Plant Biotechnology Journal

Plant Biotechnology Journal (2015) **13**, pp. 565–577 doi: 10.1111/pbi.12282

Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era

Laura Pascual[1,†], Nelly Desplat[1,‡], Bevan E. Huang[2], Aurore Desgroux[1,§], Laure Bruguier[3], Jean-Paul Bouchet[1], Quang H. Le[4], Betty Chauchard[3], Philippe Verschave[3] and Mathilde Causse[1,*]

[1]*INRA, UR1052, Génétique et Amélioration des Fruits et Légumes, Montfavet, France*
[2]*Computational Informatics and Food Futures Flagship, CSIRO, Dutton Park, Qld, Australia*
[3]*Vilmorin, Centre de La Costière, Ledenon, France*
[4]*Vilmorin & Cie, Route d'Ennezat, Chappes, France*

Received 30 July 2013;
revised 19 September 2014;
accepted 24 September 2014.
**Correspondence* (Tel +33 432722803; fax +33 4 32 72 27 02; email mathilde.causse@avignon.inra.fr)
[†]Present address: Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UA, Universidad de Barcelona, Barcelona 08193, Spain.
[‡]Present address: BIOGEMMA, Centre de Recherche de Chappes, CS 90126, Chappes, 63720, France
[§]Present address: IGEPP, Domaine de la Motte, BP 35327, Le Rheu Cedex 35653, France.

Keywords: multiparental population, *Solanum lycopersicum*, resequencing, QTL, SNP.

Summary

Identification of the polymorphisms controlling quantitative traits remains a challenge for plant geneticists. Multiparent advanced generation intercross (MAGIC) populations offer an alternative to traditional linkage or association mapping populations by increasing the precision of quantitative trait loci (QTL) mapping. Here, we present the first tomato MAGIC population and highlight its potential for the valorization of intraspecific variation, QTL mapping and causal polymorphism identification. The population was developed by crossing eight founder lines, selected to include a wide range of genetic diversity, whose genomes have been previously resequenced. We selected 1536 SNPs among the 4 million available to enhance haplotype prediction and recombination detection in the population. The linkage map obtained showed an 87% increase in recombination frequencies compared to biparental populations. The prediction of the haplotype origin was possible for 89% of the MAGIC line genomes, allowing QTL detection at the haplotype level. We grew the population in two greenhouse trials and detected QTLs for fruit weight. We mapped three stable QTLs and six specific of a location. Finally, we showed the potential of the MAGIC population when coupled with whole genome sequencing of founder lines to detect candidate SNPs underlying the QTLs. For a previously cloned QTL on chromosome 3, we used the predicted allelic effect of each founder and their genome sequences to select putative causal polymorphisms in the supporting interval. The number of candidate polymorphisms was reduced from 12 284 (in 800 genes) to 96 (in 54 genes), including the actual causal polymorphism. This population represents a new permanent resource for the tomato genetics community.

Introduction

Identifying the genes responsible for the variation of adaptation and agronomic traits is a main goal for plant geneticists. The frequent polygenic control of these traits complicates the identification of the causal molecular variants (Morell *et al.*, 2012). Two main approaches, family-based quantitative trait loci (QTLs) mapping and genome-wide association studies (GWAS), have been employed to elucidate their genetic architecture (Mitchell-Olds, 2010).

Traditionally, family-based QTL mapping relies on populations derived from biparental crosses. These populations, like F2 and backcrosses, can be directly analysed (Clarke *et al.*, 1995; Grandillo and Tanksley, 1996), or studied after reaching homozygosity (Keurentjes *et al.*, 2007). This approach has led to the identification and positional cloning of several major genes underlying QTLs (Price, 2006). However, such populations allow only the analysis of alleles differing between two lines, and the resolution is limited to 10–30 cM, as the analysis mainly relies on recombination events taking place during the F1 meiosis (Hall *et al.*, 2010). This is true even in recombinant inbred lines (RILs), as the number of efficient recombination decreases in advanced generations. GWAS overcome the limitations of biparental crosses. Based on collections of unrelated individuals, GWAS screen a wide range of diversity and take advantage of the historical recombination events that have accumulated over thousands of generations (Korte and Farlow, 2013). This approach has been useful to identify genetic associations with complex agronomic traits (Huang *et al.*, 2012a; Sauvage *et al.*, 2014). However, GWAS are limited by linkage disequilibrium (LD), which may vary greatly from one region to the other in the genome, and by population substructure, which can result in false positive or false negative results (Mitchell-Olds, 2010; Visscher *et al.*, 2012). Moreover, some interesting phenotypes might be caused by rare alleles, which are difficult to identify by GWAS (Kover and Mott, 2012). More complex experimental populations offer alternatives to these designs and address their limitations. To increase the number of recombinations, Darvasi and Soller (1995) proposed the development of advanced intercross lines (AILs). After the development of an F2 progeny, successive generations of random mating allow the accumulation of recombination break points. To increase the genetic variation

analysed, nested association mapping (NAM) populations were developed in maize from a diverse set of parental lines crossed with a reference line (Yu *et al.*, 2008). However, the effect of genetic background and epistasis may affect QTL detection and are not taken into account in these populations (Rakshit *et al.*, 2012). To overcome these limitations, the AIL methodology was extended to multiple parent populations. This approach was first used to develop the mice heterogeneous stock (Yalchin *et al.*, 2005). Since then, it has been described by several acronyms, and many breeding designs (Rockman and Kruglyak, 2008; Valdar *et al.*, 2006). To avoid confusion, we will refer to them as multiparent advanced generation intercross (MAGIC) populations (Cavanagh *et al.*, 2008). MAGIC populations have been set up in the model species *Arabidopsis* (Kover *et al.*, 2009) and several cereal crops (Bandillo *et al.*, 2013; Huang *et al.*, 2012b) demonstrating the power of such resource to detect QTLs underlying quantitative traits.

Tomato (*Solanum lycopersicum*) is one of the most important vegetables consumed worldwide, but also the model species for studying fleshy fruit development (Giovannoni, 2004). During its domestication, the diversification of fruit aspect, as well as the adaptation to a wide range of environmental conditions, was simultaneous to a strong reduction of molecular diversity (Blanca *et al.*, 2012; Miller and Tanksley, 1990). This lack of genetic variation in the cultivated species limited the exploitation of intraspecific variation and thus led geneticists to study trait variation mostly in progenies of distant crosses involving wild species (Zamir, 2001). Recent association studies including cherry tomato accessions (*Solanum lycopersicum* var. *cerasiforme*) that have an intermediate position between cultivated tomato and its closest wild relative species (Ranc *et al.*, 2008) have shown the potential of this material to detect QTLs by GWAS (Ranc *et al.*, 2012; Xu *et al.*, 2013). The potential of MAGIC populations to include a wide range of variation and the recent publication of the tomato genome (Tomato Genome Consortium, 2012) open new avenues for the exploitation of this variation.

Here, we present the first tomato MAGIC population and describe its potential for (i) intraspecific variation exploitation, (ii) QTL mapping and (iii) causal polymorphism identification. We have constructed the MAGIC population crossing eight tomato lines, selected to include a wide range of the genetic diversity of *S. lycopersicum* species. These eight founder lines have been deeply characterized following a systems biology approach (Pascual *et al.*, 2013) and their whole genomes resequenced allowing the identification of more than 4 million SNPs (Causse *et al.*, 2013). We used this information to develop a subset of markers especially designed to analyse the MAGIC population. The selected markers were employed to develop the first intraspecific saturated map in tomato. We phenotyped the population, discovered a wide range of variation through new allelic combinations and mapped QTLs. Finally, a strategy to fine map QTLs and identify the causal polymorphisms is proposed. We demonstrate the power of the MAGIC population when coupled with available genome sequence to restrict the number of putative causal polymorphisms underlying the QTLs.

Results

A tomato MAGIC population composed of 397 MAGIC lines was constructed as described in Figure 1, following four generations of crosses and three of selfing.

