

Fine genetic mapping of *cp*: a recessive gene for compact (dwarf) plant architecture in cucumber, *Cucumis sativus* L.

Yuhong Li · Luming Yang · Mamta Pathak ·
Dawei Li · Xiaoming He · Yiqun Weng

Received: 10 April 2011 / Accepted: 13 June 2011 / Published online: 7 July 2011
© Springer-Verlag (outside the USA) 2011

Abstract The compact (dwarf) plant architecture is an important trait in cucumber (*Cucumis sativus* L.) breeding that has the potential to be used in once-over mechanical harvest of cucumber production. Compact growth habit is controlled by a simply inherited recessive gene *cp*. With 150 F_{2:3} families derived from two inbred cucumber lines, PI 308915 (compact vining) and PI 249561 (regular vining), we conducted genome-wide molecular mapping with microsatellite (simple sequence repeat, SSR) markers. A framework genetic map was constructed consisting of 187 SSR loci in seven linkage groups (chromosomes) covering 527.5 cM. Linkage analysis placed *cp* at the distal half of

the long arm of cucumber Chromosome 4. Molecular markers cosegregating with the *cp* locus were identified through whole genome scaffold-based chromosome walking. Fine genetic mapping with 1,269 F₂ plants delimited the *cp* locus to a 220 kb genomic DNA region. Annotation and function prediction of genes in this region identified a homolog of the cytokinin oxidase (CKX) gene, which may be a potential candidate of *compact* gene. Alignment of the CKX gene homologs from both parental lines revealed a 3-bp deletion in the first exon of PI 308915, which can serve as a marker for marker-assisted selection of the compact phenotype. This work also provides a solid foundation for map-based cloning of the compact gene and understanding the molecular mechanisms of the dwarfing in cucumber.

Communicated by Y. Xue.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1640-6) contains supplementary material, which is available to authorized users.

Y. Li · D. Li
Horticulture College, Northwest Agriculture & Forestry
University, Yangling 712100, China

Y. Li · L. Yang · M. Pathak · D. Li · X. He · Y. Weng (✉)
Horticulture Department, University of Wisconsin,
Madison, WI 53706, USA
e-mail: weng4@wisc.edu; yiqun.weng@ars.usda.gov

M. Pathak
Department of Vegetable Crops,
Punjab Agricultural University, Ludhiana 141 004, India

X. He
Vegetable Research Institute, Guangdong Academy
of Agricultural Sciences, Guangzhou 510640, China

Y. Weng
USDA-ARS, Vegetable Crops Research Unit,
1575 Linden Drive, Madison, WI 53706, USA

Introduction

Cucumber, *Cucumis sativus* L. ($2n = 2x = 14$) is an important specialty vegetable crop worldwide. In the United States, cucumber yield has been stagnant since the 1980s (Gusmini and Wehner 2008). With mixed results obtained when selecting directly for yield, it was proposed that the most effective approach to breeding for yield may be selecting for yield components with a higher heritability (Wehner 1989; Cramer and Wehner 2000). Such traits may include the number of harvests per plant, stem length, number of branches per plant, number of flowering nodes per branch, time to anthesis, percentage of pistillate flowers, and percentage of fruit set (Cramer and Wehner 2000). These traits can be manipulated to create various genotypes that possess an array of architectural habits.

One such trait is the compact (dwarf) plant growth habit, which in cucumber generally refers to mutants with reduced vine length (or plant height). Several types of

compact cucumber mutants have been described. The first was identified in two plant introduction lines PI 308915 and PI 308916, which exhibited significantly reduced internode length (super dwarf) and was due to a recessive gene, *cp* (Kauffman and Lower 1976). Later, from EMS-treated cucumber plants, Kubicki et al. (1986) identified a second compact mutant *compact-2* (*cp-2*) with shortened internodes similar to PI 308916. The allelism between *cp* and *cp-2* is not known; but *cp-2* is required to interact with the ‘bushy’ gene to produce the dwarf phenotype (Kubicki et al. 1986). The EMS-induced mutant of ‘super compact’ cucumber was controlled by a recessive gene, *scp* (Niemirowicz-Szczytt et al. 1996) with drastically reduced main stem length and no lateral branches, which may be identical to the rosette (*ro*) mutant described by de Ruiter et al. (1980). Crien et al. (2009) discovered yet another compact cucumber, which is different from either *cp* or *cp-2* in their genome locations. One unique aspect of this mutant is that this compact gene expresses a compact phenotype when homozygous but an intermediate compact phenotype when heterozygous.

The compact vining growth habit in PI 308916 was suggested as an alternative to the standard vining phenotype for use in high density plantings for once-over machine harvested pickling cucumber production (Kauffman and Lower 1976); while the new compact cucumber by Crien et al. (2009) has been proposed to use in high-wire cultivation of European glasshouse cucumbers. Another benefit for the compact plant phenotype is the reduction of disease incidences. Fruit rot caused by *Phytophthora capsici* is an increasingly serious disease affecting pickling cucumber production in the United States, but no resistant cultivars are available. Ando and Grumet (2006) found that PI 308916, which has a tendency to hold young fruits off the ground, exhibited lower disease occurrence. The reduced disease occurrence has not been attributed to genetic resistance, suggesting that architecture which allows less contact of fruit with the soil could be useful for *P. capsici* control in pickling cucumber.

Marker-assisted selection (MAS) is an important tool in supplementing classical plant breeding to increase selection efficiency and expedite delivery of new inbred lines or cultivars. MAS is especially advantageous for selection of recessively inherited genes like *cp*, which may eliminate the need for self-pollination thus speeding the development of target genotypes. In addition, although breeding lines incorporating the compact gene have shown significant yield advantages, there is a negative association between the PI 308916-derived compact cucumbers with poor seedling emergence (Edwards and Lower 1981). Cloning of the compact gene may help to understand its structure and function, which may in turn help to address the issues of poor seedling establishment associated with the compact phenotype.

With recent advances in technology and instrumentation for sequencing plant genomes, genetic and genomic resources in cucumber are being accumulated rapidly. The whole genomes of three cucumber lines have been sequenced including the Northern China fresh market type ‘9930’ (Huang et al. 2009), the North American pickling type ‘Gy14’ (Cavagnaro et al. 2010), and a European inbred line ‘B10’ (<http://csgenome.sggw.pl/data/>). Tens of thousands of SSR markers have been developed from the whole genome sequences (Ren et al. 2009; Cavagnaro et al. 2010). These resources are providing an invaluable tool for high-resolution genetic mapping and gene cloning in cucumber (e.g., Ren et al. 2009; Weng et al. 2010; Kang et al. 2011; Miao et al. 2011). The objective of this research is to make full use of these resources to conduct fine genetic mapping of the *cp* locus to identify molecular markers that could be used in MAS in cucumber breeding and map-based cloning of this cucumber plant architecture gene.

