Targeted transcript mapping for agronomic traits in potato


Wageningen University and Research Centrum, Department of Plant Sciences, Laboratory of Plant Breeding, PO Box 386, 6700 AJ Wageningen, The Netherlands

Received 5 January 2007; Revised 21 April 2007; Accepted 29 May 2007

Abstract

A combination of cDNA-amplified fragment length polymorphism (AFLP) and bulked segregant analysis (BSA) was used to identify genes co-segregating with earliness of tuberization in a diploid potato population. This approach identified 37 transcript-derived fragments with a polymorphic segregation pattern between early and late tuberizing bulks. Most of the identified transcripts mapped to chromosomes 5 (19 markers) and 12 (eight markers) of the paternal map. Quantitative trait locus (QTL) mapping of tuberization time also identified earliness QTLs on these two chromosomes. A potato bacterial artificial chromosome (BAC) library was screened with four of the markers linked to the main QTL. BAC contigs containing the markers showing the highest association with the trait have been identified. One of these contigs has been anchored to chromosome 5 on an ultradense genetic map of potato, which could be used as a starting point for map-based cloning of genes associated with earliness.

Key words: Bulked segregant analysis, cDNA-AFLP, potato tuberization, QTL, transcript mapping.

Introduction

Tuber formation in potato (Solanum tuberosum) represents an important developmental switch in which an underground stem, the stolon, differentiates into a storage and propagation organ, the tuber. Tuberization involves a sequence of stages including tuber induction, initiation, growth, and maturation. Induction is defined as the acquisition of the competence to form tubers (Ewing and Struik, 1992). A signal is produced in the leaves and transmitted through the phloem to the stolon tips, where initiation takes place (Gregory, 1956). This stimulus is graft transmissible and thought to be analogous to the flowering signal (Chailakyan et al., 1981).

Tuber induction is controlled by both endogenous and external cues, which ensure that the time of tuber formation coincides with favourable developmental and environmental conditions (Ewing and Struik, 1992). Several compounds, including gibberellic acid (GA), cytokinin (CK), jasmonic and tuberonic acids, auxin, abscisic acid, and sucrose, have been reported to participate in the regulation of tuber formation (Jackson, 1999; Struik et al., 1999; Fernie and Willmitzer, 2001; Rodríguez-Falcón et al., 2006). Among these, however, only GAs provide enough experimental evidence to support their definitive role in the control of induction. Tuberization is suppressed by the addition of exogenous GA, while it is enhanced by the reduction (chemical or genetic) of endogenous active GA (Menzel, 1980; Balamini and Poovaiah, 1985; Jackson and Prat, 1996; Carrera et al., 2000). A GA-deficient mutant, andigena ga 1, can form tubers in long days, a condition that inhibits tuber formation in the wild type (Van den Berg et al., 1995). Furthermore, a drop in GA level has been observed in the stolon tips upon transfer to inducing conditions (Xu et al., 1998; Carrera et al., 1999).

Photoperiod is the major environmental factor controlling the induction of tubers. Induction is generally
promoted by short days, but there is a large genetic variation among different cultivars in the strength of the response. Cultivars with long critical photoperiods are classified as early, while those with short critical photoperiods are classified as late. PHYTOCHROME B (PHYB), CONSTANS (CO), and, more recently, FLOWERING LOCUS T (FT) have been identified as three of the components of the photoperiodic control of tuber induction (Jackson et al., 1996, 1998, 2000; Martínez-García et al., 2002; Rodríguez-Falcón et al., 2006). Additionally, several other genes associated with the tuberization process have been identified, mostly following one of two approaches: testing of candidate genes or analysis of differential gene expression on a single genotype. These include genes involved in the regulation of GA metabolism or signal transduction, such as GA 20-oxidase, SgAN, PHORI, POTH1, StBEL5, and GA 2-oxidase (Carrera et al., 2000; Amador et al., 2001; Bachem et al., 2001; Rosin et al., 2003; Chen et al., 2003, 2004a; Kloosterman et al., 2005), lipoygenases (Bachem et al., 1996; Kolomiets et al., 2001), and calcium-dependent protein kinases (Raíces et al., 2001, 2003a, b). While these studies have contributed towards an insight into the molecular basis of tuberization, a causal relationship between the identified genes and tuberization is not always clear, and there is no evidence either that they are responsible for the observed genetic variation in tuberization time.

Quantitative trait locus (QTL) mapping represents an alternative approach in the identification of genes responsible for the naturally occurring allelic variation in complex traits. Several QTLs affecting the ability to form tubers under long photoperiods (earliness) have been identified (van den Berg et al., 1996; Simko et al., 1999). However, the genes responsible for those QTLs remain unknown. Map-based cloning is still the most common approach for the identification of the specific gene accounting for a QTL. Saturation of the QTL region with DNA markers tightly linked to the target gene becomes an essential step prior to positional cloning. In recent years, the benefits of using markers representing the transcribed region of the genome, as opposed to anonymous genomic DNA markers, have been emphasized (Gupta and Rustgi, 2004). The combination of cDNA-amplified fragment length polymorphism (AFLP) (Bachem et al., 1996) and bulked segregant analysis (BSA) (Michelmore et al., 1991) is a particularly suitable method for saturation of QTL regions with markers that represent expressed genes. It was recently shown that polymorphic cDNA-AFLP fragments detected in a segregating population can be directly used as genetic markers in the construction of a linkage map (Brugmans et al., 2002). In combination with BSA, the cDNA-AFLP analysis of mapping populations should lead to the identification of those polymorphic genes linked to the trait of interest, and it could potentially result in the direct identification of candidate genes.

In this report, the use of cDNA-AFLP and BSA to map polymorphic transcripts between early and late tuberizing progeny pools, as a first step towards the identification of the genes controlling this complex trait, is described. The association between the expressed genes and the trait was tested by QTL mapping. Four of the transcripts mapping on the main QTL have been used to identify bacterial artificial chromosomes (BACs) carrying these genes, which could be used as a starting point for map-based cloning.

### Materials and methods

#### Plant material

The genetic map and quantitative data used in this study are based on a diploid mapping population of 250 F1 genotypes descending from the non-inbred parents C (Hanneman and Peloquin, 1967) and E (Jacobsen, 1980). Population details can be found in van Eck et al. (1995).

