

Next-generation sequencing, FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in *Cucumis*

Luming Yang^{1,†}, Dal-Hoe Koo^{1,†}, Dawei Li^{1,2}, Tao Zhang¹, Jiming Jiang¹, Feishi Luan³, Susanne S. Renner⁴, Elizabeth Hénaff⁵, Walter Sanseverino⁶, Jordi Garcia-Mas⁶, Josep Casacuberta⁵, Douglas A. Senalik^{1,7}, Philipp W. Simon^{1,7}, Jinfeng Chen⁸ and Yiqun Weng^{1,7,*}

¹Horticulture Department, University of Wisconsin, Madison, WI 53706, USA,

²Horticulture College, Northwest A & F University, Yangling 712100, China,

³Horticulture College, Northeast Agricultural University, Harbin 150030, China,

⁴Department of Biology, University of Munich, 80638 Munich, Germany,

⁵Centre for Research in Agricultural Genomics Consejo Superior de Investigaciones Científicas-Institut de Recerca i Tecnologia Agroalimentàries-Universitat Autònoma de Barcelona-Universitat de Barcelona, 08193 Barcelona, Spain,

⁶Institut de Recerca i Tecnologia Agroalimentàries, Centre for Research in Agricultural Genomics Consejo Superior de Investigaciones Científicas-Institut de Recerca i Tecnologia Agroalimentàries-Universitat Autònoma de Barcelona-Universitat de Barcelona, 08193 Barcelona, Spain,

⁷US Department of Agriculture/Agricultural Research Service, Vegetable Crops Research Unit, 1575 Linden Drive, Madison, WI 53706, USA, and

⁸College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China

Received 27 July 2013; revised 7 October 2013; accepted 10 October 2013.

*For correspondence (e-mail weng4@wisc.edu).

†These authors contributed equally to this work.

SUMMARY

In the large Cucurbitaceae genus *Cucumis*, cucumber (*C. sativus*) is the only species with $2n = 2x = 14$ chromosomes. The majority of the remaining species, including melon (*C. melo*) and the sister species of cucumber, *C. hystrix*, have $2n = 2x = 24$ chromosomes, implying a reduction from $n = 12$ to $n = 7$. To understand the underlying mechanisms, we investigated chromosome synteny among cucumber, *C. hystrix* and melon using integrated and complementary approaches. We identified 14 inversions and a *C. hystrix* lineage-specific reciprocal inversion between *C. hystrix* and melon. The results reveal the location and orientation of 53 *C. hystrix* syntenic blocks on the seven cucumber chromosomes, and allow us to infer at least 59 chromosome rearrangement events that led to the seven cucumber chromosomes, including five fusions, four translocations, and 50 inversions. The 12 inferred chromosomes (AK1–AK12) of an ancestor similar to melon and *C. hystrix* had strikingly different evolutionary fates, with cucumber chromosome C1 apparently resulting from insertion of chromosome AK12 into the centromeric region of translocated AK2/AK8, cucumber chromosome C3 originating from a Robertsonian-like translocation between AK4 and AK6, and cucumber chromosome C5 originating from fusion of AK9 and AK10. Chromosomes C2, C4 and C6 were the result of complex reshuffling of syntenic blocks from three (AK3, AK5 and AK11), three (AK5, AK7 and AK8) and five (AK2, AK3, AK5, AK8 and AK11) ancestral chromosomes, respectively, through 33 fusion, translocation and inversion events. Previous results (Huang, S., Li, R., Zhang, Z. *et al.*, 2009, *Nat. Genet.* 41, 1275–1281; Li, D., Cuevas, H.E., Yang, L., Li, Y., Garcia-Mas, J., Zalapa, J., Staub, J.E., Luan, F., Reddy, U., He, X., Gong, Z., Weng, Y. 2011a, *BMC Genomics*, 12, 396) showing that cucumber C7 stayed largely intact during the entire evolution of *Cucumis* are supported. Results from this study allow a fine-scale understanding of the mechanisms of dysploid chromosome reduction that has not been achieved previously.

Keywords: chromosome evolution, comparative genome mapping, *Cucumis*, *de novo* genome sequence, dysploid chromosome number reduction, synteny.

INTRODUCTION

The genus *Cucumis* (family Cucurbitaceae) contains 52 species (Schaefer, 2007; Sebastian *et al.*, 2010), including two important vegetable crops, cucumber (*C. sativus* L., $2n = 2x = 14$) and melon (*C. melo* L., $2n = 2x = 24$). Both species are of Asian origin, and diverged from a common ancestor approximately 10 million years ago (Sebastian *et al.*, 2010) (Figure 1). Among the approximately 40 *Cucumis* species for which chromosome counts have been obtained, *C. sativus* is the only species with $2n = 14$ chromosomes; all other species have $2n = 24$, and occasionally $2n = 48$ or 72 chromosomes (Kirkbride, 1993). The origin of the seven cucumber chromosomes has been a long-standing question, with opposite hypotheses being proposed to explain it: a fragmentation hypothesis that postulated *de novo* regeneration of centromeres from $n = 7$ to $n = 12$ (Kozhukhow, 1930; Whitaker, 1933; Bhaduri and Bose, 1947) and a fusion hypothesis that postulated that $n = 7$ was derived from $n = 12$ via unequal translocation or fusion of non-homologous chromosomes (Trivedi and Roy, 1970). More recently, Han *et al.* (2009) proposed the involvement of centromere repositioning in evolution of cultivated cucumber chromosome C7 based on cucumber/melon comparative FISH mapping data. The results of molecular phylogenetic studies suggest that $n = 12$ is ancestral in the genus *Cucumis* (Ghebretinsae *et al.*, 2007; Renner *et al.*, 2007; Sebastian *et al.*, 2010).

Despite their distant phylogenetic relationship and sexual incompatibility, the genome sequences of melon and cucumber appear to be highly conserved. The cross-species transferability of molecular markers between cucumber and melon is approximately 50% (e.g. Neuhausen, 1992; Katzir *et al.*, 1996; Danin-Poleg *et al.*, 2000; Park *et al.*, 2004; Gonzalez *et al.*, 2010). The genome size of melon is approximately 425 Mb and that of cucumber is approximately 367 Mb, but the number of protein-coding

genes in both genomes is similar (Li *et al.*, 2011b; Garcia-Mas *et al.*, 2012), and the size difference is believed to be mainly due to expansion of intergenic regions and proliferation of transposable elements in the melon genome (Gonzalez *et al.*, 2010; Garcia-Mas *et al.*, 2012). Comparative genetic mapping and whole-genome alignment have enabled establishment of syntenic relationships of 12 melon and seven cucumber chromosomes, showing that six of the seven chromosomes arose from fusions and intra-chromosome rearrangements, while one has remained largely unchanged between cucumber and melon (Huang *et al.*, 2009; Li *et al.*, 2011a; Garcia-Mas *et al.*, 2012).

Inferring the history of chromosome reduction in *Cucumis* based on comparative studies between melon and cucumber only is insufficient for a detailed understanding because at least 24 other species fall between these two species in the most complete phylogeny of the genus (Sebastian *et al.*, 2010). Of these 24 species, the one closest to *C. sativus* is *C. hystrix* ($2n = 2x = 24$), from which it diverged approximately 4.6 million years ago (Sebastian *et al.*, 2010). Inter-specific hybrids between *C. hystrix* and *C. sativus* have been generated by embryo rescue (Chen *et al.*, 1997). Its position as the sister species of *C. sativus* confers a critical role for *C. hystrix* in understanding the process of chromosome reduction in *Cucumis*, and the family-wide phylogenies now available also establish the distance of *C. sativus* and *C. hystrix* to *C. melo* and other species in *Cucumis* and related genera.

Studying chromosome synteny and colinearity among species for which a phylogeny establishes the direction of evolution allows reconstruction of the likely events that led to present karyotypes. Events of particular interest are reductions in chromosome number (dysploidy), which have been analyzed using this approach in Brassicaceae

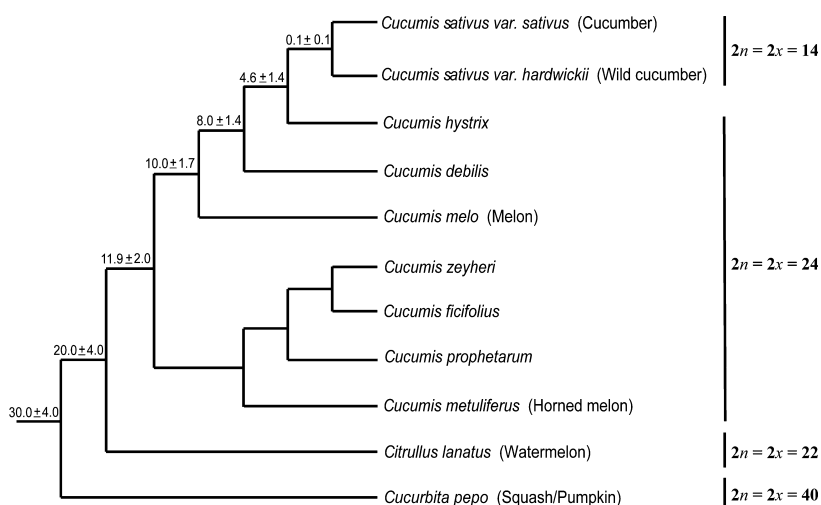


Figure 1. Chronogram of selected Cucurbitaceae species used in the present study. The estimations of divergence (in millions of years ago) are from Sebastian *et al.* (2010).

(Koch and Kiefer, 2005; Yogeewaran *et al.*, 2005; Lysak *et al.*, 2006; Mandakova and Lysak, 2008; Cheng *et al.*, 2013), Poaceae (Luo *et al.*, 2009) and Rosaceae (Vilanova *et al.*, 2008; Illa *et al.*, 2011; Jung *et al.*, 2012). Here we performed whole-genome sequencing and *de novo* genome assembly of two *C. hystrix* accessions to develop a high-density *C. hystrix* linkage map, which was then used to establish genome-wide chromosome synteny and marker colinearity by alignment of shared markers to the cucumber and melon draft genome assemblies. The melon and cucumber synteny was refined through whole-genome alignment. We also performed large-scale comparative fluorescence *in situ* hybridization (FISH) mapping among the melon, cucumber and *C. hystrix* genomes to validate or reveal syntenic blocks or structural arrangements. The previously inferred conservation and retention of cucumber chromosome 7 (C7) during evolution was validated among several *Cucumis* species, and micro-synteny analyses and annotation of an 8 Mb region in C7 was performed for cucumber, melon and watermelon (*Citrullus lanatus*, $2n = 2x = 22$). The resulting data shed light on the mechanisms of dysploid chromosome reduction that led from an $n = 12$ ancestor to the $n = 7$ karyotype of cucumber.

