

Association mapping reveals gene action and interactions in the determination of flowering time in barley

Silke Stracke · Grit Haseneyer ·
Jean-Baptiste Veyrieras · Hartwig H. Geiger ·
Sascha Sauer · Andreas Graner · Hans-Peter Piepho

Received: 8 February 2008 / Accepted: 9 September 2008 / Published online: 2 October 2008
© The Author(s) 2008. This article is published with open access at Springerlink.com

Abstract The interaction between members of a gene network has an important impact on the variation of quantitative traits, and can influence the outcome of phenotype/genotype association studies. Three genes (*Ppd-H1*, *HvCO1*, *HvFT1*) known to play an essential role in the regulation of flowering time under long days in barley were subjected to an analysis of nucleotide diversity in a collection of 220 spring barley accessions. The coding region of *Ppd-H1* was highly diverse, while both *HvCO1* and *HvFT1* showed a rather limited level of diversity. Within all three genes, the extent of linkage disequilibrium was variable, but on average only moderate. *Ppd-H1* is strongly associated with flowering time across four environments, showing a difference of five to ten days between the most

extreme haplotypes. The association between flowering time and the variation at *HvFT1* and *HvCO1* was strongly dependent on the haplotype present at *Ppd-H1*. The interaction between *HvCO1* and *Ppd-H1* was statistically significant, but this association disappeared when the analysis was corrected for the geographical origin of the accessions. No association existed between flowering time and allelic variation at *HvFT1*. In contrast to *Ppd-H1*, functional variation at both *HvCO1* and *HvFT1* is limited in cultivated barley.

Introduction

The timing of flowering is a key component of plant adaptation and so represents a major objective for crop breeding. Although some major genes have been identified in the control of flowering time in crop species (Cockram et al. 2007; Hay and Ellis 1998; Laurie 1997), the molecular

Communicated by I. Romagosa.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0896-y) contains supplementary material, which is available to authorized users.

S. Stracke · G. Haseneyer · A. Graner (✉)
Leibniz Institute of Plant Genetics and Crop
Plant Research (IPK), 06466 Gatersleben, Germany
e-mail: graner@ipk-gatersleben.de

Present Address:
S. Stracke
Department of Crop Sciences, Quality of Plant Products,
University of Göttingen, Carl-Sprengel-Weg 1,
37075 Göttingen, Germany
e-mail: sstrack@gwdg.de

Present Address:
G. Haseneyer
Plant Breeding, Technische Universität München/Centre of Life
and Food Sciences Weihenstephan, Am Hochanger 4,
85350 Freising, Germany

J.-B. Veyrieras
UMR, INRA UPS-XI INAPG CNRS, Génétique Végétale,
Ferme du Moulon, 91190 Gif-sur-Yvette, France

H. H. Geiger
Institute of Plant Breeding, Seed Science,
and Population Genetics, University of Hohenheim,
70593 Stuttgart, Germany

S. Sauer
Max-Planck-Institute for Molecular Genetics,
14195 Berlin-Dahlem, Germany

H.-P. Piepho
Institute for Crop Production and Grassland Research,
University of Hohenheim, 70593 Stuttgart, Germany
e-mail: piepho@uni-hohenheim.de

dissection of flowering time in the leading plant model *Arabidopsis thaliana* has shown that a complex network of regulatory pathways is involved (reviews: Ausin et al. 2005; Hay and Ellis 1998; Henderson and Dean 2004; Imaizumi and Kay 2006; Mouradov et al. 2002; Putterill et al. 2004). Among the crop species, the best understood system is rice, where five heading date genes (*Hd1–Hd5*) have been successfully tagged or isolated, and the DNA structures of these genes show a remarkable level of similarity to those in *A. thaliana* (Kojima et al. 2002; Lin et al. 2003; Yamamoto et al. 1998; Yano et al. 2000, 2001). Two environmental cues are involved in the determination of the flowering time of barley. One of these, temperature, acts through the vernalisation pathway; the other is day-length, acting through the photoperiod sensitivity *Ppd* genes. One of the latter, *Ppd-H1*, has recently been cloned (Turner et al. 2005). As it represents a homolog of *A. thaliana* *PRR7*, it may be involved in the functioning of the circadian clock (Nakamichi et al. 2005). The barley homologs of *CONSTANS* (*HvCO1*, 7HS; Griffiths et al. 2003) and *FLOWERING LOCUS T* (*HvFT1* synonym of *Vrn-H3*, 7HS; Faure et al. 2007; Turner et al. 2005; Yan et al. 2006) both act downstream of *Ppd-H1*, 2HS (Turner et al. 2005). In rice, these genes are represented by *Hd1* and *Hd3a*, respectively. In *A. thaliana*, the circadian clock regulated *CO* encodes a B-box-type zinc finger transcriptional activator (Putterill et al. 1995), which interacts with a transcription factor complex to induce the expression of *FT* (Ben-Naim et al. 2006; Cheng and Wang 2005; Kardailsky et al. 1999; Suarez-Lopez et al. 2001). While the *CO* gene product acts exclusively in the photoperiodic pathway (Yanovsky and Kay 2002), the *FT* product is involved in other flowering pathways as well (Abe et al. 2005; Ausin et al. 2005; Kardailsky et al. 1999; Wigge et al. 2005).

The identification of the three genes *Ppd-H1*, *HvCO1*, and *HvFT1* as players in the determination of flowering time in barley makes them attractive as candidates for a phenotype/genotype association analysis across a wide germplasm sample. In this study we set out to search for associations between nucleotide diversity at these three genes and variation in flowering time within a geographically widely distributed collection of spring barleys. The multi-factorial inheritance of most quantitative traits complicates their genetic analysis. Although the generally adopted assumption is that quantitative trait loci (QTL) act additively, significant QTL \times QTL interactions have been observed across a range of plant species (Alonso-Blanco et al. 1998; Caicedo et al. 2004; Eshed and Zamir 1996; Juenger et al. 2005; Li et al. 1997, 2000; Lukens and Doebley 1999). We have analyzed the present data set to examine the impact of statistical gene \times gene interactions on the outcome of association results.

Materials and methods

Plant material

The germplasm collection consisted of 220 spring barley accessions originating from Europe (EU, $n = 108$), East Asia (EA, $n = 39$), West Asia and North Africa (WA, $n = 44$), and America (AM, $n = 29$) (Supplemental Table S1). The accessions are a subset of the full Barley Core Collection (“BCC”; Knüpffer and Hintum 2003), chosen to represent the widest possible range of diversity, with some additional West Asian accessions (“HOR”) provided by the Gatersleben Genebank.

Phenotypic evaluation

Spring sown field trials were conducted at three locations, two in Southern (Stuttgart-Hohenheim and Irlbach) and one in Northern (Bergen-Wohlde) Germany. Flowering time was recorded in 2004 at Irlbach and at all three sites in 2005. Each trial was arranged in microplots in a 15×25 lattice design, with three replicates. Flowering date was recorded as days after sowing (das), using the date at which half of the ear had emerged from the flag leaf of the main tiller in 50% of the plants within a given plot.

Genotypic analysis

Genomic DNA was extracted from pooled leaf tissue of six plants per accession, as described elsewhere (Graner et al. 1991). Twenty-five barley expressed sequence tag-derived simple sequence repeat markers (EST-SSRs) (listed in Supplemental Table S2; Thiel et al. 2003) were used to characterize population structure. The criteria applied for marker selection were map position (to maximise genome coverage) and informativeness. The SSRs were grouped into multiplex sets of three to six each, and amplicons were generated with the Multiplex PCR kit (QIAGEN), using an amplification program of 95°C/15 min, followed by 40 cycles of 94°C/30 s, 60°C/30 s and 72°C/15 s, and a final extension step of 72°C/10 min. PCR products were separated on 6% polyacrylamide gels on an ABI377 system (Applied Biosystems/Applera, Darmstadt, Germany), and profiles were analyzed with GeneScan 3.7.1 und Genotyper 3.7 software.