For cDNA-AFLP analysis, a subset of 92 offspring and the two parents were used. Single-node cuttings from in vitro plants were grown in 3% sucrose MS medium (Murashige and Skoog, 1962) at 24 °C and 16 h light. After 4 weeks, the shoots were harvested, frozen in liquid N2, and stored at −80 °C. Additionally, five tubers of each genotype were grown in 5.0 l pots in the greenhouse between May and August 2003. Plants were distributed in five blocks and grown under natural photoperiod conditions (varying from 15.87 h on 16 May to 14.92 h on 13 August). The first three fully expanded leaves of each plant within a block were harvested at one of five evaluation dates [31, 41, 53, 69, and 85 days after planting (DAP)] immediately frozen in liquid N2, and stored at −80 °C.

#### Assessment of time of tuberization (earliness)

Data on tuberization time under field conditions were collected in an experiment carried out in Wageningen (The Netherlands) between the months of May and October 1999 (Celis-Gamboa et al., 2003). Two plants per genotype were harvested on each of 11 evaluation dates (29, 36, 42, 57, 64, 75, 89, 96, 110, 125, and 140 DAP), and earliness was scored as DAP until the appearance of the first swelling stolon tips.

Assessment of time of tuberization in the greenhouse was performed in an analogous way. The collected data are based on five evaluations (31, 41, 53, 69, and 85 DAP) and one plant per evaluation date. The greenhouse trial was restricted to the subset of the population used for cDNA-AFLP genotyping.

#### Bulked segregant analysis

Selection of in vitro bulks was based on the field phenotypic data. Ten individuals from each of the two earliest and two latest tuberizing categories were chosen to construct the bulks very early (VEi; 29 DAP), early (Ei; 36 DAP), late (Li; 64 DAP), and very late (VLi; 75 DAP).

For the greenhouse material, two groups of 10 individuals each were compared in order to identify genes linked to earliness. The first group, early (Ei), was composed of genotypes with a greenhouse tuberization time equal to that of the early parent (41 DAP); the second group, late (Li), consisted of genotypes with a trait value equal to that of the late parent (53 DAP). From each group, two different cDNA pools (Ei/l and 3, and Li/l and 3) were made corresponding to the first (31 DAP, non-tuberizing plants) and third harvest (53 DAP, tuberizing plants), respectively. A third group of 10 genotypes, with a tuberization time greater than that of the late
parent (very late, VL4) and that, therefore, had not yet started tuberizing on harvest 3, was used to eliminate transcripts that could be harvest specific rather than tuberization related.

Samples for BSA were prepared by pooling equal amounts of pre-amplification product from every individual in the bulk. All possible 256 combinations of primers with two selective nucleotides were used for analysis. Polymorphic transcript-derived fragments (TDFs) were visually identified, and their segregation was confirmed on the individual lines of each pool.

cDNA-AFLP analysis

cDNA-AFLP template was prepared according to Bucher et al. (1998) using Asel and TaqI restriction enzymes (New England Biolabs, Inc., USA) and two base extensions for selective amplification. The enzymes used for cDNA synthesis were obtained from Invitrogen BV (Breda, The Netherlands). Asel selective primers were fluorescently labelled (IRDye™ 700 and 800) for visualization. Oligonucleotides were obtained from Biologie BV (Malden, The Netherlands). PCR amplifications were carried out in a PE-9600 thermocycler (Perkin Elmer Applied Biosystems, Inc., Foster City, CA, USA) using SuperTaq polymerase (Sphaero Q, Leiden, The Netherlands). Amplification products were separated on an NEN® Global Edition IR® DNA Analyzer (LI-COR® Biosciences, Lincoln, NE, USA). A detailed description of the cDNA-AFLP protocol can be found at http://www.dpw.wau.nl/pv/aflp/protocol.pdf

TDF isolation and sequence analysis

For the isolation of candidate TDFs, a positive and a negative sample for each fragment (i.e. one containing the band of interest and one lacking it) were loaded next to each other on a new gel; gels were scanned with an Odyssey IR® Infrared Imaging System (LI-COR® Biosciences) and the bands of interest (or the corresponding gel regions for the negative samples) were excised from the gel and re-amplified using the same conditions as for pre-amplification (Brumans et al., 2003). The PCR product was cloned into the pGEM-T Easy vector (Promega, Madison, CA, USA) and transformed in Escherichia coli DH5α competent cells (Invitrogen BV). Cloned fragments were directly amplified from individual colonies using the T7 and SP6 promoter primers. The PCR products were sequenced using the T7 promoter primer (Baseclear BV, Leiden, The Netherlands) after purification with Microspin™ G-50 Columns (Amersham Biosciences, Piscataway, NJ, USA). For each TDF, four individual clones of a positive and a negative sample (which produced a PCR product) were sequenced, and only when the four sequences from the positive sample were the same and different from any sequence obtained from the negative sample were they included for further analysis.

Sequences were analysed for similarity to the expressed sequence tag (EST) and the non-redundant protein databases at NCBI (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov) and the Unigene database at SGN (Sol Genomics Network, http://www.sgn.cornell.edu) using the BLASTN and BLASTX 2.2.9 programs (Altschul et al., 1997), and to the plant Unique Gene Indices at TIGR (The Institute for Genomic Research, http://www.tigr.org) using BLASTN 2.0MP-WashU (Gish 1996–2000, http://blast.wustl.edu).

Mapping of candidate TDFs and QTL analysis

Selected TDFs were scored as dominant genetic markers. Ambiguous bands were recorded as missing values. The segregation data for the TDFs were added to the available genomic mapping data (Van Eck et al., 1995; Celis-Gamboa, 2002), and used for the construction of separate maternal and paternal maps according to the double pseudo-testcross approach for cross-pollinating populations (Garrattaglaga and Sederoff, 1994). Linkage analysis and map construction were performed using the mapping software JoinMap® 3.0 (van Ooijen and Voorrips, 2001; Plant Research International, Wageningen, The Netherlands). Markers were assigned to linkage groups at a minimum LOD value of 5.0. The calculations of the linkage maps were done using all pairwise recombination estimates <0.4 and LOD score >1 using the Kosambi’s mapping function.

