BMC
Genomics

RESEARCH ARTICLE
Open Access

Use of targeted SNP selection for an improved anchoring of the melon (*Cucumis melo* L.) scaffold genome assembly

Jason M Argyris[1], Aurora Ruiz-Herrera[2,3], Pablo Madriz-Masis[1], Walter Sanseverino[1,5], Jordi Morata[1], Marta Pujol[1], Sebastián E Ramos-Onsins[4] and Jordi Garcia-Mas[1*]

Abstract

Background: The genome of the melon (*Cucumis melo* L.) double-haploid line DHL92 was recently sequenced, with 87.5 and 80.8% of the scaffold assembly anchored and oriented to the 12 linkage groups, respectively. However, insufficient marker coverage and a lack of recombination left several large, gene rich scaffolds unanchored, and some anchored scaffolds unoriented. To improve the anchoring and orientation of the melon genome assembly, we used resequencing data between the parental lines of DHL92 to develop a new set of SNP markers from unanchored scaffolds.

Results: A high-resolution genetic map composed of 580 SNPs was used to anchor 354.8 Mb of sequence, contained in 141 scaffolds (average size 2.5 Mb) and corresponding to 98.2% of the scaffold assembly, to the 12 melon chromosomes. Over 325.4 Mb (90%) of the assembly was oriented. The genetic map revealed regions of segregation distortion favoring SC alleles as well as recombination suppression regions coinciding with putative centromere, 45S, and 5S rDNA sites. New chromosome-scale pseudomolecules were created by incorporating to the previous v3.5 version an additional 38.3 Mb of anchored sequence representing 1,837 predicted genes contained in 55 scaffolds. Using fluorescent in situ hybridization (FISH) with BACs that produced chromosome-specific signals, melon chromosomes that correspond to the twelve linkage groups were identified, and a standardized karyotype of melon inbred line T111 was developed.

Conclusions: By utilizing resequencing data and targeted SNP selection combined with a large F2 mapping population, we significantly improved the quantity of anchored and oriented melon scaffold genome assembly. Using genome information combined with FISH mapping provided the first cytogenetic map of an *inodorus* melon type. With these results it was possible to make inferences on melon chromosome structure by relating zones of recombination suppression to centromeres and 45S and 5S heterochromatic regions. This study represents the first steps towards the integration of the high-resolution genetic and cytogenetic maps with the genomic sequence in melon that will provide more information on genome organization and allow for the improvement of the melon genome draft sequence.

Keywords: Melon, SNP, Genome, Scaffold, Pseudomolecules, FISH, Karyotype

* Correspondence: jordi.garcia@irta.cat
[1]IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, 08193 Barcelona, Spain
Full list of author information is available at the end of the article

Argyris *et al. BMC Genomics* (2015) 16:4

Page 2 of 14

Background

Melon (*Cucumis melo* L.) is a highly diversified species that is cultivated worldwide, with more than 31 million tons produced in 2011 (http://faostat.fao.org). It is a eudicot diploid species (2n = 2x = 24) that belongs to the Cucurbitaceae family which includes other important vegetable crops such as cucumber, watermelon and squash. In recent years melon has become a model system for studying important biological processes as plant sex determination [1,2], phloem transport [3], and fruit ripening [4]. At the same time, several genetic and genomic tools are available in melon, such as saturated genetic maps [5] or EST databases [6].

The advent of high throughput next generation sequencing (NGS) technologies combined with declining cost has led to an explosion in the number of sequenced plant genomes in the past several years [7]. The majority have used a whole genome shotgun (WGS) strategy and a hybrid approach to genome assembly, incorporating both short reads typically generated by NGS platforms, and Sanger-derived end sequences from large insert BAC and fosmid clones [8]. The genome of melon was recently sequenced, producing a genome assembly with N50 scaffold size of 4.68 Mb and N90 index of 78, of good quality when compared with other plant genomes recently reported based on NGS [9]. The quality of the genome assembly has an impact on applications of the genome sequence by providing, among others, a reference genome for resequencing analysis. Comparisons of resequenced individuals to a reference genome offer a means to identify and characterize genetic polymorphisms, of which single nucleotide polymorphisms (SNPs) are the most important and abundant. SNP discovery using NGS approaches has been reported in many crop species [10], and it has been described that SNPs identified through whole genome resequencing have a low false discovery rate compared to other methods [11]. The abundance of SNPs in the genome, coupled with the diversity of technologies for performing multiplex assays that can range from genotyping single SNPs at a time to over one million in parallel [12], make them a powerful tool for genetic mapping and marker assisted breeding.

The assembly of genomic scaffolds to generate chromosome-scale sequences, or pseudomolecules (PM), is done by integrating them with genetic or physical maps [13]. A genetic map of sufficient accuracy and marker density is therefore essential for anchoring genomic scaffolds to linkage maps. High density genetic maps developed with single sequence repeat (SSR) and SNP markers have been used in many crop species to anchor genome assemblies. Recent examples include peach (*Prunus persica*) [14], diploid strawberry (*Fragaria vesca*) [15] and apple (*Malus x domestica*) [16] as well species of the

Cucurbitaceae family such as watermelon (*Citrullus lanatus*) [17], and cucumber (*Cucumis sativus*) [18]. A set of 602 polymorphic SNPs, derived from a *C. melo* expressed sequence tag (EST) collection [19] were used to produce a high resolution genetic map based on 72 doubled haploid lines (DHLs) of the cross "Piel de Sapo" (PS) [T111] x Songwhan Charmi' (SC) [PI 161375] to anchor and orient 87.5% (316.3 Mb) and 80.8% (291.9 Mb) of the melon scaffold genome assembly v3.5, respectively [9].

Karyotype analysis in *Cucumis* is difficult due to the small size and poor stainability of chromosomes [20]. Despite this, karyotypes based on chromosome banding and morphology have been developed for several melon types with varying results [21-23]. Fluorescence in situ hybridization (FISH) allows direct mapping of both single-copy and repetitive DNA sequences on chromosomes, and is a more robust and reliable method of identifying chromosomes and establishing an accurate karyotype. FISH is an important tool not only in cytogenetics, but also in genomic applications to estimate the size and positions of gaps in genome assemblies following sequencing, and in characterizing regions of recombination suppression where the resolution of genetic linkage maps are insufficient [24]. Probes commonly used in FISH karyotyping in plant species are the 45S and 5S rRNA genes, tandemly repeated sequences near telomeres, centromere-specific repeats, and large insert DNA clones such as bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs), and fosmids [25]. Extensive cytogenetic studies have been conducted using FISH with these probes in cucumber, including the development of integrated molecular cytogenetic maps [26-28]. Relatively less has been done in melon, but several studies using FISH with 45S, 5S, and centromere specific probes have been performed to develop karyotypes [29-31]. Cross species FISH, based on sequence similarity of probes between closely related species, has been used with cucumber fosmids to make inferences on melon chromosome structure and karyotype [32-34]. However, in some instances, fosmids mapped to several non-target melon chromosomes, or failed to map. Thus, the importance of FISH in genomic applications, and non-target specific hybridization using cross-species FISH in melon, highlights the need for an efficient FISH methodology using melon specific large insert DNA probes.

Although the percentage of assembly effectively anchored to chromosomes of the melon genome described in [9] was of a comparable level to some recently sequenced crop species [16,17], insufficient marker coverage in the genetic map left several large, gene-rich scaffolds unanchored. Also, due to the small population size and insufficient recombination, many anchored scaffolds contained just one mapped marker, or contained

Argyris *et al. BMC Genomics* (2015) 16:4

Page 3 of 14

multiple tightly linked markers whose map order was undefined, which made orientation of some scaffolds impossible. To address these deficiencies, we detail the use of resequencing data between of PS and SC to develop a new, better distributed set of SNP markers, and in combination with a larger segregating population, a high-density genetic map to augment the amount of anchored and oriented melon genome scaffolds. Also, we describe the construction of more complete PMs required to accurately represent the scaffold genome assembly. We also describe the use of BAC probes to identify the 12 melon chromosomes and in combination with two color FISH, orient them with respect to the genetic map and PMs. This is an important first step in developing an integrated molecular cytogenetic map which will be a key tool in improving the quality of the draft genome of melon for future uses.