The custom-made genotyping platform is highly efficient to predict founder haplotypes

To genotype the MAGIC population, we designed a specific SNP platform to enhance the haplotype prediction and recombination detection in the MAGIC population. From more than four million SNPs detected in the founder lines when compared to the tomato reference genome (Causse *et al.*, 2013), we selected a subset of 1536 markers using a filtering pipeline in three steps (Table 1). First, based on general quality score criteria, we retained 408 795 SNPs. Second, we removed SNPs providing successive similar profiles over the eight founders, as these SNPs would provide redundant information, and reduced the number of SNPs to 149 808. Finally, we selected 1536 SNPs taking into account their physical and genetic position and the profile of the adjacent SNPs to enhance the founder imputation power. A total of 1486 SNPs were finally used for genotyping the MAGIC population (Table

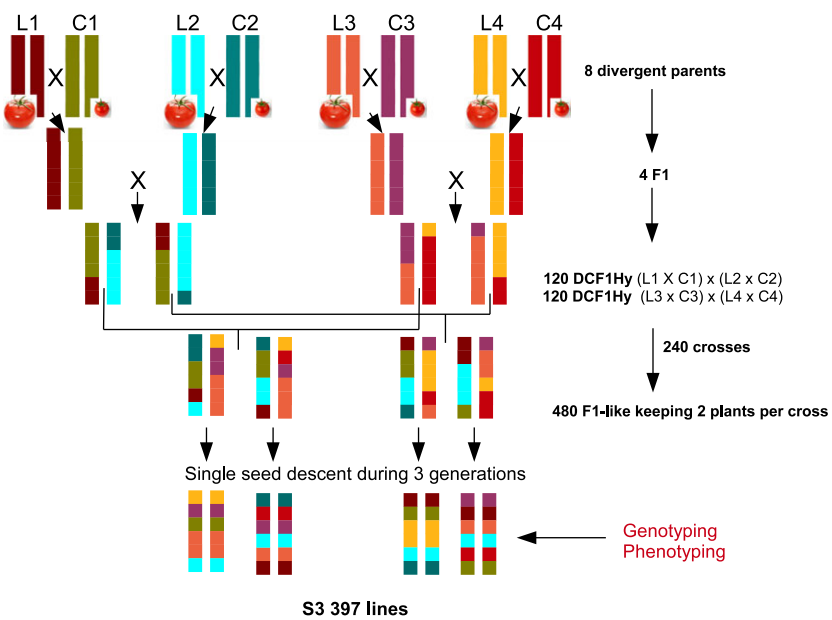

Figure 1 Construction of a tomato 8-way MAGIC population. Large fruited founders noted as L1 Levovil, L2 Stupicke PR, L3 LA0147, L4 Ferum. Small fruited founders noted as C1 Cervil, C2 Criollo, C3 Plovdiv24A, C4 LA1420. DCF1Hy: double cross F1 hybrid.

S1). The selection process allowed improving the haplotype prediction from 67% if the markers were randomly selected to 93.4% with the set of 1536 markers selected, as shown in Figure S1.

A genetic map twice as long as biparental maps

The MAGIC tomato population was then used to construct a genetic map. The final map included 1345 markers (Table S2), representing 524 unique map positions (genetic bins) with average intervals of 1.68 cM and 0.6 Mb (Table 2). The total map measured 2156 cM and covered 758 Mb (84% of the 900 Mb tomato genome size), and almost all the 760 Mb assembled genome (Tomato Genome Consortium, 2012). The 12 chromosomes were covered by 28 to 64 genetic bins. We did not find any clear correlation between genetic and physical map length, as, for example, chromosome 3 (215.2 cM), the longest, and chromosome 10 (120.96 cM), the smallest, covered 64.77 Mb and 64.8 Mb, respectively. When we compared physical and genetic positions, high recombination rates were found on the distal regions, while recombination was almost suppressed in large centromeric regions that comprised around 70% of the chromosomes (Figure 2, Figure S2).

We compared the MAGIC genetic map and the biparental tomato high-density genetic maps constructed by Sim *et al.* (2012a) (Table 2, Figure S2). The EXPEN 2012 map includes 3687 markers and was based on 160 F2 from a cross between Moneymaker (*S. lycopersicum*) and LA0716 (*S. pennellii*). The EXPIM 2012 map includes 4792 markers mapped with 183 F2 individuals from a cross between Moneymaker and LA0121 (*S. pimpinellifolium*). The MAGIC tomato map was 87% and 105% longer than the EXPEN 2012 and EXPIM 2012 map, respectively. This increase was not the same for all chromosomes, ranging from 43% to 155% with respect to EXPEN 2012 and 60% to 185% with respect to EXPIM 2012 (Table S3). Figure 2 illustrates the relationship between physical and genetic positions for the first three chromosomes. The recombination increase is limited to distal parts of the chromosomes as recombination in the centromeric regions is almost suppressed in all three maps (Figure S2). Genetic recombination also increased when compared to intraspecific tomato biparental populations. The MAGIC tomato map was 69% larger than the map constructed from a RIL population developed from a cross between Cervil and Levovil, the two most distant founders of the MAGIC population (data not shown).

Table 1 Summary of the SNP selection procedure for the construction of the genetic map

Chromosome	Total (1)	Quality filtering (2)	Successive profile filtering (3)	Final (4)
Ch1	140 192	22 790	9724	172
Ch2	274 273	37 292	9795	149
Ch3	357 900	47 148	5388	181
Ch4	505 272	45 926	24 068	166
Ch5	616 803	44 364	21 093	124
Ch6	109 945	14 480	4533	104
Ch7	385 516	48 566	16 409	112
Ch8	540 631	44 049	15 141	87
Ch9	411 088	18 348	7072	122
Ch10	63 524	10 232	3888	91
Ch11	499 546	29 416	13 984	106
Ch12	293 639	46 184	18 713	122
Total	4 198 329	40 8795	149 808	1536

(1) Total number of SNPs detected in the eight founders compared to the reference genome, (2) number of SNPs after filtering based on quality criteria, (3) number of SNPs after filtering against successive identical profiles and (4) final selection based on physical and genetic distances.

No clear structure remained in the MAGIC population

The structure and LD in the MAGIC population will determine the power to detect genetic associations. The population structure was assessed using the 1345 SNP markers included in the genetic map. According to the Evanno *et al.* (2005) test, the most probable number of groups was one, indicating the absence of subgroups in the MAGIC population (Figure S3). LD was analysed between pairs of markers within each chromosome, and pairwise r^2 was plotted against genetic and physical distances between loci (Figure S4). With respect to genetic distances, LD within chromosomes fell to < 0.7 within 5 cM, and < 0.3 within 25 cM, intersecting with LD baseline value only at 90 cM. With respect to physical distance, LD decayed quickly from an average of 0.47 at 1 kb to < 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. The kinship tended to be small with a third quartile of the values lower than 0.042, even though this value reached 0.8 for some pairs of lines (0.01%) (Table S4).

Finally, the haplotype structure of each line was analysed by identifying their founder allele at each marker (Figure 3, Figure S5). We predicted, on average, the marker origin for 89% of the genome (Figure 3a). This value was higher for all the chromosomes but 5 and 11, where three parental lines (Stupicke PR, LA0147 and Levovil) were difficult to distinguish. Along the chromosomes, haplotypes were predicted with lower accuracy at the end of the centromeric regions, probably due to the augmentation in recombination rates (Figure S5). Founder contribution was close to the expected value (0.125) along all the chromosomes, showing an increase in the contribution of LA1420 and Criollo coupled with a decrease of Cervil at the end of chromosome 4 (Figure 3b).

We then estimated the haplotype (segment inherited from a single founder) size and the number of haplotypes per chromosome for each line. With respect to genetic distances, haplotypes tended to be small (Figure 4a), with 25% of them being smaller than 11 cM and a median of 30 cM when all the chromosomes are considered together. The physical size distribution showed a bimodal distribution (Figure 4b), where most of the haplotypes were smaller than 5 Mb. An average size of 16 Mb was due to the large centromeric regions. The median number of haplotypes per chromosome was 4 or 3 depending of the chromosome. When all the chromosomes were considered together, the average number of recombination break points was 29.3, with all the tomato lines carrying variable parts of the eight founder lines.

Fruit weight QTL detection in the MAGIC population

Fruit weight (FW) is a key trait selected since domestication which has been widely studied in tomato. Several FW QTLs and associations have been already described (Grandillo *et al.*, 1999; Xu *et al.*, 2013) and two of them positionally cloned

Table 2 Characteristics of the genetic map. Number of SNP markers, coverage in cM and Mb for each chromosome. Comparison of map length with tomato biparental maps

Chr.	Number of markers	Unique bins	Coverage (cM)	Marker interval (cM)		Coverage (Mb)	Marker interval (Mb)		Expansion in MAGIC (cM)	
				Max.	Average		Max	Average	EXPEN 2012	EXPIM 2012
1	156	58	214.4	13.20	1.38	90.16	7.13	0.58	83%	68%
2	131	47	200.2	14.53	1.54	49.57	4.22	0.38	82%	150%
3	161	64	215.2	25.58	1.34	64.77	9.52	0.40	104%	99%
4	147	64	200.1	13.31	1.37	63.37	6.08	0.43	85%	115%
5	111	40	168.2	22.93	1.53	64.99	6.36	0.59	76%	89%
6	85	36	140.9	20.94	1.68	45.96	7.28	0.55	61%	111%
7	99	35	191.0	21.02	1.95	65.11	7.68	0.66	155%	130%
8	79	33	129.4	17.68	1.66	62.97	7.69	0.81	68%	67%
9	109	40	206.1	16.69	1.91	67.64	7.57	0.63	113%	185%
10	73	28	120.9	13.83	1.68	64.80	10.21	0.90	43%	60%
11	90	40	183.3	13.50	2.06	53.18	10.23	0.60	86%	99%
12	104	39	186.2	19.55	2.00	65.47	13.80	0.64	88%	121%
Total	1345	524	2156	–	–	758.00	–	–	–	–

