Plant Science 242 (2016) 120–130

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Dissecting quantitative trait variation in the resequencing era: complementarity of bi-parental, multi-parental and association panels

Laura Pascual[a,1,5], Elise Albert[a,5], Christopher Sauvage[a], Janejira Duangjit[a,2], Jean-Paul Bouchet[a], Frédérique Bitton[a], Nelly Desplat[a,3], Dominique Brunel[b], Marie-Christine Le Paslier[b], Nicolas Ranc[a,4], Laure Bruguier[c], Betty Chauchard[c], Philippe Verschave[c], Mathilde Causse[a,*]

[a] INRA, UR1052, Centre de Recherche PACA, 67 Allée des Chênes CS60094, 84143 Montfavet Cedex, France
[b] INRA, US1279, Etude du Polymorphisme des Génomes végétaux (EPGV), CEA-IG/CNG, 2 rue Gaston Crémieux, 91057 Evry, France
[c] Vilmorin S.A. – Groupe Limagrain, Centre de Recherche de La Costière, Route de Meynes, 30210 Ledenon, France

ARTICLE INFO

Article history:
Received 13 April 2015
Received in revised form 12 June 2015
Accepted 16 June 2015
Available online 23 June 2015

Keywords:
Tomato
QTL mapping
Genome-wide association
Fruit quality
Resequencing

ABSTRACT

Quantitative trait loci (QTL) have been identified using traditional linkage mapping and positional cloning identified several QTLs. However linkage mapping is limited to the analysis of traits differing between two lines and the impact of the genetic background on QTL effect has been underlined. Genome-wide association studies (GWAs) were proposed to circumvent these limitations. In tomato, we have shown that GWAs is possible, using the admixed nature of cherry tomato genomes that reduces the impact of population structure. Nevertheless, GWAs success might be limited due to the low decay of linkage disequilibrium, which varies along the genome in this species.

Multi-parent advanced generation intercross (MAGIC) populations offer an alternative to traditional linkage and GWAs by increasing the precision of QTL mapping. We have developed a MAGIC population by crossing eight tomato lines whose genomes were resequenced. We showed the potential of the MAGIC population when coupled with whole genome sequencing to detect candidate single nucleotide polymorphisms (SNPs) underlying the QTLs. QTLs for fruit quality traits were mapped and related to the variations detected at the genome sequence and expression levels. The advantages and limitations of the three types of population, in the context of the available genome sequence and resequencing facilities, are discussed.

1. Introduction

Agronomic traits are usually under the control of several genes with variable effects modulated by the environment. Since the pioneer work of Paterson and colleagues [1], deciphering the genetic control of quantitative traits into quantitative trait loci (QTLs) has been studied through QTL mapping [2,3]. Quantitative trait loci have been mapped in many crops in biparental populations segregating after one (F2 populations) or a few selfing generations (in recombinant inbred lines, RIL), when selfing is possible, or on advanced backcross progenies. Populations of introgression lines covering the whole genome are also helpful to identify QTLs from wild species in a cultivated genetic background [4]. Among hundreds of QTLs mapped, only a few were identified following positional cloning [5]. Nevertheless such populations allow the identification of the QTLs differing only between the two parental lines. The confidence intervals around QTLs are usually large as they only rely on one or two efficient recombination generations. Until the recent advent in genome sequencing, the number of available molecular markers was also limiting the power of this approach, particularly to fine map genes and QTLs. Since the discovery of SNP markers, thousands of markers are available, drastically changing the paradigm of QTL mapping. In the early 2000s, it was proposed

* Corresponding author. Tel.: +33 432722803.
E-mail address: mathilde.causse@avignon.inra.fr (M. Causse).
[1] Present address: Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UA, Universidad de Barcelona, Barcelona 08193, Spain.
[2] Present address: Department of Horticulture, Faculty of Agriculture, Kasetsart University, 10900 Bangkok, Thailand.
[3] Present address: Institut Bergonié, 229, Cours de l'Argonne, 33000 Bordeaux, France.
[4] Present address: Syngenta Seeds 12, 31790 St Sauveur, France.
[5] These authors contributed equally to the article.

http://dx.doi.org/10.1016/j.plantsci.2015.06.017

to extend the QTL mapping approach to panels of unrelated lines through Genome- Wide Association Studies (GWAs) as first used in human genetics. The GWAs allow the discovery of QTLs in broad panels. It is particularly efficient in species with low linkage disequilibrium (LD) [6,7]. The population structure of the studied panel must be taken into account as it can lead to false positive association discovery [8,9]. If LD is sufficiently low and the number of markers is high, GWAs can land on the causal polymorphism [10].

Multi-parental populations represent intermediate populations, with more equilibrated allelic frequencies than GWAs panels and higher efficient recombination than biparental populations. Two main types of populations were proposed, Nested Association Mapping, mainly used in maize [11] and Multi-allelic Genetic Intercross (MAGIC), which have been developed in *Arabidopsis* [12], rice [13], wheat [14], barley [15] and tomato [16]. Multi-parental populations constitute a unique resource that can overcome the main limitations of GWAs and RIL studies and provide complementary information [17]. Generating new phenotypes by mixing different gene alleles permits the exploitation of QTL effects on the different founders of the population and quickly identifies causal variants [16]. Additionally, these new phenotypes constitute a highly valuable pre-breeding resource and a potential tool to develop genomic selection models. Evaluating GWAs offers unique information by allowing the analysis of a wider range of diversity, and usually provide greater precision, as they are based on recombination that has taken place during a greater number of generations. Other connected population designs were proposed [18,19] with related interests. We recently developed a tomato MAGIC population based on eight cultivated lines and showed its potential to map QTLs for fruit weight [16]. Furthermore, the genomes of the eight parental lines were sequenced [20] and the list of candidate genes was reduced by combining the predicted allelic effect at the QTLs with SNP haplotypes.

To illustrate the pros and cons of each of the three strategies, QTL mapping (in RIL and MAGIC populations) and GWAs (in a panel of accessions), we used the cultivated tomato (*Solanum lycopersicum* L.) as a model. Tomato is commonly cultivated vegetables worldwide and a model species for fruit quality and development [21]. For years QTL mapping among cultivated accessions of tomato was hampered by the low polymorphisms in the species [22], but many progenies involving distant related species were characterized [23]. Several QTL controlling fruit weight or fruit composition were mapped and characterized [24,25]. A high quality tomato genome sequence is now available [26] allowing the resequencing of several accessions [27–29] and the detection of several million of SNPs, which aids in the development of a SNP chip for diversity analyses [30]. In cultivated tomato, the molecular polymorphism is low and LD is high, although varied along the chromosomes [31]. Using a panel of highly variable cherry tomato accessions, we showed that GWAs were possible in tomato for fruit metabolite traits [32]. It was also particularly helpful to identify causative SNPs for a QTL identified by map based cloning [33].

In the present article, we compare original results of QTL and association mapping experiments using three populations: (1) a RIL population that was first mapped using RFLP markers [34]. The resequencing of the parental lines allowed the construction of a saturated map and QTL mapping using this new map; (2) a MAGIC population derived from eight lines whose genomes were resequenced and (3) a GWAs experiment based on a core collection. QTL were mapped for fruit quality and agronomic traits and their locations and effects were compared. Finally we discuss and compare these populations for QTL mapping and characterization in the new genome era.

2. Materials and methods

2.1. RIL mapping population

A population of 124 F7 recombinant inbred lines (RIL) was developed from the intraspecific cross of two inbred lines Cervil and Levovil as described in [34]. Cervil is a cherry tomato (*Solanum lycopersicum* var. *cerasiforme*) with small fruits (6–10 g) and high aroma intensity, whereas Levovil (*S. lycopersicum*) has much larger fruits (90–160 g) with common taste. In 1996, the RIL were phenotyped for plant and fruit quality traits in a fully randomized trial in a greenhouse at Chateaurenard in Southern France. Plant traits were flowering date of the first flower on the third truss (FLW) and height of the 6th truss on plant stem (HT). The quality traits measured on red fruits were: fresh weight (FW), firmness (FIR), external color (COB, corresponding to the b parameter – blue to yellow – of L, a*, b* parameters), soluble solids content (SSC), pH and titratable acidity (TA), as detailed in [34].

2.2. Genetic data and mapping in RIL

Following the resequencing of the parental lines [20], 754 polymorphic markers were genotyped on the progeny: 679 SNP from parent re-sequencing, 2 RAPD (random amplified polymorphic DNA) and 73 RFLP (restriction fragment length polymorphism) mapped in the previous genetic map from this progeny [22]. The average rate of missing data per marker was estimated at 3% while 98% of the markers passed the Chi-square test ($\alpha = 0.0001\%$). Markers with significant segregation distortion were excluded. Linkage analysis was performed using JoinMap 4.1 [35]. The 12 linkage groups (LG) corresponding to the 12 chromosomes of the tomato genome were built with a grouping logarithm of odds (LOD)-threshold of 4.0, except LG05 for which the grouping threshold was lowered to 3.0. The regression-mapping algorithm was used to order markers within each LG. Genetic distances between markers were calculated using the Haldane mapping function. When several markers colocalized, only the one with the lower rate of missing data was conserved.