Results

SNP validation and construction of the genetic map

Melon genome assembly v3.5 is composed of 1,599 scaffolds, of which 87 were previously anchored to the genetic map, representing 87.5% of the scaffold assembly [9]. As 98.5% of the scaffold assembly is contained in the 150 largest scaffolds (Additional file 1: Table S1) and SNPs between PS and SC lines are available [9], we mined SNPs from 147 of the 150 scaffolds to construct a new genetic map in an F2 population derived from the PS x SC cross. The scaffolds ranged in size from 9.16 Mb (CM3.5_scaffold00004) to 31.5 kb (CM3.5_scaffold00144). Three scaffolds (CM3.5_scaffold00134, 147, and 150) were excluded due to the inability to extract high-quality SNPs from the genome sequence. 768 SNPs were selected for the design of a GoldenGate assay, 435 SNPs from unanchored scaffolds (sub-set one) and 333 SNPs from already anchored scaffolds (sub-set two) (Additional file 1: Table S2). Overall, 599/768 (78%) of SNPs were successfully genotyped in the F2 population. Of the 435 new SNPs comprising sub-set one, 288 (66%) were validated. The remaining 147 SNPs from this set either failed to amplify, presented fused or extra clusters (null alleles), or were false SNPs. The average designability rank score for failed SNPs (0.86) was markedly lower compared to that for successful SNPs (0.95). Fifty percent of failed SNPs (74/147) were designed from the 31 smallest scaffolds that had a mean size of 121 kb. Therefore, SNPs that were designed to just 1/5 of the 147 scaffolds originally selected for re-anchoring the scaffold genome assembly accounted for half of the failure rate. The success rate for the second, previously validated sub-set of SNPs was much higher, with 311 of 332 (94%) functioning. The successful SNPs from the two sub-sets were combined into a single set of 599 SNPs that was entered into JoinMap for construction of the genetic map. Nineteen SNPs were excluded from

the mapping process due to extreme segregation distortion, missing data, or implausible fit in a linkage group (LG). Following this step, a core set of 580 SNPs (279 from set one + 301 from set two) and 139 PS x SC F2 lines were used to construct a high resolution genetic map spanning 1,153 cM of the 12 melon LGs with a mean of 49 SNPs per LG (Figure 1, Table 1). The size of LGs ranged from 113.6 cM (LG6) to 67.2 cM (LG10) with the number of SNPs per LG ranging from 29 to 65. Average spacing of markers genome-wide was 1.99 cM/SNP.

In constructing the genetic linkage map, major areas of segregation distortion ($p < 0.05$) were evident on LGs I and IV spanning 41.4 and 38 cM, respectively (Additional file 1: Table S3). In both cases, marker alleles were skewed toward the SC parent. Segregation distortion was reported previously in DHL and F2 mapping populations of melon derived from the PS x SC cross [35,36] but did not correspond to the regions detected in this study. Conversely, segregation distortion was not detected in backcross populations used for NIL development [37].

Genome re-anchoring

The genetic map was used to perform a re-anchoring of the scaffold genome assembly to the 12 LGs (Figure 1, Table 1). By anchoring the genetic map, six chimeric scaffolds (CM3.5_scaffold00022, CM3.5_scaffold00029, CM3.5_scaffold00045, CM3.5_scaffold00053, CM3.5_scaffold00063, CM3.5_scaffold00072), each mapping in two different locations in the genome (LGV/LGVIII, LGVII/X, LGVIII/LGXI, LGVII/LGX, LGIV/LGX, LGI/LGIV, and LGVII/LGX, respectively) were detected (Additional file 1: Table S4). These were due to single misassemblies within scaffolds where paired-end links between contig sequences were erroneously joined during the genome assembly. After splitting, newly created scaffolds were designated as CM3.5.1_scaffold01600-1605. Additionally, a misassembly was identified in CM3.5_scaffold01599. CM3.5_scaffold00056 was first detected as chimeric in assembly v3.4, and split into CM3.5_scaffold01599 for v3.5 [9]. As the split was likely not performed properly, 91.6 kb was trimmed from the latter and appended to the former. The melon genome assembly was updated to v3.5.1, identical to v3.5 except for the above modifications and a slightly decreased final size of the assembled genome (375.47 Mb).

Each of 141 scaffolds (135 from assembly v 3.5 + 6 newly created from chimeras), containing over 354.84 Mb, or 98.2% of the scaffold genome assembly, was anchored to the genetic map with a minimum of 1 and a maximum of 12 SNPs (Additional file 1: Table S2) The number of scaffolds anchored per LG ranged from 6 (LGV) to 18 (LGIV) with a size ranging from 9.16 Mb to 41 kb. The size of newly anchored scaffolds ranged from
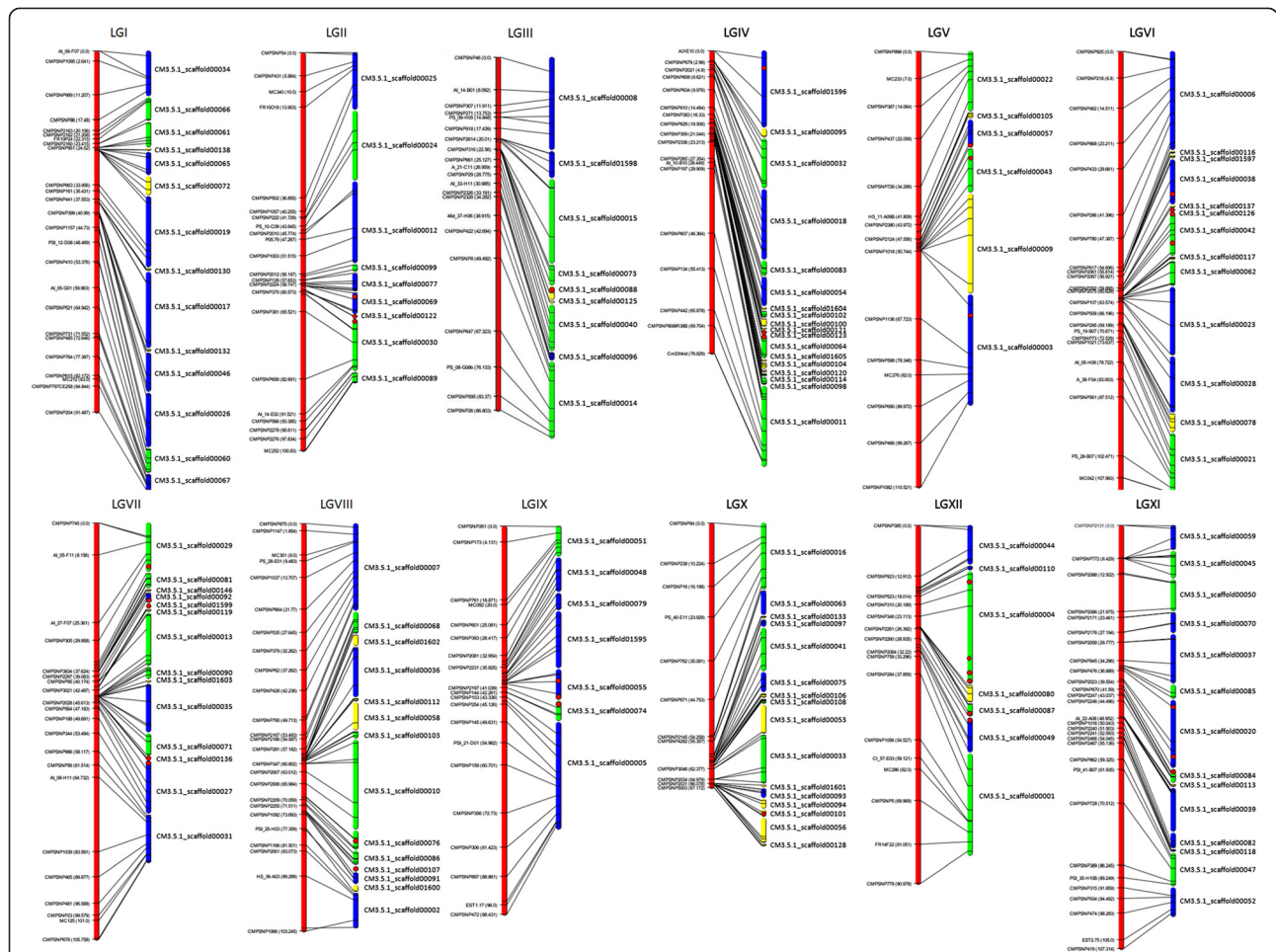
Argyris *et al. BMC Genomics* (2015) 16:4

Page 4 of 14

Figure 1 Anchoring of the melon scaffold genome assembly to the PS x SC F2 genetic map. Red bars represent the 12 melon linkage groups; SNPs are located according to genetic distance (cM). Melon genome scaffolds were positioned in each linkage group with corresponding genetic markers. Blue, scaffolds in positive orientation; green, scaffolds with negative orientation (reverse and complemented); yellow, scaffolds that were anchored but not oriented. Red dots represent locations of centromere-specific repeats inferred by *in silico* analysis.* Not all SNP names are represented in the genetic map.