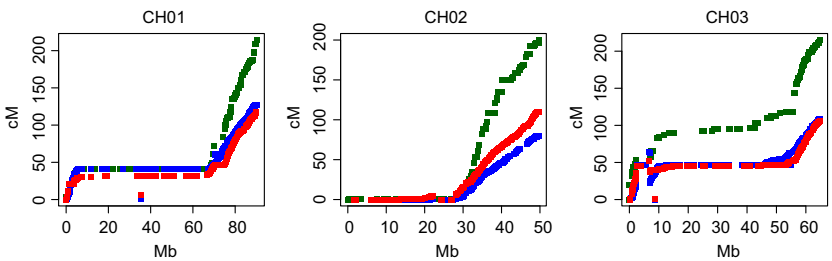

Figure 2 Relationship between the genetic and physical positions for the first three chromosomes. Positions are indicated in green for the MAGIC map, blue for the EXPIM 2012 and red for the EXPEN 2012 (adapted from Sim *et al.*, 2012a,b).

(Chakrabarti *et al.*, 2013; Frary *et al.*, 2000). We thus chose FW as an example to analyse the power and precision of the MAGIC population for QTL mapping. The phenotypic characterization was performed at two locations in the south of France. In each location, the complete set of 397 RIL MAGIC lines (one plant per line) and five replicates of each founder were characterized (Table S5). FW distributions (Figure 5) illustrated the large range of phenotypic variation in the population, including transgressive lines, as well as a difference in average FW among locations. We thus analysed the data separately and then compared the QTLs obtained in each location. To detect QTLs and genetic associations, two approaches were tested, interval mapping adapted to MAGIC populations and GWAS.

To map QTLs by simple interval mapping (IM), we performed a joint Wald test for the significance of all founder effects at putative QTL positions along the genome. At location A, nine QTLs on chromosomes 2, 3 5, 7, 8 and 11 were detected (Table 3, Figure 6b). Support intervals (SI) ranged from 6 to 78 cM. The 78 cM interval actually corresponded to the second QTL peak detected on chromosome 11, including the first QTL peak inside the SI. At physical scale, SI ranged from 0.94 Mb to 5.61 Mb, except for the QTL located on chromosome 5 (57 Mb) that covered the centromeric region. Finally, after fitting all the QTLs in a unique model, we determined that the nine QTLs together explained 51% of the trait variation. At location B, three QTLs were detected on chromosomes 2, 3 and 11 (Table 3, Figure 6d). All these QTLs colocalized with those of location A QTLs (Table 3). For the QTLs on chromosomes 2 and 3, SI were 3 and 4 cM larger than for location A, but this difference is buffered when translated to physical distance. For the QTL on chromosome 11, SI was 14 cM (0.48 Mb) larger. The three QTLs explained 34% of the trait variation when fitted all together in a unique model.

According to the population structure analysis, the MAGIC population did not present any clear structure in subgroups; thus, we performed GWAS taking into account the kinship (Table S4) in a mixed linear model (MLM). To reduce the false positive associations, p-values were corrected to account for multiple testing, and only associations with corrected p-value lower than 0.05 were considered significant. At location A, 35 significant associations were detected (Figure 6a, Table S6), located on chromosomes 1, 2, 3, 5, 11 and 12 (Figure S6). When we compared QTL SI and the associations detected by MLM, no significant markers were detected for the QTL SI on chromosomes 7 and 8. By contrast, associations were detected by MLM on chromosomes 1 and 12 where no QTLs were called by IM (Table S4, Figure 6a,b).

At location B, we identified 30 associations on chromosomes 2, 3, 11 and 12 (Table S6, Figure 6c, Figure S6), 60% also identified at location A. When we compared the IM results with GWAS analysis, we detected associated markers along all the QTLs SI and associations were also found on chromosome 12 (Table S6, Figure 6c,d).

QTL effects and causal polymorphisms detection

The effect of each founder allele was calculated by IM for each QTL with respect to a reference founder, LA0147 (*S. lycopersi-*

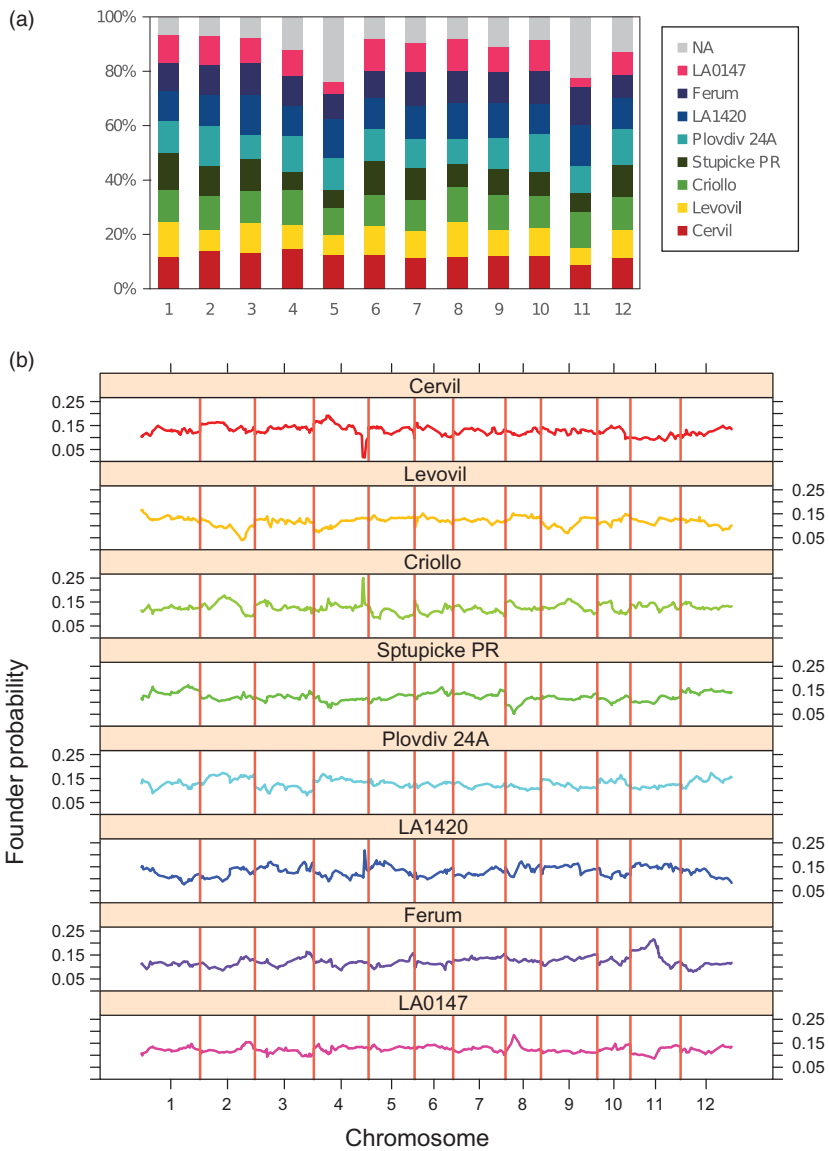

Figure 3 (a) Proportion of founder allele predicted in each chromosome. (b) Percentage of genome-wide founder assignment along the chromosomes.

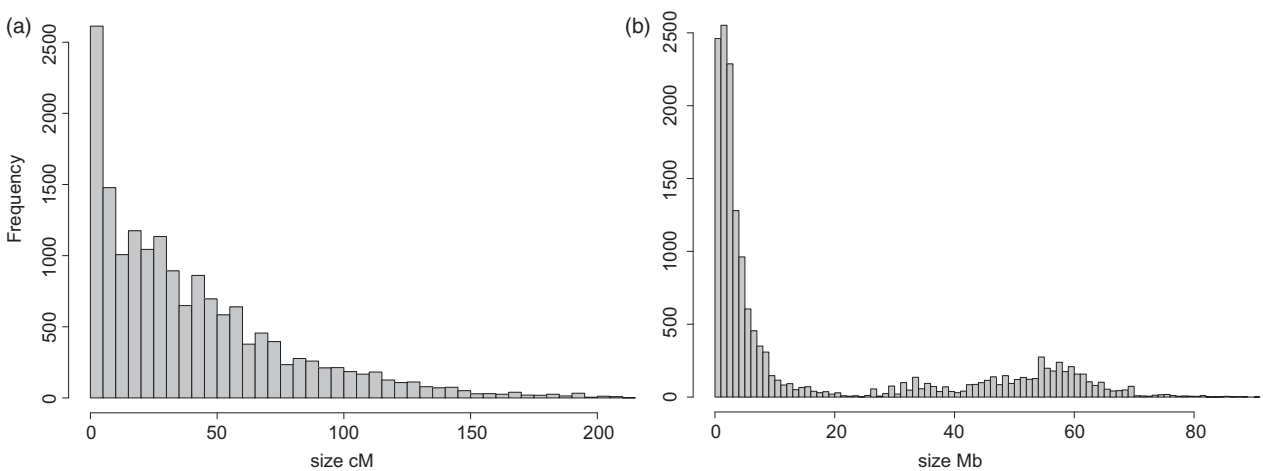

Figure 4 Size of haplotype blocks for all the MAGIC lines and chromosomes, relative to (a) genetic distances (cM), (b) physical distances (Mb).