2.3. QTL detection in RIL

Quantitative trait loci detection was performed by simple interval mapping [36] using the expectation maximization (EM) algorithm method implemented in R/QTL package [37]. A $\log_{10}$ transformation was applied to FW, FIR and COB as trait distributions deviated from normality. A 1000-permutation test was performed to estimate significant threshold. The LOD threshold was 2.76, corresponding to a genome-wise significance level of $\alpha = 0.10$. For each detected QTL, position, LOD score, confidence interval (CI – for a decrease in the LOD score of one unit), average phenotypic values of the two parental alleles and percentage of phenotypic variation explained (PVE) were displayed. The genetic-CIs were translated into physical intervals (Physical-CI) onto the tomato genome (assembly 2.4).

2.4. MAGIC population

The MAGIC population (397 lines) was obtained by crossing eight tomato lines (including the two parents of the RIL population), selected to include a wide range of genetic diversity of the species as described in [16]. The population was grown in two locations in the South of France in Avignon (location INRA) and La Costière (location VCo). In each location, the 397 lines (one plant per line) and five replicates of each founder were grown in greenhouses during spring-summer 2012, as described in [16]. The traits measured were truss height at second truss (HT), flowering date at third truss

(FLW) and fruit quality traits. Red-fruit quality traits were fresh weight (FW), firmness (FIR), external color (COB), soluble solids content (SSC), pH and titratable acidity (TA) as in RILs. Traits were evaluated from a minimum of five ripe fruits per genotype, collected during two different harvests from truss two to six. A $\log_{10}$ transformation was carried out to normalize the data for FW and HT.

2.5. Genetic data, mapping and QTL detection in MAGIC population

The whole genomes of the eight founder lines were re-sequenced allowing the identification of more than 4 million SNPs [20]. Polymorphism information was used to select a subset of 1486 markers specially designed to analyze the MAGIC population. The selected markers were employed to develop a saturated map [16]. Briefly, genetic distances were estimated with the '*computemap*' *mpMap* function, using a 15-marker window and Haldane distances were computed. This map and the genotype of the parental lines were used to predict the haplotype origin of each locus along the MAGIC population lines genome (using '*mpprob*' function from *mpMap*) [16]. Based on this information QTLs were detected by Simple Interval Mapping using the *R/MpMap* package [38]. QTLs were called when *p*-values were smaller than the empirical threshold *p*-value (1.31×10^{-4}) derived after computing 1000 permutations, to reflect a genome-wide significance threshold of 0.05. When the QTL profile showed more than one QTL peak per chromosome, multiple QTLs were considered significant when peaks were separated by more than 20 cM and the LOD score dropped by more than one. In order to compare QTLs detected in several populations, a second less stringent QTL detection was performed, where QTLs were called when *p*-values were smaller than 10^{-3} corresponding to a LOD 3 value. As in the RIL population, QTL support intervals (SI) were determined with a 1-LOD drop support and translated in physical intervals (physical-SI) on the tomato genome (assembly 2.4).

Recombination events were imputed at locations where the parent/founder allele/haplotype changed along the chromosome in the RIL/MAGIC lines. To calculate haplotype size, recombination locations were translated to physical positions according to their position on the tomato genome (assembly 2.4). Then, haplotypes were defined as the blocks between the beginning/end of the chromosome and the closer recombination, or two consecutive recombinations. Using the haplotype predictions along the MAGIC line genomes, we performed a joint Wald-test for the significance of all founder effects at putative QTL positions.

2.6. GWAs panel

The tomato diversity panel consisted of 163 accessions composed of 28 *S. lycopersicum* (S.L.), 119 *S. lycopersicum* cv. *cerasiforme* (S.C.) and 16 *S. pimpinellifolium* (S.P.) as described in [32], and [39]. Plants (four replicates) were grown in plastic tunnel in Avignon, France during summers 2007 and 2008. At least 10 fruits per plot were measured for the same traits, as in the RIL and MAGIC populations, as described in [34,39], except for flowering time and plant height. Phenotypic data collected in 2007 and 2008 were averaged over the two years and $\log_{10}$ transformed when the Shapiro–Wilk test evidenced a non-normal distribution.

2.7. Genetic data and GWAs analysis

Genotyping was performed using the Infinium assay (Illumina Inc.), developed by the Solanaceae Coordinated Agricultural Project (SolCAP) [40]. After filtering for quality, missing data and low allele frequency, a set of 5995 SNP markers remained. Briefly, pairwise kinship coefficients (*K* matrix) were estimated using the Ritland formula implemented in SPAGeDI [41]. For population stratification, the most likely number of clusters K in all simulations were assumed to be in the range of $K = 1$ to $K = 10$. Ten replicates were conducted in the structure software [42] for each *K* with a burn-in period of 1×10^6, followed by 5×10^6 Markov chain Monte Carlo (MCMC) steps. Then, the Evanno correction was applied [43]. GWA analyses were performed with correction for population structure (Q for FIR, pH and SSC or PCoA for FW, COB and TA) and modeling phenotypic covariance with the kinship (*K*) matrix. These matrices were implemented into a modified version of the multi-locus mixed model (MLMM) described in [44]. The analysis followed the same steps as in [32]. Levels of significance were assessed according to [44].

2.8. Comparison of QTLs and screen for candidate genes and polymorphisms

We projected all markers on the SL2.40 genome sequence and thus mapped all the QTLs/associations on the same framework. We compared QTL SI and decided that a single QTL was present when the SI overlapped or when an association lay in the SI.

To screen for candidate genes and polymorphisms, we selected QTLs from the MAGIC population with SI lower than 1 Mb and listed all the polymorphisms detected among the parental lines in the interval. Then based on founder allelic effects at the QTLs, we identified two successive conditions (pairs of lines with identical or different alleles at the QTL) and listed polymorphisms corresponding to the conditions.

2.9. Data availability

Input RIL data (genotypes and phenotypes) are provided as Supplemental data S1. MAGIC map details and genotype data are available in [16]. MAGIC phenotypes are provided in Supplemental data S2. MAGIC SNPs, polymorphisms and QTLs are deposited on the GNPis repository hosted at https://urgi.versailles.inra.fr/gnpis [45]. Genome-wide association input data (K and Q matrices as well as genotypes and phenotypic data) and results are deposited on the GNPis repository hosted at https://urgi.versailles.inra.fr/association

3. Results

3.1. QTLs in RIL population

3.1.1. A saturated map of the intraspecific RIL population

After resequencing the two parental lines of the RIL population, the new genetic map constructed with SNP markers included 501 distinct loci covering 1090 cM. The average number of markers per chromosome was 42, with an average distance between markers of 2.60 cM. The map covered 98% of the assembled tomato genome, against 70% for the genetic map obtained earlier with the same progeny [22]. In particular, coverage of chromosomes 1, 7 and 8 was improved from 9% to 99%, 15% to 99% and 30% to 99%, respectively (Supplemental data S3). This map was then used to map QTLs with the phenotype data earlier described [34] and unpublished data for FLW and HT.

3.1.2. QTLs in the RIL population

In the RIL population, 25 QTLs were detected for eight traits, explaining 8 (*flw5.1*) to 36% (*fir4.1*) of the phenotypic variation (Table 1 and Supplemental data S4). The percentage of variation explained per trait ranged from 20 (for pH) to 67% (for FW). Several clusters of QTLs were identified, particularly on chromosomes 2, 4 and 9. Most of the QTLs detected earlier [34], using the same phenotypic data and a genetic map with a lower coverage rate, were confirmed. The TA and pH QTLs, on chromosome 12, were no longer

Table 1
Characteristics of QTLs and associations detected in the 3 populations. The number of individuals and markers per population are indicated. The total number of QTLs and the number of QTLs per trait are indicated, together with the percentage of variation explained (PVE) and the minimum and maximum support intervals (CI or SI, and distance of LD decay for GWAs). The numbers of QTL common to two or three populations are indicated.

