

RESEARCH ARTICLE

Open Access

Mapping in the era of sequencing: high density genotyping and its application for mapping TYLCV resistance in *Solanum pimpinellifolium*

Marcela Viquez-Zamora^{1,2,3†}, Mylaska Caro^{1,3†}, Richard Finkers^{1,2}, Yury Tikunov^{1,2}, Arnaud Bovy^{1,2}, Richard GF Visser^{1,2}, Yuling Bai¹ and Sjaak van Heusden^{1,2*}

Abstract

Background: A RIL population between *Solanum lycopersicum* cv. MoneyMaker and *S. pimpinellifolium* G1.1554 was genotyped with a custom made SNP array. Additionally, a subset of the lines was genotyped by sequencing (GBS).

Results: A total of 1974 polymorphic SNPs were selected to develop a linkage map of 715 unique genetic loci. We generated plots for visualizing the recombination patterns of the population relating physical and genetic positions along the genome.

This linkage map was used to identify two QTLs for TYLCV resistance which contained favourable alleles derived from *S. pimpinellifolium*. Further GBS was used to saturate regions of interest, and the mapping resolution of the two QTLs was improved. The analysis showed highest significance on Chromosome 11 close to the region of 51.3 Mb (*qTy-p11*) and another on Chromosome 3 near 46.5 Mb (*qTy-p3*). Furthermore, we explored the population using untargeted metabolic profiling, and the most significant differences between susceptible and resistant plants were mainly associated with sucrose and flavonoid glycosides.

Conclusions: The SNP information obtained from an array allowed a first QTL screening of our RIL population. With additional SNP data of a RILs subset, obtained through GBS, we were able to perform an *in silico* mapping improvement to further confirm regions associated with our trait of interest. With the combination of different ~ omics platforms we provide valuable insight into the genetics of *S. pimpinellifolium*-derived TYLCV resistance.

Keywords: SNPs, *S. pimpinellifolium*, *In silico*, TYLCV, Flavonoids, Hexose, Genotype by sequencing (GBS)

Background

Solanum pimpinellifolium is a source for introgression breeding in tomato (*S. lycopersicum*). This species is one of the closest wild relatives of *S. lycopersicum*, and it is present in the pedigree lineage of some commercial cultivars such as the sequenced 'Heinz 1706' [1]. Linkage maps from crosses between *S. lycopersicum* and *S. pimpinellifolium* were generated by various researchers [2-8]. Their work represents a small piece of the successful

use of genome-wide linkage analyses to map underlying genetic factors of traits between the two species.

Recombinant inbred lines (RILs) derived from inter-specific crosses consist of individuals with parental mosaics and are an efficient resource for mapping quantitative trait loci (QTL) [9]. Genotyping with molecular markers allows the visualization of recombination patterns which is crucial for the elucidation of loci associated with segregating traits [10,11]. This has become more efficient due to the availability of vast numbers of markers such as single nucleotide polymorphisms (SNPs). In tomato, the availability of high throughput SNP arrays allows massive parallel whole-genome screening of genotypes [8,12].

Nowadays, next generation sequencing technologies are offering new ways to increase genotyping throughput by several orders of magnitude [13]. Even more, it is

* Correspondence: sjaak.vanheusden@wur.nl

†Equal contributors

¹Wageningen UR Plant Breeding, Wageningen University & Research Centre, P.O. Box 386, Wageningen, AJ 6700, the Netherlands

²Centre for Biosystems Genomics, P.O. Box 98, Wageningen, AB 6700, the Netherlands

Full list of author information is available at the end of the article

possible to combine different genotyping platforms to increase the power of the analyses. Furthermore, due to published complete tomato genomes [1], next generation re-sequencing approaches can be applied in related germplasm [14]. Studies on evolutionary and domestication, as well as the genetic basis underlying important traits can be benefited from these genomic tools [15].

TYLCV is the causal agent of an aggressive tomato disease that can result in production losses up to one hundred percent, and its rapid spread worldwide is threatening the production of tomatoes. Development of TYLCV resistant tomato cultivars is an important strategy to avoid the damage caused by TYLCV. However, no TYLCV resistance has been identified in the cultivated tomato germplasm, except for the resistance allele of *ty-5* which is possibly originated from a mutation in the cultivated tomato [16]. Breeding for resistance to TYLCV has been focused on the introgression of tolerance or resistance genes from tomato wild relatives such as *S. pimpinellifolium*, *S. chilense*, *S. habrochaites* and *S. peruvianum* [17,18]. Several *S. pimpinellifolium* accessions are known to confer resistance to the virus [19-24], but attempts to map the causal factor in this species were not very successful. Thus, *S. pimpinellifolium*-derived TYLCV resistance is currently not well-exploited in tomato breeding programs [25]. In our study we genotyped a RIL population between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554 with a custom made SNP array [12], and a subset of 60 lines was also genotyped by sequencing using Illumina HiSeq 2000 (150 Tomato Genome ReSequencing project; www.tomatogenome.net). Furthermore, we explored the population with an untargeted metabolic profiling and compared resistant vs. susceptible lines in order to get more insights on compounds that might play a role in the resistance. Our study shows how we can combine different omics approaches to identify genetic loci underlying resistance to Tomato Yellow Leaf Curl Virus (TYLCV) in *S. pimpinellifolium* using a RIL population.

Methods

Recombinant Inbred Lines (RILs)

From a cross between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554 (CGN reference CGN 15528) a set of 100 RILs was generated through single seed descent (SSD) until the sixth generation [26]. These RILs, which have been used for many different experiments e.g. Khan *et al.* [27], were used in this study.

DNA extraction

Genomic DNA from young leaflets was extracted using a CTAB based protocol [28,29] adjusted for high throughput isolation. Two young leaflets were ground with a Retsch 300 mm shaker (Retsch BV, Ochten, The Netherlands)

using 1 ml micronic tubes (Micronic BV, Lelystad, The Netherlands). DNA pellets were washed in 76% EtOH with 10 mM NH₄Ac before re-suspending the DNA in TE buffer.

Genome wide genotyping

Genome wide genotyping was done as described by Viquez-Zamora *et al.* [12]. In short, DNA samples were sent to ServiceXS (<http://www.servicexs.com/>), Leiden, The Netherlands. A custom made Infinium HD Ultra Assay protocol [30] was used for hybridization onto a BeadChip. The Genotyping Module 1.9.4 of Illumina's GenomeStudio® V2011.1 software package was used to analyse the genotyping results under default settings. All samples corresponding to the RIL population and the parents were selected for a separate analysis in which manual inspection and adjustment were performed in order to discard questionable SNPs for the population and to optimize call rates. All polymorphic SNPs for the RIL population were named after their position on the SL2.40 version (<http://solgenomics.net/>) of the tomato genome sequence published online [1].