Materials and methods

Plant materials and mapping populations

Two cucumber plant introduction lines, PI 308915 and PI 249561, were used in the present study. Both lines were obtained originally from the USDA North Central Regional Plant Introduction Station (Ames, IA, USA) and have undergone three generations of selfing. PI 308915 is a monoecious compact (*cp**cp*) cucumber line with indeterminate growth habit that has light green immature fruit, and light green to creamy white mature fruit with white spines. PI 308915 has the same phenotype as PI 308916 (Kauffman and Lower 1976; Ando and Grumet 2006) and both lines presumably carry the same compact gene. PI 249561 is an early flowering, monoecious line with regular vining habit (*CpCp*) bearing green immature and yellow mature fruit with black spines (Fig. 1).

A single F₁ plant from the cross between PI 249561 (female) and PI 308915 (male) was self-pollinated to produce F₂ progeny. 150 F₂-derived F₃ families were generated for progeny test to infer F₂ genotypes at the *compact* locus. For high-resolution mapping, an additional 1,123 F₂ plants from the same cross were used.

Phenotypic data collection

Parental lines, F₁, F₂, and F₃ families were evaluated in both the field (Hancock Agriculture Research Station, Hancock, Wisconsin) and the greenhouse of the University of Wisconsin at Madison, WI, USA. Three F₂ populations with a total of 1,569 plants from PI 249561 × PI 308915, designated as F₂-S (small), F₂-M (medium) and F₂-L



Fig. 1 Greenhouse and field performance of PI 308915 cucumber line with compact growth habit. **a** PI 249561 adult plant with regular vine (*left*) and PI 308915 plant with compact vining habit (*right*) in

the greenhouse. **b** PI 308915 dwarf plant with significantly reduced internode length. Fruit setting (**c**) and vine (**d**, front) of compact F_2 plants from PI 249561 \times PI 308915 cross in the field

(large), respectively, hereafter, were used at different stages of this project. The F_2 -S mapping population contained 150 $F_{2:3}$ families, 46 of which were used in initial mapping to develop a whole genome framework map and identify markers linked with the *compact* locus. Individual F_2 plants of F_2 -S population were grown in the greenhouse during 2009 winter season. The vining growth habit (compact or regular) of each plant was recorded, and each F_2 was self-pollinated to generate 150 F_3 families. Fifteen plants from each F_3 family were planted in Hancock field in 2010 summer growing season (June–September) to observe segregation of vine growing habit. Also planted in the Hancock field were 406 F_2 plants of the F_2 -M population, but only compact plants (homozygous *cpcp*) in this population were used for linkage analysis with molecular markers.

After closely linked markers flanking the *cp* locus were identified, a larger F_2 population (F_2 -L), with 1,013 plants was grown in the greenhouse in 96-cell plastic trays. Leaf samples from the F_2 -L population were collected in corresponding 96 deep-well plates for high throughput DNA extraction to identify recombinants defined by selected markers flanking the *cp* locus.

Several F_2 plants from the F_2 -S population were identified to be critical recombinants in the fine mapping stage. Their respective F_3 families were re-planted in the greenhouse to confirm segregation at the *cp* locus. At least 30 plants per family were phenotyped for each F_3 family.

Fine genetic mapping strategy

We started construction of a whole genome framework map with 46 F_2 plants of the F_2 -S population. One SSR

marker linked with the *cp* locus was identified. The Gy14 (Cavagnaro et al. 2010) and 9930 (Huang et al. 2009) scaffold sequences harboring this *cp*-linked marker were used to identify new markers. Meanwhile, other markers in this region on published genetic maps (Ren et al. 2009; Weng et al. 2010; Miao et al. 2011) were also tested. New polymorphic markers were mapped in the 46 F_2 plants, which led to identification of two markers flanking the *cp* locus that resided in the same scaffold. These markers were then applied to all 150 F_2 plants in the F_2 -S population and the compact-only plants in the F_2 -M population to confirm the relative locations of these markers. Lastly, two markers flanking the *cp* gene were applied to the F_2 -L population to identify all recombinants which served as the high-resolution population for fine mapping of *cp*.

Development of molecular markers for fine genetic mapping of the compact locus

During the initial mapping stage, published SSR markers that were developed from whole genome sequences of cucumber inbred lines 9930 (Ren et al. 2009) and Gy14 (Cavagnaro et al. 2010) were randomly chosen for polymorphism screening and development of a framework map. After flanking SSR markers were identified a whole genome scaffold-based chromosome walking was initiated to find more closely linked markers with the *cp* locus. In the target region of the scaffold under consideration, SSR markers in the Gy14 scaffold were selected from a collection of 83,689 SSRs developed by Cavagnaro et al. (2010). If no suitable SSRs were available, new SSRs were developed from the target region using the SSR primer design software SSR Locator (<http://minerva.ufpel.edu.br/~lmaia.faem/>)

(da Maia et al. 2008). The uniqueness of expected PCR products from these SSRs was verified through in silico (or virtual) PCR (Cavagnaro et al. 2010) using the Gy14 and 9930 draft genome assemblies as templates. This process was repeated for several rounds to identify markers toward the target gene. If all SSRs were exhausted, DNA fragments in the target region were PCR amplified and sequenced from both parental lines to identify insertion/deletion (Indels), or single nucleotide polymorphisms (SNPs). To visualize and map SNPs, cleaved amplified polymorphic sequence (CAPS) (Neff et al. 1998) or derived CAPS (dCAPS) (Michaels and Amisino 1998) markers were designed based on these SNPs. Information of all markers newly developed from the present study including primer sequences and their scaffold locations in the Gy14 and 9930 draft genomes is provided in Tables S1 and S2 of Supplemental file 1.

DNA sequencing, gene annotation and predication of gene function

Several genes that were potential candidates of the *compact* locus were sequenced during the fine mapping process. In the predicted gene region, sequencing primers were designed using Primer 3.0 with expected amplicon size from 600 to 900 bp. Neighboring primer pairs were designed to amplify PCR products with at least 100 bp overlap with its preceding fragment. Gene annotation was performed with the computer program FGENESH (<http://sunl.softberry.com/>) and function prediction was conducted with BLASTx at the NCBI (National Center for Biotechnology Information) website (<http://blast.ncbi.nlm.nih.gov>).