RESULTS

Next-generation sequencing and *de novo* assembly of the *C. hystrix* genome

We sequenced two accessions of *C. hystrix*, WI7001 and WI7002, using both Roche/454 GS FLX Titanium and Illumina HiSeq 2000 technologies (Table 1). The 454 pyrosequencing generated 651.8 and 418.0 Mb high-quality reads from WI7001 and WI7002, respectively. From approximately 252 million raw HiSeq 2000 reads, we obtained 5218 Mb WI7001 and 4432 Mb WI7002 sequences. Hybrid assembly of Illumina contig sequences and Roche/454 paired-end reads resulted in 11 649 scaffolds containing total 209 Mb sequences, with an N50 scaffold size of 50 831 bp and the largest scaffold of 560 870 bp (N50 is a weighted median statistic such that 50% of the entire assembly is contained in a scaffolds equal to or larger than this value) (Table 2).

To assess the quality of this *C. hystrix* draft assembly, the 40 largest scaffolds were BLAST-aligned against the

Gy14 cucumber and DHL92 melon genome scaffold assemblies (Table S1). Of the 40 scaffolds, 38 showed a high degree of sequence homology with the scaffolds of cucumber or melon or both. This *C. hystrix* draft genome assembly provides insights into micro-synteny among the *C. melo*, *C. hystrix* and *C. sativus* genomes. Of the approximately 12.2 Mb sequences in the 40 scaffolds, 11.5 Mb aligned to 13.1 Mb of melon scaffold sequences, indicating that the melon genome is approximately 13.9% larger than that of *C. hystrix* in these regions. Meanwhile, the syntenic regions in the cucumber genome in the 11.7 Mb *C. hystrix* scaffold regions were slightly longer (by 543 kb) (Table S1), suggesting that the *C. hystrix* genome may not be significantly larger than the cucumber genome.

Among the 40 *C. hystrix* scaffolds, ten had one melon/two cucumber or one cucumber/two melon alignment patterns (Table S1). The sequences of these ten *C. hystrix* scaffolds could be used to improve the cucumber or melon draft genome assemblies by estimating the gap size between two adjacent scaffolds or anchoring scaffolds that had not been placed to the cucumber or melon draft genome assembly. Conversely, the melon and cucumber scaffold assemblies can be used to anchor more *C. hystrix* scaffolds for draft genome assembly.

Development of an SSR-based *C. hystrix* genetic map

For genetic map construction, we screened 2826 cucumber and 1004 melon microsatellite markers (SSRs), and identified 185 (7%) and 63 (6%) polymorphic SSRs, respectively, between WI7001 and WI7002. Polymorphic SSRs were also identified *in silico* between the two *C. hystrix* contig assemblies. We first performed genome-wide identification of microsatellite sequences in WI7001 and WI7002 contig assemblies. In 128 257 and 117 711 SSRs identified, respectively, from WI7001 and WI7002, 373 were selected based on *in silico* polymorphism between the two parents and maximum sequence alignment with the Gy14 cucumber genome. The polymorphism of 312 (84%) SSRs was empirically validated. In contrast, among 89 randomly selected SSRs, only 28 (33%) were empirically polymorphic, suggesting that our high-throughput whole-genome screening strategy based on *in silico* polymorphisms was highly efficient with respect to time and cost for linkage map development.

Table 1 Summary of data for whole-genome sequencing of the *C. hystrix* genome

<i>C. hystrix</i> accessions	Next-generation sequencing method	Number of raw reads	Total length (bp)	Mean length (bp)	Total length of clean reads (Mbp)
WI7001	Illumina HiSeq 2000	135 870 754	9 782 694 288	72	5218
	454 GS FLX Titanium	2 034 004	651 804 005	320	652
WI7002	Illumina HiSeq 2000	117 104 464	8 431 521 408	72	4432
	454 GS FLX Titanium	1 192 482	418 001 825	351	418
Total		256 201 704	19 284 021 526		10 720

Table 2 Summary statistics of *C. hystrix* genome assembly

Assembly	Total number	Total bases (bp)	Mean size (bp)	Maximum size (bp)	Minimum size (bp)	N50
Scaffolds	11 649	209 044 452	17 945	560 870	1825	50 831
Large contigs	124 535	210 428 975	1689	21 932	500	2249
All contigs	200 296	232 662 438	1162	21 932	100	2012

The resulting *C. hystrix* genetic map contained 416 markers (see Table S2 for details), of which 215, 151 and 50 were derived from *C. hystrix*, cucumber and melon, respectively. Using a logarithm of odds (LOD) score of 7.0, the 416 markers were grouped into 12 linkage groups, designated H01–H12 based on their synteny with the respective melon chromosomes (I to XII). Key statistics for this genetic map are presented in Table 3 and Figure S1. The 1001.5 cM map length was comparable to two previous melon genetic maps (Diaz *et al.*, 2011; Li *et al.*, 2011a), suggesting near-complete coverage of the *C. hystrix* genome.

Syntenic and chromosome rearrangements between the *C. hystrix* and *C. melo* genomes

Primer sequences of the 416 markers on the *C. hystrix* genetic map were used as queries for *in silico* PCR or BLAST searches with the DHL92 melon assembly as the template, and 261 (75%) had *in silico* PCR amplicons or BLAST alignments covering approximately 294 Mb (93%) of the DHL92 melon genome. The alignment of the 261 shared markers between the 12 *C. hystrix* and corresponding melon chromosomes is shown in Figure 2(a). The physical locations of these markers in the DHL92 melon assembly are shown in Table S3.

Ten of the 12 *C. hystrix* chromosomes were highly syntenic with melon, thus retaining the ancestral condition. In *C. hystrix* chromosome H02, the region from 53.1–97.5 cM (12 marker loci) was syntenic to melon chromosome II,

whereas the 0–42.8 cM block (21 loci) was syntenic to approximately 7.2 Mb in the distal (lower) end of melon chromosome VIII (Table S3). Seven of the eight markers in H08 were shared with chromosome VIII, but one (NR39) was located in a syntenic region of melon chromosome II, indicating a reciprocal translocation between H02 and H08, as supported by sequence alignment of *C. hystrix* scaffolds anchored to H02 and H08 with the DHL92 assembly (Figure 3a).

Marker colinearity between the *C. hystrix* and *C. melo* genomes was inconsistent in many syntenic blocks (Table S3), which may be due to chromosome rearrangements between the two species or to errors in genetic mapping or whole-genome assembly. We performed large-scale comparative pachytene FISH mapping among *C. sativus*, *C. hystrix* and *C. melo* to assess the structural rearrangements. Molecular markers mapped to seven cucumber chromosomes were used to screen a cucumber fosmid library to identify positive clones as FISH probes. We identified 128 fosmid clones (Table S4), all of which were unambiguously mapped to the seven cucumber chromosomes. Among 122 clones tested in *C. hystrix*, all resulted in single and excellent FISH signals; of 119 tested in *C. melo*, all but three gave good FISH signals (see Figure S2 for examples). The results from comparative FISH mapping of 12 *C. hystrix* and melon chromosomes are shown in Figure 4, and revealed at least 14 inversions between the two genomes, five of which appear to be nested within

Table 3 Information for the *C. hystrix* linkage map and the syntenic relationships with cucumber and melon

Linkage group (chromosome)	Number of loci	Map length (cM)	Map interval (cM)	Number of markers anchored		Synteny to melon	Synteny to cucumber
				Cucumber	Melon		
H01	37	97.9	2.6	33	27	I	C7
H02	47	97.5	2.1	36	33	II + VIII	C1 + C6
H03	35	84.9	2.4	30	20	III	C2 + C6
H04	32	101.5	3.2	25	18	IV	C3
H05	47	80.4	1.7	37	34	V	C2 + C4 + C6
H06	36	112.4	3.1	35	20	VI	C3
H07	34	86.5	2.5	27	25	VII	C4
H08	24	56.1	2.3	17	8	II + VIII	C1 + C4 + C6
H09	27	71.9	2.7	24	22	IX	C5
H10	29	51.3	1.8	26	18	X	C5
H11	39	89.2	2.3	30	21	XI	C2 + C6
H12	30	71.9	2.4	28	16	XII	C1
Sum	417	1001.5	2.4	348	261		