4.17 Mb (CM3.5.1_scaffold00036) to 41.2 kb (CM3.5.1_ scaffold00146) with an average size of approximately 730 kb. We were able to anchor 99 scaffolds with 2 or more SNPs and thus determine orientation for over 327 Mb (90.5%) of the scaffold assembly. Forty-two scaffolds, totaling 27.8 Mb, remained unoriented (Additional file 1: Table S5). Furthermore, the ordering of 20 of the unoriented scaffolds, located on six PMs (III, IV, VI, VII, VIII, and X) was uncertain. Most of these were less than 1 Mb in size and were anchored in recombination suppression regions (Figures 1 and 2). Although the average marker spacing was 694 kb/SNP, large gaps were present on the physical map; e.g. 6.4 Mb in CM3.5.1_scaffold00004 on LGXII between CMSNP348 and CMPSNP211 and a 5.1 Mb gap in CM3.5.1_scaffold00017 on LGI between CMPSNP1111CE74 and CMPSNP399. We failed to anchor 12 scaffolds targeted from assembly v3.5 containing 1.52 Mb of sequence

ranging in size from 410–31 kb with at least 1 SNP marker (Additional file 1: Table S1, Additional file 1: Table S2). Anchored scaffolds were assembled into 12 PMs according to the nomenclature and established orientation of the melon LGs of the consensus linkage map [5]. These were supplemented by a virtual chromosome 0 containing 20.6 Mb of sequence. The PMs ranged in size from 35.9 (LGVI) to 24.1 (LGIX) Mb and the number of newly anchored scaffolds ranged from 0 (PMIX) to 11 (PMX) per PM (Figure 1, Table 1, Additional file 1: Table S5).

The genome-wide recombination rate (cM/Mb) calculated for the melon genome was 3.3 cM/Mb. The recombination rates among PMs varied widely from a minimum of 2.3 cM/Mb on LGIV, to 4.1 cM/Mb on LGIX (Table 1) and, with the exception of PMX, was correlated to physical size, with the shortest chromosomes exhibiting higher recombination rates, and vice versa. The ratio between

Argyris *et al. BMC Genomics* (2015) 16:4

Page 5 of 14

Table 1 Anchoring of melon genome assembly v 3.5.1 to the PS x SC F2 genetic map

LG	LG size (cM)	SNPs used in map construction	Scaffolds anchored*	Genome anchored (bp)	% Scaffold assembly	Oriented scaffolds	Genome oriented (bp)	% Scaffold assembly	Recombination rate (cM/Mb)
I	91.5	53	14	35,370,099	9.8	11	35,197,227	9.7	2.6
II	100.8	40	9	26,185,771	7.2	8	26,070,550	7.2	3.9
III	86.8	39	9	29,379,469	8.1	7	28,223,253	7.8	3.0
IV	76.8	65	18	33,106,231	9.2	10	30,831,195	8.5	2.3
V	110.5	29	6	28,332,775	7.8	4	19,742,536	5.5	3.9
VI	113.6	54	13	35,927,859	9.9	7	33,605,504	9.3	3.2
VII	105.8	57	14	26,760,857	7.4	9	26,066,430	7.2	4.0
VIII	103.2	56	14	32,500,408	9.0	9	28,100,052	7.8	3.2
IX	98.4	39	7	24,101,567	6.7	7	24,101,567	6.7	4.1
X	67.2	56	16	25,347,316	7.0	8	18,851,255	5.2	2.7
XI	107.3	60	14	31,429,130	8.7	13	31,190,979	8.6	3.4
XII	91	32	7	26,394,393	7.3	6	25,034,488	6.9	3.8
Total	**1162**	**580**	**141**	**354,835,875**	**98.2**	**99**	**327,015,036**	**90.5**	**3.3**

*Includes newly created scaffolds (see text).

genetic and physical distances localized regions of high recombination mostly concentrated around the ends of the PMs, with recombination suppression increasing as distance from the ends increased (Figure 2, Additional file 1: Table S6). An exception was PMX, which showed high recombination only on one end of the chromosome, and a large zone of recombination suppression extending for approximately 16 Mb to the opposite end. Recombination rates within chromosomes also varied widely, reflective of the size and distribution of zones. BLAST analysis with four distinct centromere specific repeats: sSat107 and CentSpA, B, C as well as repeats specific to the nucleolus organizer region (NOR) or 45S rDNA and 5S rDNA sites, corresponded with zones of recombination suppression (Figure 1, Additional file 1: Table S7, Additional file 1: Table S8).
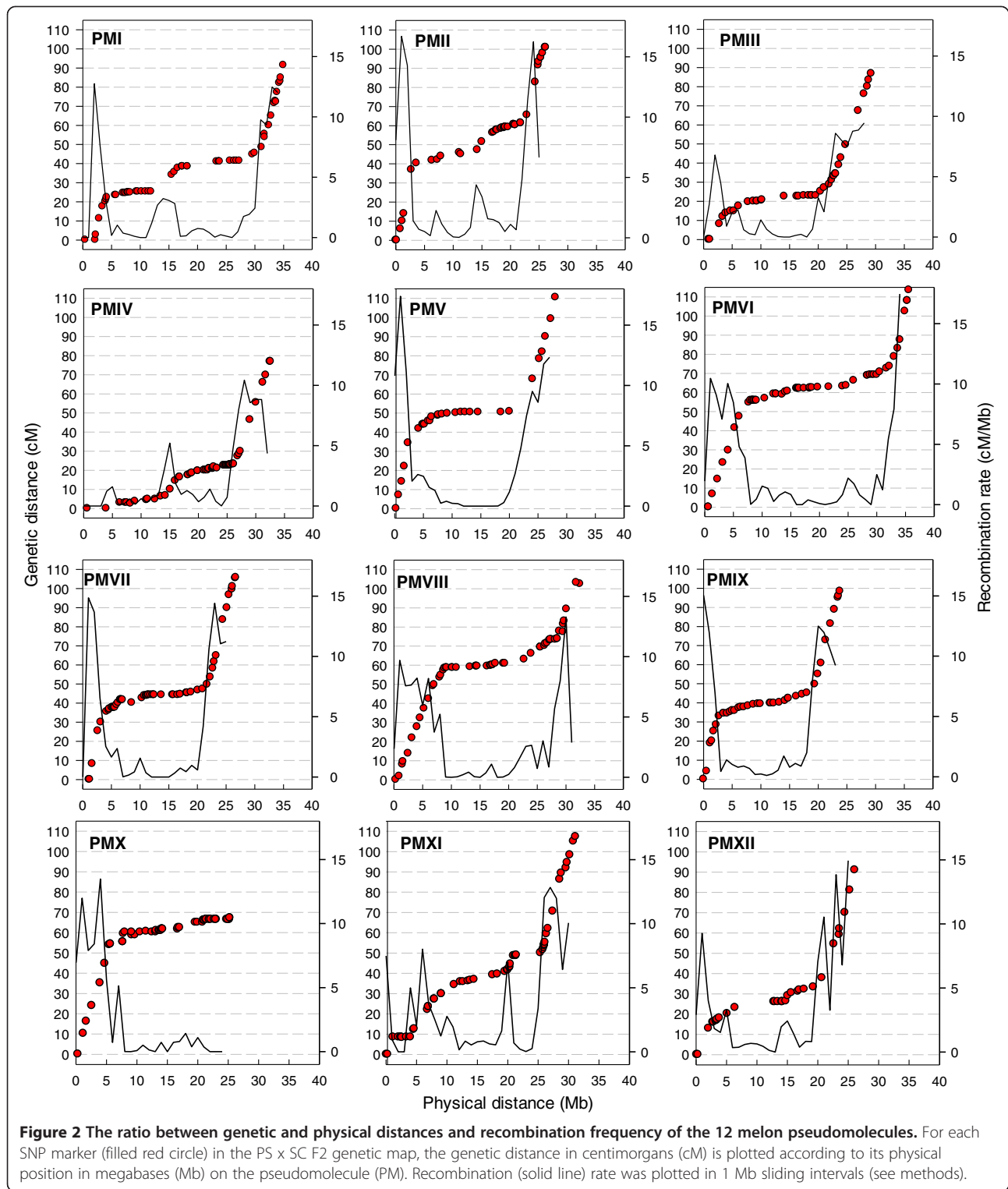
Karyotype of PS

Nineteen single copy BACs were used to identify the 12 metaphase chromosomes of melon designated CME 1–12 (Table 2, Figure 3A). Chromosomes were then assigned to their corresponding LGs according to the map positions of 19 genetic markers (13 RFLP, 6 SNP) contained in these BACs. Seven chromosomes (CME 2, 4–7, 9 and 11) were labeled with 2 BAC probes located in ends of each chromosome, allowing the identification of the short (p) and long (q) arms, and orientation of chromosomes (Table 2, Additional file 2: Figure S1). Single BACs were hybridized to the long arms of CME 1, 3, and 12 and to the short arm of CME 10. The designation for CME 8 was uncertain. Chromosomes were categorized morphologically as metacentric (CME 12), sub-metacentric (CME 1, 3, 4, 5, 6, 8, 9, 11) or acrocentric (CME 2, 7, 10) and a

standardized melon karyotype constructed (Table 2, Figure 3A). The relative sizes of chromosomes was visible in the karyotype, with CME 9, 10 and 12 being the smallest and most compact, while CME 1, 4, and 11 were among the largest. The structure of PMs corresponded to the morphology of CME 2, 4, and 6 (Figure 3B-D). Centromere positions indicated by pSat107 and CentSp sequences correlated well to zones of recombination suppression visible in the karyotype, and to the presence of many, relatively small scaffolds on the physical map. A large constriction was evident on CME4 (Figure 3C). FISH mapping of BAC clones on chromosomes confirmed the correct positioning of genomic scaffolds within PMs as the hybridization signals were located on the extremes of chromosome arms, as predicted by their anchoring to the genetic map (Figure 3B-D).

