Theor Appl Genet (2013) 126:1337–1356
DOI 10.1007/s00122-013-2056-2

ORIGINAL PAPER

Combined linkage and association mapping of flowering time in Sunflower (*Helianthus annuus* L.)

Elena Cadic · Marie Coque · Felicity Vear · Bruno Grezes-Besset · Jerôme Pauquet · Joël Piquemal · Yannick Lippi · Philippe Blanchard · Michel Romestant · Nicolas Pouilly · David Rengel · Jerôme Gouzy · Nicolas Langlade · Brigitte Mangin · Patrick Vincourt

Received: 21 September 2012 / Accepted: 20 January 2013 / Published online: 23 February 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Association mapping and linkage mapping were used to identify quantitative trait loci (QTL) and/or causative mutations involved in the control of flowering time in cultivated sunflower *Helianthus annuus*. A panel of 384 inbred lines was phenotyped through testcrosses with two tester inbred lines across 15 location × year combinations. A recombinant inbred line (RIL) population comprising 273 lines was phenotyped both per se and through testcrosses with one or two testers in 16 location × year combinations. In the association mapping approach, kinship estimation using 5,923 single nucleotide polymorphisms was found to be the best covariate to correct for effects of panel structure. Linkage disequilibrium decay ranged from 0.08 to 0.26 cM for a threshold of 0.20, after correcting for structure effects, depending on the linkage group (LG) and the ancestry of inbred lines. A possible hitchhiking effect is hypothesized for LG10 and LG08. A total of 11 regions across 10 LGs were found to be associated with flowering time, and QTLs were mapped on 11 LGs in the RIL population. Whereas eight regions were demonstrated to be common between the two approaches, the linkage disequilibrium approach did not detect a documented QTL that was confirmed using the linkage mapping approach.

Communicated by J. Wang.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2056-2) contains supplementary material, which is available to authorized users.

E. Cadic (✉) · Y. Lippi · N. Pouilly · D. Rengel · J. Gouzy · N. Langlade · P. Vincourt (✉)
Laboratoire des Interactions Plantes-Microorganismes (LIPM), INRA, UMR441, 31326 Castanet-Tolosan, France
e-mail: elena.cadic@toulouse.inra.fr

P. Vincourt
e-mail: patrick.vincourt@toulouse.inra.fr

E. Cadic · Y. Lippi · N. Pouilly · D. Rengel · J. Gouzy · N. Langlade · P. Vincourt
Laboratoire des Interactions Plantes-Microorganismes (LIPM), CNRS, UMR2594, 31326 Castanet-Tolosan, France

E. Cadic · M. Coque · B. Grezes-Besset · J. Pauquet
BIOGEMMA SAS, Domaine de Sandreau, Mondonville, 31700 Blagnac, France

M. Coque · J. Piquemal
SYNGENTA SEEDS, 12 Chemin Hobit, 31790 Saint Sauveur, France

F. Vear
INRA, UMR 1095, Domaine de Crouelle, 234, Ave du Brezet, 63000 Clermont-Ferrand, France

P. Blanchard
EURALIS Semences, Domaine de Sandreau, Mondonville, 31700 Blagnac, France

M. Romestant
RAGT 2N, Site de Bourran, 12000 Rodez, France

B. Mangin
INRA, Unité de Biométrie et Intelligence Artificielle UR875, 31326 Castanet-Tolosan, France

Introduction

Plant breeding requires an understanding of the genetic architecture of agronomic traits, such as yield, grain quality and resistance to biotic and abiotic stresses. Many of these traits are quantitative, governed by multiple interactions of loci with effects that depend on environment. Thus, connection of marker and molecular information to phenotypes constitutes a major challenge for breeders and molecular biologists. The most frequently used approach to study this problem, initiated in the 1980's (Lander and Botstein 1989), is quantitative trait locus (QTL) mapping, referred to in this study as linkage mapping. It requires experimental populations derived from a known pedigree, the simplest of which is a cross between two parental inbred lines.

Despite the proven usefulness of this technique to identify many genomic regions involved in complex traits (reviewed in Mackay 2001), the lack of an accurate method to localize and estimate the effects of a QTL represents a serious limitation for its application to marker-assisted selection (MAS) (Holland 2004). In particular, the accuracy of QTL detection depends on the amount of genetic variability present in the experimental population.

Often presented as an alternative approach, association mapping, based on a large number of genotypes representing a broader germplasm, has several advantages over linkage mapping: (a) better resolution due to the accumulation of historical recombination events, (b) a larger number of alleles surveyed and (c) the use of already existing material, such as elite germplasm, of practical value for breeding programs (Breseghello and Sorrells 2006). These advantages should facilitate MAS, with respect to the level of linkage disequilibrium (LD), in a population. However, association mapping has two major drawbacks compared with linkage mapping. First, it does not enable the detection of rare variants, which can reduce the power of this technique, depending on the genetic architecture underlying the trait of interest. In the case of the genetics of flowering time, two types of results have been reported. In Arabidopsis, genes with large effects on this trait were most commonly detected (Atwell et al. 2010). In contrast, in maize, numerous QTLs with common or uncommon genes, each explaining a small part of phenotypic variation, were detected (Buckler et al. 2009). In the majority of species, most alleles are rare (Myles et al. 2009) and can remain undetected with association mapping. A second common issue that limits the use of association mapping is the occurrence of spurious associations due to structured populations. This is particularly the case for adaptive traits such as flowering time, which is strongly correlated with population structure in *Arabidopsis thaliana* and maize (Aranzana et al. 2005; Flint-Garcia et al. 2005). Statistical methodologies have been developed to take population structure into account, but they could restrict the power of association mapping if the population is highly structured (Zhao et al. 2007). Given the complementary strengths of these detection techniques, research has been made to analyze results of association mapping and linkage mapping to combine the advantages of the two methods: resolution for the former and robustness for the latter (Nordborg and Weigel 2008).

Brachi et al. (2010) validated association peaks detected for flowering time in *Arabidopsis thaliana* when QTLs detected by linkage mapping co-localized with these peaks. Using this strategy, they were able to distinguish true from false positives and identify false negatives (causative loci lost when accounting for the population structure in the model.). Similarly, combined association and linkage mapping made possible fine mapping of QTLs in rice (Famoso et al. 2011) and in wheat (Mir et al. 2012). Association mapping methods first focused on candidate gene strategies, based on prior knowledge of the pathway controlling the trait of interest in model plants. Indeed, these strategies were guaranteed to have sufficient power, especially when the LD was high (Yan et al. 2011). Recently, the development of second generation sequencing and high throughput genotyping technologies has enabled considerable progress to be made in genetic mapping of agronomic traits. With the increasing availability of markers, the candidate gene approach has evolved towards whole genome scans, i.e., GWAS (genome wide association studies), enabling many SNP to be queried at the same time.

Sunflower (*Helianthus annuus* L.) is a species for which massive genomic resources have been recently developed (Kane et al. 2011). It is the fourth most widely grown oilseed crop in the world. It is of major economic importance as it produces healthy oil and has low input requirements (nitrogen, water and fungicides). Sunflower production increased by 32 % over the past 20 years, reaching 32 million tons in 2010 with acreage of 23 million hectares (FAO). It has also been used as a model species to study speciation and interspecific hybridization (Rieseberg and Willis 2007). Moreover, genomic sequence of sunflower should be available soon (Kane et al. 2011). High density genetic maps are now becoming available in sunflower (Bowers et al. 2012). These maps should help genetic dissection of many important agricultural traits. Numerous linkage mapping studies for most traits, including flowering time, have been published for sunflower (detailed in this study), but only one association study is available to date, focusing on *Sclerotinia* head-rot resistance (Fusari et al. 2012).

Flowering time is a major event in the plant life cycle. This trait is controlled by both genetics and environmental

stimuli. It is related to genetic adaptation to a range of abiotic stresses (e.g., drought, light and temperature) and affects susceptibility to diseases such as *Sclerotinia* head rot. It is therefore of great interest to evolutionary biologists, eco-physiologists and breeders who need to assess flowering time in studies of crop domestication (Burke et al. 2002; Wills and Burke 2007; Baack et al. 2008; Blackman et al. 2011), response to photoperiod (Leon et al. 2001), and relations with other agronomic traits (Mestries et al. 1998; Leon et al. 2000; Mokrani et al. 2002; Bert et al. 2003; Poormohammad Kiani et al. 2009). However, the genetic architecture of this trait remains poorly understood.

In this study, we combined association and linkage mapping approaches in sunflower, based on the evaluation of genotypes per se and/or in testcrosses over several location × year combinations. First, we evaluated the structure and LD within a large panel comprising elite lines. Then, we compared statistical models to reduce the confounding effects of associations caused by panel structures. Finally, we combined and compared linkage and association results to identify the genetic basis of flowering time.

Materials and methods

Plant material

The linkage mapping study was conducted on a population of 273 RILs obtained through single seed descent (to at least F8) from a cross between two INRA lines: XRQ and PSC8 (Vear et al. 2008). XRQ is a maintainer line originating from a cross between the founder USDA line HA89 and the Russian open pollinated variety "Progress", which confers tolerance to phomopsis and resistance to downy mildew. PSC8 was obtained from a restorer gene pool improved for *Sclerotinia* head-rot resistance by recurrent selection. Both parental lines were included in the association panel.

Association mapping was carried out on a core collection of 384 inbred lines from INRA and sunflower breeding companies, chosen for its diversity from an initial set of 752 inbred lines (Coque et al. 2008), see also https://www.heliagene.org/Web/public/core/Core_collections_list.html). It was comprised both elite lines, parents of commercial hybrids, and lines with introgressions from several wild *Helianthus* accessions, including *H. annuus, H. argophyllus and H. petiolaris*. In this core collection, 176 lines are publicly available whereas the others are proprietary lines provided by three breeding companies: Soltis (73 lines), R2N (46 lines) and Syngenta Seeds (89 lines).

Testcross progeny were obtained by crossing association panel lines and RILs with one or two of seven testers according to their status (maintainers of cytoplasmic male sterility [B-lines] or fertility restorers [R-lines].These testers were chosen with the purpose of obtaining single headed, and if possible male-fertile hybrids resembling modern cultivars. They also were selected for their ability to confer standard agronomic value and resistance to major diseases to their hybrids, thus avoiding artifact effects while at the same time allowing the expression of phenotypic variability.

Testers for the RIL were CMS-PGF650 for restorer genotypes and 83HR4gms for maintainer or unbranched restorers genotypes. 83HR4gms was bred by introgressing genetic male sterility into the INRA restorer line 83HR4 (Table 1).

For the association panel, R-lines were crossed with the two CMS PET1 counterparts of B-line testers (T1 or T3) while the B-lines were crossed with two R-line testers: 83HR4gms and T2, a proprietary line carrying PEF1 cytoplasmic male sterility (Crouzillat et al. 1991) which it maintains, although it is a restorer for classical PET1 cytoplasm (Table 2). The groups of B-lines and R-lines were named B-pool and R-pool, respectively.