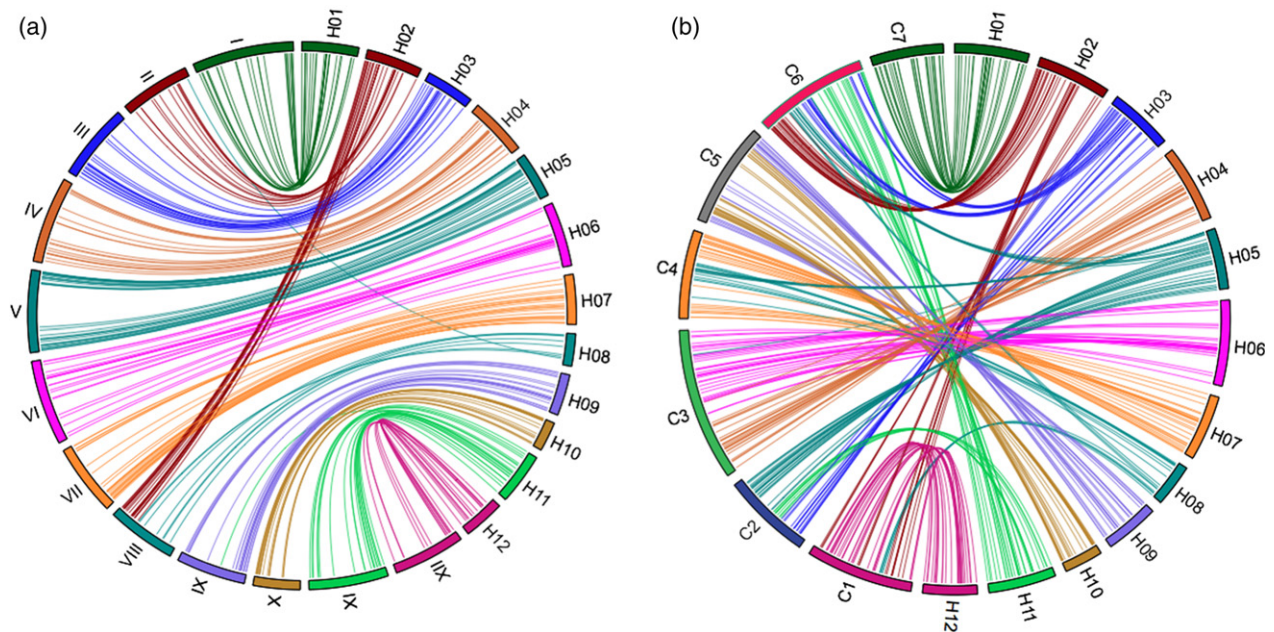


Figure 2. Syntenic relationships of *Cucumis hystris* (H01–H12) with melon (I–XII) (a) and cucumber (C1–C7) (b) chromosomes. Synteny between *C. hystris* and melon is based on alignment of 262 shared markers on the *C. hystris* genetic map and the DHL92 melon draft genome assembly (a); synteny between *C. hystris* and cucumber is based on alignment of 348 shared markers on the *C. hystris* genetic map and the Gy14 cucumber draft genome assembly. Syntenic blocks are connected by lines (shared markers) with the same color.

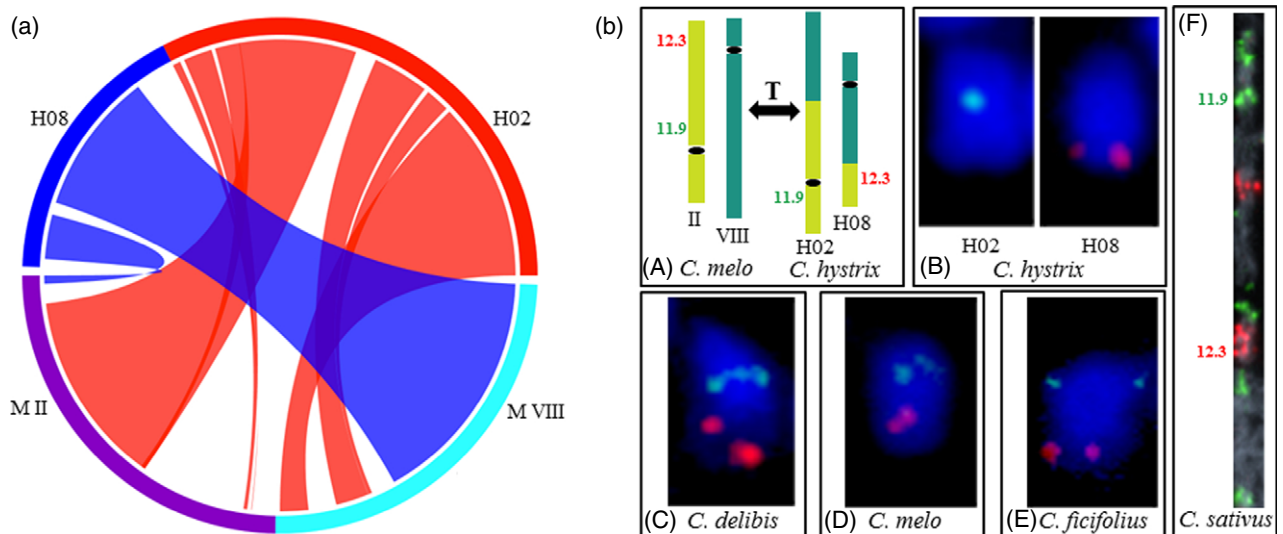


Figure 3. *Cucumis hystris*-specific reciprocal translocation.

(a) The H02–H08 translocation in relation to melon chromosome II (MII) and chromosome VIII (M VIII) revealed from sequence alignment of *C. hystris* scaffolds and the DHL92 melon genome assembly.

(b) Comparative genetic mapping between *C. hystris* and melon identified the H02–H08 translocation (A). Two fosmid clones 11.9 (green) and 12.3 (red) were used to probe mitotic (B–E) or meiotic pachytene (F) chromosomes of five *Cucumis* species (B–F).

additional inversions. All fosmid loci in chromosomes H09 and H10 were colinear with those in melon chromosomes IX and X, respectively, suggesting that the two chromosomes were conserved during evolution.

The H02–H08 reciprocal translocation between *C. hystris* and *C. melo* was cross-validated with comparative FISH in

five additional species representing the *Cucumis* phylogenetic tree, namely the Asian species *C. debilis*, the African species *C. metuliferus*, *C. ficifolius* and *C. zeyheri*, and the African/Pakistani species *C. prophetarum*, as well as *C. melo* and *C. sativus*. Two fosmid probes, 11.9 and 12.3, which mapped on cucumber chromosome C1 and melon

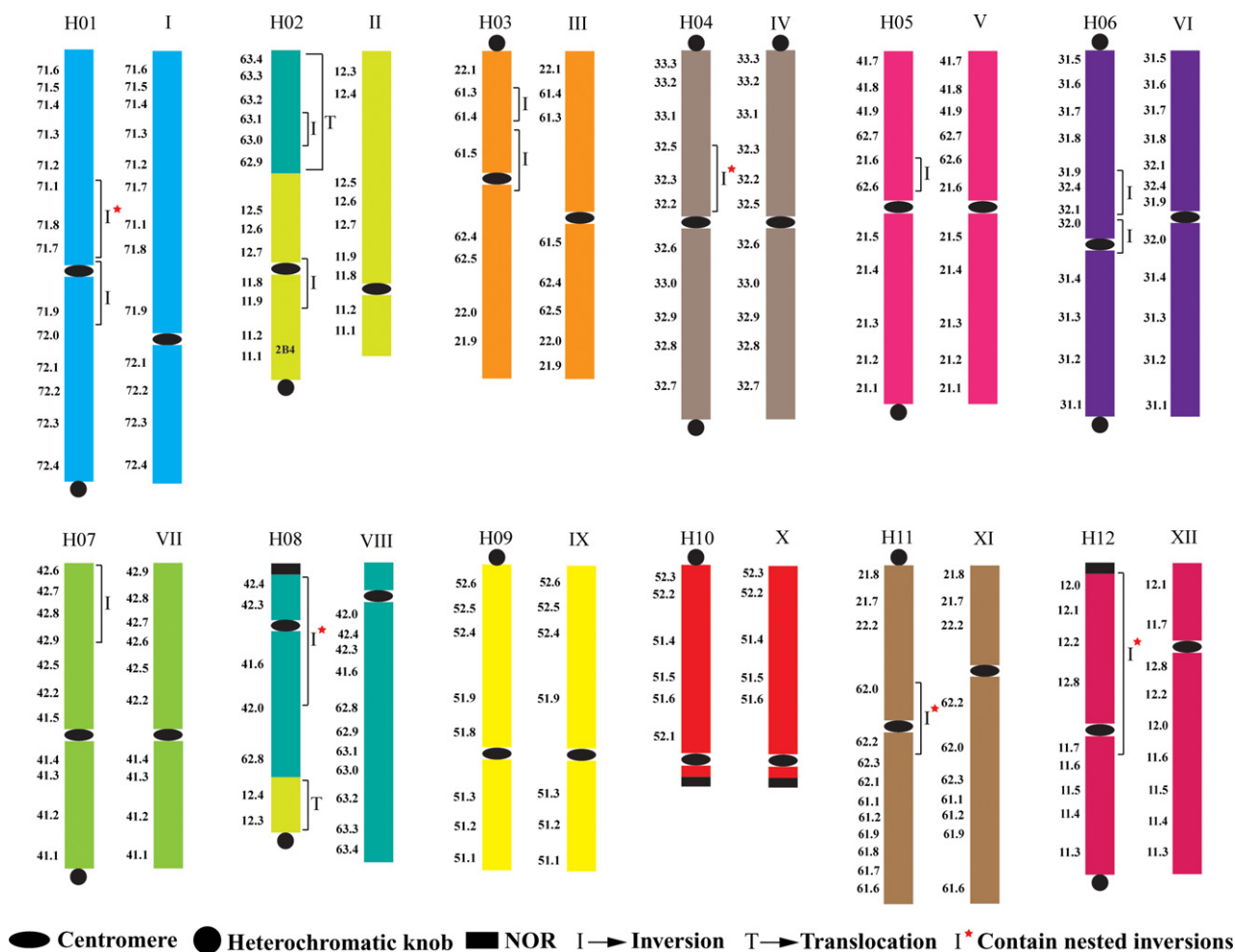


Figure 4. Syntenic relationships between *Cucumis hystris* and melon based on comparative genetic and pachytene FISH mapping. Syntenic chromosomes are arranged side by side. Black ovals indicate putative centromere locations.

chromosome II (Figure 4), were used in FISH of somatic chromosomes of these species. We found that both fosmid loci were located on the same chromosome in all species except *C. hystris* (Figure 3b), suggesting that this reciprocal translocation is specific to *C. hystris*. Both fosmid probes gave signals in cucumber C1, which appears to be the result of a fusion between *C. hystris* chromosomes H02/H08 and H12 (see below).

Syteny and chromosome rearrangements between the *C. hystris* and *C. sativus* genomes

Of the 416 markers on the *C. hystris* genetic map, 348 (84%) aligned to the *C. sativus* genome, covering 95% (181/191 Mb) of the Gy14 draft genome assembly. Based on the chromosome syteny inferred from this alignment (Figure 2b) and the physical locations of these markers in the Gy14 genome (Table S5), the syntenic relationships between *C. hystris* and cucumber chromosomes may be expressed as follows: H01 = C7, H02 = C1 + C6, H03 = C2 + C6, H04 = C3, H05 = C2 + C4 + C6, H06 = C3, H07 = C4,

H08 = C1 + C4 + C6, H09 = C5, H10 = C5, H11 = C2 + C6 and H12 = C1 (Table 3).

