Mol Breeding (2011) 28:25–35
DOI 10.1007/s11032-010-9457-6

A *CONSTANS*-like gene candidate that could explain most of the genetic variation for flowering date in *Medicago truncatula*

Jean-Baptiste Pierre · Matthieu Bogard · Doris Herrmann · Christian Huyghe · Bernadette Julier

Received: 11 January 2010 / Accepted: 3 May 2010 / Published online: 21 May 2010

Abstract Flowering is a critical period during a plant's life cycle, and thus the identification and characterization of genes involved in flowering date control are of great importance in agronomy, breeding, population genetics and ecology. The model species *Medicago truncatula* can be used to detect genes explaining the variation for flowering date, which could also explain this variation in legume crops. The objective of this study was to identify the most promising candidate gene explaining a major quantitative trait locus (QTL) for flowering date previously found in three *M. truncatula* mapping populations. Fine mapping and bioinformatic analysis of bacterial artificial chromosomes (BACs) in the confidence interval of the QTL showed six genes potentially involved in the control of flowering date. Two of these genes, similar to *CONSTANS* and *FT* of *Arabidopsis thaliana* respectively, had genomic mutations when compared to the parents. The transcriptomic study of these genes by semi-quantitative RT-PCR in leaves and flowers sampled at two developmental stages showed that the *CONSTANS*-like gene was differentially expressed in the two parental lines. A gene belonging to the *CONSTANS*-like family could explain the major QTL for flowering date segregating in *M. truncatula* progenies.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-010-9457-6) contains supplementary material, which is available to authorized users.

J.-B. Pierre · M. Bogard · D. Herrmann · C. Huyghe · B. Julier (✉)
INRA, UR 4, Unité de Recherche Pluridisciplinaire Prairies et Plantes Fourragères, BP6, 86600 Lusignan, France
e-mail: bernadette.julier@lusignan.inra.fr

Present Address:
J.-B. Pierre
ARVALIS-Institut du végétal, Pôle Variétés, Semences et Génétique, 91720 Boigneville, France

Present Address:
M. Bogard
INRA, UMR1095 Génétique Diversité et Ecophysiologie des Céréales, Site de Crouël 234 avenue du Brézet, 63100 Clermont-Ferrand, France

Present Address:
D. Herrmann
Ecole Polytechnique Fédérale de Lausanne, Indo Swiss Collaboration in Biotechnology, Lausanne, Switzerland

Present Address:
C. Huyghe
INRA, Présidence Centre Inra Poitou-Charentes, BP6, 86600 Lusignan, France

Keywords Fine mapping · Expression study · Legume · QTL

Introduction

An alfalfa (*Medicago sativa*) relative, *Medicago truncatula*, was proposed as a model plant for legume genomics because it possesses a number of interesting characteristics for both molecular and classical genetics (Barker et al. 1990). Phylogenetically close to the Vicieae tribe that includes alfalfa, clovers (*Trifolium* sp.), pea (*Pisum sativum*) and faba bean (*Vicia faba*), *M. truncatula* has a strong degree of synteny with tetraploid or diploid alfalfa and pea (Julier et al. 2003; Choi et al. 2004; Aubert et al. 2006). Key attributes of *M. truncatula* include diploidy and autogamous fertilization, a small genome size of 470 Mb/1C, a rapid reproductive cycle, a high level of diversity, a large number of available populations and a well-characterized nitrogen-fixing symbiont, *Sinorhizobium meliloti* (Cook 1999; Young et al. 2005). Several tools for molecular genetics and genomics have recently been developed, including mapping populations, markers and genetic maps (Huguet et al. 2007), mutant collections (Tadege et al. 2005), cDNA and genomic libraries (Nam et al. 1999; Gamas et al. 2006; Park and Nam 2006), sequences (Young et al. 2005) and efficient methods of transformation (Chabaud et al. 1996; Trinh et al. 1998). Genetic and genomic resources offer the possibility of studying this model species before investigating genetic bases of traits in the cultivated legume species (Bell et al. 2001).

The flowering process has been analyzed in the model plant *Arabidopsis thaliana*. Four major pathways were identified using genetic and molecular dissections of flowering date mutants: the photoperiod, the autonomous, the gibberelin and the vernalization pathways (Komeda 2004). Many genes are described for their effect on a specific step leading to flowering in interaction with growing conditions.

Genetic variation for flowering date was described in a range of accessions of *M. truncatula* (Delalande et al. 2004; Julier et al. 2007). Genetic determinism of flowering date was studied in a recombinant inbred lines (RILs) population obtained from the cross of two parental lines (Jemalong6 and DZA315.16) which differed in flowering date (Julier et al. 2007). It showed that a major quantitative trait locus (QTL) on chromosome 7 explained between 35.8 and 59.2% of the variation for flowering date. More recently, a QTL at the same location was identified in two other mapping populations, in which it explained between 11 and 34% of the variation (Pierre et al. 2008).

The objectives of this study were (1) to carry out a fine mapping analysis on a population derived from Jemalong6 and DZA315.16, in order to determine the position of the QTL more accurately on chromosome 7 and to reduce its confidence interval, (2) to search for candidate genes in the confidence interval using available *M. truncatula* bacterial artificial chromosome (BAC) sequences and their annotations, and (3) to analyze the sequence and the expression of these candidate genes in both parental lines in order to finally produce a strong hypothesis of the most probable one to cause the flowering date QTL.