Field experiments

For the RILs, testcross hybrid progeny were evaluated in four locations in 2001, three in 2002 and six in 2010. For the association panel, there were 15 location × year combinations, (designated as "environments") from 2008 to 2010. While 70 RILs were evaluated on two testers per environment, the other RIL and association panel hybrids all had the same tester in any one environment. Each environment–tester combination or environment was considered to be a trait for the RIL (Table 1) or association panel (Table 2), respectively, and was analyzed separately. Each experiment was formed of blocks, with 24 or 30 entries replicated in two sub-blocks. Each sub-block was randomized separately and contained two to four check hybrids.

RILs were also evaluated per se in three additional environments, two in 2001 and one in 2004. Environments CF01 and CF04 consisted of single rows of 13–15 plants per genotype without replication, whereas there were two replications of 25 plants for CF01_I.

In all trials, flowering time was measured as the number of days after sowing when 50 % of the plants had started anthesis.

Phenotypic data analysis

Observations made in 2001, 2002 and 2004 on tester × RIL combinations were first subjected to 2-way ANOVA to check statistical validity (data not shown).

Table 1 Details on environments, testers and effective used for the 23 traits evaluated on RILs

Trait	Environment combination		Genetic profile of RILs[b]	Tester[c]	Number of RILs under evaluation
	Location[a]	Years			
CF01_83	CF	2001	NR, Mild.R	83HR4gms	115
CF01_PG	CF	2001	Rest.	CmsPGF650	163
CF02_83	CF	2002	NR, Mild.R	83HR4gms	115
CF02_PG	CF	2002	Rest.	CmsPGF650	155
SC01_83	SC	2001	NR, Mild.R	83HR4gms	115
SC01_PG	SC	2001	Rest.	CmsPGF650	154
SC02_83	SC	2002	NR, Mild.R	83HR4gms	115
SC02_PG	SC	2002	Rest.	CmsPGF650	154
SL01_83	SL	2001	NR, Mild.R	83HR4gms	115
SL01_PG	SL	2001	Rest.	CmsPGF650	162
RN01_83	RN	2001	NR, Mild.R	83HR4gms	115
RN01_PG	RN	2001	Rest.	CmsPGF650	163
RN02_83	RN	2002	NR, Mild.R	83HR4gms	115
RN02_PG	RN	2002	Rest.	CmsPGF650	155
AI10_I	AI_I	2010	Rest.	CmsPGF650	134
AI10_NI	AI_NI	2010	Rest.	CmsPGF650	134
GA10_I	GA_I	2010	NR	PSC8RMgms	110
GA10_NI	GA_NI	2010	NR	PSC8RMgms	110
AU10_I	AU_I	2010	Rest.	CmsXRQ	110
AU10_NI	AU_NI	2010	Rest.	CmsXRQ	110
CF01 per se	CF	2001	–	–	243
CF01_I per se	CF	2001	–	–	243
CF04 per se	CF	2004	–	–	241

The RILs were evaluated per se in 2001 and 2004, and in combination with testers in 2001, 2002 and 2010

[a] The locations covered the range of environments where sunflowers are cultivated in France, could be irrigated (I) or not (NI), and were designated as follows: *CF* Clermont-Ferrand (center), *SC* Longre (middle west), *SL* Baziege (south west), *RN* Villampuy (north), *GA* Gaillac (south west), *AI* Aigrefeuille (middle west), Auzeville (south west). Each trait refers to a {location × year} × tester (or per se) combination

[b] When evaluated in combination, a subset of RILs was chosen according to the RILs genetic profile (*NR* non-branched, *Mild.R* conferring the resistance to the race710 of downy mildew, *Rest* restauration of the male fertility)

[c] 83HR4gms is a modification of the restorer line 83HR4 that had been converted to genetic male sterility. PSC8RMgms is a modification of PSC8 with resistance to race 710 of downy mildew and genetic male sterility. Cms XRQ and Cms PGF650 are classical female lines

Then, to make possible comparisons between all trials, data were expressed, for each genotype, as a percentage of the check mean.

In 2010, the data collected in 15 environments for the association panel and from six environment–tester combinations for RILs were analyzed with ASReml-R (Butler et al. 2007) using the following mixed model (naïve model):

$$Y_{ijk} = \mu + G_i + b_j + c_{k(j)} + e_{ijk}$$

where Y_{ijk} is the phenotypic observation for the ith genotype in the kth sub-block of the jth block, μ is the intercept term, G_i is the genetic effect of the ith genotype considered to be random, b_j is the effect of the jth block, $c_{k(j)}$ is the effect of sub-block k nested in block j and e_{ijk} is the residual. Block and sub-block effects were treated as fixed.

The naïve model was enhanced in two alternative ways: (a) by the addition of random effects of row and column for the "row × column" model, or (b) including a first-order autoregressive process in the residuals to take into account autocorrelation between neighbor plots for the "ar1 × ar1" model. To compare these three models, the Aikake criterion was calculated (AIC), and its significance assessed using log ratio tests between the naïve models and the two spatial models in succession. For each trait, once the best model had been selected, the best linear unbiased predictors (BLUP) of genotypes were extracted for the next step of analysis.

For each trait, broad sense heritability (h^2) was calculated with the following formula $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_e^2}{r}}$ where σ_g^2 being the genetic variance, σ_e^2 the residual variance, and r the

Table 2 Details on environment, testers and effective used for the 15 traits evaluated on the association panel

Trait	Environment combination		Tester for B-pool	Tester for R-pool	Number of lines under evaluation
	Location	Years			
AI08_I	AI_I	2008	83HR4gms	T1	171
AI08_NI	AI_NI	2008	83HR4gms	T1	172
CO09_I	CO_I	2009	83HR4gms	T1	262
CO09_NI	CO_NI	2009	83HR4gms	T1	262
GA09_I	GA_I	2009	83HR4gms	T1	261
GA09_NI	GA_NI	2009	83HR4gms	T1	260
LO10	LO	2010	83HR4gms	T1	270
AI09_I	AI_I	2009	T2	T3	263
AI09_NI	AI_NI	2009	T2	T3	263
VE09_I	VE_I	2009	T2	T3	257
VE09_NI	VE_NI	2009	T2	T3	257
CA10	C1	2010	T2	T3	290
CO08_I	CO_I	2008	T2	T3	230
CO08_NI	CO_NI	2008	T2	T3	229
SE10	SE	2010	T2	T3	272

The traits are designated using the same principles as in Table 1. Within a same location × year combination, the lines of the association panel were evaluated in testcross with the testers 83HR4gms or T2, for the lines belonging to the B-pool (See "Materials and methods"), and with the testers T1 or T3 for those belonging to the R-pool

[a] The locations were designated as follows: *AI* Aigrefeuille (Middle West), *CA* Castelnaudary (South West), *CO* Cornebarrieu (South West), *GA* Gaillac (South West), *VE* Verdun (South West), *LO* Loudun (Middle West), *SE* Segoufielle (South West)

average number of replicates per genotype, which was actually close to the number of expected replicates.

Genetic and residual variances were estimated using the naïve model to compare trial precisions, independent of any specific improvements with spatial models. Phenotypic Pearson correlations and principal component analysis (PCA) were performed between traits for each type of material using R (R Development Core Team 2012).

Association mapping

Genotyping

The association panel was genotyped for 12,136 single nucleotide polymorphism (SNP) markers using the Illumina BeadXpress and Infinium platforms. Polymorphism was initially identified using three different strategies: (a) transcriptome sequencing with mRNAseq technology from samples collected from whole plants, (b) genomic sequencing of targeted genes involved in hormone signaling pathways, development, stress response or transcription, (c) non-targeted genome sequencing of gene spaces. For genotyping, DNA was extracted from young leaf tissue with a Qiagen DNeasy 96 Plant Kit using a modified protocol (Horne et al. 2004). DNA concentrations were quantified with a Quant-iT PicoGreen dsDNA assay (Invitrogen, Karlsruhe, Germany). A total of 250-ng genomic DNA per sample was used to genotype SNP on the Illumina iScan platform (IntegraGen, Evry, France) or the Illumina BeadXpress platform (in collaboration with seed companies). Cluster positions for each marker were manually adjusted using Illumina GenomeStudio software. Heterozygous data expected to have low frequencies were considered as missing data.

A total of 8,844 high-quality SNP showing polymorphism across the association panel were retained for subsequent analysis. Eighty lines of the association panel with suspect genotypic data were discarded, giving a total dataset of 304 inbred lines for analysis.

Analysis of panel structure

A set of 5,923 SNP markers with less than 10 % missing data and a minor allele frequency (MAF) greater than 3 % were selected. Genotypic errors among these markers were assumed to be negligible above this threshold. Panel structure was investigated by two methods. First, a model-based approach implemented with STRUCTURE v2.2 software (Pritchard et al. 2000) was used to assign individuals to subpopulations according to correlated allele frequencies and admixture parameters. The algorithm was run for a number of subpopulations varying between one and ten. Ten replications for each subpopulation number were performed, with a burn-in time of 50,000 and 100,000

iterations. Evanno's criterion (2005) was applied to select the most likely number of subgroups. F_{ST} values similar to classical Wright's F_{ST} (1951) were used to estimate divergence between groups. Membership probabilities for each genotype in each of these subgroups were used to construct the Q matrix. Second, to perform a PCA on this dataset, each marker was centered by subtracting the average marker value across all samples and normalized by dividing by the theoretical standard deviation of the marker data at Hardy–Weinberg equilibrium (Patterson et al. 2006).

The significance of principal components was evaluated with a test based on the Tracy–Widom distribution. A PC matrix was established based on genotype coordinates of the components selected. The relatedness between all pairs of individuals were estimated using the Alikeness In State estimator (AIS) in Cocoa software (Maenhout et al. 2009), which comprised the K_{ais} matrix. To assess the relative performance to correct panel structure, Q, PC and K_{ais} matrix values were successively specified as covariates in association models.

In addition, a structure effect denoted "Tester", due to the status (B/R) of the lines, was included in model comparisons. As tester lines used to obtain hybrids were different between B and R lines, the Tester effect did not dissociate the effects of B or R status from the particular genotypic effect(s) of the tester(s). In each model, the structure was considered to exert fixed effects, and the genetic background captured by K_{ais} matrix was considered to be random, following the recommendations of Yu et al. (2006). A summary of all the models compared is presented in Table 3. Bayesian information criterion (BIC) and the p value of the significance of the fixed effects were used to determine the best-fitting model for each trait. The ability of each model to control for type I errors was compared by examining p values of association tests on a set of 1,000 random markers using ASReml-R. Finally, Student's test in R was used to check whether the mean of each phenotypic trait differed significantly between subpopulations.