Sequence analysis

Primers were designed to amplify coding and non-coding sequence intervals of the candidate genes, covering 1230 bp (*Ppd-H1*), 1772 bp (*HvCO1*) and 1974 bp (*HvFT1*). The genomic position of each amplicon was inferred from the respective reference sequence (*Ppd-H1*–

AY970701, *HvCOI*–AF490467, and *HvFT1*–DQ100327). Primer sequences and relevant reaction conditions are given in supplemental Table S3. The *Ppd-H1* fragment includes the region from intron 4 to the beginning of exon 7, and covered the entire exon 5 and 6 as well as the entire intron 5 and 6. The region contains one of the four base pair positions diagnostic for the early and the late flowering alleles *Ppd-H1* and *ppd-H1* (Turner et al. 2005). The *HvCOI* sequence covered most of the gene, with small gaps in the first intron and at the 5' and 3' ends. *HvFT1* was resequenced in its entirety, except for 100 bp at its 3' end. As *FT* regulates flowering time in a dosage-dependent manner (Kardailsky et al. 1999), an additional 1 kbp of the 5' upstream region was included. In preparation for sequencing, DNA fragments were amplified using AccuPrime DNA polymerase (Invitrogen). The PCR amplicons were purified in 384 well plates and their concentrations were equalized. Ten nanogram PCR product served as the template for cycle-sequencing using ABI BigDye Terminator 3.1 chemistry on a capillary automatic sequencing device (3730xl ABI 96; Applied Biosystems/Applera, Darmstadt, Germany). The amplification primers were used as sequencing primers. For quality scoring a software package based on the poly-phred system was used (Nickerson et al. 1997). Alignments were compiled and analyzed using Sequencher v4.1 (Gene Codes Corporation, Ann Arbor, USA) and Bioedit v4.7.8 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>).

Population structure

Population structure was analysed in two ways. First, within a Bayesian framework, a Markov chain Monte Carlo algorithm was used to sample from the joint posterior distribution of the subpopulation allele frequencies, and assignment of individuals to particular sub-populations was effected with STRUCTURE version 2.1 (Pritchard et al. 2000a; <http://pritch.bsd.uchicago.edu/structure.html>). Q_{ST} is the matrix of admixture proportions obtained with STRUCTURE. For between two and ten subpopulations, ten independent simulations were performed, using the admixture model and a burn-in of 500,000 followed by 5,000,000 iterations. The 'optimal' number of subpopulations K was chosen on the basis of the second order rate of change of the likelihood function with respect to K (Evanno et al. 2005). The second approach has been developed by Veyrieras et al. (2006) from the Frequentist framework devised by Tang et al. (2005). For a given value of K , an expectation maximization (EM) algorithm was run with ten random starting points to obtain the maximum likelihood estimates of both the admixture proportions and the population allele frequencies. Only the run with the highest likelihood was retained. For each value of K , the cells of the

allele count matrix C ($C_{ij} = 1$ if individual i carries allele j , 0 otherwise) were adjusted according to their expected frequencies under the admixture model with K populations (i.e., the estimated individual admixture proportion was multiplied by the allele frequencies in each population). Then the spectral norm of the resulting adjusted matrix M (i.e., the eigenvalue of the covariance matrix $M'M$) was used as a measure of the linkage disequilibrium (LD) between marker alleles not explained by the current admixture model. Spectral norms are expected to decrease until reaching a plate after the 'optimal' value of K . The model selection is thus based on a scree plot of these spectral norms, in order to identify the value of K after which the decrease is no longer significant (see Supplemental Figure S1). This value is referred to as K_{EM} , and the associated matrix of admixture proportions as Q_{EM} .

Allele diversity

All cultivars were treated as pure lines and the small proportion of non-homozygous data points was treated as missing data. For the purpose of statistical analysis, indels were regarded as single site polymorphisms. Only haplotypes represented by at least two accessions were considered. To assess nucleotide variability across the candidate sequences, and to measure and compare the diversity within and across geographical origins, both nucleotide (π) and haplotype (hd) diversities were estimated using DnaSP (Version 3.51; Rozas and Rozas 1999; <http://www.ub.es/dnasp/>). Estimates for LD were obtained with TASSEL software (v1.0.9; <http://www.maizegenetics.net/bioinformatics/tasselindex.htm>) applying the measurement r^2 (squared correlation coefficient; Hill and Robertson 1968) and the significance of any LD was determined using Fisher's two-tailed exact test. Values of r^2 were plotted against the pairwise distance between polymorphic sites, or presented in a matrix plot.

Association analysis

For the analysis of phenotypic data, the adjusted means per environment were first computed, and these were subjected to a mixed model analysis over trials. Individual trials were analyzed by a mixed model using fixed effects for genotype and replicate, and random block effects for recovery of information. Variance components were estimated by the REML method. For each trial, adjusted means and weights were computed from the diagonal elements of the inverse of the asymptotic variance–covariance matrix of adjusted means (Smith et al. 2001). In the subsequent step, the following terms were fitted: trial main effect (fixed), genotype main effect (random), genotype-by-trial interaction (random). While our final model assumes random genotypes,

we initially take the genotypic effect as fixed in the first step. This is necessary because taking genotypes as random already in the first step we would be estimating genotype means by best linear unbiased prediction (BLUP), which involves shrinkage. The degree of shrinkage usually varies among trials, and so BLUPs would be differently scaled and therefore not comparable across trials. For this reason, BLUPs cannot be used as data in stage two (Smith et al. 2001). Taking genotypes as fixed in stage one, we obtain unbiased estimates of genotype means, for which the model in stage two is (approximately) valid. In addition, the population structure was modeled by fixed-effects regression on an $N \times K$ matrix Q of probabilities of membership for K sub-population membership and each of N genotypes. The Q matrix was either computed using the approach of Pritchard et al. (2000a) or Veyrieras et al. (2006) as described above. For model comparison, two approaches were followed to compute a pairwise kinship matrix using the 25 background SSRs. One approach was equivalent to ridge regression for the allele effects. We generated a dummy variable for each allele taking value unity when the allele was present and zero when absent. For each dummy a random linear regression coefficient was fitted. Coefficients for all alleles were assumed to be sampled from a common normal distribution. The model was fitted by using a Toeplitz structure as implemented in the MIXED procedure of the SAS system. It corresponds to fitting a linear variance structure with a fixed matrix corresponding to a kinship matrix. For further details, including implementation in the MIXED procedure see Ruppert et al. (2003). The second kinship matrix was computed based on estimated Nason's coefficients (Loiselle et al. 1995) using SPAGeDI version 1.2 software (Hardy and Vekemans 2002). Negative kinship values were set to zero following Yu et al. (2006). Each kinship matrix was fitted for random genotype main effects and random genotype-by-trial interaction effects. Model selection with regard to admixture probabilities (Q) and kinship was performed using Akaike's Information Criterion (AIC) in "smaller-is-better" form. All these analyses used components of PROC MIXED in the SAS System (v9.1.3) package. ML was used instead of REML for model selection to enable comparison between models with different fixed effects structures. Tests of fixed effects based on selected models were then based on the REML method with denominator degrees of freedom approximated by the Kenward-Roger method. Adjusted means were compared by Wald t tests. The family-wise type I error rate was controlled by the Bonferroni-Holm procedure. As the means were not variance balanced, we used the method of (Piepho 2004) to generate a letter display showing the significance of comparisons. In order to identify gene \times gene interactions, a two-way classification model was fitted for two loci at a time. When such digenic epistasis proved statistically

significant, adjusted means were compared for each locus separately. When the interaction was non-significant, the interaction term was ignored and marginal means were computed for each locus.

Power of all tests was negatively affected by unequal cell sized in the factorial classification, which stems from the data structure of the population studied. However, the validity of a parametric test does not rest on a minimal sample size. As the provided data met the normality assumption, the tests are valid in small samples.

Results

Phenotypic variation

Flowering time ranged from 58.7 and 81.7 das with a mean of 70 das (Table 1). The EU accessions were predominantly late flowering, on average four to seven days after those from the remaining regions. Six-rowed accessions flowered on average about five days before two-rowed ones, but for both subgroups the range in flowering time was about 21 days. Both the genetic and genotype \times environment interaction variances were highly significant ($P < 0.01$), with former about eight fold larger than the latter. As a result, heritability was high (0.96).

Population structure

Using the mixed model, region of origin and ear row number explained, respectively, 26.1 and 21.6% of the genetic variation for flowering time (Table 2). A total of 173 alleles was observed across the EST-SSR loci (mean 6.9 alleles per locus). Using either the EM or the STRUCTURE approach, a two subgroup structure attracted the highest probability (Supplemental Figure S1). The Q_{EM} matrix for the mixed model, as estimated by the EM algorithm enjoyed a comparable fit relative to the one derived by

Table 1 Means, minimum (min) and maximum (max) for flowering time (days after sowing, das) of all accessions, and of subsamples grouped by geographic origin and ear type

Germplasm	Number	Mean \pm sd	Min	Max
Overall	220	70.0 \pm 5.26	58.7	81.7
EU	108	72.9 \pm 3.57	60.7	81.7
EA	39	65.8 \pm 5.22	58.7	76.8
WA	44	68.1 \pm 5.21	59.4	78.8
AM	29	68.1 \pm 4.76	58.9	75.7
Two-rowed	126	72.1 \pm 4.40	58.9	81.7
Six-rowed	94	67.2 \pm 4.97	58.7	81.0

sd standard deviation, EU Europe, EA East Asia, WA West Asia and North Africa, AM America

Table 2 Percentage of explained genetic variance, var(G), for flowering time based on different mixed models

Model	% var(G)
Q_{EM}^2	19.4
Row number	21.6
Origin	29.2
Q_{EM}^2 + row number	22.9
Q_{EM}^2 + origin	32.1
Q_{EM}^2 + origin*	32.3
Q_{EM}^2 + row number + origin*	33.9

* Subdivision into two origins: EU and non-EU

STRUCTURE (Supplemental Table S4). In the subsequent association studies, a value of 2 was taken for K_{EM} . Since both kinship matrices did not generate an improved fit (AIC) (Supplemental Table S4), this component was not included in the analyses.

