An Extended Intervarietal Microsatellite Linkage Map of Cucumber, *Cucumis sativus* L.

Yiqun Weng
USDA-ARS, Vegetable Crops Research Unit and Horticultural Department, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706

Shanna Johnson
Horticultural Department, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706

Jack E. Staub
USDA-ARS, Vegetable Crops Research Unit, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706

Sanwen Huang
Institute of Vegetables and Flowers, Chinese Academy of Agriculture Sciences, Beijing 10081, China

Additional index words. SSR, genetic mapping, little leaf, determinate growth habit

Abstract. A recombinant inbred line (RIL) population derived from two cultivated cucumber (*Cucumis sativus* var. *sativus* L., 2n = 2x = 14) lines, Gy7 (synonym G421) and H-19, was previously used to map yield and fruit quality components. However, the map consisted mainly of dominant markers (i.e., random amplified polymorphic DNAs or amplified fragment length polymorphisms) limiting its use in plant improvement and map-based gene cloning. We report here a moderately saturated genetic map derived from this RIL population that incorporates codominant microsatellite [simple sequence repeat (SSR)] markers and two architectural traits, little leaf (ll) and determinate (de), growth habit. Of 821 cucumber genomic SSR primer pairs evaluated for map construction, 140 (17.0%) were polymorphic between the mapping parents. In combination with 42 previously mapped sequence characterized amplified region (SCAR) and SSR markers, these polymorphic markers were used to construct a linkage map with 46 RILs and 176 mapped loci spanning ~400 cM across seven linkage groups (LG). The numbers of loci mapped on LG 1 through 7 were 11, 6, 35, 18, 46, 45, and 15, respectively. The *ll* locus was flanked by SSR02355 and SSR03940 (4.2 and 3.6 cM from *ll*, respectively), and *de* was flanked by CSWCTT14b and SSR13251 (1.4 and 4.2 cM from the *de*, respectively). The SSR markers linked with the *de* and *ll* genes were mapped to Chromosome 6. No recombination suppression was detected among the mapped loci examined. This Gy7 × H-19 RIL-based genetic map shared 94 marker loci with a previously reported RIL-based linkage map derived from a wide cross between *C. sativus* var. *sativus* line Gy14 and *C. sativus* var. *hardwickii* Alef. R. PI 183967. Comparative mapping supported previous findings that genome differences (likely chromosomal rearrangements) exist between Gy14 and PI 183967.

Cucumber (*Cucumis sativus* var. *sativus* L.) is an important and profitable processed vegetable product in many countries and a popular fresh market culinary component of human diets worldwide. In 2008, cucumber ranked fourth in total acres harvested among 20 major vegetable crops in the United States with a total production of more than one million tons and a farm gate value of $421 million (USDA-NASS, 2009). However, cucumber yield in the United States has not substantially increased in more than two decades (Gusmini and Wehner, 2008). Thus, increasing fruit yield remains a priority in cucumber breeding (Staub and Bacher, 1997; Staub et al., 2008).

Cucumber has a very narrow genetic base and, thus, the level of molecular polymorphism is low (3% to 12%); Dijkhuizen et al., 1996; Horejsi and Staub, 1999; Knerr et al., 1989; Meglic et al., 1996), which in turn makes the construction of saturated maps (mean marker interval less than 2 cM) difficult in this species (Bradeen et al., 2001; Serquen et al., 1997). Moreover, the majority of molecular markers placed on "low-resolution" cucumber maps have been dominant (e.g., random amplified polymorphic DNAs (RAPDs) or amplified fragment length polymorphisms (AFLPs)), which is not the marker type of choice for broad-based applications in marker-assisted selection (MAS) (e.g., Bradeen et al., 2001; Fazio et al., 2003a; Robbins et al., 2008; Serquen et al., 1997; Yuan et al., 2008a, 2008b).