To characterize the syntenic relationships between the two species in more detail, we divided the 12 *C. hystris* chromosomes into 53 syntenic blocks. Each block was defined as a region on the *C. hystris* genetic map that aligned with a continuous stretch of DNA sequences in the Gy14 genome and was anchored by at least one shared marker or fosmid position. The orientation of each syntenic block in relation to the Gy14 genome assembly was determined by the order of shared markers on the *C. hystris* genetic map and in the Gy14 assembly, and further verified by pachytene FISH. Of the 53 syntenic blocks, 20 were colinear, and 25 contained inversions between *C. hystris* and cucumber. The orientation of eight blocks remains unknown (Table S5). The resulting accurate view of the seven cucumber chromosomes in terms of their syteny with *C. hystris* (Figure 5) may be simplified as follows: C1 = H02/H08 + H12, C2 = H03 + H05 + H11, C3 = H04 + H06, C4 = H7 + H8 + H05, C5 = H09 + H10, C6 = H03 +

H11 + H08/H02 + H05 and C7 = H01. Chromosome C7 shows complete synteny with H01, while C1, C3 and C5 each appear to be the result of fusion of two *C. hystrix* chromosomes, and C2, C4 and C6 each contain syntenic blocks from more than two *C. hystrix* chromosomes. The arrangement of *C. hystrix* syntenic blocks in all but the C7 cucumber chromosome is clearly far more evolutionarily complicated than a simple fusion of two *C. hystrix* chromosomes.

Syntenic and chromosome rearrangements between the *C. sativus* and *C. melo* genomes

To refine the cucumber/melon synteny, we divided the DHL92 melon genome assembly into 91 syntenic blocks (Table S6), each >500 kb in size and exhibiting sequence homology with the Gy14 cucumber draft genome. Among the 91 blocks, 35 were anchored with two or more fosmid loci, providing additional evidence on the orientation of these syntenic blocks. Of the 91 syntenic blocks, 45 had at least one inversion between cucumber and melon. Melon chromosome V is known to be syntenic to cucumber chromosome C2 (Li *et al.*, 2011a; Garcia-Mas *et al.*, 2012). Sequence alignment identified two new syntenic blocks of melon chromosome V in cucumber chromosomes C4 and C6 (Figure S3), which was confirmed with pachytene FISH (Figure S2). A melon syntenic block view of cucumber chromosomes C1–C7 is shown in Figure 5. Except for the 14 *C. hystrix*-specific inversions and the H02/H08 translocation, the arrangements of *C. melo* and *C. hystrix* syntenic blocks in the cucumber genome were essentially the same.

Conservation of cucumber chromosome C7 with *C. hystrix* and *C. melo*

Comparative mapping and sequence alignment studies revealed a one-to-one whole chromosome synteny of

cucumber chromosome C7 with melon chromosome I (M1) and *C. hystrix* H01. To verify this conserved synteny, we performed comparative FISH mapping using 14 fosmid probes located in C7 (71.1–72.4, Figure 5) in cultivated cucumber (*C. sativus* var. *sativus*), wild cucumber (*C. sativus* var. *hardwickii*, $2n = 2x = 14$, the progenitor species of cultivated cucumber; Yang *et al.*, 2012), *C. hystrix* (Figure 6a) and melon (*C. melo*). All 14 probes detected single hybridization signals in these species except for probe 72.0, which gave no signal in melon. The large paracentric inversion spanning six fosmid loci (71.1–71.6) in C7 is specific to *C. sativus* (Yang *et al.*, 2012). The short arms of wild cucumber, *C. hystrix* and *C. melo* (defined by fosmid loci 71.1–71.8) were differentiated by one or two inversions.

The six fosmid loci (71.9–72.4) on the long arm showed complete colinearity across the three species (Figure 6a), suggesting a high level of conservation of this region in *Cucumis*. The region comprised approximately 8.0 and 13.0 Mb in the Gy14 cucumber (Yang *et al.*, 2012) and DHL92 melon (Garcia-Mas *et al.*, 2012) draft genomes, respectively. To investigate the scope of conservation of this region in more distant cucurbit species, we extracted sequences in the syntenic block of this region from the watermelon draft genome (Guo *et al.*, 2013). This block is approximately 8.4 Mb in watermelon chromosome W2. Sequence alignment confirmed a high level of micro-synteny across the three species (Figure 6b and Figure S4). We also annotated genes, DNA transposons and retrotransposons of this region in the three assembled genomes as well as the *C. hystrix* scaffolds anchored to the region. Although 70% of genes in this region were conserved among the three species, the number of transposons in melon compared with cucumber and watermelon has increased 3.6 or 2.5 times, respectively (Table 4). Consistent with this, the gene density in melon was lower than

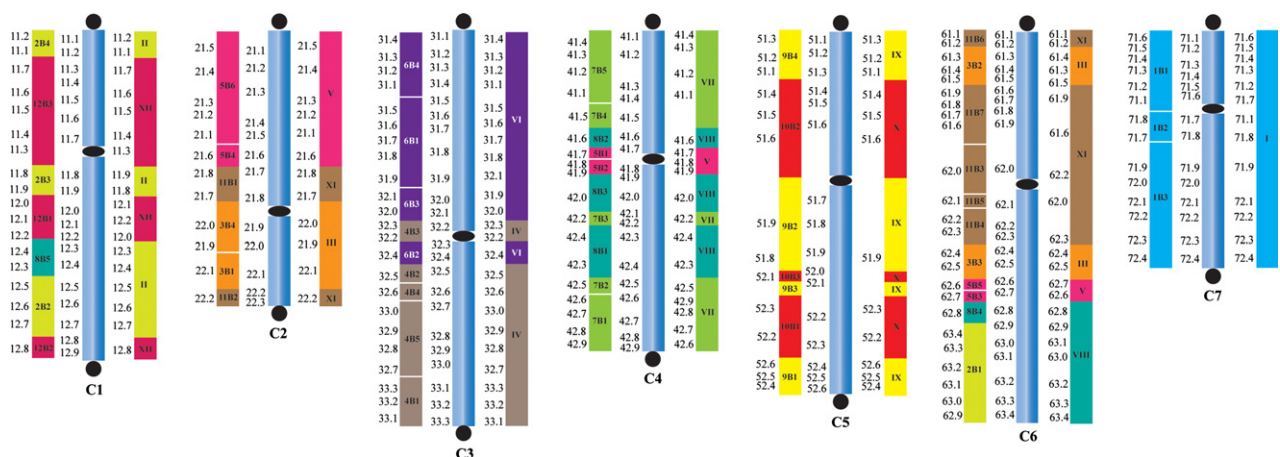


Figure 5. *Cucumis hystrix* (left) and melon (right) syntenic block views of cucumber chromosomes C1–C7 (center). Syntenic blocks that originated from the same *C. hystrix* chromosomes (1–12) or melon chromosomes (I–XII) have the same color. *C. hystrix* syntenic block boundaries are indicated by white lines or centromeres, and anchored by fosmid loci. The numbers to the left of each syntenic block are fosmid clones.

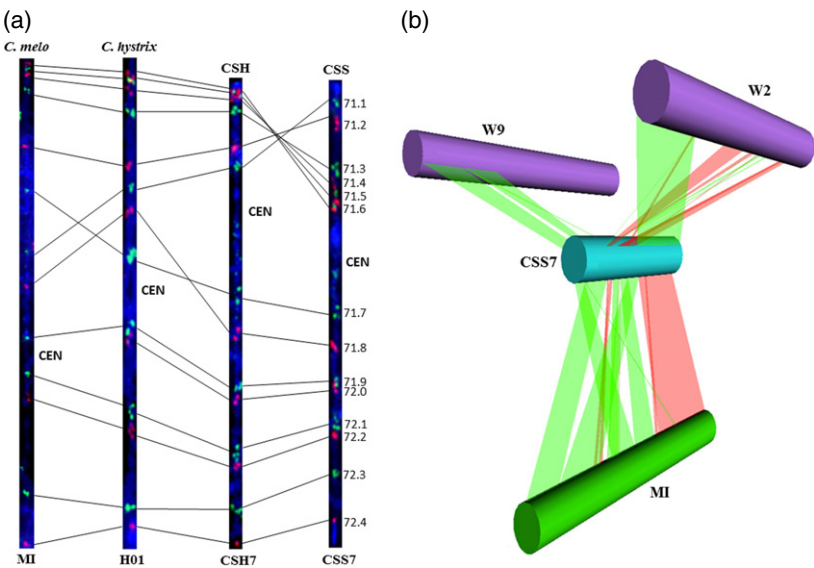


Figure 6. Synteny of cucumber chromosome 7 (C7) with other cucurbit species. (a) Pachytene FISH analysis was performed with 14 fosmid probes in three *Cucumis* species (CSS, *C. sativus* var. *sativus*; CSH, *C. sativus* var. *hardwickii*). CEN indicates the putative centromere location. (b) Alignment of draft genome sequences of C7, melon chromosome I and watermelon chromosomes W2 and W9 showing a high degree of sequence homology in this region.

that in cucumber, *C. hystris* or watermelon. However, the distribution of mobile elements across the region was not uniform. While the distal telomeric region appeared to be conserved, with approximately the same size in the three genomes, there was notable expansion in the proximal centromeric region in melon, with a higher frequency of mobile elements (Figure S4). In watermelon, distribution of genes and transposons or retrotransposons was relatively uniform across this region, while the highly conserved regions were gene-dense in both melon and cucumber, with the region corresponding to the expansion in melon also being transposon-dense (Figure S4).

DISCUSSION

Major chromosome rearrangements among *C. melo*, *C. hystris* and *C. sativus*

The present study highlights the synergy from integrated use of high-throughput genome sequencing and assembly, comparative genetic and physical mapping, and whole-genome alignment to understand karyotype evolution in non-model plant species (Rocchi *et al.*, 2006; Schubert and Lysak, 2011). In particular, using next-generation sequencing data, we developed a quick method of *in silico* polymorphism screening for development of a genetic map in *C. hystris*, for which no available genetic or genomics resource was available. This cost-effective and highly efficient approach should have wide applicability in genetic mapping studies. Comparative genetic and physical mapping revealed a high level of synteny among chromosomes of cucumber, *C. hystris* and melon. However, structural rearrangements appear to be common within syntenic blocks among the three species, with inversions as the predominant event.

