ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2011) **9**(2); 185–188
doi:10.1017/S1479262111000086

A comparison of population types used for QTL mapping in *Arabidopsis thaliana*

Joost J. B. Keurentjes[1,2], Glenda Willems[3,4], Fred van Eeuwijk[1,2], Magnus Nordborg[3,5] and Maarten Koornneef[1,4]*

[1]*Wageningen University, Wageningen, The Netherlands,* [2]*Centre for Biosystems Genomics, Wageningen, The Netherlands,* [3]*University of Southern California, Los Angeles, CA, USA,* [4]*Max Planck Institute for Plant Breeding Research, Cologne, Germany and* [5]*Gregor Mendel Institute, Vienna, Austria*

Abstract

In Arabidopsis, a variety of mapping populations have been used for the detection of quantitative trait loci (QTLs) responsible for natural variation. In this study, we present an overview of the advantages and disadvantages of the different types of populations used. To do this, we compare the results of both experimental and natural populations for the commonly analysed trait flowering time. It is expected that genome wide association (GWA) mapping will be an increasingly important tool for QTL mapping because of the high allelic richness and mapping resolution in natural populations. In *Arabidopsis*, GWA mapping becomes ever more facilitated by the increasing availability of re-sequenced genomes of many accessions. However, specifically designed mapping populations such as recombinant inbred lines and near isogenic lines will remain important. The high QTL detection power of such experimental populations can identify spurious GWA associations, and their unique genomic structure is superior for investigating the role of low-frequency alleles. Future QTL studies will therefore benefit from a combined approach of GWA and classical linkage analysis.

Keywords: *Arabidopsis thaliana*; flowering time; natural variation; quantitative trait loci mapping

Introduction

The analysis of the loci controlling genetic variation for quantitative traits is performed by so-called quantitative trait loci (QTL) analysis, where trait values are compared between molecular marker genotypes, which are linked to different alleles of causal genes. For this type of analysis, large populations with accurate and dense genetic marker maps as well as accurate phenotyping are required. The latter can be achieved when one can work with homozygous lines, which can be analysed in replications. Such materials can be derived by repeated selfing of initially heterozygous individuals or by the use of naturally occurring inbred genotypes. Immortal mapping populations have also the advantage that experiments analysing different traits and conditions can be compared on the basis of QTL map positions. This has already led to the identification of new pleiotropic functions of known genes (Koornneef *et al.*, 2004). In this study, we will describe some of the properties and characteristics of a number of immortal populations used in *Arabidopsis thaliana*.

Types of mapping populations used in *Arabidopsis*

The quantitative variation in complex traits and the control by heritable factors are broadly acknowledged in *Arabidopsis* (Alonso-Blanco and Koornneef, 2000). This observation, together with the ease to generate segregating material and the availability of genomic and

* Corresponding author. E-mail: maarten.koornneef@wur.nl

genetic resources, led to a manifold of QTL studies aiming at the detection of the genes underlying this natural variation (reviewed in Alonso-Blanco *et al.*, 2009).

Although genetic analysis can be performed on any type of segregating population, this review focuses on collections of homozygous lines. Widely used are recombinant inbred lines (RILs) derived by single seed descent from F_2 individuals, which are the progeny of a hybrid between two distinct homozygous (inbred) lines. The largest disadvantage of inbred lines, compared with natural (outbreeding) populations, is the limited number of effective meiotic recombination events. To overcome this limitation, large population sizes are needed to sample enough meiotic events. Alternatively, one can artificially increase the number of meiotic events and thereby the resolution to map loci by additional rounds of selective outcrossing. Several RIL populations have been derived after inter-crossing F_2 plants and later generations (Balasubramanian *et al.*, 2009). However, even after repeated inter-crossing, RIL populations still suffer from slowly decaying linkage disequilibrium (LD), which together with their often limited population size makes that identified QTLs have rather large confidence intervals. Resolution can be increased by using large populations, or when QTL analyses of different populations can be combined as was recently demonstrated for the analysis of seed dormancy in six RIL populations (Bentsink *et al.*, 2010).