QTL analysis was performed using the Kruskal–Wallis, interval mapping (IM), and multiple QTL mapping (MQM) methods implemented in MapQTL® 5.0 (van Ooijen, 2004; Kyazma BV, Wageningen, The Netherlands). Two-way interactions (P < 0.005) were analysed using the General Linear Model module of the statistical package SPSS version 14.0.1 (SPSS Inc., Chicago, IL, USA) using the corresponding markers as fixed factors. Chromosomal maps and QTL graphs were drawn with the graphical package MapChart (Voorrips, 2002).

BAC library screening

Four TDFs (CE4, 6, 16, and 35) located in the identified QTL region on chromosome 5 were used to screen a 10 genome equivalent HindIII/EcoRI BAC library from the diploid clone RBH99-039-16. This library is currently being used for the construction of a potato physical map and consists of 73 000 clones (average insert size >120 kb) assembled into ~7200 contigs and 7600 singletons.

Ninety-six BAC plate pools (1 plate pool=384 BACs), covering half of the library, were screened for the presence of any of the four TDFs by means of PCR amplification. After identification of positive plate pools, a PCR for the marker was carried out on quarter plate pools (1 quarter plate pool=96 BACs), column and row pools (eight and 12 BACs, respectively), and single BACs following a stepwise procedure. Plasmid DNA was extracted using a standard alkaline lysis protocol. PCR amplifications were performed in a total volume of 20 μl containing 10 ng of template DNA, 0.125 mM of each dNTP (Invitrogen BV), 4 pmol of each primer (Biologie BV, Malden, The Netherlands), and 0.5 U of SuperTaq polymerase plus 1× PCR buffer (Sphaero Q). PCRs were carried out in a PE-9600 thermocycler (Perkin Elmer Applied Biosystems, Inc., Foster City, CA, USA) using the following conditions: 3 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at 60 °C (56 °C for CE16), and 50 s at 72 °C, and a final extension of 3 min at 72 °C. Primer sequences were as follows: CE4-forward 5′-TGTTGAAAATATTCACCTTGGA-GC-3′, CE4-reverse 5′-GACAAGAATGGATGTAGCTGAG-3′; CE6-forward 5′-CTTTGTGC TTACTCGTGCTT-3′, CE6-reverse 5′-ACAGAGCTTGCCCTTG- GAAGT-3′; CE16-forward 5′-CAGTTACGTCTCCAAAAA TGG-3′, CE16-reverse 5′-AAGGAAAGAATGGCGAC-3′; and CE35-forward 5′-ACCACCAAGAAAACCAACAA-3′, CE35-reverse 5′-CAATCCCCAGAAAAATCCAC-3′.

Results

Phenotypic analysis

The data for tuberization time (earliness) in the field used in this study were obtained in an experiment carried out in Wageningen (The Netherlands) from 12 May to 20 September 1999 (Celis-Gamboa et al., 2003). Trait values for the parental clones were 36 DAP for the early parent (E) and 57 DAP for the late parent (C). The offspring showed a larger variation in earliness, with values ranging
from 29 to 75 DAP. The distribution of the population deviated significantly from normality (Shapiro–Wilk test, \( P < 0.001 \)), showing positive skewness (Fig. 1A). A large group of the progeny (44\%) had a value for earliness equal to that of \( E \), with the latter genotypes showing a wider distribution and transgressive segregation. These results suggested the presence of a major dominant gene for earliness segregating from the early parent, and several minor factors, with positive as well as negative alleles, segregating from either parent.

Assessment of tuberization time in the greenhouse was restricted to the subset of the population used for cDNA-AFLP fingerprinting. Tubers were planted on 16 May, thus, under the same photoperiod conditions as in the field trial. The earliest tuberization was detected 41 DAP. At this time, 50\% of the offspring and the early parent had started tuber formation. Swelling in the maternal clone was first observed 53 DAP, while the latest genotypes had not yet produced tubers at the end of the experimental period (Fig. 1B). Day temperatures above 35 \( ^\circ C \) were recorded during the last 4 weeks (59–85 DAP) for most of the day (>8 h) with maximums exceeding 40 \( ^\circ C \). High temperatures are known to have a negative effect on tuberization (Ewing and Struik, 1992) and are therefore likely to have caused a delay in tuber induction of the late genotypes. Since both experimental conditions and evaluation dates were different between the two trials, direct comparison of the two data sets is not possible; however, correlation between the two experiments was significant \(( r = 0.7, P < 0.001 \)).

Identification of tuberization time-related genes

In order to identify transcripts linked to a QTL by BSA, it is important to minimize all sources of variation unrelated to the trait that could lead to an increase in false positives or mask true positives. An in vitro system satisfies this criterion, and, as such, it can be regarded as an appropriate source of expressed genes directly linked to the target trait. This system would, however, fail to detect transcripts or changes in transcription that occur specifically at the time of tuber induction. We therefore decided to combine two different experimental configurations. First, in vitro plantlets grown under controlled conditions were used as the source of mRNA to test the efficacy of using cDNA-AFLP and BSA in the identification of transcripts linked to a quantitative trait. Secondly, an in vivo experiment was designed that could potentially identify changes in expression occurring at the point of induction, which would have passed undetected in the in vitro analysis.