The background structure only accounted for 19.4% of the genetic variation in flowering time, so that each subgroup maintained plenty of variation. The distribution of SSR alleles was in close accordance with the taxonomic subgroups, as defined by ear row number (Fig. 1). The model-based and the row number-based assignment were in agreement for the bulk (>88%) of the accessions. The models adjusting for row number and Q_{EM}^2 + row number explained almost the same proportion of the variation. In view of the distinct phenotypic differences between the EU and non-EU (origin*) accessions, as well as between the row number types, we adjusted the model to Q_{EM}^2 + row number + origin* for the purpose of the association test.

Molecular diversity and LD

Within the *Ppd-H1* amplicon, 12 variable base pair positions were identified (Table 4), equivalent to a mean rate of one polymorphic site per 102 bp. Single nucleotide polymorphisms (SNPs) were more frequent in the coding than in the non-coding region (Table 3). In the coding region,

five of the nine SNPs resulted in a codon replacement. The frequency of the SNP alleles was mostly well balanced, only at the ps1515 site the minor allele had a frequency (MAF) of <0.05. Six haplotypes (*PpdH1*¹ through *PpdH1*⁶) were identified, four of which were present at a frequency > 0.05 (Table 4). In the non-EU material, most of the haplotypes were present at a frequency > 0.05. However, almost 85% of the EU accessions were of haplotype *Ppd-H1*¹. This same proportion of WA and AM accessions carried three, and EA as many as five haplotypes. *Ppd-H1*¹, *-H1*² and *-H1*³ were represented in all the geographical origins, but the frequency of *Ppd-H1*² was <0.05 in the EU material. LD between the SNPs was highly variable (mean r^2 of 0.41, $P \leq 0.001$) (Fig. 2).

Within *HvCO1*, 16 polymorphic sites (one polymorphic site per 111 bp) were defined, comprising 14 SNPs and two indels (Table 5). The non-coding sequence showed a moderate level of nucleotide diversity, but within the coding regions, this was reduced by 15 fold (Table 3). All five coding SNPs and two of the 11 non-coding polymorphic sites were present at a MAF of <0.05. Two (ps427 and ps582) of the five coding region sites defined a change in amino acid residue. Ten haplotypes (*HvCO1*¹ through *HvCO1*¹⁰) present in at least two accessions were recognized, but six of these were rare (<0.05) (Fig. 3). Haplotype frequencies varied markedly between the geographical origins. *HvCO1*¹ was present in >75% of the EU accessions, while the other groups had a more balanced distribution, with the commonest haplotype present in no more than 40% of the accessions. Pairwise LD estimates produced a mean r^2 value of 0.49 ($P \leq 0.001$), with 88% of all highly significant associations ($r^2 > 0.5$) located in intronic sequence (Fig. 2).

Within *HvFT1*, 21 SNPs and four indels (one polymorphic site per 79 bp) were identified (Table 6). Among the eight haplotypes (*HvFT1*¹ through *HvFT1*⁸) identified, only three were present at >0.05 (Fig. 3; Table 6). Across all the whole genotype set, the highest level of nucleotide diversity was in the 5'-flanking interval. The mean diversity in the intron regions was moderate, while the exon sequence revealed only a single SNP (Table 3).

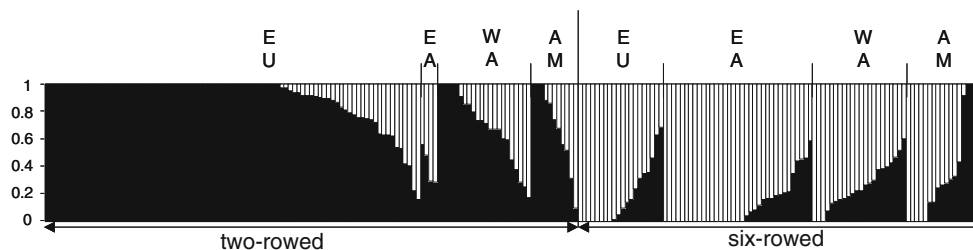


Fig. 1 Calculated population structure, based on genotype at 25 EST-SSR loci. The population was partitioned into two color-coded subpopulations. Each bar represents a single accession, and the colored segments within each bar reflect the proportional contribution of the

two subpopulations to that accession. Accessions are ordered by ear type (two-rowed and six-rowed) and by origin. EU Europe, EA East Asia, WA West Asia and North Africa, AM America

Table 3 Estimates of nucleotide diversity (π) and haplotype diversity (hd) in coding and non-coding regions of the genes *Ppd-H1*, *HvCO1*, and *HvFT1*

Genomic region	L	L* ¹	S	$\pi \times 10^{-3}$ (\pm sd)	H	hd (\pm sd)
<i>Ppd-H1</i> (N = 205)						
Total	1,230		12	2.45 (\pm 0.17)	6	0.59 (\pm 0.03)
Coding	676		9	2.81 (\pm 0.24)	6	0.59 (\pm 0.03)
Non-coding	554		3	2.00 (\pm 0.13)	5	0.59 (\pm 0.03)
<i>HvCO1</i> (N = 200)						
Total	1,772	1771	16	1.55 (\pm 0.13)	10	0.68 (\pm 0.03)
Coding	911		5	0.20 (\pm 0.06)	5	0.12 (\pm 0.03)
Non-coding	861	860	11	2.98 (\pm 0.25)	7	0.66 (\pm 0.03)
<i>HvFT1</i> (N = 195)						
Total	1,975	1969	25	4.61 (\pm 0.07)	8	0.68 (\pm 0.02)
Total* ²	972		8	1.51 (\pm 0.12)	6	0.66 (\pm 0.02)
Coding	437		1	0.12 (\pm 0.05)	2	0.05 (\pm 0.02)
Non-coding	535		7	2.67 (\pm 0.19)	6	0.66 (\pm 0.02)
5'-flanking	1,003	997	17	7.56 (\pm 0.06)	6	0.67 (\pm 0.02)

L fragment length in bp, L*¹ fragment length after treating indels as single sites, S number of polymorphic sites, H haplotype number, sd standard derivation, N number of considered genotypes, total*² 5'-flanking region excluded

Table 4 Polymorphic sites in *Ppd-H1* located in the exonic (e) and intronic (i) regions reflect their position in the AY970701 reference sequence

Haplotypes	Freq.	Polymorphic site position (ps)												
		1	1	1	1	1	1	2	2	2	2	2	2	2
		4	5	7	7	8	8	0	0	0	2	3	3	
		4	1	2	8	3	4	3	3	3	3	2	2	
		3	5	5	5	9	3	2	5	6	0	2	7	
		i4	e5	i5	e6	e6	e6	e6	e6	e6	i6	e7	e7	
					s	s	s			s		s		
AY970701		C	C	T	A	C	A	A	G	C	C	T	C	
<i>Ppd-H1</i> ¹	0.615	A	C	G	.	T	.	.	.	
<i>Ppd-H1</i> ²	0.141	.	.	C	.	.	C	.	.	.	T	.	.	
<i>Ppd-H1</i> ³	0.137	
<i>Ppd-H1</i> ⁴	0.058	.	.	.	G	T	C	.	A	.	T	C	T	
<i>Ppd-H1</i> ⁵	0.034	.	T	C	.	.	C	
<i>Ppd-H1</i> ⁶	0.015	.	.	.	G	.	C	.	.	.	T	C	T	
Model: Q_{EM}^2														
Associated		***	NS	***	**	*	**	***	*	***	***	**	**	
var(G) explained		30.7		8.4	5.0	2.0	4.4	30.7	2.0	30.7	16.0	5.0	5.0	
Model: Q_{EM}^2 + row number + origin*														
Associated		***	NS	*	*	NS	*	***	NS	***	***	*	*	
var(G) explained		20.1		5.7	6.7		6.4	20.1		20.1	11.0	6.7	6.7	

Replacement nucleotides are labeled "s". Dots indicate identity with the reference sequence. Haplotype frequencies within the entire collection are shown. In the lower part of the figure, association tests between the polymorphic site and flowering time calculated with two models are summarized (NS = non-significant, * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$). The % genetic variation explained, var(G) is shown for each significantly associated site

A strong LD block (ps606–ps1199) was detected in the 5'-flanking region, comprising at least 590 bp and flanked by sequence with a highly variable pairwise LD structure (Fig. 2). This interval had a mean r^2 of 0.88 ($P \leq 0.001$), while the mean LD declined sharply to 0.40 ($P \leq 0.001$) in the coding region. This LD block structure was attributable to a split into two tight haplogroups—haplogroup I com-

prising haplotype *HvFT1*¹ and *HvFT1*⁵, and haplogroup II with *HvFT1*² through *HvFT1*⁴, and *HvFT1*⁶.