Molecular marker analysis

For the F₂-S mapping population, unexpanded young leaves from each F₂ plant were collected into 2.0 ml microcentrifuge tubes, lyophilized in a freeze dryer, and ground into fine powder in a high-throughput homogenizer (OPS Diagnostics, Lebanon, NJ, USA). Leaf samples from the F₂-M and F₂-L populations were collected into 2.0 ml 96-deep well plates, freeze dried and ground for DNA extraction. Genomic DNAs were extracted using the CTAB method.

Each polymerase chain reaction (PCR) contained 25 ng template DNA, 0.5 μM each of forward and reverse primers, 0.2 mM dNTP mix, 0.5 unit of *Taq* DNA polymerase and 1× PCR buffer (Fermentas, Glen Burnie, MD, USA) in a total volume of 10.0 μl. A “touch-down” PCR program was employed for all primer sets; the PCR products were size-fractionated in a 9% polyacrylamide gel and band patterns were visualized with silver staining (Weng et al. 2005).

For genotyping with CAPS or dCAPS markers, after performing specific primer-based PCR, the appropriate

restriction enzyme was added to the PCR reaction and incubated for 2–16 h at the temperatures in accordance with manufacturer’s instructions. Digested products were then separated in 9% polyacrylamide gel and visualized with silver staining as described above.

Data analysis

For phenotypic data on vining habit collected from the F₂-S mapping population, χ^2 tests for goodness-of-fit were used to test for deviations from the expected 1:2:1 segregation ratio (3:1 for segregation in the F₂-M and F₂-L populations). Linkage analysis of the *cp* locus with molecular markers was performed with JoinMap 3.0. Initial linkage groups (LGs) were established at a LOD threshold of 4.0 and the Kosambi function.

Results

Inheritance of compact growth habit

PI 308915 is an extreme dwarf type plant with significantly reduced internode length. The compact phenotype could be easily identified after the third true leaf stage. In the seedling stage, one characteristic of compact plants was that the upper part of the main stem was crooked with very short internode length. Adult compact plants in general had a height of <50 cm while PI 249561 usually grew up to 200 cm. Expression of the *compact* gene was generally not affected by environment. Thus in both the greenhouse and field, it was relatively easy to recognize the compact plants (Fig. 1). No difference was found in homozygous dominant (*CpCp*) and heterozygous (*Cpcp*) states of the *compact* locus.

Based on F₃ field data, segregation of 150 F₂ plants at the *cp* locus was 36 *cpcp*:81 *Cpcp*:33 *CpCp* ($P = 0.58$ in χ^2 test against 1:2:1) in the greenhouse. Among the additional 1,013 F₂ plants observed in the greenhouse, there were 779 *Cp_*:234 *cpcp* ($P = 0.20$ in χ^2 test against 3 regular:1 compact). Also in 2010 field season, of the 406 F₂ plants examined, there were 296 *Cp_* and 110 *cpcp* ($P = 0.33$ in χ^2 test against 3:1). These results confirmed the previous conclusion that compact growth habit was controlled by a simply inherited recessive gene (Kauffman and Lower 1976).

Cucumber genetic map construction and association of *compact* gene (*cp*) with microsatellite markers

Of a total of 2,335 SSR markers screened, 273 (11.7%) were polymorphic between PI 308915 and PI 249561, and 195 were applied for linkage analysis, of which eight failed to be assigned to any linkage group (LG). Thus, 187 SSR

markers and the compact growth habit gene *cp* were mapped to seven LGs using a LOD threshold of 4.0. Eighty markers were shared between this map and the one by Ren et al. (2009). Based on common markers, LG1–LG7 could be assigned to seven chromosomes (1–7), respectively, as defined in Koo et al. (2005) and Ren et al. (2009). The linkage map constructed is presented in Fig. 2 and map information is summarized in Table 1. More detailed information of all mapped markers including primer sequences and the Gy14 and 9930 cucumber whole genome scaffold locations is provided in Supplemental file 1-Table S1.

Of the 188 mapped loci, 23 (12.2%) showed segregation distortion (SD) in the F_2 -S population in χ^2 tests ($P = 0.05$) (loci with asterisks in Fig. 1). Nearly half (11 of 23) of the SD loci were clustered in the two distal ends of Chromosome 1; of the remaining 12, seven were in two clusters in Chromosomes 3 and 6, respectively (Fig. 1). Interestingly, for all the 17 markers with segregation

distortion from LG1 and LG3, the alleles from the maternal parent (PI 249561) were significantly fewer than expected (in favor of PI 308915); whereas the six markers in LG2 and LG6 all biased toward the paternal parent PI 308915.

Linkage analysis suggested that the *compact* gene, *cp* is located in the distal end of the long arm of Chromosome 4 (Fig. 2). Four SSR markers (UW084033, UW084212, UW084662, and UW084680) were cosegregating with the *cp* locus in the F_2 -S population (46 F_2 plants). Physically, all four markers were located in the same scaffold from the 9930 cucumber draft genome assembly (scaffold000032) and two scaffolds in the Gy14 assembly (scaffold02920 and scaffold03533) (Supplemental file 1-Table S1).

Fine mapping of *cp* gene in cucumber

SSR18551 (at 85.8 cM in Chromosome 4, Fig. 2) was the first marker showing linkage with the *cp* locus, which was previously mapped in the long arm of cucumber

