Tree Genetics & Genomes (2008) 4:495–507
DOI 10.1007/s11295-007-0126-9

ORIGINAL PAPER

Sequencing and annotation of the evergrowing locus in peach [*Prunus persica* (L.) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation

Douglas Gary Bielenberg · Ying (Eileen) Wang · Zhigang Li · Tetyana Zhebentyayeva · Shenghua Fan · Gregory Lynn Reighard · Ralph Scorza · Albert Glenn Abbott

Received: 14 May 2007 / Revised: 23 October 2007 / Accepted: 27 November 2007 / Published online: 10 January 2008
© Springer-Verlag 2007

Abstract Buds are specialized structures that protect fragile meristematic regions during dormancy and are part of the mechanism that plants use to survive unfavorable environmental conditions such as low temperature or dessication stress. The *evergrowing* (*evg*) mutant of peach [*Prunus persica* (L.) Batsch] does not form terminal vegetative buds in response to dormancy-inducing conditions such as short days and low temperatures, and the terminal meristems maintain constant growth (leaf addition and internode elongation). We genetically mapped the *evg* trait and identified the corresponding genomic region in a wild-type genome. We sequenced and annotated the 132-kb region. Nineteen genes were predicted to be in the sequenced region. Ten of the predicted genes were demonstrated to be expressed in the wild-type germplasm but six of these were not expressed in mutant tissues. These six genes are a cluster of MIKC-type MADS-box transcription factors similar to genes from *Ipomoea batatas* and *Solanum tuberosum* MADS-box, which also regulate meristem growth in vegetative tissues. A 41,746-bp deletion is present in this region of the mutant genome which results in the loss of all or part of four of the six MADS-box genes. The six MADS-box genes that are not expressed in the mutant are candidates for the regulation of growth cessation and terminal bud formation in peach in response to dormancy-inducing conditions and have been named *d*ormancy-*a*ssociated *M*ADS-box (DAM) genes.

Keywords MADS-box · Bud · Dormancy · Evergrowing · Peach

Communicated by J. Davis

Financial Source This work was supported in part by United States Department of Agriculture (USDA) Cooperative State Research, Education, and Extension Service (CSREES) Special Research Grant AC 2004-06140 to GLR and USDA CSREES NRI Grant #2005-35300-15452 to AGA.

D. G. Bielenberg (✉) · Z. Li · G. L. Reighard
Department of Horticulture, Clemson University,
152 Poole Agricultural Center,
Clemson, SC 29634-0319, USA
e-mail: dbielen@clemson.edu

Y. (E.) Wang · T. Zhebentyayeva · S. Fan · A. G. Abbott
Department of Genetics & Biochemistry, Clemson University,
Clemson, SC 29634-0318, USA

R. Scorza
Appalachian Fruit Research Station, USDA-ARS,
Kearneysville, WV 25430, USA

Introduction

Perennial woody plants (particularly trees) are the dominant species in many ecosystems and represent the majority of terrestrial plant biomass. Perennial woody species native to temperate and subtropical regions alternate between periods of active growth, when conditions are favorable, and periods of dormancy, when conditions are detrimental to growth and/or survival. Several distinct physiological phenomena have been the focus of research on perennial dormancy and winter or dry season survival: the induction of bud structures and endodormancy, the induction of cold or drought hardiness separate from and in combination with

bud dormancy, the development of a quantitative requirement for cold exposure (chilling requirement), and the breaking of dormancy for the resumption of growth (Dennis 1996; Faust et al. 1997; Fuchigami and Wisniewski 1997; Kozlowski and Pallardy 1997; Rowland et al. 1999).

Current understanding of the regulatory network(s) involved in dormancy induction is limited. The basic genetic factors governing the developmental pathway(s) of bud and dormancy induction remain elusive. Several reviews have summarized the current state of the field (Rohde et al. 2000; Horvath et al. 2003; Rohde and Bhalerao 2007). In brief, there is obvious involvement of genes controlling perceptive events (e.g., phytochrome and the perception of photoperiod), as well as data supporting the downstream regulation of the cell cycle (Anderson et al. 2001; Horvath and Anderson 2002; Anderson et al. 2005) that must occur for growth arrest. In between these two endpoints of the signal and the response, less is known, although the recent analysis of the *Populus* CO/FT regulatory genes have demonstrated a role for these genes in seasonal growth dynamics in addition to the role they have in flowering regulation (Bohlenius et al. 2006). Hormones (gibberellic acid (GA), abscisic acid, and ethylene) and metabolic regulation (Gevaudant et al. 2001) are clearly involved in the regulatory network(s) connecting signal and response, as demonstrated by reverse genetic approaches (Rohde et al. 2000; Horvath et al. 2003; Molmann et al. 2003, 2005; Olsen 2003; Ruonala et al. 2006). Just as analysis of mutants for environmental responses allowed the dissection of regulatory networks in model species such as Arabidopsis (Klee and Estelle 1991; Barak et al. 2000; Michaels and Amasino 2000; Koornneef et al. 2002), available dormancy-incapable mutants (natural or transgenic) are useful for the identification of regulatory genes involved in bud and dormancy induction.