Genotype by sequencing (GBS)

A subset of 60 lines was selected for resequencing (lines with extreme values for TYLCV resistance were included). Whole genomic DNA was isolated from each line (see above). Shallow sequencing of 500 bp inserts was carried out using Illumina HiSeq 2000 (100 bp paired end reads) at an average coverage of 3x. Bases with Q < 20 were trimmed before read mapping with BWA [31,32] against the SL2.40 genome sequence of *S. lycopersicum* cv. Heinz with a maximum insert size of 750 bp (50% deviation), reporting at most 30 hits and removing PCR duplicates. SAMTOOLS [31] was used for variant calling without skipping InDels and a minimum gap distance of 5 bp. In addition, GATK [33], was used to call variants for all 60 genotypes in one single analysis.

The JBrowse by Skinner *et al.* [34] was used for the embedding and visualization of the SNP variants. The available gene models (ITAG 2.3) were obtained from the Sol Genomics Network (<http://solgenomics.net/>). Subsequently, a script was generated in order to combine the information of SNPs within the RILs. Access to the JBrowse with the information of the sequences can be obtained through: http://www.tomatogenome.net/ril_variants. Furthermore, the program Marker2sequence [35] was used to look for genes between specific genome coordinates based on their annotation.

TYLCV screening

Virus inoculation

Agrobacterium-mediated inoculation was performed to infect plants with TYLCV. Plantlets at the 3–4 leaf stage

(approximately 21 days after sowing) were inoculated with *A. tumefaciens* LBA4404 bearing a tandem repeat of an infectious TYLCV-IL (Israel isolate) clone. Bacterial growth was performed as previously described by Verlaan *et al.* [36] and bacteria were injected into true leaves using syringes without needle. Plants were grown under greenhouse conditions at 23°C, 60% humidity and 16-h/8-h day/night cycle.

Disease test

Disease symptoms were recorded 20, 25, 35 and 45 days post inoculation. Plants were scored for symptom severity according to the scale described by Friedmann *et al.* [37]. A first screening of the RILs was conducted using one plant per line. Thereafter, a second screening followed for the RILs classified as resistant to confirm the phenotype where four plants per resistant line were assessed. TYLCV disease symptoms rating was: 0 = no visible symptoms, inoculated plants show same growth and development as non-inoculated plants; 1 = very slight yellowing and minor curling of leaflet margins on apical leaf; 2 = some yellowing and minor curling of leaflet ends; 3 = a wide range of leaf yellowing, curling and cupping, with some reduction in size, yet plants continue to develop; 4 = very severe plant stunting and yellowing, and pronounced cupping and curling; plants cease to grow (Additional file 1: Figure S2).

Metabolic profiling

The RIL population was grown in triplicate under the same greenhouse conditions. Seven weeks after sowing, fully developed leaves were detached and main veins were removed. Samples were frozen in liquid nitrogen and thereafter ground into fine powder.

Untargeted metabolic profiling of leaves was performed with three platforms: 1) Liquid chromatography (LC), using a C18-reversed phase column, coupled to a Quadrupole-time-of-flight (TOF) mass spectrometer (MS) and a photodiode array detector (PDA) to detect semi-polar compounds such as flavonoids, alkaloids, phenylpropanoids, saponins, phenolic acids and polyamines according to De Vos *et al.* [38]. 2) Gas chromatography (GC) coupled to electron impact time of flight (TOF)-MS for detection of primary metabolites according to Lisec *et al.* [39]. 3) Solid phase microextraction (SPME)-GC-MS for the analysis of volatiles according to Tikunov *et al.* [40].

Metabolomics data processing

Metabolites were quantified and identified according to Tikunov *et al.* [41]. Each dataset was processed using MetAlign (www.metalignment.nl) for baseline correction, noise estimation, and ion-wise mass spectral alignment of the corresponding chromatograms. MSclust software was

used to extract compounds mass spectra and for data reduction [42].

The putative identification of metabolites was based upon their spectra, retention time, molecular weight and fragmentation patterns. For LC-MS data, compound characteristics were analysed and compared using the Dictionary of Natural Products (<http://dnp.chemnetbase.com>) and in-house tomato metabolite databases. GC-MS data were annotated using the NIST Mass Spectral Search Program v2.0 (<http://chemdata.nist.gov/mass-spc/ms-search/>) by matching mass spectra extracted to the NIST mass spectra collection and the Golm Metabolome Database (<http://gmd.mpimp-golm.mpg.de/>) for mass spectra matching followed by retention index comparison.

Linkage analysis

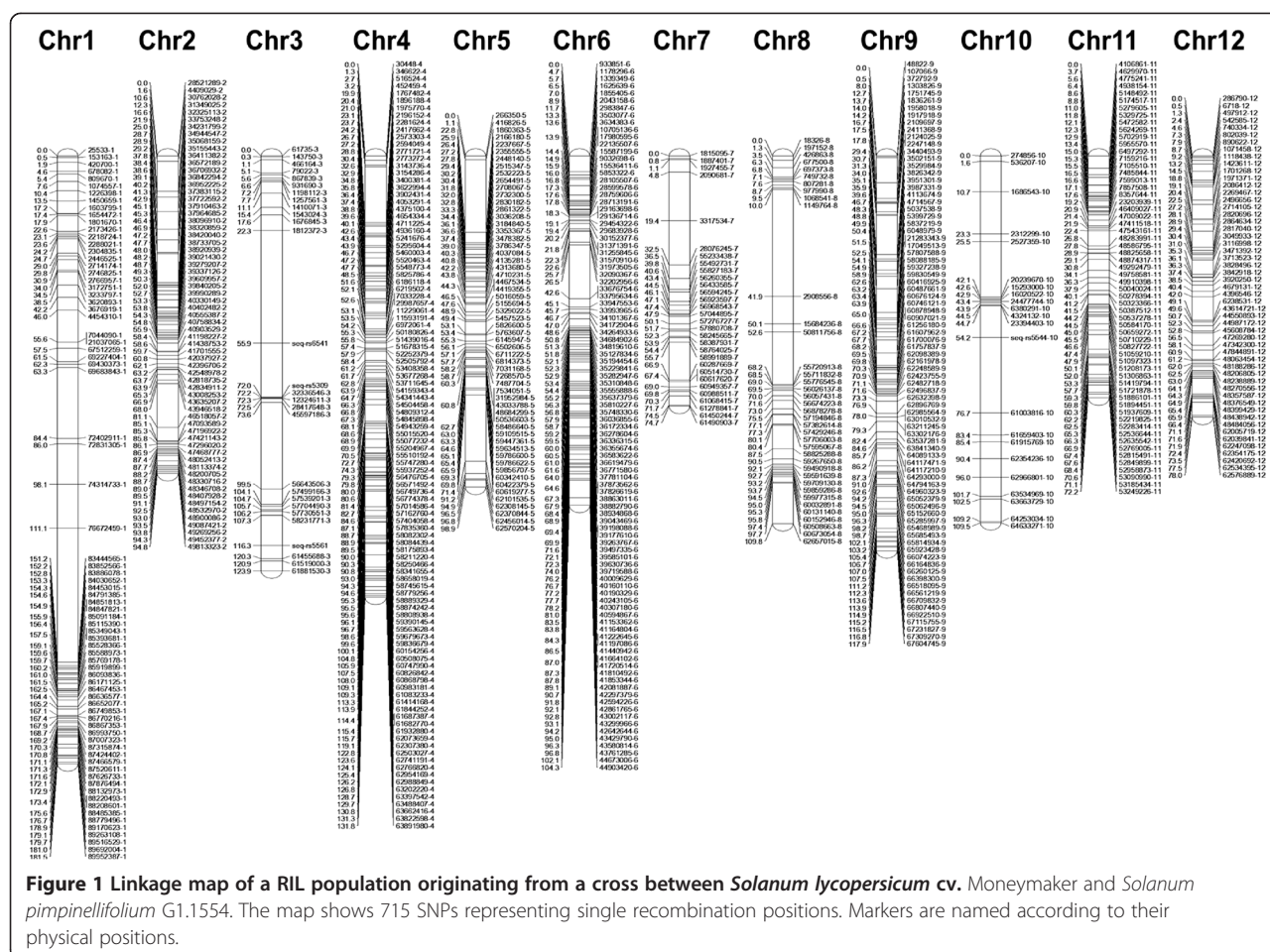
Linkage maps were constructed using JoinMap[®] 4.1 (Kiyazma[®]: <http://www.kiyazma.nl/>, [43]) with the specifications by Viquez-Zamora *et al.* [12] using the Haldane's mapping function. Genetic linkage groups were compared to the physical maps based on the tomato genome version SL2.40 using MapChart 2.2 [44]. The software GenStat 16th edition was used to perform mapping of QTLs for TYLCV resistance and the MapQTL software was used to map metabolite QTLs (mQTLs). The genotypic and phenotypic information is available at: http://www.plantbreeding.wur.nl/Publications/SNP/RILs_genotype-TYLCVphenotype.xlsx. Identified QTLs for TYLCV resistance were named according to their chromosomal position as in Kadirvel *et al.* [45]; *qTy-p3* and *qTy-p11* (*p* as from *S. pimpinellifolium*) for QTL on Chromosomes 3 and 11, respectively. The Marker2sequence application was used to mine regions for candidate genes [35].