Interlocus associations were minor compared to intralocus LD, but some differences were apparent between the locus pairs (Supplemental Figure S2). The mean level of significant LD between *Ppd-H1*–*HvFT1* ($r^2 = 0.05$; $P \leq 0.05$, MAF > 0.01) was twofold less than that observed

Fig. 2 Strength and extent of LD at *Ppd-H1*, *HvCO1*, and *HvFT1*. Polymorphic sites with a minor allele frequency of >0.01 were considered for pairwise calculation. Each point in the LD matrix represents a comparison between a pair of polymorphic sites, with the r^2 values displayed below the diagonal, and P values for Fisher’s Exact test above. Points on the diagonal correspond to comparisons of each site with itself. Black, white and hatched margins indicate, respectively, exonic, intronic and 5’-flanking regions. Color codes for r^2 and P values are given

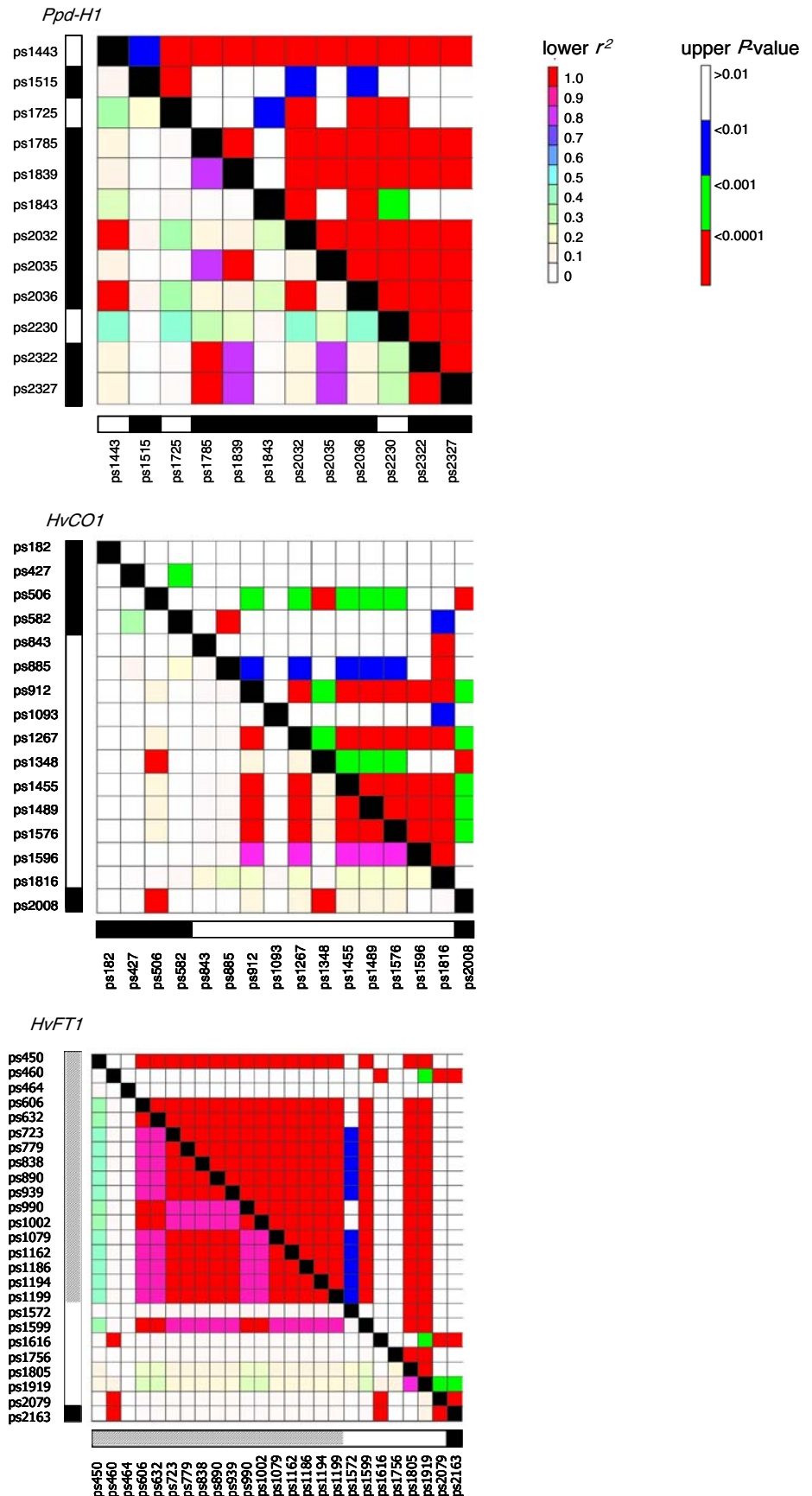


Table 5 Polymorphic sites in *HvCOI* located in the exonic (e) and intronic (i) regions reflect their position in the AF490467 reference sequence

Haplotypes	Freq.	Polymorphic site position (ps)															
		1	4	5	5	8	8	9	1	1	1	1	1	1	1	1	2
		8	2	0	8	4	8	1	0	2	3	4	4	5	5	8	0
		2	7	6	2	3	5	2	9	6	4	5	8	7	9	1	0
									3	7	8	5	9	6	6	6	8
		e1	e1	e1	e1	i1	i1	i1	i1	i1	i1	i1	i1	i1	i1	i1	e2
		s		s													
AF490467		C	G	C	C	AT	A	C	A	A	A	A	T	C	C	T	G
<i>HvCOI</i> ¹	0.520
<i>HvCOI</i> ²	0.129	G	.	G	.	C	C	A	T	C	.
<i>HvCOI</i> ³	0.129	del	C	.
<i>HvCOI</i> ⁴	0.114	del	C	.
<i>HvCOI</i> ⁵	0.030	G	C	.
<i>HvCOI</i> ⁶	0.020	.	.	.	T	.	del	C	.
<i>HvCOI</i> ⁷	0.020	.	.	T	.	.	.	G	.	G	G	C	C	A	.	C	A
<i>HvCOI</i> ⁸	0.020	C	.
<i>HvCOI</i> ⁹	0.010	T
<i>HvCOI</i> ¹⁰	0.010	.	A	.	T	.	del	C	.
Model: Q_{EM}^2																	
Associated		NS	NS	*	*	NS	NS	**	NS	**	NS	**	**	**	NS	**	*
var(G) explained				2.1	1.2			2.7		2.7		2.7	2.7	2.7		4.1	2.1
Model: Q_{EM}^2 + row number + origin*																	
Associated		NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
var(G) explained					1.2												

Replacement nucleotides are labeled “s”. Dots indicate identity with the reference sequence. The starting point of indels is indicated. Haplotype frequencies within the entire collection are shown. In the lower part of the figure, association tests between the polymorphic site and flowering time calculated with two models are summarized (NS = non-significant, * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$). The % genetic variation explained, var(G) is shown for each significantly associated site

for *Ppd-HI-HvCOI* ($r^2 = 0.11$; $P \leq 0.05$, MAF > 0.01) and *HvCOI-HvFTI* ($r^2 = 0.10$; $P \leq 0.05$, MAF > 0.01).

Haplotype–phenotype association

Assuming a population structure (defined by $K_{EM} = 2$), significant haplotype/flowering time associations were identified for all three genes (Table 7). The *Ppd-HI* association explained 32.4% of the genetic variance, and the haplotype classes revealed highly significant differences of five to ten days between accessions carrying the latest flowering *Ppd-HI*¹ haplotype and those carrying any of the five remaining ones (Table 8). One replacement (ps2036), one silent (ps2032) and three intronic sites (ps1725, ps1443, ps2230) were highly significantly ($P < 0.001$) associated with phenotype. The three sites, ps1443, ps2032, ps2036, were in perfect LD to one another (Fig. 2) and are diagnostic for *Ppd-HI*¹; these explain 30.7% of the genetic variation (Table 4).

In contrast, the *HvCOI* and *HvFTI* haplotypes were not able to differentiate between the early and the late-flower-

ing types, as, respectively, 80 and 62% of the haplotype classes were associated with an intermediate phenotype (Table 8). At *HvCOI*, only intronic sites showed any significant ($P \leq 0.01$) association with flowering time (Table 5). Five of these, all in perfect LD with one another (ps912, ps1267, ps1455, ps1489, ps1576), explained 2.7%, while ps1816 explained 4.1% of the genetic variation (Fig. 2, Table 5). At *HvFTI* the 5'-flanking region SNP at ps450 and two intronic sites (ps1805 and ps1919) in high LD with one another ($r^2 = 0.8$; $P \leq 0.0001$) were significantly ($P \leq 0.001$) associated with flowering time, but explained, respectively, only 6.6, 8.8, and 9.7% of the genetic variation (Fig. 2; Table 6).

