Theor Appl Genet (2013) 126:349–358
DOI 10.1007/s00122-012-1983-7

ORIGINAL PAPER

Combining bulk segregation analysis and microarrays for mapping of the pH trait in melon

Amir Sherman · Ravit Eshed · Rotem Harel-Beja · Galil Tzuri · Vitaly Portnoy · Shahar Cohen · Mor Rubinstein · Arthur A. Schaffer · Joseph Burger · Nurit Katzir · Ron Ophir

Received: 21 February 2012 / Accepted: 15 September 2012 / Published online: 16 October 2012

Abstract The availability of sequence information for many plants has opened the way to advanced genetic analysis in many non-model plants. Nevertheless, exploration of genetic variation on a large scale and its use as a tool for the identification of traits of interest are still rare. In this study, we combined a bulk segregation approach with our own-designed microarrays to map the *pH* locus that influences fruit pH in melon. Using these technologies, we identified a set of markers that are genetically linked to the pH trait. Further analysis using a set of melon cultivars demonstrated that some of these markers are tightly linked to the pH trait throughout our germplasm collection. These results validate the utility of combining microarray technology with a bulk segregation approach in mapping traits of interest in non-model plants.

Communicated by M. Havey.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1983-7) contains supplementary material, which is available to authorized users.

A. Sherman (✉) · R. Eshed · M. Rubinstein · R. Ophir (✉)
Genomic Unit, Institute of Plant Sciences,
Volcani Research Center, Agricultural Research Organization,
50250 Bet Dagan, Israel
e-mail: asherman@agri.gov.il

R. Ophir
e-mail: ron@agri.gov.il

R. Harel-Beja · G. Tzuri · V. Portnoy · J. Burger · N. Katzir
Institute of Plant Sciences, Newe Ya'ar Research Center,
Agricultural Research Organization, P.O. Box 1021,
30095 Ramat Yishay, Israel

S. Cohen · A. A. Schaffer
Institute of Plant Sciences, Volcani Research Center,
Agricultural Research Organization, 50250 Bet Dagan, Israel

Introduction

A high-throughput system to map phenotypic mutants would be very effective in converting the wealth of phenotypic mutants in melon, as well as in many other plants, into useful traits that can be used in plant breeding and investigations of plant biology. Ultimately, accurate mapping could lead to gene cloning, understanding plant biology at the molecular level and gene manipulation, as most agricultural organisms are expected to be sequenced in the near future.

Although high-throughput sequencing technologies are developing very rapidly, the ability to uncover genetic variation in large populations is still limited in many agricultural systems due to a lack of whole-genome reference sequences for comparison. In addition, resequencing of many individuals is still quite expensive and the analysis is very time consuming utilizing the existing technologies.

One of the tools that has been applied for high-throughput discovery of genetic variation is microarray single feature polymorphism (SFP). This approach detects differential hybridization of genomic DNA (gDNA) or RNA (Gupta et al. 2008) and uses the hybridization signal as a marker, without prior knowledge of the change that led to this difference. This approach, using gDNA for labeling, was first reported in yeast (Winzeler et al. 1998) and was then applied to a variety of different plants, such as *Arabidopsis,* rice, soybean, tomato and melon (Borevitz et al. 2003; Kaczorowski et al. 2008; Kumar et al. 2007; Ophir et al. 2010; Sim et al. 2009). A similar approach, using RNA for labeling, has been used in yeast (Ronald et al. 2005) and in a few plants, such as wheat, cowpea, barley and alfalfa (Bernardo et al. 2009; Das et al. 2008; Rostoks et al. 2005; Yang et al. 2009). Although the ability to map traits in plants using this technology was suggested several

years ago (Borevitz and Chory 2004; Zhu and Salmeron 2007), only a few examples of its use have been reported, in *Arabidopsis,* rice and soybean. Use of SFP technology for the discovery of genetic markers linked to traits is still limited by its high false discovery rate (FDR) in complex genomes (Gore et al. 2007; Kumar et al. 2007; Sim et al. 2009), the high cost of using many arrays, and the availability of arrays for only a few agricultural organisms.

Melon (*Cucumis melo* L.) is a highly polymorphic species comprising a broad range of wild and cultivated genotypes differing in fruit traits such as climactericity, sugar and acid content, secondary metabolites, taste, aroma, fruit shape, and flesh and rind color (Pitrat et al. 2000). Over the years, a few different genetic and genomic resources have been developed for melon. Different genetic maps have been constructed that comprise many diverse traits, including resistance to pathogens and fruit traits such as total soluble solids, fruit size and shape, climacteric ripening, netting, color, various metabolites, sugar metabolism, carotene and ethylene biosynthesis (Burger et al. 2009; Harel-Beja et al. 2010). On the genomic front, a large collection of expressed sequence tags (ESTs) has been integrated into the curcubit database (http://www.icugi.org), BAC libraries have been created (Luo et al. 2001) and a unified genetic map has been established (Diaz et al. 2011). However, the ability to use these for genetic mapping and cloning genes of interest is still quite limited. Only a few genes have been cloned by chromosomal walking in melon (Boualem et al. 2008; Joobeur et al. 2004; Nieto et al. 2006; Martin et al. 2009), and the use of high-throughput technologies to map traits is still in its early stages (Ophir et al. 2010). The cucumber genome, a close relative of melon, was published and can be used for comparative mapping (Huang et al. 2009).