Furthermore, the information of the sequences was embedded into JBrowse 1.11.1 [34] to visualize the detected structural variants. The SL2.40 tomato genome assembly and ITAG 2.31 tomato genome annotation was loaded together with the BAM and VCF files of the 60 genotypes.

Results

Linkage map and genome-wide visualizations

A custom made SNP Array was assembled from polymorphisms mainly found between two cherry and two round tomatoes [12]. This array was used to genotype a RIL population between *S. lycopersicum* cv. Money-maker and *S. pimpinellifolium* G1.1554. A total of 1974 polymorphic SNPs were identified between the parents. These SNPs were used to develop a linkage map based on their segregation patterns among the 100 RILs. The resulting map included 715 loci with an average distance of 1.85 cM between loci (Figure 1). The greatest gap was approximately 40 cM on Chromosome 1 and covered the region between 76 and 83 Mb.



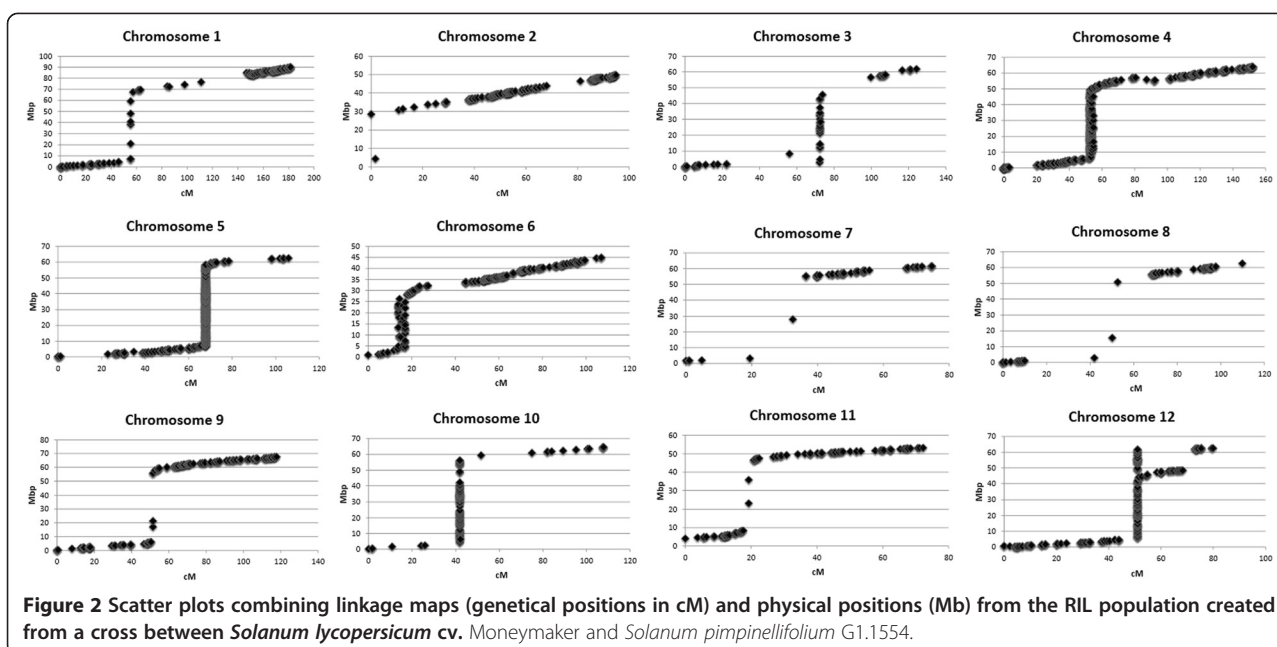
In order to visualize the recombination patterns along each chromosome, the physical positions of the SNP markers were determined using the published tomato genome [1]. For each SNP and its flanking sequence, a BLAST was performed to the genome sequence version SL2.40. Except for markers on chromosome 12, colinear orders were observed between the genetic and physical maps, as shown in scatter plots per chromosome between the linkage (cM) and physical map (Mb) (Figure 2). These scatter plots further allowed the visualization of cold- and hot-spots of recombination. When a large physical distance corresponds to only a small difference in cM, we can assume cold-spots of recombination. These cold-spots were always the heterochromatin pericentromeric regions and could be as long as 50 to 80 Mb. In contrast, hot-spots of recombination could be present if there is a large cM difference corresponding to small physical distance between markers.