Digenic epistatic interaction was highly significant for the *Ppd-HI/HvCOI* combination, weakly so for *HvCOI/HvFTI* and non-significant for *Ppd-HI/HvFTI* (Table 7). As *Ppd-HI* exerted the largest influence on flowering time, the impact of allelic distribution at this gene on the association results for *HvFTI* and *HvCOI* was evaluated. Excluding the non-significant gene \times gene interaction from the model, the fixed *HvFTI* effects were calculated by adjusting

Table 6 Polymorphic sites of the *HvFTI* gene located in the exonic (e) and intronic (i) and 5'-flanking (5') regions reflect their position in the DQ100372 reference sequence

Haplotypes	Freq.	Polymorphic site position (ps)																									
		4	4	4	6	6	7	7	8	8	9	9	1	1	1	1	1	1	1	1	1	1	2	2			
		5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'	5'				
DQ100327		C	C	A	C	G	G	del	G	A	C	G	G	T	CTTG	T	C	C	T	C	C	T	A	G	G	G	
<i>HvFTI</i> ¹	0.479	.	.	.	T	C	C	CGCG	A	G	G	del	C	C	del	C	G	T	.	T	.	.	T	C	.	.	
<i>HvFTI</i> ²	0.273	G	T	C	.	.	
<i>HvFTI</i> ³	0.124	
<i>HvFTI</i> ⁴	0.036	del	
<i>HvFTI</i> ⁵	0.031	C	CGCG	A	G	G	.	.	C	del	C	G	T	.	.	.	C	
<i>HvFTI</i> ⁶	0.026	.	G	T	.	T	.	A	T
<i>HvFTI</i> ⁷	0.021	T	C	.	.	
<i>HvFTI</i> ⁸	0.010	G	.	G	T	C	.	.	
<i>Model: Q_{EM}²</i>																											
Associated	***	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	
var(G) explained	6.6																				8.8	9.7					
<i>Model: Q_{EM}² + row number + origin*</i>																											
Associated	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	
var(G) explained	3.2																				6.0	7.2					

Dots indicate identity with the reference sequence. The starting point of indels is indicated. Haplotype frequencies within the entire collection are shown. In the lower part of the figure, association tests between the polymorphic site and flowering time calculated with two models are summarized (NS = non-significant, * = $P \leq 0.05$, ** = $P \leq 0.01$, *** = $P \leq 0.001$). The % genetic variation explained, var(G) is shown for each significantly associated site

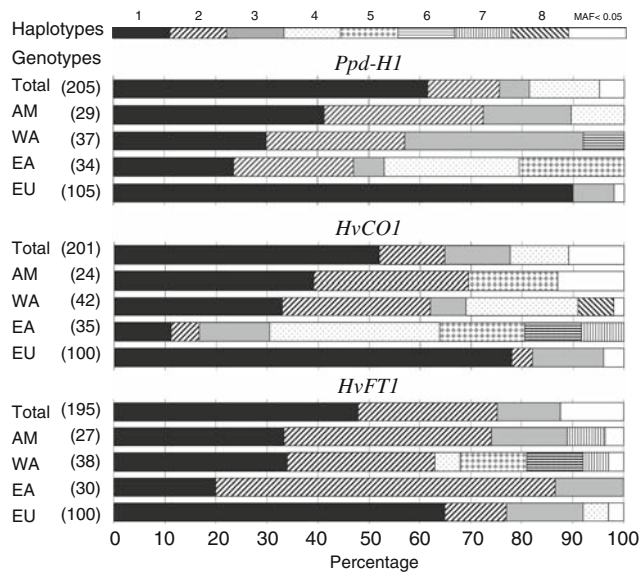


Fig. 3 Haplotype frequency and distribution at *Ppd-H1*, *HvCO1*, and *HvFTI* both across the whole collection and separately for geographical origin (EU Europe, EA East Asia, WA West Asia and North Africa, AM America). The number of relevant accessions is shown in brackets. Each hatched pattern represents a haplotype occurring at a frequency of >0.05. Rarer haplotypes are pooled, and represented by a white bar

for the effect of *Ppd-H1*, leading to a dramatic reduction in the significance of the association between *HvFTI* and flowering time. Phenotypic means of the various haplotype classes did not differ significantly from one another (Table 8). In contrast, a similar adjustment for the effect of *HvFTI* on *Ppd-H1* did not affect the differentiation between the late flowering *Ppd-H1*¹ haplotype and the remaining haplotypes (Table 8). Given the statistically significant *Ppd-H1* × *HvCO1* interaction, the *Ppd-H1* background was fixed for *Ppd-H1*¹ and the earlier flowering haplotypes *Ppd-H1*²⁻⁶ to estimate the effect of *HvCO1* within these haplotype classes (Fig. 4a). One group of haplotypes (*HvCO1*¹, *HvCO1*², *HvCO1*⁶) suggested a background-dependent change in phenotypic expression, while a second group (*HvCO1*³, *HvCO1*⁴, *HvCO1*⁵, *HvCO1*¹⁰) behaved almost independently. In the presence of *Ppd-H1*¹, *HvCO1*¹ and *HvCO1*² generated the latest flowering phenotype, differing significantly (by 8 days) from carriers of *HvCO1*⁴. This earlier flowering haplotype can be identified by a two-base pair deletion (site 843; Table 5). For all other haplotypes resulting in a significantly different phenotype it was not feasible to generate a diagnostic marker. In the *Ppd-H1*²⁻⁶ background the flowering time of *HvCO1*¹ and *HvCO1*²

Table 7 Estimated percentage of genetic variance, var(G), for flowering time explained by haplotypes of the genes *Ppd-H1*, *HvCO1*, and *HvFT1* based on two mixed models

Models for SNP genotypes	Model Q_{EM2}		Model Q_{EM2} + row number + origin*	
	% var(G)	<i>P</i> value [#]	% var(G)	<i>P</i> value [#]
<i>Ppd-H1</i>	32.4	<0.0001	22.4	<0.0001
<i>HvCO1</i>	9.3	0.0005	0	0.1112
<i>HvFT1</i>	9.5	0.0003	3.4	0.0621
<i>Ppd-H1</i> + <i>HvCO1</i> + <i>Ppd-H1</i> × <i>HvCO1</i>	51.0	<0.0001	48.8	<0.0001
<i>Ppd-H1</i> + <i>HvFT1</i> + <i>Ppd-H1</i> × <i>HvFT1</i>	41.5	0.6544	30.6	0.8584
<i>HvCO1</i> + <i>HvFT1</i> + <i>HvCO1</i> × <i>HvFT1</i>	16.9	0.0450	12.3	0.0298

* Subdivision into two origins, EU and non-EU

Significance level of Wald-*t* test; in the case of the 2-factorial models *P* values refer to the interaction

carriers was accelerated to become indistinguishable from carriers of *HvCO1*⁴.

For *HvCO1/HvFT1* the haplotype of each gene was changed in a stepwise manner. In a *HvCO1*¹ background, significant differences between the *HvFT1* haplotype classes were noted, whereas in the *HvCO1*^{2–10} backgrounds no such significant differences occurred. No *HvFT1* haplotype background was associated with any significant variation in the presence of a particular set of *HvCO1* haplotypes (data not shown).

Finally, epistatic interactions were sought using the model Q_{EM2} + row number + origin* which explained most of the phenotypic variation in flowering time (Table 2). Only *Ppd-H1* was consistently associated with flowering time (Table 7). *HvCO1* was not associated with flowering time, although the interaction between *Ppd-H1* and *HvCO1* was highly significant. Once again, significant phenotypic differences were observed between the *HvCO1* haplotype classes in the presence of *Ppd-H1*¹, but none in the *Ppd-H1*^{2–6} background (Fig. 4b).

Discussion

The genetic dissection of quantitative variation is key to the implementation of genomics-based crop improvement strategies. Association mapping can be a powerful means of understanding the genetic basis of quantitative variation, and can even succeed in identifying candidate genes involved in its determination (review: Hirschhorn et al. 2002). We have investigated here the allelic effects of three genes involved in the determination of flowering time, and have used a haplotype-based association analysis in conjunction with statistical models of differing complexity. A special focus was on the impact of gene × gene interactions on the outcome of association results.