The recent availability of the whole genome sequence in cucumber provides a platform for the development of codominant markers (Huang et al., 2009). A high-resolution simple sequence repeat (SSR)-based linkage map was developed using the recombinant inbred line (RIL) population derived from a wide cross between cultivated cucumber line, Gy14 (*C. sativus* var. *sativus*), and a wild cucumber accession, PI 183967 (*C. sativus* var. *hardwickii* Alef. R.) (Ren et al., 2009). Unfortunately, genetic recombination was extensively suppressed in several linkage groups (LG; chromosomal regions) as a result of possible structural rearrangements that may have occurred during domestication (i.e., selection and introgression of wild and domesticated genotypes). Clustering of markers may hinder mapping of horticulturally important target traits that reside in these recombination suppressed regions.

A narrow-based *C. sativus* var. *sativus* RIL mapping population was previously developed from a cross between two cultivated cucumber lines, Gy7 (synonym G421) and H-19 (Staub et al., 2002). The gynoecious (controlled by *F*), determinate (*de*) line Gy7 possesses standard-sized leaves (~80 cm²) and few (one to three) lateral branches but develops fruit with exceptional processing quality characteristics. The monoecious line H-19 is indeterminate (*De*), possesses little leaves (*ll*, ~40 cm²), and bears relatively poor quality fruit on five to 15 primary lateral branches that support a sequential fruit setting habit.

Several studies using the Gy7 × H-19 RIL population have contributed to the development of cucumber maps that positioned ~200 molecular markers (Bradeen et al., 2001; Fazio et al., 2003a; Serquen et al., 1997), which were subsequently used in MAS (Fan et al., 2006; Fazio et al., 2003b; Robbins et al., 2009). However, most of these maps consisted of dominant markers (i.e., RAPDs or AFLPs), and markers associated with *ll*, *de*, and *F* are not tightly linked (greater than 3 cM) and, thus, may be of limited value in MAS that seeks to change plant architecture. Therefore, a mapping project was designed to: 1) develop a moderately saturated linkage map using the Gy7 × H-19-derived RIL population; 2) identify SSR markers associated with *ll* and *de*; and 3) examine the syntenic relationship between this narrow-based map and a wide-based map (Gy14 × PI 183967) to determine the degree of genetic recombination suppression or segregation distortion in the narrow-based population.

Materials and Methods

Phenotypic evaluation. Seeds of 143 F₀ Gy7 × H-19 recombinant inbred lines and...
two parental lines were planted on 5 June 2009 in an open-field [Plainfield loamy sand (Typic Udipsammoil) soil] at the University of Wisconsin Agricultural Experiment Station in Hancock, WI. Ten non-replicated plants of each RIL were phenotyped for two morphological traits [little leaf (ll) versus standard-sized (Ll) leaf; determinate (de) versus indeterminate (De)] 3 weeks after planting. Plants possessing main stems that ended in a flower cluster were classified as determinant. Likewise, plants were classified as little or standard leaf type when compared with parental lines by visual examination. Plant evaluations for leaf size and growth habit were conducted three times during the growing season (21, 42, and 70 d after sowing).

Molecular marker analysis. Unexpanded young leaves of each RIL were collected from each of the 10 plants and pooled, and genomic DNA was extracted using the CTAB method (Murray and Thompson, 1980). A subset of 46 RILs was used for linkage analysis, in which 26 and 20 possessed standard-sized and little leaves and 21 and 25 were determinant and indeterminate, respectively.

Each polymerase chain reaction (PCR) contained 25 ng template DNA, 0.5 μM each of two primers, and 1 x PCR master mix (Fermentas, Glen Burnie, MD) in a total volume of 10.0 μL. A “touch-down” PCR program was used for all primer sets (Weng et al., 2005).

Two sets of molecular markers were used in this study. The first set included 821 SSR markers developed from a whole genome sequence of the Northern China fresh market-type cucumber inbred line 9930 (Huang et al., 2009). The second set included 15 SSR, 26 SCAR, and one singe nucleotide polymorphism markers previously mapped in the Gy7 × H-19 RIL population (Fazio et al., 2003a; Robbins and Staub, 2009). Each primer pair amplified a single band in each sample template DNA, in which variation in PCR product size allowed for banding morphotypes to be resolved in 3% agarose gels stained with ethidium bromide.