The mosaic pattern of each RIL was calculated and composition of lines varied between 20% and 80% of alleles coming from each parent. In addition, we calculated the SNP allele frequency within the RIL population

per marker location along each chromosome. The frequency distribution was mostly 50-50% as expected. However, we found skewness in the distribution of two regions. A preference for *S. pimpinellifolium* alleles was seen near the centromere of Chromosome 2, and a preference for *S. lycopersicum* alleles on Chromosome 9 (Additional file 2: Figure S1).

QTLs and in silico mapping

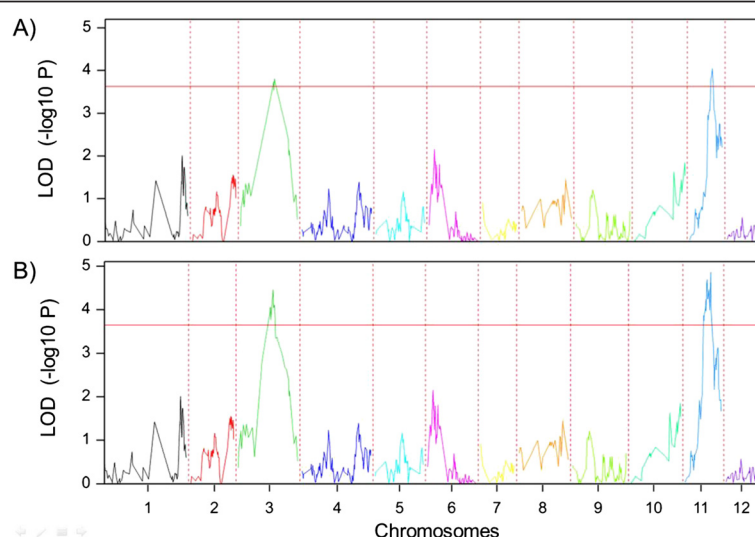
The genotypic file and the linkage map obtained above were then used to map multiple traits. One of the traits screened using our RIL population was TYLCV resistance. Eighty-one RILs were infected with TYLCV. Typical virus symptoms appeared from 30 days after inoculation (dpi); plants were scored according to their symptom development up to 45 dpi and classified as Resistant (R) or Susceptible (S). The susceptible parent 'MoneyMaker', as expected, displayed severe TYLCV symptoms such as plant stunting and reduced leaf size with upwards curling and yellowing. The resistant parent, *S. pimpinellifolium* G1.1554, remained without symptoms until the end of the experiment. Five out of 81 tested RILs showed no



symptoms after virus inoculation (disease score = 0), and four RILs showing very mild symptoms (disease score ≤ 1) were considered resistant. The remaining 72 RILs were classified as susceptible, showing clear TYLCV symptoms including the characteristic leaf curling and yellowing with disease scores ranging from 2 to 4 (Additional file 1: Figure S2).

In order to identify the genomic regions involved in the resistance, single trait QTL analysis was performed.

Two putative QTLs associated with the resistance were identified, one on Chromosome 3, hereafter referred to as *qTy-p3*, and one on Chromosome 11, hereafter referred to as *qTy-p11* (Figure 3A). For *qTy-p3*, 20 markers showed significant association with a LOD value ranging from 3.68 to 3.81, locating the QTL between 4.74 and 45.59 Mb of chromosome 3; the most significant marker for *qTy-p3* was L_45597186-3. For *qTy-p11*, 6 significantly associated markers were identified with a LOD



value from 3.79 to 4.04, in a region between 50.82 and 51.20 Mb of chromosome 11. The most significant marker for this QTL was L_51208173-11 (Figure 3; Additional file 3: Figure S3).

Sixty lines from the RIL population were re-sequenced, and the resulting genome sequences were aligned to the published tomato genome, version SL2.40 [1]. The fully resistant lines were included among the 60 sequenced RILs. JBrowse [34] was used to visualize SNP variants within the RILs and allowed us to retrieve the corresponding SNP information of all aligned reads in regions of interest.

We selected 43 additional SNPs to saturate Chromosome 3 resulting in approximately one marker per 0.6 Mb. For Chromosome 11, we included two markers in the region of 7.5-8.3 Mb and 27 in the region between 49-53Mbp. As a result, the Chromosomes 3 and 11 linkage groups were improved, as was the *in silico* mapping for the subset of 60 lines.

The outcome of the QTL analysis with the enriched genotypic data and improved genetic map is depicted in Figure 3B. Using this extended dataset, the analysis confirmed the QTLs *qTy-p3* and *qTy-p11*. The calculated threshold was very similar to the previous calculated threshold (3.64). For *qTy-p3* the LOD values ranged from 3.7 to 4.5, comprising a region with 53 significantly linked markers. The most significantly linked marker position for *qTy-p3* was then refined from 45597186 bp in the first QTL mapping to 46454095 bp and 46520535 bp (both LOD of 4.46) in the improved version. For *qTy-p11* the LOD values for the 26 significantly linked markers (in the improved map) ranged from 3.86 to 4.86, and the most significant marker position was refined from 51208173 bp to 51347236 bp and 51373277 bp (both LOD of 4.86). Together, both QTLs explained almost 28% of the phenotypic effect (13.46 for *qTy-p3* and 14.18 for *qTy-p11*).

A QTL analysis using cofactors (MQM) was performed. When the most significant markers of Chromosome 3 were used as cofactors, the LOD values of *qTy-p11* decreased but were still significant. However when the most significant markers of Chromosome 11 were used as cofactors, the values of *qTy-p3* decreased to non-significant levels. Therefore, the greater impact of *qTy-p11* for the resistance was confirmed. Although all resistant RILs were homozygous for the *S. pimpinellifolium* allele at both QTLs, 14 RILs had disease scores of 2–4 (susceptible). Thus both QTLs with the favourable alleles are necessary for resistance, but their presence did not necessarily result in resistant plants.

Identification of candidate genes

In order to identify candidate genes for TYLCV resistance, we re-explored the QTL regions using the physical positions of the SNP markers flanking the QTLs. For

Chromosome 11, we targeted the region between 50.2 and 51.4 Mb. For *qTy-p11*, a total of 124 predicted genes were identified using Marker2sequence [35] based on the tomato genome sequence (Sol Genomics Network, SGN). Four putative disease-resistance proteins were predicted in the *qTy-p11* region, three of them clustering in the region from position 51347236 to 51373277. Furthermore, approximately 74.9 kb of *qTy-p11* overlaps with the region reported to contain the *Ty-2* resistance allele from *S. habrochaites* accession B6013 [46].

The *qTy-p3* QTL region is physically large, from 2.48 to 47.44 Mb (45 Mb), including the centromeric region. This QTL region harbours more than six hundred annotated genes. In the vicinity of position 46454095 bp (the marker with the highest LOD score) there are genes related to sugars (e.g. high-affinity sugar transporters) and flavonoids (e.g. flavanone 3-hydroxylase-like protein).