Materials and methods

Plant material for QTL analysis

In a previous QTL analysis (Julier et al. 2007; Pierre et al. 2008), the RIL population LR4 from the cross Jemalong6× DZA315.16 was used. A major QTL for flowering date and other morphogenetic traits was detected on chromosome 7 at positions varying from 56.3 to 57.8 cM depending on the season and the year. A confidence interval of 7.54 cM was defined around this QTL region. Among the 199 RILs in the LR4 population, a F6 line that was heterozygous in the region of the QTL on chromosome 7 and homozygous elsewhere was identified (line 105). It was self-pollinated and produced about 2000 F7 seeds. This generation is subsequently called a pseudo-F2 generation because it segregates as a F2 generation on the heterozygous portion. These seeds were manually scarified with sandpaper and placed in Petri dishes for imbibition for 24 h at room temperature. Seedlings were vernalized for 7 days at 4°C and were transferred to a greenhouse on 3 May 2006. The experiment was stopped on 21 July 2006, when all plants had flowered. The plants were grown under natural day-length that increased from 14 h 23 min to 15 h 46 min during the experiment. The temperature in the greenhouse ranged from 18 to 35°C, the

minimum temperature varied from 9 to 25°C, and the maximum from 26 to 45°C. Flowering date was individually recorded when the first flower on the primary branch opened (Moreau et al. 2006) and expressed in degree-days (°C.D) above 0°C from the time of planting. In total, 1,640 plants were observed.

Genetic mapping

Genomic DNAs from the plants of the pseudo-F2 population were extracted according to Cheung et al. (Cheung et al. 1993). Two markers (MTIC040 and MTIC714) flanking the confidence interval (7.54 cM) of the QTL detected in the F6 generation (Julier et al. 2007) were genotyped on the 1,640 plants of the pseudo-F2 generation to identify recombinant plants. PCR was performed in a 6.5-µl volume containing 25 ng of genomic DNA, 500 nM of reverse primer, 50 nM of forward primer with an extended M13 tail, 523 nM of labelled M13 primer and 830 µM dNTP, and amplicons were visualized on a ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). Markers are described in Electronic Supplementary Material 1 along with melting temperature and PCR product length for the Jemalong6 and DZA315.16 alleles.

The markers MTIC714, which corresponds to mtgsp002g05, and MTIC040 are located on the integrated genetic map from the University of Minnesota (http://www.medicago.org/genome/map.php). All contigs inside the confidence interval of the QTL detected in F6 were listed according to the physical map of *M. truncatula* (http://www.medicago.org). One marker per contig was chosen using information collected on the website http://www.medicago.org/genome/downloads.php. Finally, ten SSR markers and two polymorphic gene sequences ("CONSTANS-like" and "ABA-responsive") were used to map the contigs (Electronic Supplementary Material 1). PCR was performed in a 25-µl volume containing 50 ng of genomic DNA, 0.2 µM of each primer (including labelled M13 primer), 200 µM dNTP, 1× Buffer 10×, 2 mM $MgCl_2$, and 0.625 units of Taq polymerase. A MJ Research PTC-100TM was used for amplification with the following PCR profile: 94°C for 4 min, followed by 35 cycles of 94°C for 1 min, optimal melting temperature for 1 min, 72°C for 1.5 min, and a final extension at 72°C for 5 min. Electrophoreses were performed on a LI-COR Gene ImagIR2 DNA analyzer system. Markers were genotyped on the recombinant plants of the pseudo-F2 population and a map was constructed with Carthagene software (de Givry et al. 2005) using a segregation ratio of 1:2:1. A χ^2 test was performed to verify that marker segregation followed the expected ratio. The linkage map was constructed using the Haldane map function and drawn with Biomercator software (Arcade et al. 2004).

QTL analysis

Four classes of genotypes were defined: one with the homozygous genotypes for the alleles Jemalong6 for the two markers MTIC040 and MTIC714, one with the homozygous genotypes for the alleles of DZA315.16, one with the heterozygous genotypes and finally a class of genotypes that showed a recombination between MTIC040 and MTIC714. An analysis of variance was performed using the first three classes to demonstrate the statistical differences between classes for flowering date. QTL analysis by simple interval mapping (Lander and Botstein 1989) with 0.1-cM intervals was performed using the package R-QTL with R software v2.4.1 (Broman et al. 2003). A Bayesian confidence interval was built to determine the position of the QTL (Manichaikul et al. 2006).

Candidate gene selection

Candidate genes were selected within the confidence interval of the QTL defined on the pseudo-F2 population, using an in-silico approach. The amino-acid sequence predictions of the 573 genes carried by the 44 BACs located in the confidence interval of the QTL were obtained from IMGAG (http://www.medicago.org/genome/downloads.php) or Fgenesh (http://www.tigr.org/tigr-scripts/medicago/gbrowse_bacs.pl). These amino-acid sequences were then analyzed with BLASTp using the Uniprot-Swissprot database with default parameters (E value = 1e-6, Blosum62 matrix). Results were analyzed considering a gene list obtained from literature dealing with flowering date in model species (Putterill et al. 2004a; Hecht et al. 2005). Only sequences with strong similarities (E value < 0.001) to genes from this list were selected. To ensure that these predicted genes were expressed, a BLASTn of these candidate

genes was performed on the Legume EST database (http://www.comparative-legumes.org/lis/blasttool.html). Genomic and coding sequences of the selected genes were obtained from the IMGAG database (http://www.tigr.org/tigr-scripts/medicago/IMGAG/imgag_annotator.pl?browse=1).