Results and Discussion

Segregation of leaf size and growth habit among recombinant inbred lines. Of the 143 lines examined, 42 were homozygous little leaf (ll), 100 homozygous standard-sized leaf (Ll), and one was heterozygous (Ll). Growth habit analysis of these RILs resulted in their classification as 57 homozygous determinate (de) and 86 homozygous indeterminate (De,De). The expression of each of these traits is controlled by single recessive genes in this mapping population (Bradeen et al., 2001; Fazio et al., 2003a; Serquin et al., 1997). However, χ² analysis revealed that although segregation at the little leaf locus deviated significantly from the expected 1:1 ratio (P = 0.001), segregation at the determinate locus marginally fit this ratio (P = 0.015). The little leaf (ll) locus has pleiotropic associations with multiple lateral branching (Fazio et al., 2003a) and is located in a chromosome region that manifests relatively strong segregation distortion that supports the dominant action of Gy7 alleles (see subsequently). The action of the little leaf allele does not affect gynoecy conditioned by the F locus, a trait that also maps to the same LG as ll. It is sometimes difficult to discriminate between leaves that have an intermediate (50 to 70 cm²) and standard (>70 cm²) leaf size (Staub et al., 2002). It is likely that the greater-than-expected standard leaf-sized plants identified here was the result of either segregation distortion or misclassification of leaf-type phenotypes.

Table 1. Distribution of genetic markers among seven cucumber (Cucumis sativus L.) chromosomes mapped with a recombinant inbred line population derived from a Gy7 × H-19 cross.¹

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>No. loci (%)</th>
<th>Map length [cM (%)]</th>
<th>Marker interval [cM]²</th>
<th>Physical length [μm (%)]³</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11 (6.3)</td>
<td>25.6 (6.4)</td>
<td>2.3</td>
<td>107.4 (14.9)</td>
</tr>
<tr>
<td>2</td>
<td>6 (3.4)</td>
<td>16.6 (4.1)</td>
<td>2.8</td>
<td>102.9 (14.3)</td>
</tr>
<tr>
<td>3</td>
<td>35 (19.9)</td>
<td>82.9 (20.7)</td>
<td>2.4</td>
<td>129.1 (17.9)</td>
</tr>
<tr>
<td>4</td>
<td>18 (10.2)</td>
<td>62.9 (15.7)</td>
<td>3.5</td>
<td>102.1 (14.2)</td>
</tr>
<tr>
<td>5</td>
<td>46 (26.1)</td>
<td>78.7 (19.7)</td>
<td>1.7</td>
<td>95.0 (13.2)</td>
</tr>
<tr>
<td>6</td>
<td>45 (25.6)</td>
<td>95.6 (23.9)</td>
<td>2.1</td>
<td>110.0 (15.3)</td>
</tr>
<tr>
<td>7</td>
<td>15 (8.5)</td>
<td>37.9 (9.5)</td>
<td>2.5</td>
<td>73.2 (10.2)</td>
</tr>
<tr>
<td>Sum</td>
<td>176 (100.0)</td>
<td>400.2 (100.0)</td>
<td>2.5</td>
<td>719.7 (100.0)</td>
</tr>
</tbody>
</table>

¹Numbers in parentheses in each column are percentages of the number of markers in each chromosome based on the sum of seven chromosomes.
²Average genetic distances between adjacent markers.
³Pachytenic chromosome length according to Koo et al. (2005).
alleles were from the exotic paternal PI 183967 parent.