	RIL			MAGIC			GWAs			Common QTL RIL/MAGIC	Common QTL MAGIC/GWAs; RIL/GWAS	Common to 3 pops
Nb genotypes	140			397			163					
Nb markers	500			1500			5500					
Map size (cM)	1200			2100			–					
QTL number	25			63			28			16	8; 5	4
	nb QTL	PVE	Min–max CI (Mb)	nb QTL[a,b]	PVE[c,d]	Min–max SI (Mb)	nb QTL	PVE	Min–max distance of LD decay (kb)			
Fruit weight	5	0.67	1.38–48.41	11 (2, 3)	0.51–0.34	0.95–57.33	9	0.80	1.1–4097	4	3; 2	2 (*fw3.2; fw12.1*)
Soluble Solid content	2	0.36	1.62–4.74	7 (1, 0)	0.41–0.13	1.10–3.39	9	0.55	1.3–269.7	2	3; 2	2 (*ssc2.1; ssc9.1*)
Titratable acidity	4 + 1	0.59	0.89–46.14	4 (0, 0)	0.36–NA	1.89–4.16	6	0.47	3.7–152.9	2	0; 1	0
pH	1 + 1	0.20	1.09–1.86	9 (5, 0)	0.48–0.32	0.90–9.29	2	0.24	11.4–181.1	1	2; 0	0
Firmness	3	0.51	1.03–11.46	8 (2, 2)	0.48–0.42	0.50–53.19	1	0.43	96.3	2	0; 0	0
Color-b	3	0.59	1.24–8.92	8 (2, 2)	0.43–NA	1.04–59.37	1	0.43	116.8	2	0; 0	0
Flowering T3	4	0.40	2.24–62.21	9 (2, 0)	0.45–0.33	0.77–51.75				2		
Plant height	1	0.35	10.63	7 (2, 1)	0.33–0.16	0.34–51.72				1		

[a] Number of QTLs just below the threshold.
[b] Number of QTLs detected as 2nd peak.
[c] PVE for INRA location.
[d] PVE for VCo location.

detected, because of a large gap without any marker. Two new QTLs were detected on chromosomes 1 and 4, for fruit firmness (*fir1.1*) and fruit fresh weight (*fw4.1*), respectively. They were located in genomic regions weakly covered by the previous map.

3.2. QTLs in the MAGIC population

3.2.1. QTL mapping in the MAGIC population

The MAGIC linkage map is composed of 1345 SNP markers, covering 758 Mb (84% of the 900 Mb tomato genome size, and almost all the 760 Mb assembled genome [26]) and 2156 cM. It thus more than doubled the map size compared to the RIL population. We could predict the haplotype origin for an average of 89% of the MAGIC line genomes [16].

A total of 63 QTLs (corresponding to 78 QTLs over the two locations) were detected for the eight traits (Table 1 and Supplemental data S5), with four to eleven QTLs per trait. The PVE per trait ranged from 13% (for SSC in VCo) to 51% (for FW in INRA location). Lower PVE in VCo compared to INRA location was due to more homogeneous growing conditions in INRA trial. For the six traits assessed in the two locations, 15 QTLs were detected in both locations, while nine were specific of VCo and 27 of INRA location. Support intervals ranged from 5.5 (*ta6.1*) to 86 cM (*cob9.2*) and from 340 kb (*ht4.1*) to 64 Mb (*flw1.1*). Segregation of different QTLs according to the founders lead to variation in allelic effects according to the QTLs. Allele effects according to the parental line are detailed in Supplemental data S5 and illustrated for FLW in Fig. 1.

If we combine the support intervals of QTLs detected at two locations to limit the interval boundaries, nine QTLs were mapped in an interval close to or less than 1 Mb. Table 2 illustrates for these QTLs the number of genes and polymorphisms detected in the regions and the number of mutations with an effect on the protein sequence. Several thousands of polymorphisms were frequently detected, but the number with an effect on the protein was much lower. By assessing the allele effect of the eight founder lines at the QTLs, it was possible to determine combinations of parental alleles that should be similar or different. This strongly reduced the number of candidate polymorphisms. Polymorphisms with an effect on the protein sequence were detected for four QTLs (for FIR and SSC) providing a short list of candidate genes to be further studied. In some cases, we could not find any polymorphism corresponding to the condition. This could be due to missing or ambiguous sequence data or the causal variant may be due to a long Indel (not detected) or a copy number variants (CNV). The analysis of the founder genome sequences revealed several regions with CNV [20] covering 35 Mb (around 4.4% of the genome). Copy number variants were detected in at least five of the nine regions scanned in Table 2. Epigenetic modifications could also account for the QTL variation as shown in the case of the CNR gene variant [46].

3.2.2. Haplotypes, recombination and linkage disequilibrium

The LD, recombination rate and haplotype sizes will determine the power to detect genetic associations. In the MAGIC population LD decayed quickly from an average of 0.47 at 1 kb to less than 0.2 at 2 Mb, reaching a minimum of 0.08 at 20 Mb. However, for more distant markers (40 Mb), LD increased again (higher than 0.13) to fall again to previous values at distances around 50 Mb [16]. This is caused by the large centromeric regions with low recombination rate in the tomato genome that comprise around 70% of the chromosomes [30]. In natural populations used for GWAs studies LD is lower especially in the centromeric regions and baseline is reached before 50 cM [31].

Higher apparent recombination rates in the MAGIC population reduced haplotype sizes and conferred greater precision to QTL detection when compared with RIL population. The MAGIC genetic map (2156 cM) is 97.8% longer than the RIL map (1090 cM). This

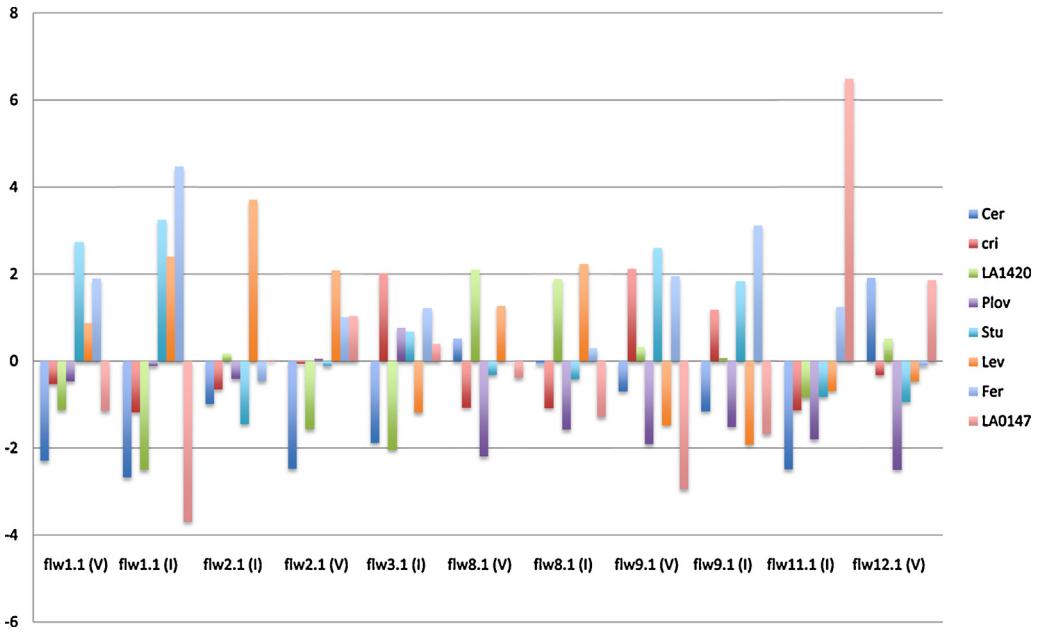

Fig. 1. Founder allelic effects at the flowering time QTLs in the MAGIC population. Centered effects for the eight parental alleles (Cervil, Levovil, Criollo, Stupicke PR, Plovdiv24A, LA1420, Ferum and LA0147, from left to right). The QTL name and location of the trial (Inra, I or VCo, V) is indicated below the QTL.

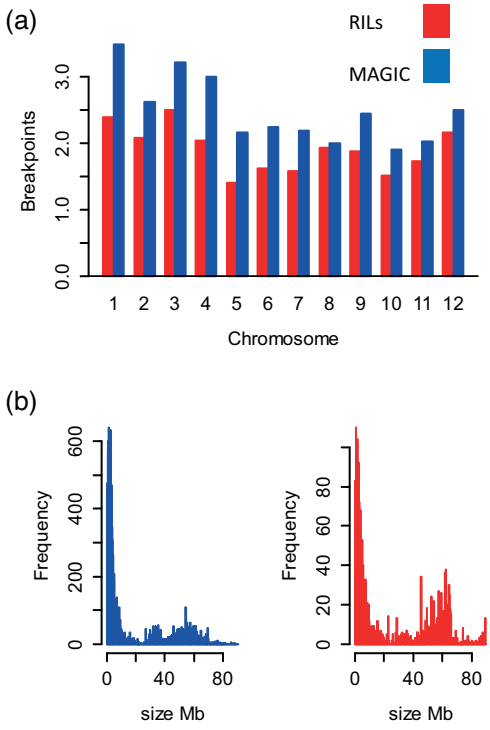

Fig. 2. Comparison of recombination in RIL and MAGIC populations. (a) Distribution of average number of break points per chromosome in RILs and MAGIC lines; (b) distribution of the haplotype block size in RILs and MAGIC lines.

is consistent with the average number of break points per line and chromosome, 2.49 and 1.46 in MAGIC and RIL population, respectively (Fig. 2a). This increase was not that obvious when we compared the haplotype block physical size, which was just 33% higher in the RIL (22.33 Mb) than in the MAGIC population (16.77 Mb). However the higher recombination rates in the MAGIC population clearly reduced the frequency of long haplotype blocks, corresponding to centromeric regions (Fig. 2b).