Sequencing of the candidate genes

In order to detect sequence polymorphism between the two parental lines, gene sequencing was conducted on leaves sampled on 5 April 2005 at the beginning of flowering on plants grown in the greenhouse, on mRNA and/or on genomic DNA. Primers (Electronic Supplementary Material 2) were defined to sequence a portion of the six candidate genes, containing exons and sometimes introns (Electronic Supplementary Material 3). PCR products were sequenced after purification by Millegen (Labège, France). The whole gene cDNA sequence, resulting from both the sequenced cDNA and the cDNA predicted from gDNA, was translated into amino acids using ExPasy (http://www.expasy.org). With the amino acid sequence, a BLASTp search was performed based on the UniProtKB database and a Blosum62 matrix (http://beta.uniprot.org) to verify the predicted protein function. For one candidate gene that was similar to the *CONSTANS*-like gene family of *A. thaliana*, it was important to know its similarity with the groups within this family (Griffiths et al. 2003). Its amino acid sequence was compared to the *CONSTANS*-like genes of *A. thaliana* using standard settings of ClustalW provided by EMBL-EBI, except that UPGMA clustering was used instead of neighbour-joining (http://www.ebi.ac.uk/Tools/clustalw2/index.html).

Plant material for transcriptomic study

Plant material was sampled in winter 2005–2006 in a growth chamber trial carried out to study line and photoperiod effects on flowering date (Pierre et al. 2008). Jemalong6 and DZA315.16 only were sampled in the 12-h day-length treatment. In this trial, a significant difference for flowering date was observed between the two lines. Samples of leaves were harvested before (29 November 2005) and at the beginning of flowering (11 January 2006), and flowers were sampled at the beginning of flowering, on six plants of each line. Samples of each organ and developmental stage for each line were placed in liquid nitrogen.

Transcriptomic study by semi-quantitative RT-PCR

The objective was to evaluate the expression of the six candidate genes included in the confidence interval. The internal control chosen was *MSC27*, a housekeeping gene frequently used for transcriptomic studies in *Medicago* sp. (Crespi et al. 1994). Total RNA extraction was conducted with RNeasy kit (Qiagen) after grinding in liquid nitrogen. RNAs were then treated with DNase and retrotranscripted using "DNeasy" and "Superscript first strand synthesis" kits (Invitrogen) respectively, according to the manufacturer's instructions. Primers (Electronic Supplementary Material 4) used to amplify candidate genes were designed to anneal in exon sequences with Oligo6 software (Molecular Biology Insights, Inc. Cascade, CO, USA). They were designed to have an average length of 20 nucleotides, melting temperature between 57 and 62°C and theoretical PCR amplicon size different from those of the control gene (*MSC27*) and the genomic DNA. PCR amplicons were sequenced after purification by Millegen (Labège, France) and the sequences were aligned with the CDS prediction sequences to make sure that the targeted genes were amplified. The PCR conditions were chosen so that none of the PCR products analyzed reached a plateau at the end of the amplification protocol, and that the two sets of primers used in each reaction did not compete with each other. We performed a competition control for each targeted gene by amplifying the same sample at the same time in the presence of specific primers for either the targeted gene or the internal control *MSC27*, and for both together (Marone et al. 2001). The samples were then run on the same agarose gel for quantification. When competition was detected, primer concentration was adjusted. PCRs were performed in PTC100 and PTC200 thermocyclers (MJ Research Inc.) using "Taq platinium" kits (Invitrogen) as follows: cDNA (2 ng/µl), buffer1×, 200 µM dNTP, 0.05 µM of each *MSC27* forward and reverse primers, 0.2 µM of forward and reverse primers of each target gene, 0.625 units of Taq polymerase, optimal concentration of $MgCl_2$, and ultra-pure water to adjust final volume

to 25 µl. Optimum PCR conditions for each candidate gene are described in Electronic Supplementary Material 4. PCR products were loaded onto ethidium bromide-stained 2% agarose gel in 0.5× TBE buffer. A 100-bp DNA ladder was run on each gel to confirm expected molecular weight of the amplification product. Each set of reactions included a no-sample negative control and a genomic DNA positive control. Images of the RT-PCR ethidium bromide-stained agarose gels were acquired with a ST-50 camera (Fisher Bioblock Scientific) and quantification of the bands was performed using ImageQuant 5.0 software (Molecular Dynamics©, 1998). Band intensity was expressed in pixel absorbance units. The expression ratio between the target gene whose expression had to be determined and the internal control gene (*MSC27*) was calculated to take into account initial variations in sample concentration. Mean and standard deviation of expression ratio on the six repetitions for each organ and developmental stage of each line were calculated after normalization. A Student's *t*-test was carried out to test the effect of lines on expression ratio.

Kinetic of *MtCO* expression

Among the six genes identified in the QTL interval, one was similar to *CONSTANS*-like (*CO*). This gene is known to be differentially expressed in the course of the day, with an increase in expression during the light period and a decrease during the dark period. The analysis of expression of *MtCO* was carried out on the two parental lines five times in the course of 1 day at a vegetative stage.

Seeds of Jemalong6 and DZA315.16 were manually scarified with sandpaper and sown in Petri dishes on 27 June 2007 for imbibition for 24 h. After imbibition, they were vernalized at 4°C for 3 days. Seeds were transplanted into pots in a greenhouse on 2 July 2007. The experimental design comprised three randomized blocks, each including five plants. On 24 July 2007, samples of leaves (youngest mature leaf on a primary branch) were collected from three individual plants of each line at 6:30 a.m., 10:30 a.m., 2:30 p.m., 6:30 p.m. and 10:30 p.m. and placed on ice. Sunrise was at 6:30 a.m. and sunset was at 9:42 p.m. Samples were then stored at −80°C. Extraction and semi-quantitative RT-PCR of these samples was performed as described above. Means and standard deviations for expression ratio at each sampling time of each line were calculated after normalization. A Student's *t*-test was carried out to test the effect of lines on expression ratio.

Results

Phenotypic data in the pseudo-F2 population

The flowering date of the population ranged from 700 to 1,302°C.D with a mean of 979°C.D. Homozygous genotypes with the Jemalong6 alleles flowered around 946°C.D, homozygous genotypes with the DZA315.16 alleles flowered around 1,026°C.D, while heterozygous genotypes flowered around 972°C.D (Electronic Supplementary Material 5). The analysis of variance of plants genetically fixed inside the QTL (homozygous with the alleles of Jemalong6, homozygous with the alleles of DZA315.16 and heterozygous at both MTIC040 and MTIC714) showed a significant difference between the classes ($P < 0.001$). This analysis also showed a partial dominant effect of the Jemalong6 allele ($P < 0.001$).