The linkage map presented here spans 400.3 cM with a mean marker distance of 2.5 cM (Table 1). The genetic distance of this map is considerably less than the expected 750 to 1000 cM length of the cucumber genome as estimated from cytological observations (Ramachandran and Seshadri, 1986; Staub and Meglic, 1993). Given the relatively unsaturated condition of historic dominant marker-based cucumber maps, addition of codominant markers is desirable for MAS. Of particular consequence in this regard is LG 1, 2, 4, and 7, in which only 11, 6, 18, and 15 SSR markers were mapped here, respectively, in which each possessed varying estimated physical map length intervals (Table 1). There is, generally, a positive correspondence between genetic (LG) and physical (putative chromosomal) lengths in cucumber (Ren et al., 2009). Chromosomes 3 (LG 3) and 6 (LG 6) are the longest among the seven (Koo et al., 2005; Table 1), and these are, in fact, the LG having the longest map length in this study (i.e., LG 3 = 82.9 and LG 6 = 95.6 cM, respectively; Table 1). In contrast, such chromosome–LG associations cannot be extrapolated for the shortest chromosomes (5 and 7), although the marker concentrations on some LG (e.g., LG 5 having 46 markers) were relatively high. More accurate map length estimation must, therefore, await greater map saturation and intensive chromosome structure analysis (i.e., fluorescent in situ hybridization).

Marker-assisted selection can enhance conventional plant improvement efforts by increasing its efficiency (Collard and Mackill, 2007). The efficiency of MAS is dependent on many factors, including the strength of marker/trait associations (i.e., linkage) and a knowledge of epistasis, correlated traits, and genotypic × environmental interactions (Wang et al., 2007). The strategic use of $II$ and $de$ trait/marker associations in MAS could increase the efficiency of selection to alter plant architecture for once-over harvest operations (Staub et al., 2008). The genes for little leaf ($ll$) and determinate ($de$) growth habit were both mapped onto LG 6 (Fig. 1), where the $II$ was flanked by SSR02355 and SSR03940 at 4.2 and 3.6 cM, respectively, and $XSCWT714b$ and $XSSR13251$ flanked $de$ at 1.4 and 4.2 cM, respectively. When compared with historic maps (Bradeen et al., 2001; Fazio et al., 2003a), the linkage map constructed here defines robust trait/marker relationships that could be exploited in MAS as has been the case with quantitative trait loci (Fan et al., 2006; Fazio et al., 2003b; Robbins and Staub, 2009). Moreover, because the whole genome sequence of cucumber (through Line 9930) is available (Huang et al., 2009), fine mapping of these genes should be feasible. For instance, intensive analysis of $II$ map position should be possible given the known whole genome scaffolds (sca) of flanking markers SSR02355 [resides in sca000028 ($\approx 1.91$ Mbp)] and SSR03849 [resides in sca000024 ($\approx 2.22$ Mbp)] (Huang et al., 2009). As fine mapping is performed, putative candidate gene(s) may be identified and further evaluated.

Comparison of narrow and wide cross-based cucumber genetic maps. Ren et al. (2009) used 995 SSR markers to saturate the Gy14 × PI 183967 wide-based map constructed by Kennard et al. (1994, a 70-point map of 480 cM in 10 LGs) using an F$_2$ population derived from the same initial parental cross. The map of Ren et al. (2009) spanned only 635.4 cM as a result of strong recombination suppression (LG 4, 5, and 7) and segregation distortion (LG 1, 4, and 6) in several genomic regions.

Ninety-four common marker loci were shared between the map constructed by Ren et al. (2009) and the one constructed here to facilitate synteny analysis. These maps shared seven, five, 18, 10, 24, 21, and nine markers in LG 1 to 7, respectively. Although the marker orders in LG 1, 2, 4, and 7 were not compared because of limited common markers, marker order concordance among 21 common markers in LG 6 indicated map colinearity (Fig. 1). Likewise, 16 of 18 common marker loci in LG 3 were collinear (SSR15419 and SSR15519 were not collinear). A comparative analysis of LG 5 is shown in Figure 2. In LG 5 of Ren et al. (2009), 99 markers clustered in two regions (63 at 6.0 cM