BSA of in vitro material

Ten individuals from each of the two earliest and two latest tuberizing categories were selected to construct the bulks ‘very early’ (\( VE_{iv} \); 29 DAP), ‘early’ (\( E_{iv} \); 36 DAP), ‘late’ (\( L_{iv} \); 64 DAP), and ‘very late’ (\( VL_{iv} \); 75 DAP). The pooled samples were fingerprinted with all 256 possible combinations of primers carrying two selective nucleotides. This resulted in >10 000 TDFs being analysed, with an average of 40 bands per primer combination. After visual inspection of the gels, 29 candidate TDFs, generated by 22 primer combinations, were selected for further analysis. The chosen transcripts could be classified in two categories. The first class (Fig. 2A), containing 17 TDFs, consisted of markers present in both bulks of one phenotype (\( VE_{iv} / E_{iv} \) versus \( VL_{iv} / L_{iv} \) and either absent or present as a significantly weaker band in the other two. Fourteen of these transcripts were associated with the \( VE_{iv} \) and \( E_{iv} \) pools, and only three were predominantly detected in the \( L_{iv} \) and \( VL_{iv} \) bulks. In addition, 12 TDFs that appeared as differentials in only one bulk were also selected (Fig. 2B, C). Eleven of these TDFs were present with the
same intensity in three of the pools and absent, or present as a faint band, in the fourth (nine in the VLiv, one in the VEiv, and one in the Eiv bulk). When looking at the individual progeny, however, two of these markers (CE18 and 19) showed a segregation pattern corresponding to that of the first group (see below). The remaining TDF (CE29) was selected because of its higher intensity in the VLiv bulk.

All the 29 candidates were subsequently examined in the individual lines of each pool along with the parents. Differences in band intensity between two pools can be the result of differences in allelic frequencies, or they may reveal variation in the expression level of the corresponding gene. Only one of the candidate TDFs, CE4, showed such expression polymorphism (Fig. 3). The PCR product could be detected in all individuals, including both parental clones, but showed a much higher intensity among those in the VEiv and Eiv, as well as the early parent. Overall, there was a good correlation between the BSA results and the observed segregation in the individuals comprising the pools. Exceptions were the above-mentioned CE18 and 19, which showed low frequency among the VEiv, Eiv, and VLiv; very early, early, late, and very late in vitro bulks. Ei, Li, and Vi: early-harvest 1, early-harvest 3, late-harvest 1, and late-harvest 3 greenhouse bulks. (A) TDFs appearing as differentials between early and late phenotypes (VEiv/Eiv versus VLiv/Liv). (B, D) TDFs with a segregation pattern different from that obtained in the BSA. (C) TDFs appearing as differentials in only one bulk. (E) Differential TDFs identified in the greenhouse bulks and not detected in the in vitro bulks.

Fig. 2. Candidate transcript-derived fragments (TDFs) identified by bulked segregant analysis (BSA). The name in parentheses indicates the primer combination giving rise to the TDF; the lower case letter at the end specifies the parental origin of the band; ‘c’ denotes a marker segregating from C; ‘h’ denotes a marker segregating from both parents. VEiv, Eiv, Liv, and VLiv: very early, early, late, and very late in vitro bulks. Ei, Li, and Vi: early-harvest 1, early-harvest 3, late-harvest 1, and late-harvest 3 greenhouse bulks. (A) TDFs appearing as differentials between early and late phenotypes (VEiv/Eiv versus VLiv/Liv). (B, D) TDFs with a segregation pattern different from that obtained in the BSA. (C) TDFs appearing as differentials in only one bulk. (E) Differential TDFs identified in the greenhouse bulks and not detected in the in vitro bulks.
BSA of in vivo material

To identify genes differentially expressed upon tuberization and associated with earliness, four pools of samples (Eg1, Eg3, Lg1, and Lg3) representing two phenotypic groups, early and late, and two treatments, non-tuberizing and tuberizing plants, were compared by cDNA-AFLP. Because the very late genotypes had not yet produced tubers at the end of the experimental period, the Lg bulk does not include the most extreme individuals but individuals with a tuberization time equal to that of the late parent. The Lg group is therefore different from the VLiv and Liv groups. A single group of early genotypes (made up of individuals included in the VEiv and Eiv pools) was also used instead of two since no major differences had been observed between the VEiv and Eiv bulks.

After cDNA-AFLP fingerprinting with all 256 primer combinations, 81 candidate TDFs showing polymorphism between the early and late bulks were selected for further analysis. These included 15 of the transcripts identified with the in vitro bulks. Five other in vitro TDFs could not be detected in the greenhouse cDNA pools (CE1, 7, 13, 14, and 15), whereas one transcript (CE18), although present, did not come out as differential. The remaining in vitro TDFs corresponded to markers CE20–29 (Fig. 2C), which, as expected, did not show any polymorphism between the greenhouse pools. Only six (Fig. 2E) of the 66 new candidate TDFs, however, appeared to be linked to earliness when tested on the individual progeny of the bulks. All six markers corresponded to alleles segregating from the male parent. None of the candidate TDFs showed changes in expression level associated with tuberization.

Over 200 bands were also identified that, although monomorphic between the early and late bulks, did show polymorphism between harvests (i.e. Eg1/Lg1 versus Eg3/Lg3). These transcripts could represent genes differentially regulated during tuberization and development, or they could reflect other environment-associated differences. In order to filter out tuberization-related from environment-specific gene expression, 10 additional genotypes (very late, VLg) with a tuberization time greater than that of the late parent, and therefore not tuberizing at harvest 3, were used to construct two new cDNA pools, VLg1 and VLg3. Comparison of their fingerprints with those obtained for Eg1, Eg3, Lg1, and Lg3 revealed 15 bands putatively associated with tuber formation. Analysis of the individual lines of each pool and the two parents, however, did not allow a clear association between an expression pattern and tuberization, and the TDFs were disregarded for further analysis.

Genetic mapping of candidate TDFs

In order to determine the chromosome position of the selected TDFs, segregation analysis was extended to a total of 92 individuals. In vitro plantlets and leaves from harvest 3 of the greenhouse experiment were used as the source of mRNA. cDNA-AFLP fragments were scored as dominant genetic markers and used directly for mapping. In the case of CE4, a clear cut-off could be established.
between high and low intensity bands, which allowed their classification into two genotypic classes: a band with high intensity was scored as present and a band with low intensity as absent (Fig. 3).