Linkage disequilibrium

The resolution of association mapping studies depends on the pattern of linkage disequilibrium (LD). For LD estimation, we used a set of 1,874 SNP mapped on the proprietary BIOGEMMA consensus map and genotyped on the association panel, with a MAF of 5 % and missing data of <10 %. LD was calculated for each chromosome between all pairs of markers using classical statistics (r^2, squared correlations, between two loci, here in their haploid state) and a new measure correcting for biases caused by structure and relatedness between individuals obtained: r^2_{vs} (Mangin et al. 2011). Classical and new LD statistics were plotted as a function of genetic distance to estimate LD decay per chromosome, using Hill and Weir's model (1988), with a threshold of 0.2.

Association tests

Association mapping was based on a set of 6,645 SNP, including markers used for structure estimation and markers localized in candidate genes. This sample was taken from the validated set of 8,844 SNP after removing data containing more than 10 % missing observations and MAF lower than 5 %. Association between single loci and traits was carried out in Emma (Kang et al. 2008) using the two mixed models that correct for genetic relatedness between lines: "K_{ais}" and "K_{ais} + Tester". The full statistical model is:

Table 3 Summary of models tested for association detection

Model	Description	Covariate specification
Naïve	No correction for population structure	–
Fixed effects		
Q	Population structure inferred by STRUCTURE	Proportion of genome assigned at each group (g1, g2, g3)
PC	Population structure resulting from PCA	Coordinates on principal components
Tester	Structure of breeding pools (B-pool, R-pool)	Binary (0/1)
Random effects		
K_{ais}	Relatedness as estimated by alikeness in state coefficient	Matrix of variance–covariance proportional to the pairwise AIS matrix
Mixed		
$K_{\text{ais}} + Q$	Mixed model with population structure as fixed effects and relatedness as random effect	
K_{ais} + PC		
K_{ais} + Tester		

$$G_i^{BLUP} = \sum_c X_{ic} a_c + M_{il}\theta_l + u_i + e_i$$

G_i^{BLUP} is BLUP for ith hybrids, X_{ic} is tester category (0–1), a_c is effect of tester category c, M_{il} is genotype of the ith hybrid at locus l, θ_l is effect of locus l. a_c and θ_l are considered to be fixed effects. u_i is the random polygenic effect modeling genetic relatedness with $\mathrm{Var}(u) = \sigma_u^2 K_{\mathrm{ais}}$ where K_{ais} is an AIS matrix and $\mathrm{Var}(e) = \sigma_e^2$.

A false discovery rate (FDR) (Benjamini and Hochberg 1995) was applied on p values to correct for multiple testing. Variance explained by each marker was estimated using the R_{LR}^2 statistics described in Sun et al. (2010).

Linkage mapping of QTLs

A consensus genetic map was built using the method explained in Vincourt et al. (2012) and QTLs detected by linkage mapping in the RIL population were designated LM-QTL in this study. This consensus map and the corresponding map for the RIL population were produced from 619 and 517 public markers, respectively, of which respectively 345 and 285 were SNP (Supplementary File 1 also available at https://www.heliagene.org/Web/public/Consensus_INEDI_FUxPAZ2_V1/mapping_INEDI_FUxPAZ2_2012-07.html). QTLs were detected for each trait (environment-tester combination) with MCQTL v.5.2.4 (Jourjon et al. 2005) using iterative composite interval mapping (Charcosset et al. 2000) with the model:

$$G_i^{BLUP} = \mu + \sum_{l=1}^{n}\sum_{g=1}^{2} P\left(G_i^l = g|M_i\right)\beta_g^l + e_i$$

where G_i^{BLUP} is the BLUP for the ith RIL, μ is the global mean, $P\left(G_i^l = gM_i\right)$ is the probability of individual i having genotype g at QTL or cofactor locus l given the whole marker information denoted M_i. β_g^l is the mean of genotype g at locus l and e_i is the residual.

The statistical significance of QTLs was assessed using the MCQTL test, which is equal to $-\log$ (p value (F test)), as described in the MCQTL version 5 reference manual (http://carlit.toulouse.inra.fr/MCQTL/). A genome-wide type I error rate of 0.05 was applied, estimated after performing 1,000 permutation tests for each trait. QTL confidence regions were determined using a two-LOD support interval.

Overlap between linkage and association signals

An 8,235 marker proprietary consensus map developed by BIOGEMMA (unpublished data) served as a reference to project the 517 public markers and the detected QTLs using BioMercator v4 (Arcade et al. 2004). When possible, associated markers were located on this map. For those unmapped, linkage disequilibrium (r_{vs}^2) with all positioned SNP was calculated to place the markers. Unmapped SNP were assigned to the same position as the mapped SNP that was in maximum LD if the LD statistic was above a threshold of 0.1. The overlap between detected QTLs and association signals was tested statistically according to Tian et al. 2011. The total genome distance covered by a QTL over the genome size (in centiMorgans, cM) was computed to determine the probability of an SNP falling into a QTL support interval. When several QTLs overlapped, the largest interval was chosen. The hypothesis that associated SNP was placed with a larger probability in a QTL region than was expected by chance was tested using a binomial distribution.

In addition, LD statistics r_{vs}^2 between markers positioned in regions where QTLs overlapped with associated markers were used to build LD heat maps from a modified code of snp.plotter function (Luna and Nicodemus 2007).

Results

Phenotypic data analysis

The period from sowing to flowering time was measured on the association panel and RILs in a variety of environments. As indicated above, flowering time in each environment (for the association panel) or each environment–tester combination (for the RIL population) was considered to be a separate trait. In a first step, statistical analyses were made on the 15 traits obtained on the association panel and the six traits obtained in RILs in 2010.

Statistics for the naïve model, i.e., including only fixed blocks and sub-block effects with a random genotype factor, were computed (Supplementary File 2). Genotypic variance differed significantly from zero for all traits. Broad sense heritability ranged from 0.55 to 0.96, with a lower mean (0.68) for the RILs than for the association panel (0.84). Spatial models displayed a significantly better fit than the naïve model with respect to the AIC criterion. The model "ar1 × ar1" was more appropriate in 13 environments, while the "row × column" model performed better in six environments and the naïve in one. However, BLUP extracted from these three models were also highly correlated (data not shown). We selected BLUP from the best model to perform further analysis.

Correlations between the 15 association panel traits were all significant ($p < 0.001$; Supplementary File 3). A higher mean correlation between traits derived from the environments allocated to testcrosses made with the same pair of testers was observed: 0.66 and 0.72 for the group of 83HR4gms/T1-related traits and for the group of T2/T3-

related traits, respectively, to be compared with 0.49 between the two groups. This result was also illustrated by the PCA (Fig. 1). We still observed a large amount of variability in flowering time occurrence and distribution within each group. For example, T2/T3 group showed flowering time differences spanning 24 days.

Correlations between traits recorded on RILs (complete set of 23 traits, Supplementary File 4) were less significant. The mean correlation coefficient between traits recorded on hybrids involving the tester 83HR4gms was 0.51, compared to 0.56 for the tester CmsPG650. Similarly to what we observed for the association panel, on the second principal axis, PCA revealed a clear distinction between traits derived from different testers, with 83HR4gms- and CmsPG650-related traits being the most distant.

Despite of the overall good correlation between traits, we chose to conduct association and linkage mapping for each trait independently to capture specific interactions associated with a particular tester or environment.

Population structure

Using the Evanno's criterion (delta Evanno), the model-based STRUCTURE approach distinguished three probable groups, g1–g3, in the panel, as illustrated in Fig. 2a, b. The second criterion, i.e., the distribution of the log likelihood of the data, was not very meaningful because it did not reach a plateau. Two of the three groups exhibited by STRUCTURE were made up of wild introgression lines, belonging either to a set of 29 B-lines for the "g1 group" or to a set of 36 R-lines for the "g2 group". A total of 27 over 29 lines were assigned to g1, with a mean percentage of 0.98, and 22 lines over 36 were assigned to g2, at a percentage of 0.90. The third group ("g3" group) contained a majority of public B-lines. The g1 and g2 groups presented higher F_{ST} values (0.57 and 0.40, respectively) than g3 (0.07). The groups inferred by STRUCTURE are also highlighted in the PCA (Fig. 2c), where in addition to g1 and g2, g3 appears clearly in the PCA as a dense block of related individuals. For 151 inbred lines, there was no evidence of clear assignment to one of these three groups at a threshold of 0.80. The set of these 151 lines will be named g4 thereafter for convenience.

R/B line divergence was more obvious when using PCA analysis (Fig. 2c) than in the STRUCTURE analysis. Based on the Tracy–Widom statistics (Patterson et al. 2006), the first three principal components were considered to be significant and explained 13.21 % of the total variability. The first principal component, explaining 5.91 % of the variability, separated the B-pool on the right side with the g1 group on top and the R-pool on the left side with the g2 group on top.

Models comparison for association mapping

The majority of traits presented a mean significant difference between R- and B-lines, as well as between groups detected by STRUCTURE, highlighting the need for structure correction in association tests. Thus, a total of eight models were compared for their ability to correct stratification for each trait (Table 3). In the first step, marker information was not taken into account when exploring BIC criteria (Table 4) and p values of fixed effects (Supplementary File 5). Among the eight models compared, two reached the lowest BIC for most of the traits, including the "K_{ais}" model for seven environments

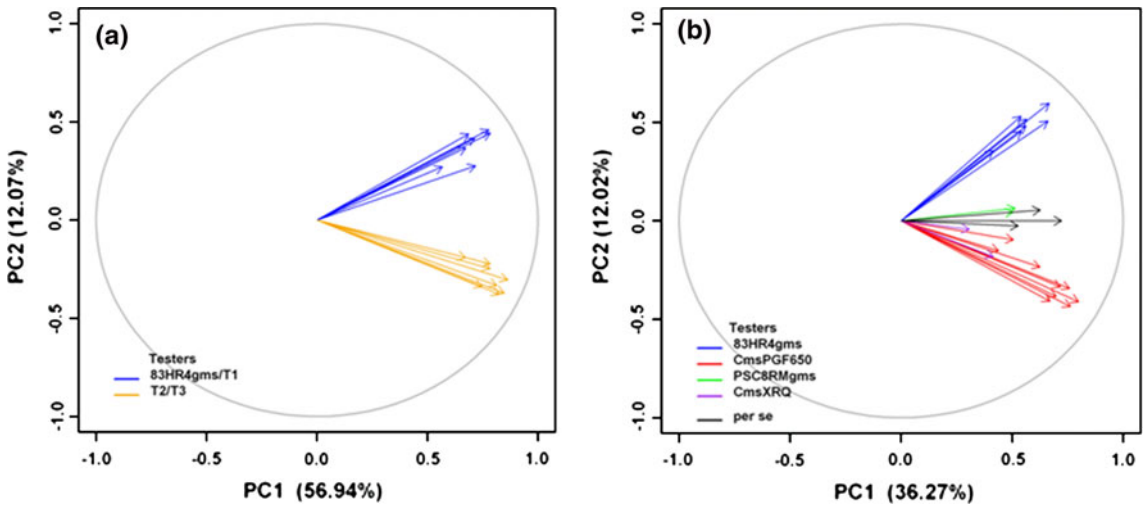

Fig. 1 Principal coordinate plots for the flowering time phenotypic traits recorded on the association panel (15 traits, **a**) and on the RIL population (23 traits, **b**). *Percentages in parentheses* refer to the proportion of variance explained by first and second principal coordinates (color figure online)