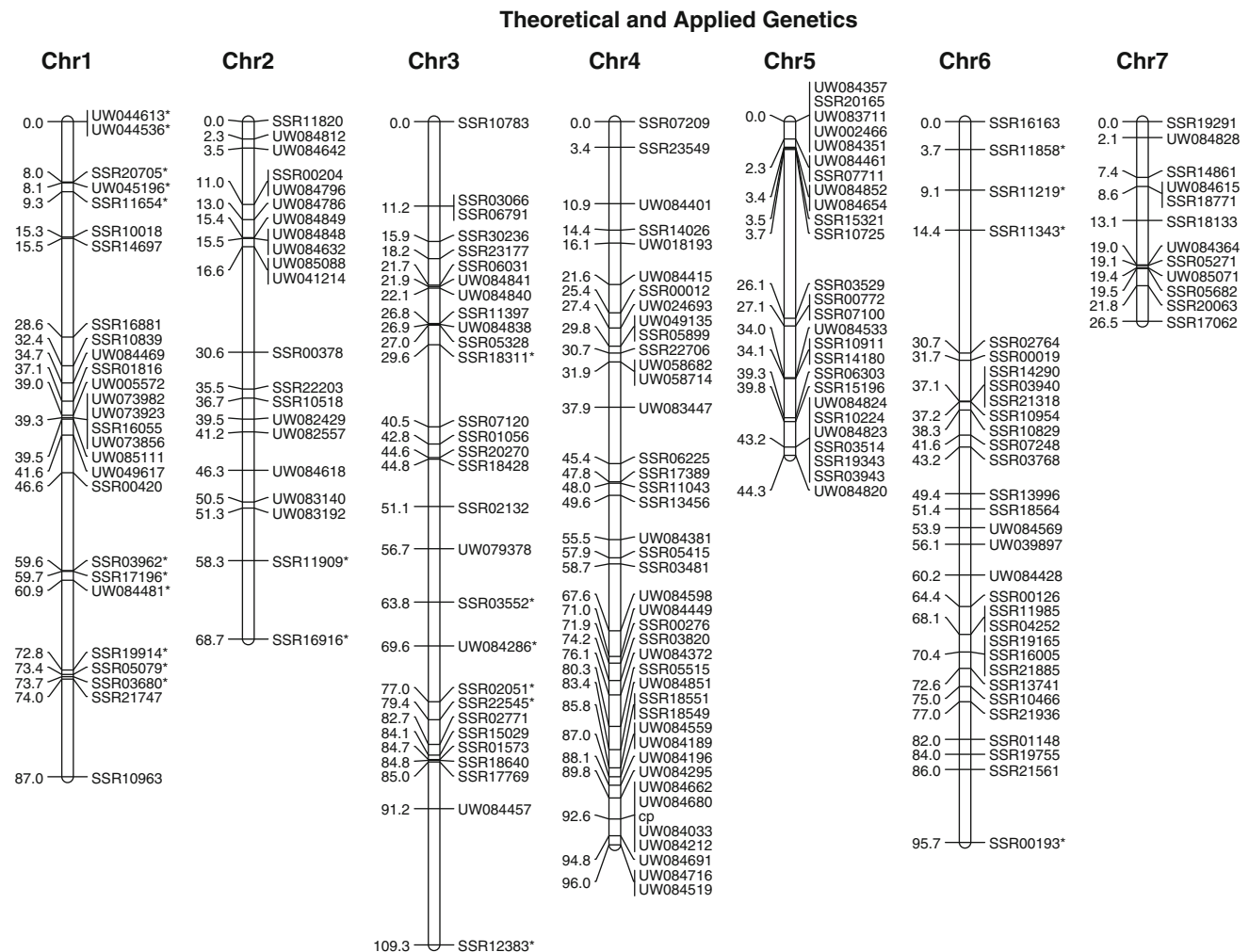


Fig. 2 A framework linkage map based on 46 F_2 plants from PI 249561 \times PI 308915 and 187 microsatellite markers. The *compact* locus (*cp*) was placed in the distal half of linkage group 4 (Chromosome 4). Loci with asterisks showed segregation distortion. *Chr* chromosome

Table 1 Distribution of genetic markers among seven cucumber chromosomes mapped with an $F_{2:3}$ population derived from PI 249561 \times PI 308915

Chromosome	# Loci (%)	Map length in cM (%)	Marker Interval (cM) ^a	Physical length [μm (%)] ^b
1	27 (14.4)	87.0 (16.5)	3.2	107.4 (14.9)
2	21 (11.2)	68.7 (13.0)	3.3	102.9 (14.3)
3	29 (15.4)	109.3 (20.7)	3.8	129.1 (17.9)
4	42 (22.3)	96.0 (18.2)	2.3	102.1 (14.2)
5	26 (13.8)	44.3 (8.4)	1.7	95 (13.2)
6	31 (16.5)	95.7 (18.1)	3.1	110 (15.3)
7	12 (6.4)	26.5 (5.0)	2.2	73.2 (10.2)
Sum	188 (100.0)	527.5 (100.0)	2.8	719.7 (100.0)

Numbers in parentheses in each column are percentages of the number in each chromosome based on the sum of seven chromosomes

^a Average genetic distances between adjacent markers

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

Chromosome 4 (Weng et al. 2010). Physically, SSR18551 is located in scaffold000032 of 9930 and scaffold03533 in Gy14. Consequently, this SSR marker was used as the starting point of our fine mapping effort.

Initially SSR18551 was mapped 3.8 cM away from the *cp* locus. Based on its scaffold location and other markers previously mapped in this region (Ren et al. 2009; Weng et al. 2010), 36 new SSRs were designed from 9930 scaffold000032 and three Gy14 scaffolds (scaffold00998, scaffold02541, and scaffold03533). One polymorphic marker, UW084033, was identified which was located at 1,132,618 bp (the start nucleotide position of left primer binding site, same hereinafter) in scaffold000032 and at the 774,214 bp position in Gy14 Scaffold03533. Linkage analysis suggested that SSR18551 and UW084033 were at the opposite sides of the *cp* locus. The physical distances between SSR18551 and UW084033 were approximately 545 kb in 9930 (scaffold000032) and 569 kb in Gy14 (scaffold03533).

Based on the genomic DNA sequences of Gy14 scaffold03533, 48 SSR markers were developed and screened for polymorphisms between PI 249561 and PI 308915. Three polymorphic markers, UW084189, UW084196, and UW084212 were identified and placed on the map. Both UW084212 and UW084033 were cosegregating with the *cp* locus (Fig. 3a). When 131 of the 150 F_2 -S $F_{2:3}$ families were used in linkage analysis, the *cp* gene was placed in between UW084196 and UW084212/UW084033, which were 2.5 and 0.6 cM away from the *cp* locus, respectively (data not shown).

The physical distance between UW084212 and UW084196 was \sim 275 kb in 9930 scaffold000032 and \sim 285 kb in Gy14 scaffold03533. To shorten the distance to the target gene (*cp*), 16 SSR markers were developed in this region and one polymorphic marker, UW084295, was mapped. The distance between UW084212 and UW084295 was

now approximately 147 kb in 9930 scaffold000032 and 143 kb in Gy14 scaffold03533. Annotation and gene function prediction of the 143 kb genomic DNA sequences in Gy14 scaffold03533 predicted 18 putative genes in this region including one encoding for the squalene epoxidase (SqE) (or squalene monooxygenase). SqE is an important enzyme in plants in the biosynthesis of brassinosteroid phytohormone, which has been shown to be responsible for dwarf phenotypes in a number of crop species (Mori et al. 2002). In the Gy14 cucumber genome, this SqE homolog was located from 725,355 to 731,870 bp (total 6,516 bp) in scaffold03533, and was very close to the SSR marker UW084212 (759,582 bp in Gy14 scaffold03533). We sequenced this gene from both parental lines. However, except for a putative SNP in the introns, no nucleotide changes were found between the two lines (data not shown). Meanwhile, we designed STS (sequence tagged site) primers to amplify genic regions between UW084212 and UW084295. We sequenced 11 DNA fragments in this region and identified one indel marker, UWSTS0125 (737,643 bp in Gy14 scaffold03533), that was polymorphic between PI 249561 and PI 308915. Linkage analysis in 150 $F_{2:3}$ families identified two recombination events between UWSTS0125 and the *cp* locus, but UWSTS0125 was cosegregating with UW084212 and UW084033 in this F_2 -S population (Fig. 3b). This result not only excluded SqE as the candidate gene of *cp*, but it was also evident that the *cp* gene was not located in the region delimited by the markers UW084295 and the block of three cosegregating markers (UW084033, UW084212, and UWSTS0125). Rather, *cp* should be located outside the three marker block toward the telomeric direction (opposite to marker UW084295).