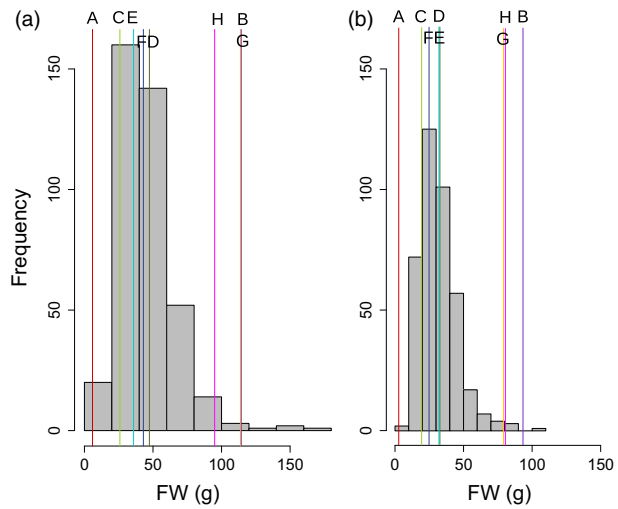

Figure 5 Distribution of fruit weight (gr) in the MAGIC lines grown in (a). Avignon. (b). La Costière. Founder trait values are indicated with vertical lines (A Cervil, B Levovil, C Criollo, D Stupicke PR, E Plovdiv24A, F LA1420, G Ferum, H LA0147).

cum). This line was chosen as reference because it was the closest to Heinz1706, the line used to sequence the tomato genome and then to detect polymorphism in the founder sequences. Effects were variable among QTLs (Figure 7). Cervil (the founder with the smallest fruits) alleles decreased or did not change fruit weight for all the QTLs. The other founder alleles decreased or increased fruit size depending on the QTL considered.

We hypothesized that coupling the QTL founder allele effects with the polymorphisms detected along the QTL SI should facilitate the identification of putative causal variants underlying the QTLs. We tested this hypothesis with two FW QTLs already cloned that colocalized with QTLs detected at both locations. For one of them, *fw3.2*, the polymorphism responsible for the phenotype has been recently identified in a cytochrome P450 gene, (Solyc03 g114940 at position 58 852 276; Chakrabarti *et al.*, 2013). The support interval for the corresponding QTL (detected at both locations) on chromosome 3 ranged between markers X03_58386463 and X03_60392846 and could correspond to the same gene. According to genomic annotations this interval comprised 800 genes (Tomato Genome Consortium, 2012). Between the two flanking markers, 12 284 SNPs and INDELs have been identified in the eight founders. We filtered out the polymorphisms according to expected allelic effects (Figure 7). Based on QTL effects, we supposed that Cervil, Criollo and LA1420 should have the same allele at the QTL, while Plovdiv24A, Stupicke PR and LA0147 should have the opposite allele. This procedure allowed us to reduce the number of polymorphisms to 96 SNPs and 3 INDELs covering 54 genes (Table S7). The list included the SNP corresponding to the one identified by Chakrabarti *et al.* (2013) as responsible for the *fw3.2* QTL.

For the other cloned QTL, *fw2.2* (Solyc02g090730, position 46 832 171), the causal polymorphism has not yet been identified (Frary *et al.*, 2000). The confidence interval for the QTL that colocalized with the gene ranged between markers X02_46353818 and X02_47498009. In this interval, 6510 polymorphisms have been identified. According to the founder effects, the alleles from the two lines, Cervil and Stupicke PR, clearly reduced fruit weight with respect to the reference. Thus, we first screened for polymorphisms common to these lines and different from the rest of the founders. We ended up with 18 SNPs and 1 INDEL. However, none of them colocalized or was near to the *fw2.2* gene (Table S8). When we analysed all the polymorphisms located inside or close to the *fw2.2* gene (Table S9), we identified a total of 43 SNPs and 3 INDELs, all of them being polymorphic only in one founder line (Cervil). This suggested the presence of two linked QTLs in the region, *fw2.2* being specific to Cervil.

Discussion

We have constructed and analysed a MAGIC population in tomato, one of the most important vegetables consumed worldwide and the model plant for fleshy fruit. Quantitative trait

Table 3 Characteristics of QTLs detected for fruit weight by interval mapping in the MAGIC population. (A) Phenotyping data from location A (Avignon). (B) Phenotyping data form location B (La Costiere)

Chr.	Pos.	LeftMrk	RightMrk	SI (cM)	SI (Mb)	*P*-value
A						
2	186	X02_47433596	X02_47498009	179–192	46.35–47.49	0
2	152	X02_42399961	X02_42773566	150–156	42.39–43.47	2.47×10^{-11}
3	178	X03_58754293	X03_58846611	170–190	57.98–60.24	8.62×10^{-7}
3	202	X03_62140362	X03_62287203	168–207	57.98–63.29	1.41×10^{-5}
5	50	X05_05638011	X05_05886227	16–72	2.8–60	1.12×10^{-4}
7	148	X07_60966290	X07_61091852	102–156	57.73–61.53	9.84×10^{-5}
8	70	X08_56902554	X08_57091589	58–86	56.26–58.33	8.63×10^{-6}
11	154	X11_51548415	X11_51631459	152–161	51.35–52.66	2.49×10^{-8}
11	118	X11_48934628	X11_49059536	105–183	47.57–53.18	5.41×10^{-6}
B						
2	184	X02_47433596	X02_47498009	179–192	46.35–47.49	4.44×10^{-16}
3	176	X03_58754293	X03_58846611	170–190	57.98–60.24	1.47×10^{-5}
11	148	X11_51176762	X11_51308212	137–160	50.5–52.29	1.16×10^{-9}

Including chromosome (Chr), position of maximum *P*-value peak (Pos in cM), left and right markers flanking the peak position, 1-LOD support interval (SI) in cM and Mb, *P*-value at the peak position.