RIL population metabolic profiling

Using the RIL population (not TYLCV infected), we performed untargeted metabolic profiling on leaf material. Primary metabolites were evaluated using GC-TOF-MS. Few differences were observed between parents and individuals of the population showing a similarity in the primary metabolism. However, the LC-TOF-MS and the SPME-GC-MS platforms uncovered more differences and revealed several QTLs for secondary metabolites and volatiles. More than 200 QTLs were found with putatively identified compounds; an mQTL for sucrose was mapped near *qTy-p11*, and several mQTLs for flavonoid glycosides were present near the region of *qTy-p3* (Additional file 4: Table S1).

Furthermore, since there were TYLCV-susceptible and resistant lines with both QTLs having the homozygous *S. pimpinellifolium* alleles, we performed a T-Test with all metabolic data in order to find metabolites that were significantly different between the two groups of RILs. Five compounds showed significant differences (p-value lower than 0.05) and had higher accumulations in the resistant plants. Three of them were putatively identified as glycosylated forms of kaempferol (LCS146), laricitrin (LCS149) and quercetin (LCS151) having a 4.3, 3.8 and 2.8-fold change, respectively. The other two compounds were acetoxymatine (C724) and sucrose (C121) with 1.6 and 1.5-fold difference, respectively.

Discussion

High-throughput genetic mapping

The custom made SNP array was designed to distinguish different *S. lycopersicum* cultivars, nevertheless a vast amount of polymorphisms were detected between *S. pimpinellifolium* and *S. lycopersicum* cv. MoneyMaker making it possible to construct a high density genetic linkage map. In general, positions on the genetic linkage map were

consistent with the physical positions on the tomato genome showing the accuracy and robustness of the map and the quality of the tomato sequence.

High and low recombination rates were consistent with the known distribution of euchromatic and heterochromatic regions, as shown by Sim *et al.* [8]. Chromosomes 1, 3, 4, 5 and 10 had large regions without recombination including the centromeres. Centromeric patterns were also observed for chromosomes 6, 7, 8, 9 and 11, but there were some possible distortions that could profit from more markers in the region. Still, the distortions of Chromosome 6 might be influenced by the distinct heterochromatin distribution that follows an alternating pattern [47]. Chromosome 12 also showed a non-recombining centromeric pattern, but this is a clear representation of the likely scaffold misalignment reported previously [12]. Strong clustering of markers on the genetic map but with a clear physical distance between these markers shows a suppression of recombination in these areas (Figure 2).

The allele frequencies showed a preference for the *S. pimpinellifolium* alleles near the centromere on Chromosome 2. This part of the chromosome is linked to rDNA genes. Therefore, there could be a preference for *S. pimpinellifolium* rDNA. A preference was also found for the 'Moneymaker' alleles on Chromosome 9 which might be related to deleterious effects of carrying the *S. pimpinellifolium* alleles in this region or to structural DNA differences. Species in the same genus can have DNA configuration differences generating structural changes in the rearrangement of chromosomes after a cross [11]. Differences in local recombination frequencies could be related to the pairing of homologous chromosomes, DNA sequence similarity or divergence, including the presence or absence of genes involved in the recombination process, chromatin conformation or to differences in timing during meiosis [48].

Actual research is enriched by the combination of different software packages. The combination of JBrowse [34], loaded with gene models from the Sol Genomics Network (<http://solgenomics.net/>), with previous information of possible genes of interest obtained from Marker2sequence [35] allowed an efficient targeted *in silico* mapping.

TYLCV resistance mapping and ~ omics platforms combination

The sequenced subset of 60 lines created suitable tools for mapping regions of interest. We enriched regions on Chromosome 3 and Chromosome 11 that were associated with TYLCV resistance, and the *in silico* approach proved to be successful in increasing the power of QTL detection. After the addition of more SNPs coming from the known sequences, we confirmed that *qTy-p3* and

qTy-p11 were not artefacts but had real effects. This allowed us to target the location of the QTL region for *qTy-p11* and it showed the most significant region for *qTy-p3* (Figure 3), even though a large region of Chromosome 3, including the centromere, looks to have an essential impact on the expression of the resistance.

The effect of both QTLs together explained only 28% of the phenotypic effect on the resistance of our RIL population, suggesting additional genetic factors playing a role on the resistance which might have been undetected in our analysis. The accuracy of QTL localization using RILs depends on population size, where a genome-wide coverage of the parents should be present in the mapping population [49]. The fact that both *qTy-p3* and *qTy-p11* were needed for resistance but their presence does not necessarily lead to resistant plants also suggests the possible interaction of extra factors. TYLCV resistance derived from a number of *S. pimpinellifolium* accessions (e.g. LA121, LA373, UPV16991) has been previously suggested to be quantitatively inherited and to show variable gene penetrance [24]. Further genotyping, targeting the regions of low marker coverage, is being assessed in order to detect the presence of one or more additional QTLs, or potential modifier genes. These interactions might be associated with the secondary metabolism of the plants.

A number of TYLCV resistance loci have been reported from different wild *Solanum* species, including *S. chilense*, *S. habrochaites* and *S. peruvianum* [25]. Recently, the *Ty-1* gene from *S. chilense* LA1969 has been cloned and is a representative for a novel class of resistance genes, an RNA-dependent RNA polymerase of the RDRy class [18,50]. TYLCV resistance in *S. chilense* accessions LA1932 and LA2779, *S. habrochaites* accession B6013 and TY172, a tomato line derived from different accessions of *S. peruvianum* have been mapped to Chromosomes 3 and 10 (*Ty-4* [25] and *Ty-6* [51]), Chromosome 11 (*Ty-2* [46]) and Chromosome 4 (*ty-5* [16]), respectively.

Several accessions from *S. pimpinellifolium* have been screened and identified to confer resistance to TYLCV [19,20,22-25]. However, the genetics of the trait are complex and only one report on mapping resistance originating from *S. pimpinellifolium* (accession 'Hirsute INRA') has been reported using RAPD markers [21]. This resistance was mapped to Chromosome 6, close to the *Ty-1* gene. The QTLs identified in the present study represent newly mapped loci conferring resistance derived from *S. pimpinellifolium* G1.1554 and provide a starting point for assessing putative candidate genes in the identified regions. A cluster of disease resistance-like proteins is present near *qTy-p11* (based on the cultivated tomato genome sequence). Furthermore, this region on Chromosome 11 overlaps with 75 kb of the upper part of the mapped region of *Ty-2*, a TYLCV resistance allele derived from

S. habrochaites accession B6013 [46]. Although *Ty-2* has not yet been cloned, annotated genes in this common region (e.g. elongation factor 1- α) might provide further insights for assessing candidate genes for TYLCV resistance derived from these wild tomato species, and/or additional genes involved in the resistance pathway. Plant defense mechanisms are the result of complex gene networks which trigger or mediate the signaling pathways leading to resistance. Besides the reported *Ty*-loci, genes playing a role in these networks have been identified from their differential expression in resistant vs. susceptible genotypes and induced by TYLCV infection, e.g. *Permease I*-like protein and the hexose transporter *LeHT1* [52,53]. Silencing these genes through Virus-induced gene silencing (VIGS) in a resistant genotype led to the collapse of the resistance, demonstrating the role and importance of these genes in the defense network of the plant.