To identify markers outside the UW084295-UW084212 region, new microsatellite markers were developed from two 9930 scaffolds, scaffold000032 (1,690 kb in length) and scaffold000063 (965 kb), as well as two Gy14

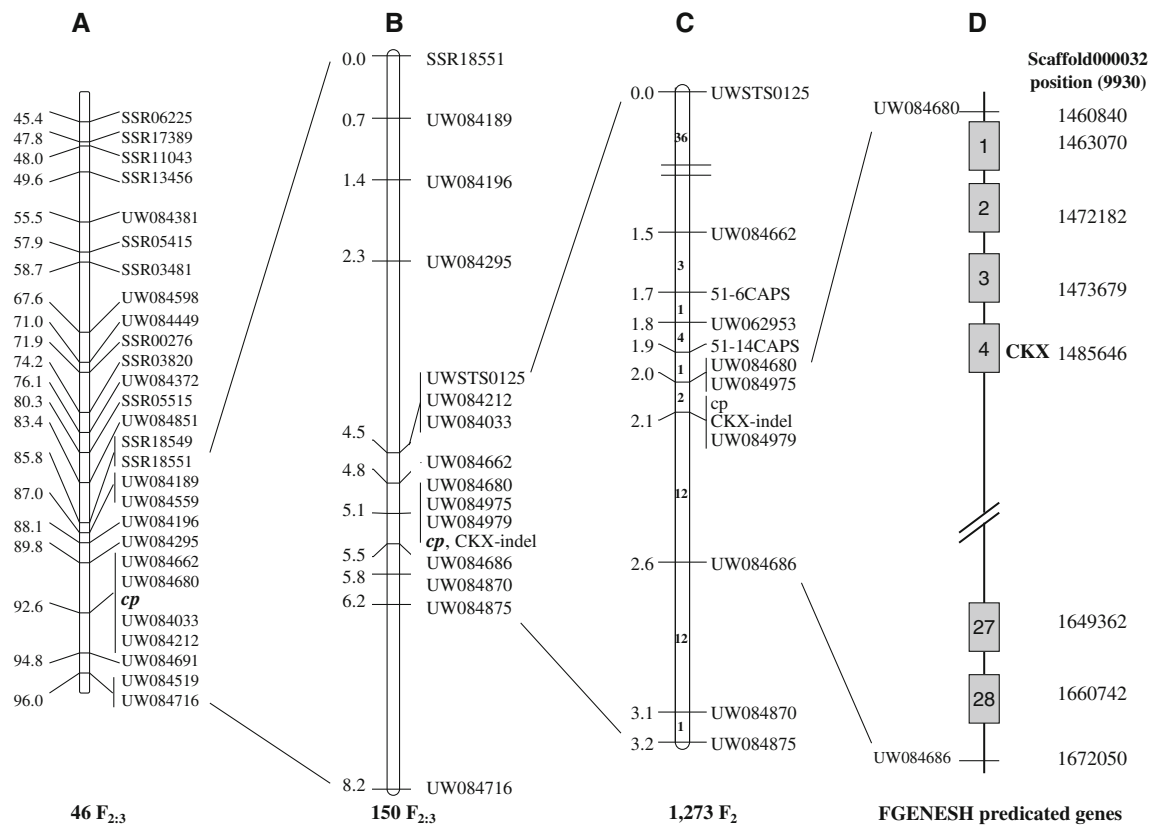


Fig. 3 Fine genetic map and physical map of *cp* locus. **a** Part of linkage group 4 taken directly from Fig. 2. **b** High-resolution linkage map based on 150 $F_{2:3}$ families. Vertical bars delimit cosegregating marker loci. **c** Fine genetic map in the *compact* gene (*cp*) region based on 1,273 F_2 plants. The number in each rectangle box (representing chromosome segments) signifies number of recombination events

occurred in this region as defined by two adjacent markers. **d** The 220-kb genomic DNA region between UW084680 and UW084686 in which 28 genes were predicted. The scaffold location (in bp) of each marker or gene is shown to the right of the box. CKX cytokinin oxidase

scaffolds, scaffold03533 (886 kb) and scaffold02500 (1080 kb). In total, 110 new SSRs were developed, from which six polymorphic markers, UW084662, UW084680, UW084686, UW084716, UW084870, and UW084875 were identified and mapped using 150 $F_{2:3}$ families (Fig. 3b). Although UW084680 was cosegregating with *cp* in the F_2 -S population, it was on the proximal side of the *cp* locus when 110 compact plants from the field (F_2 -M population) were included in linkage analysis (data not shown). Consequently these data clearly indicated that the *cp* locus was located between the two markers UW084680 and UW084686.

In the 9930 scaffold000032, UW084680 and UW084686 were ~220 kb apart. FGENESH program predicted 28 putative genes in this region (listed in Supplemental file 1-Table S3). One of the 28 genes encodes a cytokinin oxidase (CKX), which is a key regulator of the level of the phytohormone cytokinins in plants. Since cytokinins seem to be involved in reducing plant height (see “Discussion” below), CKX could be a potential candidate of the *cp* locus. Based on the sequence information in 9930 scaffold000032 of the CKX gene, which was 4,127 bp in 9930 cucumber,

we cloned this gene from both PI 249561 and PI 308915. Alignment of the CKX gene homologs among four genotypes (PI 308915, PI 249561, 9930 and Gy14) is provided in Supplemental file 2. Based on a 3-bp deletion in the predicted exonic region of the CKX gene in PI 308915, one marker, CKX-indel was developed. In addition, two new microsatellite markers, UW084975 and UW084979 from the region delimited by UW084680 and UW084686 were mapped. It turned out that UW084979 was in the CKX gene region; however, the polymorphism was not because of the number of microsatellite motifs between the two lines, but rather due to a 4-bp deletion in PI 249561 (nucleotides 3080–3083, see Supplemental file 2). No genetic recombination was found between the *cp* locus and any of the three newly developed markers (UW084975, UW084979 and CKX-indel) in the 150 $F_{2:3}$ mapping population. The resulted genetic map is shown in Fig. 3b.

