 reference proteins (Table 2). To classify the functions of the assembled oriental melon loci, GO term analysis was performed using TAIR identification information. A total of 42,386 KM and 42,743 NW transcripts were assigned to 23 functional categories: 13 'biological process' categories, 7 'cellular component' categories, and 3 'molecular function' categories (Fig. 2). For both the KM and NW transcripts, 'cellular process', 'cell and cell part', and 'catalytic activity' were the most common terms in the 'biological process,' 'cellular component,' and 'molecular function'

categories, respectively. These GO term data will be used for further studies of the characteristics of oriental melon by functional profiling, prediction of gene function, and functional categorization of genes (Rhee et al., 2008).

### Expression of tissue-specific locus candidates

Constitutive promoters, such as those for ubiquitin and 35S, are used in plant genetic engineering to express genes of interest in a wide range of species (Brisson, 1984; Cornejo et al., 1993). However, overexpression using constitutive promoters may lead to undesirable pleiotropic effects in transgenic plants (Hsieh et al., 2002; Kasuga et al., 1999). The use of tissue-specific promoters with particular developmental expression patterns has been suggested as a strategy to avoid such undesirable pleiotropic effects (Kasuga et al., 2004). Therefore, the development of tissue-specific promoters capable of driving transgene expression is an important area of research in plant genetic engineering (Potenza, 2004).

For unbiased detection of tissue-specific expressed transcripts, statistical analysis of a large number of raw reads was performed. Fisher's exact test was used to compare the proportion of given transcripts among all transcripts in the different tissues. Non-normalized oriental melon cDNA libraries were



**Fig. 3.** Number of SNPs. Synonymous and non-synonymous SNPs were distributed into 12 linkage groups using the published melon genome (*Cucumis melo*. L) as the reference.

prepared from the female and male flowers, fruit, leaves, and roots of each of the oriental melon cultivars. The sequence tags of short reads were mapped to each transcript using Bowtie, and the number of mapped reads for each transcript was then determined. The mapped reads were normalized using the DESeq library in R script. Statistical analyses identified 1,169 and 2,504 tissue-specific transcripts in the KM and NW cultivars, respectively. For the KM cultivar, 144, 107, 231, 256, and 431 transcripts were specific to the female flower, male flower, fruit, leaves, and roots, respectively. In the case of the NW cultivar, 75, 1,121, 368, 150, and 790 transcripts were specific to the female flower, male flower, fruit, leaves, and roots, respectively (Supplementary Tables S2 and S5). The functions of the KM and NW transcripts were predicted by identifying orthologues using melon mRNA sequences. In total, 16,873 of 27,427 melon mRNAs were identified as orthologues of 48,144 KM transcripts and 66,216 NW transcripts. Among these transcripts, 914 were KM-specific, identified based on 746 melon mRNAs, whereas 1,070 NW-specific transcripts were identified based on 573 melon mRNAs (Supplementary Table S3). These tissue-specific candidates will be validated using RT-PCR, and the promoter regions will be investigated for cloning tissue-specific promoters. Furthermore, functional studies of tissue-specific