Population structure, the target genomic region, and the number of polymorphic sites can all have a major impact on

the pattern of LD and consequently, on the outcome of association studies (Akey et al. 2003; Ke et al. 2004; Pritchard et al. 2000b). The population structure of the barley collection is largely consistent with the phenotype-based grouping into two- and six-rowed ear types, as has been noted in other sets of spring (Brantestam et al. 2007; Hayes and Szucs 2006) and winter cultivars (Matus and Hayes 2002; Stracke et al. 2007; Thiel et al. 2003). Ear type is encoded by a single major gene *v*, located on chromosome 2H (Komatsuda et al. 2007), which cannot account for the existence of genetically distinct subpopulations. However, breeders tend to avoid intermixing these types, as six-rowed barleys are usually destined for feed, and the two-rowed ones for malting (Hayes and Szucs 2006; Kjaer and Jensen 1996; Matus and Hayes 2002).

While the level of nucleotide diversity in the non-coding regions of the three genes investigated was comparable, coding sequences revealed marked differences. The coding region of *Ppd-H1* is particularly diverse, with a substantial number of replacement alleles. This is consistent with the reported role of the *A. thaliana* pseudo-response regulator gene family (*PRR*), where both *PRR7* (orthologous to *Ppd-H1*) and *PRR5* are rich in allelic variation allowing for the wide adaptation of the species (Michael et al. 2003). The late flowering haplotype *Ppd-H1*¹ is synonymous with the *ppd-H1* allele present in the cultivar Triumph (Turner et al. 2005). Surprisingly, in addition to the diagnostic SNPs for *Ppd-H1*¹, two polymorphic intronic sites were also significantly associated with flowering time, but these are independent of the dimorphism between the late and the earlier flowering haplotypes. Given that the assignment to subgroups only imperfectly reflects the genetic background of the collection (Balding 2006), this further association may be attributable to residual background structure within the subgroups.

Many rare exonic SNPs are present in *HvCO1*, and this resulted in the definition of many haplotypes and a low average nucleotide diversity. *A. thaliana* *CO* belongs to a

Table 8 Means and their 95% confidence intervals of the haplotype classes of *Ppd-H1*, *HvCO1* and *HvFT1* for flowering time based on two models

Haplotypes	Model			Model		
	Q_{EM2}			$Q_{EM2} + \text{row number} + \text{origin}^*$		
	Mean	CI ₉₅	P^{\S}	Mean	CI ₉₅	P^{\S}
<i>Ppd-H1</i> ¹	72.56	0.73	. b	72.06	0.86	. b
<i>Ppd-H1</i> ²	66.20	1.50	a .	67.05	1.54	a .
<i>Ppd-H1</i> ³	67.38	1.47	a .	67.85	1.46	a .
<i>Ppd-H1</i> ⁴	65.67	2.54	a .	66.56	2.52	a .
<i>Ppd-H1</i> ⁵	66.88	3.21	a .	67.80	3.18	a b
<i>Ppd-H1</i> ⁶	61.66	4.53	a .	62.47	4.42	a .
<i>HvCO1</i> ¹	71.61	1.15	. b	70.85	1.29	a .
<i>HvCO1</i> ²	68.21	1.94	a b	69.25	1.90	a .
<i>HvCO1</i> ³	71.25	2.13	. b	70.20	2.12	a .
<i>HvCO1</i> ⁴	67.09	2.22	a b	68.33	2.14	a .
<i>HvCO1</i> ⁵	67.83	3.81	a b	68.99	3.58	a .
<i>HvCO1</i> ⁶	63.66	4.23	a .	64.67	3.98	a .
<i>HvCO1</i> ⁷	63.08	5.30	a b	64.31	4.98	a .
<i>HvCO1</i> ⁸	70.67	4.49	a b	71.41	4.34	a .
<i>HvCO1</i> ⁹	70.14	6.33	a b	68.64	6.10	a .
<i>HvCO1</i> ¹⁰	68.23	6.35	a b	67.23	5.94	a .
<i>HvFT1</i> ¹	70.50	0.99	a b	69.82	1.00	a .
<i>HvFT1</i> ²	68.40	1.33	a .	68.99	1.26	a .
<i>HvFT1</i> ³	73.70	1.95	. b	72.39	1.92	a .
<i>HvFT1</i> ⁴	71.61	3.38	a b	70.09	3.30	a .
<i>HvFT1</i> ⁵	72.53	3.99	a b	72.92	3.92	a .
<i>HvFT1</i> ⁶	71.61	4.12	a b	71.57	3.86	a .
<i>HvFT1</i> ⁷	65.31	4.50	a .	67.32	4.42	a .
<i>HvFT1</i> ⁸	70.09	6.34	a b	68.85	5.96	a .
Adjusted for <i>PpdH1</i>						
<i>HvFT1</i> ¹	67.37	1.42	a .	67.48	1.40	a .
<i>HvFT1</i> ²	65.93	1.30	a .	66.19	1.40	a .
<i>HvFT1</i> ³	66.26	2.46	a .	65.76	2.56	a .
<i>HvFT1</i> ⁴	65.88	3.11	a .	66.34	3.08	a .
<i>HvFT1</i> ⁵	60.61	5.25	a .	61.37	5.22	a .
<i>HvFT1</i> ⁶	69.33	3.89	a .	68.93	3.88	a .
<i>HvFT1</i> ⁷	67.47	3.89	a .	68.83	3.96	a .
<i>HvFT1</i> ⁸	62.60	5.43	a .	62.71	5.36	a .
Adjusted for <i>HvFT1</i>						
<i>Ppd-H1</i> ¹	71.76	1.28	. b	72.00	1.38	. b
<i>Ppd-H1</i> ²	63.40	2.14	a .	63.67	2.24	a .
<i>Ppd-H1</i> ³	65.49	1.83	a .	65.99	1.86	a .
<i>Ppd-H1</i> ⁴	65.11	3.00	a .	65.43	3.00	a .
<i>Ppd-H1</i> ⁵	67.24	3.70	a b	67.36	3.70	a b
<i>Ppd-H1</i> ⁶	61.09	4.58	a .	61.25	4.54	a .

[§] Means followed by a common letter are not significantly different according to the Bonferroni-Holm procedure

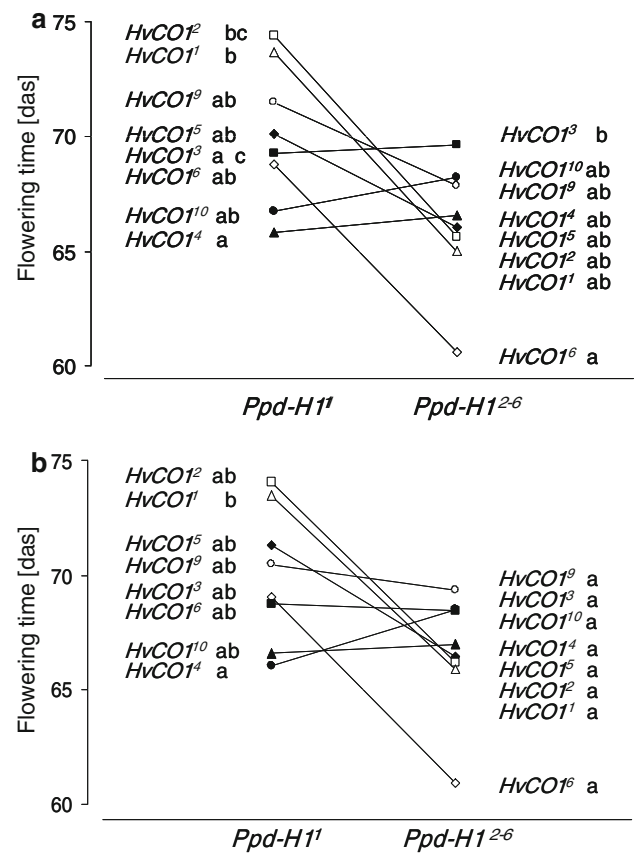


Fig. 4 The interaction between *Ppd-H1* and *HvCO1*. The estimated mean phenotypic values are shown for *HvCO1* haplotypes in carriers of the *PpdH1*¹ and *Ppd-H1*²⁻⁶ haplotypes, applying **a** the model Q_{EM2} , and **b** $Q_{EM2} + \text{row number} + \text{origin}^*$. Haplotypes are denoted by exponents attached to the gene symbol. Means followed by a common letter do not differ significantly from one another, using the Bonferroni-Holm procedure

family of putative transcription factors characterized by two conserved domains. B-boxes near the N-terminus regulate protein-protein interactions, and a CCT domain near the C-terminus controls the nuclear localization of the CO protein (Robson et al. 2001) and binds to a trimeric CCAAT factor (Ben-Naim et al. 2006; Robson et al. 2001). All of the functional *CO* mutations reported to date are located in one or other of these two domains, which underlines their importance to the control of flowering time (Ben-Naim et al. 2006; Lagercrantz and Axelsson 2000; Robson et al. 2001). In *HvCO1* one polymorphic non-replacement site was identified in each of the two domains. As resequencing was incomplete (76% coverage), it is possible that a number of other polymorphic sites was not identified. Nevertheless, because of LD structure, we assume that all possible associations have been identified.