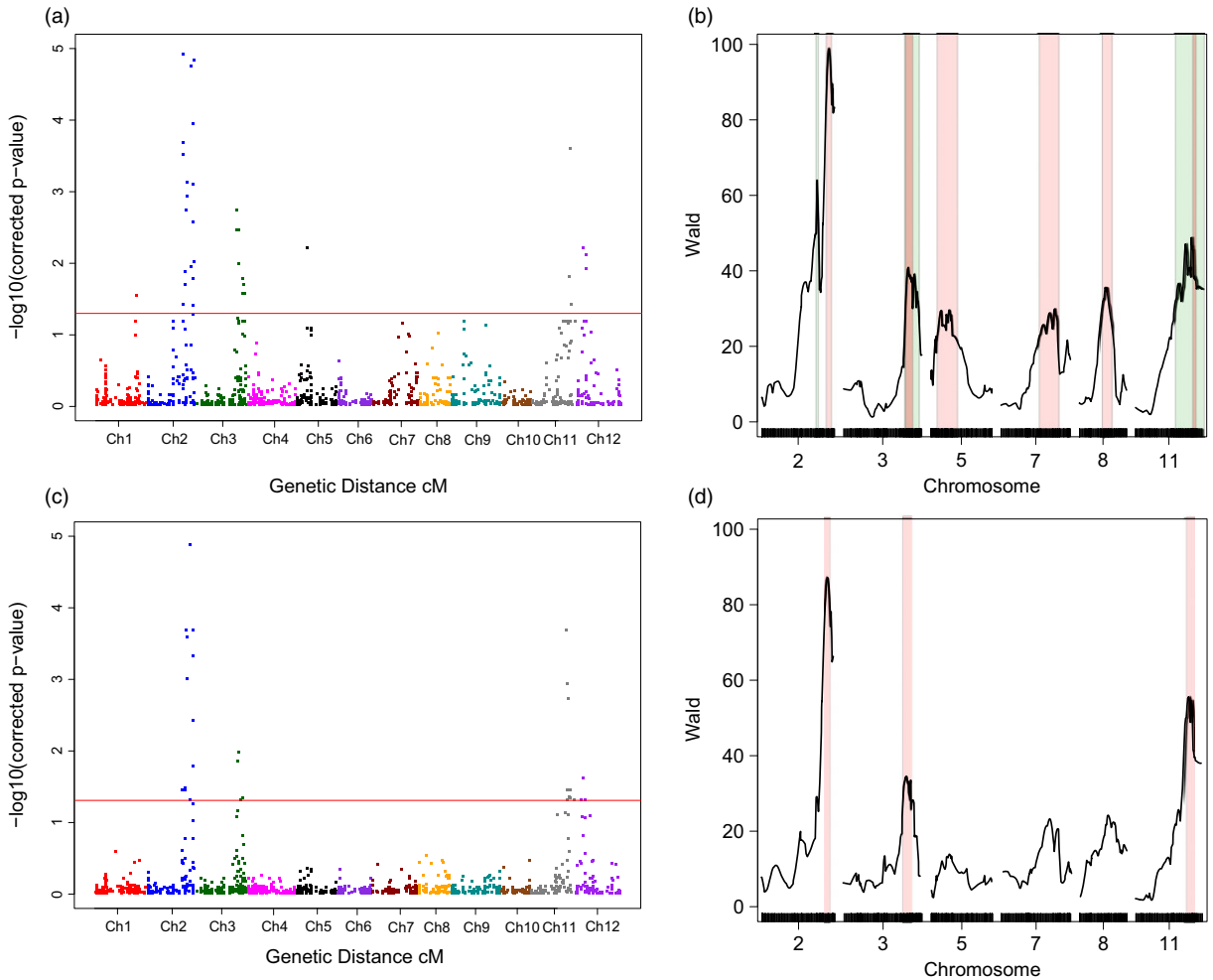

Figure 6 QTL detection in the MAGIC population for FW. (a and c) Manhattan plot (MLM) showing corrected p-values, red line indicates significance threshold. (b and d) Wald test profile on chromosomes with significant QTL. Red regions indicate 1-LOD support intervals, green regions indicate 1-LOD support intervals for QTLs detected from the original QTL profile with the function *findqtl2*. (a and b): location A (Avignon). (c and d): location B (La Costière).

dissection has been especially challenging for species like tomato, with low genetic diversity and strong population structure. The lack of intraspecific genetic variation has led to the study of interspecific progenies involving related species, but the potential of intraspecific variation remains poorly explored. The MAGIC population enabled the construction of the first high-density intraspecific map in tomato. We assessed the power of such population to generate new variation through new allelic combinations and to map QTLs. Finally, we showed its potential to identify putative causal variants and closely linked QTLs.

The population presented here together with the whole genome sequences of the founder lines constitute a useful resource for the scientific community, which overcomes the main disadvantages linked to collections of accessions employed for GWAS or to the interspecific biparental populations developed until now (Table 4). The main issues when using natural populations to detect genetic associations are (i) the presence of population substructure, (ii) variable LD block length along chromosomes and (iii) unbalanced allele frequencies (Visscher *et al.*, 2012). In tomato, the structure is especially strong (Blanca *et al.*, 2012; Ranc *et al.*, 2008). Indeed, the species can be divided in two major groups: *S. lycopersicum* and *S. lycopersicum var. cerasiforme,* and FW is highly correlated to the classification (Ranc *et al.*, 2012). We used four accessions of each group as founders of the population. The population showed a wide range of phenotypes increasing the phenotypic range of the founders, without any remaining subgroups. The design used to develop the MAGIC population effectively mixed the genomes of all the founder genotypes, leading to balanced allele frequencies along the genome. In contrast to other MAGIC populations (Huang *et al.*, 2012b), < 1.5% of the markers showed segregation distortion, even though some founder lines carried introgressions from distant wild species (Causse *et al.*, 2013).

LD extent, as well as the specific size and genomic structure of the centromeres, impacts the number of markers needed to detect QTLs and should be taken into account when designing genotyping strategies. The tomato genome structure characterized by very low recombination rates in the centromeric regions that comprise around 70% of the chromosomes (Sim *et al.*, 2012a) greatly affected the LD among markers. LD decay is slower in the MAGIC population than in natural populations, where the LD baseline is reached before 50 cM (Sauvage *et al.*,

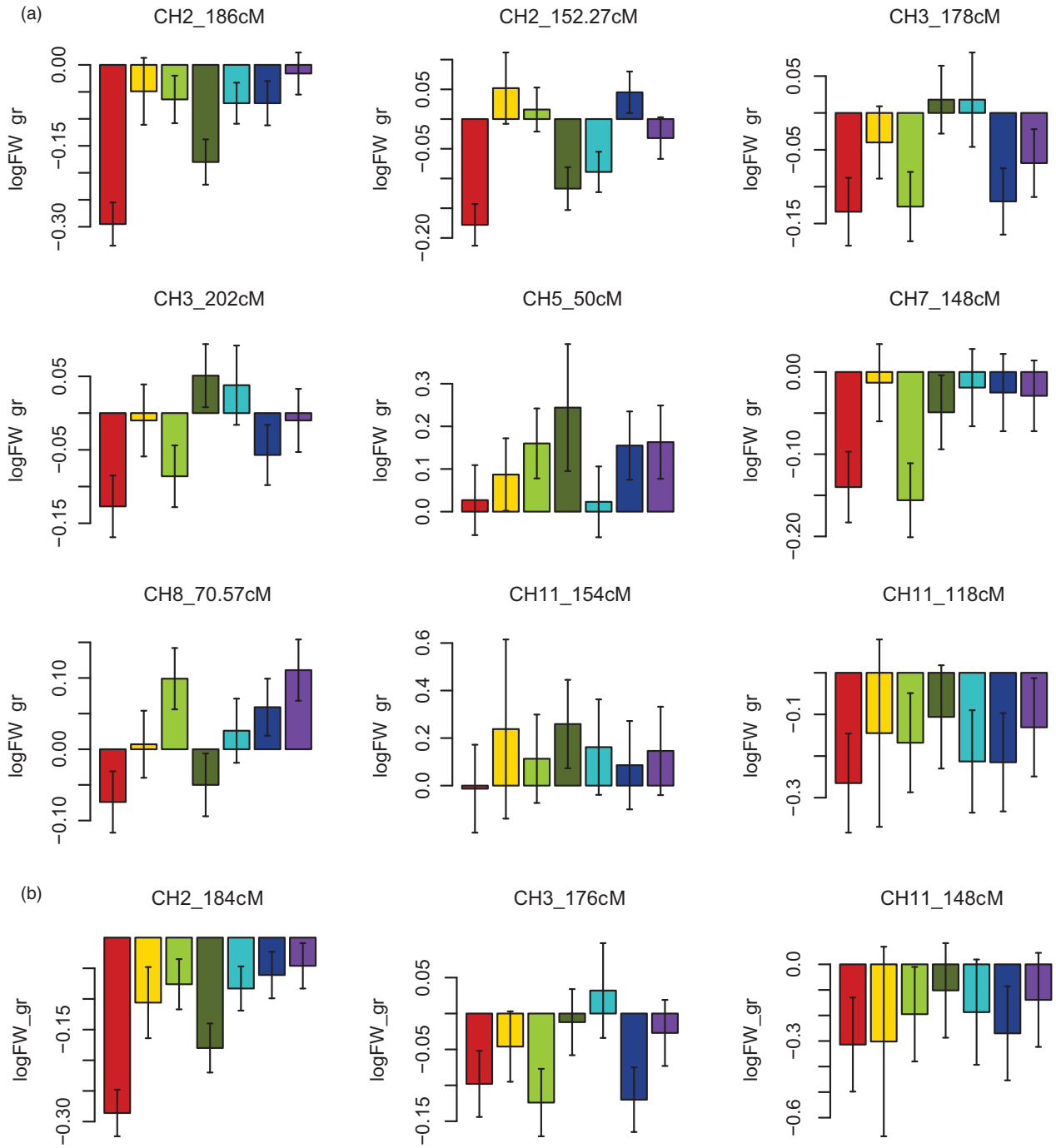

Figure 7 Founder allelic effects at the fruit weight QTL. Effects and standard errors for the alleles of Cervil, Levovil, Criollo, Stupicke PR, Plovdiv24A, LA1420 and Ferum (from left to right), relative to LA0147. QTLs detected with data from location A (Avignon) and B. (La Costière).

2014; Sim *et al.*, 2012b). Compared to MAGIC populations developed in other species like *Arabidopsis,* where the LD baseline (0.05) is reached at 15 Mb (Kover *et al.*, 2009), our population reached the baseline at longer physical distances (40–60 Mb) due to the size of centromere regions.