3.3. GWAs

3.3.1. Linkage disequilibrium, kinship and population stratification

Briefly, within the GWA experiment, we took benefit from the most recent development achieved in the estimation of LD decay by using the *LDcorSV* measurement that takes into account kinship and population stratification in the studied population [47]. Based on this measurement, the average intra-chromosomal LD estimations ranged from 0.337 in *S. pimpinellifolium* to 0.567 in *S. lycopersicum*. This demonstrated that selection tends to increase LD level especially in cultivated accessions. As earlier reported [32], the average degree of relatedness was low with an average value of 0.074 while the number of ancestral populations was estimated to be two ($K = 2$).

3.3.2. Phenotype–genotype associations

A total of 41 associations were detected for the six traits also measured in the RIL and MAGIC populations (Table 1). The number of associated loci ranged from one (for COB and FIR) to nine (for FW and SSC). In terms of genomic location, chromosomes carried varying numbers of associated SNP with chromosome 7 carrying only one association (i.e. SSC) while up to five associations were detected onto chromosome 2 for FW, SSC and TA. The estimated heritability (estimated at step 0 of the model, based on the variance component σ_g^2 computed for all markers and g representing the estimated genetic variance of the trait) and PVE ranged from 0.42 (FIR) to 0.88 (FW) and from 0.24 (pH) to 0.80 (FW), respectively. Detailed information regarding these results, such as peak SNP annotation, is reported in Supplemental data S6 and S7.

3.4. Comparison of QTLs across populations and candidate gene identification

Table 1 presents the total number of QTLs per population and QTLs detected in overlapping intervals across populations. For the six common traits, 17 of 95 QTLs were detected in at least two populations. Fig. 3 summarizes the QTLs detected in the three populations. A few chromosome regions present clusters of QTLs,

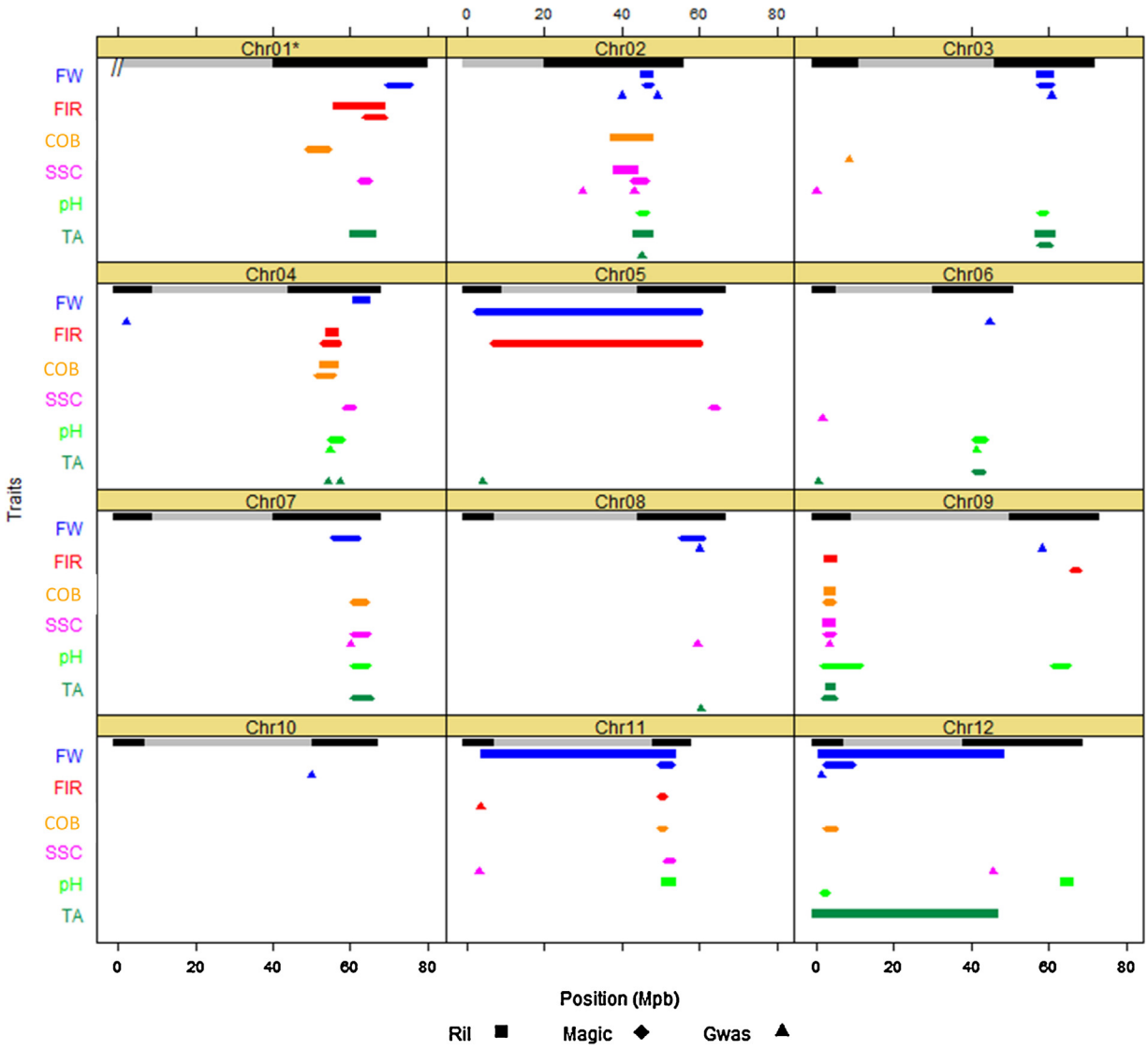

Fig. 3. Overview of fruit quality QTLs identified on the tomato genome by mapping analysis in RIL, MAGIC and GWA populations. At the top of the twelve panels, lines proportional to chromosome physical size in million base pairs (Mb) represent tomato chromosomes. Chromosome 1 is truncated of the first 20 Mbp for representation comfort (marked by // and *). Centromeric parts with low recombination frequency are indicated in gray and peripheral parts in black (according to [30]). QTLs are represented by square, diamond and triangle symbols in the RIL, MAGIC and GWAs populations, respectively. Color codes correspond to the six fruit quality traits: fresh weight (FW) in blue, firmness (FIR) in red, b color parameter (COB) in orange, soluble solids content (SSC) in pink, pH in light green and titrable acidity (TA) in dark green. For the MAGIC population, only QTLs identified by simple interval mapping are represented. Besides, when a QTL was found in two locations, only the one with the shorter confidence interval is represented.

particularly on chromosomes 2, 4, 7, 9 and 11. We then reviewed the traits considering a single QTL when SI overlapped. For FW, a total of 20 different QTLs were detected. The two previously cloned QTLs, *fw2.2* [24] and *fw3.2* [33] were detected in the three populations, confirming their major role in the difference between cherry and large-fruited tomato accessions. The QTL *fw11.2* was detected in both RIL and MAGIC populations and probably corresponds to a QTL close to the fasciated (*fas*) locus, which has been fine mapped to 149 kb [48]. On chromosome 2, several linked QTLs seem to be present in a small region as already showed [16] and this region should be precisely dissected as it contains many QTLs and major genes for fruit size, shape and sugar composition [49].

For SSC, 13 QTLs were detected. The locus *ssc9.1* was detected in the three populations. It likely corresponds to a previously cloned QTL (*Brix9.2.5*) exhibiting a polymorphism in a cell wall invertase gene (*lin5* [25]). Another QTL was detected on chromosome 2 in RIL and GWA panel around 42 Mb, and it seems linked to another QTL (*ssc2.2*) detected in the MAGIC population with a peak around 45 Mb. The smallest support interval in the MAGIC population concerned *ssc2.2*, which covered 5.8 cM and 430 kb. When looking at the allelic effect of the founders, we poorly reduced the list of candidate polymorphisms (1368 in 46 genes) as the allelic effects of the parents corresponded to the major haplotype in the region. A total of 24 polymorphisms had an effect on the coding sequence in 12 genes but we could not identify any specific candidate gene based on their annotation.