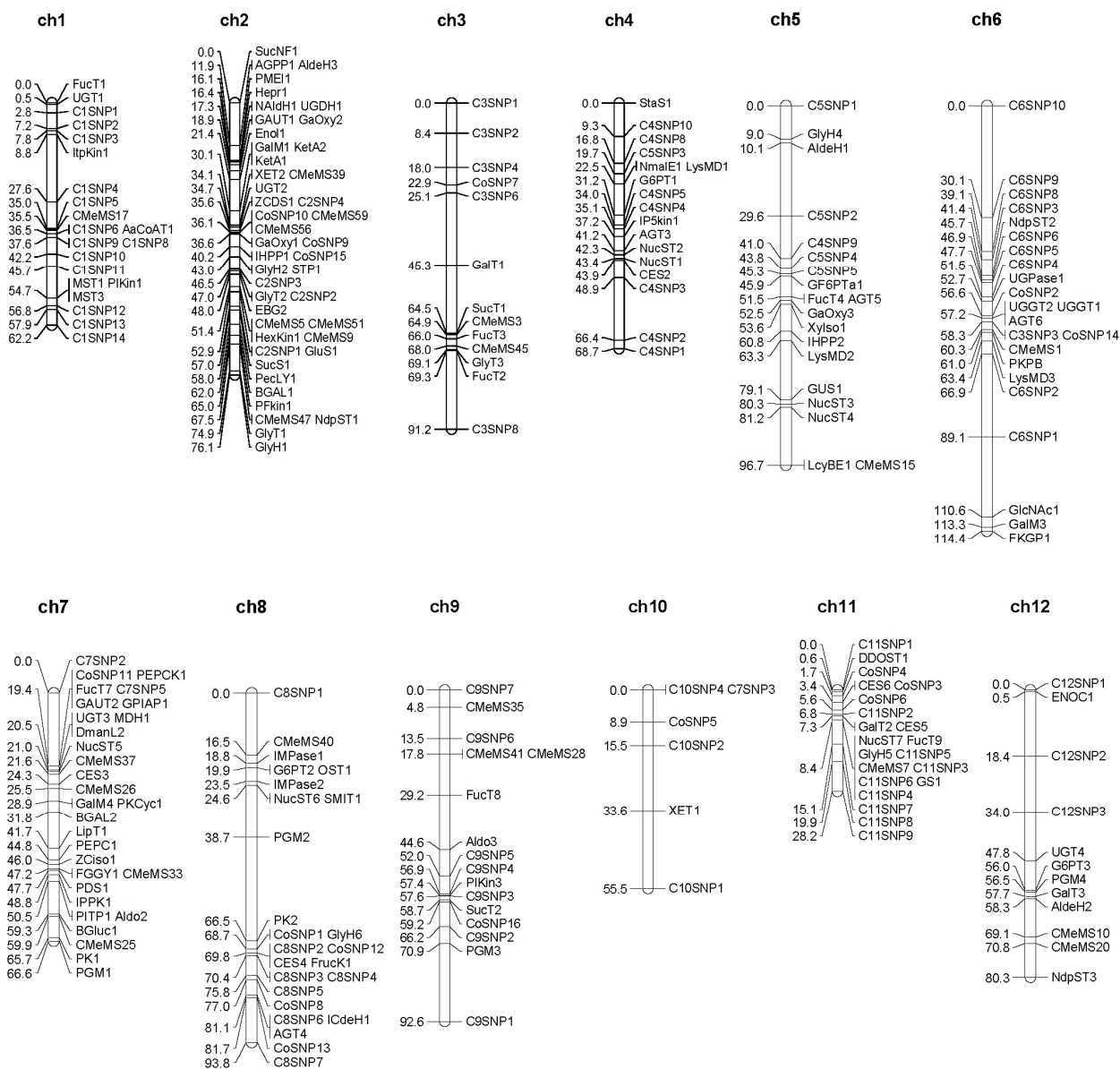
genes will provide additional insights into plant development.

#### Identification of SNPs and SSRs

Molecular markers are important resources for constructing high-density genetic maps such as those used in crop breeding and for the identification of traits of interest. Since NGS technology was developed, many plant genomes have been sequenced, including that of melon (Garcia-Mas et al., 2012). In addition, a large amount of sequence data for melon has been accumulated over the past several years (Gonzalez-Ibeas et al., 2007; 2010; Portnoy et al., 2011; Rodriguez-Moreno et al., 2011). SSRs and SNPs are increasingly used in the construction of melon genetic maps (Blanca et al., 2011; 2012; Diaz et al., 2011; Kong et al., 2011). SNPs and SSRs were identified among the KM, NW, and melon transcripts using the assembled transcripts and melon mRNA sequences. A total of 7,871 SNPs covering 2,156 loci and 3,110 transcripts were identified between the KM and NW cultivars (KM/NW), and 4,752 SNPs were identified in exon regions. Between the KM and melon sequences (KM/melon), 3,730 SNPs were identified covering 1,063 loci and 1,547 transcripts, and 2,297 SNPs were identified in exons (Table 3; Supplementary Table S4).