The selection of the SNPs greatly impacted the power of haplotype prediction along the MAGIC line genomes and thus the chances to detect genetic associations through the haplotype reconstruction (Huang and George, 2011; Mott *et al.*, 2000). The availability of the founders' sequences allowed haplotype imputation in most of the genomic regions. To characterize the remaining problematic regions, it may be necessary to characterize the population by genotyping by sequencing (GBS) as it was performed in rice (Bandillo *et al.*, 2013), as long as the founder lines are not identical by descent. In such a case, clustering similar founder haplotypes may increase statistical power to detect QTLs, as shown by Bardol *et al.* (2013).

The MAGIC map size was 87% larger than biparental maps (Sim *et al.*, 2012a), showing the effectiveness of MAGIC population to enhance recombination and admixture. This increase

Table 4 Comparison of advantages and limits of biparental, MAGIC and association populations

	Advantages	Limits
Biparental progeny	Rapid to set up Useful for mapping rare alleles Easy analysis	Limited to two contrasting alleles Few recombination generations Large QTL support interval
MAGIC population	Several alleles and QTLs segregating Higher precision than biparental population Rapid fine mapping Useful for candidate SNP screening No population structure Suitable for selection	Time to establish Require more markers and larger populations than biparental population
Association panel	Existing collections, high diversity Natural recombination When LD limited, recise mapping	Require many markers Population structure When high LD, coarse mapping Rare alleles poorly identified

subsequently reduces the QTL support intervals when translated to physical size, facilitating QTL fine mapping and candidate gene selection. However, recombination was not increased in the centromere regions, comprising around 70% of the chromosomes but <40% of the genes (Sim *et al.*, 2012a). Therefore, any QTL detected in/over the centromere regions will still have a very large support interval in physical distance, and the absence of recombination will make positional cloning impossible. Luckily, the detection of QTLs in these regions should be an exception, as most of the genes are located in the chromosome extremities (Tomato Genome Consortium, 2012).

To test the power of this resource to map QTLs, we analysed the fruit weight distribution in the population grown in two locations. Genetic associations were detected by IM and GWAS (accounting for population kinship). Using both methods and locations, we detected associations with markers close to the already cloned FW QTLs (Chakrabarti *et al.*, 2013; Frary *et al.*, 2000), showing the precision of QTL mapping. By IM, we detected 9 QTLs on six chromosomes at location A, among which three were also detected at location B. Five of the QTLs detected at location A (one at location B) were not detected in a biparental population derived from the cross between the two most distant founders of the MAGIC population (Saliba-Colombani *et al.*, 2001), but colocalized with QTLs already detected in interspecific populations (reviewed by Grandillo *et al.*, 1999). This showed that increasing the variability in the founders allows the discovery of new QTL.

GWAS analysis by MLM method taking into account the kinship may avoid the detection of false positives (Visscher *et al.*, 2012) but failed to identify two of the QTLs identified by IM at location A (on chromosomes 7 and 8) that colocalized with already known QTLs (Causse *et al.*, 2004; Grandillo *et al.*, 1999; van der Knaap and Tanksley, 2003). On the other hand, we detected new associations on chromosome 1 at location A and on chromosome 12 at both locations. A QTL located at the top of chromosome 12 has been detected in the biparental progeny derived from two of the founders (Saliba-Colombani *et al.*, 2001), suggesting, that at least in this case, MLM method was more powerful than IM. This might occur in genomic regions where it is not possible to distinguish among several founders, as MLM relies on a biallelic model, while IM is comparing eight different haplotypes. For QTLs where the markers have an allele shared by founders with the same phenotypic value, MLM might be more powerful. On the contrary, if the allele is not shared among founders with the same phenotype, it will be impossible to detect association by MLM. Working with haplotypes, IM avoids this problem and permits the calculation of each founder allele effect for each QTL.

Allelic effects varied among QTLs, lines and locations. The same allelic effect was never shared by all the small or all the large fruited founders. The final fruit size is thus obtained from a specific combination of founder alleles. The knowledge of QTL allelic effects, coupled with the availability of the MAGIC population, constitutes a highly valuable resource to develop strategies for breeding and develop models to conduct genomic selection. On one hand, it is composed by a set of highly admixed lines from which we can obtain the genomic breeding values and develop models that encompass most of the species diversity (Morell *et al.*, 2012). On the other hand, it allows the validation of model predictions, as it contains most of the possible allelic combinations. Additionally, the discovery of a large number of QTLs on different chromosomes allows the development of interesting breeding schemes as alternative allelic combinations could be used to create specific phenotypes (Rosyara *et al.*, 2013).

Using a MAGIC population derived from resequenced founders allows to design strategies to identify causal polymorphisms. We showed that the list of candidate genes for a QTL interval can be strongly reduced and putative causal variants identified by coupling QTL effects with the founders' genome sequences. We analysed the genomic sequences underlying the two QTLs that colocalized with already cloned genes (*fw3.2*, Solyc03 g114940, Chakrabarti *et al.*, 2013 and *fw2.2*, Solyc02 g090730, Frary *et al.*, 2000). For *fw3.2,* we discarded 99.21% of the SNPs and INDELs located in the QTL SI by selecting variants differing between the three founders that equally decreased fruit weight and the three with opposite effect. The residual variants included the causal SNP identified by Chakrabarti *et al.* (2013). For the region around *fw2.2*, the causal polymorphism has not yet been identified. However, Frary *et al.* (2000) indicated that the phenotype was probably caused by one or more changes upstream in the promoter region. When we selected the variants that differed in the SI between the two founders that decreased fruit weight and the other founders, we did not find any variant linked to Solyc02 g090730. The analysis of all the variants surrounding the gene revealed that only one founder was different from the other lines in the region. However, there was another founder whose allele reduced fruit size but to a smaller extent. This might be caused by another gene located in close proximity. The presence of two linked QTLs might bias the estimation of founder effects and should be taken into account when looking for causal variants.

However, even if MAGIC populations allow QTL mapping at a subcentimorgan scale, this range might correspond to hundreds of kilobases. Thus, to identify the polymorphism underlying a QTL, it is still necessary to produce new recombinant plants and conduct positional cloning. The tomato MAGIC population was characterized after three selfing generations. The lines still carry residual heterozygosity that can be directly used to that end.

In conclusion, we have created the first intraspecific population of highly recombinant lines in tomato. This population segregates for many traits and can be analysed in different environments, providing a permanent resource to analyse the basis of phenotypic traits. We have illustrated its power for future fine mapping experiments. Our study has also highlighted the potential of the availability of the founder genome sequences. On the one hand, it enabled the efficient selection of a subset of markers especially designed to analyse the MAGIC population. On the other hand, it permitted to drastically restrict the number of putative causal polymorphisms underlying the QTLs.

Experimental procedures

Founder lines selection and population construction

Eight tomato lines, thoroughly characterized at different molecular and physiological levels (Pascual *et al.*, 2013), were selected as founders of an eight-way MAGIC population. Founders included four *S. lycopersicum* (Levovil, Stupicke PR, LA0147 and Ferum) and four *S. lycopersicum var. cerasiforme* (Cervil, Criollo, Plovdiv24A and LA1420) lines. These lines were chosen to maximize the genetic diversity, based on a previous molecular characterization of 360 tomato accessions (Ranc *et al.*, 2008).

Four crosses were performed between one *S. lycopersicum* line and one *S. lycopersicum var. cerasiforme* (Figure 1). F1 hybrids Levovil × Cervil and Stupicke PR × Criollo were crossed to obtain 120 DCF1Hy (Double Cross F1 hybrid) plants, while LA0147 × Plovdiv24A was crossed with Ferum × LA1420 to obtain another set of 120 DCF1Hy plants. The two subsets of 120 plants were intercrossed via 240 independent crosses using each DCF1Hy plant once as father and once as mother. Two offsprings per cross were then kept, producing 480 individuals (F1-like), each bearing parts of the eight founder genomes. These plants were propagated by single seed descent during three selfing generations, to create the set of 397 MAGIC lines (F4-like) employed in this study (Figure 1).