Discussion

The resequencing of PS and SC, the parents of the DHL92 melon reference genome, identified 2.1 million putative SNPs occurring with a frequency of one per 176 bp [9]. We took advantage of this resource to identify a new set of 435 polymorphic SNPs between PS and SC to develop a high-resolution genetic map and anchor 354.98 Mb of sequence contained in 141 scaffolds, representing 98.2% of the ~361 Mb melon scaffold genome assembly v3.5.1 (Table 1). We oriented 325.4 Mb of sequence, representing 90.3% of the assembly.

The quantity of anchored, and especially oriented, genome was comparable to other recently sequenced crop plants [14-18,38-45] (Additional file 1: Table S9) and highest among sequenced cucurbit species [17,18]. It also represented a significant improvement from the anchoring of

Argyris *et al. BMC Genomics* (2015) 16:4

Page 6 of 14

Figure 2 The ratio between genetic and physical distances and recombination frequency of the 12 melon pseudomolecules. For each SNP marker (filled red circle) in the PS x SC F2 genetic map, the genetic distance in centimorgans (cM) is plotted according to its physical position in megabases (Mb) on the pseudomolecule (PM). Recombination (solid line) rate was plotted in 1 Mb sliding intervals (see methods).

assembly v3.5 [9]. As SNPs used to construct the genetic map and anchor assembly v3.5 were selected at random from a collection of melon ESTs, their chromosomal location in the assembly was unknown [19]. Furthermore, SNPs derived from ESTs have a high false discovery rate

compared to other sources as previously described in [11]. Therefore, marker coverage of the genome was likely incomplete because some scaffolds were either not represented by ESTs and subsequently, SNPs; the SNPs failed because they were false; or they failed to be validated for

Argyris *et al. BMC Genomics* (2015) 16:4

Page 7 of 14

Table 2 Markers and BAC clones used to identify and orient the 12 melon chromosomes

Chr	LG	Marker*	Scaffold	Scaffold position (bp)	BAC clone	Chr arm	Chr form
CME1	I	MC216	CM3.5scaffold00034	434,148 - 433,875	39G23	q-arm	sm
CME2	II	MC252	CM3.5scaffold00089	58,157 - 57,875	27K12	p-arm	ac
		MC340	CM3.5scaffold00025	1,148,058 - 1,146,502	2N3	q-arm	
CME3	III	MC127	CM3.5scaffold00008	924,766 - 924,972	5P10	p-arm	sm
CME4	IV	CmEthlmd	CM3.5scaffold00011	486,913 - 486,496	14C18	p-arm	sm
		A13E10	CM3.5scaffold01596	644,885 - 645,310	41C15	q-arm	
CME5	V	MC233	CM3.5scaffold00022	4,179,585 - 4,180,293	58B10	p-arm	sm
		MC276	CM3.5scaffold00003	6,499,792 - 6,499,614	01N03	q-arm	
CME6	VI	MC042	CM3.5scaffold00021	551,528 - 550,701	20H14	p-arm	sm
		MC069	CM3.5scaffold00006	671,768 - No hits	24I09	q-arm	sm
CME7	VII	MC125	CM3.5scaffold00031	3,442,638 - 3,442,244	12H24	p-arm	ac
		MC373	CM3.5scaffold00029	3,679,447 - 3,768,638	49P09	q-arm	
CME8	VIII	F080	CM3.5scaffold00007	305,608 - 306,861	22K19	–	sm
CME9	IX	MC092	CM3.5scaffold00051	1,034,849 - 1,035,091	24H03	p-arm	sm
		EST1.17	CM3.5scaffold00005	No hits - 8,239,117	53P08	q-arm	
CME10	X	CmEXP3	CM3.5scaffold00041	2,803,910 - 2,803,572	54E01	p-arm	ac
CME11	XI	MC326	CM3.5scaffold00045	1,714,833 - 1,715,049	51H20	p-arm	sm
		EST2.75	CM3.5scaffold00052	1,878,167 - 1,877,623	33O17	q-arm	
CME12	XII	MC286	CM3.5scaffold00001	2,781,343 - 2,783,316	20F17	q-arm	m

*From reference [49].
The chromosome (chr) form was assigned according the centromere position: acrocentric (ac), submetacentric (sm) and metacentric (m).

various other reasons as was seen in this study (Additional file 1: Table S1) and previously [19]. An additional problem was that SNP selection was biased to larger scaffolds, as none smaller than 575 kb were anchored in v3.5. We addressed these deficiencies and achieved a high level of anchoring first, by utilizing the resequencing data and the SNP calling pipeline, combined with an "inverse mapping" strategy of planned genetic map construction. As opposed to creating a genetic map at random, the inverse mapping strategy consisted of targeted SNP selection using data from the melon genome *a priori* to identify 147 of 150 largest scaffolds comprising the N98 index (Additional file 1: Table S1) and focusing SNP discovery to these scaffolds. This permitted SNP development for large and smaller scaffolds equally well. Second, to maximize the probability of orientation, multiple SNPs were chosen at the extremes of targeted scaffolds for the GoldenGate genotyping assay to maximize the physical distances between them. This strategy was combined with a doubling of the mapping population size to 139 F2 individuals to increase the probability of recombination, and subsequent orientation of the genomic scaffolds in the relation to the genetic map. In this way, both large, gene rich scaffolds (e.g. CM3.5.1_scaffold00036, 38, and 50) which accounted for most of the newly anchored sequence, as well as many smaller scaffolds less than 1 Mb in size could be anchored and oriented (Figure 1, Additional file 1:

Table S5). A similar strategy of employing resequencing data for SNP discovery combined with a larger population was used to improve the anchoring and orientation of the soybean genome [44]. One disadvantage to choosing markers principally from the extremes of scaffolds was uneven marker distribution, and the presence of several large physical gaps between SNPs, e.g. on LGI and LGXII. This drawback was balanced by the precision of SNP selection and effectiveness of the genome anchoring and orientation using the inverse mapping approach, and should be an effective strategy for anchoring other plant genomes given the availability of a high quality reference genome and resequencing data from parental lines.

We constructed PMs of scaffold genome assembly v3.5.1 corresponding to the 12 melon chromosomes that were significantly more complete than v3.5, incorporating an additional 38.3 Mb of anchored sequence representing 1,837 predicted genes contained in 55 scaffolds. An additional 33.5 Mb of sequence was newly oriented. Compared to assembly v3.5, the 12 PMs were augmented by a mean of 3.2 Mb, and a maximum of 9.1 Mb in the case of LGX, together representing 25,065 (91.4%) of the predicted 27,427 melon genes. The quality of the anchoring of scaffold genome assembly v3.5.1, and subsequent construction of PM builds was verified by comparing it to assembly v3.5 [9]. The scaffold order was almost completely conserved, with some exceptions being
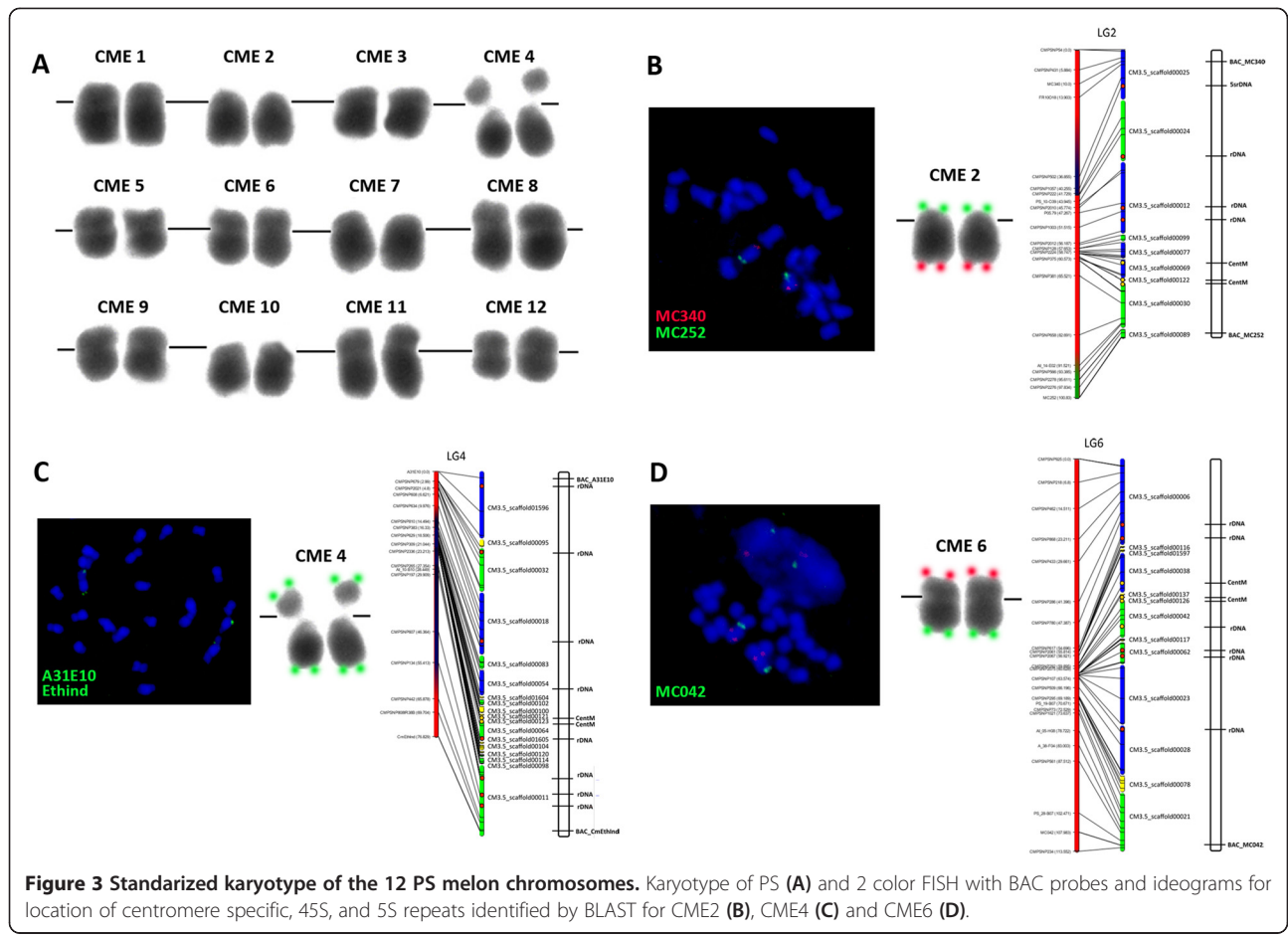
Argyris *et al. BMC Genomics* (2015) 16:4