Studying only one QTL at the time avoids complications of the segregation of multiple loci (e.g. epistasis). This can be achieved by the construction of introgression or near isogenic lines (NILs), in which small chromosomal regions from a donor parent are introduced in the genetic background of a recurrent parent (Koumproglou *et al.*, 2002; Keurentjes *et al.*, 2007; Torjek *et al.*, 2008). NILs, although often providing less accurate map positions, allow the detection of minor QTLs that cannot be detected in RILs (Keurentjes *et al.*, 2007). The use of heterogeneous inbred line families selected from early generation inbred lines that are still heterozygous in the region of interest (Tuinstra *et al.*, 1997) achieves the same goal, which is often referred too as 'Mendelizing' a QTL. Apart from the inaccuracy of the map positions, biparental populations have as disadvantage that they allow only the analysis of genetic variation present between the two parents.

The solution for both mapping accuracy and the analysis of more variation is the use of genome wide association (GWA) mapping (Nordborg and Weigel, 2008). This method uses (in practice a subset of) all variants available and aims to associate trait differences with specific genotypes. Due to the fast LD decay in natural *Arabidopsis* accessions, often less than 10 Kbp (Kim *et al.*, 2007), this procedure requires dense genetic maps (Weigel and Mott, 2009). Although a powerful approach, a number of factors limit the successful use of GWA. One is the often present population structure resulting in false positive associations. Compensating for the effects of population structure often also removes true positives. These so-called false negatives (Brachi *et al.*, 2010) do show up and are therefore much better studied in experimental populations. Another important limitation is that low-frequency alleles, even when having strong effects, remain undetected. An example of this is the *CRY2* allele from the Cvi accession that results in a major QTL in crosses (El-Assal *et al.*, 2001; Brachi *et al.*, 2010) but is not detected in GWA analyses (Atwell *et al.*, 2010).

Intermediate mapping populations that combine the advantages of both approaches consist of experimental crossings with increasing numbers of parents, sizes of populations and recombination events. One solution is the combined analysis of multiple populations as was demonstrated by nested association mapping in maize, where 25 RIL populations were generated by crossing diverse parents with a common reference (Buckler *et al.*, 2009). In *Arabidopsis*, this was demonstrated at a smaller scale, using L*er* (Bentsink *et al.*, 2010) or Col (Brachi *et al.*, 2010) as a common parent, respectively. The use of multiple parents in different combinations is another alternative, and two of such multiple parent combinations are now described for *Arabidopsis*. One is the so-called MAGIC population developed by Kover *et al.*

Table 1. QTL analyses of flowering time in studies using different types of mapping populations

Type	Parents	Size	QTL	R^2	References
NIL	L*er* × Cvi	92	5	83	Keurentjes *et al.* (2007)
RIL	L*er* × Div.[a]	120–164	3–8	68–85	El-Lithy *et al.* (2006); Keurentjes *et al.* (2007)
RIL	Col × Div.[a]	223–456	2–7	31–70	Brachi *et al.* (2010)
MAGIC	19 Inter-crossed	527	4	64	Kover *et al.* (2009)
AMPRIL	8 Inter-crossed	550	4	88	Huang *et al.* (2011)
GWA	Wild accessions	96–256	>20[b]	<50[b]	Atwell *et al.* (2010); Brachi *et al.* (2010)

R^2, the total explained variance.
[a] Data of diverse populations using a common reference accession (L*er* or Col). [b] Number of detected QTLs and total explained variance in GWA analyses depends strongly on the applied methods and threshold levels.

(2009). This population is derived by inter-crossing 19 parents for several generations, after which RILs were developed. The other, the so-called AMPRIL population (Paulo *et al*., 2008; Huang *et al*., 2011), was derived from inter-crossing four pairwise hybrids from eight founder lines, followed by RIL construction via single seed descent up to the F_5.

