Plant Science xxx (2015) xxx–xxx

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 xxx

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
0168-9452/Published by Elsevier Ireland Ltd.

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