Page 8 of 14

Figure 3 Standarized karyotype of the 12 PS melon chromosomes. Karyotype of PS **(A)** and 2 color FISH with BAC probes and ideograms for location of centromere specific, 45S, and 5S repeats identified by BLAST for CME2 **(B)**, CME4 **(C)** and CME6 **(D)**.

inversions of adjacent scaffolds on PM IX (CM3.5.1_scaffold00055 and CM3.5.1_scaffold00074) and PM XI (CM3.5.1_scaffold00059 and CM3.5.1_scaffold00045) and changes in the orientation of others. The anchoring of the 55 new scaffolds to their corresponding LGs, and their ordering within PMs, was also largely verified with data from another melon genetic map (Nurit Katzir, personal communication). This provided a further independent measure of the quality and accuracy of the anchoring of assembly v3.5.1, for example, by confirming the new position of CM3.5.1_scaffold00078 on LGVI. Lastly, by adding markers to regions of scaffolds that were previously without them, the new genetic map helped to verify the integrity of existing scaffold builds, as well as improving the quality of the assembly through the identification of six mis-assemblies and correction in the size of two others (Additional file 1: Table S4).

While pseudomolecule build v 3.5.1 was significantly more complete compared to v 3.5, approximately 5.4 Mb of sequence contained in 1,464 scaffolds ranging from 2–20 kb in size and containing 448 genes remained unanchored. Despite this, expending the effort to anchor smaller scaffolds beyond those in the

N98 index would have been prohibitive compared to the benefit (Additional file 3: Figure S2). Another feature of the new PM build was uncertainty in order and orientation of many mostly small scaffolds, principally in large zones of recombination suppression on LGIV, VI, and X. Recombination suppression zones are usually heterochromatic and gene poor, typically containing a high percentage of repeat sequences [46]. To anchor remaining scaffolds, and resolve uncertainties in order and position, above all, in zones of recombination will require the development of a FISH protocol on meiotic chromosomes. This will facilitate the comparison between the molecular and cytogenetic map as has been performed in cucumber [27], but only partially in melon [32]. In spite of these drawbacks, PM builds of v3.5.1 will be indispensable as a more complete reference sequence for comparative mapping and syntenic studies among Cucurbits [34], for downstream applications such as genotyping by sequencing (GBS) [47], and for optical mapping that has been performed in other species to improve draft genome quality [48]. The links established by a more comprehensive anchoring of the melon genome sequence and the genetic map

Argyris *et al. BMC Genomics* (2015) 16:4

Page 9 of 14

developed here and elsewhere [5,49] will be vital in moving from quantitative trait loci (QTL) for important traits, to cloning of the underlying genes.

The karyotype of the commercially important 'Piel de Sapo' Spanish type melon was 2n = 2x = 24 = 16sm + 6ac + 2m (Figure 3A). This was distinct from what has been reported in other types including muskmelon (2n = 2x = 24 = 16m + 4sm + 4stm) [22]; American muskmelon (*Cucumis melo* subsp. *melo* group *reticulatus*) and 'Hetian' (*Cucumis melo* subsp. *melo* Pang) as 2n = 2x = 24 = 20m + 4sm and 2n = 2x = 24 = 22m + 2sm, respectively [23]; and two other *C. melo* types, 'Jiashi' thick skin and 'Huangjin' thin skin type (both with 2n = 2x = 24 = 14m + 10st) [21]. The difference in karyotype may reflect a fundamental difference in the chromosome morphology due to changes in centromere position within the *inodorus* group, or misinterpretation associated with the difficulty of establishing a karyotype with small chromosomes and the lack of a landmark (CentM repeat, for example) to clearly demarcate centromeres. One feature of the PS karyotype was similar to those of other studies, namely the presence of a satellite produced by a secondary constriction in the pericentromeric region on the short arm of CME4 [23]. Large insert DNA clones, in this case BACs, were used for the first time as reliable cytogenetic landmarks to identify the 12 melon chromosomes and in combination with two-color FISH, orient them with respect to the genetic map and PMs (Figure 3B, C, D). Although probes for the 45S and 5S genes and centromere specific repeats were not mapped by FISH to melon chromosomes as in other karyotyping studies [29-31], their positions were inferred after *in silico* analysis. CentSp repetitive sequences, a family of satellite DNAs tandemly arranged in large arrays that are the primary components of the centromeres [29] were located in large zones of reduced recombination and high SNP marker density (Figures 1 and 2). This indicated that, lacking FISH probes for these features, their *in silico* positioning was a good approximation. This work represents the first steps towards the integration of the high-resolution genetic and cytogenetic map with the genomic sequence in melon and provides more information on genome organization.

Although both the number of SNPs obtained and genotyping efficiency can be improved by using an NGS-based genotyping alternative such as GBS [50], we chose the Illumina GoldenGate system in v3.5.1 in order to re-use some of the SNPs and thus have common anchoring points with the previous version v3.5 [9]. Utilizing this system and resequencing data from PS and SC lines, 66% (288/435) of newly designed SNPs called with the simply unified pair-end resequencing (SUPER) pipeline [51] were validated (Additional file 1: Table S2), adding to the more than 1,200 SNP markers validated in the

PS x SC population to date (Garcia-Mas, unpublished). The PS line represents the predominant and most economically important melon type grown in Spain and is a parental line in DHL, NIL, and other F2 mapping populations [35,37,52] used to identify QTL for an extensive variety of fruit quality and disease resistance traits [5]. Thus, the additional SNP markers developed here will add to the genotyping resources available for the mapping populations in the same genetic background [53] and facilitate fine mapping and cloning of QTL currently under development [53,54]. Furthermore, despite the fact that newly designed SNPs were selected to be polymorphic between PS and SC, a high percentage can be expected to function in other accessions and outside of the *inodorus* group, as was found for other SNPs developed for the PS x SC cross described in Esteras et al. [19]. The failure rate of SNPs in the GoldenGate assay was higher than in previous studies with melon where validation rates varied between 78 and 91% [9,19]. There was a high correlation between SNP failure and the corresponding size of the unanchored scaffold for which they were designed, illustrating the difficulty of anchoring small scaffolds with highly repetitive sequence, using diminished SNP calling parameters.

Understanding recombination landscape within a species is important for plant breeding applications. A detailed genome-wide and local depiction of recombination rates allows for accurate scaling of map-based cloning projects, marker assisted selection (MAS) strategies for trait introgression, and crossing programs where unfavorable linkage between traits needs to be broken, since regions with high or low recombination rates require correspondingly higher or lower marker densities. At a finer scale, understanding recombination in crop species can help to define recombinationally hyperactive regions or individual genes controlling recombination rate [55] as well as understanding plant genome variability [56]. The genome wide recombination rate calculated in this study (3.3 cM/Mb) for melon was higher than previous (3.1 cM/Mb) [9] primarily reflecting the increased quantity of anchored genomic sequence and correspondingly larger genetic map distances. It was within the range of other cucurbit species, 2.3 cM/Mb and 3.2 cM/Mb for watermelon and cucumber, respectively [17,18]. The recombination rate between chromosomes varied widely, reflecting the size and distribution of recombination suppression zones (Figure 2, Additional file 1: Table S6). For example, positioning of repetitive sequences by BLAST analysis supported the idea that zones of recombination suppression covering large physical distances corresponded to some components of the NOR on LGIV (18S and 5.8S) and LGX (5.8S) as well as the 5S rDNA on LGXII (Additional file 1: Table S7 and Additional file 1: Table S8), as has been reported [30,31,33]. Furthermore, we detected components of NOR on LGI (5.5S, 26S, and

Argyris *et al. BMC Genomics* (2015) 16:4

Page 10 of 14

18S), which has not previously been reported. Han et al. [32] detected a large, heterochromatic segment in the pericentromeric region of the long arm of melon chromosome I that accounted for approximately 47% of its length, which may correspond to this zone. Overall, the differences in recombination rates between chromosomes, and the collocation of recombination suppression zones with the genomic scaffolds identified here will aid in the planning for MAS projects and development of mapping populations in melon. For example, it would be necessary to increase the population size to increase the possibilities of obtaining recombinant individuals on LGIV and LGX.