Fine mapping and QTL detection

A new map of the QTL region was built with the addition of 12 molecular markers genotyped on the 217 recombinant genotypes. This map has a length of 7.5 cM with a distance between markers ranging from 0 to 4.3 cM (Fig. 1). Some markers were located at the same position as CONSTANS-like, ABA and MTIC720 on one hand and MTGSP2e05 and MTIC705 on the other hand. A QTL for flowering date was detected at 1.3 cM near markers ABA/CONSTANS-like/MTIC720 with a LOD score of 40.6. A confidence interval, representing a (1–LOD) unit decrease from the maximum LOD score, spanned 2.4 cM between markers MTIC719 and 19l15a (Fig. 1).

Selection of candidate genes

Within the confidence interval of the QTL delimited by MTIC719 and 19l15a markers, 44 BACs were identified, carrying 573 genes predicted by IMGAG. After a BLASTp analysis of all 573 genes, six genes promoting flowering were selected as candidates (Electronic Supplementary Material 2). *MtCO* has a strong similarity to *CONSTANS* of *A. thaliana,* in

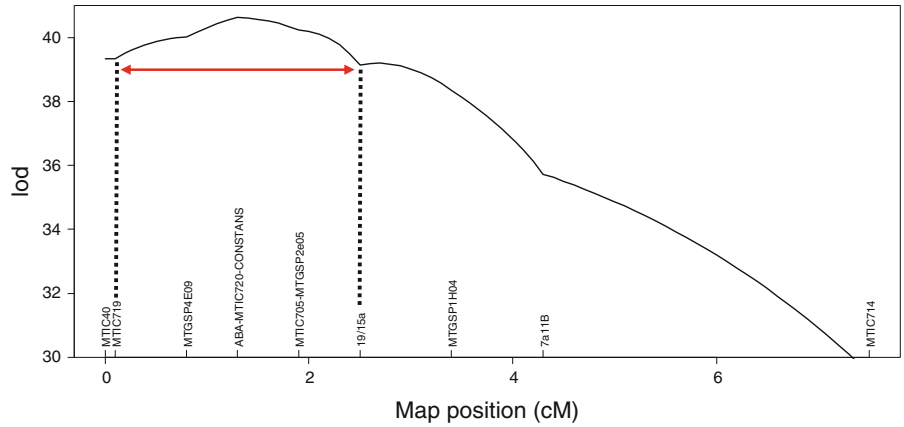

Fig. 1 Fine mapping of the major QTL for flowering date on chromosome 7 in a pseudo-F2 population from the cross Jemalong6× DZA315.16. The position of the markers is presented in centiMorgans. The *arrow* indicates the confidence interval

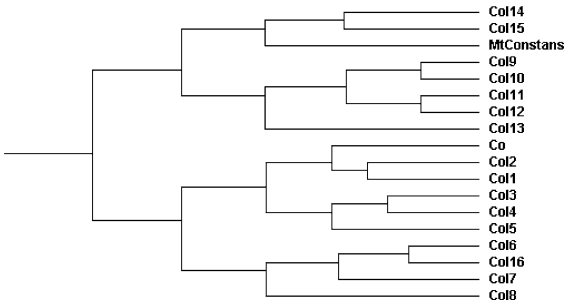

Fig. 2 Clustering of *CONSTANS*-like of *M. truncatula* genotype Jemalong6 (MtConstans) with *CONSTANS* (Co) or *CONSTANS*-like genes (Col) of *A. thaliana* on the basis of amino acid sequences

which it is considered a key gene controlling the flowering switch by being involved in photoperiodism perception (Griffiths et al. 2003). The amino-acid sequence of MtCO was clustered together with COL14 and COL15 of *A. thaliana* (Fig. 2), i.e. among *CONSTANS*-like-genes declared as group III (Griffiths et al. 2003). In addition, the BLASTp search revealed a minimum E-value of 8E-30 and 3E-17 for COL14 and COL15, respectively, whereas CONSTANS protein of *A. thaliana* (Q39057) showed a considerably higher E-value of 8E-4. *MtFD* showed a strong similarity with *FD* of *A. thaliana*. This gene encodes a transcription factor (b-ZIP) interacting with Flowering Locus T (FT) protein in the shoot apical meristem to induce sequential expression of floral identity genes (Blazquez 2005). Three *FT* genes (*MtFTa*, *MtFTb*, *MtFTc*) located in a cluster on BAC AC123593 showed high similarity with *FT* of *A. thaliana* and contained a RKIP (Raf Kinase Inhibitor Protein) domain. The lengths of the genomic sequences of the three *FT* genes were different (1,704, 588 and 2,139 bp, respectively), and the predicted mRNA was shorter for *MtFTb* (396 bp) than for *MtFTa* and *MtFTc* (533 and 520 bp respectively), producing a predicted protein with a truncated RKIP domain in MtFTb (about 40 amino acids are missing). In this model species, FT proteins are translocated from leaves to the shoot apical meristem. This process is thought to be involved in florigen signalization (Huang et al. 2005; Lifschitz et al. 2006). *MtPKS* has a strong similarity with Phytochrome Kinase Substrate 1. Its protein has a negative impact on phytochrome B which is involved in flowering inhibition (Fankhauser et al. 1999). Expressed sequence tags (ESTs) were detected for *MtPKS* (BF644398), *MtCO* (TC86982) and *MtFD* (TC99455), but no ESTs corresponding to *MtFTa*, *MtFTb* and *MtFTc* were specifically identified (Hecht et al. 2005), leading to doubt their expression in plants. *MtCO* was mapped exactly at the peak of the QTL, and *MtFTa*, *MtFTb* and *MtFTc* were also very close to the peak. *MtFD* and *MtPKS* were mapped at the extreme of QTL confidence interval.