Table 4 Summary of genes, DNA transposons and retotransposons annotated in a highly conserved syntenic region across four cucurbit species

	Melon	Cucumber	<i>C. hystris</i>	Watermelon
Length of syntenic block (Mbp)	13.0	8.0	1.1	8.4
Genes	1083	843	105	898
Mean (number of genes/Mb)	83.3	105.4	95.5	106.9
Mean (kb/gene)	12.0	9.5	10.5	9.5
DNA transposons				
CACTA	141	9	–	0
MULE	172	23	–	5
RIF	52	4	–	1
hAT	10	3	–	0
Unclassified	208	139	–	196
Total	583	178	–	202
Mean (number/Mb)	44.8	22.3	–	24.0
Retrotransposons				
gypsy	590	79	–	35
copia	403	56	–	57
Unclassified	245	188	–	418
Total	1238	323	–	510
Mean (number/Mb)	95.2	40.4	–	60.7
Grand total	1821	501	–	712
Overall mean (number/Mb)	140.1	62.6	–	84.8

– indicates that no estimation was made because the contiguous *C. hystris* genome sequences were too short.

For example, whole-genome pachytene FISH revealed 14 inversions between *C. hystris* and melon (Figure 4), and 20 of the 53 syntenic blocks (47%) between *C. hystris* and cucumber also contained inversions (Table S5). We previously identified six inversions differentiating wild

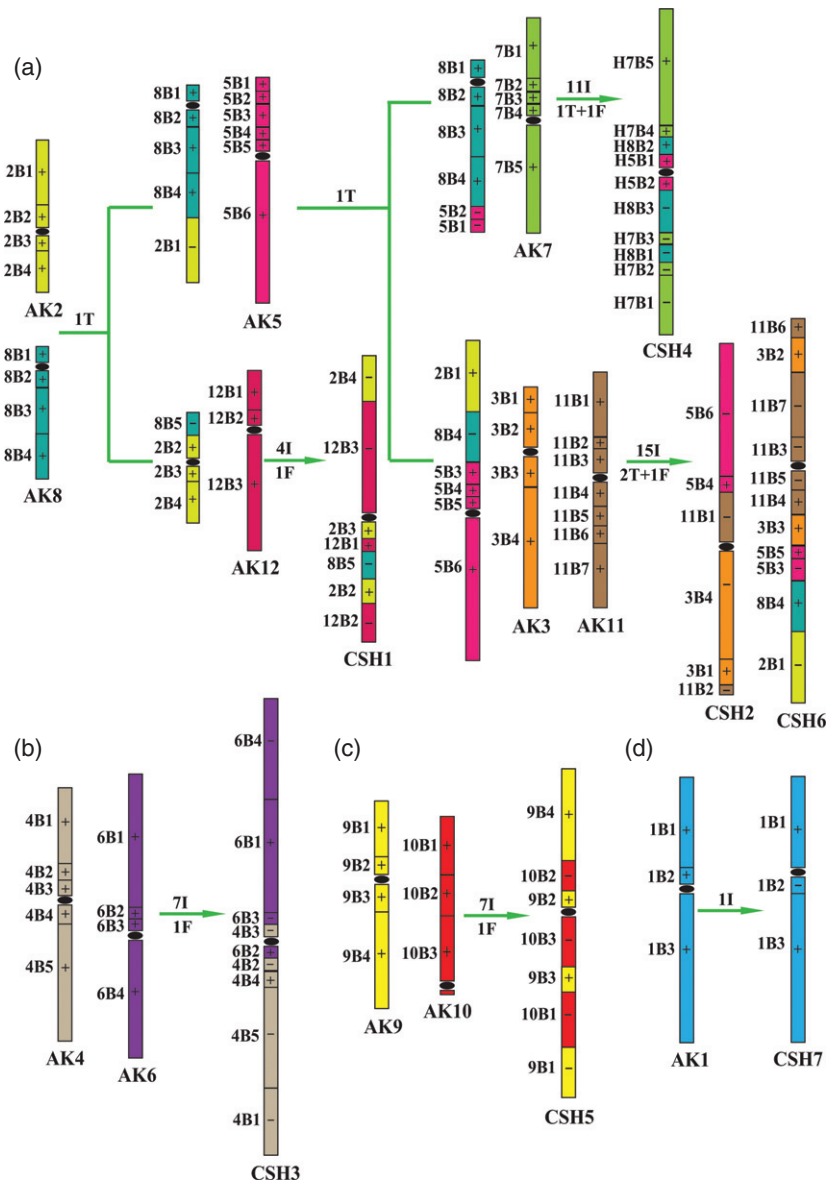
(*C. sativus* var. *hardwickii*) and cultivated (*C. sativus* var. *sativus*) cucumbers (Yang *et al.*, 2012), and these appear to correspond well with *C. hystrix* syntenic blocks 7B1 and 7B5 (C4), 9B2, 9B4 and 10B2 (C5), and 1B1 (C7) in the present study (Figure 5). Comparison of the orientations of these six syntenic blocks in *C. melo*, *C. hystrix* and *C. sativus* suggests that four inversions, 7B5 in C4, 9B2 and 9B4 in C5, and 1B1 in C7, may have occurred during domestication of cucumber (Figure S5). Furthermore, genome alignment and comparative FISH identified 45 of the 91 melon syntenic blocks (50%) as having at least one inversion in each syntenic block compared with cucumber, covering almost half of the cucumber and melon genomes (Table S6). The predominance of inversions found here matches findings from Rosaceae (Vilanova *et al.*, 2008), Solanaceae

(Livingstone *et al.*, 1999; Wu and Tanksley, 2010) and Brassicaceae (Lagercrantz, 1998; Yogeewaran *et al.*, 2005). Inversions are believed to play an important role in karyotype evolution, speciation and local adaptation by reducing recombination and thereby protecting genomic regions from introgression (Hoffmann and Rieseberg, 2008; Kirkpatrick, 2010; Lowry and Willis, 2010; Ruiz-Herrera *et al.*, 2012).

We also identified a reciprocal translocation involving *C. hystrix* chromosomes H02–H08 that sets this species and *C. sativus* apart from the other species analyzed (Figure 3). Comparative FISH in *C. debilis*, *C. prophetarum*, *C. metuliferus*, *C. zeyheri* and *C. ficifolius* suggests that this translocation may have arisen in the common ancestor of *C. hystrix* and *C. sativus* (Figure 1).

Figure 7. Reconstruction of karyotype evolution from the $n = 12$ ancestor to the $n = 7$ wild cucumber (CSH).

The 12 ancestral chromosomes (AK1-AK12) are divided into 53 syntenic blocks that are the same as in *Cucumis hystrix*. The origin of each syntenic block may be tracked by its name and color. T, translocation; I, inversion; F, fusion. The + or - symbol within each block represents its orientation (+, colinear; -, inversion) between *C. hystrix* and cucumber. Black ovals indicate centromeres as inferred from pachytene FISH.



Retention of cucumber chromosome C7 during *Cucumis* evolution

We confirmed the conservation of cucumber chromosome C7 with melon chromosome I and extended it to *C. hystrix* H01 (Figure 6a). Chromosome C7 is also syntenic to chromosomes W2 and W9 of watermelon (Huang *et al.*, 2009; Guo *et al.*, 2013) (Figure 6b). However, we found that the degree of conservation varied in the long and short arms of C7 among *C. sativus*, *C. hystrix*, *C. melo* and watermelon. Their short arms were differentiated by one or two inversions, probably due to their high content of heterochromatin as indicated by DAPI-stained dark blue blocks in pachytene chromosomes (Figure 6a). By contrast, the long arms in these species showed complete colinearity that spanned 8.0, 8.4 and 13.0 Mb in the cucumber, watermelon, and melon draft genomes, with high levels of microsynteny and colinearity of genes despite the genome size differences (Figure S4 and Table 4). In the two *Cucumis* species, genes and transposable elements were clustered, with higher transposable element density in the pericentromeric region and higher gene density in the distal telomeric region; by contrast, in watermelon, the distribution of transposable elements and genes was relatively uniform (Figure S4).

Inferred mechanisms of dysploid reduction from $n = 12$ to $n = 7$ in *Cucumis*

The evolution from an ancestral karyotype may be modeled through synteny-based comparative analysis in extant species (Abrouk *et al.*, 2010; Schubert and Lysak, 2011; Salse, 2012). Geographically, the range of *C. hystrix* [found in Myanmar (Burma), North and West Thailand and Southwest China] overlaps that of wild cucumber (*C. sativus* var. *hardwickii*) (Sebastian *et al.*, 2010), and phylogenetic studies have shown that *C. hystrix* is the sister species to *C. sativus* (Ghebretinsae *et al.*, 2007; Renner *et al.*, 2007; Sebastian *et al.*, 2010). It is reasonable to assume that the common ancestor of *C. sativus* and *C. hystrix* had 12 chromosomes, AK1–12, with a structure that is similar, but not identical, to chromosomes H01–H12 of *C. hystrix*. Therefore, we modeled the ancestral karyotype of cucumber using both *C. melo* and *C. hystrix*, and reconstructed the history of karyotype changes from this $n = 12$ ancestor to $n = 7$ in cucumber. The main events during this process are summarized in Figure 7, and details are presented in Figure S5.

Cucumber chromosome C7

C7 is highly conserved with chromosome I of melon and H02 of *C. hystrix* (Figure 6). C7 differentiated from AK1 by only one pericentric inversion in syntenic block 1B2. Another paracentric inversion in syntenic block 1B1 occurred during cucumber domestication (Figure S5A).