Conclusions

We provided a significantly improved version of the anchored melon genome by developing a new set of targeted SNP markers with better distribution to construct a high-resolution genetic map. The improved anchoring of the melon genome will permit faster map-based cloning of genes underlying QTL for agronomically important traits, diversity assessment through comparative genomics, and provide important insights on phylogenetic relationships among cucurbits. With the improved anchoring, it was possible to make *in silico* inferences on likely chromosome structure by relating zones of recombination suppression to centromeres and 45S and 5S heterochromatic regions. The FISH mapping with the melon specific BAC probes allowed us to orient the 12 melon chromosomes and develop a karyotype. This represents the first steps towards the integration of the high-resolution genetic and cytogenetic map with the genomic sequence in melon that will provide more information on genome organization and allow for the improvement of the melon genome draft sequence.

Methods

Mapping population

An F2 population derived from the cross "Piel de Sapo" T111 (PS) x 'Songwhan Charmi' PI 161375 (SC) was grown in winter of 2012. F2 seeds were planted in trays in the greenhouse and leaf samples taken at the 2 leaf stage. DNA was extracted using the CTAB method [57]. DNA concentration was estimated with a Qubit® 2.0 Fluorometer (Life Technologies).

SNP selection and genotyping

A set of 768 SNPs was selected to perform the new anchoring. These were dived into two sub-sets (Additional file 1: Table S2). The first sub-set was composed of 435 new SNPs selected from the 63 largest unanchored scaffolds of the N98 index predicted to contain up to 40.5 Mb of the scaffold assembly of melon genome version 3.5, or scaffolds that were previously anchored, but unoriented

(Additional file 1: Table S1). All SNPs in each unanchored scaffold were first called using the SUPER pipeline [51] based on resequencing data of SC, PS and the reference sequence of DHL92 generated from Illumina paired-end sequencing averaging $70*10^6$ reads per sample at a read length of 150 bp and 22x coverage / 10Gb per sample [9]. SNP calling was done in iterations starting with maximum global quality 999 (phred-scaled quality score for the assertion. -10log_10 prob), minimum read depth of 15x, and absence of other SNPs in 50 bp flanking regions. For smaller scaffolds, it was necessary to decrease the stringency of selection parameters to obtain SNP calls. To increase the likelihood of anchoring and orienting scaffolds, a novel strategy was utilized by selecting at least two SNPs from each scaffold end as determined by their physical position from assembly version 3.5. From 1 to 5 new SNPs were selected per unanchored scaffold, depending on the size. Thus, together with previously validated SNPs, from 1 to 13 SNPs/scaffold were utilized for the re-anchoring. SNPs plus flanking sequences were submitted and evaluated with Illumina Assay Design Tool (ADT) (Illumina, Inc.) and assays selected to have the highest probability of success according to ADT results. The second sub-set contained 333 SNPs previously utilized to anchor the melon genome [9] as detailed in [19]. This second group served as a reference to make inferences across this experiment and the previous genome anchoring study v3.5.

The F2 lines described above were genotyped using the Illumina GoldenGate genotyping assay [58] and a custom 768plex Illumina Golden Gate Panel with Beadarray technology (GoldenGate Universal-32 BeadChips). Results were analyzed in a BeadArray Reader (Illumina, Inc.). SNP genotypes were scored with the Genotyping Module of the GenomeStudio Data Analysis software (Illumina, Inc.) using default parameters. F2 individuals with a call rate ≤ 85% were eliminated from further analysis. Genotype clusters were manually edited when necessary. Data was outputted and entered into the genotyping pipeline as described in [59]. SNPs with GenTrain and GeneCall 10% higher than 0.4 and 0.2, respectively, and at least two genotypic classes were classified as polymorphic and usable. Usable SNPs were then hand inspected and eliminated from the analysis if the genotype of either the PS or SC parent was not unequivocal.

Genetic map construction and genome anchoring

Following the filtration as described above, the genotyping data was entered into JoinMap v4.1 [60]. Markers were examined by Chi-square analysis and those showing extreme segregation distortions (p = 0.0001) from expected 1:2:1 ratios were eliminated from the analysis. Linkage groups were then calculated using a LOD score of 10 and the ML mapping algorithm with default parameters. Further pruning of the data within LGs was conducted by eliminating

Argyris *et al. BMC Genomics* (2015) 16:4

Page 11 of 14

individuals with >10% missing genotypes. Markers with segregation ratios that differed at p = 0.05 were classified as displaying segregation distortion. To anchor the genome assembly to the genetic map, BLASTN analysis was performed with e-value cutoff of 1E-30 for every SNP marker in the genetic map against the genome assembly. Scaffolds were then assigned to linkage groups accordingly. When more than one marker had hits on the same scaffold, it was possible to orientate this scaffold on the map. A small group of markers that were not located consecutively on the genetic map according to their physical location in a scaffold were eliminated from further analysis when there were more than two other markers with recombination to accurately establish the orientation of a scaffold (Additional file 1: Table S1). For scaffolds that were not oriented, information from the genetic map of the first version of the melon genome anchoring was used as a reference [9]. The genome anchoring to the pseudomolecules was drawn with the Harry Plotter software [61].

Correction of genome assembly and construction of chromosome-scale pseudomolecules
Manual correction of 6 chimeric scaffolds consisted of splitting them in two separate scaffolds, with the splitting point selected between the two contigs that showed most inconsistencies in paired-end links. The chosen cut point was located in Ns tracks, except in CM3.5_scaffold00063. An error in the cut point of CM3.5_scaffold01599, which was originally created by splitting CM3.5_scaffold00056 [9], was also identified. The 5′ split fragment of the former was re-joined to the 3′ end of the latter. According to the order and orientation of the scaffolds in the new physical map and the corrections made in the genome assembly, scaffolds were ordered into chromosome-scale PMs where each scaffold was joined by a 1,000 N track. The non-anchored scaffolds and contigs were joined in the non-ordered chromosome zero, also joined by a 1,000 N track. A new gff3 file with new gene coordinates in corrected scaffolds was also obtained. All these files can be accessed at http://www.melonomics.net/files/Genome/Melon_genome_v3.5.1.

Estimation of recombination
The average genome-wide recombination rate (GWRR) and the average recombination rate for each chromosome were expressed as the ratio of total genetic map length in centiMorgans divided by the genome size and physical length of the chromosome in megabase pairs, respectively. Recombination rates along each LG were estimated by first plotting Marey maps of the genetic positions of molecular markers (in centimorgans, cM) against their physical position (in Megabase pairs, Mb). Cumulative recombination curves for each chromosome were then estimated using a cubic spline method (implemented in the standard library of R, http://cran.r-project.org) and a sliding window of 1 Mb. The recombination value per position was obtained calculating the slope per window (that is, the derivative) and their curves were plotted for each of the chromosomes.

Demarcating centromeres and pericentromeric boundaries of melon pseudomolecules
The putative locations of the centromere of melon PMs were identified by aligning *C. melo* DNA for the 352 bp centromere specific pSat107 satellite (pSat107) [GI:3929695] [62] and these clustered sequences: *Cucumis melo* centromere-specific repeat A sequence (CentSpA) [GI:212961153], *Cucumis melo* centromere-specific repeat B sequence (CentSpB) [GI:212961171] and *Cucumis melo* centromere-specific repeat C sequence (CentSpC) [GI:212961196] [29] to the melon genome assembly v3.5.1 with blastn (version: blast 2.2.27). We retained the alignments that co-localized 2 or more different centromeric-specific sequences (pSat107 and any of CentSp), with a coverage of 80% or more and sequence identity of 80% or more. Putative locations of 45S and 5S rDNA were identified using the ribosomal gene annotations obtained previously [9] and by supplementing this data with RNAmmer predictions [http://www.cbs.dtu.dk/services/RNAmmer/] [63] with the default parameters.

Cytogenetics - plant material and chromosome preparation for cytogenetic analysis
PS seeds were germinated in Petri dishes. Root tips were cut, directly fixed in methanol/acid acetic (3:1) and subsequently digested in 4% (wt/vol) cellulose R-10 Onozuka, 1% (wt/vol) pectolyase Y23 and 4% (wt/vol) hemicellulase in 0.01 M citrate buffer (1/1 volume of 0.01 M citric acid and 0.01 M tri-sodium citrate) for one hour at 37°C. Enzyme solution was removed and tips were washed two times in 0.01M citrate buffer for 10 minutes at room temperature. The material was centrifuged during 5 minutes at 4,000 rpm and the resulting pellet was diluted in fresh ethanol/acid acetic (3:1). Slides were air dried at room temperature and stored at –20°C until use.