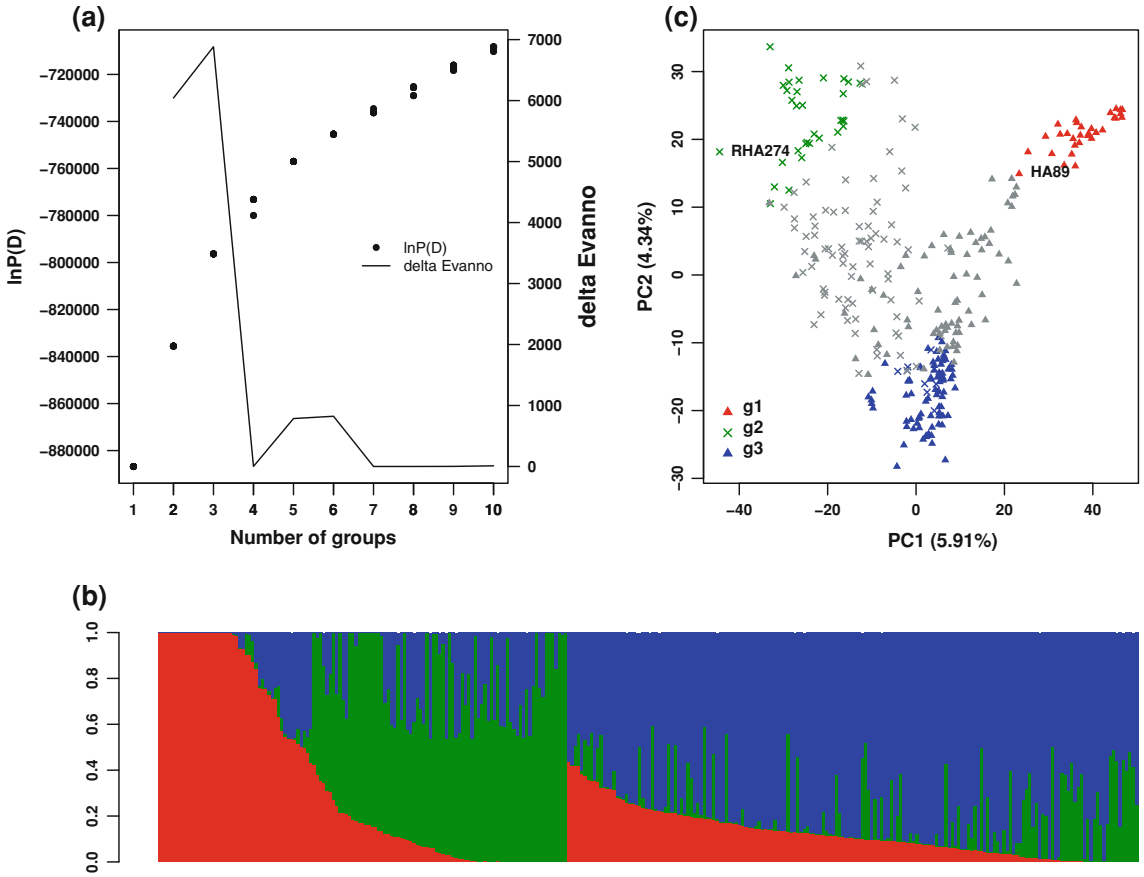

Fig. 2 Stratification of the 304 association panel lines. **a** Log-likelihood and delta Evanno statistics for a number of putative populations ranging from 1 to 10, **b** STRUCTURE output for three groups, **c** top two principal components in PCA. *Percentages in parentheses* refer to the proportion of variance explained by the principal coordinate. *Symbols* represent the two breeding pools (*x* for R lines and *filled triangle* for B lines). In **b** and **c**, each genotype is *colored* according to its STRUCTURE group (color figure online)

Table 4 Comparison of Bayesian information criteria (BIC) among eight models

Traits	Naïve	Fixed effects			Random effects	Mixed effects		
		Q	PC	Tester	K_{ais}	K_{ais} + *Q*	K_{ais} + PC	K_{ais} + Tester
AI08_I	528.35	526.73	529.14	515.93	513.61	523.18	528.17	*511.46*
AI08_NI	410.47	417.78	422.13	410.24	*409.08*	419.94	425.35	412.02
AI09_I	1,283.27	1,258.12	1,260.24	1,249.42	1,227.81	1,238.26	1,241.99	*1,221.57*
AI09_NI	1,221.61	1,194.95	1,198.11	1,184.06	1,145.96	1,156.85	1,161.30	*1,138.55*
CO08_I	1,038.25	1,031.42	1,036.03	1,032.15	*1,003.58*	1,014.32	1,019.02	1,008.59
CO08_NI	1,051.91	1,049.51	1,054.98	1,054.94	*1,018.95*	1,030.00	1,034.65	1,024.43
CO09_I	888.63	889.51	894.05	881.06	*877.57*	888.51	893.89	879.44
CO09_NI	941.38	926.07	929.03	918.22	916.22	926.24	930.62	*915.62*
GA09_I	988.50	947.65	946.65	937.98	941.09	947.65	949.88	*933.47*
GA09_NI	946.87	907.37	909.53	902.56	901.20	908.62	912.24	*896.73*
VE09_I	1,155.77	1,165.35	1,169.99	1,161.12	*1,120.86*	1,131.60	1,137.58	1,126.38
VE09_NI	1,180.22	1,189.04	1,192.37	1,185.16	*1,173.77*	1,185.10	1,190.19	1,179.23
CA10	1,409.67	1,411.75	1,416.63	1,399.70	*1,343.59*	1,353.56	1,359.90	1,344.40
LO10	1,081.28	1,038.51	1,042.78	*1,023.55*	1,042.25	1,045.34	1,049.39	1,029.12
SE10	*1,072.14*	1,080.56	1,085.58	1,074.08	1,075.05	1,086.33	1,091.69	1,079.79

Values in italics correspond to the lowers BIC for each trait

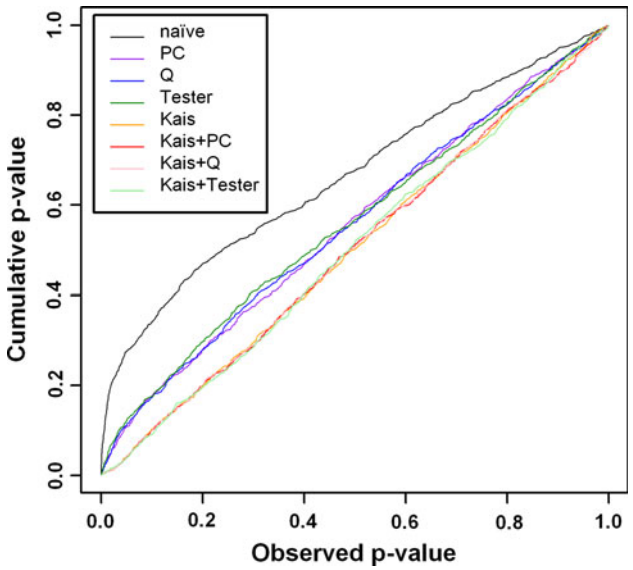

Fig. 3 Cumulative *p* value distribution of the association scan over a set of 1,000 random SNPs. The same eight models as in Table 4 were compared for the trait AI09_I (color figure online)

and the "K_{ais} + Tester" model for six environments. BIC values were found to be quite similar between these two models.

We compared *p* values of fixed effects in six out of the eight models ("K_{ais}" and naïve models were excluded). Significant structure effects were observed for most of the traits in the fixed models. When relatedness was taken into account in these models, *p* values of STRUCTURE and PCA covariates became non-significant (except for LO10). In contrast, the Tester effect remained significant for six traits, including those best fitted by the "K_{ais} + Tester" or "Tester" models, according to the BIC criterion. For CO09_NI, the Tester effect was not significant, although this trait showed best fit with the "K_{ais} + Tester" model. These results show that principal components and STRUCTURE covariates were not needed to correct for structure effects, as kinship probably retained a large amount of this information.

In the second step of analysis, we incorporated a set of 1,000 SNP into the investigated models to search for an excess of significant associations over those predicted by the null hypothesis. The quantile–quantile (Q–Q) plots looked very similar between traits; a representative example is presented in Fig. 3. For the naïve association model, the rate of false positives was clearly inflated, suggesting that correction was necessary. Models specifying structure effects performed better, and only models including kinship matched the diagonal, showing good control of false positives. The above analysis demonstrates that both the "K_{ais}" and "K_{ais} + Tester" models, considered to be the best models according to BIC and *p* value criteria, gave similar reductions in false positives. As a consequence, we decided to conduct association tests using these two models.

LD estimation

We first investigated the pairwise LD for each chromosome in the entire panel, with (Mangin et al. 2011) or without correcting for B/R line structure and kinship confounding effects. Mean LD decay, defined by the Hill and Weir model (1988) with a threshold of 0.20, was estimated across each LG. The value decreased from 0.41 cM without correction to 0.14 cM with kinship and structure correction. As illustrated in Fig. 4a for LG08, long distance pairwise LD (over 10 cM) was not maintained after correction. However, we observed considerable differences in LD decay, ranging from 0.08 to 0.26 after correction, between LGs. LG08 and LG10 presented a specific pattern, with high LD values when no correction was applied (Fig. 4b).

In a second step, we computed the corrected LD statistics accounting only for kinship effect: r^2_{vs}. This was done separately for the R-pool (121 accessions) and the B-pool (183 accessions), as maintainer and restorer lines are considered to belong to distinct breeding pools. For most chromosomes, the B-lines presented a mean decay of LD that was similar to that of the entire panel. In contrast, the R-lines showed higher LD values, with a mean decay of 0.31. LGs having the largest LD differences between the two breeding pools, with a higher value for R-lines pool, included chromosomes 8, 10, 13, 14 and 17 (Fig. 4c).