Fig. 1. A cucumber ($Cucumis$ sativus L.) linkage map derived from a Gy7 × H-19 recombinant inbred line (RIL). Numbers on top of the map are linkage groups (LG) (1 through 7), which are also corresponding to chromosomes numbers as defined in Ren et al. (2009). Numbers in parentheses after LG numbers correspond to linkage groups on the map by Fazio et al. (2003a). Cumulative map distance (cM) is shown to the left of each linkage group and marker designation is on the right. SCAR markers are designated by –S or –M, L19-SNP-H19 is a single nucleotide polymorphism (SNP) marker, and the remaining markers are simple sequence repeat (SSR) markers. Two morphological genes are determinate ($de$) and little leaf ($ll$). Markers with asterisks showed segregation distortion.
and 36 at 7.9 cM). Among 11 markers defined in these two cluster regions, 10 were distributed across the Gy7 × H-19 RIL map (Fig. 2). For example, although two marker loci, X SSR14247 and X SSR15321, mapped to the same cluster in LG 5 (at 6.0 cM) in the wide-based map (i.e., Gy14 × PI 183967), their map interval was 50.8 cM in the narrow-based Gy7 × H-19 map, suggesting that recombination suppression did not occur in this region of the latter map. Moreover, the marker map locations across the two clusters in LG 5 of the wide-based map were distributed across the length of this LG in the narrow-based map (Fig. 2). The disparity between these maps may be caused by possible chromosomal structural changes (e.g., translocations, inversions) that have occurred in cultivated and wild cucumbers (Huang et al., 2009; Ren et al., 2009), although genotyping errors could not be excluded.

The position of marker SSR05819 was not concordant between the two maps examined (Fig. 2). Marker SSR05819 is located in the line 9930-derived cucumber whole genome scaffold designated sca000104. Several markers from this scaffold mapped to the 6.0 cM cluster in LG 5 in the wide-based cross map (Ren et al., 2009). Thus, it is likely that the location of this marker on the wide-based map was the result of an error during marker-map assignment. In addition to SSR05819, the order of four adjacent markers (SSR03529, SSR07100, SSR00772, and SSR17464) at the distal end of LG 5 of the Gy7 × H-19 map was not syntenic with that of the Gy14 × PI 183967 map. This marker order disparity may be indicative of chromosomal rearrangements that have occurred during domestication (i.e., cultivated versus feral cucumber comparisons). However, because these four and other two markers [SSR01498 (6.5 cM location) and SSR00170 (69.2 cM location)] in LG 5 (Fig. 2) belong to the same scaffold (sca000003) of the line 9930 whole genome assembly (Huang et al., 2009), it is also possible that the map position of these markers is incorrect in the narrow-based cross. Thus, additional fine map analysis will be needed to elucidate this apparent disparity.

The comparative mapping conducted here confirmed marker order differences between C. sativus var. sativus and var. hardinckii identified by Ren et al. (2009) and lent support to the hypothesis of chromosomal structural changes in cucumber. This study also identified the determinate (de) and little leaf (l) genes to be located in the long and short arms of chromosome 6, respectively, as tagged by flanking SSR markers.

The 174-point map developed here improves the marker saturation of the reference Gy7 × H-19 map developed by Fazio et al. (2003a) by 24.7%. It adds 140 codominant SSR markers to this narrow-based map. Moreover, unlike the case in the wide-based genetic map of Ren et al. (2009; Gy14 × PI 183967), no suppression of genetic recombination or significant segregation distortion was detected in the Gy7 × H-19 genetic map.

It has been long thought that determinate, gynoecious, multiple lateral branching U.S. processing cucumber phenotypes may improve yield potential in once-over mechanical harvest operations (Staub et al., 2008). Gain from selection for some yield and quality components have not been fully realized as a result of a lack of map saturation (weak trait/marker linkages) and a meager understanding of epistatic interactions (Robbins et al., 2008). For instance, the little leaf trait is pleiotrophic to multiple lateral branching, but also confers poor fruit quality. It may be possible, using previously identified marker/trait associations (Fazio et al., 2003b) and those identified here for de and l, to apply MAS in the seedling stage to recover gynoecious, determinate, little leaf types that possess multiple lateral branching and then evaluate fruit quality characteristics in the open field and by salt stress evaluation, thus increasing selection efficiency and effectiveness.

**Literature Cited**