Sequencing of candidate genes (Electronic Supplementary Material 3)

MtCO was entirely sequenced and showed a single exon polymorphism between Jemalong6 and DZA315.16 in position +389 that induced no modification of protein sequence. Other polymorphisms

were detected in non-coding regions. The promoter sequencing was conducted 1,400 bp before the ATG codon. The cis-element prediction revealed the presence of a TATA-box similar to the structure of a pea TATA-box (Tjaden et al. 1995) in position −76 before the ATG codon. There were three TATA motifs in Jemalong6 sequence and four motifs for DZA315.16. *MtFTa* and *MtFTc* were fully sequenced and no polymorphism was detected among the parents. For *MtFTb*, an exonic insertion of 2 bp in the RKIP domain in DZA315.16 produced a protein reduced to 123 amino acids instead of 131. For *MtFD* and *MtPKS*, the cDNAs were partly sequenced in Jemalong6 (103 bp in the first exon for *MtFD* and 294 bp in the unique exon between positions 536 and 828 bp for *MtPKS*).

Transcriptomic study

MtFTa and *MtFTc* could not be amplified in our samples harvested in a growth chamber; this fact has to be considered in relation to the absence of ESTs for these genes, even though cDNAs were obtained in samples collected in the greenhouse in spring 2005. No significant variations of *MtFD* and *MtFTb* expression in leaves and flowers were recorded between parental lines, whatever the stage (Table 1). Except in the case of leaves harvested after the beginning of flowering, for which a significantly over-expression of *MtPKS* was recorded in Jemalong6 compared to DZA315.16, expression level of all these three genes (*MtFTa*, *MtFD* and *MtPKS*) was similar in the two lines in both stages.

For *MtCO* in leaves, a decrease of its expression in Jemalong6 was recorded at the beginning of flowering compared to the vegetative stage. Conversely, the expression of *MtCO* did not change in DZA315.16 according to the stage.

Comparison of *MtCO* expression in the two parental lines showed a significantly higher expression of this gene in leaves harvested before flowering in Jemalong6 than in DZA315.16 ($P < 0.05$) with an expression level in Jemalong6 about twice that of DZA315.16. Furthermore, a slightly higher expression of this gene in DZA315.16 than in Jemalong 6 was recorded in leaves harvested at flowering ($P < 0.05$) (Table 1). No difference of *MtCO* expression was recorded in flowers between the two lines.

In the transcriptomic study performed five times in the course of the day, expression of *MtCO* showed a wave-like profile with the highest level at 6:30 p.m. (12 h after sunrise) for the two lines. At this time, a significantly higher expression of this gene was observed in Jemalong6 than in DZA315.16 ($P < 0.05$) (Fig. 3).

Discussion

Fine mapping and in-silico approach for candidate gene selection

The strategy used in this study to select candidate genes combined genetics and bioinformatics. In order to map the locus of interest and to select candidate genes, we first reduced the confidence interval of the QTL and then used information available from *M. truncatula* genome sequencing. This fine-mapping strategy based on 1640 pseudo-F2 plants produced a confidence interval of the QTL of 2.4 cM. In comparison, for the same QTL position, we used three connected RIL mapping populations of about

Table 1 Expression of candidate genes by semi-quantitative RT-PCR

Gene	Leaves before flowering		Leaves at flowering		Flowers	
	Jemalong6	DZA315.16	Jemalong6	DZA315.16	Jemalong6	DZA315.16
MtFTb	0.63 (0.04)	0.62 (0.06)	0.61 (0.06)	0.62 (0.10)	0.71 (0.03)	0.65 (0.05)
MtPKS	0.89 (0.21)	1.03 (0.13)	0.98 (0.07)*	0.90 (0.06)*	0.86 (0.12)	0.90 (0.06)
MtFD	0.81 (0.14)	0.85 (0.27)	0.77 (0.19)	0.73 (0.06)	0.77 (0.03)	0.74 (0.04)
MtCO	1.17 (0.38)*	0.63 (0.07)*	0.49 (0.08)*	0.64 (0.09)*	0.60 (na)	0.60 (0.10)

Results are expressed as the mean of expression ratio between each candidate gene and an internal control (*MSC27*) followed by standard error in brackets

* Significant difference ($P < 0.05$) between the two parental lines

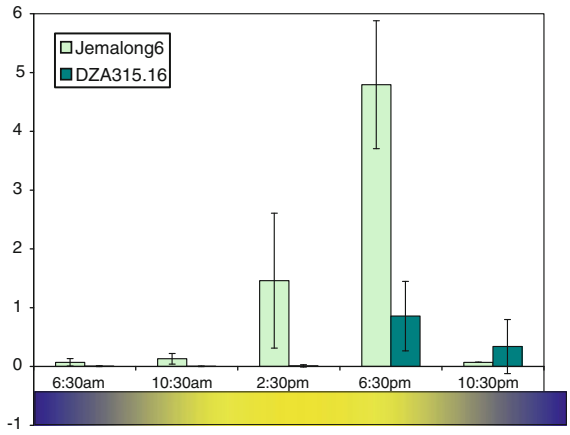

Fig. 3 *MtCO* expression in leaves during a day prior to flowering for two parental *M. truncatula* lines (Jemalong6 and DZA315.16) as the mean of expression ratio between each candidate gene and an internal control (MSC27). *Bars* indicate standard deviations

200 lines each in another study (Pierre et al. 2008). In this case, each population carried this QTL, and the QTL confidence interval calculated over the three populations was only 0.9 cM. Thus a higher precision was obtained in a multiple population QTL experiment of 600 RILs than in a F2 population of 1,640 individuals. A F2 population is known to give a lower QTL precision than a RIL population because heterozygous F2 individuals are not useful to estimate the additive effect of a QTL (De Vienne 1998). However, the chance of observing a common QTL in three different populations is uncertain, so the strategy of gathering RIL populations cannot be recommended in all situations. Furthermore, as the framework map was built on a population of only 199 RILs, the position of the markers delimiting the confidence interval of the QTL in the multiple population experiment cannot be used for an in-silico candidate gene search.