**Table 4.** Types of SSRs according to motif length in KM and NW transcripts

Number of repeat units	KM					NW				
	Length of motifs					Length of motifs				
	Di-	Tri-	Tetra-	Penta-	Hexa-	Di-	Tri-	Tetra-	Penta-	Hexa-
5	2,807	1,505	180	83	65	3,917	2,533	208	96	108
6	1,137	806	77	30	21	1,529	1,300	119	40	40
7	686	485	27	8	10	874	794	33	16	13
8	515	298	15	4	2	654	454	30	2	2
9	392	197	10	-	1	491	358	11	-	3
10	248	124	3	-	-	342	202	1	3	-
11	152	96	1	1	-	221	169	5	-	-
12	125	59	2	1	-	178	116	4	-	-
13	92	44	1	-	-	103	63	-	-	-
14	62	29	2	-	-	78	52	1	-	-
15	45	17	-	-	-	83	35	-	-	-
16	37	17	-	-	-	40	33	-	-	-
17	35	16	-	-	-	52	33	-	-	-
18	31	25	-	-	-	44	54	-	-	-
19	11	18	-	-	-	25	23	-	-	-
20	15	-	-	-	-	11	-	-	-	-
21	14	-	-	-	-	13	-	-	-	-
22	8	-	-	-	-	18	-	-	-	-
23	2	-	-	-	-	11	-	-	-	-
24	6	-	-	-	-	4	-	-	-	-
25	3	-	-	-	-	7	-	-	-	-
26	2	-	-	-	-	4	-	-	-	-
27	-	-	-	-	-	2	-	-	-	-
28	2	-	-	-	-	5	-	-	-	-
29	2	-	-	-	-	2	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-
Total	6,429	3,736	318	127	99	8,708	6,219	412	157	166



**Fig. 4.** Distribution of genetic markers in the oriental melon genetic map. Linkage groups are numbered at the top, and markers are listed to the right of each linkage group. Map distances are given in cM from the top of each linkage group on the left.

The distribution of synonymous and non-synonymous SNPs among the 12 melon chromosomes was also investigated. The number of synonymous and non-synonymous SNPs in chromosome 2 was significantly larger in KM/NW and KM/melon (Fig. 3). The frequency of SNP occurrence between KM and NW was expected to be low, as the sequences were derived from two near-isogenic lines. However, the number of SNPs between KM/NW was larger than between KM/melon. The melon samples used for genome sequencing were double-haploid line, derived from the cross between PI 161375 (Songwhan Charmi) (SC) and the 'Piel de sapo' (PS) (Oliver et al., 2001). The NW line was bred from the cross between EunCheon type commercial FI variety and Chinese landrace melon. Furthermore, EunCheon is derived from the cross be-

tween Japanese landrace Charmi and small melon. Half of genome sequences in melon reference were Songwhan Charmi (PI 161375) while NW genome sequence were consisted with lots of melon landrace genome. Thus genome variations of KM and NW were higher than those of KM and melon reference. Consequently, fewer SNPs were identified in KM/melon compared with KM/NW. Data regarding the synonymous and non-synonymous SNPs of NW and melon compared with KM are provided in Supplementary Tables S6 and S7. The GO terms for the KM and NW transcripts or KM and melon mRNA sequences with SNPs were sorted, and the rate of synonymous and non-synonymous SNPs was determined (Supplementary Table S8).

dCAPS primers were designed based on 277 SNPs of KM

and NW transcripts and used to screen polymorphic markers. dCAPS primer sets for 245 of the 277 SNPs were moderately amplified in both oriental melon cultivars. The amplified products were digested using restriction enzymes specific to sites in the primer sequences. A total of 234 PCR products exhibited polymorphism between the KM and NW cultivars, and 16 SNP markers are shown in Supplementary Fig. S4.

A general screening of the KM and NW transcript dataset and melon mRNA sequences for unigenes was performed for the presence of di-, tri-, tetra-, penta-, and hexa-SSR motifs. Motifs ranging from dimers to hexamers with more than five repeat units were selected (Garg et al., 2011; Kong et al., 2007). A total of 10,709 SSRs were identified among 9,938 KM transcripts (Supplementary Table S9). The major types of SSRs identified were dinucleotides (6,429), followed by trinucleotides (3,736), tetranucleotides (318), pentanucleotides (127), and hexanucleotides (99) (Table 4). The most frequent SSR motif was GA/TC (1,776), followed by AG/CT (1,574), AT/AT (1,171), TA/TA (1,148), and GAA/TTC (778) (Supplementary Table S10). In NW, 15,662 SSRs were identified among 14,436 transcripts (Supplementary Table S9). The major SSR motifs were dinucleotides (8,708) and trinucleotides (6,219), followed by tetranucleotides (412), hexanucleotides (166), and pentanucleotides (157) (Table 4). The GA/TC motif exhibited the highest frequency (2,731), followed by AG/CT (2,401), GAA/TTC (1,399), and AT/AT (1,310) (Supplementary Table S10).

For NW, primers were designed for SSR marker candidates from the transcripts. Transcripts containing one SSR marker candidate were selected for screening polymorphisms among KM, NW, and melon using virtual PCR (Supplementary Table S11). The presence and size of amplicons resulting from the virtual PCR analyses were compared between NW and KM or NW and melon (Supplementary Table S12) for use in further marker development. Of 8,052 SSR marker candidates, 64 were selected for PCR analysis to determine the presence of polymorphisms between KM and NW. A total of 25 SSR markers exhibited a polymorphic pattern between the two cultivars; the results for 16 markers are shown in Supplementary Fig. S5.