Mapping results

Linkage mapping

QTLs were detected for 20 out of 23 traits. The QTLs accounted for 6–29 % of the phenotypic variation, with an additive effect varying from 0.1 to 5.0 days (Detailed statistics in Supplementary File 6, linkage map with QTL available at https://www.heliagene.org/Web/public/linkage_mapping_FT_INEDI.html). Considering the 2001 and 2002 environments in which hybrids were tested, the crosses with CmsPGF650 led to the identification of more QTLs (23) than for those with 83HR4gms (13). For environments where the two testers were used, we identified QTLs on both of them, especially on linkage groups 6 (LG06) and LG14 (Fig. 5). In contrast, several regions were only detected with one tester. For example, the LG09 region was identified in material crossed to the tester CmsPGF650. This region, with an average confidence interval of 6.60 cM, was highly significantly associated with nine traits (average MCQTL test = 9.44 and explained variance of 24 %), with allelic effects reaching

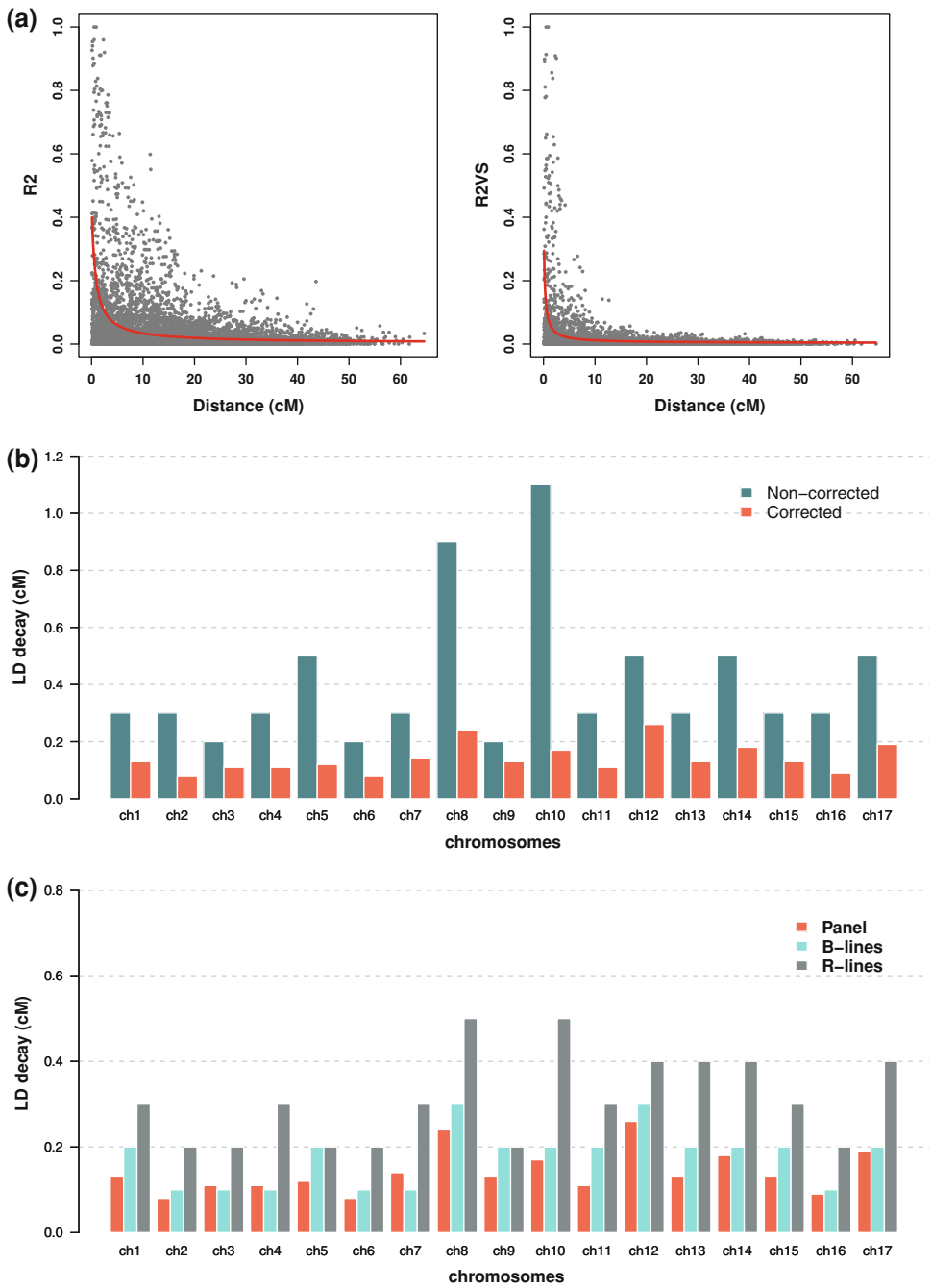

Fig. 4 **a** LD statistics plotted over genetic distance (cM) on LG08: classical LD (*left*), r^2_{vs} (LD corrected from relatedness and B/R line structure, *right*); **b** distribution of classical LD and r^2_{vs} statistics across chromosomes; **c** distribution of r^2_{vs} across chromosomes for the entire panel and for each breeding pool (B-lines and R-lines) (color figure online)

5 days. In addition to this QTL, three regions were particularly highlighted by overlapping of QTL support intervals for a large number of traits. LG06 contained a cluster of QTLs detected in 12 environments overlapping in a 21-cM support interval. A cluster of QTLs corresponding to 10 environments and overlapping within a 33 cM support interval was found on LG14. LG12 presented a smaller confidence region of 14 cM, where QTLs for seven environments were mapped. While the allele conferring late flowering were derived from XRQ for the QTL located on LG09, LG12 and LG14, it came from PSC8 for the QTL located on LG06. Fourteen QTLs were detected in environments where RILs per se were evaluated. Overall, they presented a higher significance level than the rest of QTLs detected (MCQTL test = 8.00 vs. 5.3) but with a lower additive effect (0.12 vs. 2.10 for the QTL detected on hybrids). Compared to results obtained for hybrids, new regions appeared to be involved in the flowering trait evaluated on RILs per se on LG05 and LG10, while regions on LG06, LG08, LG12 and LG14 were confirmed.

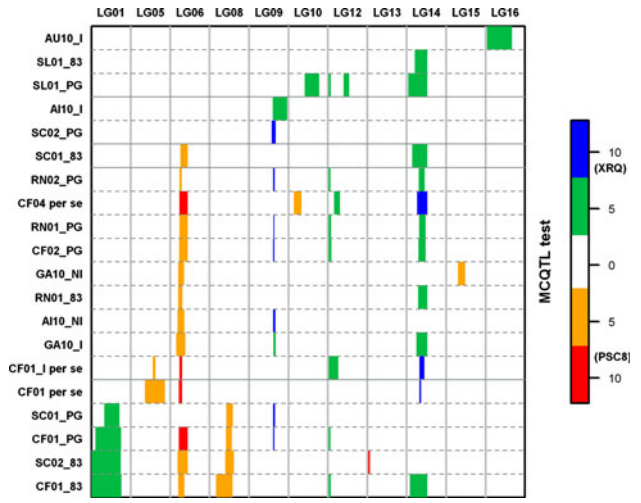

Fig. 5 Heat map of detected QTLs. Only chromosomes (*columns*) and traits (*row*) where QTLs were detected at a threshold of 3.80 are represented. A *color scale* is used to indicate QTL significance (MCQTL test). Positive values (*orange and red*) denote a later occurrence of flowering time for RILs carrying PSC8 alleles (color figure online)

Linkage disequilibrium mapping

Figure 6 presents the results of marker trait association analysis for "K_{ais}" and "K_{ais} + Tester" models and their localization on the proprietary map. Eleven of the 15 traits presented significant associations at the FDR threshold of 10 %. Most were detected using both models. However, several signals found associated with a trait using the "K_{ais}" model were not identified using the "K_{ais} + Tester" model.

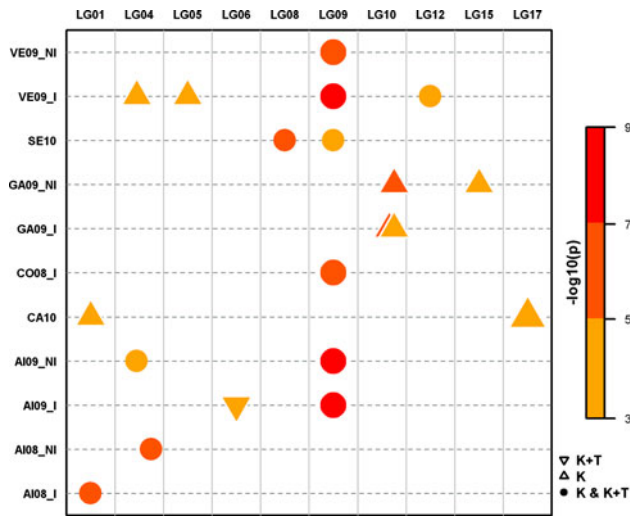

Fig. 6 Heat map of significantly associated markers mapped (or localized with LD) on the consensus map for each trait and both tested models ("K_{ais}" or "K_{ais} + Tester"). A *color scale* is used to indicate SNP significance based on p values (−log 10 (p value)) (color figure online)

To summarize, 11 regions on 10 LG were found to be associated with flowering time. One of these regions, on LG09, was noticeable due to the number of traits (6) for which an association was detected. In this region, very low p values were found for some of the markers. For example, SNP HS117598 presented a p value of 7.65×10^{-10} and additive effect of 3 days on flowering time in the AI09_I environment. Moreover, the markers spanned an interval greater than 2 cM, thus showing the linkage disequilibrium in this region. The allelic effect was found in the same direction for each trait. Apart from this region and the three markers mapped on LG01, 04 and 10, respectively (detected in two environments each), all association signals were specific to one environment. We also observed poor consistency in mapping results between highly correlated environments, such as those differing only by irrigation treatment. Mapping results were not consistent between traits obtained with 83HR4gms/T1 and those obtained with T2/T3, as only one marker on LG01 was detected in both groups of traits. Details of MAF, p values, allelic effects and variance explained for each SNP detected (range 4.5–13.3 %) are provided in supplementary data (Supplementary File 7).

Overlap between association signals and LM-QTL

SNP identified using association mapping were compared with the positions of QTLs detected in the RIL population. For this purpose, a consensus map was built by projecting the LM-QTLs detected on the public map onto the proprietary map. The significance of overlapping between association peaks and LM-QTLs was assessed as follows. We first identified independent SNP among the 27 SNP found associated with traits by calculating LD r^2_{vs} (corrected statistics) for each pair of markers. Among the 13 SNP considered to be independent r^2_{vs} (threshold of 0.1) and mapped by recombination or LD on the consensus map, nine were positioned in LM-QTL support intervals. Binomial tests proved that the overlap observed was significant ($p = 3.49 \times 10^{-6}$), compared to the probability of an SNP falling into an LM-QTL region by chance (0.17).

A total of eight chromosomes carried QTLs on which associated SNP were also detected (Table 5). Figure 7 describes the pattern of LD combined with associated p values for two regions of interest: LG09, which displays highly significant associations consistent across environments, and LG10, on which one marker is positioned in a candidate gene for flowering time (detailed in the discussion). On the latter chromosome, two markers were significantly associated with this trait. One of these markers was not localized in the QTL regions (Fig. 7a) but was in LD to the second marker.