Many genes were identified in the QTL region using BAC sequences and annotations. This was a powerful method to select candidate genes but, in the present situation, this strategy could lack exhaustiveness. Firstly, the presence of gaps in the physical map of *M. truncatula* implies that parts of genome were not used to select candidate genes. Secondly, BAC selection relies on their mapping within the QTL interval, but this mapping is made by the *M. truncatula* consortium on a F2 map calculated from a small population (69 individuals; http://www.medicago/org), so some BACs and even contigs may accidentally not be mapped inside this confidence interval, although they should be. Consequently some of the genome sequences of *M. truncatula* were probably not used to select candidate genes. However, we listed all BACs that spanned the confidence interval of the QTL detected in F6 to perform the fine mapping of this locus. This strategy allowed us to select six candidate genes potentially involved in flowering date control within the relatively small genetic distance of 2.5 cM (*MtCO*, *MtFTa*, *MtFTb* and *MtFTc*, *MtPKS*, *MtFD*). Interestingly, these six genes interact in the plant to induce flowering and they are located very close together in the genome.

Genomic DNA sequencing of these six genes in Jemalong6 and DZA315.16 showed only synonymous mutations for *MtFTa*, *MtFTc*, *MtPKS* and *MtFD*. However, in *MtCO*, the modification of the TATA-box observed in DZA315.16 compared to Jemalong6 may have impacts on the transcription efficiency. Deletions or mutations in the TATA-box do not break the transcription completely; they can induce either a decrease of the transcript rate or the loss of fidelity for transcription site (O'Shea-Greenfield and Smale 1992). In *MtFTb*, a gene with a RKIP domain that was truncated, a mutation was detected in an exon in DZA315.16, leading to an even shorter protein sequence. If the protein of Jemalong6 is possibly active, that of DZA315.16 could have a modified activity.

Transcriptomic study

The objective of this transcriptomic study was to compare, for two parental lines, the expression level of the genes included in the confidence interval of the QTL. Differential expression of a gene would be an indication that this gene could be involved in the phenotypic variation for flowering date (Marone et al. 2001).

Among the six genes listed in the confidence interval as possibly involved in flowering date, we failed to amplify *MtFTa* and *MtFTc* cDNA in plants grown in the growth chamber with a photoperiod of 12 h, despite a large range of PCR conditions tested. Surprisingly, cDNA was obtained and sequenced from samples collected in the greenhouse. The absence of amplification could be due to an absence or a very low rate of expression in the samples.

Similarly, Hecht et al. (2005) failed to recover these sequences by BLASTp on ESTs database. These *FT* genes may not be functional.

For the other four genes, only a differential expression of *MtCO* between Jemalong6 and DZA315.16 was observed in leaves harvested before flowering. In addition, expression during the light period in the day was much higher in Jemalong6 than in DZA315.16. *MtCO* expression could explain the differences in flowering time between these two lines. This result is in agreement with QTL detection which shows that the peak of the QTL was located on markers ABA/CONSTANS-like/MTIC720. Differences between the parental lines in both expression level and DNA sequence supports the hypothesis that *MtCO* is the most probable gene that explains the QTL for flowering date in this cross. *MtCO* shows a strong similarity with *CONSTANS* of *A. thaliana* which is well known for its role of flowering promoter (Hayama and Coupland 2004). This gene is involved in light duration perception, according to the model of "external coincidence" (Imaizumi and Kay 2006); its expression shows a circadian oscillation under control of the circadian clock. Its activity is proved to be effective only under light. Thus, the coincidence of a high expression level of *CONSTANS* and light leads to the induction of genes called "integrators of flowering". It has been shown that the direct target of *CONSTANS* is an FT protein whose transfer from leaves to the shoot apical meristem constitutes the "florigen" signal that induces "floral identity genes".

Based on the dendrogram and the BLASTp search, *MtCO* was close to *COL14* and *COL15* that belong to group III of a *CONSTANS*-like gene family (Griffiths et al. 2003). This classification was further confirmed by the structure of *MtCO* with three introns, which correspond to the structure of group III *COL* genes and was in contrast to group I *COL* genes that have a single intron (Griffiths et al. 2003). This allocation was rather unexpected as active *CONSTANS* or *CONSTANS*-like genes were mainly described as belonging to group I (Hayama and Coupland 2004; Skot et al. 2005). In addition, increasing expression of *CONSTANS* normally affects *FT* expression, which was not observed in this study. However, research was often restricted to group I genes (e.g. in legumes (Putterill et al. 2004b; Hecht et al. 2005)) and results in *A. thaliana* with *COL9* point out the integration in flowering pathway for group III genes (Cheng and Wang 2005). A major role of orthologues of *CONSTANS* in flowering date control has been shown in other species. A comparative mapping study in *A. thaliana* and *Brassica nigra* showed the presence of an orthologue of *CONSTANS* under the peak of a QTL for flowering date in *B. nigra* (Lagercrantz et al. 1996b). Moreover, in order to explain a major QTL, a bulk segregant analysis on three RIL populations of *Avena sativa* was performed and provided evidence for the role of locus *Di1,* an orthologue of *Hd1* in *Oriza sativa* and *CONSTANS* in *A. thaliana* (Locatelli et al. 2006).