The polymorphisms of the 64 SSR markers and 277 dCAPS markers designed from SNP markers were analyzed in the parent lines, and 25 SSR and 234 SNP markers exhibited polymorphisms. All of these polymorphic markers were screened in 94 F<sub>2</sub> population plants and exhibited co-dominant type Mendelian segregation. The genotypes of these 259 markers were used to construct an oriental melon genetic linkage map consisting of 12 linkage groups covering 926 cM, with an average map distance between markers of 3.7 cM. A total of 248 markers were assigned to the 12 linkage groups, and 11 markers were unlinked (Fig. 4). The chromosome numbers for all of the linkage groups were determined based on alignment of our DNA marker sequences with the melon genome sequence data (Garcia-Mas et al., 2012) regarding the markers are provided in Supplementary Table S13.

The oriental melon genetic linkage constructed based on the 25 SSR and 234 dCAPS SNP markers spanned 926 cM. Although not all of the markers were distributed evenly among the 12 linkage groups, analysis using the  $\chi^2$  goodness-of-fit test showed no significant distortions from the expected Mendelian ratio for any of the markers. The largest linkage group was on Ch6, spanning 114 cM, whereas the smallest was on Ch11, spanning 28.2 cM. Because no markers common to other previously published melon linkage maps were used in our study, it was difficult to estimate the consensus between the genome structures. A previously reported integrated melon genetic map was constructed using 1,592 markers from 8 independent

mapping experiments and spanned 1,150 cM across the 12 linkage groups (Diaz et al., 2011). In this integrated map, the genetic length of the linkage groups ranged from 73 to 119 cM. The significantly shorter genetic length on Ch11 and lower resolution of linkage groups (8 gaps of more than 20 cM) in our map compared with the previously reported integrated map are most likely due to our development of marker types only from transcriptome data or to the use of an insufficient number of markers. It will therefore be necessary to use more markers derived from the genome sequence database and consensus markers positioned based on several mapping experiments to construct a linkage map in which the markers are evenly distributed and cover all of the genome.

*Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).*

## ACKNOWLEDGMENTS

This work was supported by a grant from the Biogreen21 program (PJ11088) of the Rural Development Administration, Korea, and the KRIBB Research Initiative Program. Transcriptome sequences from the female and male flowers, leaves, root, and fruit of Korean melon are available at NABIC under accession NN-0070-000001-NN-0070-000005.

## REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.
- Blanca, J.M., Cañizares, J., Ziarsolo, P., Esteras, C., Mir, G., Nuez, F., Garcia-Mas, J., and Picó, M.B. (2011). Melon transcriptome characterization: simple sequence repeats and single nucleotide polymorphisms discovery for high throughput genotyping across the species. *Plant Genome* 4, 118-131.
- Blanca, J., Esteras, C., Ziarsolo, P., Perez, D., Ferná Ndez-Pedrosa, V., Collado, C., Rodra Guez de Pablos, R., Ballester, A., Roig, C., Canizares, J., et al. (2012). Transcriptome sequencing for SNP discovery across *Cucumis melo*. *BMC Genomics* 13, 280.
- Brisson, N., Paszkowski, J., Penswick, J.R., Gronenborn, B., Potrykus, I., and Hohn, T. (1984). Expression of a bacterial gene in plants by using a viral vector. *Nature* 310, 511-514.
- Cornejo, M.J., Luth, D., Blankenship, K.M., Anderson, O.D., and Blechl, A.E. (1993). Activity of a maize ubiquitin promoter in transgenic rice. *Plant Mol. Biol.* 23, 567-581.
- Cox, M.P., Peterson, D.A., and Biggs, P.J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinf.* 11, 485.
- Diaz, A., Fergany, M., Formisano, G., Ziarsolo, P., Blanca, J., Fei, Z., Staub, J.E., Zalapa, J.E., Cuevas, H.E., Dace, G., et al. (2011). A consensus linkage map for molecular markers and quantitative trait loci associated with economically important traits in melon (*Cucumis melo* L.). *BMC Plant Biol.* 11, 111.
- da Maia, L.C., Palmieri, D.A., de Souza, V.Q., Kopp, M.M., de Carvalho, F.I., and Costa de Oliveira, A. (2008). SSR locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int. J. Plant Genomics* 2008, 412696.
- Gonzalez, V.M., Benjak, A., Henaff, E.M., Mir, G., Casacuberta, J.M., Garcia-Mas, J., and Puigdomenech, P. (2010). Sequencing of 6.7 Mb of the melon genome using a BAC pooling strategy. *BMC Plant Biol.* 10, 246.
- González-Ibeas, D., Blanca, J., Roig, C., Gonzalez-To, M., Pico, B., Triniger, V., Gomez, P., Deleu, W., Cano-Delgado, A., Arus, P., et al. (2007). MELOGEN: an EST database for melon functional genomics. *BMC Genomics* 8, 306.
- Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., Gonzalez, V.M., Henaff, E., Camara, F., Cozzuto, L., Lowy, E., et al. (2012). The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. USA* 109, 11872-11877.
- Garg, R., Patel, R.K., Tyagi, A.K., and Jain, M. (2011). De novo assembly of chickpea transcriptome using short reads for gene