Mutants that fail to cease growth and enter dormancy under dormancy-inducing conditions have been described in only two woody perennial species, *Corylus avellana* L. (Thompson et al. 1985) and *Prunus persica* (L.) Batsch (Rodriguez et al. 1994). The best-described mutant of these two species is the *evergrowing* (*evg*; previously known as evergreen, United States Department of Agriculture PI442380) peach mutant, a non-dormant genotype identified from southern Mexico (Rodriguez et al. 1994; Werner and Okie 1998). Growth of terminal meristems on *evg* trees is continuous, and terminal buds do not form in response to short days or cold temperatures. The non-dormant *evg* trait segregates as a single recessive nuclear gene (Rodriguez et al. 1994). Previous work by our group used a segregating full sibling F_2 population to generate a genetic map of the *evg* trait. This map localized *evg* to an 11.7-cM region of linkage group one (LG1) of the general *Prunus* genetic map (Rodriguez et al. 1994; Wang et al. 2002a, b). On our genetic map, the amplified fragment length polymorphism (AFLP) marker EAT/MCAC is linked to the EVG locus at a distance of only 1.0 cM (Wang et al. 2002a). The EAT/MCAC band was cloned, sequenced, and converted to a sequence tagged site (STS) marker (Wang et al. 2002a). The EAT/MCAC STS fragment was used to screen a bacterial artificial chromosome (BAC) library for clones representing the *evg* genomic locus.

Here, we report the identification of two additional simple sequence repeat (SSR) markers from a BAC that hybridized with the EAT/MCAC STS fragment. One of these SSR markers flanks the *evg* trait at a distance of 2.0 cM and the other marker cosegregates with the *evg* trait. We subsequently physically mapped, sequenced, and annotated 132,177 base pairs of genomic DNA that encompasses the region controlling the *evg* trait. The sequenced region contained several predicted genes, including a cluster of six MADS-box transcription factors. Southern hybridization analysis using probes developed from these genes demonstrated that a portion of the EVG locus comprising at least four of these MADS-box genes is deleted in the mutant (Bielenberg et al. 2004). We have determined the exact breakpoints of this deletion in the mutant by inverse polymerase chain reaction (PCR). Reverse transcription and PCR amplification of predicted genes in the locus reveals that the *evg* mutation eliminates expression of all six of the MADS-box transcription factors and does not affect the expression of any other transcribed genes in the locus. These six genes are the first-gene candidates identified by forward genetics and map-based cloning for a perennial life history trait. These candidates are genes not previously thought to have a role in dormancy induction in perennials.

Materials and methods

Fine-mapping of the EVG locus

New SSR loci pchgms40 ($[GA]_7$ repeat, GenBank ID DQ847415, pchgms40F 5'-TCAAGCTCAAGGTAC CAGCA-3', pchgms40R 5'-AAGGCACTCTCC CTCTCCTC-3') and pchgms41 ($[AG]_{10}$ repeat, GenBank ID DQ847416, pchgms41 F 5'-GGAAATTC CCTGTCCTTCCT-3', pchgms41 R 5'-CCTCGAAC TAGTTGCCTTTGA-3') were identified from subclones of BAC PpN018F12 using the screening technique previously described (Wang et al. 2002b). A genetic map of the EVG locus was generated from a segregating F_2 population of 109 individuals derived from a cross between wild-type dormant ('Empress' dwarf) and mutant, non-dormant (*evg*, PI442380) parents (Rodriguez et al. 1994; Wang et al. 2002a). SSR analysis was performed as described by Wang

et al. (2002b). The linkage map was generated using JoinMap® v3.0 (Van Ooijen 2001). Marker positions were calculated using a minimum logarithm of odds score threshold of 8.0 with application of Kosambi mapping function.