Genetic marker selection and construction of a MAGIC custom-made SNP platform

More than four million SNPs were detected by resequencing the genomes of the eight founder lines (Causse *et al.*, 2013; Raw sequences deposited in ENA, accession numbers ERR327646 to ERR327656; SNPs and INDELs identified deposited in ENA, accession numbers ERZ015686 to ERZ015701). Among all these polymorphisms, 1536 SNPs were selected to construct the genetic map. The selection of the SNPs was performed with custom Python scripts (available upon request to the corresponding author) using a three steps filtering strategy. First, we selected 'the best quality SNPs' based on nucleotide prediction reliability, whether the SNPs were biallelic and the quality of the flanking sequences (60 bp). In the second step, we kept only one position for each set of successive positions with the same allelic profile over the eight lines, as they would provide redundant information. Finally, SNPs were selected from successive intervals with a maximum genetic distance of 8 cM. Genetic distances were assessed for each physical position by a linear regression using only the five nearest markers from EXPEN 2000 map (solgenomics.net) and giving to each marker a weight proportional to the physical distance from the candidate position. For this step, we performed a recursive search to select a combination of physical positions that had between them a minimal genetic distance (1–0.5 cM). Each combination should have balanced allele frequencies, enhancing founders with less SNPs. Besides, the combination of marker alleles should be specific for each founder, enhancing haplotype prediction.

DNA isolation and molecular marker genotyping

DNA was isolated from young leaves of each founder line and the 397 lines of the MAGIC population using DNeasy 96 Plant kit supplied by Qiagen (Hilden, Germany). From the 1536 SNPs selected to develop the genotype platform, 1486 passed the KASPar manufacturing quality control (K-Biosciences/LGC Genomics, Molsheim, France) and were finally used for genotyping the MAGIC population (Table S1). Genotyping was performed using the Fluidigm 96.96 Dynamic Arrays according to the manufacturer's protocol using the genotyping EP1 System (San Francisco, CA). Fluorescence intensity was measured with the EP1 reader (Fluidigm Corp, San Francisco, CA), and genotypic calls were made using the Fluidigm SNP Genotyping Analysis program (Fluidigm, 2011). All genotype calls were manually checked for accuracy and ambiguous data points that failed to cluster were scored as missing data. Heterozygous markers (caused by the residual heterozygosity in the MAGIC lines) were also scored as missing data.

Genetic map construction, recombination event prediction and LD analysis

The genetic map was constructed using the R package *mpMap* (Huang and George, 2011) version 1.24.3 and the available information of the physical location of the SNPs from the tomato reference genome version SL2.40 (Tomato Genome Consortium, 2012). First, we filtered out the markers with missing data in the founders, markers with more than 20% missing data in the population and markers with a segregation distortion *P*-value $< 5 \times 10^{-9}$. MAGIC lines with more than 20% missing markers were also removed. These lines were later included for QTL analysis.

Second, recombination fraction between each pair of markers was estimated with the '*mpestrf*' *mpMap* function. This function maximizes the likelihood of observing data from a pair of markers over discrete values of recombination fractions (default values). Then, markers were grouped with the '*mpgroup*' *mpMap* function. Markers in each groups were checked against the tomato reference genome and groups assigned to the different tomato chromosomes. Markers that were not in their expected chromosome were discarded. Then, markers were ordered within linkage groups, based on their physical position in the tomato reference genome.

Third, genetic distances were estimated with the '*computemap*' *mpMap* function, using a 15-marker window and Haldane distances computed. The genetic distances among markers were plotted against the physical distances for each chromosome, and when inconsistency was found, problematic markers were removed and genetic distances re-estimated.

For each line, we calculated the multipoint probability that the genotype at a marker location (and positions spaced every 2 cM) was inherited from each of the eight founders ('*mpprob*' function from *mpMap*). Recombination events were imputed at locations where the founder allele changed along the chromosome. To assess the differences between MAGIC and biparental populations, we compared our genetic map with the tomato high-density genetic maps, constructed by Sim *et al.* (2012a) using interspecific biparental F2 tomato populations.

Multiallelic linkage disequilibrium (LD) r^2 was estimated between each pair of markers using the multipoint probabilities with the *'mpcalcld'* function as described by Huang *et al.* (2012b). We compared the values for r^2 between markers along the chromosomes and analysed the LD decay over genetic and physical distances, plotting r^2 values against the distances by chromosomes. Values were fitted by nonlinear regression.

Phenotypic data

The population was grown in two locations in the south of France in Avignon (location A) and La Costière (location B). 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 Pascual *et al.* (2013). Fruit weight (FW) was evaluated from a minimum of 10 ripe fruits per genotype, harvested from truss two to six. Before detecting the FW QTLs, a $\log_{10}$ transformation was carried out to normalize the phenotype.

Population structure and association analysis

Population structure was inferred with Structure v2.1 software (Pritchard *et al.*, 2000) in the complete MAGIC population using the 1345 mapped SNP markers. We used the admixture model for the ancestry of individuals and linkage to correct for the effect of nearby markers. The structure was modelled with a burn-in of 2.5×10^5 cycles followed by 10^6 Markov chain Monte Carlo repeats. Probabilities were estimated for population number (k) between 1 and 20, computing 10 replicates for each k. The Evanno transformation method (Evanno *et al.*, 2005) was then used to detect the number of subgroups in the MAGIC population. Kinship matrix was calculated with the SPAGeDi software (Hardy and Vekemans, 2002). According to Yu *et al.* (2006), the diagonal matrix was set to two and the negative values were set to zero. Association analysis was performed independently for FW measured in each location. Analyses were conducted with the TASSEL version 3.0 software (Bradbury *et al.*, 2007) employing mixed linear models (MLM) incorporating the kinship among individuals. The p-values were adjusted with the Benjamini and Hochberg (2000) procedure using the R package '*multtest*' (Pollard *et al.*, 2012). Associations with an adjusted P-value < 0.05 were considered significant.

QTL detection and identification of putative causal variants

QTLs were called by simple interval mapping (IM) with the *'mpIM'* function from the R package *mpMap* (Huang and George, 2011). This function computes founder effects at a step size of 2 cM with a regression approach, based on the multipoint probabilities computed with '*mpprob*' function. Then, it performs a joint Wald test for the significance of all founder effects at each putative QTL position along the genome. QTLs were called when p-values were smaller than the empirical threshold p-value (1.72×10^{-4}) derived using the function '*sim.sigthr*' after computing 1000 permutations, to reflect a genome-wide significance threshold of 0.05. This approach called a single QTL by chromosome, so when the QTL profile showed more than one QTL peak by chromosome, multiple QTLs were considered significant when peaks were separated by more than 20 cM and the LOD score dropped by more than one. QTL support intervals were determined with a 1-LOD drop support. After QTL detection, QTL effects were simultaneously fitted in a single model with the function '*fit*' from R *mpMap* package to estimate the percentage of phenotypic variation explained by the QTLs. In order to test the potential of the MAGIC population to detect causal variants, we analysed two QTLs that colocalized with already cloned genes. All the polymorphisms (Causse *et al.*, 2013) present in the QTL support intervals were analysed and filtered according to the estimated founder effects.

Acknowledgements

We thank Yolande Carretero, Justine Gricourt, Frédérique Bitton, Esther Pelpoir and Renaud Duboscq for their help in phenotyping. We also thank the experimental team and Yolande Carretero for taking care of the plants in the greenhouse. This work was supported by the ANR MAGIC-Tom SNP project 09-GENM-109G. LP was supported by a postdoctoral INRA fellowship.

References

Bandillo, N., Raghavan, C., Muyco, P.A., Sevilla, M.A.L., Lobina, I.T., Dilla-Ermita, C.J., Tung, C.W., McCouch, S., Thomson, M., Mauleon, R., Singh, R.K., Gregorio, G., Redoña, E. and Leung, H. (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice*, **6**, 11.

Bardol, N., Ventelon, M., Mangin, B., Jasson, S., Loywick, V., Couton, F., Derue, C., Blanchard, P., Charcosset, A. and Moreau, L. (2013) Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementaries between models based on parental haplotype and single locus polymorphism. *Theor. Appl. Genet.* **126**, 2717–2136.

Benjamini, Y. and Hochberg, Y. (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Stat.* **25**, 60–83.

Blanca, J., Canizares, J., Cordero, L., Pascual, L., Diez, M.J. and Nuez, F. (2012) Variation revealed by SNP genotyping and morphology provides insight into the origin of the tomato. *PLoS ONE*, **7**, e48198.

Bradbury, P.J., Zhang, D., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Causse, M., Duffe, P., Gomez, M.C., Buret, M., Damidaux, R., Zamir, D., Gur, A., Chevalier, C., Lemaire-Chamley, M. and Rothan, C. (2004) A genetic map of candidate genes and QTLs involved in tomato fruit size and composition. *J. Exp. Bot.* **55**, 1671–1685.

Causse, M., Desplat, N., Pascual, L., Le Paslier, M.C., Sauvage, C., Bauchet, G., Bérard, A., RémiBounon, R., Tchoumakov, M., Brunel, D. and Bouchet, J.P. (2013) Whole genome profiles of nucleotide variations reveal breeding and introgression events in 8 lines of tomato. *BMC Genomics*, **14**, 791.

Cavanagh, C., Morell, M., Mackay, I. and Powell, W. (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* **11**, 215–221.