It is interesting to note that this QTL on linkage group 7 of *M. truncatula* exhibited partial dominance. Similarly, in *A. thaliana* and *B. nigra*, *CONSTANS* has been identified as one of the few flowering time loci that exhibit partial dominance when the majority of loci influencing flowering time have alleles that are clearly dominant or recessive (Koornneef et al. 1991; Lagercrantz et al. 1996a).

The difference in the genomic sequence of *MtFTb* between the parental lines and its expression in leaves and flowers indicate that this gene is a candidate. As the protein produced in DZA315.16 is shorter than in Jemalong6, it could be less active in promoting the flowering process. However, as the RKIP domain of this protein is not entire even in Jemalong6, the activity of this gene is questionable. No other *FT* gene is described in the *M. truncatula* genome, even if members of related families (*MFT*, *Mother of FT and TFL*, *Terminal Flower*) exist.

MtFD expression was noticed in this experiment. In *A. thaliana*, this gene is not expressed in leaves. This *MtFD* gene was less similar to *AtFD* than a gene present on chromosome 6 of *M. truncatula* (score 6E-5 vs. 2E-14). *MtFD* on chromosome 7 may not be involved in the flowering pathway.

It seems that even if flowering date pathways are similar across plant species (Komeda 2004), the genes involved in genetic variation within species are different. Indeed, *FRIGIDA* and *FLOWERING LOCUS C*, which are respectively involved in the vernalization and the autonomous pathways, seem to play a leading part in *A. thaliana* (Salathia et al. 2006; Li et al. 2006) whereas fine mapping of a QTL explaining 73% of flowering date variability in *Glycine max* enabled the identification of *FT1*, an orthologue of *FT* (Yamanaka et al. 2005). Further studies are needed to test the *MtCO* candidate gene hypothesis. Such investigations could involve RNAi

or complementation of *A. thaliana* mutants with Jemalong6 allele. The next step will be to evaluate the part of flowering date variation covered by the *MtCO* gene in a large range of genetic material, as already described for *FRI* in *A. thaliana* (Le Corre 2005). *M. truncatula* accessions are already available (Ronfort et al. 2006) for an association genetics analysis (Flint-Garcia et al. 2003).

The fine-mapping strategy together with a bio-analysis of sequencing data was efficient in identifying six candidate genes involved in the major QTL for flowering date on chromosome 7 of *M. truncatula*. Sequence polymorphism and differential expression among the two parental lines indicated that a gene belonging to the *CONSTANS*-like family could explain this QTL.

Acknowledgments The authors are very grateful to F. Durand, C. Gibelin and D. Cadier for their help in genotyping recombinants, gene sequencing and RT-PCRs, to J.F. Bourcier and J. Jousse for phenotyping the plants, to R. Minault and F. Gelin for greenhouse management and to M.R. Perretan and G. Boutet at INRA of Clermont-Ferrand for pseudo-F2 genotyping. J.B. Pierre received a Ph.D. grant from Région Poitou-Charentes.

References

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

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Henaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112: 1024–1041

Barker D, Bianchi S, Blondon F, Dattée Y, Duc G, Essad S, Flament P, Gallusci P, Génier G, Guy P, Muel X, Tourneur J, Denarie J, Huguet T (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. Plant Mol Biol Rep 8:40–49

Bell CJ, Dixon RA, Farmer AD, Flores R, Inman J, Gonzales RA, Harrison MJ, Paiva NL, Scott AD, Weller JW, May GD (2001) The Medicago Genome initiative: a model legume database. Nucleic Acids Res 29:114–117

Blazquez MA (2005) The right time and place for making flowers. Science 309:1024–1025

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890

Chabaud M, Larsonneau C, Marmouget C, Huguet T (1996) Transformation of barrel medic (*Medicago truncatula* Gaertn) by *Agrobacterium tumefaciens* and regeneration via somatic embryogenesis of transgenic plants with the MtENOD12 nodulin promoter fused to the gus reporter gene. Plant Cell Rep 15:305–310

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

Cheung W, Hubert N, Landry B (1993) A simple and rapid DNA microextraction method for plant, animal, and insect suitable for RAPD and other PCR analyses. PCR Methods Appl 3:69–70

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101:15289–15294

Cook DR (1999) *Medicago truncatula*—a model in the making!. Curr Opin Plant Biol 2:301–304

Crespi MD, Jurkevitch E, Poiret M, Daubentoncarafa Y, Petrovics G, Kondorosi E, Kondorosi A (1994) Enod40, a gene expressed during nodule organogenesis, codes for a nontranslatable RNA involved in plant-growth. EMBO J 13:5099–5112

de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T (2005) CAR(H)(T)AGene: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics 21:1703–1704

De Vienne D (1998) Les marqueurs moléculaires en génétique et biotechnologies végétales. INRA, Paris

Delalande M, Ronfort J, Prosperi JM (2004) Diversity for flowering time in a large collection of *Medicago truncatula* Gaertn. North American Alfalfa Improvement Conference, Québec

Fankhauser C, Yeh KC, Lagarias JC, Zhang H, Elich TD, Chory J (1999) PKS1, a substrate phosphorylated by phytochrome that modulates light signaling in Arabidopsis. Science 284:1539–1541

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

Gamas P, Debellé F, Berges H, Godiard L, Niebel A, Journet EP, Gouzy J (2006) *Medicago truncatula* cDNA and genomic libraries. *Medicago truncatula* Handbook

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

Hayama R, Coupland G (2004) The molecular basis of diversity in the photoperiodic flowering responses of *Arabidopsis* and rice. Plant Physiol 135:677–684