In sweet melons, the main component of fruit quality is sugar content (sucrose being the major component). Organic acid levels play only a small role in determining fruit quality (Wang et al. 1996) as the pH of mature fruit is around 6.0. This phenomenon is very different from many other fruits, in which one of the main fruit quality components is determined by the combination of sugar and organic acid levels (Sweeney et al. 1970). In some other melon accessions such as a 'Faqqous', 'Chito', 'Conomon', the fruit pH is around 5.0 (Stepansky et al. 1999). These melons contain a high level of organic acids and their fruits are mostly consumed when they are still young, before organic acid accumulation (sour taste). The sour taste trait of fruit flesh has been reported to be dominant over non-sour taste (Kubicki 1962) and controlled by a single gene. In a recent work, this phenotype was redefined by measuring fruit flesh pH, with similar results (Danin-Poleg et al. 2002). We mapped the pH trait to linkage group 8 of melon (Danin-Poleg et al. 2002; Harel-Beja et al. 2010; Diaz et al. 2011). The pH trait is important for creating new varieties of melon with new taste combinations (Burger et al. 2003). In this work, we combine a self-designed microarray for melon (Ophir et al. 2010) with a bulk segregation approach (Michelmore et al. 1991) in order to fine map the pH trait. Using these tools, we successfully discovered a set of markers around the *pH* locus. We used our melon germplasm collection to confirm the value of this type of approach. The use of this approach and similar approaches for mapping (and cloning) of traits of interest by high-throughput genomic technologies and classical breeding in agricultural plants is discussed.

Materials and methods

Plant material

Three melon populations were analyzed: (1) A recombinant inbred (RI) population (designated 414xDul) (Harel-Beja et al. 2010) was developed from a cross between PI 414723 (S_5) (*Cucumis melo* var. *momordica*) and 'Dulce' (*C. melo* var. *reticulatus*) consisting of 98 RI lines (RILs) comprising a mix of F_6, F_7 and F_8 generations. (2) An F_3 segregating population was developed from the same cross consisting of 33 families (that were used for this study) and (3) a collection consisting of 50 diverse melon cultivars from different origins.

Evaluation of pH trait

The 414xDul RI population was evaluated as described by Harel-Beja (2010) in Newe Ya'ar, Israel, during the summers of 2005–2007. Each line was represented by 8, 12 and 10 plants in 2005, 2006 and 2007, respectively. The two parents were each grown in three replications of 10–12 plants, randomly distributed in the field. A single fruit per plant was harvested when the abscission layer developed. The F_3 families were evaluated based on their F_2 phenotype and markers selection of 10–20 plants/family and identification of homozygous plants. In the two populations, the pH values were obtained by squeezing the fruit flesh and measuring it using a pH meter (Danin-Poleg et al. 2002).

pH phenotype call

To reestablish the pH calls for the 414xDul RILs, we performed a non-parametric statistical test. Initially, to remove measurement bias among years, we scaled the data using quantile normalization. For the scaled values, we calculated the median of RILs overall years and used this as the expected parameter for low and high pH mixture (data not shown). For each line, we tested whether the

median of the line measurements is statistically different from the expected mixture by running a Wilcoxon-ranked test. If the median of the line measurements was significantly greater than the mixture, it was defined as high pH (D) (low acidity level). If the median was significantly smaller, the line was defined as low pH (PI) (high acidity level). A non-significant difference was set as not defined (ND) (Online resource 1). ND represents families that are still segregating for the pH trait.

Array design

The array was based on 16,637 Unigenes that were fetched from the melon database of the International Cucurbit Genomics Initiative (ICuGI; melon v2). The probes were designed to variable lengths, ranging between 45 and 55 bp, to be T_m optimized to 76.6 °C. The T_m, as well as other features of the probes such as cross hybridization, folding, and low complexity, were calculated using OligoWIZ 2.0 (Wernersson et al. 2007). The array in the present study was derived from a melon SFP custom array (Ophir et al. 2010) that was narrowed to 103,507 probes by filtering out some of the probes that did not contain a SFP in the previous study (Ophir et al. 2010) by hybridization to the two parents PI 414723 and 'Dulce'. The designed array was loaded onto E-array and printed on an Agilent 105K array (http://earray.chem.agilent.com/earray/).

Genomic DNA preparation and microarray hybridization

DNA was extracted from young leaf tissue from pools of ten plants (of F_3, RIL families and our germplasm collection) according to the preparation procedure described by Harel-Beja (2010). Pools of DNA were prepared by mixing equal amounts of DNA according to the phenotypic analysis and generations. Four different pools were prepared: F_3AA (16 families—low pH phenotype), F_3aa (17 families—high pH phenotype), F_8 RILs AA (34 families—low pH phenotype) and F_8 RILs aa (31 families—high pH phenotype). The DNA pools were treated with RNaseH (Sigma, USA) and purified using phenol/chloroform/isoamyl alcohol (25:24:1) followed by ethanol precipitation. Labeling of DNA pools and hybridization were performed at the Weizmann Institute's DNA array unit following Agilent CGH protocols for comparative genomic hybridization (http://www.agilent.com).

Microarray statistical analysis

Probe signal pre-processing and fitting were performed with the R-package LIMMA (Smyth 2004) following background subtraction. For within-array normalization, we applied the "lowess" method (Yang et al. 2002) and for between-array normalization, the "Aquantile" method (Yang and Thorne 2003). To select for statistical significance of signal ratios, two subsequent steps were taken: a least-square fitting to the linear model

$$\text{Log signal ratio} = \mu + L + NL$$

where L stands for linkage and is the ratio between the gDNA of the differential pools of dominant and recessive allele plants and NL stands for no linkage and is the ratio between biological replicates of gDNA within-recessive and dominant allele pool plants (Biological replicates F_3, F_8 with the same phenotype); then, after calculating the contrast $L - NL$ to remove false signals, such as those from copy number variation, the contrast fitting followed by an empirical Bayesian correction for better variance estimation were performed. Finally, a multiple comparison correction was performed on the P values using the Benjamini and Hochberg (1995) method.

Mapping the single nucleotide polymorphism (SNP) markers around the *pH* locus

Fourteen markers from the top 50 absolute moderated t values were chosen. The genomic area of these SFPs were PCR amplified with specific oligos and direct sequencing from the two parents of the mapping populations (PI 414723 and Dulce) were preformed identifying genetic variation (Online Resource 3). We used variation (SNPs) that were homozygous in the two parents. These SNPs were retested on the different F_8 RIL pools to validate their linkage to the pH trait using direct sequencing. SNP assays were performed by KASPar technology (KBioscience, England) on the RIL population and the germplasm collection. The RIL genotype and the pH call were integrated into a 13 (12 markers plus pH call, KASPar assay failed in two SNPs) × 98 (RILs) table (Online resource 4). Mapping was performed using JoinMap® 3.0 software (Van Ooijen and Voorrips 2001) integrating the new data set into our melon genetic map (Harel-Beja et al. 2010) by rebuilding the map with the original and the new markers. The new markers did not change the known order.