QTL analysis of flowering time

Table 1 summarizes the results of different population types. Many QTLs were identified at similar positions and most likely represent the same loci. Only when field conditions were employed predominantly different, QTLs were detected (Brachi *et al*., 2010), indicating the importance of genotype × environment interactions. When considering multiple parent populations, the number of QTLs was surprisingly low in the MAGIC and AMPRIL populations (Kover *et al*., 2009; Huang *et al*., 2010). Partially this can be explained by the reduced power due to the multiple alleles that segregate at multiple loci in multi-parent populations. A recent GWA study included flowering time analyses from various experiments (Atwell *et al*., 2010). The study showed the complication of population structure that is relatively strong for flowering time, as well as the known issues of significance thresholds and false positives. By combining the RIL and GWA approaches in the same experiment, these alleged false positives could be reduced and clear candidates could be indicated (Brachi *et al*., 2010). The recent developments in GWA mapping together with the relatively ease to generate experimental mapping populations, either as F_2/F_3 and/or the efficient development of doubled haploids using centromere-mediated genome elimination (Ravi and Chan, 2010), will continue to contribute to our understanding of complex traits. We therefore argue that multiple types of populations and approaches are needed for a thorough understanding of the complex genetic architecture of quantitative traits.

References

Alonso-Blanco C and Koornneef M (2000) Naturally occurring variation in *Arabidopsis*: an underexploited resource for plant genetics. *Trends in Plant Science* 5: 22–29.

Alonso-Blanco C, Aarts MG, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D and Koornneef M (2009) What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* 21: 1877–1896.

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J and Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.

Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, Maloof JN, Loudet O, Trainer GT, Dabi T, Borevitz JO, Chory J and Weigel D (2009) QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. *Public Library of Science ONE* 4: e4318.

Bentsink L, Hanson J, Hanhart CJ, Blankestijn-de Vries H, Coltrane C, Keizer P, El-Lithy M, Alonso-Blanco C, de Andres MT, Reymond M, van Eeuwijk F, Smeekens S and Koornneef M (2010) Natural variation for seed dormancy in Arabidopsis is regulated by additive genetic and molecular pathways. *Proceedings of the National Academy of Sciences USA* 107: 4264–4269.

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, Bergelson J, Cuguen J and Roux F (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *Public Library of Science Genetics* 6: e1000940.

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S and McMullen MD (2009) The genetic architecture of maize flowering time. *Science* 325: 714–718.

El-Assal S, Alonso-Blanco C, Peeters AJ, Raz V and Koornneef M (2001) A QTL for flowering time in *Arabidopsis* reveals a novel allele of CRY2. *Nature Genetics* 29: 435–440.

El-Lithy ME, Bentsink L, Hanhart CJ, Ruys GJ, Rovito D, Broekhof JL, van der Poel HJ, van Eijk MJ, Vreugdenhil D and Koornneef M (2006) New Arabidopsis recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. *Genetics* 172:1867–1876.

Huang X, Paulo MJ, Boer M, Effgen S, Keizer P, Koornneef M and van Eeuwijk F (2011) Analysis of natural allelic variation in Arabidopsis, using a new multi-parent recombinant inbred line population. *Proceedings of the National Academy of Sciences USA* in press.

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, 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.

Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D and Nordborg M (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics* 39: 1151–1155.

Koornneef M, Alonso-Blanco C and Vreugdenhil D (2004) Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annual Review of Plant Physiology and Plant Molecular Biology* 55: 141–172.

Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ and Kearsey MJ (2002) STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. *Plant Journal* 31: 355–364.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C and Mott R (2009) A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *Public Library of Science Genetics* 5: e1000551.

Nordborg M and Weigel D (2008) Next-generation genetics in plants. *Nature* 456: 720–723.

Paulo M, Boer M, Huang X, Koornneef M and van Eeuwijk F (2008) A mixed model QTL analysis for a complex cross population consisting of a half diallel of two-way hybrids in *Arabidopsis thaliana*: analysis of simulated data. *Euphytica* 161: 107–114.

Ravi M and Chan SW (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature* 464: 615–618.

Torjek O, Meyer RC, Zehnsdorf M, Teltow M, Strompen G, Witucka-Wall H, Blacha A and Altmann T (2008) Construction and analysis of 2 reciprocal *Arabidopsis* introgression line populations. *Journal of Heredity* 99: 396–406.

Tuinstra MR, Ejeta G and Goldsbrough PB (1997) Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theoretical and Applied Genetics* 95: 1005–1011.

Weigel D and Mott R (2009) The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biology* 10: 107.