Table 5 Significant markers detected with association mapping and mapped within the QTL regions found by linkage mapping

LG	Map position	Marker	QTL interval on the consensus map	Arabidopsis homolog locus	Arabidopsis homolog gene name (if any)	Arabidopsis homolog TAIR description
LG01	33.5	HS136120	−6.00–72.71	AT5G18120	APRL7	Encodes a protein disulfide isomerase-like (PDIL) protein, a member of a multigene family within the thioredoxin (TRX) superfamily. This protein also belongs to the adenosine 5′-phosphosulfate reductase-like (APRL) group.
LG05	44.8[a]	HS107108	34.22–66.11	AT3G17590	BSH	Encodes the Arabidopsis homolog of yeast SNF5 and represents a conserved subunit of plant SWI/SNF complexes.
LG06	32.2	HS113607	18.8–35.7	Not found		
LG08	29.0	HS097037	−20–30	AT1G75560		Zinc knuckle (CCHC-type) family protein
LG09	31.3	HS117040	27.08–35.97	AT1G67430		Ribosomal protein L22p/L17e family protein
LG09	31.6	HS090401		AT1G24030		Protein kinase superfamily protein
LG09	31.6	HS117598		AT1G67580		Protein kinase superfamily protein
LG09	31.8	HS095606		AT5G50260	CEP1	Cysteine proteinases superfamily protein
LG10	53.2	HS061549	41.12–66.8	AT3G63010	GID1B	Encodes a gibberellin (GA) receptor ortholog of the rice GA receptor gene (OsGID1). Has GA-binding activity, showing higher affinity to GA4. Interacts with DELLA proteins in vivo in the presence of GA4.
LG10	57.2	HS073886		AT5G47780	GAUT4	Encodes a protein with putative galacturonosyltransferase activity.
LG12	34.8	HS067214	45–66.8	AT3G14310	PME3	Encodes a pectin methylesterase, targeted by a cellulose binding protein (CBP) from the parasitic nematode Heterodera schachtii during parasitism.
LG15	42.2[a]	HS057257	36.48–51.04	Not found		

Only markers corresponding to different genes are indicated. Map positions refer to the consensus map built by projection of the public map onto the proprietary map. The putative *Arabidopsis thaliana* homologs were obtained on TAIR by blasting (BLASTX) the *Helianthus annuus* contig sequence carrying the SNP

[a] These markers were mapped on the consensus map based on their LD with mapped markers

Discussion

Panel structure and breeding history

As a prerequisite for association mapping, panel population structure was assessed using a PCA and a Bayesian model in STRUCTURE software. Results with the two analyses were quite similar and confirmed by the significant correlation between the *Q* and *P* matrices (data not shown).

The panel appeared to be made up of three groups (Fig. 8), the two most divergent groups g1 and g2 contained respectively B and R lines. The proximity among the g1 lines in one hand and among the g2 lines in other hand is probably due to founder effect which arose from the introgression of wild *Helianthus* accessions into an elite B line and an elite R line, respectively. This suggests that it should be possible to enhance genetic variability in both B and R gene pools by increasing the use of wild *Helianthus* accessions in breeding programs. It should be pointed out that the two well-known USDA lines HA89 and RHA274 were located in the g1 and g2 groups, respectively.

The third group (g3) was constituted with a large proportion of the public maintainer lines (B-lines).

The other genotypes, both public or proprietary elite, maintainer or restorer lines were not assigned to a group by STRUCTURE, but for clarity will be designated group g4 ($g4_B$ and $g4_R$ for B lines and R lines, respectively).

The group g3 of B-lines showed a lower level of divergence from g4 compared to g1 and g2. In addition, g3 presented less diversity than the public R lines belonging to $g4_R$ as shown in Fig. 2b. This is in agreement with observations of Mandel et al. (2011), who used 86 of the set of public lines included in our panel as part of a collection of 433 cultivated accessions. Taking into account genotyping data from two SSR markers per linkage group, these authors observed less diversity among B lines than among the restorer lines ("INRA RHA" in their work). A large fraction of the restorer lines involved in breeding programs were derived from crosses between one of the main sources of fertility restoration genes (wild *H.annuus* carrying *Rf1*) like RHA274, and B lines (or their CMS counterpart). The founder effect of the line RHA274 could

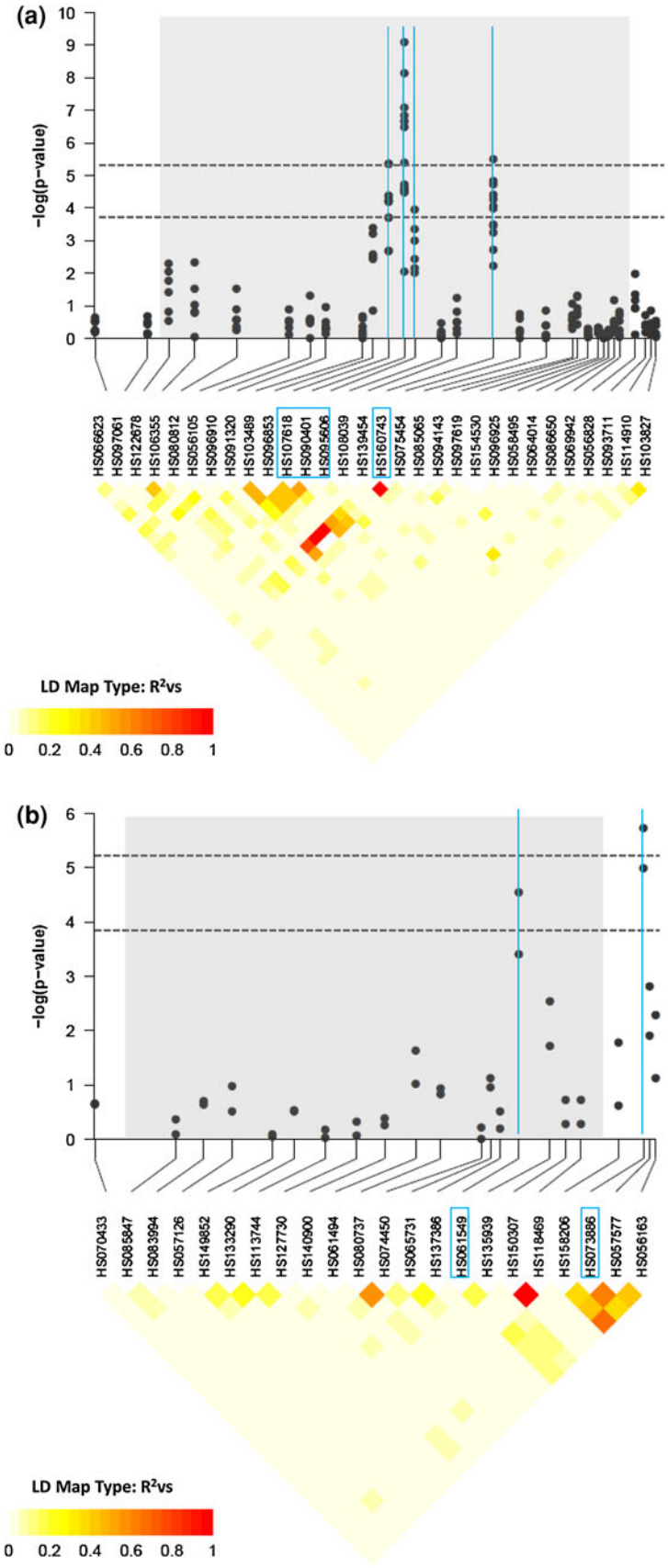

Fig. 7 LD heat map with corrected LD statistics (r^2_{vs}). Only QTL regions (*shaded area*) and markers tested in association mapping with their *p* values are represented. Bonferroni (*lower*) and FDR thresholds are indicated by *dashed lines*. **a** LG10, over 10.7 cM **b** LG09, over 18 cM (color figure online)

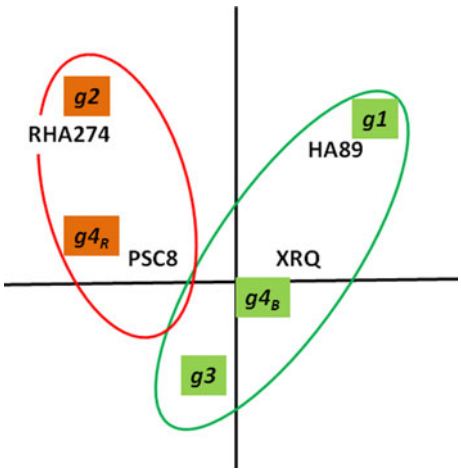

Fig. 8 Simplified representation of genetic groups according to Fig. 2b (color figure online)

explain why the R lines belonging to g4 have been found wide spread between RHA274, found close to g2, and more or less closely related to the g3 B line group.

In contrast, the large distance between HA89, which was derived from the Russian open pollinated population VNIIMK 8931, and the g3 group suggests that, in our panel, the set of B lines is divided into two subsets. One of these is more or less related to the HA89, and the other (g3) could be related to another particular founder effect which is not known. According to the pedigree information we have, this founder could have originated from Eastern European breeding programs, with either specific selections from Russian open pollinated varieties other than VNIIMK 8931, or introgressions from wild *Helianthus*, or exotic germplasm very different from g1 and g2.

Divergence between B and R elite gene pools has previously been mentioned (Gentzbittel et al. 1994).The fact that it did not appear as a major factor of our panel when using STRUCTURE is probably due to the presence of introgression lines. In addition, the three groups revealed by STRUCTURE or the PCA did not provide the best model for association detection. Indeed, the best models were obtained with the nearly statistically equivalent "K_{ais}" or "K_{ais} + Tester" models. Although these two models aimed at correcting different layers of stratification, the BIC values were quite similar. Furthermore, both of these models led to a good control of the rate of false positives and they detect the same most significant SNP. All but two of the detected associations were confirmed to be true positives, as they were located within regions detected through linkage mapping. It was not possible to check the validity of the two remnant associations, as the RIL parents were not polymorphic at these loci.

These results show that kinship probably accounted for most of the information provided by the structure of the panel. They also validate our strategy of keeping a sub-optimal model such as "K_{ais}" for association detection.

Linkage disequilibrium

Because the resolution of association mapping depends on the LD between genotyped markers and causative polymorphism, it was necessary to understand the nature of LD in our panel. LD decay is a good predictor of what is expected, given the present marker density. Standard measurements of LD have been described in the literature (Flint-Garcia et al. 2003). However, long range LD, which results from the presence of subpopulations with different allelic frequencies, can bias these estimates. We therefore used the modified statistics proposed by Mangin et al. (2011). This modified estimation, when compared to classical r^2 statistics, lead to a considerably increased rate of LD decay by neglecting r^2 between unlinked loci. When considering our design, this modification took kinship and the B/R differentiation into account. This is reminiscent of the "K_{ais} + Tester" model which was found overall the best for association detection. Moreover, we investigated the LD decay in each breeding pool using the LD statistics, thereby correcting for the relatedness between lines in each group.

When considering all of the lines together, large variations of LD were observed between linkage groups. The extent of LD is influenced by many factors, such as recombination and selection (Gaut and Long 2003) through hitchhiking with selected loci (Maynard Smith and Haigh 1974; Mackay and Powell 2007). The breeding history of the parental lines of sunflower hybrids might have resulted in a high mean LD on LG08 and LG10 compared to other LGs. Indeed, the recessive branching gene is located on chromosome 10. Thus, positive or negative selection for this trait, depending on the breeding purpose (B or R lines, respectively), might have caused the extent of LD near this locus. Similarly, a cluster of resistance genes to downy mildew (including *Pl2*, Bouzidi et al. 2002), mapped on LG08, has been an important target of selection, particularly for R lines during the 1975–1995 period.

