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

Yiqun Weng²

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. Staub¹

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 $\times$ 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 genomic 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 gynodioecious (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 $\times$ 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 and Staub, 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 $\times$ 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 $\times$ 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 F₉ Gy7 $\times$ H-19 recombinant inbred lines and

Received for publication 4 Mar. 2010. Accepted for publication 20 Apr. 2010.

We thank Linda Crubaugh for technical assistance and Dr. Michael J. Havey for valuable comments and suggestions during this study.

¹Current address: USDA-ARS Forage and Range Research Laboratory, 690 N. 1100 East, Logan, UT 84322.

²To whom reprint requests should be addressed; e-mail yiqun.weng@ars.usda.gov or weng4@wisc.edu.

two parental lines were planted on 5 June 2009 in an open-field [Plainfield loamy sand (Typic Udipasamment) 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 determinate. 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 determinate and indeterminate, respectively.

Each polymerase chain reaction (PCR) contained 25 ng template DNA, 0.5 μ M each of two primers, and 1 $\times$ PCR master mix (Fermentas, Glen Burnie, MD) in a total volume of 10.0 μ L. 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 single nucleotide polymorphism markers previously mapped in the Gy7 $\times$ H-19 RIL population (Fazio et al., 2003a, 2003b; 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.

Data analysis and map construction. An RIL-based linkage map was constructed using the computer program JOINMAP 3.0 based on SSR and morphological marker segregation data that were analyzed for fit to the expected 1:1 ratios by χ^2 analyses. Marker data were assigned to LGs using a minimum logarithm of odds (LOD) score of 3.0, and the Kosambi map function (Kosambi, 1944) was used to calculate the genetic distance between markers.

Markers were assigned to LG according to shared molecular markers with the high-resolution cucumber genetic map of Ren et al. (2009). For synteny analysis between linkage maps, the order of 94 common markers on the map of Ren et al. (2009) and the Gy7 $\times$ H-19 RIL-based map was compared.

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 (*dede*) and 86 homozygous indeterminate (*DeDe*), respectively. 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; Serquen et al., 1997). However, χ^2 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 gynoecey 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 ($\approx$ 80 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.

Slight distortions in segregation, however, usually do not affect trait-marker association during linkage analysis (Hackett and Broadfoot, 2003). Nevertheless, pleiotropy and epistatic interactions have been observed in cucumber and, thus, care must be taken during phenotypic classification for segregation analyses (Robbins et al., 2008).

Polymorphism screening and linkage analysis. Of the 821 cucumber genomic SSR primer sets tested, 17.0% (143 of 821) were polymorphic between Gy7 and H-19. This degree of polymorphism was higher than that reported in previous DNA-based studies (3% to 12% polymorphism) examining diversity among elite (*C. sativus* var. *sativus*) and exotic (*C. sativus* var. *hardwickii*) cucumbers (Dijkhuizen et al., 1996; Staub and Bacher, 1997). This disparity can be explained by the fact that when compared with other elite commercial cucumber germplasm (lines and hybrids), Gy7 and H-19 are genetically distinct (Staub and Bacher, 1997). Moreover,

the types of markers used here differ from those used by Dijkhuizen et al. (1996; used RFLP and isoenzyme markers) and Horejsi and Staub (1999; used RAPD markers). Moreover, the microsatellite markers used here were developed from a whole genome sequence of cucumber inbred line 9930 through pre-selection of long repeat motifs (Ren et al., 2009), which are generally more polymorphic than those having shorter motifs (Huang et al., 2009).

Forty-two markers (26 SCARs, 15 SSRs, and one SNP; Fazio et al., 2003a; Robbins and Staub, 2009) previously mapped in cucumber were also used for RIL analysis. Each primer pair amplified a single band, and, thus, 184 marker or gene loci were included in linkage mapping. One hundred seventy-six markers were mapped to seven LG using a LOD threshold of 3.0. Of these loci, 94 were shared with the cucumber genetic map constructed by Ren et al. (2009). Based on common-marker alignments between these maps, markers could be mapped to the seven LG (2n = 2x = 14), which corresponded to cucumber Chromosomes 1 to 7, respectively, as defined in Ren et al. (2009) and Table 1 (distribution information) and depicted in Figure 1.

Of the 176 molecular marker loci examined by χ^2 test analysis, significant derivation from the expected 1:1 segregation (i.e., segregation distortion) was detected for 11 marker loci. Of these 11 markers, eight were concentrated in two chromosomal regions [segregation distortion region (SDR)], which included four (*XSSR01609*, *XII-S*, *XN6-A-S*, and *XNR66*) in LG 3 (from 11.9 to 18.9 cM) and four (*XSSR03940*, *XSSR16451*, *XSSR17624*, and *XSSR13884*) in the middle of LG 6 adjacent to the *ll* locus (from 35.5 to 40.4 cM; Fig. 1). Three other loci showing segregation distortion (*XSSR17023*, *XSSR14934*, and *XSSR06653*) were scattered in different locations of LG 6 (Fig. 1). Although the SDR presented in LG 3 exhibited excess numbers of marker genotypes associated with H-19 (paternal parent), Gy7 maternal parent alleles were more frequently associated with *ll* and other four adjacent SSR loci. The SDR region adjacent the *ll* locus was also identified by Ren et al. (2009) using a Gy14 $\times$ PI 183967 mapping population. However, the preferentially transmitted

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

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

^xNumbers in parentheses in each column are percentages of the number of markers in each chromosome based on the sum of seven chromosomes.