Mapping melon EST's to cucumber alignment

The 12 EST's harboring the genetic markers in melon were also aligned to the cucumber genome at ICuGI (http://www.icugi.org) by blast search (default parameters).

Linkage disequilibrium on melon germplasm

Fourteen markers were genotyped, by KASPar technology, for 50 melon cultivars from different backgrounds based on

phenotypic criteria (Pitrat et al. 2000) and pH values (Online Resource 5). A distance matrix was calculated by setting 1, 2 and 3 for alleles of high (dominant), heterozygote and low (recessive) acid levels, respectively (Online Resources 5) and then calculating Euclidean distance. This matrix was used to create an ordered heatmap by agglomerative hierarchical clustering using Ward's linkage method (Ward 1963) (Fig. 4).

Results

Establishing pH phenotype in the PI 414723 × Dulce RI population

The acid level (low or high) acts as a monogenic trait in melon (*Cucumis melo L.*) (Kubicki 1962; Danin-Poleg et al. 2002), with low pH being dominant over high pH. To reestablish this phenotype in the RI segregation population (PI 414723 × Dulce RI), we used the measurements of the RI population in three consecutive years (Harel-Beja et al. 2010). The distribution of pH measurements within the RI population is illustrated in Fig. 1. As references, we utilized fruits of the parents of the RI population, PI 414723 and 'Dulce', which were measured under the same conditions (Fig. 1, circles). Clearly, the pH distribution illustrates two separate modes with a range of 4.5–6.4 and a median of 5.21, as expected. We tested each line for its statistically significant difference from the overall median (Wilcoxon-rank test; $P < 0.05$). If the difference was significantly lower or higher than the overall median, we classified it as low pH (PI 414723) or high pH ('Dulce'), respectively. Otherwise, the pH call (listed in Online Resource 1) was set to undefined (mix of genotypes). This analysis verifies that pH behaves in a Mendelian segregation manner. We used these results in our experimental design and in the analysis of the F_3 families' results.

Bulk segregation analysis: experimental design

To isolate a set of markers linked to the *pH* locus, we based our experimental design on the combined use of microarray and bulk segregation analysis approach (Michelmore et al. 1991). The basic assumption in this approach is that microarrays can recapitulate information from a pool of families (Brauer et al. 2006; Wenger et al. 2010). Our F_3 families and RILs are progenies of a cross between two melon accessions: 'Dulce', which shows high pH in its mesocarp in mature fruits and PI 414723, which shows low pH in its mesocarp in mature fruits (Danin-Poleg et al. 2002). We pooled DNA from 16 F_3 families with low pH (dominant allele phenotype) and 17 F_3 families from high pH plants (recessive allele phenotype). As a biological

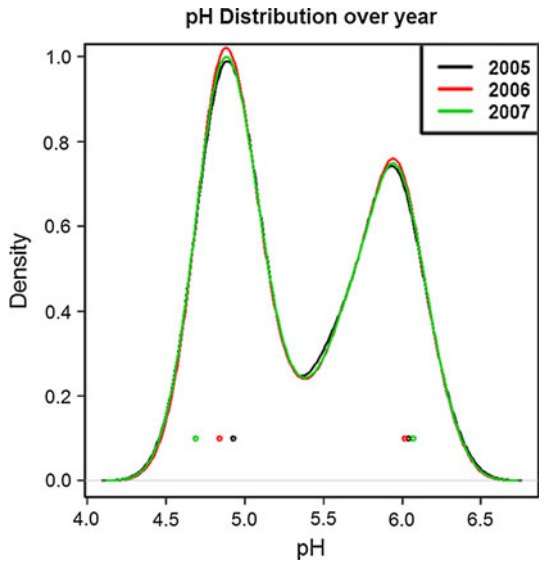

Fig. 1 Distribution of pH level in melon 414xDul RI population. The measurements show Mendelian segregation of two phenotypes (high pH—dominant allele, low pH—recessive allele).The pH of melon fruits was measured for 87 RILs (8, 12, 10 replicates/line per year) over 3 years (2005, 2006, 2007; *black*, *red*, *green*, respectively). The measurement method is described in "Materials and methods". The graph illustrates the density (kernel smoothing) of all measurements in each year after quintile normalization. The *y*-axis shows the proportion values and the *x*-axis, the pH values. The *circles* are the average measurements of the parents PI 414723 and 'Dulce', which were used as references. RILs pH call can be found in Online Resource 1 (color figure online)

replicate, we used bulk pools of the RILs (F_8) (Harel-Beja et al. 2010), 34 families from low pH and 31 families from high pH plants (DNA prepared from them as well). These four pools were labeled with Cy3 and Cy5 (Online Resource 2). The pools from the same generation were hybridized on the same array. Pools from a different generation with the same phenotype were hybridized to another array. We then hybridized these arrays competitively against each other in a loop design (Yang and Speed 2002), which optimizes the variance of comparisons in a factorial design. This is done by averaging over the dye effect and between arrays effect while performing direct comparisons for the contrasts under the study.

By executing the experiment in this design, we created two comparisons: one between dominant allele pools and recessive allele pools and the other between different generations (F_3 and F_8 RILs) within *pH* allele pools (recessive and dominant). The comparison between the dominant and recessive trait pools should exhibit increasing differences in signal, as the probes that are genetically in close proximity to the *pH* locus are expected to be of parental origin, due to the genetic linkage between the *pH* locus and the markers around it. A comparison between the dominant phenotypes (F_3 and F_8 RILs) and recessive phenotypes (F_3 and F_8 RILs) should detect no difference

when averaging within-dominant allele pools (F_3 and F_8 RILs) and within-recessive allele pools (F_3 and F_8 RILs), except for false positives (probably related to copy number variations, Online Resource 2). In fact, the later comparison is equivalent to calculating the expected allele frequency with no linkage.