Sequencing of BAC clones

All BAC clones used in this study were obtained from a genomic library of the peach rootstock Nemared (wild-type dormant; Georgi et al. 2002). BAC DNA was isolated as previously described (Bielenberg et al. 2004). Sequencing of BAC PpN018F12 was performed with a transposon insertion system. Cesium chloride purified PpN018F12 was inserted in vitro with the Tn5 transposon insertion system of the EZ:TN™ <KAN2> Insertion Kit (Epicenter, Madison, WI, USA) Transformed (kan^R) Tn5-containing BACs were sequenced at the Clemson University Genomics Institute using the forward and reverse primers located on the inserted transposon. The significant duplication in the *evg* region caused difficulty with automated assembly of the sequence derived solely from transposon inserted BAC DNA. Therefore, *Hin*dIII and *Eco*RI subclones of PpN018F12 were made and fully sequenced by primer walking to provide large stretches of known contiguous DNA sequence to seed the sequence assembly. BACs PpN089G02 and PpN018G07 were completely sequenced by subcloning restriction enzyme fragments and fully sequencing the resulting clone by primer walking. Sequencing reactions were performed using Big Dye™ v2.0 terminator chemistry (Applied Biosystems, Inc., Foster City, CA, USA) according to the manufacturer's instructions for large-insert DNA. Sequencing reactions were analyzed on an ABI PRISM 377 sequencer (Applied Biosystems, Inc., Foster City, CA, U.S.A.).

Sequence analysis and gene prediction

DNA sequence was assembled with Sequencher v4.1.2 (Gene Codes Corp., Ann Arbor, MI, USA). Artifacts generated during subcloning of BACs such as the presence of bacterial IS10 sequences were manually removed from sequence assemblies (Kovarík et al. 2001). Potential genes were identified by parallel methods. Four-gene-prediction programs were used to identify potential exon–intron structure: Fgenesh (Salamov and Solovyev 2000), GeneMark (Besemer and Borodovsky 2005), GlimmerHMM (Majoros et al. 2004), and GenScan (Burge and Karlin 1997). The Arabidopsis-trained prediction algorithms were used in all cases where available. Predicted cDNAs and polypeptides were compared against expressed sequence tags (EST) and protein databases for sequence similarity with tBLASTx or tBLASTn algorithms (Altschul et al. 1990; Gish and States 1993).

In parallel, 1,000-bp segments of the sequenced region in steps consisting of 500 bp overlaps were compared against the National Center for Biotechnology Information (NCBI) GenBank, SwissProt, and translated European Molecular Biology Laboratory (TREMBL) nucleotide and protein databases using the tBLASTx algorithm (Altschul et al. 1990; Gish and States 1993). Genomic DNA sequences with significant tBLASTx similarities to ESTs, genes, or proteins in the databases were cross-referenced with the output of the gene-prediction programs. Repetitive element sequences were detected with Repbase (Jurka et al. 2005). The annotated 132,177 bp sequence was deposited in GenBank as accession number DQ863257.

cDNA isolation and expression analysis

For the purpose of detecting if predicted genes were expressed, a pooled sample of RNA was made representing the wild-type tissues and another representing the mutant tissues. The pooled samples contained RNA from root and shoot (stem, leaf, branch tips, flowers, and fruit) tissues grown in the greenhouse and RNA from shoot tissues grown in the field and collected at four times during the annual growth cycle: post-chilling, pre-budbreak (March); actively growing, long days (June); shortening days, budset (September); and dormant (December).

Total RNA was extracted by the procedure outlined in Gevaudant et al. (1999) and modified for use in our laboratory. DNase-treated total RNA samples were reverse-transcribed using an anchored 5'-$(T)_{17}$V-3' oligomer as a primer with SuperScript III First Strand Synthesis System for reverse transcriptase (RT)-PCR, where V = A, G, or C (Invitrogen, Carlsbad, CA, U.S.A). Subsequent amplification of gene-specific products was accomplished following van der Linden et al. (2002). In brief, this method entails an initial round of linear amplification with a forward primer and subsequent logarithmic amplification with a second nested forward primer and a downstream reverse primer (van der Linden et al. 2002). Forward and reverse PCR primers for each of the amplifications were designed from BAC sequences using Primer3_www.cgi v 0.2 (Rozen and Skaletsky 2000) and obtained from Integrated DNA Technologies (Coralville, IA, USA). The amplification products were gel-purified and isolated with the QIAquick® Gel Extraction Kit (Qiagen, Inc., Valencia, CA, USA). Isolated cDNAs were A-tailed and cloned into pGEM-T using the pGEM-T cloning kit (Promega, Madison, WI, USA). Full-length cDNAs were isolated by 5' Rapid Amplification of cDNA Ends (RACE) and 3'RACE using gene-specific primers for each of the target genes, according to Sambrook and Russell (2002). Full-length cDNA sequences were deposited in GenBank and assigned the following accession numbers: DAM1, DQ863253; DAM2, DQ863255; DAM3,

DQ863256; DAM4, DQ863250; DAM5, DQ863251; and DAM6, DQ863252.