In general, the presence of compounds such as amino acids and organic acids was very similar between the two species. Differences are more pronounced in the secondary metabolism. Our metabolic data show that the compounds present at higher amounts in the resistant plants are mainly flavonoid glycosides. Flavonoids are phenolic compounds known to be involved in resistance to diverse stress conditions, including plant viruses [54]. For instance quercetin, one of the metabolites detected at higher levels in the resistant lines is a flavonoid known to inhibit HSP70 (Heat-shock protein 70) transcription in animal and plant cells. In *N. benthamiana*, Tomato yellow leaf curl Sardinia virus (TYLCSV) had a delayed infection speed after silencing a member of the HSP70 family, showing that high levels of this protein are required for infection of the virus [55]. Inhibition of HSP70 expression by quercetin resulted in decreased amounts of nuclear TYLCV coat protein in tomato, demonstrating the potential involvement of this flavonoid in the virus resistance pathway [56]. Furthermore, an additional QTL analysis suggests that glycosides of the flavonoid kaempferol co-localise with the TYLCV resistance QTL on Chromosome 3 and that sucrose could be related to the QTL on Chromosome 11 (Additional file 4: Table S1). Kaempferol is known for its antibacterial properties. Besides, we observed the presence of this compound and other flavonoids attached to hexoses in the resistant RILs; transporters of hexoses have been reported to play crucial roles in disease resistance [53,57]. Some of these compounds likely linked to the resistance also showed an mQTL on chromosome 1 besides the one on chromosome 3, and the mQTL of sucrose also showed significance on chromosome 7. These regions will be further targeted in a fine mapping effort following up this research.

It should be noted that the different concentrations of the compounds observed in resistant vs. susceptible lines

were measured prior to TYLCV infection. Sade et al. [58] showed that the expression of genes controlling the synthesis of these phenolic compounds is associated with TYLCV resistance. Genes in the flavonoid biosynthesis pathway of a resistant line derived from *S. habrochaites* increased their expression after TYLCV infection leading to the accumulation of flavonoids and contributing to the resistance.

Conclusion

A RIL population obtained from a cross between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554 was successfully genotyped with a custom made SNP array. Furthermore, the re-sequencing of a subset of the RILs allowed the possibility of *in silico* mapping of TYLCV resistance. Two QTLs were related to the resistance, one showing the highest significance on Chromosome 11 close to the region of 51.3 Mb and the other close to 46.5Mbp on Chromosome 3. However, there might be extra loci or genetic factors playing a role that could be unravelled if the population size is increased or when advanced populations are further explored. The resistance towards TYLCV suggests an interaction between flavonoids and hexoses favouring the trait.

We concluded that investments in sequencing can redeem the value of screenings of germplasm due to the fact that both SNPs and sequences can be targeted at the same time. Therefore, screenings can start with a defined number of retrieved SNPs per chromosome, and thereafter, regions of interest can be further targeted. However, data storage, software acquisition and qualified human resources for data analysis and interpretation of combined ~ omics platforms are going to make the difference to get robust analyses.

Additional files

Additional file 1: Figure S2. Disease scores of TYLCV symptom development. Plants were scored according to symptom severity: 0, no visible symptoms; 1, very slight yellowing and minor curling of leaflet margins; 2, yellowing and minor curling of leaflet ends; 3, leaf yellowing, curling and cupping; 4, severe leaf yellowing, curling and cupping, plant stunting (Friedmann et al., [37]).

Additional file 2: Figure S1. Probabilities of marker frequencies calculated in GenStat. A skewness in the direction of the chromosome region from *S. pimpinellifolium* G1.1554 is observed for Chromosome 2. A skewness in the direction of the chromosome region from *S. lycopersicum* cv. Moneymaker is observed for Chromosome 9.

Additional file 3: Figure S3. Physical map of Chromosome 3 and Chromosome 11 between 49–53 Mb. A) Chromosomes with initial SNPs. B) Chromosomes with incorporated SNPs in green. Black arrows indicate the most significant marker related to TYLCV resistance on each case. Black frames indicate the length covering significant markers for *qTy-p3* and *qTy-p11*.

Additional file 4: Table S1. QTLs found in non-infected leaves among the population between *S. lycopersicum* cv. Moneymaker and *S. pimpinellifolium* G1.1554.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MVZ conceived the study, carried out the analysis of the SNP data, performed the metabolomics experiments and wrote the manuscript. MC conceived the study, performed the experiments for TYLCV resistance and wrote the manuscript. RF carried out the bioinformatics analysis of the sequences. YT gave support for the metabolomic analysis. AB verified the analysis. RGFV lead the entire project. YB coordinated the experimental set-up related to TYLCV resistance and co-wrote the manuscript. AWH conceived the entire study and co-wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This project was carried out within the research programme of the Centre of BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research. And the metabolic profiling of the leaves was carried out within a project developed in collaboration with the Netherlands Metabolomics Centre (NMC). Furthermore, we received support from the COST action 'Quality Fruit' (COST FA1106).

We would like to thank Fien Meijer-Dekens for collecting and maintaining all the tomato samples at Wageningen UR-Plant Breeding, Ben Vosman for his support with the NMC project and Maarten Verlaan for his support and guidance in the elaboration of the TYLCV resistance experiments. We also would like to thank various people within the department of Bioinformatics of Wageningen UR for their support in the development of the SNP markers and in allocating them into the different versions of the Tomato Genome Sequence. We would like to make a special acknowledgement to the department of Bioscience at Wageningen UR, specially to Ric de Vos, Henriette van Eekelen, Harry Jonker, Roland Mumm and Robert Hall for all their help and support in running the samples. Additionally, MVZ acknowledges the 'Instituto Nacional de Innovación y Transferencia en Tecnología Agropecuaria' in Costa Rica, for their support through part of the doctoral project.

Author details

¹Wageningen UR Plant Breeding, Wageningen University & Research Centre, P.O. Box 386, Wageningen, AJ 6700, the Netherlands. ²Centre for Biosystems Genomics, P.O. Box 98, Wageningen, AB 6700, the Netherlands. ³Graduate School Experimental Plant Sciences, Wageningen 6708 PB, the Netherlands.