Chakrabarti, M., Zhang, N., Sauvage, C., Muños, S., Blanca, J., Cañizares, J., Diez, M.J., Schneider, R., Mazourek, M., McClead, J., Causse, M. and van der Knaap, E. (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proc. Natl Acad. Sci. USA*, **110**, 17125–17130.

Clarke, J.H., Mithen, R., Brown, J.K. and Dean, C. (1995) QTL analysis of flowering time in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **248**, 278–286.

Darvasi, A. and Soller, M. (1995) Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*, **141**, 199–1207.

Evanno, G., Regnaut, S. and Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620.

Fluidigm (2011) *Fluidigm SNP Genotyping Analysis*. San Francisco: Fluidigm Corp.

Frary, A., Nesbitt, T.C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B. and Tanksley, S.D. (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 547685–547688.

Giovannoni, J.J. (2004) Genetic regulation of fruit development and ripening. *Plant Cell*, **16**, S170–S180.

Grandillo, S. and Tanksley, S.D. (1996) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. *Theor. Appl. Genet.* **92**, 935–951.

Grandillo, S., Ku, H.M. and Tanksley, S.D. (1999) Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theor. Appl. Genet.* **99**, 978–987.

Hall, D., Tegstrom, C. and Ingvarsson, P.K. (2010) Using association mapping to dissect the genetic basis of complex traits in plants. *Brief. Funct. Genomics*, **9**, 157–165.

Hardy, O.J. and Vekemans, X. (2002) SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes*, **2**, 618–620.

Huang, B.E. and George, A.W. (2011) R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics*, **27**, 727–729.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., Fan, D., Lu, Y., Weng, Q., Liu, K., Zhou, T., Jing, Y., Si, L., Dong, G., Huang, T., Lu, T., Feng, Q., Qian, Q., Li, J. and Han, B. (2012a) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39.

Huang, B.E., George, A.W., Forrest, K.L., Kilian, A., Hayden, M.J., Morell, M.K. and Cavanagh, C.R. (2012b) A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol. J.* **10**, 826–839.

Keurentjes, J.J., Bentsink, L., Alonso-Blanco, C., Hanhart, C.J., Blankestijn-De Vries, H., Effgen, S., Vreugdenhil, D. and Koornneef, M. (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics*, **175**, 891–905.

van der Knaap, E. and Tanksley, S.D. (2003) The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato. *Theor. Appl. Genet.* **107**, 139–147.

Korte, A. and Farlow, A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, **9**, 29.

Kover, P.X. and Mott, R. (2012) Mapping the genetic basis of ecologically and evolutionarily relevant traits in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **15**, 212–217.

Kover, P.X., Valdar, W. and Trakalo, J. (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5**, 7.

Miller, J.C. and Tanksley, S.D. (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor. Appl. Genet.* **80**, 437–448.

Mitchell-Olds, T. (2010) Complex-traits analysis in plants. *Genome Biol.* **10**, 113.

Morell, P.L., Buckler, E.S. and Ross-Ibarra, J. (2012) Crop genomics: advances and applications. *Nat. Rev. Genet.* **13**, 85–96.

Mott, R., Talbot, C.J., Turri, M.G., Collins, A.C. and Flint, J. (2000) A method for fine mapping quantitative trait loci in outbred stocks. *Proc. Natl Acad. Sci. USA*, **97**, 12649–12654.

Pascual, L., Xu, J., Biais, B., Maucourt, M., Ballias, P., Bernillon, S., Deborde, C., Jacob, D., Desgroux, A., Faurobert, M., Bouchet, J.P., Gibon, Y., Moing, A. and Causse, M. (2013) Deciphering genetic diversity and inheritance of tomato fruit weight and composition through a systems biology approach. *J. Exp. Bot.* **64**, 5737–5752.

Pollard, K.S., Gilbert, N.H., Ge, Y., Taylor, S. and Dudoit, S. (2012) *multtest: Resampling-based multiple hypothesis testing*. R package version 2.14.0.

Price, A.H. (2006) Believe it or not, QTLs are accurate!. *Trends Plant Sci.* **11**, 213–216.

Pritchard, J.K., Stephens, M. and Donelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rakshit, S., Rakshit, A. and Patil, J.V. (2012) Multiparent intercross populations in analysis of quantitative traits. *J. Genet.* **91**, 111–117.

Ranc, N., Munos, S., Santoni, S. and Causse, M. (2008) A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (Solanaceae). *BMC Plant Biol.* **8**, 130.

Ranc, N., Muños, S., Xu, J., Le Paslier, M.C., Chauveau, A., Bounon, R., Rolland, S., Bouchet, J.P., Brunel, D. and Causse, M. (2012) Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *cerasiforme*. *G3*, **2**, 853–864.

Rockman, M.V. and Kruglyak, L. (2008) Breeding designs for recombinant inbred advanced intercross lines. *Genetics*, **179**, 1069–1078.

Rosyara, U.R., C.A.M. Bink, M., van de Weg, E., Zhang, G., Wang, D., Sebolt, A., Dirlewanger, E., Quero-Garcia, J., Schuster, M. and Iezzoni, A.M. (2013) Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. *Mol. Breed.* **32**, 875–887.

Saliba-Colombani, V., Causse, M., Langlois, D., Philouze, J. and Buret, M. (2001) Genetic analysis of organoleptic quality in fresh market tomato. 1. Mapping QTLs for physical and chemical traits. *Theor. Appl. Genet.* **102**, 259–272.

Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Thi Do, P., Nikoloski, Z., Fernie, A.R. and Causse, M. (2014) Genome wide association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol.* **165**, 1120–1132.

Sim, S.C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., Van Deynze, A., Hamilton, J.P., Buell, C.R., Causse, M., Wijeratne, S. and Francis, D.M. (2012a) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE*, **7**, e45520.

Sim, S.C., Van Deynze, A., Stoffel, K., Douches, D.S., Zarka, D., Ganal, M.W., Chetelat, R.T., Hutton, S.F., Scott, J.W., Gardner, R.G., Panthee, D.R., Mutschler, M., Myers, J.R. and Francis, D.M. (2012b) High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS ONE*, **7**, e40563.

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Valdar, W., Flint, J. and Mott, R. (2006) Simulating the collaborative cross: power of QTL detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics*, **172**, 1783–1797.

Visscher, P.M., Brown, M.A., McCarthy, M.I. and Yang, J. (2012) Five years of GWAS Discovery. *Am. J. Hum. Genet.* **90**, 7–24.

Xu, J., Ranc, N., Muños, S., Rolland, S., Bouchet, J.P., Desplat, N., Le Paslier, M.C., Liang, Y., Brunel, D. and Causse, M. (2013) Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species. *Theor. Appl. Genet.* **126**, 567–581.

Yalchin, B., Flint, J. and Mott, R. (2005) Using progenitor strain information to identify quantitative trait nucleotides in out bred mice. *Genetics*, **171**, 673–681.

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S. and Buckler, E.S. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203–208.

Yu, J., Holland, J.B., McMullen, M.D. and Buckler, E.S. (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics*, **178**, 539–551.

Zamir, D. (2001) Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* **2**, 983–989.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 Power of founder allele prediction with simulated data from: (a) subset of SNP randomly chosen among the original set of 4 million SNPs without filtering, (b) subset of SNPs randomly chosen among the 149 808 SNPs after filtering by quality criteria and removing successive SNPs with equal profile, (c) subset of SNPs after filtering by quality criteria and removing successive SNPs with equal profile with MAF value > 0.375, (d) final SNPs selection.

Figure S2 Relationship between genetic and physical positions within each chromosome.
Figure S3 Analysis of the structure in the MAGIC population.
Figure S4 Estimates of LD vs. distance for all the markers within chromosomes relative to (a) physical distance, (b) genetic distance. Plots were fitted by nonlinear regression (red curve).
Figure S5 Founder haplotypes estimated for each line at each position on chromosome 1.
Figure S6 Linkage disequilibrium among the markers of the chromosomes with significant associations.
Table S1 Characteristics of SNP markers included in the genotyping platform.
Table S2 MAGIC Tomato map.
Table S3 Comparison of MAGIC and bi-parental F2 maps.
Table S4 Kinship values between each pair of lines in the MAGIC population.
Table S5 Average fruit weight values for each line in the MAGIC populations and the founder lines (5 replicates each), at location A (Avignon) and location B (La costière).
Table S6 Significant associations for fruit weight estimated with MLM model incorporating the kinship among individuals.
Table S7 Remaining polymorphisms located in the QTL region on chromosome 3 (position 178 cM) after filtering by founder effects.
Table S8 Remaining polymorphisms located in the QTL region on chromosome 2 (position 186 cM) after filtering by founder effects.
Table S9 Polymorphisms located around the fw2.2 gene (Solyc02g090730).