RT-PCR Primers designed from a *P. persica* tubulin-α EST were used (DY650410) as a positive control for amplification of predicted genes from wild-type and mutant cDNA. Template cDNA was synthesized from 2.5 µg of total RNA using the SuperScript first-strand synthesis system as above. First strand cDNAs were diluted 1:4 with nuclease-free water and aliquots of the cDNA sample were amplified using gene-specific primers (Table 1). PCR reactions were performed in a 25 µl volume containing 1 U of Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA, USA), 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 0.4 µM of each primers and 1×Taq polymerase reaction buffer, and 4 µl of cDNA sample. PCR was performed with the following parameters: 95°C for 60 s; 26 cycles of 94°C for 30 s, 62°C for 30 s, 72°C for 40 s; followed by 72°C for 120 s. Amplification products were size-fractionated on a 3% (*w*/*v*) agarose gel, stained with SYBR Green (Invitrogen, Carlsbad, CA, USA), and photographed using Fujifilm Science Laboratory 99 Image Gauge system (Fujifilm USA, Inc., Vahalla, NY, USA).

Cloning of mutant deletion sequence by inverse PCR

Five micrograms of mutant genomic DNA was digested overnight in a 20 µl reaction containing *Hin*dIII enzyme (5 U enzyme per µg of DNA; Promega, Madison, WI, U.S.A.) followed by a 20-min heating to 65°C to destroy the enzyme activity. One microgram of digested genomic DNA was self-ligated overnight at room temperature in a total volume of 500 µl following mixing with 30 U of T4 DNA Ligase (Promega, Madison, WI, USA) and ligation buffer at a 1× concentration. Circularized DNA was amplified using primers 18HB09-iPCR1 (5'- CGGTTTTGGGTGAAGTT TAGTGTG -3'), and 18HB09-iPCR2 (5'- CTGCGAGTT GGGTTTGTTGCTG -3'). PCR was performed with 2 µl of the self-ligation in a volume of 25 µl, containing 1 U of Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA, USA.), 0.4 µM of each primer, 200 µM dNTPs, 1.5 mM $MgCl_2$, and 1× *Taq* polymerase buffer. An initial denaturation step of 30 s at 95°C was followed by 37 cycles of 94°C for 30 s, 62°C for 60 s, and 72°C for 180 s. The resulting amplified band was gel-excised, isolated and TA-cloned into the pGEM-T vector. Sequencing was performed as above following plasmid isolation.

Results

Fine genetic mapping of evg

The EAT/MCAC STS (GenBank AF516359) was used to identify a BAC clone, PpN018F12, in the putative physical locus of the *evg* trait. PpN018F12 was subcloned following *Hin*dIII digestion. Clones containing *Hin*dIII digestions fragments of PpN018F12 were dot-blotted and screened for additional SSRs by hybridization with radiolabeled oligonucleotides representing different di- or tri-nucleotide repeats. Positively hybridizing subclones were end-sequenced to verify the presence of an SSR and identify primer sites for amplification. All potential markers were amplified from parental DNA and a subset of F_2 offspring of both genotypes.

Two new SSR markers polymorphic in the F_2 mapping population were identified. The SSR marker pchgms40 (see "Materials and methods" for SSR primer sequences) was dominant and polymorphic in the F2 mapping progeny. The pchgms40 amplicon was always present in the wild-type progeny and missing in the mutant progeny. Pchgms41 was co-dominant and polymorphic in the F_2 mapping progeny. Pchgms41 mapped to a distance of 2 cM from the *evg* trait, closer than the nearest SSR markers previously identified

Table 1 Oligonucleotide sequences for gene-specific amplification of the six MADS-box genes found in the EVG locus