Discovery of a set of pH-linked SFPs using bulk segregation analysis

We expected enrichment in homozygous alleles linked to the *pH* locus in both the dominant and recessive pools. Since the largest difference in signal would result from a comparison between two different homozygous alleles, we anticipated that the differential signal would increase as the SFP markers got closer to the *pH* locus (the signal is correlated with the genetic distance from the locus of interest (Borevitz et al. 2003). Therefore, we selected for 383 significant SFPs with higher absolute t-values, i.e., those expected to be closer to the trait (Fig. 2, black dots). We calculated the absolute moderated *t* value since the directionality of the differential signal is dependent on the probe sequence on the array, whereas the source of these probes is not unified (Ophir et al. 2010). To reduce the number of false positives, we subtracted the non-linked differential signal (Fig. 2, red dots) from the linked differential signal (Fig. 2, black dots). The non-linked fraction can be estimated by summing all families from the dominant and recessive lineages. Thus, averaging over the dominant and recessive allele pools (Online Resource 2) should reflect the genetic background with no linkage. The absolute moderated *t* values of linked differential to non-linked differential difference (Fig. 2, green dots) were sorted increasingly and the linked and non-linked moderated *t* values were sorted, respectively. Doing so, we would expect the curve to be closer to the theoretical log-odds score (Borevitz et al. 2003). Using this approach (Fig. 2), we can estimate that the markers are linked to the *pH* locus but not on their order on the chromosome as the melon genome was not published at the time of the research and therefore we cannot refer the markers to the physical map.

Mapping *pH*-linked markers to the melon genetic map and cucumber physical map

To test the pH-linked SFP markers identified by bioinformatics analysis (Fig. 2), we validated a list of markers with top moderated *t*-value scores (i.e., the ones closest to the trait) by PCR and direct sequencing on the parents of the mapping population, PI 414723, 'Dulce', and the different DNA pools that were used for the comparative hybridizations (Online Resource 2). Fourteen SFPs in different Unigenes that contained *pH*-linked markers (SNPs) were

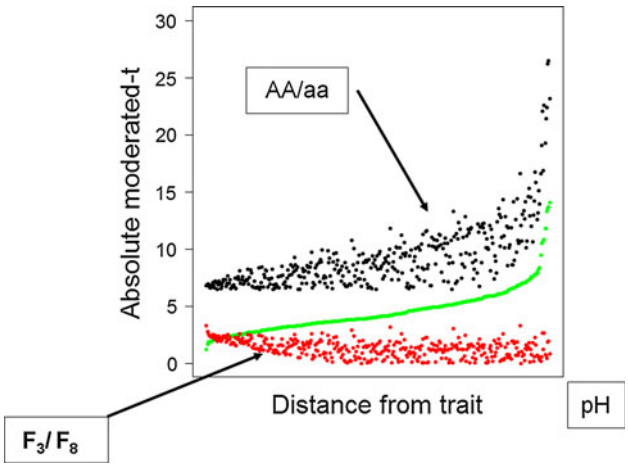

Fig. 2 The differential signal of low pH versus high pH phenotypes as a measure of linkage. The low pH (AA dominant)-to-high pH (aa recessive) signal ratio was used as a measure of linkage. As the SFP gets closer to the trait, the differential signal increases. The absolute moderated *t* value was calculated (*y*-axis) for low pH versus high pH hybridizations (average of F_3 and F_8 hybridizations, *black dots*) and the 383 statistically significant values are presented. The equivalent values of F_3-to-F_8 ratio (average over low pH and high pH phenotypes) are in *red*. Calculating the absolute moderated *t* contrast of low pH-to-high pH log ratio (linked, *black dots*) to the F_3-to-F_8 log ratio (not linked, *red dots*) results in a linked-to-non-linked log ratio (*green dots*). This is an empirical approximation of the log-odds score. All values are ranked by the linked to non-linked moderated *t* values (*x*-axis) (color figure online)

used for further analysis (Online Resource 4 and "Materials and methods"). SNP assays were developed based on KASPar technology (KBioscience): 98 F_8 RILs were genotyped using this set of 12 markers (two markers, MU3250 and MU7281, failed in the SNP assays on the RILs). The RILs genotyping data were utilized to map the SNPs (SFP markers) relative to known markers for the *pH* locus (CMAT141, MU45362, CMAG59, SYS_8.04 and CMCTTN181) using JoinMap 3.0 (Van Ooijen and Voorrips 2001) and based on the genetic map that we previously published (Harel-Beja et al. 2010). Eleven new markers (MU8350 defines as unlink to the *pH* locus) reside on both sides of the *pH* locus (Fig. 3a, Online Resource 3). As an additional test for our mapping in melon, we compared our results to the cucumber genome draft (Huang et al. 2009). We mapped by sequence similarity the Unigenes containing the SFPs and a known marker (CMCTTN181) around the *pH* locus on the physical cucumber map (http://www.icugi.org). Most of the markers from melon were mapped on cucumber chromosome 4, in close vicinity to each other, as three different groups with two large gaps between them (Fig. 3b). As expected (Huang et al. 2009; Wu and Tanksley 2010), the order of the markers was more conserved in the level of microsynteny, reinforcing the results of our pH-linked marker selection and genetic mapping of the *pH* locus in melon.