With cosegregating markers identified for the *cp* locus in the F_2 -S population, next, we employed 1,123 F_2 plants for fine genetic mapping of *cp*, which included 110 compact plants from 406 F_2 plants (F_2 -M population) grown in the

field and 1,013 F₂ plants from the F₂-L population grown in the greenhouse. Two markers flanking the *cp* locus, UWSTS0125 and UW084875 (Fig. 3b) were used to genotype these F₂ plants. Among the 1,273 F₂ plants (150 from F₂-S, 110 from F₂-M and 1,013 from F₂-L), there were 72 recombinants between UWSTS0125 and UW084875. All markers mapped in this region in the F₂-S population (Fig. 3b) were used to genotype the 72 recombinants. The high-resolution map of *cp* with the number of recombinants in each chromosomal segment defined by adjacent markers is shown in Fig. 3c. The order of all markers on this fine genetic map was completely consistent with their positions in the scaffolds of either the 9930 or Gy14 cucumber draft genomes (see Table S2 for details). Among the 1,273 F₂ plants examined, no recombination was found between the *compact* locus and the two markers, CKX-indel and UW084979, within the CKX gene region (Fig. 3c).

Predicted genes in the *cp* region

Fine genetic mapping placed the *cp* locus in the region delimited by markers UW084680 and UW084686, which was approximately 220 kb in length in the 9930 genome scaffold000032. Using FGENESH program, 28 genes were predicted in this region. Information about the position and predicted function of each gene is presented in Table S3 (Supplemental file 1). Among them, the fourth gene that encodes a putative cytokinin oxidase (CKX) seems to be a good candidate of the *cp* gene.

We cloned the CKX gene homologs from both PI 308915 and PI 249561 and the whole length of its genomic DNA sequence was 4,127 and 4,229 bp, respectively, in the two parental lines. FGENESH predicted four exons (coding regions, CDS) in this gene. Alignment of CKX gene sequences between PI 249561 and PI 308915 revealed the only difference between the two parental lines in the CKX gene exonic regions was a 3-bp deletion (CKX-indel in Fig. 3) in PI 308915. However, there were 11 SNPs and 6 indels (1–5 bp) in the introns between the two lines (Supplemental file 2).

Since most dwarf phenotypes in plants are related with plant hormones (see “Discussion” below), we also examined genomic regions surrounding the *cp* locus for other possible plant hormone-related genes. Two such putative genes were identified: the first encoded a CGA1 (Cytokinin-responsive GATA Factor 1) transcription factor, which was also residing in 9930 scaffold000032 and was ~244 kb before the CKX gene; the other was a CTR1-like protein kinase that was approximately 42 kb before the CKX gene. We partially sequenced the two genes and designed two CAPS markers: 51-6CAPS for the cytokinin responsive factor, and 51-14CAPS for the CTR1-like

protein kinase gene. Both markers were mapped in the region delimited by UW084680 and UW084662 (Fig. 3c) and were excluded as possible candidates of the compact gene.

Discussion

Genome wide SSR polymorphism in cultivated cucumber

Genetic diversity studies revealed that cucumber has a very narrow genetic base with a polymorphism level of 3–12% (Knerl et al. 1989; Dijkhuizen et al. 1996; Serquen et al. 1997). In the present study, we screened 2,335 microsatellite primer pairs between the two intra-varietal parental lines of cultivated cucumber and only 11.1% were polymorphic; which was similar to that of Weng et al. (2010) between Gy7 and H-19 (17.0%) and Zhang et al. (2010) between 9110Gt and 9930 (12.6% for SSRs in Chromosome 2) providing further evidence of low genetic diversity in cultivated cucumbers. The overall polymorphism level of SSR markers among cultivated cucumbers is likely in the range 10–20% at the whole genome level, which gives us some general idea in preparing for a genome mapping project. The polymorphism may be even lower at fine scale or the DNA sequence level. In the present study, we developed 210 new SSR markers in the region between SSR18551 and UW084716 (Fig. 3b); 11 were polymorphic with a polymorphism level of 5.2%. At the DNA sequence level, polymorphism varies significantly depending on the genome regions under investigation. In the sequenced 4,149 bp CKX gene region (Supplemental file 2), 8 indels (1–5 bp) and 12 SNPs were identified between PI 308915 and PI 249561, which translated to 4.82 polymorphisms per kb. However, in the *Psm* (paternal sorting of mitochondria) region, Al-Faifi et al. (2008) only found six polymorphisms across two genomic regions with a total of 150,649 bp genomic DNA between two parental lines (0.04 SNP per kb). These observations suggest that genetic mapping and map-based cloning in cucumber still require non-trivial efforts. Nevertheless, as evidenced from the present study, the availability of whole genome sequences is making this work much easier than before.

The cucumber whole genome framework linkage map

The framework genetic map was composed of 188 loci in seven linkage groups (Fig. 2). Although the map was constructed with only 46 F₂ plants, clustering of marker loci on the map was not obvious. The order of loci on this map seems highly reliable and accurate as evidenced from

the following two facts. First, the present map shared 80 and 22 markers with those developed by Ren et al. (2009) and Weng et al. (2010), respectively. Except in very few cases (see below for “Discussion”), the order of shared loci in each linkage group was largely co-linear. Second, molecular markers with in silico PCR amplicons located in the same draft genome scaffold were in general mapped in adjacent locations on the genetic map (see Supplemental Table S1 for details).

When compared with previously published SSR-based cucumber genetic maps (Ren et al. 2009; Weng et al. 2010; Zhang et al. 2010), some minor discrepancies in map locations (marker loci were not consistent within 5 cM mapping distances) were observed. For example, eight of the first 17 markers in the distal end of Chromosome 4 short arm from this study (Fig. 2) were shared with those developed by Ren et al. (2009) with 77 RILs from a wide cross between Gy14 and wild cucumber line PI 183967. The eight SSR loci spanned 48.0 cM on the present map (Fig. 2), but they were in two clusters (4.4 and 5.0 cM) on the wide cross map, which was due to strong recombination suppression in this region caused by possible changes in chromosome structure between Gy14 and PI 183967 (Ren et al. 2009). Since the map locations on the present map were consistent with their relative positions in the whole genome scaffolds (Table S1), the inconsistent order of mapped loci between the two maps was likely due to the possible inversion or other minor chromosomal structure changes between the two parental lines revealed in by Ren et al. (2009) although mapping errors could not be excluded in the early study.