Six QTLs revealed colocations for TA and pH, which are assumed to be related traits. Thus, we considered the two traits as a single one. A total of 14 QTLs were detected, with five in at least two out of the three populations analyzed. The strongest effects concerned *ta9.1, ta3.1* (in RIL and MAGIC), *ph6.1* (in MAGIC and GWA populations) and the association on chromosome 2 (position around 45 Mb which colocalized with a QTL for TA in RIL and pH in MAGIC populations). The association on chromosome 6 could be related with that detected on the same panel for citrate (position 41,345,468) in the close vicinity of two malate transporters previously identified [32].

Table 2
Screening for candidate SNP in the MAGIC population. Number of possible causal polymorphisms underlying the MAGIC QTLs with SI smaller or around 1Mb. QTL start and end are the SI limits translated to physical bp, calculated by LOD decrease of one. Nb genes and nb pol correspond to the number of genes and polymorphisms in the SI without selection, after selection based on founder effects (as described by the different contrasts according to allelic effects), in the SI and nb pol high/moderate the number of polymorphisms with an effect on the protein sequence.

QTL	Intervals			Total polymorphisms			Selection on allelic effect						
	Start	End	Size (bp)	nb genes	nb pol	nb pol high/moderate	cond1	Nb cond1	cond2	Nb cond2	nb both conditions	in nb genes	Nb pol high/moderate
fir1.1	86,982,244	87,980,000	997,756	126	3998	224 (in 91 genes)	Cer ≠ Cri	2615	Cri = Plo	2190	1400	61 genes	71 (in 30 genes)
fir4.1	54,840,000	55,870,000	1,030,000	70	3173	171 (in 60 genes)	Fer ≠ Lev	1333	Plo = Lev	952	109	31 genes	4 (in 4 genes)
fir11.1	49,998,845	50,507,387	508,542	64	2374	133 (in 41 genes)	Fer = La0	2350	Cer = Cri	168	157	54 genes	5 (in 3 genes)
ssc2.2	43,049,064	43,480,000	430,936	46	1763	63 (in 31 genes)	Cri = Lev	1707	Cer ≠ Lev	1404	1368	46 genes	24 (in 12 genes)
ph6.1	41,150,000	42,150,000	1,000,000	129	3803	153 (in 65 genes)	Lev ≠ La0	5	La0 = Plo	3711	1	1 gene	0
ph12.1	1,921,369	2,821,857	900,488	177	4236	197 (in 105 genes)	Fer ≠ La01	0	La0 = Plo	841	0	0	0
flw9.1	64,036,532	65,053,302	1,016,770	108	4623	206 (in 76 genes)	Stu ≠ La0	82	Cri = La0	1814	27	2 genes	0
flw11.1	50,290,034	51,066,236	776,202	87	3276	107 (in 48 genes)	La0 ≠ Plo	17	Plo = Cer	240	3	2 genes	0
ht4.1	62,567,262	62,907,824	340,562	28	584	40 (in 16 genes)	LA0 ≠ Stu	0	–				

For fruit firmness, seven QTLs were detected, among which two were detected in RIL and MAGIC populations on chromosomes 1 and 4. The QTL *fir11.1*, which colocalized with *cob11.1*, had the smallest support interval (0.5 Mb). The selection based on allelic effect of founders allowed reducing the list of candidate polymorphisms from 2374 in 64 genes to 157 in 54 genes. Three genes presented polymorphisms with coding effects (Table 2). Among them a vacuolar sorting protein (Solyc11g067230) also showed a strong correlation ($r = -0.81$; $p = 0.002$) between its expression in growing fruits of parental lines and their allele effect at the QTL (unpublished data). Again, for fruit firmness, GWA signal and QTL overlapped especially on chromosome 11 supporting our results.

For COB, eight QTLs were detected in total, among which two were detected both in RIL and MAGIC populations on chromosome 4 and 9, respectively, but with large confidence intervals, while GWAs detected other associations on chromosome 3.

FLW and HT were assessed only in the RIL and MAGIC populations.

For FLW, 11 QTLs were detected, with two common to both populations on chromosomes 2 and 12. In the MAGIC population, taking advantage of the allelic effects of the QTLs with the smallest support interval allowed identifying 3 and 27 candidate polymorphisms for *flw11.1* and *flw9.1*, respectively (Table 2). For HT, seven QTLs were detected with one common to both populations on chromosome 6. The QTL *ht4.1* had a small support interval, carrying only 584 polymorphisms. None of them corresponded to the allelic pattern of the founders at the QTL, suggesting either low coverage, presence of an undetected large Indel, missing data or epigenetic effect. The QTL on chromosome 3 for these two traits could be related to a pleiotropic effect of *fw3.2* as this QTL was shown to affect also earliness and plant vigor [33].

4. Discussion

4.1. Common QTLs in the three populations

The results describe QTLs and associations detected for fruit and plant traits in three panels of (or derived from crosses between) large fruited and cherry tomato accessions. Using three different panels, we detected 71 QTLs for the six traits evaluated in the three populations, among which 17 were at least detected in two populations (Table 1). The large proportion of QTLs detected in RILs also detected in the MAGIC population was expected as the two parents of the first population were among the parents of the second. On the contrary in the GWAs panel a larger set of QTLs may segregate explaining the number of differences (32% of associations mapped in the support interval of a QTL). We must underline that we supposed that overlapping support interval corresponded to a single QTL although only fine mapping experiments could prove that two linked QTLs do not segregate as sometimes shown after fine mapping experiments [49,50].

4.2. The benefit from the genome sequence

For the first time, the availability of the reference tomato genome sequence [26] allowed the projection of the QTLs and their support intervals onto the physical map of the tomato. This allows comparing QTL positions even in populations with maps constructed with different marker sets. Several clusters and most of the QTLs fall in regions where a recent diversity study based on 360 resequenced accessions [28] identified selective sweeps due to the rise in frequency of favorable haplotypes and leading to a drastic reduction of the nucleotide diversity when comparing cherry accessions to large fruited lines. Although these 133 regions only cover 7% (54.5 Mb) of the assembled genome, 52% (33 of 63) of the

QTLs detected in the MAGIC population have their support interval in one of these regions. In the future, a special attention should be brought to these regions as they contain important genes for breeding.

The uneven distribution of crossovers in tomato, with large chunks of chromosome around centromeres which almost do not recombine, leads to a few QTLs encompassing more than 50 Mbp. Luckily such QTLs are not frequent as they represent less than 10% of the QTL. Fewer genes are present in these regions but the low recombination frequency hampers their use in tomato breeding.

The availability of a high quality reference genome sequence and the development of next generation sequencing technologies [51,52] eased the resequencing the genomes of the parental lines of the RIL and MAGIC populations and the discovery of more than 4 million polymorphisms among the eight founders with a very high level of accuracy [20]. Thus, combining founder allelic effects and SNP catalogs reduced the number of candidate polymorphisms and allowed targeting candidate genes or regions more precisely than ever. Furthermore, polymorphisms with an effect on the coding sequences are quite rare, but many QTLs have been discovered in non-coding sequences [53].

4.3. Interests and limitations of the three population types

Table 3 summarizes the main pro and cons of the use of each type of panel.

4.3.1. RIL and MAGIC populations

Briefly, RIL population are easy to set up and to analyze. They are interesting for mapping rare alleles such as disease resistance genes or other specific traits, but they lead to large support intervals due to the low recombination while the genetic background effect (epistasis) may hamper QTL detection. MAGIC populations are more complex to set up and to analyze and population size need to be much larger (at least 50 individuals per founder [17]). However, MAGIC allows better detection as a larger set of QTLs segregate among the population founders (63 QTLs *vs* 25 in RIL in our experiments performed on the same traits). Several methodologies were proposed to analyze such populations. On RILs, Composite Interval Mapping provided results very similar to Simple Interval Mapping (data not shown). For the MAGIC data, we used a regression of phenotypic values on the predicted haplotypes of the lines, as the percentage of prediction was high. We could have used the SNP alleles as in GWAs approaches [14], or intermediate approach grouping the haplotypes [18]. These approaches may lead to small differences as shown for FW in tomato [16]. One of the main interests in detecting QTLs in MAGIC population is the dissection of allelic effects when the founder genomes are sequenced. This is useful for combining positive alleles through marker-assisted selection or genomic selection and for QTL identification. In the MAGIC population, support intervals are smaller in average than in RILs (from 10 to 7 Mb on average, in our experiment). Successive inter-crossing before selfing generations could even increase the apparent growth in recombination [12,50]. The increase in recombination reduces support interval size and subsequently reduces the number of candidate genes or polymorphisms to be studied. If the population is large enough (e.g. about 1000 *Arabidopsis* accessions) then the QTLs can be directly identified [12].