- discovery and marker identification. *DNA Res.* 18, 53-63.
- Hsieh, T.H., Lee, J.T., Chang, Y.Y., and Chan, M.T. (2002). Tomato plants ectopically expressing arabidopsis CBF1 show enhanced resistance to water deficit stress. *Plant Physiol.* 130, 618-626.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1-13.
- Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44-57.
- Kosambi, D.D. (1943). The estimation of map distance from recombination values. *Ann. Eugen.* 12, 172-175.
- Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat. Biotechnol.* 17, 287-291.
- Kasuga, M., Miura, S., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). A combination of the Arabidopsis DREB1A gene and stress-inducible rd29A promoter improved drought- and low-temperature stress tolerance in tobacco by gene transfer. *Plant Cell Physiol.* 45, 346-350.
- Kim, J.E., Oh, S.K., Lee, J.H., Lee, B.M., and Jo, S.H. (2014). Genome-wide SNP calling using next generation sequencing data in tomato. *Mol. Cells* 37, 36-42.
- Kong, Q., Xiang, C., Yu, Z., Zhang, C., Liu, F., Peng, C., and Peng, X. (2007). Mining and characterizing microsatellites in Cucumis melo expressed sequence tags from sequence database. *Mol. Ecol. Notes* 7, 281-283.
- Kong, Q., Xiang, C., Yang, J., and Yu, Z. (2011). Genetic variations of Chinese melon landraces investigated with EST-SSR markers. *Hort. Environ. Biotechnol.* 52, 163-169.
- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newberg, L.A. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1, 174-181.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Liu, L., Kakihara, F., and Kato, M. (2004). Characterization of six varieties of Cucumis melo L. based on morphological and physiological characters, including shelf-life of fruit. *Euphytica* 135, 305-313.
- Mliki, A., Staub, J.E., Zhangyong, S., and Ghorbel, A. (2001). Genetic diversity in melon (*Cucumis melo* L.): an evaluation of African germplasm. *Genet. Resour. Crop Evol.* 48, 587-597.
- Neff, M.M., Neff, J.D., Chory, J., and Pepper, A.E. (1998). dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in Arabidopsis thaliana genetics. *Plant J.* 14, 387-392.
- Oliver, M., Garcia-Mas, J., Cardus, M., Pueyo, N., Lopez-Sese, A.L., Arroyo, M., Gomez-Paniagua, H., Arus, P., and de Vicente, M.C. (2001). Construction of a reference linkage map for melon. *Genome* 44, 836-845.
- Potenza, C., Aleman, L., and Sengupta-Gopalan, C. (2004). Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. In *Vitro Cell. Dev. Biol. Plant* 40, 1-22.
- Portnoy, V., Diber, A., Pollock, S., Karchi, H., Lev, S., Tzuri, G., Harel-Beja, R., Forer, R., Portnoy, V.H., Lewinsohn, E., et al. (2011). Use of non-normalized, non-amplified cDNA for 454-Based RNA sequencing of fleshy melon fruit. *Plant Genome* 4, 36-46.
- Rodriguez-Moreno, L., Gonzalez, V.M., Benjak, A., Marti, M.C., Puigdomenech, P., Aranda, M.A., and Garcia-Mas, J. (2011). Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics* 12, 424.
- Rhee, S.Y., Wood, V., Dolinski, K., and Draghici, S. (2008). Use and misuse of the gene ontology annotations. *Nat. Rev. Genet.* 9, 509-515.
- Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086-1092.
- Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821-829.