The only marker mapped in this study with obvious discrepancy in map location was SSR03962 in Chromosome 1 (at 59.6 cM, Fig. 2). Ren et al. (2009) placed this marker in Chromosome 6 (49.0 cM or bin 6B73). This discrepancy was probably due to the fact that there are at least two copies of this marker in the cucumber genome. No in silico PCR product was identified from SSR03962 in the 9930 draft genome probably because it did not meet one or more of the criteria for in silico PCR. In the Gy14 draft genome, the amplicon was from scaffold01543 (Supplemental Table S1). However, when the DNA sequence from in silico PCR product was used to run BLASTn against the 9930 genome assembly, there was one hit from the 9930 scaffold000089. Multiple SSR markers from this scaffold have been mapped in both cucumber Chromosomes 1 and 6 (Ren et al. 2009; Yang et al. unpublished data). This observation also suggests possible incorrect assembly of scaffold000089 in the 9930 draft genome. Incorrect assembly into scaffolds in either the 9930 or Gy14 draft genome could also explain a number of discrepancies for some markers between their map positions and scaffold locations. In Table S1, it can be seen that

several markers were mapped in different locations of the same chromosome but belong to the same scaffold. For example, in Chromosome 6, the three markers SSR11219 (at 9.1 cM), SSR11343 (at 14.4 cM), and SSR16005 (at 70.4 cM) were placed in different locations, but in silico PCR indicated that they were from the same Gy14 scaffold03611. This is likely due to incorrect assembly of this scaffold because in the 9930 draft genome, its in silico PCR amplicon was annotated as a repetitive sequence (scaffold_repeat018375) (Huang et al. 2009). These observations point out the importance of a high quality cucumber genetic map of densely spaced markers that could be used to verify the accuracy of whole genome assemblies and affirm the correct placement of scaffolds in each chromosome.

Fine genetic mapping and potential candidate gene of the compact locus

For the first time, we located the *cp* gene onto the distal end of the long arm of cucumber Chromosome 4 (Fig. 2). Cucumber whole genome scaffold-based chromosome walking enabled us to identify cosegregating markers with stepwise increase of the sizes of mapping populations (from 46 to 150, and 1,273 F₂ plants). Gene prediction in the 220 kb genomic sequence encompassing the *cp* locus identified the cytokinin oxidase (CKX) gene as a possible candidate of the *compact* locus.

Manipulation of plant architecture and especially reduced plant height (dwarf or compact growth habit) plays an important role in plant breeding. A number of dwarf genes have been cloned from several crop species (reviewed by Hedden 2003; Wang and Li 2006). In most cases, two classes of plant hormones, gibberellin acid (GA) and brassinosteroid (BR) are involved in the reduced plant height (Mandava 1988; Clouse and Sasse 1998; Fujioka and Yokota 2003). Often the dwarf mutants are due to malfunctioning genes in GA or BR biosynthesis, or signal transduction pathways (Vogler and Kuhlemeier 2003; Mori et al. 2002; Wang and Li 2006; Li et al. 2011). In a few cases, defects in the transporter involved in polar movement of the phytohormone auxins were found to be responsible for dwarf mutants in sorghum and Arabidopsis (Multani et al. 2003; Dai et al. 2006).

Direct association of the cytokinin phytohormone with dwarf phenotypes has not been well documented. In the case of extreme dwarf wheat, ‘Tibet Dwarf’ was found to accumulate elevated quantities of cytokinins (Knauber and Banowitz 1992; Banowitz 1997; Banowitz et al. 1999). The levels of cytokinins in plant tissues are regulated by cytokinin oxidase/dehydrogenase (CKX, EC 1.5.99.12) enzymes. The irreversible degradation of cytokinins catalyzed by CKX is an important and well-known mechanism

of cytokinin regulation (Jones and Schreiber 1997). In barley, lower CKX enzyme activity in transgenic plants correlated with significantly lower transcript accumulation in developing spiklets. The mean plant heights in transgenic lines of the CKX gene were lower than those of the control plants (Zalewski et al. 2010).

Among the 28 predicted genes between UW084680 and UW084686 (Supplemental file 1-Table S3), CKX seems to be the most probable candidate for the *compact* gene. However, additional work is needed to confirm the identity of the CKX gene and the *compact* locus. We plan to genotype an additional 1,200 F₂ plants from the cross between PI 308195 and PI 249561 with four markers (UW084680, UW084686, CKX-indel, and UW084979). Recombinants will be selfed to obtain F₃ families which will be used to infer F₂ genotypes at the *cp* locus. Once the identity of the CKX gene as the *cp* candidate is verified, additional experiments will be conducted for function validation of the candidate gene.

Acknowledgments The authors thank Linda Crubaugh for technical assistance, Michael Havey for critical reading of the manuscript. Yuhong Li was supported by a scholarship sponsored by the China Scholarship Council, China. Mamta Pathak was supported by a training grant from Punjab Agricultural University, Ludhiana, India. Xiaoming He was supported by the Guangdong Academy of Agricultural Sciences, China. The authors greatly appreciate the support of these sponsors.