Gene	Accession number	Oligonucleotide	Product size (bp)
DAM1	DQ863253	F-GGGGACGATGAAAATGACGAGGGAG R-CAATCACCCGGCCAAGGCTTGCATC	428
DAM2	DQ863255	F-CAGTCAGCCAGCAGGAGAAGCAGCC R-ACCAGCTCAGCTCCCTCAGT	546
DAM3	DQ863256	F-ACCAGCTAAGGCAGACGATGA R-GAGGGAGAGAGACTGAGAGCA	166
DAM4	DQ863250	F-TGTGGCACTTGAGAAAAAGGGA R-CAGGTTACTTTCCCCAGGCCAC	419
DAM5	DQ863251	F-CCCCGAAACCCACCAACGAAGATG R-CAGCACTGTTGCAGGTGGTG	685
DAM6	DQ863252	F-CCAACAACCAGTTAAGGCAGAAGA R-GGAAGCCCCAGTTTGAGAGA	229

Each primer pair is separated by one or more introns in the genomic sequence.

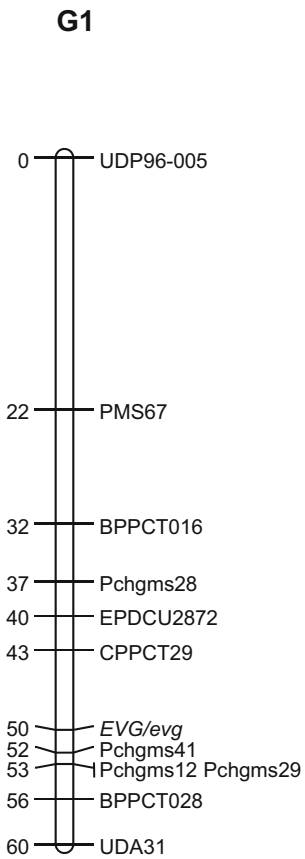

Fig. 1 Local genetic map of *evg* with SSR markers known to be present in linkage group 1 (LG1) of the *Prunus* general reference map and with the new SSR marker pchgms41. The *vertical line* is a representation of the linkage group and the *numbers to the left of the line* are cM distances positions of the genetic markers indicated on the right

(pchgms12 and pchgms29) in an earlier study (Wang et al. 2002a; Fig. 1). A local map of the *evg* region was constructed using SSR markers previously anchored a *Prunus* reference map (Joobeur et al. 1998; Dirlewanger et al. 2004): UDA031 (Messina et al. 2004), UDP96–005 (Cipriani et al. 1999), BPPCT016 and BPPCT028 (Dirlewanger et al. 2002), PMS67 (Cantini et al. 2001), CPPCT029 (Aranzana et al. 2002), and EPDCU2872 (Dirlewanger et al. 2004).

The close mapping of pchgms41 to the *evg* trait and cosegregation of pchgms40 with the *evg* trait indicated that PpN018F12 was likely to contain the sequence controlling *evg*. However, previous analysis by southern hybridization revealed that there appeared to be a deletion in the mutant locus that extended beyond PpN018F12 and contained repeated sequence (Bielenberg et al. 2004). Therefore, two BAC clones which overlapped either end of BAC PpN018F12 were identified by a hybridization walk, and these clones were also sequenced.

Sequence of the EVG locus

Three BACs putatively encompassing the EVG locus were sequenced. BACs PpN018F12, PpN089G02, and PpN018G07 were 54,998 bp, 61,930 bp, and 29,547 bp in length, respectively. BAC PpN018F12 was identified as containing the STS marker EAT/MCAC and served as the start point of the physical walk through the region to identify markers closely linked to *evg*. BACs PpN089G02 and PpN018G07 overlap PpN018F12 at the SP6 and T7 ends of PpN018F12, respectively. The T7 end of PpN089G02 overlaps the SP6 end of PpN018F12 by 6,804 bp. The T7 end of PpN018G07 overlaps the T7 end of PpN018F12 by 7,495 bp (Fig. 2a). The final contiguous length of sequenced genomic region was 132,177 bp (Fig. 2).

Several inverted repeat (IR) sequences were found in the sequenced region (Table 2; Fig. 2b). These repeats were between 18 and 60 bp in length with 10 to 3,974 bp of intervening sequence (Table 2). In the case of IR4 and IR5, each 'side' of the repeat is composed of a palindrome. The inverted repeats are clustered in two locations, one cluster is between 3,100 and 6,500 bp, (IR1, IR2, and IR3) and another between 33,500 and 45,200 bp (IR4, IR5, and IR6; Table 2; Fig. 2).