Probe preparation and Fluorescent in situ Hybridization (FISH)
BAC clones for FISH analysis (Table 2) were selected from an existing 23,040 BAC library from the double-haploid melon line DHL92 described previously [64]. BACs were previously anchored by SSR, SNP, and RFLP markers that mapped on extreme ends of each of the 12 melon LGs [49]. The positions of anchored markers on the genetic map were then used to establish a link between the scaffold-based physical map and the cytogenetic map. BLASTN analysis was conducted against the

Argyris *et al. BMC Genomics* (2015) 16:4

Page 12 of 14

melon genome assembly (http://www.melonomics.net/tools/blast/run/) on BAC end sequences when available or marker sequences, to ensure that they were located in the expected LG and present in only a single copy in the genome. DNA from BACs was extracted according to standard protocols using the QIAGEN Plasmid kit (QIAGEN), and a PCR with specific primers was performed to confirm the correct selection of each BAC clone before FISH analysis.

Fluorescence in situ hybridization (FISH) with specific BAC clones was performed on metaphase chromosomes as previously described [65] with modifications. Briefly, 1 μg of the BAC DNA was labelled with digoxigenin-11-dUTP or biotin-11-dUTP by Nick Translation technique (Abbot kit) and ethanol precipitated with competitor DNA (melon genomic DNA), salmon sperm DNA (Invitrogen, 10 mg/ml) and 1/10 volume of 3 M sodium acetate overnight at −20°C. For two-colour FISH, two different probes were labelled with digoxinenenin-11-dUTP and biotin-11-dUTP, respectively. About 250 ng of labelled DNA was used for each hybridization experiment. The precipitated probe mix was resuspended in 14 μl hybridization buffer (50% deionised formamide, 10% dextran sulfate, 2xSSC and 0.5 M sodium phosphate), denatured 80°C for 10 min and preannealed at 37°C for 1 h. After overnight hybridization, slides were washed two times in 50% formamide/2xSSC at 42°C for 10 min followed by three washes in 2xSSC at 42°C for 5 min each. Chromosomes were counterstained with DAPI (4′, 6-diamidino-2-phenylindole) in Vectashield antifade.

Microscopy
Preparations were visualized using a Zeiss Axioskop epifluorescence microscope equipped with the appropriate filters and a charged coupled device camera (ProgRes® CS10plus, Jenoptik). Chromosomes were ordered in a standard karyotypic layout, with the short arms on top. For simplicity, we decided not to change the alignment and orientation of LGs and PMs to match the orientation of chromosomes of the karyotype, but to conserve it according to the standard nomenclature and orientation of melon LGs established in Périn et al. [66]. The three-letter acronym CME was used to designate the genus and species (*Cucumis melo*), followed by the Arabic numeral that denotes the chromosome number.

Availability of supporting data
Anchored scaffold and pseudomolecule sequences are deposited in the European Nucleotide Archive (ENA) and can be accessed at http://www.ebi.ac.uk/ena/data/view/PRJEB68 with accession numbers LN681792-LN713266. The v3.5.1 sequences can also be found at http://melonomics.net/. All other supporting data are included as additional files.

Additional files

Additional file 1: Table S1. Largest scaffolds of melon genome assembly v3.5. **Table S2.** 768 SNPs selected to reanchor melon genome assembly 3.5. **Table S3.** Segregation distortion in the PS x SC F2 mapping population. **Table S4.** Chimeric and misassembled scaffolds identified in assembly v3.5. **Table S5.** Scaffolds of assembly v3.5.1 anchored to the genetic map. **Table S6.** Recombination frequencies along 12 melon PMs measured in 1 Mb windows. **Table S7.** BLAST analysis with four distinct centromere specific repeats: sSat107 and CentSpA, B, C. **Table S8.** BLAST predictions of locations of 45S and 5S rDNA in melon scaffolds using ribosomal gene annotations obtained previously (Garcia-Mas et al. [9]) and by supplementing with RNAmmer predictions [http://www.cbs.dtu.dk/services/RNAmmer/]. **Table S9.** Metrics of recently sequenced plant genomes. Numbers in parentheses are following reanchoring of genome.

Additional file 2: Figure S1. Fluorescence in situ hybridization (FISH) with BAC clones to melon metaphase chromosomes labelled with digoxinenenin-11-dUTP (green) or biotin-11-dUTP (red).

Additional file 3: Figure S2. Percentage of the scaffold genome assembly anchored as a function of the number of scaffolds. Vertical line represents scaffolds contained in the N98 index anchored in this study.

Abbreviations
NGS: Next generation sequencing; WGS: Whole-genome shotgun; SNP: Single nucleotide polymorphism; PM: Pseudomolecule; DHLs: Doubled haploid lines; FISH: Fluorescence in situ hybridization; BACs: Bacterial artificial chromosome; YAC: Yeast artificial chromosome; PS: Piel de Sapo; SC: Songwhan charmi; LG: Linkage group; EST: Expressed sequence tag; MAS: Marker assisted selection; NOR: Nucleolus organizer region; GBS: Genotyping-by-sequencing.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JMA conducted DNA extractions, chose the new SNP markers for reanchoring the assembled genome, constructed the genetic map, participated in the design of the study, and drafted the manuscript. ARH, PM and MP conducted DNA extractions, designed and performed the FISH experiments and analyses of results. WS performed SNP mining. JM carried out analysis of genotypic data and performed bioinformatic analyses and data curation. SRM performed the analysis of recombination. JGM conceived, designed and coordinated the study, and helped to draft the manuscript. All authors read and approved the final manuscript.

Author details
[1]IRTA, Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, 08193 Barcelona, Spain. [2]Departament de Biologia Cellular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Campus UAB, 08193 Barcelona, Spain. [3]Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona, Campus UAB, 08193 Barcelona, Spain. [4]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, 08193 Barcelona, Spain. [5]Present Address: Sequentia Biotech, Campus UAB - Edifici CRAG, Bellaterra - Cerdanyola del Vallès, 08193 Barcelona, Spain.