Received: 18 August 2014 Accepted: 12 December 2014

Published: 20 December 2014

References

1. The Tomato Genome Consortium: **The tomato genome sequence provides insights into fleshy fruit evolution.** *Nature* 2012, **485**(7400):635–641.
2. Grandillo S, Tanksley SD: **Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium*.** *Theor Appl Genet* 1996, **92**(8):957–965.
3. Chen FQ, Foolad MR: **A molecular linkage map of tomato based on a cross between *Lycopersicon esculentum* and *L. pimpinellifolium* and its comparison with other molecular maps of tomato.** *Genome* 1999, **42**(1):94–103.
4. Lippman Z, Tanksley SD: **Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *Lycopersicon pimpinellifolium* and *L. esculentum* var. Giant Heirloom.** *Genetics* 2001, **158**(1):413–422.
5. Doganlar S, Frary A, Ku HM, Tanksley SD: **Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589).** *Genome* 2002, **45**(6):1189–1202.
6. Sharma A, Zhang L, Nio-Liu D, Ashrafi H, Foolad MR: **A solanum lycopersicum solanum pimpinellifolium linkage map of tomato displaying genomic locations of R-genes, RGAs, and candidate resistance/defense-response ESTs.** *International Journal of Plant Genomics* 2008, **2008**:1–18.
7. Ashrafi H, Kinkade M, Foolad MR: **A new genetic linkage map of tomato based on a *Solanum lycopersicum* × *S. pimpinellifolium* RIL population displaying locations of candidate pathogen response genes.** *Genome* 2009, **52**(11):935–956.
8. Sim S-C, Durstewitz G, Plieske J, Wieseke R, Ganai MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis D: **Development of a large SNP genotyping array and generation of high-density genetic maps in tomato.** *PLoS One* 2012, **7**(7):e40563.
9. Broman KW: **The genomes of recombinant inbred lines.** *Genetics* 2005, **169**(2):1133–1146.
10. Paran I, Goldman I, Tanksley SD, Zamir D: **Recombinant inbred lines for genetic mapping in tomato.** *Theor Appl Genet* 1995, **90**(3–4):542–548.
11. Mézard C: **Meiotic recombination hotspots in plants.** *Biochem Soc Trans* 2006, **34**(4):531–534.
12. Viquez-Zamora M, Vosman B, van de Geest H, Bovy A, Visser RGF, Finkers R, van Heusden AW: **Tomato breeding in the genomics era: Insights from a SNP array.** *BMC Genomics* 2013, **14**(1):354.
13. Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B: **High-throughput genotyping by whole-genome resequencing.** *Genome Res* 2009, **19**(6):1068–1076.
14. Causse M, Desplat N, Pascual L, Le Paslier MC, Sauvage C, Bauchet G, Bérard A, Bounon R, Tchoumakov M, Brunel D, Bouchet J: **Whole genome resequencing in tomato reveals variation associated with introgression and breeding events.** *BMC Genomics* 2013, **14**(1):791.
15. Afilitos S, Schijlen E, de Jong H, de Ridder D, Smit S, Finkers R, Wang J, Zhang G, Li N, Mao L, Bakker F, Dirks R, Breit T, Gravendeel B, Huits H, Struss D, Swanson-Wagner R, van Leeuwen H, van Ham RCHJ, Fito L, Guignier L, Sevilla M, Ellul P, Ganko E, Kapur A, Reclus E, de Geus B, van de Geest H, te Lintel Hekkert B, van Haarst J, Smits L, Koops A, Sanchez-Perez G, van Heusden AW, Visser R, Quan Z, Min J, Liao L, Wang X, Wang G, Yue Z, Yang X, Xu N, Schranz E, Smets E, Vos R, Rauwerda J, Ursem R, Schuit C, Kerns M, van den Berg J, Vriezen W, Janssen A, Datema E, Jahrman T, Moquet F, Bonnet J, Peters S: **Exploring genetic variation in the tomato (*Solanum section Lycopersicon*) clade by whole-genome sequencing.** *Plant Journal* 2014, **80**(1):134–148.
16. Anbinder I, Reuveni M, Azari R, Paran I, Nahon S, Shlomo H, Chen L, Lapidot M, Levin I: **Molecular dissection of Tomato leaf curl virus resistance in tomato line TY172 derived from *Solanum peruvianum*.** *Theor Appl Genet* 2009, **119**(3):519–530.
17. Pico B, Ferriol M, Diez MJ, Vinals FN: **Agroinoculation methods to screen wild *Lycopersicon* for resistance to Tomato yellow leaf curl virus.** *J Plant Pathol* 2001, **83**(3):215–220.
18. Verlaan MG, Hutton SF, Ibrahim RM, Kormelink R, Visser RG, Scott JW, Edwards JD, Bai Y: **The Tomato Yellow Leaf Curl Virus Resistance Genes Ty-1 and Ty-3 Are Allelic and Code for DFDGD-Class RNA-Dependent RNA Polymerases.** *PLoS Genet* 2013, **9**(3):e1003399.
19. Banerjee MK, Kalloo MK: **Sources and inheritance of resistance to leaf curl virus in *Lycopersicon*.** *Theor Appl Genet* 1987, **73**(5):707–710.
20. Kasrawi MA, Suwwan MA, Mansour A: **Sources of resistance to tomato-yellow-leaf-curl-virus (TYLCV) in *Lycopersicon* species.** *Euphytica* 1988, **37**(1):61–64.
21. Chagué V, Mercier J, Guenard M, De Courcel A, Vedel F: **Identification of RAPD markers linked to a locus involved in quantitative resistance to TYLCV in tomato by bulked segregant analysis.** *Theor Appl Genet* 1997, **95**(4):671–677.
22. Pico B, Sifres A, Elia M, Diez MJ, Nuez F: **Searching for new resistance sources to tomato yellow leaf curl virus within a highly variable wild *Lycopersicon* genetic pool.** *Acta Physiologiae Plantarum* 2000, **22**(3):344–350.
23. Pilowsky M, Cohen S: **Screening additional wild tomatoes for resistance to the whitefly-borne tomato yellow leaf curl virus.** *Acta Physiologiae Plantarum* 2000, **22**(3):351–353.
24. PérezdeCastro A, Diez MJ, Nuez F: **Inheritance of Tomato yellow leaf curl virus resistance derived from *Solanum pimpinellifolium* UPV16991.** *Plant Disease* 2007, **91**(7):879–885.
25. Ji Y, Scott JW, Hanson P, Graham E, Maxwell DP: **Sources of resistance, inheritance, and location of genetic loci conferring resistance to members of the tomato-infecting begomoviruses.** In *Tomato Yellow Leaf Curl Virus Disease: Management, Molecular Biology, Breeding for Resistance*. Edited by Czosnek H: Springer Netherlands; 2007: 343–362.
26. Voorrips RE, Verkerke W, Finkers R, Jongerius R, Kanne J: **Inheritance of taste components in tomato.** *Acta Physiologiae Plantarum* 2000, **22**(3):259–261.
27. Khan N, Kazmi RH, Willems LAJ, van Heusden AW, Ligterink W, Hilhorst HWM: **Exploring the Natural Variation for Seedling Traits and Their Link with Seed Dimensions in Tomato.** *PLoS One* 2012, **7**(8):e43991.