Of the 27 markers included in the male map, 18 were assigned to chromosome 5 and eight to chromosome 12 (Fig. 4). Only CE9, a heterozygous marker, fell outside these two linkage groups, being placed on chromosome 1. Markers on chromosome 5 clustered in two groups 16 cM apart from each other. The first group, consisting of 11 TDFs, spans a region of 7 cM, while the second group, with seven loci, spreads over a 4 cM window; the remaining marker on chromosome 5, CE15, maps between these two groups at a distance of 10 cM and 6 cM respectively. Several of these markers showed complete linkage in the analysed individuals (CE33–CE4–CE5, CE6–CE34–CE35–CE36, CE2–CE12–CE11, and CE14–CE10). As expected, markers predominantly detected in the late phenotype bulks (CE15, CE16, CE36, and CE37), were in repulsion to the other TDFs mapped on E5. Most of the TDFs on chromosome 12 (seventy out of eight) stretch over a 30 cM window on the long arm of this chromosome. No recombinant between CE30, CE31, and the genomic AFLP marker E39M60-29 was detected in the genotyped offspring.

Markers originating from the female parent were scattered over five different chromosomes. Three linked TDFs were located on chromosome 1 within a 5 cM distance (CE13, CE20, and CE25), although approximately 29 cM apart from the mentioned heterozygous marker (CE9) on the same chromosome. A second set of three TDFs was positioned on linkage group 11; two of these markers, CE23 and 29, showed complete linkage, while the third, CE22, mapped at a 4 cM distance. Of the remaining four TDFs, two of them mapped on chromosome 2 (CE21 and CE24) at a distance of approximately 14 cM, whereas the other two markers were positioned on chromosomes 8 (CE28) and 12 (CE26).

QTL mapping

The effect of the individual TDFs on earliness was examined using the Kruskal–Wallis test. Significant associations ($P < 0.005$) were detected in this way only for markers on chromosome 5 of the male map. The strongest effect was associated with CE6. Individuals carrying this allele tuberized on average 21 d earlier in the field and 13 d earlier in the greenhouse than those without it, which is close to the observed differences between the two parents (21 d and 10 d, respectively).

For IM and MQM, a natural log transformation was applied to the raw data in order to improve normality. IM identified one QTL above the significance threshold (LOD 2.6, $P < 0.05$, 1000 permutations) on chromosome E5. By applying MQM, a second QTL peak (LOD 2.6) just reaching the significance threshold was detected on chromosome E12 for field earliness (Fig. 4). The main QTL on chromosome 5 had the highest LOD at CE6 when field tuberization data were used and at CE16 for greenhouse tuberization time. This QTL maps at the same

---

**Fig. 4.** Chromosomal location of transcript-derived fragments (TDFs) and QTL likelihood maps for earliness (obtained by multiple QTL mapping) on linkage groups E5 and E12. The vertical dashed line indicates the LOD score threshold of 2.6. The linkage group number is indicated on the top right corner of each map. f, field tuberization time; gh, greenhouse tuberization time.
position as previously described QTLs in this population for onset and end of senescence (Celis-Gamboa, 2002; Malosetti et al., 2006), foliage maturity type, and late blight resistance (Visker et al., 2003). The second QTL on chromosome 12 (LOD 2.6 at the co-segregating markers CE30, CE31, and E39M60-29) had a smaller effect and mapped at the same position as QTLs for onset and midpoint of senescence described earlier (Malosetti et al., 2006). This second QTL was not significant for greenhouse tuberization time.

Epistasis between the QTLs and the remaining TDFs was analysed by two-way analysis of variance (ANOVA). One significant interaction (P <0.005) was detected between the QTLs on chromosome E5 and E12 for field tuberization time. The total variance explained by these two chromosomal regions accounts for 70% as estimated by ANOVA. The interaction term was not significant when greenhouse tuberization time was used as the dependent variable.

Sequence analysis of candidate genes

All candidate TDFs were excised from the gels, reamplified, and sequenced. DNA sequences were obtained for 29 fragments. A comparison of the obtained sequences (using BLASTX) with those in the non-redundant protein database of GenBank resulted in significant similarities (e-value <10^-4) being found for 16 of the transcripts. A search against the EST and the Unigene databases of TIGR and SGN, however, generated positive identifications for all the remaining TDFs, confirming that the isolated bands correspond to expressed genes. The Unigene or EST (if not included in a contig) with the highest similarity to each of these TDFs was then used to search the databases for a second time. The results of the BLAST search are summarized in Table 1.

TDFs CE23 and 29, that show complete linkage in the analysed subset of the population, differ only by an indel of 27 bp. Likewise, CE36 differs from CE35 by an insertion of 9 bp, and no recombinant was found among the offspring. These two pairs of TDFs are likely to represent two allelic forms of the same gene, or alternatively two copies of the same gene tightly linked.

BAC library screening

As a first step in the construction of a physical map covering the earliness region of chromosome 5, gene-specific primers for six of the mapped TDFs (CE4, CE5, CE6, CE16, CE35, and CE37) were designed to screen a potato BAC library by PCR amplification. Two of these markers (CE5 and CE37) did not render a good amplification product when tested on genomic DNA from the clone used to construct the library and were discarded from further analysis. In an on-going project aimed at the construction of a potato physical map, AFLP patterns of all BAC clones have been generated, and the fingerprints used to assemble BACs into contigs on the basis of overlapping patterns.

The screening of the library identified three candidate contigs for CE16, two for CE4 and CE35, and one for CE6, with CE6 and CE35 sharing one contig in common (Table 2).

Recent progress in the physical mapping of potato has confirmed the genetic mapping location of CE16. One of the contigs identified by this TDF has in fact been anchored to chromosome 5 of the ultradense genetic linkage map of potato (http://potatodbase.dpw.wau.nl/UHData.html), constructed on an F1 mapping population of 136 individuals and comprising >10 000 markers (van Os et al., 2006). The position of CE16 on the ultradense map (Fig. 5) is one recombination event from the restriction fragment length polymorphism marker GP21, previously associated with earliness (Collins et al., 1999; Visker et al., 2003).

Discussion

The aim of this work was to identify possible candidate expressed genes linked to an important agronomic trait, earliness of tuberization in potato. We have used a combination of QTL mapping, transcript analysis, and BSA, followed by direct mapping of the selected transcripts, to generate a series of possible candidates that are linked to earliness in potato. The putative functions of these candidates in connection with the trait are discussed.