Even using the corrected statistics, the overall LD in our panel (0.14 cM at 0.20 threshold) remained high. It extended over approximately 272 kb, with a 3.5-Gb genome and a genetic map of 1,800 cM. Comparisons with previous studies are difficult. LD extends from 50 to 250 kb in *Arabidopsis thaliana*, a self-pollinated species (Nordborg et al. 2005). In crops, elite lines can present a slower decay of LD (Rafalski 2002). For example, in maize, different studies have shown that LD persists over 1 kb for landraces (Tenaillon et al. 2001) to 500 kb for commercial elite inbred lines (Jung et al. 2004).

In sunflower, Kolkman et al. (2007) analyzed a set of ten elite inbred lines and two wild accessions and estimated LD decay for 30 loci. In this study, LD was found to extend over 5,500 bp for a r^2 threshold of 0.32, suggesting that the threshold we used in this study would have led to a larger extent of LD. Fusari et al. (2008) assessed the LD over 28 candidate genes (<1 kb) in a panel of 19 elite inbred lines. They estimated that LD decays over 643 bp for an r^2 of 0.64 in the entire set or 0.48 in one of the subpopulations identified by STRUCTURE. This result highlights the need to take into account the presence of subpopulations when estimating LD. These earlier studies assessed LD over short distances (<1 kb for the latter), whereas we investigated the LD in a broader sample that accounted for structure. A slow LD decay usually confers poor resolution for association mapping but requires fewer markers. In this study, given the LD estimated, 12,857 SNP would be required to cover the genome, but we only met half of this requirement.

On all LGs, the LD was found to decay more rapidly in the entire panel than within the B or R group. When considering all of the lines together, because selection pressures were different for B and R lines depending on trait, a different number of recombination events may have occurred for B lines and for R lines. Second, estimates of within-population rates of LD decay are subject to much larger standard errors than those based on whole populations, due to the smaller number of polymorphic sites (Ingvarsson 2005).

Mapping results

Most significant associations were specific to a set of testers (83HR4gms/T1 or T2/T3). The consistency of mapping results between the testcross results is of great importance, especially when they are applied to breeding programs (Melchinger et al. 1998). Among the hypotheses that explain the lack of common QTLs between testers, the most common is that dominant alleles of tester lines, especially when they are elite lines (Hallauer and Miranda 1988; Austin et al. 2000), can exert masking effects. In the testing design we used, only dominant alleles from the panel are expected to be detected when combined with recessive alleles from the testers.

In the linkage mapping study, rather low correlations between environments were observed, while a relatively high congruency was found between the detected QTL. Due to this congruency, shorter interval supports for congruent QTL were found when running the QTL detection on the year or tester averages (data not shown). In contrast, in this association study, even when raw correlations between environments where high r^2 (values ranging 0.42–0.93), only a few signals were repeated across environments. In an attempt to explain this intriguing result, we first examined the impact of the panel relatedness to the high raw correlations. Using an approach similar to that of Mangin et al. (2011), we calculated a modified r^2 using the kinship matrix K as a metrics. These modified r^2 values were considerably lower than raw r^2 values (0.00–0.58). This suggests that kinship—which also accounts for B-lines vs. R-lines—drove the common responses of the panel across the environments, whereas association peaks were the driving forces behind genotype × environment interactions. However, this does not explain why we found, in the linkage mapping study, lower correlations between environments together with more congruency of QTL. It should be pointed out that in the association study, only hybrids between B lines and R-type testers (83HR4gms or T2) or between R lines and B-type testers (T1 or T3) were evaluated ("unrelated" context). In contrast, each RIL involved in the linkage mapping study is a patchwork including B and R background. This patchwork was evaluated either in testcrosses with B-type and R-type testers, or per se, thus leading at least to local B × B or R × R combinations ("related" context). On another cross-pollinated species (*Medicago sativa*), it has been shown that part of the genetic variability expressed differs between "related" and "unrelated" contexts (Gallais 1984). Moreover, hybrids ("unrelated" context) generally show better stability across environments, in the sense of Allard and Bradshaw (1964), than inbred lines ("related context"). This could explain why testcross data did not result in the same level of correlation between environments in the association study and in the linkage mapping study. Finally, we hypothesize that the covariance between environments in the linkage mapping study, which is still significant as shown in Fig. 1b, accounts for congruency between QTL.

All together these results indicate that despite its well-documented weakness, the linkage mapping approach is still relevant, because it is robust when relevant genetic variability exists between the parental lines of the RIL population.

Recently, several studies integrating association and linkage mapping have demonstrated the power and resolution of this approach to identify loci of interest. These joint-linkage association mapping methods are based on controlled crosses that provide equilibrated allelic frequencies (Myles et al. 2009). Such a design was developed in maize with a nested association mapping (NAM) population consisting of 25 RIL populations derived from crosses between 25 diverse lines and a tester line B73, which was used to dissect flowering times (Buckler et al. 2009).

Taking advantage of our combined linkage mapping analysis, we were able to determine whether the association

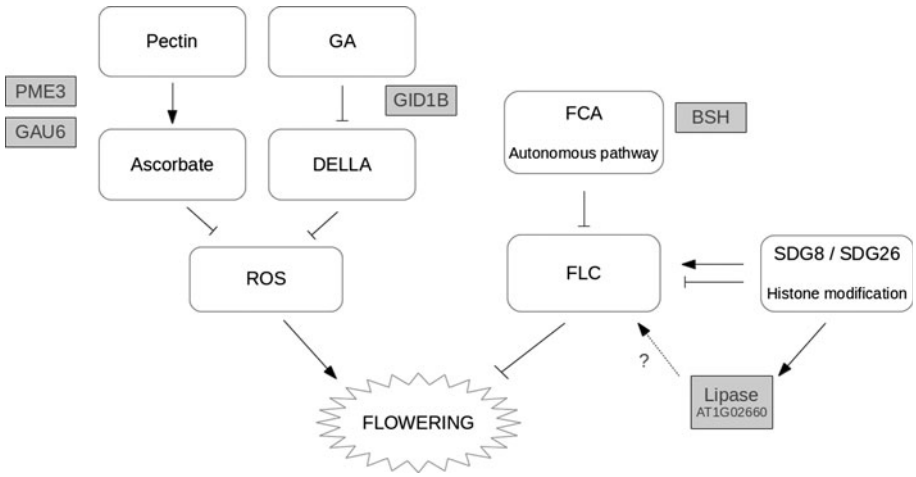

Fig. 9 Candidate pathways involved in flowering time variation in sunflower. Genes carrying associated polymorphisms are indicated in *gray boxes*

signals we detected were false positives. A true association at a given locus implies that a QTL overlapping this locus should exist if population parents carry different alleles for that locus (Zhao et al. 2007). In this study, all of the SNP detected through the association mapping approach, and for which alleles differed between RIL parents, were located in LM-QTL regions. In the case of LG10 and LG15 (detected only through the suboptimal model "K_{ais}"), we confirmed that the SNP detected were true positives whereas those detected on LG04, LG12 and LG17 could not be confirmed, as the RIL parents were not polymorphic at these loci.

On the contrary, one major QTL identified with linkage mapping on LG14 and also reported in the literature (Poormohammad Kiani et al. 2009) was not tagged through association mapping. This situation can occur when one of the alleles at the locus concerned is present at low frequency in the association panel. It can also occur when the structure of the panel is correlated with polymorphism at this locus, thus inducing a "false negative" (Famoso et al. 2011). The 106 SNP tested in association mapping were mapped within the 20-cM support interval of this LG14 QTL. We observed large blocks of LD in this region when no structure correction was applied, thus helping to confirm the hypothesis of a false negative in the association approach.

According to the genetic profile of the two parental lines XRQ and PSC8 (Fig. 8), the RIL population enabled the detection of differences resulting from their respective derivations: g1 × g3 for XRQ, and g2 × g3 for PSC8, which could make it possible to identify loci differentiating g1 and g2. In contrast, in the association mapping design, kinship accounted for breeding history and it may have precluded the detection of association peaks on LG14. This also suggests that the QTL located on LG14 is an important feature distinguishing B-type and R-type sunflower lines for flowering time.

The polygenic pattern of inheritance of the flowering trait in sunflower has been reported in the literature (Leon et al. 2000). Taking into consideration the overlap between association and linkage mapping results, eight regions involved in the inheritance of flowering time were identified in this study. Several of these are in good agreement with other linkage mapping results from studies concerning the same trait.

In these eight regions, five potential candidates identified through association peaks appear to be functionally related to flowering time in other species (Fig. 9). The peptide predicted from the sunflower sequence HaT13l016684 and corresponding to the associated marker HS107108 (localized on LG05) is homologous to the Arabidopsis BSH protein. Interestingly, BSH interacts with SWI3B to form a complex with the regulator of flowering time FCA (Sarnowski et al. 2002). The epigenetic control of FLC via H3K36 histone modification is controlled by SDG8 and SDG26 in Arabidopsis and could also be involved in sunflower. In fact, in their study, Xu et al. (2008) identified in both SDG8 and SDG26 mutant backgrounds a modification of expression of a lipase homologous to a gene associated with flowering time variation in our study: HaT13l002875. More strikingly, HaT13l048245, which carries the marker HS061549 on LG10, is homologous to Arabidopsis GID1B. This gibberellin receptor regulates DELLA proteins by targeting the proteins to the proteasome. It is well known that the gibberellin/DELLA pathway controls flowering time in Arabidopsis (Sun 2010). Several studies have confirmed that this pathway is prone to genetic variation, which explains flowering time variations in crops such as maize (Andersen et al. 2005; Thornsberry et al. 2001) and, most likely, canola (Raman et al. 2012). However, due to differences in genetic material, Blackman et al. (2011) were unable to confirm the presence of this pathway in sunflower. DELLA repressed ROS accumulation by increasing of the transcription of

genes encoding ROS scavenging enzymes (Achard et al. 2008). Interestingly, two other genes that regulate ROS accumulation via the ascorbate pathway in Oncidium (Shen et al. 2009) are carrying confirmed SNP: HaT13l001758 and HaT13l024508, which are homologous to GAUT6 and PME3, respectively. The functional analysis of genes associated with flowering time should still be considered with extreme caution. However, not surprisingly, different pathways have been identified in which ROS may play a role and further analysis is needed to resolve this issue.

Conclusion

This study has shown, for observations of flowering time in cultivated sunflower across several genetic and agronomic environments, a large similarity between association peaks detected in a wide panel and QTL mapped in a RIL population. In the linkage disequilibrium mapping approach, comparison of models demonstrated that the kinship provided the best fit. However, both due to the genetic design which aimed to analyze other agronomic traits and to the breeding history, the detection of associations using this model did not identify a QTL documented in the literature and confirmed in this study. The results have provided novel information on whole genome linkage disequilibrium in cultivated sunflower, including a possible hitchhiking effect on two linkage groups carrying loci under strong selection during the breeding process.