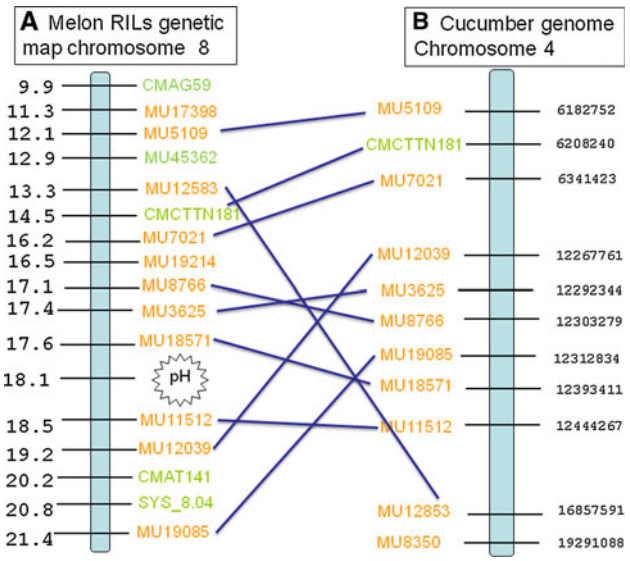

Fig. 3 Maps of pH-linked markers in melon and cucumber. **a** A genetic map of pH-linked markers was set relative to the pH trait (encapsulated by *star*) in melon. The order of the map was created based on genotyping 98 melon RILs using JoinMap® 3.0 software. Markers colored in *green* are known markers from the melon genetic map. Markers in *orange* are new markers based on this work. The numbers on the melon genetic map are in cM. *pH* locus (encapsulated by *star*) is marked based on the genetic mapping. **b** A synteny map of the same markers on the cucumber genome draft. The map was set by running of the corresponding UniGenes from cucumber against the cucumber genome (see "Materials and methods"). The *numbers* on the cucumber physical map are the genome positions in base pairs. The putative synteny is presented by *blue lines* (color figure online)

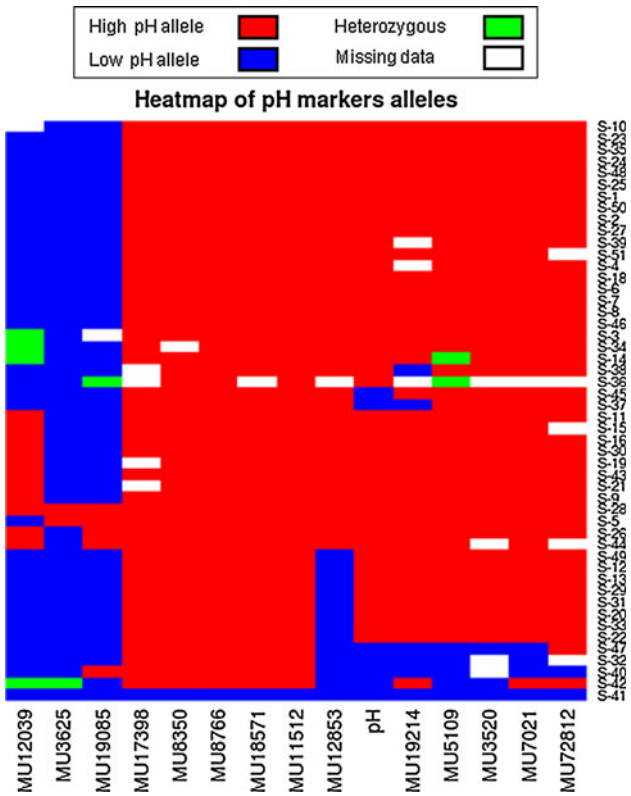

Fig. 4 Association study of a set of markers around the pH trait in a set of melon cultivars. Heatmap of 50 melon cultivars. A set of melon cultivars were genotyped for the 14 marker alleles (SNPs) using KASPar technology. Fruit flesh pH was measured using a pH meter. We classified the cultivars and the markers by two-way agglomerative hierarchical clustering. Five markers (MU5109, MU7021, MU19214, MU7281, and MU3520) showed strong linkage to the pH trait in most of the germplasm collection. For a full list of the melon germplasm collection cultivars used and their pH, see Online Resource 5. Low *pH* allele marked in *blue*. High *pH* allele marked in *red*. Heterozygous marked in *green*. No data marked in *white* (color figure online)

Nevertheless, cucumber and melon are not the same plant, and therefore many differences could be found at the macrosynteny level (Huang et al. 2009; Ren et al. 2009).

Transferability of the pH-linked markers

One of the features anticipated from a genetic marker that is tightly linked to a trait is to be stable and therefore the marker-trait linkage will be transferable within the germplasm collection. We challenged the transferability of our markers among our germplasm collection. Fifty cultivars of melons, classified as *Cucumis melo* sub sp. *melo* ($N = 40$) and *C. melo* subsp. *agrestis* ($N = 10$). Most of the cultivars were phenotyped as high pH (>5.346; $N = 43$) and the rest were phenotyped as low pH (<5.3; $N = 7$) (Online Resource 5). Fourteen markers were used to genotype the 50 cultivars (Fig. 4, Online Resource 5). PI 414723 (S-41) and "Dulce" (S-28) demonstrated full linkage between the different markers and the pH phenotype as expected. Few markers did not present any variation and were not informative (MU8359, MU8766, MU17938, MU18571, MU11512). Some of the markers lost the linkage with the *pH* locus. Several markers were strongly linked to the *pH* locus (MU19214, MU5109, MU3520, MU7021, MU7281) based on the heatmap (Fig. 4). Markers MU5109, MU7021 and MU19214 are linked to the pH trait in the north side of the melon map (Fig. 3a). Markers MU3520 and MU7281 SNP assay did not work on the RILs and are not present on the melon genetic map (Fig. 3a). A linkage disequilibrium test was not done as the number of low pH cultivars in the collection was too small for such a test (Online Resource 5).

Discussion

Combining our self-designed SFP array for melon with a bulk segregation approach, we succeeded in fine mapping the pH trait of melon fruit at very high accuracy with a limited number of arrays and limited sequence information. In addition, some of the validated markers demonstrated a strong linkage to the *pH* locus in a large set of germplasm cultivars demonstrating the value of this approach for plant

breeding. In this specific case, combining fruit with low pH and high sucrose creates melons with new taste (Burger et al. 2003).