Cucumber chromosomes C1, C3 and C5

A common theme in the evolution of these three cucumber chromosomes is that each derives from the fusion of two ancestral chromosomes. For C1, this process was probably initiated by a reciprocal translocation between AK2 and AK8 producing two translocated chromosomes (AK2/AK8 and AK8/AK2). Then, AK12 was inserted into the centromeric region of AK2/AK8 followed by two pericentric and two paracentric inversions to form chromosome C1 (Figure S5B). For C3, a pericentric inversion occurred in both AK4 and AK6, creating one telocentric and one acrocentric chromosome, which then underwent a fusion event (Robertsonian-like translocation). This translocation generated a product consisting of telomeric and AK6 centromeric fragments, which was probably unstable during meiosis and therefore eliminated. The fused chromosome went through two para- and two pericentric inversions leading to C3 (Figure S5C), which is the longest cucumber chromosome. Similar to C3, the origin of C5 involved fusion of AK9 and AK10, followed by nine inversions. C5 of cultivated and wild cucumbers then evolved via three more inversions during domestication (Figure S5D).

Cucumber chromosomes C2, C4 and C6

The evolution of these three chromosomes appears to be more complex because each contains blocks that are syntenic to at least three ancestral chromosomes. Evolution may have started with a reciprocal translocation between AK5 and AK8/AK2, resulting in two intermediates (AK8/AK5 and AK5/AK8/AK2) (Figure S5E). Two translocations between AK7 and AK8/AK5 and a series of inversions created C04 of wild cucumber (Figure S5F). After one paracentric inversion, a reciprocal translocation between AK5/AK8/AK2 and AK11 may have resulted in two new intermediates, each of which recombined with a segment from AK3 leading to C02 (by translocation) and C06, respectively (Figure S5E).

This model of dysploid reduction explains the location and orientation of the 53 *C. hystrix* syntenic blocks in the seven cucumber chromosomes identified herein (Figure 5). The 59 chromosome rearrangement events included five fusions, four translocations and 50 inversions (Table 5). Assuming all fusion and translocations were reciprocal translocations, up to 114 breakpoints may explain the 59 rearrangement events. The majority of these breakpoints were presumably located in heterochromatic regions including centromeric, telomeric, pericentromeric or nucleolus organizer (NOR) regions. Indeed, almost half of the breakpoints (53/114) involve centromeric or telomeric positions (Table 5) that are rich in repetitive sequences (e.g. Figure 6 and Figure S2). For example, *C. hystrix* chromosomes H08 (AK8) and H10 (AK10) are acrocentric, and H08, H10 and H12 (AK12) have NORs in the telomeric ends

(Figure 4). It makes sense, therefore, that AK08, AK10 and AK12 were involved in three of the five fusion events forming cucumber chromosomes C01, C05 and C04 (Figure S5B,D,F).

The mechanism of formation of cucumber C1 through nested chromosome fusion revealed here resembles the situation in some grasses (Luo *et al.*, 2009), in which centromeric or telomeric illegitimate recombination between non-homoeologous chromosomes has been demonstrated (Murat *et al.*, 2010). In *Arabidopsis*, chromosome rearrangements leading to descending dysploidy from an ancestral $n = 8$ species to *A. thaliana* ($n = 5$) also involved inversions, fusions and translocations (Lysak *et al.*, 2006). Given the constraints on chromosome structure, most or all chromosome fusions are probably mediated by reciprocal translocations with or without preceding para- and pericentric inversions, resulting in a fusion chromosome and a meiotically unstable (a)centric fragment that is eliminated (Schubert and Lysak, 2011). Therefore, in addition to the expansion of intergenic regions and proliferation of transposable elements in the melon genome (Gonzalez *et al.*, 2010; Garcia-Mas *et al.*, 2012), the size difference between the cucumber and melon genomes may be partly explained by loss of these (a)centric fragments.

The centromere is the point of spindle attachment in cell division, and is essential for faithful segregation of chromosomes. The insertion of an ancestral chromosome AK12 into the centromeric region of translocated AK2/AK8 (Figure S5B) probably involved inactivation of the AK8 centromere, and the eight additional translocations or fusions inferred here probably led to loss of centromeres from AK2, AK3, AK6, AK7 (or AK8) and AK10, with cucumber chromosomes C1–C7 instead inheriting their centromeres from the remaining seven ancestral chromosomes

(Table 5). Such weak constraints on the centromeres fit with speculations that centromeres are activated and maintained primarily by epigenetic mechanisms (Karpen and Allshire, 1997). Cucumber chromosome C7 has remained largely intact in the evolution of *Cucumis* (Figure 6). The centromere positions in C7, melon chromosome M1 and *C. hystrix* H01 (Figure 6) appear to be consistent with a centromere repositioning event as proposed by Han *et al.* (2009). On the other hand, the different centromere positions between cucumber chromosome C6 and melon chromosome VIII (Mel in Han *et al.*, 2009) may be the result of multiple rearrangements that may have occurred during the complex evolutionary history of C6 (Figure S5E).

Among the three *Cucumis* species, *C. hystrix* and melon diverged approximately 10 million years ago, and are differentiated by 14 inversions and a reciprocal translocation (Figure 4); the subsequent dysploid reduction from $x = 12$ to $x = 7$ in the ancestor of *C. sativus* then occurred at some time after that species diverged from *C. hystrix* approximately 4.6 million years ago (Sebastian *et al.*, 2010). The factors that caused the accelerated karyotype change in the *C. sativus* lineage remain unknown. It would also be interesting to know why cucumber C7 remained largely unchanged during at least 12 million years ago of evolution. However, the syntenic relationships revealed here provide a fresh perspective in understanding chromosome evolution in *Cucumis*, and the newly sequenced genome of *C. hystrix* adds further information on a wild relative of cucumber.

EXPERIMENTAL PROCEDURES

Plant materials

An F_2 mapping population of 91 F_2 plants was developed for construction of a linkage map in *C. hystrix*, which was derived from a

Table 5 Summary of cytogenetic changes during dysploid reduction from the $n = 12$ ancestor to $n = 7$ wild cucumber

Cucumber chromosome	Ancestor donor chromosome ^a	Number of breakpoints ^c	Number of translocations ^d	Number of inversions	Number of fusions	AK centromere inherited by cucumber chromosome
	t(AK2/AK8) ^b	2 (0)	1	0		
C1	t(AK2/AK8), AK12	11 (6)	0	4	1	AK12
C3	AK4, AK6	15 (8)	0	7	1	AK4
C5	AK9, AK10	19 (7)	0	10	1	AK9
C7	AK1	2 (1)	0	1	0	AK1
C2	AK3, AK5, AK11	65 (31)	3	28	0	AK5
C4	AK5, AK7, t(AK8/AK2)				1	AK7 or AK8
C6	AK3, AK5, t(AK2/AK8), AK11				1	AK11
Sum		114 (53)	4	50	5	

^aThe ancestral karyotype is similar to or the same as the *C. hystrix* karyotype. AK1–AK12 correspond to H01–H12, respectively, in terms of chromosome structure and organization.

^bt(AK2/AK8) = result of the chromosome translocation between AK2 and AK8 that preceded all other events.

^cBreakpoint counts assume all that fusions are due to reciprocal translocations. The numbers in parentheses indicate breakpoints that occurred in centromeric/telomeric regions.

^dReciprocal translocations excluding those mediating chromosome fusions.

cross between two accessions WI7001 and WI7002 originally collected from Thailand and Southwest China, respectively. Additional species in the genus *Cucumis* used in comparative cytological analysis included inbred or plant introduction (PI) lines of cultivated cucumber (*C. sativus* var. *sativus* L., PI 249561), wild cucumber (*C. sativus* var. *hardwickii* Alef., PI 183967), melon (*C. melo* L., Top Mark), *C. debilis* W.J. de Wilde & Duyfjes, *C. prophetarum* L. (PI 193967), *C. metuliferus* E. Mey. ex Naudin (PI 482443), *C. zeyheri* Sond. (PI 364472) and *C. ficifolius* A. Rich. (PI 299570). Seeds of all PI lines were obtained from the US National Plant Germplasm System (Ames, IA, USA).

C. hystrix whole-genome sequencing and assembly

Sequencing of the genomes of WI7001 and WI7002 using both the Roche/454 (Branford, CT, USA) and Illumina systems (San Diego, CA, USA) was performed in the Biotechnology Center of the University of Wisconsin at Madison. For the 454 GS FLX Titanium pyrosequencing, 8 kb paired-end libraries were used; whole-genome shotgun sequencing was performed on the Illumina's HiSeq 2000 system. For the HiSeq 2000 raw reads, sequences with quality scores <28 and length <41 were removed to obtain high-quality reads. A hybrid assembly strategy was used to integrate all sequences from both platforms. First, the trimmed Illumina reads of WI7001 and WI7002 were assembled into contigs using the ABySS assembler (version 1.3.0) (Simpson *et al.*, 2009) with *k*-mer size = 51. Then, for all contigs larger than 1 kb in length, each contig was chopped into 1 kb fragments, with neighboring ones having an 800 bp overlap. These artificial reads were combined with approximately 3.2 million 454 pair-end reads using Newbler 2.6 in heterozygotic mode (Margulies *et al.*, 2005), with a minimum overlap length of 50 bp and a minimum overlap identity of 95%. To assess the quality of the resulting genome assembly, the largest 40 *C. hystrix* scaffolds were aligned to the Gy14 cucumber and DHL92 melon draft genome assemblies using BLASTN (Altschul *et al.*, 1990) (cut-off *E* value $\leq 10^{-100}$). Visual alignment of these scaffolds with the cucumber and melon draft genomes was performed using the progressiveMauve procedure in MAUVE version 2.3.1 (Darling *et al.*, 2004).