Since swelling of the stolon tip is the earliest visible sign and most direct indicator of tuber induction, assessment of earliness was carried out by monitoring the time of appearance of the first swelling stolon tip. The destructive character of this method implied that a new set of plants had to be monitored for each time point, limiting the number of possible observations. The occurrence of either swellings or tubers in all plants (of a given genotype) at all evaluation dates subsequent to that in which swollen stolons were first observed, however, supports the reliability of the collected data. The consistency of the field data was further confirmed by the results of a second trial including a reduced set of early and late extreme genotypes (data not shown). Likewise, the significant correlation found between field and greenhouse earliness corroborates the reliability of the evaluation method and implies that a large fraction of the observed variance can be explained by genetic factors.

The central aim of this work was to identify expressed genes linked to the QTL regions underlying earliness in potato. The adopted strategy, combining cDNA-AFLP with BSA, proved effective, and 26 TDFs mapping to linkage groups 5 and 12, both harbouring earliness QTLs, were identified. An in silico analysis of the TCs (tentative
consensus sequences) collected in the Potato Gene Index database at TIGR (release 7.0) indicates that around 20% of all potato cDNAs would render scorable cDNA-AFLP fragments (between 50 bp and 500 bp in length) after restriction with the enzyme combination AseI/TaqI, and that, on average, those cDNAs would produce 1.5 cDNA-AFLP fragments (unpublished data). Based on 40–50 bands per primer combination and a rate of polymorphism of 10 polymorphic markers per primer pair, with around 6% of those representing either allelic variants of the same gene or two fragments of the same allele, it is estimated that the total number of genes surveyed by our BSA is in the order of 7000, of which ~2400 are expected to segregate in the population as sequence polymorphisms. While this should be sufficient to provide a satisfactory number of markers to start BAC library screening and chromosome walking, we acknowledge that the probability of landing directly on the gene responsible for the QTL is limited.

Almost all the identified earliness-associated TDFs, using both in vitro plantlets and greenhouse material, segregated as presence/absence polymorphisms in the population. Only one transcript, CE4, segregated as a clear expression polymorphism in both growing conditions. Although these presence/absence polymorphisms can also represent differences in the expression level of the gene, our previous experience suggests that the majority of them result from either single nucleotide polymorphisms in the restriction site or selective nucleotides or indels in the amplified region (Brugmans et al., 2002), and that their map position therefore corresponds to the physical position of the gene. Differences in band intensity were, as expected, much more frequent (involving ~30% of the

Table 1. Accession numbers, e-values, and annotated function of genes with the highest similarity to the sequenced TDFs

<table>
<thead>
<tr>
<th>TDF and GenBank accession no.</th>
<th>BLASTN&lt;sup&gt;a&lt;/sup&gt;</th>
<th>BLASTX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unigene/EST number</td>
<td>e-value</td>
</tr>
<tr>
<td>CE1 (C-E5) EL737965</td>
<td>TC160981</td>
<td>1e-31</td>
</tr>
<tr>
<td>CE2 (E5) EL737966</td>
<td>DNS89121</td>
<td>4e-05</td>
</tr>
<tr>
<td>CE3 (E5) EL737967</td>
<td>TC153785</td>
<td>9e-16</td>
</tr>
<tr>
<td>CE4 (E5) EL737968</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE5 (E5) EL737969</td>
<td>CV302561</td>
<td>5e-11</td>
</tr>
<tr>
<td>CE6 (E5) EL737970</td>
<td>SGN-U281605</td>
<td>3e-53</td>
</tr>
<tr>
<td>CE8 (E12) EL737971</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE9 (C-E1) EL737972</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE10 (E5) EL737973</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE16 (E5) EL737974</td>
<td>TC137946</td>
<td>2e-34</td>
</tr>
<tr>
<td>CE17 (E12) EL737975</td>
<td>CN213753</td>
<td>7e-143</td>
</tr>
<tr>
<td>CE18 (E12) EL737976</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE19 (E12) EL737977</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE20 (C1) EL737978</td>
<td>SGN-U271429</td>
<td>7e-38</td>
</tr>
<tr>
<td>CE21 (C2) EL737979</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE22 (C11) EL737980</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE23 (C11) EL737981</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE24 (C2) EL737982</td>
<td>SGN-U282937</td>
<td>1e-123</td>
</tr>
<tr>
<td>CE25 (C1) EL737983</td>
<td>TC144731</td>
<td>2e-48</td>
</tr>
<tr>
<td>CE26 (C12) EL737984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE28 (C8) EL737985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE29 (C11) EL737986</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE30 (E12) EL737987</td>
<td>CK293066</td>
<td>3e-17</td>
</tr>
<tr>
<td>CE31 (E12) EL737988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE32 (E5) EL737989</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE33 (E5) EL737990</td>
<td>SGN-U274734</td>
<td>7e-42</td>
</tr>
<tr>
<td>CE35 (E5) EL737991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE36 (E5) EL737992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE37 (E5) EL737993</td>
<td>SGN-U285874</td>
<td>4e-19</td>
</tr>
</tbody>
</table>

<sup>a</sup> Chromosome location of the TDF is indicated in parentheses.

<sup>b</sup> BLASTN was carried out using the Plant Unique Gene Indexes (TIGR, The Institute for Genomics Research), Unigene database (SGN, Sol Genomics Network), or the EST others database (NCBI; National Center for Biotechnology Information) only in cases when the Blastx of the TDF against the non-redundant protein database (NCBI) yielded no significant homologies (e-value >10^-4). The Unigene or EST was then used to search the non-redundant protein database (NCBI) for a second time using the program Blastsx.
transcripts) when leaves from greenhouse-grown plants were used as the source of mRNA, reflecting the variation in rate of growth and development among individuals and less uniform growing conditions. This variation in gene expression resulted in a high number of false positives, highlighting the importance of minimizing all source of variation unrelated to the trait for an efficient application of BSA using cDNA-AFLP. The sample heterogeneity did not, however, hinder the identification of earliness-linked transcripts, and six new TDFs that were not detectable in the in vitro bulks were found. None of the observed intensity differences were associated with earliness and, consequently, they were not analysed further. We cannot conclude therefore how much of the observed variation had a genetic origin.