28. Stewart CN Jr, Via LE: A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *Biotechniques* 1993, **14**(5):748–750.
29. Kabelka E, Franchino B, Francis DM: Two loci from *Lycopersicon hirsutum* LA407 confer resistance to strains of *Clavibacter michiganensis* subsp. *michiganensis*. *Phytopathology* 2002, **92**(5):504–510.
30. Illumina: *Infinium® HD Assay: Ultra Protocol Guide*. California, USA: ©Illumina, Inc; 2009:1–224. Catalog #WG-901-4007.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009, **25**(16):2078–2079.
32. Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010, **26**(5):589–595.
33. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010, **20**(9):1297–1303.
34. Skinner ME, Uzilov A, Stein LD, Mungall CJ, Holmes IH: JBrowse: A next-generation genome browser. *Genome Res* 2009, **19**(9):1630–1638.
35. Chibon P, Schoof H, Visser RG, Finkers R: Marker2sequence, mine your QTL regions for candidate genes. *Bioinformatics* 2012, **28**(14):1921–1922.
36. Verlaan MG, Szinay D, Hutton SF, de Jong H, Kormelink R, Visser RG, Scott JW, Bai Y: Chromosomal rearrangements between tomato and *Solanum chilense* hamper mapping and breeding of the TYLCV resistance gene Ty-1. *Plant J* 2011, **68**(6):1093–1103.
37. Friedmann M, Lapidot M, Cohen S, Pilowsky M: A novel source of resistance to tomato yellow leaf curl virus exhibiting a symptomless reaction to viral infection. *J Am Soc Horticultural Sci* 1998, **123**:1004–1007.
38. De Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD: Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2007, **2**(4):778–791.
39. Liscic J, Schauer N, Kopka J, Willmitzer L, Fernie AR: Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat Protoc* 2006, **1**(1):387–396.
40. Tikunov Y, Lommen A, De Vos CHR, Verhoeven HA, Bino RJ, Hall RD, Bovy AG: A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles. *Plant Physiol* 2005, **139**(3):1125–1137.
41. Tikunov YM, de Vos RC, Paramás AMG, Hall RD, Bovy AG: A role for differential glycoconjugation in the emission of phenylpropanoid volatiles from tomato fruit discovered using a metabolic data fusion approach. *Plant Physiol* 2010, **152**(1):55–70.
42. Tikunov YM, Laptinok S, Hall RD, Bovy A, de Vos RCH: MSCLust: A tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data. *Metabolomics* 2012, **8**(4):714–718.
43. Van Ooijen JW: Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Res* 2011, **93**(5):343–349.
44. Voorrips RE: Mapchart: Software for the graphical presentation of linkage maps and QTLs. *J Heredity* 2002, **93**(1):77–78.
45. Kadirvel P, de la Peña R, Schafleitner R, Huang S, Geethanjali S, Kenyon L, Tsai W, Hanson P: Mapping of QTLs in tomato line FLA456 associated with resistance to a virus causing tomato yellow leaf curl disease. *Euphytica* 2013, **190**(2):297–308.
46. Yang X, Caro M, Hutton SF, Scott JW, Guo Y, Wang X, Rashid MH, Szinay D, de Jong H, Visser RGF et al: Fine mapping of the tomato yellow leaf curl virus resistance gene Ty-2 on chromosome 11 of tomato. *Molecular Breeding* 2014, **34**:749–760.
47. Iovene M, Wielgus SM, Simon PW, Buell CR, Jiang J: Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato. *Genetics* 2008, **180**(3):1307–1317.
48. Tam SM, Hays JB, Chetelat RT: Effects of suppressing the DNA mismatch repair system on homeologous recombination in tomato. *Theor Appl Genet* 2011, **123**(8):1445–1458.
49. Keurentjes JJ, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankstijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M: Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* 2007, **175**(2):891–905.
50. Butterbach P, Verlaan MG, Dullemans AM, Lohuis D, Visser RGF, Bai Y, Kormelink R: Tomato yellow leaf curl virus resistance by Ty-1 involves increased cytosine methylation of viral genomes and is compromised by cucumber mosaic virus infection. *Proc Natl Acad Sci* 2014, **111**(35):12942–12947.
51. Hutton SF: Fine-Mapping and Cloning of Ty-1 and Ty-3; and Mapping of a New TYLCV Resistance Locus, “Ty-6” [abstract]. Tomato Breeders Round Table 2013: http://tgc.ifas.ufl.edu/2013/abstracts/SamOrchardAbstract%20TBRT%202013.pdf.
52. Eybishtz A, Peretz Y, Sade D, Akad F, Czosnek H: Silencing of a single gene in tomato plants resistant to Tomato yellow leaf curl virus renders them susceptible to the virus. *Plant Mol Biol* 2009, **71**(1–2):157–171.
53. Eybishtz A, Peretz Y, Sade D, Gorovits R, Czosnek H: Tomato yellow leaf curl virus infection of a resistant tomato line with a silenced sucrose transporter gene LeHT1 results in inhibition of growth, enhanced virus spread, and necrosis. *Planta* 2010, **231**(3):537–548.
54. Bol J, Linthorst H, Cornelissen B: Plant pathogenesis-related proteins induced by virus infection. *Annu Rev Phytopathol* 1990, **28**(1):113–138.
55. Czosnek H, Eybishtz A, Sade D, Gorovits R, Sobol I, Bejarano E, Rosas-Díaz T, Lozano-Durán R: Discovering host genes involved in the infection by the tomato yellow leaf curl virus complex and in the establishment of resistance to the virus using Tobacco Rattle Virus-based post transcriptional gene silencing. *Viruses* 2013, **5**(3):998–1022.
56. Gorovits R, Moshe A, Ghanim M, Czosnek H: Recruitment of the host plant heat shock protein 70 by tomato yellow leaf curl virus coat protein is required for virus infection. *PLoS One* 2013, **8**(7):e70280.
57. Sade D, Brotman Y, Eybishtz A, Cuadros-Inostroza Á, Fernie AR, Willmitzer L, Czosnek H: Involvement of the hexose transporter gene LeHT1 and of sugars in resistance of tomato to tomato yellow leaf curl virus. *Mol Plant* 2013, **6**(5):1707–1710.
58. Sade D, Shriki O, Cuadros-Inostroza A, Tohge T, Semel Y, Haviv Y, Willmitzer L, Fernie AR, Czosnek H, Brotman Y: Comparative metabolomics and transcriptomics of plant response to Tomato yellow leaf curl virus infection in resistant and susceptible tomato cultivars. *Metabolomics* 2014, 1–17. http://link.springer.com/article/10.1007/s11306-014-0670-x.

doi:10.1186/1471-2164-15-1152

Cite this article as: Viquez-Zamora et al.: Mapping in the era of sequencing: high density genotyping and its application for mapping TYLCV resistance in *Solanum pimpinellifolium*. *BMC Genomics* 2014 **15**:1152.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