Gene content of the EVG locus

Gene content of the EVG locus was predicted by combined analysis of the results of GenScan, GeneMark.hmm,

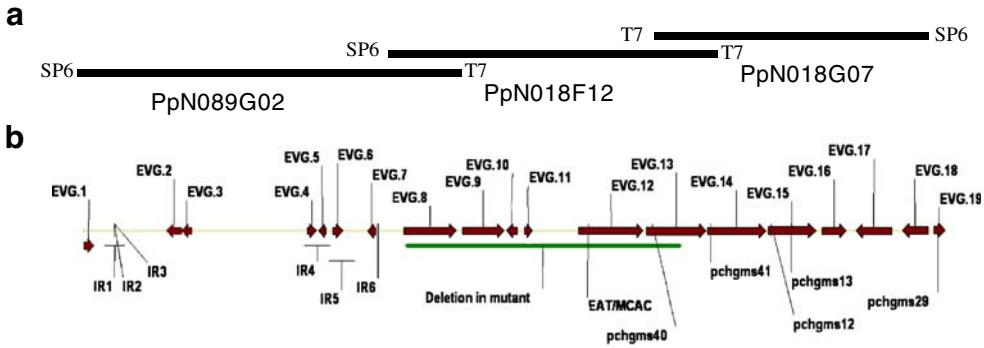

Fig. 2 Map and annotation of the EVG locus. **a** Relative size and overlapping sequence of the three BACs sequenced in the EVG locus. **b** Physical map of the EVG locus. Arrows indicate the sense orientation of the indicated genes and the relative genomic (exons and introns) size of the gene from transcription start site to transcription stop site (if known). Gene labels are as listed in Table 2. IR indicates the presence of an inverted repeat sequence, different inverted repeat sequences are numbers sequentially from left to right

Table 2 Inverted repeat (IR) sequences located in the *EVG* locus

Repeat	Left repeat position	Right repeat position	Size (bp)	Internal size (bp)	Notes
IR1	3,145–3,185	6,280–6,320	41	3,094	5-base mismatch in repeat
IR2	4,642–4,684	4,843–4,885	43	159	Identical repeat
IR3	4,687–4,745	4,795–4,853	59	49	2-base mismatch in repeat
IR4	33,748–33,787	37,762–37,802	40	3,974	3-base mismatch in repeat; repeat is palindromic
IR5	37,463–37,519	41,473–41,531	57	3,953	1-base mismatch in repeat; repeat is palindromic
IR6	45,101–45,161	45,172–45,232	60	10	8-base mismatch in repeat

GlimmerHMM, and fgenesh gene-prediction programs as described in the "Materials and methods" section (Burge and Karlin 1997; Salamov and Solovyev 2000; Majoros et al. 2004; Besemer and Borodovsky 2005). Nineteen genes were predicted in the *evg* region and were assigned a number based on their sequential position within the sequenced region from EVG.1 to EVG.19 (Table 3; Fig. 2).

The most prominent gene features in the sequenced region were six MIKC-type MADS-box transcription factors at EVG.8, EVG.9, EVG.12, EVG.13, EVG.14, and EVG.15 (Table 3, Fig. 2). These genes are arranged on the same coding strand and in tandem with a segment of intervening sequence between EVG.9 and EVG.12. Expression of each of the MADS-box genes were confirmed by amplification of full-length cDNAs from pooled total RNA from wild-type tissues (see "Materials and methods" for a description of the pooled samples). The full-length cDNAs of the MADS-box genes were translated in silico and the sequences aligned using the ClustalW algorithm with the default alignment parameters (Fig. 3). The predicted protein lengths were between 235 and 242 amino acids in length and similarity scores for pairwise comparisons of the amino acid sequences were between 73% and 84%. A tBLASTx similarity search against the NCBI non-redundant database showed that the six MADS-box genes are all similar to the StMADS11/AtAGL24/AtSVP group of MADS-box transcription factors which are primarily expressed in vegetative tissues (Table 4).

Gene EVG.1 was predicted to be comprised of one exon with similarity to the *a*uxin-*i*nduced *r*esponse 12 protein from Arabidopsis. The gene prediction for EVG.1 was confirmed by isolation of the cDNA containing the 3'UTR of this gene. Gene EVG.2 was predicted to be on the opposite strand from EVG.1 and to encode for a carbonic anhydrase protein (Table 3; Fig. 2). EVG.2 was confirmed by cDNA isolation and given the name *P. persica c*arbonic *a*nhydrase (PpCA). Gene EVG.3 was predicted to be a basic helix–loop–helix (EF-hand) protein (Table 3; Fig. 2) and was confirmed by cDNA isolation.