References

- Ando K, Grumet R (2006) Evaluation of altered cucumber plant architecture as a means to reduce *Phytophthora capsici* disease incidence on cucumber fruit. *J Am Soc Hortic Sci* 131:491–498
- Al-Faifi S, Meyer JDF, Garcia-Mas J, Monforte AJ, Havey MJ (2008) Exploiting synteny in *Cucumis* for mapping of *Psm*: a unique locus controlling paternal mitochondrial sorting. *Theor Appl Genet* 117:523–529
- Banowetz GM (1997) Cultivars of hexaploid wheat of contrasting stature and chlorophyll retention differ in cytokinin content and responsiveness. *Ann Bot* 79:185–190
- Banowetz GM, Ammar K, Chen DD (1999) Temperature effects on cytokinin accumulation and kernel mass in a dwarf wheat. *Ann Bot* 83:303–307
- Cavagnaro PF, Senalik DA, Yang LM, Simon PW, Harkins TT, Kodira CD, Huang SW, Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics* 11:569
- Clouse SD, Sasse JM (1998) BRASSINOSTEROIDS: essential regulators of plant growth and development. *Annu Rev Plant Physiol Plant Mol Biol* 49:427–451
- Cramer CS, Wehner TC (2000) Path analysis of the correlation between fruit number and plant traits of cucumber populations. *HortScience* 35:708–711
- Crienen J, Reuling G, Segers B, van de Wal M (2009) New cucumber plants with a compact growth habit. Patent, International publication number WO 2009/059777 A1
- da Maia LC, Palmieri DA, de Souza VQ, Kopp MM, de Carvalho FIF, Oliveira AC (2008) SSR Locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int J Plant Genomics* 2008:412696
- Dai Y, Wang HZ, Li BH, Huang J, Liu XF, Zhou YH, Mou ZL, Li JY (2006) Increased expression of *MAP KINASE KINASE7* causes deficiency in polar auxin transport and leads to plant architectural abnormality in Arabidopsis. *Plant Cell* 18:308–320
- de Ruiter AC, van der Knapp BJ, Robinson RW (1980) Rosette, a spontaneous cucumber mutant arising from cucumber-musk-melon mentor pollen. *Cucurbit Genet Coop Rep* 3:4
- Dijkhuizen A, Kennard WC, Havey MJ, Staub JE (1996) RFLP variation and genetic relationships in cultivated cucumber. *Euphytica* 90:79–89
- Edwards MD, Lower RL (1981) Investigations into the characteristics of seeds from compact cucumber plants. *Cucurbit Genet Coop Report* 4:2–4 (article 1)
- Fujioka S, Yokota T (2003) Biosynthesis and metabolism of brassinosteroids. *Annu Rev Plant Biol* 54:137–164
- Gusmini G, Wehner TC (2008) Fifty-five years of yield improvement for cucumber, melon, and watermelon in the United States. *HortTechnology* 18:9–12
- Hedden P (2003) The genes of the green revolution. *Trends Genet* 19:5–9
- Huang S, Li RQ, Zhang ZH, Li L, Gu XF, Fan W, Lucas WJ, Wang XW et al (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* 41:1275–1281
- Jones RJ, Schreiber BMN (1997) Role and function of cytokinin oxidase in plants. *Plant Growth Regul* 23:123–134
- Kang HX, Weng T, Yang YH, Zhang ZH, Zhang SP, Mao ZC, Cheng GH, Gu XF, Huang SW, Xie BY (2011) Fine genetic mapping localizes cucumber scab resistance gene *Ccu* into an *R* gene cluster. *Theor Appl Genet* 122:795–803
- Kauffman CS, Lower RL (1976) Inheritance of an extreme dwarf plant type in the cucumber. *J Am Soc Hortic Sci* 101:150–151
- Knauber DC, Banowetz GM (1992) Phenotypic analysis of a dwarf wheat (*Triticum aestivum* L.) with altered phytochrome-mediated growth responses. *Plant Physiol* 100:1901–1908
- Knerr LD, Staub JE, Holder DJ, May BP (1989) Genetic diversity in *Cucumis sativus* L. assessed by variation at 18 allozyme coding loci. *Theor Appl Genet* 78:119–128
- Koo DH, Choi HW, Cho J, Hur Y, Bang JW (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
- Kubicki B, Soltysiak U, Korzeniewska A (1986) Induced mutations in cucumber (*Cucumis sativus* L.) V. Compact type of growth. *Genet Pol* 27:289–298
- Li J, Jiang JF, Qian Q, Xu YY, Zhang C, Xiao J, Du C, Luo W, Zou GX, Chen ML, Huang YQ, Feng YQ, Cheng ZK, Yuan M, Chong K (2011) Mutation of rice *BC12/GDD1*, which encodes a kinesin-like protein that binds to a GA biosynthesis gene promoter, leads to dwarfism with impaired cell elongation. *Plant Cell* 23:628–640
- Mandava NB (1988) Plant growth-promoting brassinosteroids. *Annu Rev Plant Physiol Plant Mol Biol* 39:23–52
- Miao H, Zhang SP, Wang XW, Zhang ZH, Li M, Mu SQ, Cheng ZC, Zhang RW, Huang SW, Xie BY, Fang ZY, Zhang ZX, Weng Y, Gu XF (2011) A linkage map of cultivated cucumber (*Cucumis sativus* L.) with 248 microsatellite marker loci and seven genes for horticulturally important traits. *Euphytica*. doi:10.1007/s10681-011-0410-5
- Michaels SD, Amisino RM (1998) A robust method for the detecting single-nucleotide changes as polymorphic markers by PCR. *Plant J* 14:381–385

- Mori M, Nomura T, Ooka H, Ishizaka M, Yokota T, Sugimoto K, Okabe K, Kajiwara H, Satoh K, Yamamoto K, Hirochika H, Kikuchi S (2002) Isolation and characterization of a rice dwarf mutant with a defect in brassinosteroid biosynthesis. *Plant Physiol* 130:1152–1161
- Multani DS, Briggs SP, Chamberlin MA et al (2003) Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. *Science* 302:81–84
- Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant J* 14:387–392
- Niemirowicz-Szczytt K, Rucinska M, Korzeniewsia A (1996) An induced mutation in cucumber: super compact. *Cucurbit Genet Coop Rep* 19:1–3 (article 1)
- Ren Y, Zhang ZH, Liu JH, Staub JE, Han YH, Cheng ZC, Li XF, Lu JY, Miao H, Kang HX, Bie BY, Gu XF, Wang XW, Du YC, Jin WW, Huang SW (2009) An integrated genetic and cytogenetic map of the cucumber genome. *PLoS One* 4:e5795
- Serquen FC, Bacher J, Staub JE (1997) QTL mapping of horticultural traits in a narrow cross in cucumber using random-amplified polymorphic DNA markers. *Mol Breed* 3:257–268
- Vogler H, Kuhlemeier C (2003) Simple hormones but complex signaling. *Curr Opin Plant Biol* 6:51–56
- Wang YH, Li JY (2006) Genes controlling plant architecture. *Curr Opin Biotechnol* 17:123–129
- Wehner TC (1989) Breeding for improved yield in cucumber. *Plant Breed Rev* 6:323–359
- Weng Y, Li W, Devkota RN, Rudd JC (2005) Microsatellite markers associated with two *Aegilops tauschii*-derived greenbug resistance loci in wheat. *Theor Appl Genet* 110:462–469
- Weng Y, Johnson S, Staub JE, Huang S (2010) An extended microsatellite genetic map of cucumber, *Cucumis sativus* L. *HortScience* 45:880–886
- Zalewski W, Galuszka P, Gasparis S, Orczyk W, Nadolska-Orczyk A (2010) Silencing of the *HvCKX1* gene decreases the cytokinin oxidase/dehydrogenase level in barley and leads to higher plant productivity. *J Exp Bot* 61:1839–1851
- Zhang SP, Miao H, Gu XF, Yang YH, Xie BY, Wang XW, Huang SW, Du YC, Sun RF (2010) Genetic mapping of the scab resistance gene in cucumber. *J Am Soc Hortic Sci* 135:53–58