HvFT1 has a divergent diversity structure both up- and downstream of the start codon. The extended LD block in the 5'-flanking region has been noted previously in both

spring and winter barley cultivars (Yan et al. 2006). As this feature is associated neither with flowering time nor with growth habit, it may either play a role in some other signaling pathway, or be an artifact of genetic drift. In *A. thaliana*, the *FT* promoter region is thought to be a potential target for selection for early flowering (Roux et al. 2006) and *FT* is one of a set of genes which is regulated by photoperiodic, vernalization, gibberellin, and developmental signals (Abe et al. 2005; Ausin et al. 2005; Kardailsky et al. 1999; Wigge et al. 2005). Induced mutations at *FT* have revealed the importance of binding sites at the promoter (Helliwell et al. 2006; Huang et al. 1995), the first intron (Helliwell et al. 2006) and of potential binding pockets in the coding sequence (Hanzawa et al. 2005). A motif search in *HvFTI* revealed that ps460 and ps464 in the 5'-flanking region are both located at a potential binding site of a MADS-box transcription factor. However, the minor allele generates a change in the potential binding motif at both sites, but occurred at a frequency of <0.05 and neither site was associated with flowering time. Interestingly, the two intron SNPs that have been recently proposed to be diagnostic between spring and winter barleys (Yan et al. 2006; *Vrn-H3*, SNP270 and SNP384) are identical to the segregating sites SNP1805 and SNP1819 in our spring barley collection. As shown in our study, the potential effect on flowering time of both SNPs loses on significance when adjusting for *Ppd-H1*. These contradictory results show that sampling effects (number of genotypes and genetic background) have a strong influence on the results of association studies.

Overall, our data demonstrate that most of the variation in flowering time in the collection is not caused by functional mutations in either *HvCO1* or *HvFTI*. This seems to be consistent with QTL results from the literature. The location of *HvCO1* did not coincide with the maximum-likelihood peak for flowering QTL (Griffiths et al. 2003), and though *HvFTI* was located in a QTL for the earliness factor *eps7S*, no polymorphism between the cross parents Igri and Triumph was detected (Faure et al. 2007). However, it is still possible that induced mutations in the putative functional domains of both genes could show an analogous effect on flowering time in barley as they do in *A. thaliana*.

The phenotypic differences between the *HvFTI* haplotype classes were mainly attributable to an imbalanced distribution of *Ppd-H1* alleles (Supplemental Table S5). The proportion of the late-flowering *Ppd-H1*¹ carriers in the three most frequent *HvFTI* haplotype classes was 70.6, 35.8 and 95.8% for, respectively, *HvFTI*¹, *HvFTI*² and *HvFTI*³, which led to the identification of spurious significant differences among the haplotype classes. Although *HvFTI* is a key gene for the control of flowering time (Turner et al. 2005), none of the variants exerted any effect

when the data were corrected for the effect of *Ppd-H1*. Direct biological inferences from statistical tests are notoriously risky (Cordell 2002), but our conclusions are strongly supported by the lack of difference in the expression level of *Ppd-H1* between cultivar Triumph (*Ppd-H1*¹) and a closely related line carrying the early flowering *Ppd-H1*³ allele of cultivar Igri (Turner et al. 2005). However, the expression level of *HvFTI* was strongly dependent on the identity of the allele at *Ppd-H1*. *HvCO1* expression was high in the presence of both *Ppd-H1* alleles, but was significantly lower in the *Ppd-H1*¹ background at some measurement time points. As the contrasting genotypes were monomorphic at *HvCO1*, biological evidence for the *Ppd-H1* × *HvCO1* interaction remains to be collected. It is of interest, however, that in rice, epistatic interactions affect all combinations of *Hd1* (*HvCO1*), *Hd2* (*Ppd-H1*) and *Hd3a* (*HvFTI*) (Lin et al. 2000).

As shown in our study, genotypes with identical *Ppd-H1* background revealed significant differences in the flowering time depending on the allelic constitution at *HvCO1*. Hence, it might be useful for practical breeding purposes to select not only for *Ppd-H1* but also for *HvCO1* alleles. With exception of *HvCO1*⁴, however, it was not feasible to generate diagnostic markers for the sequences at *HvCO1* in our data set. Thus, to date re-sequencing might be the only one, but cost-intensive method for the identification of the desirable haplotype.

It has not yet been possible to decide, whether the highly significant *Ppd-H1* × *HvCO1* interaction is a result of co-evolution, or whether it is due to a sampling artifact. Inter-locus LD was significant, but rather weak (mean values of $r^2 = 0.05$ and 0.11). The imbalanced distribution of the alleles in the EU and non-EU (EA, WA, AM) materials may reflect a degree of local adaptation (Supplemental Table S6). Late flowering favors vegetative growth, and thus enhances the amount of assimilate accumulated by the plant where moisture and temperature remain non-limiting; in contrast, early flowering is advantageous where terminal drought or temperature stress is commonplace (Roux et al. 2006). The strikingly high frequencies of the *Ppd-H1*¹ and *HvCO1*¹ haplotypes across the whole collection (76% of the accessions are *Ppd-H1*¹/*HvCO1*¹) is mainly due to their predominance within the EU set, which contributed 76.5% of all entries with this combination. The effect of narrow domestication-related bottlenecks on haplotype diversity between modern barley cultivars and its closest wild relative *H. spontaneum* has been described repeatedly (Badr et al. 2000; Matus and Hayes 2002; Piffanelli et al. 2004).

Consequences for association studies

The current study illustrates the potential of association mapping but also highlights some common pitfalls. The

outcomes agree well with a QTL-based analysis in which variation at *Ppd-H1* accounted for some 60% of the genetic variation in flowering time (Laurie et al. 2004). Note that the association mapping approach delivered a much lower estimate for this proportion. This might be partly attributable to the fact that in the QTL approach only two alleles of each locus were considered from parent lines strongly contrasting in flowering time, whereas our association approach was based on a wide range of germplasm with different allelic compositions.

Results from the present study and from literature (Laurie 1997; Laurie et al. 1994, 2004) indicated both, QTL analysis and association approach, are suitable tools for mapping of quantitative loci with strong effects on flowering time. For future research gene identification might greatly facilitate by the recently suggested integrative method nested association mapping (Yu et al. 2008) where both approaches are combined by using first QTL mapping in sets of multi-line mapping populations to detect chromosomal regions of interest and second, association mapping with high-density markers for fine dissection. Traits controlled by polygenic networks involving a large number of small-effect QTL are probably not readily amenable to the association approach. A recent theoretical treatment that has highlighted the impact of epistasis on QTL detection in association studies was provided by Jannink (2007). Our analysis also underlines the need to consider gene \times gene interactions, both to highlight spurious associations and to identify hidden variation.