The sequenced region contains a group of predicted genes, EVG.4, EVG.5, EVG.6, and EVG.7, that have high similarity to the two open reading frames (ORFs) found in the AtIS112A autonomous transposon in Arabidopsis named for the similarity to the bacterial IS112 autonomous transposon (Jurka et al. 2005). Predicted genes EVG.4 and EVG.6 are near identical in sequence and both encode a transposase-like protein similar to the putative AtIS112A-2p protein (Jurka et al. 2005). EVG.5 and EVG.7 are highly similar to each other and encode a protein with similarity to the putative DNA-binding protein (1p) of AtIS112A (Jurka et al. 2005). Expression of the predicted ORFs of EVG.4 through EVG.7 could not be confirmed by RT-PCR of cDNAs prepared from the pooled wild-type RNA samples (Table 3). Similarly, primers designed for EVG.4, EVG.5, EVG.6, and EVG.7 amplified products from the pooled wild-type cDNA samples but cloned fragments did not correspond to the genomic DNA sequence flanked by the primer sites. Southern hybridization of digested genomic DNA with labeled EVG.7 amplification product revealed at least eight copies of the target sequence in the wild-type peach genome (data not shown).

EVG.10 is a putative ORF with similarity to the polyprotein encoding sequence from the *Copia* class of retrotransposons. No long terminal repeat (LTR) sequences were detected around this putative ORF. cDNA corresponding to the predicted exons was isolated from the wild-type pooled samples, suggesting that the gene was expressed.

EVG.11 is a predicted ORF comprised of two exons, the first of which is predicted to be a calcium-binding domain with a single EF-hand motif (Table 3). The second exon in the predicted gene is primarily composed of GGT repeats and no significant similarity to any protein or nucleotide sequence in the NCBI databases. Expression of EVG.11 could not be confirmed by RT-PCR, as above (Table 3). Attempts to isolate EVG.11 were unsuccessful but resulted in consistent amplification of another cDNA with very high similarity to the predicted first exon of EVG.11. Southern hybridization of wild-type genomic DNA with the predicted calcium binding domain sequence of EVG.11 revealed two hybridizing bands, indicating that another highly similar sequence is present in the wild-type genome (data not shown).

EVG.16 was predicted to contain a domain belonging to the protein family PF03138, which does not have a currently defined function. Expression of EVG.16 was confirmed by isolation of partial cDNA spanning multiple

Table 3 Gene content of the *EVG* locus of peach

ID	Designation	Strand	Region	Predicted gene product	Expressed in wild-type?	Expressed in mutant?
EVG.1	Expressed gene	+	103–1079	Similar to air 12 protein in Arabidopsis	Yes	Yes
EVG.2	Putative PpCA	−	14,934–12,683	Similar to Arabidopsis carbonic anhydrase	Yes	Yes
EVG.3	Putative bHLH	−	16,312–15,498	Putative basic helix-loop-helix domain	Yes	Yes
EVG.4	Putative transposase	+	34,253–35,629	Similar to transposase domain of AtIS112A; 99% nucleotide similarity to EVG.6	No	n.d.
EVG.5	DNA-binding domain	−	37,142–36,039	Similar to DNA-binding domain of AtIS112A; nucleotide similarity to EVG.7	No	n.d.
EVG.6	Putative transposase	+	38,274–39,650	Similar to transposase domain of AtIS112A; 99% nucleotide similarity to EVG.4	No	n.d.
EVG.7	DNA-binding domain	−	44,721–43,768	Similar to DNA-binding domain of AtIS112A; nucleotide similarity to EVG.5	No	n.d.
EVG.8	DAM1	+	49,186–57,052	MIKC-type MADS-box transcription factor of the StMADS11/SVP subfamily; full-length cDNA isolated.	Yes	No
EVG.9	DAM2	+	58,142–64,448	MIKC-type MADS-box transcription factor of the StMADS11/SVP subfamily; full-length cDNA isolate	Yes	No
EVG.10	Retro	−	65,515–64,934	Copia-like retrotransposon protein, no identifiable LTR	Yes	No
EVG.11	CaBD	+	67,735–68,878	Ca^{2+}-binding domain, single EF-hand	No	n.d.
EVG.12	DAM3	+	76,049–85,785	MIKC-type MADS-box transcription factor of the StMADS11/SVP subfamily; full-length cDNA isolated	Yes	No
EVG.13	DAM4	+	86,381–95,391	MIKC-type MADS-box transcription factor of the StMADS11/SVP subfamily; full-length cDNA isolated	Yes	No
EVG.14	DAM5	+	95,667–104,606	MIKC-type MADS-box transcription factor of the StMADS11/SVP subfamily; full-length cDNA isolated	Yes	No
EVG.15	DAM6	+	105,077–112,304	MIKC-type MADS-box transcription factor of the StMADS11/SVP subfamily; full-length cDNA isolated	Yes	No
EVG.16	Expressed gene	+	113,332–116,939	Expressed gene with similarity to At4g24530, protein contains domain from Pfam PF03138	Yes	Yes
EVG.17	Expressed gene	−	123,887–118,564	Similar to mechano-sensitive ion channel in *E. coli*; *Prunus* EST match DY653679	Yes	Yes
EVG.18	Predicted gene	−	129,469–125,672	No similarity in database	No	n.d.
EVG.19	Predicted gene	+	131,386–131,396	Similar to At5g60930 from Arabidopsis	No	n.d.