4.3.2. Genome-wide association study

Genome-wide association scans have the potential to detect more precisely loci underlying the variation of traits due to the high density of markers and the rapid decay of LD even though the studied population is stratified. Thus, this approach is complementary to linkage-based approaches (either the population is bi-parental or multi-parental). In the present study, a total of 28 associations were identified for the six studied traits explaining varying part of the variance. This confirmed the polygenic architecture of the traits with a large number of small effect loci that we were not able to detect. Stringent threshold, lack of statistical significance for the control of false negatives caused by small effect sizes [8] may explain these results. The estimations of the missing heritability (Supplemental data S6) tend to support our observations and means that small to medium effect loci remain to be identified. However, for the fruit weight, the associated loci explained 80% of the variation suggesting strong effect QTLs, but the missing heritability remains high. Furthermore, identifying associated loci that have been previously cloned (*lin5*; *fw2.2*) validates our methodological approach.

When examining the annotation of the peak SNPs related to the traits (Supplemental data S6), few of these seemed to be directly related to the traits they are associated to. This means that peak SNPs are in LD with the candidate gene or polymorphism. Defining the shortest physical distance that contains the candidate gene is much more complex in GWAs than in classical linkage populations, where methods such as 1-LOD support interval or bootstrapping are commonly used to assess QTL confidence interval [36] followed by polymorphism examination. As the number of SNPs is limited (5595), we examined LD decay around each peak SNP to define a 'LD bin' in which looking for putative candidate genes provided by the tomato genome annotation. In our case, when considering an arbitrary LD decay of 0.2 around the peak SNP, the estimated length of the LD bins were ranging from 1.1 kb to 4.1 Mb with a median value of 57 kb, reflecting the different degree of LD decay in the tomato genome. Thus, looking for candidate loci in a LD bin may be time consuming or nearly impossible as hundreds of genes may be included within the same LD bin [54]. To circumvent this inherent problem in GWAs, two approaches have been proposed. The first one tends to predict a minimal genomic region around a genetic association signal within a LD bin with a high degree of accuracy by observing around an association signal LD between polymorphic markers that is known to be stronger in cases compared to controls [55]. The second approach tends to recover power in regions of high LD by whether estimating the kinship with all the markers that are not located on the same chromosome as the tested SNP or taking into account the correlation between markers to weight the contribution of each marker to the kinship [56]. Thus, as previously stated [57], the method chosen to define an associated chromosomal region influences GWAs reliability and this issue remains under investigated.

Regarding the design of the GWA study, much improvement could be achieved, especially through (1) optimizing the panel population by choosing the individuals on the basis of their relatedness to maximize its reliability, and (2) increasing the SNP density. In the context of genomic selection, an approach that discriminates which individuals must be included in the calibration set was proposed [58]. Applying this approach to define the optimal association panel would be worth testing. In parallel, increasing the number of markers would definitely help to detect more associations and reduce the missing heritability part. However, this statement largely relies onto the LD patterns of the species. In the cultivated tomato, several studies reported that LD decays over large genomic regions (up to several Mb [31,32,39,59]) limiting the interest of high-density SNP arrays, in addition to the ascertainment bias introduced by the use of SNP arrays. Increasing the SNP density would be of interest in recombining regions of the genome. However, larger set of SNP would imply more stringent threshold due to correction for multiple testing. To overcome this limitation, one way is to test for genotype–phenotype associations using haplotypes (blocks of LD) rather than single markers [60]. Thanks to NGS and imputation methods, the use of haplotypes has already been tested [61] and applied [62] in plants, demonstrating its increasing interest.

Table 3
Comparison of advantages and limitations of the RIL, MAGIC and GWA populations.

	RIL	MAGIC	GWA
Time to develop	Intermediate	Long	Short
Precision in mapping common alleles	+	++	+++
Precision in mapping rare alleles	+++	+	–
Access to recombination	+	++	+++
Nb of markers needed	+	++	+++
Population structure	No	No	+
Main advantages	Rare allele mapping Easy analysis	Several alleles segregating Founder allele effect for MAS and QTL identification	Precision due to historical recombination
Main limitations	Large QTL confidence intervals	Time to set up Large population needed	High LD limits the precision Pop structure responsible for false positive

4.3.3. Combining populations to close the genotype–phenotype gap

Overall, the interest of this study relies on combining results from linkage mapping experiments in RIL and MAGIC populations with GWA analysis to decipher the genetic architecture of traits related to the fruit quality in tomato. The combination of the populations seems efficient as not all the loci affecting these complex traits are expected to be detected in a single population because of allele specific effects. Such combination may reinforce QTL relevance and restrict the support interval for their characterization [63], as association signal as well as QTL support interval may span over large genomic regions (see above) directly reflecting the patterns of LD decay.

From a wider point of view, closing the gap between the genotype and the phenotype is not a recent idea. The pioneer work of Mitchell-Olds on *Arabidopsis* [64] clearly demonstrate the interest of combining quantitative genetics and population genetics to decipher the genetic architecture of adaptive traits and solve the "genotype-to-phenotype problem". In the same ecological context, Stinchcombe and Hoekstra [65] reviewed the advances of this combination demonstrating its power to identify candidate genes. More recently, Mitchell-Olds [66] dissected two studies in *Arabidopsis* related to flowering traits, respectively based on population [67] and pedigree (MAGIC) [12]. He suggested that an increasing number of small effect loci will be detected but also that the combination of pedigree and natural populations will elucidate the patterns of trait variation. In addition, the combination of quantitative and population genetics makes sense as breeding system, effective population size, selective history and population demography influence the genetic architecture of traits, as illustrated in *Arabidopsis* to detect genes associated to the resistance to the PPV virus [68].

4.4. Prospects: QTL characterization in genome and resequencing era

Nowadays genome sequences and next generation sequencing technologies provide a number of changes in QTL detection strategies: a number of wild tomato relatives and cultivated accessions have been resequenced [27–29]. Polymorphism discovery is no longer limiting and Genotyping-by-Sequencing [69] may allow the rapid discovery of SNPs necessary for the construction of new genetic maps at a reduced cost. It is thus possible to map new QTLs at the intraspecific level for traits differing in cultivated accessions without the large effect of major QTLs, which distinguish wild from cultivated accessions. For example, among the eight founders of the MAGIC population, the cherry tomato accessions differ from the reference genome by 1–2 million SNPs, while the four large fruited lines only differ by 180,000–350,000 SNPs. The availability of the catalog of polymorphisms among parental lines also considerably facilitated the fine mapping of candidate genes. If two lines only differ in a small number of large effect QTLs, then Bulk Segregant Analysis can be combined to NGS to speed up gene discovery. This has been shown efficient to directly identify the polymorphism responsible of major mutations [70,71] but also used for QTL mapping as illustrated in tomato for FW [28].

Genome information is also important to compare and identify QTLs. We have shown that physical positions of markers allow the projection of QTLs on a reference map independent of the progeny. It is thus possible to perform meta-analysis on a large number of studies using dedicated software [72,73] and thus reducing the support interval around the QTLs [45]. Thus, managing all the phenotype data in a common database is highly important today [74]. In tomato, the Sol Genomics Network [75] (http://solgenomics.net/) concentrates Solanaceae genome sequences, polymorphisms as well as a few QTLs and phenotypic data. Genome annotation provides gene catalog under the QTLs. Thus high quality gene annotation is also strongly needed, as a large percentage of genes are still with unknown function. Finally transcriptome (RNAseq) data on several organs, developmental stages and genotypes [76] also provide cues for the identification of candidate genes. The identification of candidate genes underlying a QTL relies on a set of arguments related to gene location, function, expression and polymorphism. When a candidate gene has expression variation linked to the phenotype in parental lines, eQTLs can be mapped and the colocation of a trait QTL, a related gene and its eQTL may confirm that a polymorphism close to the gene is responsible for its variation and putatively responsible for trait variation [77]. The validation of such guilty-by-association candidate polymorphism may not be easy by traditional transgenic approaches as knockout or overexpressing a gene may have an effect on the phenotype even though it is not the QTL. Today, the genome editing technologies make possible to precisely perform genome modifications in plants and thus validate a specific polymorphism [78]. Screening and characterizing mutants in a candidate gene in Tilling populations is another way to validate the effect of a candidate gene [79].

Taken together, all these results illustrate that finding the genes underlying the phenotype of interest is only feasible in species for which genetic information is abundant and even in model organism, such as tomato, the ability to move from QTL to QTN is still not that easy [80]. However, compared to the decade needed to clone the first QTL responsible for fruit weight (*fw2.2*) in tomato [24], biologists gained the power to prove that a variant is responsible for the trait variation with a much larger variety of genomic tools and experimental designs that speed up the process. Thus, even though the statistical approaches used in QTL and GWAs present some caveats, we are still discovering and understanding new molecular determinants underlying traits of economical and agronomic interest in crops. However, major challenges remain especially toward the understanding of the role of non-coding 'junk' DNA and epigenetic marks [81] onto the regulatory landscape of genomes and the adaptation of crops to their environmental conditions notably in their response to biotic and abiotic stresses.