Acknowledgments We would like to thank M.C. Boniface and D. Varès (INRA Toulouse), H. Bony, G. Joubert, F. Serre, S. Roche and J. Philippon (INRA Clermont-Ferrand), Th. André (SOLTIS), S. Châtre (RAGT), P. George and M. Barthes (BIOGEMMA) and colleagues from SYNGENTA Seeds for their involvement in sunflower trial management. This work benefited from the GENOPLANTE program "HP1" (2001–2004), the "SUNYFUEL" project, financially supported by the French National Research Agency (2008–2011), and the "OLEOSOL" project (2009–2012) with the financial support from the Midi Pyrénées Region, the European Fund for Regional Development (EFRD), and the French Fund for Competitiveness Clusters (FUI).

References

Achard P, Renou JP, Berthomé R, Harberd NP, Genschik P (2008) Plant DELLAs restrain growth and promote survival of adversity by reducing the levels of reactive oxygen species. Curr Biol 18:656–660

Allard RW, Bradshaw AD (1964) Implications of genotype-environmental interactions in applied plant breeding. Crop Sci.4:503–508

Andersen JR, Schrag T, Melchinger AE, Zein I, Lübberstedt T (2005) Validation of Dwarf8 polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). Theor Appl Genet 111:206–217

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M et al (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLoS Genet 1:e60

Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J (2004) BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. Bioinformatics 20:2324–2326

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y et al (2010) Genome-wide association study of 107 traits in *Arabidopsis thaliana* inbred lines. Nature 465:627–631

Austin DF, Lee M, Veldboom LR, Hallauer AR (2000) Genetic mapping in maize hybrid progeny across testers and generations: grain yield and grain moisture. Crop Sci 40:30–39

Baack EJ, Sapir Y, Chapman MA, Burke JM, Rieseberg LH (2008) Selection on domestication traits and QTLs in crop-wild sunflower hybrids. Mol Ecol 17:666–677

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B Met 57:289–300

Bert PF, Jouan I, Tourvielle de Labrouhe D, Serre F, Philippon J, Nicolas P, Vear F (2003) Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.). 3. Characterisation of QTL involved in developmental and agronomic traits. Theor Appl Genet 107:181–189

Blackman BK, Rasmussen DA, Strasburg JL, Raduski AR, Burke JM, Knapp SJ, Michaels SD, Rieseberg LH (2011) Contributions of flowering time genes to sunflower domestication and improvement. Genetics 187:271–287

Bouzidi MF, Badaoui S, Cambon F, Vear F, De Labrouche DT, Nicolas P, Mouzeyar S (2002) Molecular analysis of a major locus for resistance to downy mildew in sunflower with specific PCR-based markers. Theor Appl Genet 104:600–952

Bowers JE, Bachlava E, Brunick RL, Rieseberg LH, Knapp SJ, Burke JM (2012) Development of a 10,000 locus genetic map of the sunflower genome based on multiple crosses. Genes Genomes Genetics 2:721–729

Brachi B, Faure N, Horton M, Flahauw E, Vazquez A et al (2010) Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet 6:e1000940

Breseghello F, Sorrells ME (2006) Association analysis as a strategy for improvement of quantitative traits in plants. Crop Sci 46:1323–1330

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ et al (2009) The genetic architecture of maize flowering time. Science 325:714–718

Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. Genetics 161:1257–1267

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2007) ASReml-R reference manual. The State of Queensland, Department of Primary Industries and Fisheries, Brisbane

Charcosset A, Mangin B, Moreau L, Combes L, Jourjon MF (2000) Heterosis in maize investigated using connected RIL populations. In: Quantitative genetics and breeding methods: the way ahead. INRA, Paris, pp 89–98

Coque M, Mesnildrey S, Romestant M, et al (2008) Sunflower lines core collections for association studies and phenomics. In: Proceedings ASTA Conference, Cordoba

Crouzillat D, De la Canal L, Perrault A, Ledoigt G, Vear F, Serieys H (1991) Cytoplasmic male sterility in sunflower: comparison of molecular biology and genetic studies. Plant Mol Biol 16:415–426

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

Famoso AN, Zhao K, Clark RT, Tung C, Wright MH, Kochian LV, McCouch SR (2011) Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. PLoS Genet 7:e1002221

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Flint-Garcia SA, Thuillet A-C, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J 44:1054–1064

Fusari CM, Lia VV, Hopp HE, Heinz RA, Paniego NB (2008) Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. BMC Plant Biol 8:7

Fusari CM, Rienzo JA, Troglia C (2012) Association mapping in sunflower for *Sclerotinia* head rot resistance. BMC Plant Biol 12:93

Gallais A (1984) An analysis of heterosis vs. inbreeding effects with an autotetraploid cross-fertilized plant *Medicago sativa* L. Genetics 106:123–137

Gaut BS, Long AD (2003) The lowdown on linkage disequilibrium. Plant Cell 15:1502–1506

Gentzbittel L, Zhang YX, Vear F, Griveau B, Nicolas P (1994) RFLP studies of genetic relationships among inbred lines of the cultivated sunflower, *Helianthus annuus* L.: evidence for distinct restorer and maintainer germplasm pools. Theor Appl Genet 89:419–425

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding. Iowa State Univ Press, Ames

Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol 33:54–78

Holland JB (2004) Implementation of molecular markers for quantitative traits in breeding programs—challenges and opportunities. In: Fischer T, Turner N, Angus J, McIntyre L, Robertson M, Borrell A, Lloyd D (eds) New directions for a diverse planet: proceedings for the 4th international crop science congress. Brisbane, Australia

Horne EC, Kumpatla SP, Patterson KA, Gupta M, Thompson SA (2004) Improved high-throughput sunflower and cotton genomic DNA extraction and PCR fidelity. Plant Mol Biol Rep 22:83a–83i

Ingvarsson PK (2005) Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European aspen (*Populus tremula* L., Salicaceae). Genetics 169:945–953

Jourjon MF, Jasson S, Marcel J, Ngom B, Mangin B (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21:128–130

Jung M, Ching A, Bhattramakki D, Dolan M, Tingey S et al (2004) Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. Theor Appl Genet 109:681–689

Kane NC, Gill N, King MG, Bowers JE, Berges H et al (2011) Progress towards a reference genome for sunflower. Botany 89:429–437

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, Gao W et al (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics 177:457–468

Lander ES, Botstein DB (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Leon AJ, Andrade FH, Lee M (2000) Genetic mapping of factors affecting quantitative variation for flowering in sunflower. Crop Sci 40:404–407

Leon AJ, Lee M, Andrade FH (2001) Quantitative trait loci for growing degree days to flowering and photoperiod response in Sunflower (*Helianthus annuus* L.). Theor Appl Genet 102:497–503

Luna A, Nicodemus KK (2007) snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics 23:774–776

Mackay TFC (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35:303–339

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

Maenhout S, De Baets B, Haesaert G (2009) Marker-based estimation of the coefficient of coancestry in hybrid breeding programmes. Theor Appl Genet 118:1181–1192

Mandel JR, Dechaine JM, Marek LF, Burke JM (2011) Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. Theor Appl Genet 123:693–704

Mangin B, Siberchicot A, Nicolas S, Doligez A, This P et al (2011) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. Heredity 108:285–291

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. Genet Res 23:23–35

Melchinger AE, Utz HF, Schon CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. Genetics 149:383–403

Mestries E, Gentzbittel L, Labrouhe DT, Nicolas P, Vear F, Am S (1998) Analyses of quantitative trait loci associated with resistance to *Sclerotinia sclerotiorum* in sunflowers (*Helianthus annuus* L.) using molecular markers. Mol Breed 4:215–226

Mir RR, Kumar N, Jaiswal V, Girdharwal N, Prasad M, Balyan HS, Gupta PK (2012) Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. Mol Breed 29:963–972

Mokrani L, Gentzbittel L, Azanza F, Fitamant L, Al-Chaarani G, Sarrafi A (2002) Mapping and analysis of quantitative trait loci for grain oil and agronomic traits using AFLP and SSR in sunflower (*Helianthus annuus* L.). Theor Appl Genet 106:149–156

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21:2194–2220

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

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall J, Wang J, Zhao K, Kalbfleisch T, Schultz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol 3:e196

Patterson N, Price AL, Reich D (2006) Population structure and Eigen analysis. PLoS Genet 2:e190

Poormohammad Kiani S, Maury P, Nouri L, Ykhlef N, Grieu P, Sarrafi A (2009) QTL analysis of yield-related traits in sunflower under different water treatments. Plant Breeding 128:363–373

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

R Development Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0

Rafalski JA (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci 162:329–333

Raman H, Raman R, Eckermann P et al (2012) Genetic and physical mapping of flowering time loci in canola (*Brassica napus L.*). Theor Appl Genet 126:119–132

Rieseberg LH, Willis JH (2007) Plant speciation. Science 317:910–914

Sarnowski TJ, Świez S, Pawlikowska K, Kaczanowski S, Jerzmanowski A (2002) AtSWI3B, an Arabidopsis homolog of SWI3, a core subunit of yeast Swi/Snf chromatin remodeling complex, interacts with FCA, a regulator of flowering time. Nucl Acids Res 30:3412–3421

Shen CH, Krishnamurthy R, Yeh KW (2009) Decreased L-ascorbate content mediating bolting is mainly regulated by the galacturonate pathway in Oncidium. Plant Cell Physiol 50:935–946

Sun TP (2010) Gibberellin-GID1-DELLA: a pivotal regulatory module for plant growth and development. Plant Physiol 154:567–570

Sun G, Zhu C, Kramer MH, Yang SS, Song W, Piepho HP, Yu J (2010) Variation explained in mixed-model association mapping. Heredity 105:333–340

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF et al (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc Natl Acad Sci USA 98:9161–9166

Thornsberry JM, Goodman MM, Doebley J et al (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nature genet 28:286–289

Tian F et al (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

Vear F, Serre F, Jouan-Dufournel, Bert I, Roche PF, Walser SP, de Labrouhe DT, Vincourt P (2008) Inheritance of quantitative resistance to downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). Euphytica 164:561–570

Vincourt P, As Sadi F, Bordat A, Langlade N, Gouzy J, Pouilly N, Lippi Y, Serre F, Godiard L, Tourvieille de Labrouhe D, Vear F (2012) Consensus mapping of major resistance genes and independent QTL for quantitative resistance to sunflower downy mildew. Theor Appl Genet 5:909–920

Wills DM, Burke JM (2007) Quantitative trait locus analysis of the early domestication of sunflower. Genetics 176:2589–2599

Wright S (1951) The genetic structure of populations. Ann Eugen 15:323–354

Xu L, Zhao Z, Dong A et al (2008) Di-and tri-but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. Mol Cell Biol 28:1348–1360

Yan J, Warburton ML, Crouch J (2011) Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. Crop Sci 51:433–449

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4