Hecht V, Foucher F, Ferrandiz C, Macknight R, Navarro C, Morin J, Vardy ME, Ellis N, Beltran JP, Rameau C, Weller JL (2005) Conservation of *Arabidopsis* flowering genes in model legumes. Plant Physiol 137:1420–1434

Huang T, Bohlenius H, Eriksson S, Parcy F, Nilsson O (2005) The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. Science 309:1694–1696

Huguet T, Gherardi M, Chardon F, Sartorel E, Prosperi JM, Chennaoui-Kourda H, Aouani ME (2007) Creation of a consensus genetic-physical (CGPM) for the identification of *Medicago truncatula* genes involved in natural variation. In: Huguet T, Aouani ME (eds), Model legume

congress, 24–28 mars 2007 et Tunis. LILM-CECB, Tunis, pp 12

Imaizumi T, Kay SA (2006) Photoperiodic control of flowering: not only by coincidence. Trends Plant Sci 11:550–558

Julier B, Flajoulot S, Barre P, Cardinet G, Santoni S, Huguet T, Huyghe C (2003) Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. BMC Plant Biol 3:9

Julier B, Huguet T, Chardon F, Ayadi R, Pierre JB, Prosperi JM, Barre P, Huyghe C (2007) Identification of quantitative trait loci influencing aerial morphogenesis in the model legume *Medicago truncatula*. Theor Appl Genet 114:1391–1406

Komeda Y (2004) Genetic regulation of time to flower in *Arabidopsis thaliana*. Annu Rev Plant Biol 55:521–535

Koornneef M, Hanhart CJ, Vanderveen JH (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. Mol Gen Genet 229:57–66

Lagercrantz U, Putterill J, Coupland G, Lydiate D (1996a) Comparative mapping in Arabidopsis and Brassica, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9:13–20

Lagercrantz U, Putterill J, Coupland G, Lydiate D (1996b) Comparative mapping in Arabidopsis and Brassica, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9:13–20

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

Le Corre V (2005) Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. Mol Ecol 14:4181–4192

Li Y, Roycewicz P, Smith E, Borevitz JO (2006) Genetics of local adaptation in the laboratory: flowering time quantitative trait loci under geographic and seasonal conditions in Arabidopsis. PLOS One 1:e105

Lifschitz E, Eviatar T, Rozman A, Shalit A, Goldshmidt A, Amsellem Z, Alvarez JP, Eshed Y (2006) The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. Proc Natl Acad Sci USA 103:6398–6403

Locatelli AB, Federizzi LC, Milach SCK, Wight CP, Molnar SJ, Chapados JT, Tinker NA (2006) Loci affecting flowering time in oat under short-day conditions. Genome 49:1528–1538

Manichaikul A, Dupuis J, Sen S, Broman KW (2006) Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics 174:481–489

Marone M, Mozzetti S, De Ritis D, Pierelli L, Scambia G (2001) Semiquantitative RT-PCR analysis to assess the expression levels of multiple transcripts from the same sample. Biol Proced Online 3:19–25

Moreau D, Salon C, Munier-Jolain N (2006) Using a standard framework for the phenotypic analysis of *Medicago truncatula*: an effective method for characterizing the plant material used for functional genomics approaches. Plant Cell Environ 29:1087–1098

Nam YW, Penmetsa RV, Endre G, Uribe P, Kim D, Cook DR (1999) Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene-response genes. Theor Appl Genet 98:638–646

O'Shea-Greenfield A, Smale ST (1992) Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. J Biol Chem 267:1391–1402

Park SY, Nam YW (2006) Construction of a bacterial artificial chromosome library containing large BamHI genomic fragments from *Medicago truncatula* and identification of clones linked to hypernodulating genes. J Microbiol Biotechnol 16:256–263

Pierre JB, Huguet T, Barre P, Huyghe C, Julier B (2008) Detection of QTLs for flowering date in three mapping populations of the model legume species *Medicago truncatula*. Theor Appl Genet 117:609–620

Putterill J, Laurie R, Macknight R (2004a) It's time to flower: the genetic control of flowering time. Bioessays 26:363–373

Putterill J, Laurie R, Macknight R (2004b) It's time to flower: the genetic control of flowering time. Bioessays 26:363–373

Ronfort J, Bataillon T, Santoni S, Delalande M, David JL, Prosperi JM (2006) Microsatellite diversity and broad scale geographic structure in a model legume: building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. BMC Plant Biol 6:28

Salathia N, Davis SJ, Lynn JR, Michaels SD, Amasino RM, Millar AJ (2006) FLOWERING LOCUS C-dependent and -independent regulation of the circadian clock by the autonomous and vernalization pathways. BMC Plant Biol 6:10

Skot L, Humphreys MO, Armstead I, Heywood S, Skot KP, Sanderson R, Thomas ID, Chorlton KH, Hamilton NRS (2005) An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Mol Breed 15:233–245

Tadege M, Ratet P, Mysore K (2005) Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*. Trends Plant Sci 10:229–235

Tjaden G, Edwards JW, Coruzzi GM (1995) Cis-elements and trans-acting factors affecting regulation of a nonphotosynthetic light-regulated gene for chloroplast glutamine-synthetase. Plant Physiol 108:1109–1117

Trinh TH, Ratet P, Kondorosi E, Durand P, Kamate K, Bauer P, Kondorosi A (1998) Rapid and efficient transformation of diploid *Medicago truncatula* and *Medicago sativa* ssp *falcata* lines improved in somatic embryogenesis. Plant Cell Rep 17:345–355

Yamanaka N, Watanabe S, Toda K, Hayashi M, Fuchigami H, Takahashi R, Harada K (2005) Fine mapping of the FT1 locus for soybean flowering time using a residual heterozygous line derived from a recombinant inbred line. Theor Appl Genet 110:634–639

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005) Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiol 137:1174–1181