Acknowledgments The field trials were supported by the breeding companies Lochow Petkus GmbH (Bergen-Wohlde) and Dr. J. Ackermann & Co (Irlbach). The authors thank K. Trnka, J. Jesse, S. Koch and A. Smyra for their excellent technical assistance, and M. Mohr for her valuable help. This work was financially supported by the German Federal Ministry of Education, and Research (grants 0313098A-C).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309:1052–1056
- Akey JM, Zhang K, Xiong MM, Jin L (2003) The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium. *Mol Biol Evol* 20:232–242
- Alonso-Blanco C, El-Assal SE-D, Coupland G, Koornneef M (1998) Analysis of natural allelic variation at flowering time loci in the *Landsberg erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics* 149:749–764
- Ausin I, Alonso-Blanco C, Martinez-Zapater JM (2005) Environmental regulation of flowering. *Int J Dev Biol* 49:689–705
- Badr A, Muller K, Schafer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F (2000) On the origin and domestication history of barley (*Hordeum vulgare*). *Mol Biol Evol* 17:499–510
- Balding DJ (2006) A tutorial on statistical methods for population association studies. *Nat Rev Genet* 7:781–791
- Ben-Naim O, Eshed R, Parnis A, Teper-Bamnolker P, Shalit A, Coupland G, Samach A, Lifschitz E (2006) The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J* 46:462–476
- Brantestam AK, Bothmer RV, Dayteg C, Rashal I, Tuveesson S, Weibull J (2007) Genetic diversity changes and relationships in spring barley (*Hordeum vulgare* L.) germplasm of Nordic and Baltic areas as shown by SSR markers. *Genet Resour Crop Ev* 54:749–758
- Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proc Natl Acad Sci USA* 101:15670–15675
- Cheng XF, Wang ZY (2005) Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in *Arabidopsis thaliana*. *Plant J* 43:758–768
- Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ (2007) Control of flowering time in temperate cereals: genes, domestication, and sustainable productivity. *J Exp Bot* 58:1231–1244
- Cordell HJ (2002) Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11:2463–2468
- Eshed Y, Zamir D (1996) Less-than-additive epistatic interactions of quantitative trait loci in tomato. *Genetics* 143:1807–1817
- 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
- Faure S, Higgins J, Turner A, Laurie DA (2007) The *FLOWERING LOCUS T*-like gene family in barley (*Hordeum vulgare*). *Genetics* 176:599–609
- Graner A, Jahoor A, Schondelmaier J, Siedler H, Pillen K, Fischbeck G, Wenzel G, Herrmann RG (1991) Construction of an RFLP map of barley. *Theor Appl Genet* 83:250–256
- Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of CONSTANS-like gene families in barley, rice, and *Arabidopsis*. *Plant Physiol* 131:1855–1867
- Hanzawa Y, Money T, Bradley D (2005) A single amino acid converts a repressor to an activator of flowering. *Proc Natl Acad Sci USA* 102:7748–7753
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2:618–620
- Hay RKM, Ellis RP (1998) The control of flowering in wheat and barley: What recent advances in molecular genetics can reveal. *Ann Bot-Lond* 82:541–554
- Hayes P, Szucs P (2006) Disequilibrium and association in barley: thinking outside the glass. *Proc Natl Acad Sci USA* 103:18385–18386
- Helliwell CA, Wood CC, Robertson M, Peacock WJ, Dennis ES (2006) The *Arabidopsis* FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. *Plant J* 46:183–192
- Henderson IR, Dean C (2004) Control of *Arabidopsis* flowering: the chill before the bloom. *Development* 131:3829–3838
- Hill WG, Robertson A (1968) Effects of inbreeding at loci with heterozygote advantage. *Genetics* 60:615–628
- Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. *Genet Med* 4:45–61
- Huang H, Tudor M, Weiss CA, Hu Y, Ma H (1995) The *Arabidopsis* MADS-box gene *Agl3* is widely expressed and encodes a sequence-specific DNA-binding protein. *Plant Mol Biol* 28:549–567

- Imaizumi T, Kay SA (2006) Photoperiodic control of flowering: not only by coincidence. *Trends Plant Sci* 11:550–558
- Jannink J-L (2007) Identifying QTL by genetic background interactions in association studies. *Genetics* 176:553–561
- Juenger TE, Sen S, Stowe KA, Simms EL (2005) Epistasis and genotype–environment interaction for quantitative trait loci affecting flowering time in *Arabidopsis thaliana*. *Genetica* 123:87–105
- Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D (1999) Activation tagging of the floral inducer FT. *Science* 286:1962–1965
- Ke XY, Hunt S, Tapper W, Lawrence R, Stavrides G, Ghorji J, Whitaker P, Collins A, Morris AP, Bentley D, Cardon LR, Deloukas P (2004) The impact of SNP density on fine-scale patterns of linkage disequilibrium. *Hum Mol Genet* 13:577–588
- Kjaer B, Jensen J (1996) Quantitative trait loci for grain yield and yield components in a cross between a six-rowed and a two-rowed barley. *Euphytica* 90:39–48
- Knüpffer H, Hintum TV (2003) Summarised diversity—the barley core collection. In: Rv Bothmer, Knüpffer H, Hintum Tv, Sato K (eds) *Diversity in barley (Hordeum vulgare)*. Elsevier, Amsterdam, pp 259–267
- Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) Hd3a, a rice ortholog of the *Arabidopsis* FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol* 43:1096–1105
- Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper i-class homeobox gene. *Proc Natl Acad Sci USA* 104:1424–1429
- Lagercrantz U, Axelsson T (2000) Rapid evolution of the family of CONSTANS like genes in plants. *Mol Biol Evol* 17:1499–1507
- Laurie DA (1997) Comparative genetics of flowering time. *Plant Mol Biol* 35:167–177
- Laurie DA, Pratchett N, Bezant JH, Snape JW (1994) Genetic analysis of a photoperiod response gene on the short arm of chromosome 2 (2H) of *Hordeum vulgare* (barley). *Heredity* 72:619–627
- Laurie DA, Griffiths S, Dunford RP, Christodoulou V, Taylor SA, Cockram J, Beales J, Turner A (2004) Comparative genetic approaches to the identification of flowering time genes in temperate cereals. *Field Crop Res* 90:87–99
- Li ZK, Pinson SRM, Park WD, Paterson AH, Stansel JW (1997) Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* 145:453–465
- Lin HX, Yamamoto T, Sasaki T, Yano M (2000) Characterization and detection of epistatic interactions of 3 QTLs, Hd1, Hd2, and Hd3, controlling heading date in rice using nearly isogenic lines. *Theor Appl Genet* 101:1021–1028
- Lin HX, Liang ZW, Sasaki T, Yano M (2003) Fine mapping and characterization of quantitative trait loci Hd4 and Hd5 controlling heading date in rice. *Breeding Sci* 53:51–59
- Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 82:1420–1425
- Lukens LN, Doebley J (1999) Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genet Res* 74:291–302
- Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. *Genome* 45:1095–1106
- Michael TP, Salome PA, Yu HJ, Spencer TR, El Sharp, Mcpeek MA, Alonso JM, Ecker JR, McClung CR (2003) Enhanced fitness conferred by naturally occurring variation in the circadian clock. *Science* 302:1049–1053
- Mouradov A, Cremer F, Coupland G (2002) Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14:S111–S130
- Nakamichi N, Kita M, Ito S, Sato E, Yamashino T, Mizuno T (2005) The *Arabidopsis* pseudo-response regulators, PRR5 and PRR7, coordinately play essential roles for circadian clock function. *Plant Cell Physiol* 46:609–619
- Nickerson DA, Tobe VO, Taylor SL (1997) Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* 25:2745–2751
- Piepho HP (2004) An algorithm for a letter-based representation of all pairwise comparisons. *J Comput Graph Stat* 13:456–466
- Piffanelli P, Ramsay L, Waugh R, Benabdelmouna A, D’Hont A, Hollricher K, Jorgensen JH, Schulze-Lefert P, Panstruga R (2004) A barley cultivation-associated polymorphism conveys resistance to powdery mildew. *Nature* 430:887–891
- Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. *Am J Hum Genet* 67:170–181
- Putterill J, Laurie R, Macknight R (2004) It’s time to flower: the genetic control of flowering time. *Bioessays* 26:363–373
- Putterill J, Robson F, Lee K, Simon R, Coupland G (1995) The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc-finger transcription factors. *Cell* 80:847–857
- Robson F, Costa MMR, Hepworth SR, Vizir I, Pineiro M, Reeves PH, Putterill J, Coupland G (2001) Functional importance of conserved domains in the flowering-time gene constans demonstrated by analysis of mutant alleles and transgenic plants. *Plant J* 28:619–631
- Roux F, Touzet P, Cuguen J, Le Corre V (2006) How to be early flowering: an evolutionary perspective. *Trends Plant Sci* 11:375–381
- Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
- Ruppert D, Wand MP, Carroll RJ (2003) *Semiparametric regression*. Cambridge University Press, Cambridge
- Smith A, Cullis B, Gilmour A (2001) The analysis of crop variety evaluation data in Australia. *Aust NZ J Stat* 43:129–145
- Stracke S, Prestel T, Stein N, Perovic D, Ordon F, Graner A (2007) Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding bymovirus resistance in barley. *Genetics* 175:805–817
- Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* 410:1116–1120
- Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 28:289–301
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422
- Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* 310:1031–1034
- Veyrieras J-B, Camus-Kulandaivelu L, Charcosset A (2006) Mining population structure using principal component analysis framework. In: Veyrieras J-B (ed) *Etude du déterminisme génétique de caractères quantitatifs chez les végétaux: Méta-analyse de qtl et études d’association dissertation thesis bioinformatique, UMR*

- INRA CNRS INAP-G UPS-XI, Agro. Paris [INAPG] 2006INAP0015, pp 64–88
- Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* 309:1056–1059
- Yamamoto T, Kuboki Y, Sy Lin, Sasaki T, Yano M (1998) Fine mapping of quantitative trait loci Hd-1, Hd-2 and Hd-3, controlling heading date of rice, as single Mendelian factors. *Theor Appl Genet* 97:37–44
- Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valarik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *vrn3* is an orthologue of *ft*. *Proc Natl Acad Sci USA* 103:19581–19586
- Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* 12:2473–2483
- Yano M, Kojima S, Takahashi Y, Lin HX, Sasaki T (2001) Genetic control of flowering time in rice, a short-day plant. *Plant Physiol* 127:1425–1429
- Yanovsky MJ, Kay SA (2002) Molecular basis of seasonal time measurement in *Arabidopsis*. *Nature* 419:308–312
- 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 Genetics* 38:203–208
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178:539–551