^yAverage genetic distances between adjacent markers.

^zPachytene chromosome length according to Koo et al. (2005).

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 are simple sequence repeat (SSR) markers. Two morphological genes are determinate (*de*) and little leaf (*ll*). Markers with asterisks showed segregation distortion.

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 struc-

ture 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 genotype × environment interactions (Wang et al., 2007). The strategic use of *ll* 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 *ll* was flanked by SSR02355 and SSR03940 at 4.2 and 3.6 cM, respectively, and *XCSWCTT14b* and *XSSRI3251* 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 *ll* map position should be possible given the known whole genome scaffolds

(sca) of flanking markers SSR02355 [resides in sca000028 (≈1.91 Mbp)] and SSR03849 [resides in sca000024 (≈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_{2:3} 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

- Koo, D.H., H.W. Choi, J. Cho, Y. Hur, and J.W. Bang. 2005. A high-resolution karyotype of cucumber (*Cucumis sativus* L. 'Winter Long') revealed by C-banding, pachytene analysis, and RAPD-aided fluorescence in situ hybridization. *Genome* 48:534–540.
- Kosambi, D.D. 1944. The estimation of map distance from recombination values. *Ann. Eugen.* 12:172–175.
- Meglic, V., F. Serquen, and J.E. Staub. 1996. Genetic diversity in cucumber (*Cucumis sativus* L.): I. A reevaluation of the U.S. germplasm collection. *Genet. Resources Crop Evol.* 43: 533–546.
- Murray, M.G. and W.F. Thompson. 1980. Rapid isolation of high molecular weight DNA. *Nucleic Acids Res.* 8:4321–4325.
- Ramachandran, C. and V.S. Seshadri. 1986. Cytological analysis of the genome of cucumber (*Cucumis sativus* L.) and muskmelon (*Cucumis melo* L.). *Z. Pflanzenzuchtung* 96: 25–38.
- Ren, Y., Z.H. Zhang, J.H. Liu, J.E. Staub, Y.H. Han, Z.C. Cheng, X.F. Li, J.Y. Lu, H. Miao, H.X. Kang, B.Y. Bie, X.F. Gu, X.W. Wang, Y.C. Du, W.W. Jin, and S.W. Huang. 2009. An integrated genetic and cytogenetic map of the cucumber genome. *PLoS One* 4:e5795.
- Robbins, M.D., M. Casler, and J.E. Staub. 2008. Pyramiding QTL for multiple lateral branching in cucumber using nearly isogenic lines. *Mol. Breed.* 22:131–139.
- Robbins, M.D. and J.E. Staub. 2009. Comparative analysis of marker-assisted and phenotypic selection for yield components in cucumber. *Theor. Appl. Genet.* 119:621–634.
- Serquen, F.C., J. Bacher, and J.E. Staub. 1997. Mapping and QTL analysis of horticultural traits in a narrow cross in cucumber using random-amplified polymorphic DNA markers. *Mol. Breed.* 3:257–268.
- Staub, J.E. and J. Bacher. 1997. Cucumber as a processed vegetable, p. 129–193. In: Smith, D.S., J.N. Cash, W.K. Nip and Y.H. Hui (eds.). *Processing vegetables: Science and technology IV*. Technomic Publishing Co., Inc., Lancaster, PA.
- Staub, J.E., L.K. Crubagha, and G. Fazio. 2002. Cucumber recombinant inbred lines. *Cucurbit Genet. Coop. Rpt.* 25:1–2.
- Staub, J.E. and V. Meglic. 1993. Molecular genetic markers and their legal relevance for cultigen discrimination: A case study in cucumber. *HortTechnology* 3:291–300.
- Staub, J.E., M.D. Robbins, and T.C. Wehner. 2008. Cucumber, p. 241–282. In: Prohens, J. and F. Nuez (eds.). *Vegetables I: Asteraceae, brassicaceae, chenopodiaceae, and cucurbitaceae*. Springer, New York, NY.
- USDA-NASS. 2009. National statistics: Cucumbers. 1 Mar. 2010. <<http://www.nass.usda.gov/>>.
- Wang, Y.H., T. Joobeur, R.A. Dean, and J.E. Staub. 2007. Cucurbits, p. 315–324. In: Kole, C. (ed.). *Genome mapping and molecular breeding in plants, vegetables*. Vol. 5. Springer, New York, NY.
- Weng, Y., W. Li, R.N. Devkota, and J.C. Rudd. 2005. Microsatellite markers associated with two *Aegilops tauschii*-derived greenbug resistance loci in wheat. *Theor. Appl. Genet.* 110: 462–469.
- Yuan, X.J., X.Z. Li, J.S. Pan, G. Wang, S. Jiang, X.H. Li, S.L. Deng, H.L. He, M.X. Si, L. Lai, A.Z. Wu, L.H. Zhu, and R. Cai. 2008a. Genetic linkage map construction and location of QTLs for fruit-related traits in cucumber. *Plant Breed.* 127:180–188.
- Yuan, X.J., J.S. Pan, R. Cai, Y. Guan, L.Z. Liu, W.W. Zhang, Z. Li, H.L. He, C. Zhang, L.T. Si, and L.H. Zhu. 2008b. Genetic mapping and QTL analysis of fruit and flower related traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. *Euphytica* 164: 473–491.