The use of microarray for genotyping and mapping in agricultural organisms is underutilized due to a lack of genomic information for array design (an issue which will be resolved by high-throughput sequencing technologies), the lack of commercial arrays for crops that can be solved by self-design (Ophir et al. 2010) and high FDRs, even in medium-size plant genomes (Gore et al. 2007).

One of the main obstacles of the SFP approach is high FDRs, as high as 40 % in some studies (Gore et al. 2007). In our array, direct sequencing of the two parents found that FDR was around 19 % (Ophir et al. 2010). Different statistical methods allow us to control the FDR. High FDR is a problem when one would like to use these markers as a tool for creating a full genetic map or for large-scale marker discovery. In this case, reducing the high rate of false discoveries will leave only few markers for downstream analysis. In a case of genetic marker-trait linkages, the density of the SFP markers is more important than their discovery rate: as one should have only a few reliable markers from both sides of the trait and a true positive rate of 80 % is more than sufficient, as demonstrated in our mapping experiment (Fig. 3). Although some of the markers could not be evaluated (failed in the SNP assay development), a few were successfully mapped to the close flanking vicinity of the trait (Figs. 2, 3a). Similar results were presented in a soybean study using SFPs for linkage analysis (Kaczorowski et al. 2008). The cucumber synteny mapping was robust and most of the markers align on cucumber chromosome 4 in few different blocks (probably as a result of rearrangement) (Fig. 3b), confirming the accuracy of the melon genetic mapping. One can speculate that a similar *pH* locus exist in the cucumber genome between markers MU11512 and MU18571 (Fig. 3a, b). As these two markers are very close to each other in melon genetic map; cucumber physical map and the *pH* locus is found between them in melon (Fig. 3a). By combining accurate mapping with a full melon genome, one might envision being able to predict gene candidates for the trait in silico. We suggest that these types of experiments can tolerate much higher than expected FDRs that are part of this technology (gene families, technical reasons and bioinformatics analysis) as they are used as a discovery tool and are validated by direct mapping (Fig. 3; Rafalski 2002), rather than as a basis for the construction of detailed genetic maps that cannot tolerate such high FDRs.

Efficient genetic mapping requires the ability to screen large sets of markers in a cost-efficient manner. This is also true for genome-wide association studies (GWASs), which are broadly used in mammalian systems but have only recently been implemented in plants (Aranzana et al. 2005; Gupta et al. 2005; Chiang et al. 2010; Rafalski 2002). Although many technologies that seem to be more accurate than SFP arrays are available, such as Illumina bead array or even some combinations of direct sequencing (Walsh et al. 2010), the SFP approach still has one major advantage: it can be used for variation discovery without prior knowledge of the variation at a reasonable price (this can also be done by direct sequencing, but the cost is prohibitive). This is relevant to many agricultural systems for which sequence information is still limited, and for the breeding of elite cultivars with a very narrow genetic base (as full sequencing of organisms over and over again is not practical). Usually, when linkage studies are done on specific segregating population, many markers are not applicable for the entire germplasm. The effectiveness of the genome-wide SFP approach is presented in this study by the fact that, among the many markers that were found for PI 414723 × 'Dulce' population, we found a few markers that were linked to the *pH* locus in most of our germplasm collection. These markers can be used for introgression of the pH trait from PI 414723 and other low pH cultivars into elite cultivars. We assume that in some cases the linkage was separated or maybe another locus influenced the pH phenotype (Fig. 4).

Utilizing the bulk segregation approach, as has been done in many biological systems (Harris et al. 2009; Kang and Mian 2010; Pandit et al. 2010; Zhang et al. 2010), makes the experiments much more cost effective. Moreover, pooling reduces the noise due to recapitulation of background alleles. Surprisingly, the combined approach of bulk segregation in microarrays is underrepresented in the literature, with its use having been reported mainly in yeast and *Arabidopsis* (Borevitz et al. 2003; Borevitz and Nordborg 2003; Brauer et al. 2006; Ehrenreich et al. 2010; Wenger et al. 2010) and more recently in a bulk expression analysis for mapping a quantitative trait locus (QTL) in potato (Kloosterman et al. 2010). Borevitz et al. (2003) employed bulk segregation analysis to compare two very small pools of F_2 (dominant and recessive alleles) plants. The likelihood ratio was then calculated by simulation studies and modeling the markers in both F_2 pools, and was validated against a known genome. However, the model could only be applied when the array design was based on one of the parents. We present a simpler approach that can be applied to any array. We use a linear model to estimate the marker linkage to the *pH* locus by having two sets of pools, from the dominant allele phenotype and from the recessive allele phenotype. To estimate the expected variance of non-linked alleles (heterozygous), we average the F_3 pool-to-F_8 pool ratio of both phenotypes. Utilizing this approach, we expect to approximate the theoretical model of expected values in log likelihood ratio. Although in our case, we could use a pool of F_8 lines as a replicate for the

F_3 family pool, we realize that such germplasm collection is not always available. Alternatively, one can simply create more than one set of F_3 pools.

One interesting question is whether the scope of the associations between the markers and the pH trait is wider than the segregating population. The transferability of the markers polymorphism was reported to be 54–66 % in barley genera *H. chilense* (Castillo et al. 2008) and 30–45 % in *Citrus* among *Citrus* species (Luro et al. 2008). In this study, we surveyed the transferability of marker linkage rather than marker polymorphism among our *Cucumis melo* germplasm collection. Of the 14 markers that were genotyped in 50 cultivars, 5 were tightly linked to the pH trait (Fig. 4). These markers were not found most closely linked to the *pH* locus in the melon genetic map (Fig. 3a). One explanation for these results is that many of the markers we found were unique to the cross we used PI 414723 (S_5) × 'Dulce' (Fig. 4). The ability to identify few markers that still associate with the pH trait in our germplasm collection enhances our confidence in the mapping around the *pH* locus and proves the utility of these markers in different cultivars. A linkage disequilibrium analysis would have given an indication to the level of the linkage of these markers, however, the germplasm collection is highly biased to the high pH cultivars and we anticipate that the results would not be indicative enough.