Argyris *et al. BMC Genomics* (2015) 16:4

Page 13 of 14

References

1. Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, et al. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science. 2008;321(5890):836–8.
2. Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, et al. A transposon-induced epigenetic change leads to sex determination in melon. Nature. 2009;461(7267):1135–U1237.
3. Zhang BC, Tolstikov V, Turnbull C, Hicks LM, Fiehn O. Divergent metabolome and proteome suggest functional independence of dual phloem transport systems in cucurbits. Proc Natl Acad Sci U S A. 2010;107 (30):13532–7.
4. Pech JC, Bouzayen M, Latche A. Climacteric fruit ripening: ethylene-dependent and independent regulation of ripening pathways in melon fruit. Plant Sci. 2008;175(1–2):114–20.
5. Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei ZJ, et al. A consensus linkage map for molecular markers and Quantitative Trait Loci associated with economically important traits in melon (Cucumis melo L.). Bmc Plant Biology. 2011;11:111.
6. Clepet C, Joobeur T, Zheng Y, Jublot D, Huang MY, Truniger V, et al. Analysis of expressed sequence tags generated from full-length enriched cDNA libraries of melon. BMC Genomics. 2011;12:252.
7. Michael TP, Jackson S. The First 50 Plant Genomes. Plant Genome. 2013;6(2):1–7.
8. Hamilton JP, Buell CR. Advances in plant genome sequencing. Plant J. 2012;70(1):177–90.
9. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, et al. The genome of melon (Cucumis melo L.). Proc Natl Acad Sci U S A. 2012;109(29):11872–7.
10. Kumar S, Banks TW, Cloutier S. SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics. 2012;2012:831460.
11. Ganal MW, Altmann T, Roder MS. SNP identification in crop plants. Curr Opin Plant Biol. 2009;12(2):211–7.
12. Perkel J. SNP genotyping: six technologies that keyed a revolution. Nat Methods. 2008;5(5):447–53.
13. Lewin HA, Larkin DM, Pontius J, O'Brien SJ. Every genome sequence needs a good map. Genome Res. 2009;19(11):1925–8.
14. Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, et al. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet. 2013;45(5):487–94.
15. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, et al. The genome of woodland strawberry (Fragaria vesca). Nat Genet. 2010;43(2):109–16.
16. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, et al. The genome of the domesticated apple (Malus x domestica Borkh.). Nat Genet. 2010;42(10):833–9.
17. Ren Y, Zhao H, Kou Q, Jiang J, Guo S, Zhang H, et al. A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. PLoS One. 2012;7(1):e29453.
18. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, et al. The genome of the cucumber, Cucumis sativus L. Nat Genet. 2009;41(12):1275–81.
19. Esteras C, Formisano G, Roig C, Diaz A, Blanca J, Garcia-Mas J, et al. SNP genotyping in melons: genetic variation, population structure, and linkage disequilibrium. Theor Appl Genet. 2013;126(5):1285–303.
20. Chen J, Xiao-Hui Z. Cucumis. In: Kole C, editor. Wild Crop Relatives: Genomic and Breeding Resources. Heidelberg: Springer; 2011. p. 67–90.
21. Zhang Y, Chen J, Yi H, Feng J, Wu M. Staining and slide-preparing technique of mitotic chromosomes and its use in karyotype determination in Cucumis melo L. Acta Botany Boreal-Occident Sinica. 2005;25:1735–9.
22. Ramachandran C, Seshadri VS. Cytological analysis of the genome of cucumber (Cucumis-Sativus L) and muskmelon (Cucumis-Melo L). Zeitschrift Fur Pflanzenzuchtung. 1986;96(1):25–38.
23. Ma D, Guo Z, Zhang C, Gao S, Wang M. A study on chromosome number and karyotype of melons (Cucumis melo L.). Acta Horticulturae. 1995;402:61–5.
24. Jiang J, Gill BS. Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome. 2006;49(9):1057–68.
25. Kato A, Vega JM, Han F, Lamb JC, Birchler JA. Advances in plant chromosome identification and cytogenetic techniques. Curr Opin Plant Biol. 2005;8(2):148–54.
26. Sun JY, Zhang ZH, Zong X, Huang SW, Li ZY, Han YH. A high-resolution cucumber cytogenetic map integrated with the genome assembly. BMC Genomics. 2013;14:461.
27. Lou Q, He Y, Cheng C, Zhang Z, Li J, Huang S, et al. Integration of high-resolution physical and genetic map reveals differential recombination frequency between chromosomes and the genome assembling quality in cucumber. PLoS One. 2013;8(5):e62676.
28. Lou QF, Zhang YX, He YH, Li J, Jia L, Cheng CY, et al. Single-copy gene-based chromosome painting in cucumber and its application for chromosome rearrangement analysis in Cucumis. Plant J. 2014;78(1):169–79.
29. Koo DH, Nam YW, Choi D, Bang JW, de Jong H, Hur Y. Molecular cytogenetic mapping of Cucumis sativus and C. melo using highly repetitive DNA sequences. Chromosome Res. 2010;18(3):325–36.
30. Chen JF, Staub JE, Adelberg JW, Jiang JM. Physical mapping of 45S rRNA genes in Cucumis species by fluorescence in situ hybridization. Can J Bot. 1999;77(3):389–93.
31. Hoshi Y, Kido M, Yagi K, Tagashira N, Morikawa A, Nagano K. Somatic chromosome differentiation in Cucumis melo L. and C. metuliferus E.Mey. ex Naudin. Chromosome Bot. 2013;8:7–12.
32. Han Y, Zhang Z, Liu C, Liu J, Huang S, Jiang J, et al. Centromere repositioning in cucurbit species: implication of the genomic impact from centromere activation and inactivation. Proc Natl Acad Sci U S A. 2009;106 (35):14937–41.
33. Liu C, Liu J, Li H, Zhang Z, Han Y, Huang S, et al. Karyotyping in melon (Cucumis melo L.) by cross-species fosmid fluorescence in situ hybridization. Cytogenet Genome Res. 2010;129(1–3):241–9.
34. Yang LM, Koo DH, Li DW, Zhang T, Jiang JM, Luan FS, et al. Next-generation sequencing, FISH mapping and synteny-based modeling reveal mechanisms of decreasing dysploidy in Cucumis. Plant J. 2014;77(1):16–30.
35. Gonzalo MJ, Claveria E, Monforte AJ, Dolcet-Sanjuan R. Parthenogenic haploids in Melon: generation and molecular characterization of a doubled haploid line population. J Am Soc Hort Sci. 2011;136(2):145–54.
36. Oliver M, Garcia-Mas J, Cardus M, Pueyo N, Lopez-Sese AL, Arroyo M, et al. Construction of a reference linkage map for melon. Genome. 2001;44(5):836–45.
37. Eduardo I, Arus P, Monforte AJ. Development of a genomic library of near isogenic lines (NILs) in melon (Cucumis melo L.) from the exotic accession PI161375. Theor Appl Genet. 2005;112(1):139–48.
38. Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, et al. The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nat Genet. 2012;45(1):51–8.
39. Potato Genome Sequencing Consortium T. Genome sequence and analysis of the tuber crop potato. Nature. 2011;475(7355):189–95.
40. Tomato Genome Consortium T. The tomato genome sequence provides insights into fleshy fruit evolution. Nature. 2012;485(7400):635–41.
41. Sharma SK, Bolser D, de Boer J, Sonderkaer M, Amoros W, Carboni MF, et al. Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. G3 (Bethesda). 2013;3(11):2031–47.
42. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010;463(7278):178–83.
43. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, et al. The genome of Theobroma cacao. Nat Genet. 2011;43(2):101–8.
44. Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, et al. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics. 2010;11:38.
45. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, et al. The genome of the mesopolyploid crop species Brassica rapa. Nat Genet. 2011;43(10):1035–9.
46. Renauld H. Heterochromatin: a meiotic matchmaker? Trends Cell Biol. 1997;7(5):201–5.
47. Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, et al. An improved genome release (version Mt4.0) for the model legume Medicago truncatula. BMC Genomics. 2014;15(1):312.
48. Chamala S, Chanderbali AS, Der JP, Lan T, Walts B, Albert VA, et al. Assembly and validation of the genome of the nonmodel basal angiosperm Amborella. Science. 2013;342(6165):1516–7.
49. Gonzalez VM, Garcia-Mas J, Arus P, Puigdomenech P. Generation of a BAC-based physical map of the melon genome. BMC Genomics. 2010;11:339.
50. Sonah H, Bastien M, Iquira E, Tardivel A, Legare G, Boyle B, et al. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One. 2013;8(1):e54603.

Argyris *et al. BMC Genomics* (2015) 16:4

Page 14 of 14

51. Sanseverino W, Henaff E, Casacuberta JM, Garcia-Mas J. Genetic Variability in Melon Resequenced Varieties. Poster session presented at: Plant and Animal Genome Conference XXI; 2013 Jan 12-16; San Diego, CA.

52. Monforte AJ, Oliver M, Gonzalo MJ, Alvarez JM, Dolcet-Sanjuan R, Arus P. Identification of quantitative trait loci involved in fruit quality traits in melon (Cucumis melo L.). Theor Appl Genet. 2004;108(4):750–8.

53. Vegas J, Garcia-Mas J, Monforte AJ. Interaction between QTLs induces an advance in ethylene biosynthesis during melon fruit ripening. Theor Appl Genet. 2013;126(6):1531–44.

54. Essafi A, Diaz-Pendon JA, Moriones E, Monforte AJ, Garcia-Mas J, Martin-Hernandez AM. Dissection of the oligogenic resistance to Cucumber mosaic virus in the melon accession PI 161375. Theor Appl Genet. 2009;118(2):275–84.

55. Schnable PS, Hsia AP, Nikolau BJ. Genetic recombination in plants. Curr Opin Plant Biol. 1998;1(2):123–9.

56. Gaut BS, Wright SI, Rizzon C, Dvorak J, Anderson LK. Opinion - Recombination: an underappreciated factor in the evolution of plant genomes. Nat Rev Genet. 2007;8(1):77–84.

57. Doyle JJ, Doyle JL. A rapid total DNA preparation procedure for fresh plant tissue. Focus. 1990;12:13–5.

58. Gonzalez-Neira A. The GoldenGate genotyping assay: custom design, processing, and data analysis. Methods Mol Biol. 2013;1015:147–53.

59. Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, et al. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PLoS One. 2012;7(4):e35668.

60. Stam P. Construction of Integrated genetic-linkage maps by means of a new computer package - joinmap. Plant J. 1993;3(5):739–44.

61. Moretto M, Cestaro A, Troggio M, Costa F, Velasco R. Harry Plotter: A user friendly program to visualize genome and genetic map features. In: ECCB10, 9th European Conference Computational Biology; 2010 Sep 26-29; Ghent, Belgium.

62. Brennicke A, Hemleben V. Sequence-analysis of the cloned cucumis-melo highly repetitive satellite DNA. Zeitschrift Fur Naturforschung C. 1983;38(11–1):1062–5.

63. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.

64. van Leeuwen H, Monfort A, Zhang HB, Puigdomenech P. Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between Cucumis melo and Arabidopsis thaliana. Plant Mol Biol. 2003;51(5):703–18.

65. Ruiz-Herrera A, Garcia F, Fronicke L, Ponsa M, Egozcue J, Caldes MG, et al. Conservation of aphidicolin-induced fragile sites in Papionini (Primates) species and humans. Chromosome Res. 2004;12(7):683–90.

66. Perin C, Hagen S, De Conto V, Katzir N, Danin-Poleg Y, Portnoy V, et al. A reference map of Cucumis melo based on two recombinant inbred line populations. Theor Appl Genet. 2002;104(6–7):1017–34.