The QTL analysis of earliness in the CxE population identified a major locus in chromosome 5, which alone explains the tuberization behaviour differences between the parental lines. Additionally, a small QTL showing epistatic effects was detected on chromosome 12 for field tuberization time. The lack of significance of this second QTL for greenhouse earliness is most probably due to the small population size of the greenhouse trial, which may have limited our ability to detect minor QTLs. Van den Berg et al. (1996) identified earliness QTLs on seven of the 12 potato chromosomes, whereas Simko et al. (1999) reported four main effect QTLs for greenhouse tuberization time. Although the total number of QTLs was larger in these populations, alleles linked at chromosome 5 appeared as the main ones responsible for the tuberization differences in both cases. A definitive comparison between the QTL position found in this study and that mapped by van den Berg et al. (1996) and Simko et al. (1999) is, however, not possible due to the lack of common markers. Alleles at chromosome 5 have also appeared, in this and several other populations, to be responsible for the main differences in foliage maturity, a trait intimately related to earliness of tuberization and that can in fact be considered as an alternative way to measure tuberization time (Collins et al., 1999; Visker et al., 2003; Bradshaw et al., 2004). It is generally considered that the use of BSA for QTL detection is only successful in detecting QTLs of large effect. This is partially in agreement with our results. A large fraction of the identified TDFs (18 out of 37) indeed map on chromosome 5. Prior to this study there was a scarcity of markers at this chromosomal region in our genetic map, with a distance of 18 cM between the two flanking genomic markers. As a result of the BSA, 10 new markers representing expressed genes and showing strong linkage to the trait have been placed in this interval. However, a gap of 10 cM between the TDFs showing the strongest association with earliness and the next marker on the distal part of the chromosome still remains. This could be due to a scarcity of expressed genes in this region or to the existence of a high recombination rate that results in large genetic distances. Although fewer markers mapping to this region were obtained, we could also identify a second QTL explaining a small proportion of the total variance on chromosome 12 of the paternal map. Additionally, our BSA identified markers on five other chromosomes of the maternal map, mainly associated with the latest genotypes, which could contribute secondarily to the transgression observed in the progeny. However, no

<table>
<thead>
<tr>
<th>TDF</th>
<th>Contig</th>
<th>BAC clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>230</td>
<td>RH017L09</td>
<td>RH203F04</td>
</tr>
<tr>
<td></td>
<td>RH070A08</td>
<td>RH057H17</td>
</tr>
<tr>
<td></td>
<td>RH055G13</td>
<td>RH184C02</td>
</tr>
<tr>
<td></td>
<td>RH166I17</td>
<td>RH009I10</td>
</tr>
<tr>
<td></td>
<td>RH020M24</td>
<td>RH179K23</td>
</tr>
<tr>
<td></td>
<td>RH080G03</td>
<td>RH152B04</td>
</tr>
<tr>
<td></td>
<td>RH038E10</td>
<td>RH028E01</td>
</tr>
<tr>
<td></td>
<td>RH065C02</td>
<td>RH040E13</td>
</tr>
<tr>
<td></td>
<td>RH102A09</td>
<td>RH070G22</td>
</tr>
<tr>
<td></td>
<td>RH088O15</td>
<td>RH011N05</td>
</tr>
<tr>
<td>735</td>
<td>RH036E01</td>
<td>RH173D01</td>
</tr>
<tr>
<td></td>
<td>RH101L13</td>
<td>RH015H12</td>
</tr>
<tr>
<td></td>
<td>RH129D11</td>
<td>RH159N05</td>
</tr>
<tr>
<td></td>
<td>RH086A13</td>
<td>RH008E01</td>
</tr>
<tr>
<td>941</td>
<td>RH017I23</td>
<td>RH047I19</td>
</tr>
<tr>
<td></td>
<td>RH091B22</td>
<td>RH120H10</td>
</tr>
<tr>
<td></td>
<td>RH057A13</td>
<td>RH011G03</td>
</tr>
<tr>
<td>CE4</td>
<td>RH017I23</td>
<td>RH136H10</td>
</tr>
<tr>
<td></td>
<td>RH047I19</td>
<td>RH091B22</td>
</tr>
<tr>
<td></td>
<td>RH120H10</td>
<td>RH057A13</td>
</tr>
<tr>
<td></td>
<td>RH011G03</td>
<td>RH017I23</td>
</tr>
<tr>
<td>1954</td>
<td>RH091M15</td>
<td>RH136H10</td>
</tr>
<tr>
<td></td>
<td>RH136H10</td>
<td>RH091M15</td>
</tr>
<tr>
<td>6918</td>
<td>RH144I18</td>
<td>RH105B16</td>
</tr>
<tr>
<td></td>
<td>RH141D14</td>
<td>RH126O17</td>
</tr>
<tr>
<td>CE35</td>
<td>RH165G01</td>
<td>RH003L15</td>
</tr>
<tr>
<td></td>
<td>RH003L15</td>
<td>RH165G01</td>
</tr>
<tr>
<td></td>
<td>RH131C17</td>
<td>RH160D20</td>
</tr>
<tr>
<td>5116</td>
<td>RH169D20</td>
<td>RH165G11</td>
</tr>
<tr>
<td></td>
<td>RH165G11</td>
<td>RH169D20</td>
</tr>
<tr>
<td></td>
<td>RH003L15</td>
<td>RH131C17</td>
</tr>
</tbody>
</table>
significant QTL was detected associated with these markers. This may reflect a limitation in detecting minor QTLs in populations where a QTL with a large effect is segregating, or it could indicate a non-causal linkage with the individuals in the bulk.

The existence of at least two major tuberization control pathways has been proposed, the gibberellin and the photoperiod pathways (for a review, see Rodríguez-Falcón et al., 2006). In the proposed photoperiod pathway, analogous to the well characterized flowering pathway, phytochromes and, most probably, cryptochromes play a role in setting the phase of the internal circadian clock. The clock regulates the abundance of GIGANTEA (GI), a nuclear-localized protein that activates transcription of CO. CO then regulates the expression of FT that triggers tuberization transition. Under long day conditions, the peak of CO expression occurs during the day; interaction of CO with PHYB would then result in repression of FT expression and so of tuberization. Under short day conditions, CO accumulates in the night when PHYB is in an inactive form, allowing CO-mediated induction of FT and thus promotion of tuberization. Evidence for a role for PHYB and CO in photoperiodic control of tuberization has been obtained by transgenic studies. Preliminary results also support a function for GI and FT in the day-length control of tuberization induction (Rodriguez-Falcón et al., 2006). One of the TDFs identified in this study, CE3, on chromosome E5, shows homology to the Arabidopsis GI gene. Based on its putative function, CE3 would be an attractive candidate gene to pursue further. However, although a single marker analysis reveals a highly significant association with earliness, its position 16 cM away from the highest LOD value for QTL E-5 does not support a major role for this TDF, and therefore for GI, in explaining the variation in tuberization time present in the population.