Development of molecular markers and linkage map construction in *C. hystrix*

To construct a *C. hystrix* genetic map, we first performed polymorphism screening between WI7001 and WI7002 using 2826 cucumber and 1004 melon microsatellite (SSR) markers (Yang *et al.*, 2012). We also performed *in silico* identification of polymorphic SSRs between the two parental lines from the Illumina genome sequences. First, all microsatellites with 2–8 bp repeat motifs were identified from both WI7001 and WI7002 HiSeq 2000 contig assemblies using MISA (Thiel *et al.*, 2003). As SSRs with dinucleotide motifs and longer repeats of motifs tend to be more polymorphic (Cavagnaro *et al.*, 2010), a subset of SSRs was selected from these identified SSRs, and primer pairs flanking these SSRs were designed using PRIMER3 software (<http://frodo.wi.mit.edu/>). Next, virtual PCR was performed using WI7001-derived SSRs with the WI7002 contig assembly as the template, or *vice versa*, resulting in SSRs that were polymorphic between the two parental genomes. To facilitate comparative analysis, these selected SSRs were subjected to *in silico* PCR using the Gy14 cucumber draft genome assembly as the template. Finally, 373 SSRs (prefixed with CHXSSR in marker names) that were polymorphic between WI7001 and WI7002 and had amplicons in the cucumber genome were selected by empirical polymorphism screening and linkage mapping in the *C. hystrix* F₂ population. However, several regions

on the *C. hystrix* map lacked markers to detect or anchor syntenic blocks in the cucumber or melon draft genome. To fill these gaps, *C. hystrix* scaffolds or contigs that were orthologous to those melon or cucumber syntenic regions were identified by BLAST sequence alignment. SSRs were selected in these *C. hystrix* regions for *in silico* PCR between the WI7001 and WI7002 contig assemblies to assure that there were PCR products in both parents. Finally, 89 such SSRs were identified for empirical polymorphism screening.

For marker analysis, unexpanded young leaves from embryo culture-generated seedling plants were collected for DNA extraction using the CTAB (cetyl trimethylammonium bromide) method (Murray and Thompson, 1980). The PCR procedure and electrophoresis of the PCR products were performed as described by Yang *et al.* (2012). Linkage analysis was performed using JOINMAP 3.0 software (van Ooijen and Voorrips, 2001). Linkage groups were determined using a minimum LOD score of 7.0. Genetic distance was calculated using the Kosambi mapping function (Kosambi, 1944). H01–H12 were assigned as the 12 *C. hystrix* linkage groups and chromosomes that were syntenic to melon chromosomes I–XII, respectively.

Establishment of chromosome synteny among the *C. hystrix*, *C. sativus* and *C. melo* genomes

The *C. hystrix* genetic map was used to infer syntenic relationships of *C. hystrix* with cucumber and melon chromosomes as described by Li *et al.* (2011b). Briefly, for each marker mapped on the *C. hystrix* linkage map, *in silico* PCR was implemented using the Gy14 cucumber and DHL92 melon draft genome assemblies as templates to align this marker and associated scaffold or contig on the respective cucumber and melon chromosome. This was performed using a custom Perl script that used the National Center for Biotechnology Information NCBI, (<http://www.ncbi.nlm.nih.gov/>) BLASTN program as the search engine (Cavagnaro *et al.*, 2010). The order and physical locations of shared markers were used to detect macrosynteny and rearrangements among chromosomes. The resulting syntenic relationships were visualized using the CIRCOs program (Krzywinski *et al.*, 2009). For comparison of the synteny between *C. hystrix* and cucumber chromosomes, each *C. hystrix* chromosome was divided into 3–7 syntenic blocks, each of which was a discrete linkage block on the *C. hystrix* genetic map occupied by several marker loci that had continuous DNA sequence matches in the Gy14 cucumber genome. These syntenic regions were the building blocks for inferring the evolutionary history of dysploid reduction from the *n* = 12 ancestor to *n* = 7 cucumber genomes.

The Gy14 cucumber and DHL92 melon draft genome sequences were used to refine the syntenic relationships identified from comparative genetic mapping. Whole-genome comparison was performed with SYMAP version 3.5 (Soderlund *et al.*, 2011) using the *nucmer* algorithm of the MUMER package (<http://mummer.sourceforge.net/>) with parameter identity $\geq 90\%$ and match length ≥ 1 kb. For simplicity, the melon genome assembly was divided into 91 syntenic blocks, where each block was defined as a region of ≥ 500 kb with continuous sequence matches in the Gy14 cucumber genome, without considering small gaps or micro-inversions that were common in genome sequence alignment.

Micro-synteny in a syntenic region across cucumber, melon and watermelon genomes

Our data suggested that cucumber chromosome C7 was largely conserved during chromosome evolution in several Cucurbitaceae species, including *C. sativus*, *C. hystrix*, *C. melo* and watermelon.

To obtain insights into micro-synteny at the DNA sequence level, we performed detailed comparative analysis in a highly conserved region across the four species. This region in the long arm of C7 spanned 8 Mb (11–19 Mb) in the Gy14 draft genome, and the corresponding syntenic region in the DHL92 melon chromosome I (MI) comprised 13 Mbp (22–35 Mb in version cm3.5) (Garcia-Mas *et al.*, 2012), and that in watermelon chromosome 9 (W9) comprised 8.4 Mbp (0–8.4 Mb) (Guo *et al.*, 2013). Four scaffolds of the *C. hystris* assembly of approximately 1.5 Mb were anchored to this syntenic block, which included scaffold00001, scaffold00005, scaffold00011 and scaffold00041. As the four *C. hystris* scaffolds only accounted for approximately 12.5% of the 8 Mb cucumber syntenic region of cucumber (excluding Ns from paired-end sequencing), the four *C. hystris* scaffolds were merged with a 100 kb spacer of Ns to avoid syntenic redundancy. The 8.4 Mbp watermelon sequences were reverse-complemented for easy plotting of the syntenic blocks.

We first aligned the sequences of this syntenic block from the four species using *MUSCLE* version 3 (Edgar, 2004) and *PROGRESSIVE-MULVE* using the non-colinear option (Darling *et al.*, 2010). We next performed gene annotation for these sequences. Genome annotations for melon and watermelon in this region were downloaded directly from www.icugi.org/ (watermelon) and www.melonomics.net (melon), respectively. The four *C. hystris* scaffolds and the cucumber genomic region were annotated using the Eugene tool (Foissac *et al.*, 2008) with *Arabidopsis thaliana* as a training set. We further performed transposon annotation for sequences of this region in cucumber, melon and watermelon. The transposon annotations for the melon and cucumber genome sequences were performed as described by Garcia-Mas *et al.* (2012). The *LTR_FINDER* program (Xu and Wang, 2007) was used to identify LTR retrotransposons *de novo* in the watermelon genome. The predicted elements were clustered (complete linkage, 80% similarity along 50% of sequence length), and a representative was taken from each cluster. The resulting set of sequences were further classified as *gypsy* or *copa* according to their similarity (tblastx, >80% identity along >200 bp) (Altschul *et al.*, 1990) to known LTR retrotransposons in Repbase (<http://www.girinst.org/repbase>). Copies of these elements in watermelon sequences were estimated using a copy finding program based on MITE-HUNTER (Han and Wessler, 2010) and modified as described by Garcia-Mas *et al.* (2012), which annotates similar sequences and joins fragmented copies. As this procedure may fail to identify some ancient copies, we performed BLAST (tblastn, $e > 1e^{-10}$) using transposase and reverse transcriptase protein queries to identify DNA transposon- and retrotransposon-related sequences, respectively. There were therefore two levels of resolution of the transposon annotation: copies that were classified into super-families and families, and copies that were identified as transposon-related or retrotransposon-related.

Comparative fluorescence *in situ* hybridization (FISH) for cytological investigation of chromosome rearrangements

We performed large-scale comparative FISH using dual color-labeled (red and green) fosmid probes to investigate genomic rearrangements in cucumber, melon and *C. hystris*. We first anchored fosmid loci onto the seven cucumber chromosomes. In a previous study, we identified 76 fosmid clones that detect chromosome differentiation between cultivated and wild cucumber chromosomes (Yang *et al.*, 2012), all of which were used in the present study. Furthermore, 207 SSR markers at strategic map locations or target syntenic blocks were selected to screen the Straight 8 cucumber fosmid library (Yang *et al.*, 2012), and 52 additional fosmid clones were identified. Fifty-two of the 53

C. hystris syntenic blocks were anchored by at least one fosmid clone, which greatly increased the accuracy of alignment of syntenic blocks among the three species.

All FISH experiments were performed on meiotic pachytene chromosomes that were prepared from pollen mother cells, except for screening of the H02–H08 reciprocal translocation in several *Cucumis* species, for which mitotic metaphase chromosomes prepared from root tips cells were used. The karyotype of the cucumber chromosomes was first established using centromere- and telomere-specific repeat probes (Yang *et al.*, 2012). Then all fosmid clones were anchored to the seven cucumber chromosomes. The physical order of adjacent fosmid clones in each chromosome was determined by two-color FISH. To improve the efficiency of performing large-scale comparative FISH, two strategies were used in the present study: multi-fosmid cocktail for probe preparation and multi-probing of the same pachytene chromosome preparation (up to eight times) (Cheng *et al.*, 2001). The FISH procedure was performed as described by Koo *et al.* (2010).

ACKNOWLEDGEMENTS

This research was supported by US Department of Agriculture Current Research Information System Project 3655-21000-048-00D and a US Department of Agriculture Specialty Crop Research Initiative grant (project number 2011-51181-30661) to Y.W.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Graphic view of the *C. hystris* genetic map developed using the W17001 × W17002 F₂ mapping population.

Figure S2. Examples of comparative pachytene FISH among cucumber, *C. hystris* and melon.

Figure S3. Sequence alignment between Gy14 cucumber and DHL92 melon draft genomes revealed syntenic blocks of melon chromosome V with three regions in cucumber chromosomes C2, C4 and C6.

Figure S4. Comparative analysis in an 8 Mb region in the long arm of cucumber chromosome C7 and its syntenic blocks in melon (13 Mbp) and watermelon (8.4 Mbp).

Figure S5. Reconstruction of karyotype evolution history of seven wild (CSH) or cultivated (CSS) cucumber chromosomes (C1–C7) from 12 ancestral chromosomes (AK1–AK12).

Table S1. Alignment of the 40 largest *C. hystris* scaffolds with the Gy14 cucumber and DHL92 melon draft genome scaffolds.

Table S2. Summary statistics of the *C. hystris* genetic map with 416 SSR marker loci.

Table S3. Physical locations in the DHL92 melon draft genome assembly of molecular markers placed on the *C. hystris* genetic map.

Table S4. Information on cucumber fosmid clones used for FISH in this study.

Table S5. Syntenic relationships of *C. hystris* and cucumber genomes revealed by genome alignment, and comparative genetic and FISH mapping.

Table S6. Syntenic blocks (>500 kb) between melon and cucumber genomes based on whole-genome sequence alignment.