Mapping genetic traits in plants is extremely important for breeding and for exploring plant biology. The pH trait is an example of a monogenic trait with wide implications for fruit quality (Burger et al. 2003). As many of the traits are more complex, being composed of a few loci that influence phenotype (QTLs), the question that arises is whether SFP and similar approaches can be performed for these traits as well. There are several reports of SFPs used to locate QTLs in yeast (DeCook et al. 2006; Gupta et al. 2008; Kim et al. 2006; Marullo et al. 2007; West et al. 2006). We assume that the SFP tool can have some advantage in QTL linkage in plants as it has the capacity to show linkages to several different loci simultaneously, identifying few QTL's in the same time.

Acknowledgments This work was supported by grant from ARO manager fund; Binational Agriculture Research and Development (BARD) Grant IS-4341-10. This paper is journal series #107-12 of the Agricultural Research Organization. The authors want to thank to David Pilzer and Shirley Horn-Saban from the Weizmann institute for their excellent work in processing the microarrays.

References

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:e60

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57:289–300

Bernardo A, Bradbury P, Ma H, Hu S, Bowden R, Buckler E, Bai G (2009) Discovery and mapping of single feature polymorphisms in wheat using Affymetrix arrays. BMC Genomics 10:251

Borevitz JO, Chory J (2004) Genomics tools for QTL analysis and gene discovery. Curr Opin Plant Biol 7:132–136

Borevitz JO, Nordborg M (2003) The impact of genomics on the study of natural variation in Arabidopsis. Plant Physiol 132:718–725

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res 13:513–523

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari MA, Collin F, Flowers JM, Pitrat M, Purugganan MD, Dogimont C, Bendahmane A (2008) A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science 321:836–838

Brauer MJ, Christianson CM, Pai DA, Dunham MJ (2006) Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. Genetics 173:1813–1816

Burger Y, Saar U, Katzir N, Paris HS, Yeselson Y, Levin I, Schaffer AA (2003) Development of sweet melon (*Cucumis melo*) genotypes combining high sucrose and organic acid content. J Am Soc Hortic Sci 128:537–540

Burger Y, Paris HS, Cohen R, Katzir N, Tadmor Y, Lewinsohn E, Schaffer AA (2009) Genetic diversity of *Cucumis melo*. Hortic Rev 36:165–198

Castillo A, Budak H, Varshney RK, Dorado G, Graner A, Hernandez P (2008) Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. BMC Plant Biol 8:97

Chiang CW, Gajdos ZK, Korn JM, Kuruvilla FG, Butler JL, Hackett R, Guiducci C, Nguyen TT, Wilks R, Forrester T, Haiman CA, Henderson KD, Le Marchand L, Henderson BE, Palmert MR, McKenzie CA, Lyon HN, Cooper RS, Zhu X, Hirschhorn JN (2010) Rapid assessment of genetic ancestry in populations of unknown origin by genome-wide genotyping of pooled samples. PLoS Genet 6:e1000866

Danin-Poleg Y, Tadmor Y, Tzuri G, Reis N, Hirschberg J, Katzir N (2002) Construction of a genetic map of melon with molecular markers and horticultural traits, and localization of genes associated with ZYMV resistance. Euphytica 125:373–384

Das S, Bhat P, Sudhakar C, Ehlers J, Wanamaker S, Roberts P, Cui X, Close T (2008) Detection and validation of single feature polymorphisms in cowpea (*Vigna unguiculata* L. Walp) using a soybean genome array. BMC Genomics 9:107

DeCook R, Lall S, Nettleton D, Howell SH (2006) Genetic regulation of gene expression during shoot development in *Arabidopsis*. Genetics 172:1155–1164

Diaz A, Fergany M, Formisano G, Ziarsolo P, Blanca J, Fei Z, Staub JE, Zalapa JE, Cuevas HE, Dace G, Oliver M, Boissot N, Dogimont C, Pitrat M, Hofstede R, van Koert P, Harel-Beja R, Tzuri G, Portnoy V, Cohen S, Schaffer A, Katzir N, Xu Y, Zhang H, Fukino N, Matsumoto S, Garcia-Mas J, Monforte AJ (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

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464:1039–1042

Gore M, Bradbury P, Hogers R, Kirst M, Verstege E, Oeveren Jv, Peleman J, Buckler E, Eijk Mv (2007) Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. Crop Sci 47:S135–S148

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18

Harel-Beja R, Tzuri G, Portnoy V, Lotan-Pompan M, Lev S, Cohen S, Dai N, Yeselson L, Meir A, Libhaber SE, Avisar E, Melame T, van Koert P, Verbakel H, Hofstede R, Volpin H, Oliver M, Fougedoire A, Stalh C, Fauve J, Copes B, Fei Z, Giovannoni J, Ori N, Lewinsohn E, Sherman A, Burger J, Tadmor Y, Schaffer AA, Katzir N (2010) A genetic map of melon highly enriched with fruit quality QTLs and EST markers, including sugar and carotenoid metabolism genes. Theor Appl Genet 121:511–533

Harris KR, Ling KS, Wechter WP, Levi A (2009) Identification and utility of markers linked to the zucchini yellow mosaic virus resistance gene in watermelon. J Am Soc Hortic Sci 134:529–534

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan WuZ, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Fang L, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Li S, Zhang X, Wang J, Sun R, Zhang B, Jiang S, Du Y (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275–1281

Joobeur T, King JJ, Nolin SJ, Thomas CE, Dean RA (2004) The Fusarium wilt resistance locus *Fom-2* of melon contains a single resistance gene with complex features. Plant J 39:283–297

Kaczorowski KA, Ki-Seung Kim KS, Diers BW, Hudson ME (2008) Microarray-based genetic mapping using soybean near-isogenic lines and generation of snp markers in the *Rag1* aphid-resistance interval. Plant Genome 1:89–98