In addition to the photoperiod pathway, a gibberellin-dependent pathway is also believed to play a key role in the control of tuber induction. GAs have long been implicated in the control of tuberization as negative regulators. More recent evidence, however, indicates that while in the stolon high levels of GAs inhibit tuber formation, this does not necessarily apply to the shoot. In this respect, it seems that a high conversion rate of GA20 to GA1 in the shoot could favour tuber formation, possibly by lowering the levels of GA20 in the aerial part of the plant and thus its transport to the stolon (Rodriguez-Falcón et al., 2006). This observation would fit with the results obtained with transgenic lines showing downregulation or overexpression of GA 20-oxidase, the enzyme catalysing the three steps from GA53 to GA20, which has been shown to play a role in the regulation of tuber formation (Carrera et al., 2000; Jackson et al., 2000). Transport of GA20 from the shoot to the stolon could also explain the observed increase in GA20 levels in the stolon in GA 2-oxidase silencing lines (Kloosterman, 2006). A GA 20-oxidase homologue is present in clone PGEC417F14 (AC149289) from Solanum demissum chromosome 11; this clone also contains a sequence homologous to one of the TDFs mapping on C11, CE22. We
could not, detect any association between this region in chromosome C11 and earliness, despite the fact that the BSA identified three TDFs at this location. Interestingly, CE31, linked to the QTL on E12, shares similarity with ent-kaureonic acid oxidase, an enzyme catalysing the early steps of GA biosynthesis from ent-kaureonic acid to GA\textsubscript{12}. Mutants in this gene have been identified in several plant species including maize, pea, and rice (Helliwell et al., 2001; Davidson et al., 2003; Sakamoto et al., 2004). These mutants show a characteristic dwarf phenotype due to GA deficiency, with shorter internodes and smaller and darker leaves. There are no reports on the function of this gene on tuberization since most studies have focused on downstream genes in the pathway.

Among the TDFs with homology to proteins with known function mapping closer to the main QTL on chromosome 5, CE6, the marker with the highest LOD value for field tuberization time, shows similarity to a quinolinate phosphoribosyl transferase (QPRT). QPRT is a key enzyme in NAD biosynthesis, catalysing the entry point step in the pyridine nucleotide cycle. In addition, QPRT is essential for the supply of nicotinic acid for pyridine alkaloid synthesis, a biosynthetic pathway induced by methyl jasmonates in response to wounding. NAD plays a major role in metabolism as coenzyme and substrate of numerous enzymatic reactions, being essential for a whole range of biological processes from regulation of energy metabolism to transcription (Lin and Guarente, 2003). QPRT therefore occupies a key position in primary metabolism. However, a specific connection to tuber formation is unclear. Two other TDFs mapping to the QTL on E5, CE5 and CE33, show homology to a tRNA-isopentenyl transferase, a key enzyme in CK biosynthesis (Zubko et al., 2002). In addition to promoting cell division, several lines of evidence support a role for CKs in the regulation of assimilate partitioning and sink–source relations (Roitsch and Ehneb, 2000). CE5 and CE33, show homology to a tRNA-isopentenyl transferase, a key enzyme in CK biosynthesis (Zubko et al., 2002). In addition to promoting cell division, several lines of evidence support a role for CKs in the regulation of assimilate partitioning and sink–source relations (Roitsch and Ehneb, 2000; Guivarc’h et al., 2002; Balibrea Lara et al., 2004). CKs may therefore be important for cell proliferation during the initial stages of tuber formation, and the development of sink strength. Guivarc’h et al. (2002) reported that overproduction of CK in axillary buds of tobacco led to the formation of short lateral branches with swelling internodes that accumulate starch grains in a process that resembles tuberization. Additionally, the \textit{POTH1} and \textit{BEL5} genes, which have been suggested to play a role in the control of tuber formation by modifying GA levels through the regulation of \textit{GA 20-oxidase}, also seem to regulate CK levels (Chen et al., 2003, 2004b; Rosin et al., 2003). Regulation of CK biosynthesis by KNOX genes has also been reported recently by Yanai et al. (2005) and Jasinski et al. (2005), suggesting that KNOX proteins may act as growth regulators by simultaneously regulating GA and CK levels. All these results support a possible role for CK in the control of tuber formation.

The ultimate goal of the study of natural variation in tuberization time is the identification of the gene(s) responsible for the existing phenotypic variation. At this point, we do not have enough evidence to identify any of the mapped TDFs as the gene responsible for the QTL. They are, however, a good starting point for the cloning of the earliness locus. An initial screening of a potato BAC library with some of these markers has identified those clones harbouring the TDFs, some of which have been anchored to chromosome 5 of the potato physical map. Further refinement of the physical map of the region and BAC sequencing will aid in the identification of new candidates and the elucidation of the molecular mechanisms of the tuberization process.

**Acknowledgements**

The authors would like to thank Dr Jan de Boer and Ir Theo Borm for providing the BAC library and for their advice on the screening procedure.

**References**


Raices M, Chico JM, Tellez-Inon MT, Ulloa RM. 2001. Molecular characterization of StCDPK1, a calcium-dependent protein kinase from *Solanum tuberosum* that is induced at the onset of tuber development. *Plant Molecular Biology* 46, 591–601.


Raices M, Ulloa RM, MacIntosh GC, Crespi M, Tellez-Inon MT. 2003b. StCDPK1 is expressed in potato stolon tips.
and is induced by high sucrose concentration. *Journal of Experimental Botany* 54, 2589–2591.


---