REFERENCES

- Abrouk, M., Murat, F., Pont, C. *et al.* (2010) Palaeogenomics of plants: synteny-based modelling of extinct ancestors. *Trends Plant Sci.*, **15**, 479–487.

- 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.
- Bhaduri, P.N. and Bose, P.C. (1947) Cytogenetical investigations in some common cucurbits, with special reference to fragmentation of chromosomes as physical basis of speciation. *J. Genet.* **48**, 237–256.
- Cavagnaro, P.F., Senalik, D.A., Yang, L., Simon, P.W., Harkins, T.T., Kodira, C.D., Huang, S. and Weng, Y. (2010) Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC Genomics*, **11**, 569.
- Chen, J.F., Staub, J.E., Tashiro, Y., Isshiki, S. and Miyazaki, S. (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. *Euphytica*, **96**, 413–419.
- Cheng, Z.K., Presting, G.G., Buell, C.R., Wing, R.A. and Jiang, J.M. (2001) High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. *Genetics*, **157**, 1749–1757.
- Cheng, F., Mandakova, T., Wu, J., Xie, Q., Lysak, M.A. and Wang, X. (2013) Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell*, **25**, 1541–1554.
- Danin-Poleg, Y., Reis, N., Baudracco-Arnas, S., Pitrat, M., Staub, J.E., Oliver, M., Arus, P., deVicente, C.M. and Katzir, N. (2000) Simple sequence repeats in *Cucumis* mapping and map merging. *Genome*, **43**, 963–974.
- Darling, A.E., Mau, B., Blattner, F.R. and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403.
- Darling, A.E., Mau, B. and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, **5**, e11147.
- Diaz, A., Fergany, M., Formisano, 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.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.
- Foissac, S., Gouzy, J., Rombauts, S., Mathe, C., Amselem, J., Sterck, L., van de Peer, Y., Rouze, P. and Schiex, P. (2008) Genome annotation in plants and fungi, EuGene as a model platform. *Curr. Bioinform.* **3**, 87–97.
- Garcia-Mas, J., Benjak, A., Sanseverino, W. et al. (2012) The genome of melon (*Cucumis melo* L.). *Proc. Natl Acad. Sci. USA*, **109**, 11872–11877.
- Ghebreinsae, A.G., Thulin, M. and Barber, J.C. (2007) Relationships of cucumbers and melons unraveled: molecular phylogenetics of *Cucumis* and related genera (Benincaseae, Cucurbitaceae). *Am. J. Bot.* **94**, 1256–1266.
- 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.
- Guo, S., Zhang, J., Sun, H. et al. (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat. Genet.* **45**, 51–58.
- Han, Y.J. and Wessler, S.R. (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199.
- Han, Y., Zhang, Z., Liu, C., Liu, J., Huang, S.W., Jiang, J. and Jin, W.W. (2009) Centromere repositioning in cucurbit species: implication of the genomic impact from centromere activation and inactivation. *Proc. Natl Acad. Sci. USA*, **106**, 14937–14941.
- Hoffmann, A.A. and Rieseberg, L.H. (2008) Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Evol. Syst.* **39**, 21–42.
- Huang, S., Li, R., Zhang, Z. et al. (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* **41**, 1275–1281.
- Illa, E., Sargent, D.J., Girona, E.L. et al. (2011) Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. *BMC Evol. Biol.* **11**, 9.
- Jung, S., Cestaro, A., Troglio, M. et al. (2012) Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceae subfamilies. *BMC Genomics*, **13**, 129.
- Karpen, G.H. and Allshire, R.C. (1997) The case for epigenetic effects on centromere identity and function. *Trends Genet.* **13**, 489–496.
- Katzir, N., Danin-Poleg, Y., Tzori, G., Karchi, Z., Lavi, U. and Cregan, P.B. (1996) Length polymorphism and homologies of microsatellites in several Cucurbitaceae species. *Theor. Appl. Genet.* **93**, 1282–1290.
- Kirkbride, J.H. (1993) *Biosystematic Monograph of the Genus Cucumis* (Cucurbitaceae): Botanical Identification of Cucumbers and Melons. Boone, NC: Parkway Publishers.
- Kirkpatrick, M. (2010) How and why chromosome inversions evolve. *PLoS Biol.* **8**, e1000501.
- Koch, M.A. and Kiefer, M. (2005) Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species – *Capsella rubella*, *Arabidopsis lyrata* subsp. *petraea*, and *A. thaliana*. *Am. J. Bot.* **92**, 761–767.
- Koo, D.-H., Nam, Y.-W., Choi, D., Bang, J.-W., de Jong, H. and Hur, Y. (2010) Molecular cytogenetic mapping of *Cucumis sativus* and *C. melo* using highly repetitive DNA sequences. *Chromosome Res.* **18**, 325–336.
- Kosambi, D.D. (1944) The estimation of map distance from recombination values. *Ann. Eugen.* **12**, 172–175.
- Kozhukhov, S.A. (1930) Karyological investigations of the genus *Cucumis*. *Bull. Appl. Bot. Plant Breed.* **23**, 357–366.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645.
- Lagercrantz, U. (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics*, **150**, 1217–1228.
- Li, D., Cuevas, H.E., Yang, L. et al. (2011a) Syntenic relationships between cucumber (*Cucumis sativus* L.) and melon (*C. melo* L.) chromosomes as revealed by comparative genetic mapping. *BMC Genomics*, **12**, 396.
- Li, Z., Zhang, Z.H., Yan, P.C., Huang, S.W., Fei, Z.J. and Lin, K. (2011b) RNA-Seq improves annotation of protein-coding genes in the cucumber genome. *BMC Genomics*, **12**, 540.
- Livingstone, K.D., Lackney, V.K., Blauth, J.R., van Wijk, R. and Jahn, M.K. (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. *Genetics*, **152**, 1183–1202.
- Lowry, D.B. and Willis, J.H. (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, e1000500.
- Luo, M.C., Deal, K.R., Akhunov, E.D. et al. (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc. Natl Acad. Sci. USA*, **106**, 15780–15785.
- Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K. and Schubert, I. (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl Acad. Sci. USA*, **103**, 5224–5229.
- Mandakova, T. and Lysak, M.A. (2008) Chromosomal phylogeny and karyotype evolution in $x = 7$ crucifer species (Brassicaceae). *Plant Cell*, **20**, 2559–2570.
- Margulies, M., Egholm, M., Altman, W.E. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
- Murat, F., Xu, J.H., Tannier, E., Abrouk, M., Guilhot, N., Pont, C., Messing, J. and Salse, J. (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res.* **20**, 1545–1557.
- Murray, M.G. and Thompson, W.F. (1980) Rapid isolation of high molecular weight DNA. *Nucleic Acids Res.* **8**, 4321–4325.
- Neuhausen, S.L. (1992) Evaluation of restriction fragment length polymorphism in *Cucumis melo*. *Theor. Appl. Genet.* **83**, 379–384.
- van Ooijen, J.W. and Voorrips, R.E. (2001) *JoinMap Version 3.0 Software for the calculation of genetic linkage maps*. Plant Research International, Wageningen.
- Park, Y., Katzir, N., Brotman, Y., King, J., Bertrand, F. and Havey, M. (2004) Comparative mapping of ZYMV resistances in cucumber (*Cucumis sativus* L.) and melon (*Cucumis melo* L.). *Theor. Appl. Genet.* **109**, 707–712.
- Renner, S.S., Schaefer, H. and Kocyan, A. (2007) Phylogenetics of *Cucumis* (Cucurbitaceae): cucumber (*C. sativus*) belongs in an Asian/Australian clade far from melon (*C. melo*). *BMC Evol. Biol.* **7**, 58.
- Rocchi, M., Archidiacono, N. and Stanyon, R. (2006) Ancestral genomes reconstruction: an integrated, multi-disciplinary approach is needed. *Genome Res.* **16**, 1441–1444.
- Ruiz-Herrera, A., Farre, M. and Robinson, T.J. (2012) Molecular cytogenetic and genomic insights into chromosomal evolution. *Heredity*, **108**, 28–36.

- Salse, J.** (2012) *In silico* archeogenomics unveils modern plant genome organisation, regulation and evolution. *Curr. Opin. Plant Biol.* **15**, 122–130.
- Schaefer, H.** (2007) *Cucumis* (Cucurbitaceae) must include *Cucumella*, *Dicelospermium*, *Mukia*, *Myrmecosicyos*, and *Oreosyce*: a recircumscription based on nuclear and plastid DNA data. *BLUMEA*, **52**, 165–177.
- Schubert, I. and Lysak, M.A.** (2011) Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* **27**, 207–216.
- Sebastian, P., Schaefer, H., Telford, I.R.H. and Renner, S.S.** (2010) Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc. Natl Acad. Sci. USA*, **107**, 14269–14273.
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M. and Birol, i.** (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123.
- Soderlund, C., Bomhoff, M. and Nelson, W.M.** (2011) SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* **39**, e68.
- Thiel, T., Michalek, W., Varshney, R.K. and Graner, A.** (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **106**, 411–422.
- Trivedi, R.N. and Roy, R.P.** (1970) Cytological studies in *Cucumis* and *Citrullus*. *Cytologia*, **35**, 561–671.
- Vilanova, S., Sargent, D.J., Arus, P. and Monfort, A.** (2008) Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). *BMC Plant Biol.* **8**, 67.
- Whitaker, T.W.** (1933) Cytological and phylogenetic studies in the Cucurbitaceae. *Bot. Gaz.* **94**, 780–790.
- Wu, F.N. and Tanksley, S.D.** (2010) Chromosomal evolution in the plant family Solanaceae. *BMC Genomics*, **11**, 182.
- Xu, Z. and Wang, H.** (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268.
- Yang, L., Koo, D.-H., Li, Y., Zhang, X., Luan, F., Havey, M.J., Jiang, J. and Weng, Y.** (2012) Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant J.* **71**, 895–906.
- Yogeeswaran, K., Frary, A., York, T.L., Amenta, A., Lesser, A.H., Nasrallah, J.B., Tanksley, S.D. and Nasrallah, M.E.** (2005) Comparative genome analyses of *Arabidopsis* spp.: inferring chromosomal rearrangement events in the evolutionary history of *A. thaliana*. *Genome Res.* **15**, 505–515.