Kang ST, Mian MAR (2010) Genetic map of the powdery mildew resistance gene in soybean PI 243540. Genome 53:400–405

Kim S, Zhao K, Jiang R, Molitor J, Borevitz JO, Nordborg M, Marjoram P (2006) Association mapping with single-feature polymorphisms. Genetics 173:1125–1133

Kloosterman B, Oortwijn M, Uitdewilligen J, America T, de Vos R, Visser RG, Bachem CW (2010) From QTL to candidate gene: genetical genomics of simple and complex traits in potato using a pooling strategy. BMC Genomics 11:158

Kubicki B (1962) Inheritance of some characters in muskmelons (*Cucumis melo*). Genet Polonica 3:265–274

Kumar R, Qiu J, Joshi T, Valliyodan B, Xu D, Nguyen HT (2007) Single feature polymorphism discovery in rice. PLoS One 2:e284

Luo MZ, Wang YH, Frisch D, Joobeur T, Wing RA, Dean RA (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon Fusarium wilt (Fom-2). Genome 44:154–162

Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R (2008) Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina Hort ex Tan*) to other Citrus species and their effectiveness for genetic mapping. BMC Genomics 9:287

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461:1135–11388

Marullo P, Aigle M, Bely M, Masneuf-Pomarede I, Durrens P, Dubourdieu D, Yvert G (2007) Single QTL mapping and nucleotide-level resolution of a physiologic trait in wine *Saccharomyces cerevisiae* strains. FEMS Yeast Res 7:941–952

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Nieto C, Morales M, Orjeda G, Clepet C, Monfort A, Sturbois B, Puigdomenech P, Pitrat M, Caboche M, Dogimont C, Garcia-Mas J, Aranda MA, Bendahmane A (2006) An *eIF4E* allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. Plant J 48:452–462

Ophir R, Eshed R, Harel-Beja R, Tzuri G, Portnoy V, Burger Y, Uliel S, Katzir N, Sherman A (2010) High-throughput marker discovery in melon using a self-designed oligo microarray. BMC Genomics 11:269

Pandit A, Rai V, Bal S, Sinha S, Kumar V, Chauhan M, Gautam RK, Singh R, Sharma PC, Singh AK, Gaikwad K, Sharma TR, Mohapatra T, Singh NK (2010) Combining QTL mapping and transcriptome profiling of bulked RILs for identification of functional polymorphism for salt tolerance genes in rice (*Oryza sativa* L.). Mol Genet Genomics 284:121–136

Pitrat M, Hanelt P, Hammer K (2000) Some comments on infraspecific classification of cultivars of melon. Acta Hortic 510:29–36

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Ren Y, Zhang Z, Liu J, Staub JE, Han Y, Cheng Z, Li X, Lu J, Miao H, Kang H, Xie B, Gu X, Wang X, Du Y, Jin W, Huang S (2009) An integrated genetic and cytogenetic map of the cucumber genome. PLoS One 4:e5795

Ronald J, Akey JM, Whittle J, Smith EN, Yvert G, Kruglyak L (2005) Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. Genome Res 15:284–291

Rostoks N, Borevitz JO, Hedley PE, Russell J, Mudie S, Morris J, Cardle L, Marshall DF, Waugh R (2005) Single-feature polymorphism discovery in the barley transcriptome. Genome Biol 6:R54

Sim SC, Robbins MD, Chilcott C, Zhu T, Francis DM (2009) Oligonucleotide array discovery of polymorphisms in cultivated tomato (*Solanum lycopersicum* L.) reveals patterns of SNP variation associated with breeding. BMC Genomics 10:466

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3:Article 3

Stepansky A, Kovalski I, Perl-Treves R (1999) Intraspecific classification of melons (*Cucumis melo L.*) in view of their phenotypic and molecular variation. Plant Syst Evol 217:313–332

Sweeney JP, Chapman VJ, Hepner PA (1970) Sugar, acid, and flavor in fresh fruits. J Am Diet Assoc 57:432–435

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen

Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB, Swisher EM, King MC (2010) Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci USA 107:12629–12633

Wang Y, Wyllie SG, Leach DN (1996) Chemical Changes during the Development and Ripening of the Fruit of Cucumis melo (Cv. Makdimon). J Agric Food Chem 44:210–216

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:236–244

Wenger JW, Schwartz K, Sherlock G (2010) Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. PLoS Genet 6:e1000942

Wernersson R, Juncker AS, Nielsen HB (2007) Probe selection for DNA microarrays using OligoWiz. Nat Protoc 2:2677–2691

West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge RW, St Clair DA, Michelmore RW (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in *Arabidopsis*. Genome Res 16:787–795

Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ, McCusker JH, Stevens DA, Wodicka L, Lockhart DJ, Davis RW (1998) Direct allelic variation scanning of the yeast genome. Science 281:1194–1197

Wu F, Tanksley SD (2010) Chromosomal evolution in the plant family Solanaceae. BMC Genomics 11:182

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet 4:e1000212

Yang YH, Speed T (2002) Design issues for cDNA microarray experiments. Nat Rev Genet 3:579–588

Yang YH, Thorne NP (2003) Normalization for two-color cDNA microarray data. Science and Statistics In: Goldstein DR (ed) A festschrift for terry speed-monograph series, Chapman and Hall/CRC press virginia beach, VA, USA, pp 403–418

Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 30:e15

Yang SS, Xu WW, Tesfaye M, Lamb JFS, Jung H-JG, Samac DA, Vance CP, Gronwald JW (2009) Single-feature polymorphism discovery in the transcriptome of tetraploid alfalfa. Plant Genome 2:224–232

Zhang XM, Wu JA, Guo AG, Zhang H, Ma YA, Fang ZY, Wang XW (2010) Molecular mapping of MS-cd1 gene in Chinese kale. Afr J Biotechnol 9:4550–4555

Zhu T, Salmeron J (2007) High-definition genome profiling for genetic marker discovery. Trends Plant Sci 12:196–202