5. Conclusion

Genome sequences and NGS technologies provided flood of genomic information such as genetic variants responsible of quantitative traits that we have to manage. But the remaining limitation is no longer genotyping or sequencing but is to properly phenotype in a high throughput and reproducible way. Relevant populations are now of high importance together with the phenotype precision. It is urgent to gather all QTL data in databases in order to be able to perform meta-analyses to decipher the genetic determinants of agronomic traits. Thus, we clearly demonstrated that the combination of QTL analysis (in RIL and MAGIC populations) and GWAs precisely mapped and identified the QTLs and avoided false positives. Combined to data of polymorphisms in large populations and expression profiles we should quickly identify new causal variants responsible for the variation of important traits.

Acknowledgements

We acknowledge the experimental team of INRA GAFL for their help in experimentation, Yolande Carretero, Justine Gricourt, Esther Pelpoir and Renaud Duboscq for their help in phenotyping. We thank the experimental and informatic INRA-EPGV team: Aurélie Bérard, Auréle Chauveau, Rémi Bounon, Maria Tchumakov and Elodie Marquand. This work was supported by CEA-IG/CNG, by performing the DNA QC and providing access to INRA-EPGV to their Illumina Sequencing Platform. We acknowledge groups of Anne Boland (DNA and Cell Bank service) and Marie-Thérèse Bihoreau (Illumina HT Sequencing). The ANR MAGIC-Tom SNP project 09-GENM-109G and the European Solanaceae Integrated Project EUSOL (Food-CT-2006-016214) supported this work. LP was supported by a postdoctoral INRA fellowship, EA by an INRA PhD fellowship and JD by a grant from the Embassy of France in Thailand in Junior Research Fellowship Program 2014.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.plantsci.2004.08.011

References

[1] A.H. Paterson, E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, S.D. Tanksley, Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms, Nature 335 (1988) 721–726.
[2] M.J. Kearsey, A.G.L. Farquhar, QTL analysis in plants; where are we now? Heredity (Edinb.) 80 (1998) 137–142.
[3] J.B. Holland, Genetic architecture of complex traits in plants, Curr. Opin. Plant Biol. 10 (2007) 156–161.
[4] Z.B. Lippman, Y. Semel, D. Zamir, An integrated view of quantitative trait variation using tomato interspecific introgression lines, Curr. Opin. Genet. Dev. 17 (2007) 545–552.
[5] A. Price, Believe it or not, QTLs are accurate!, Trends Plant Sci. 11 (2006) 213–216.
[6] A. Korte, A. Farlow, The advantages and limitations of trait analysis with GWAS: a review, Plant Methods 9 (2013) 29.
[7] P.K. Gupta, S. Rustgi, P.L. Kulwal, Linkage disequilibrium and association studies in higher plants: present status and future prospects, Plant Mol. Biol. 57 (2005) 461–485.
[8] P.M. Visscher, M.A. Brown, M.I. McCarthy, J. Yang, Five years of GWAS discovery, Am. J. Hum. Genet. 90 (2012) 7–24.
[9] A.L. Price, N.A. Zaitlen, D. Reich, N. Patterson, New approaches to population stratification in genome-wide association studies, Nat. Rev. Genet. 11 (2010) 459–463.
[10] I. Baxter, J.N. Brazelton, D. Yu, Y.S. Huang, B. Lahner, E. Yakubova, et al., A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1;1, PLoS Genet. 6 (2010) e1001193.
[11] J. Yu, J.B. Holland, M.D. McMullen, E.S. Buckler, Genetic design and statistical power of nested association mapping in maize, Genetics 178 (2008) 539–551.
[12] P.X. Kover, W. Valdar, J. Trakalo, N. Scarcelli, I.M. Ehrenreich, M.D. Purugganan, et al., A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*, PLoS Genet. 5 (2009) e1000551.
[13] N. Bandillo, C. Raghavan, P.A. Muyco, M.A.L. Sevilla, I.T. Lobina, C.J. Dilla-Ermita, et al., Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding, Rice (N.Y.) 6 (2013) 11.
[14] B.E. Huang, A.W. George, K.L. Forrest, A. Kilian, M.J. Hayden, M.K. Morell, et al., A multiparent advanced generation inter-cross population for genetic analysis in wheat, Plant Biotechnol. J. 10 (2012) 826–839.
[15] W. Sannemann, B.E. Huang, B. Mathew, J. Léon, Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept, Mol. Breed. 35 (2015), http://dx.doi.org/10.1007/s11032-015-0284-7
[16] L. Pascual, N. Desplat, B.E. Huang, A. Desgroux, L. Bruguier, J.-P. Bouchet, et al., Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era, Plant Biotechnol. J. (2014), http://dx.doi.org/10.1111/pbi.12282
[17] C. Cavanagh, M. Morell, I. Mackay, W. Powell, From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants, Curr. Opin. Plant Biol. 11 (2008) 215–221.
[18] N. Bardol, M. Ventelon, B. Mangin, S. Jasson, V. Loywick, F. Couton, et al., Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism, Theor. Appl. Genet. 126 (2013) 2717–2736.
[19] J.R. Klasen, H.-P. Piepho, B. Stich, QTL detection power of multi-parental RIL populations in *Arabidopsis thaliana*, Heredity (Edinb.) 108 (2012) 626–632.
[20] M. Causse, N. Desplat, L. Pascual, M.-C. Le Paslier, C. Sauvage, G. Bauchet, et al., Whole genome resequencing in tomato reveals variation associated with introgression and breeding events, BMC Genomics 14 (2013) 791.
[21] J.J. Giovannoni, Genetic regulation of fruit development and ripening, Plant Cell 16 (2004) 170–180.
[22] V. Saliba-Colombani, M. Causse, L. Gervais, J. Philouze, Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome, Genome 43 (2000) 29–40.
[23] L.M. Labate JA, S. Grandillo, T. Fulton, S. Muños, A.L. Caicedo, I. Peralta, Y. Ji, R.T. Chetelat, J.W. Scott, M.J. Gonzalo, D. Francis, W. Yang, E. van der Knaap, A.M. Baldo, B. Smith-White, M.C. Eller, J.P. Prince, N.E. Blanchard, D.B. Storey, M.R. Stevens, M.D. Robbins, J. Fen Wang, B.E. Liedl, M.A. O'Connell, J.R. Stommel, K. Aoki, Y. Iijima, A.J. Slade, S.R. Hurst, D. Loeffler, M.N. Steine, D. Vafeados, C. McGuire, C. Freeman, A. Amen, J. Goodstal, D. Facciotti, J. Van Eck, Labate_et_al_2007.pdf, in: C. Kole (Ed.), Genome Mapp. Mol. Breed. Plants, vol. 5, Veg., Springer-Verlag, Berlin Heidelberg, 2007, pp. 11–135.
[24] A. Frary, T.C. Nesbitt, S. Grandillo, E. van der Knaap, B. Cong, J.P. Liu, et al., fw2.2: a quantitative trait locus key to the evolution of tomato fruit size, Science 289 (2000) 85–88.
[25] E. Fridman, T. Pleban, D. Zamir, A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 4718–4723.
[26] Tomato-Genome-Consortium, The tomato genome sequence provides insights into fleshy fruit evolution, Nature 485 (2012) 635–641.
[27] S. Aflitos, E. Schijlen, H. de Jong, Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing, Plant J. 80 (2014) 136–148.
[28] T. Lin, G. Zhu, J. Zhang, X. Xu, Q. Yu, Z. Zheng, et al., Genomic analyses provide insights into the history of tomato breeding, Nat. Genet. (2014), http://dx.doi.org/10.1038/ng.3117
[29] A. Bolger, F. Scossa, M.E. Bolger, C. Lanz, F. Maumus, T. Tohge, et al., The genome of the stress-tolerant wild tomato species *Solanum pennellii*, Nat. Genet. 46 (2014) 1034–1038.
[30] S.C. Sim, G. Durstewitz, J. Plieske, R. Wieseke, M.W. Ganal, A. Van Deynze, et al., Development of a large SNP genotyping array and generation of high-density genetic maps in tomato, PLoS ONE 7 (2012) e40563.
[31] S.-C. Sim, A. Van Deynze, K. Stoffel, D.S. Douches, D. Zarka, M.W. Ganal, et al., High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding, PLoS ONE 7 (2012) e45520.
[32] C. Sauvage, V. Segura, G. Bauchet, R. Stevens, P.T. Do, Z. Nikoloski, et al., Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits, Plant Physiol. 165 (2014) 1120–1132.
[33] M. Chakrabarti, N. Zhang, C. Sauvage, S. Muños, J. Blanca, J. Cañizares, et al., A cytochrome P450 regulates a domestication trait in cultivated tomato, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 17125–17130.
[34] V. Saliba-Colombani, M. Causse, D. Langlois, J. Philouze, M. Buret, Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits, Theor. Appl. Genet. 102 (2001) 259–272.
[35] J.W. van Ooijen, JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations, Kyazma B.V., Wageningen, Netherlands, 2006.
[36] E. Lander, D. Botstein, Mapping Mendelian factors underlying quantitative traits using Rflp linkage maps, Genetics 121 (1989) 185–199.
[37] K.W. Broman, H. Wu, S. Sen, G.A. Churchill, R/qtl. QTL mapping in experimental crosses, Bioinformatics 19 (2003) 889–890.
[38] B.E. Huang, A.W. George, R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines, Bioinformatics 27 (2011) 727–729.

[39] J. Xu, N. Ranc, S. Muños, S. Rolland, J.P. Bouchet, N. Desplat, et al., Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species, Theor. Appl. Genet. 126 (2013) 567–581.
[40] S.-C. Sim, G. Durstewitz, J. Plieske, R. Wieseke, M.W. Ganal, A. Van Deynze, et al., Development of a large SNP genotyping array and generation of high-density genetic maps in tomato, PLoS ONE 7 (2012) e40563.
[41] O.J. Hardy, X. Vekemans, SPAGeDI. A versatile computer program to analyse spatial genetic structure at the individual or population levels, Mol. Ecol. Notes 2 (2002) 618–620.
[42] D. Falush, M. Stephens, J.K. Pritchard, Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, Genetics 164 (2003) 1567–1587.
[43] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software structure: a simulation study, Mol. Ecol. 14 (2005) 2611–2620.
[44] V. Segura, B.J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren, Q. Long, et al., An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations, Nat. Genet. 44 (2012) 825–830.
[45] D. Steinbach, M. Alaux, J. Amselem, N. Choisne, S. Durand, R. Flores, et al., GnpIS: an information system to integrate genetic and genomic data from plants and fungi, Database (Oxford) 2013 (2013) bat058.
[46] K. Manning, M. Tör, M. Poole, Y. Hong, A.J. Thompson, G.J. King, et al., A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening, Nat. Genet. 38 (2006) 948–952.
[47] B. Mangin, A. Siberchicot, S. Nicolas, A. Doligez, P. This, C. Cierco-Ayrolles, Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness, Heredity (Edinb.) 108 (2012) 285–291.
[48] Z. Huang, E. van der Knaap, Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11, Theor. Appl. Genet. 123 (2011) 465–474.
[49] L. Lecomte, V. Saliba-Colombani, A. Gautier, M.C. Gomez-Jimenez, P. Duffe, M. Buret, et al., Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato, Mol. Breed. 13 (2004) 1–14.
[50] Y.-F. Huang, D. Madur, V. Combes, C.L. Ky, D. Coubriche, P. Jamin, et al., The genetic architecture of grain yield and related traits in *Zea maize* L. revealed by comparing intermated and conventional populations, Genetics 186 (2010) 395–404.
[51] J.W. Davey, P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, M.L. Blaxter, Genome-wide genetic marker discovery and genotyping using next-generation sequencing, Nat. Rev. Genet. 12 (2011) 499–510.
[52] E.L. van Dijk, H. Auger, Y. Jaszczyszyn, C. Thermes, Ten years of next-generation sequencing technology, Trends Genet. 30 (2014) 418–426.
[53] A.A. Pai, J.K. Pritchard, Y. Gilad, The genetic and mechanistic basis for variation in gene regulation, PLoS Genet. 11 (2015) e1004857.
[54] M.I. McCarthy, J.N. Hirschhorn, Genome-wide association studies: potential next steps on a genetic journey, Hum. Mol. Genet. 17 (2008) 157–165.
[55] Z. Bochdanovits, J. Simon-Sanchez, M. Jonker, W.J. Hoogendijk, A. van der Vaart, P. Heutink, Accurate prediction of a minimal region around a genetic association signal that contains the causal variant, Eur. J. Hum. Genet. 22 (2013) 238–242.
[56] R. Rincent, L. Moreau, H. Monod, E. Kuhn, A.E. Melchinger, R.A. Malvar, et al., Recovering power in association mapping panels with variable levels of linkage disequilibrium, Genetics 197 (2014) 375–387.
[57] F. Cormier, J. Le Gouis, P. Dubreuil, S. Lafarge, S. Praud, A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (*Triticum aestivum* L.), Theor. Appl. Genet. 127 (2014) 2679–2693.
[58] R. Rincent, D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, et al., Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (*Zea mays* L.), Genetics 192 (2012) 715–728.
[59] M.D. Robbins, S.C. Sim, W. Yang, A. Van Deynze, E. van der Knaap, T. Joobeur, et al., Mapping and linkage disequilibrium analysis with a genome-wide collection of SNPs that detect polymorphism in cultivated tomato, J. Exp. Bot. 62 (2011) 1831–1845.
[60] J. Ross-Ibarra, P.L. Morrell, B.S. Gaut, Plant domestication, a unique opportunity to identify the genetic basis of adaptation, Proc. Natl. Acad. Sci. U. S. A. 104 (Suppl.) (2007) 8641–8648.
[61] A.J. Lorenz, M.T. Hamblin, J.-L. Jannink, Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley, PLoS ONE 5 (2010) e14079.
[62] I. Gawenda, P. Thorwarth, T. Günther, F. Ordon, K.J. Schmid, Genome-wide association studies in elite varieties of German winter barley using single-marker and haplotype-based methods, Plant Breed. 134 (2015) 28–39.
[63] B. Brachi, N. Faure, M. Horton, E. Flahauw, A. Vazquez, M. Nordborg, et al., Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature, PLoS Genet. 6 (2010) e1000940.
[64] T. Mitchell-Olds, The molecular basis of quantitative genetic variation in natural populations, Trends Ecol. Evol. 10 (1995) 324–328.
[65] J.R. Stinchcombe, H.E. Hoekstra, Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits, Heredity (Edinb.) 100 (Suppl.) (2008) 158–170.
[66] T. Mitchell-Olds, Complex-trait analysis in plants, Genome Biol. 11 (2010) 113.
[67] S. Atwell, Y.S. Huang, B.J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, et al., Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines, Nature 465 (2010) 627–631.
[68] G. Pagny, P.S. Paulstephenraj, S. Poque, O. Sicard, P. Cosson, J.-P. Eyquard, et al., Family-based linkage and association mapping reveals novel genes affecting Plum pox virus infection in *Arabidopsis thaliana*, New Phytol. 196 (2012) 873–886.
[69] J.A. Poland, T.W. Rife, Genotyping-by-sequencing for plant breeding and genetics, Plant Genome J. 5 (2012) 92–102.
[70] R.S. Austin, D. Vidaurre, G. Stamatiou, R. Breit, N.J. Provart, D. Bonetta, et al., Next-generation mapping of *Arabidopsis* genes, Plant J. 67 (2011) 715–725.
[71] H. Candela, R. Casanova-Sáez, J.L. Micol, Getting started in mapping-by-sequencing, J. Integr. Plant Biol. (2014), http://dx.doi.org/10.1111/jipb.12305
[72] O. Sosnowski, A. Charcosset, J. Joets, Biomercator V3: an upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms, Bioinformatics 28 (2012) 2082–2083.
[73] F.S. Khowaja, G.J. Norton, B. Courtois, A.H. Price, Improved resolution in the position of drought-related QTLs in a single mapping population of rice by meta-analysis, BMC Genomics 10 (2009) 276.
[74] D. Zamir, Where have all the crop phenotypes gone? PLoS Biol. 11 (2013) e1001595.
[75] A. Bombarely, N. Menda, I.Y. Tecle, R.M. Buels, S. Strickler, T. Fischer-York, et al., The sol genomics network (solgenomics.net): growing tomatoes using Perl, Nucleic Acids Res. 39 (2011) D1149–D1155.
[76] L.B.B. Martin, Z. Fei, J.J. Giovannoni, J.K.C. Rose, Catalyzing plant science research with RNA-seq, Front. Plant Sci. 4 (2013) 66.
[77] B.G. Hansen, B.A. Halkier, D.J. Kliebenstein, Identifying the molecular basis of QTLs: eQTLs add a new dimension, Trends Plant Sci. 13 (2008) 72–77.
[78] M.G. Palmgren, A.K. Edenbrandt, S.E. Vedel, M.M. Andersen, X. Landes, J.T. Østerberg, et al., Are we ready for back-to-nature crop breeding? Trends Plant Sci. 20 (2014) 155–164.
[79] F. Piron, M. Nicolai, S. Minoia, E. Piednoir, A. Moretti, A. Salgues, et al., An induced mutation in tomato eIF4E leads to immunity to two potyviruses, PLoS ONE 5 (2010).
[80] R. Mauricio, Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology, Nat. Rev. Genet. 2 (2001) 370–381.
[81] S. Zhong, Z. Fei, Y.-R. Chen, Y. Zheng, M. Huang, J. Vrebalov, et al., Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening, Nat. Biotechnol. 31 (2013) 154–159.