Region values refer to the position within the sequence. Region intervals are from transcription start site to transcription end site, if known, or the translation start codon to translation stop codon, if not known. Expression data is based on attempted amplification from pooled first strand cDNA from multiple tissues and environmental conditions for each the wild-type and mutant genotypes. Amplification products were sequenced to verify identity.
n.d. Not determined

predicted intron–exon junctions from pooled first strand cDNA from RNA samples of wild-type plants.

The EVG.17 predicted amino acid sequence was similar to a mechano-sensitive ion channel protein from *Escherichia coli*. A peach fruit EST (DY653679) from the NCBI database matched the exact sequence of EVG.17. Expression of EVG.17 in the wild-type samples from the mapping population was confirmed by isolation of partial cDNA spanning multiple predicted intron–exon junctions from pooled first strand cDNA samples of RNA from wild-type plants.

EVG.18 was identified as a potential ORF by gene-prediction programs. The predicted sequence had no similarity to any genes or proteins in the public databases. Expression of EVG.18 could not be confirmed by RT-PCR of cDNA from wild-type RNA samples. EVG.19 was predicted to encode a putative microtubule-associated kinesin. Expression of EVG.19 could not be confirmed by isolation of cDNA from wild-type samples.

Expression of the predicted genes in the EVG locus

Of the 19 predicted genes in the sequenced *evg* region, 12 were verified as expressed in tissue samples from wild-type trees (Table 3). RT-PCR was repeated for the confirmed genes with pooled cDNA from an *evg* mutant tree as a template to assay transcriptional activity of the confirmed genes in the mutant locus. RNA was extracted and first strand cDNA was synthesized from tissues of trees growing in identical conditions to that of the wild-type trees used for confirmation of the predicted genes above (see "Materials and methods" for details). Amplification products were

Fig. 3 ClustalW(1.83) alignment of conceptual translation of the six MADS-box genes located in the EVG region

```
DAM1    -MKMTREKIK IKKIDNLPAR QVTFSKRRRG IFKKAAELSV LCESEVAVVI FSATGKLFDY
DAM2    MVKTMRKKIK IKKIDYLPAR QVTFSKRRRG IFKKAAELSV LCESEVAVVI FSATGKLFDY
DAM3    MVKMMRKKIK IKKIDYLPAR QVTFSKRRRG IFKKAAELSV LCESKVAVVI FSATGKLFDY
DAM4    MVKMMREKIK IKKIDYLPAR QVTFSKRRRG IFKKAAELSV LCESEVAVVI FSATGKLFDY
DAM5    ---MMRNKIK IKKIDYLPAR QVTFSKRRRG LFKKAAELSV LCESEVAVVI FSATGKLFDY
DAM6    ---MMREKIK IKKIDYLPAR QVTFSKRRRG LFKKAAELSV LCESEVAVII FSATDKLFDY

DAM1    SSSSMKDVIE RYQEHINGAE KFDEPSIELQ PEKENHIRLS KELEEKSRQL RQMKGEDLEE
DAM2    SSSSTKDVVE RYQAHTNGVE KSDEPSVELQ LEIENHIRLT KELEEKSCQL RQIKGEDLEE
DAM3    SSSSIKDVIE RYKAHTNGVE KSDKPSVELQ LENENQIGLS KELKEKSHQL RQMKAEDLEE
DAM4    SSSSIKDVIE RYEVRTNGVE KSDEQSLELQ LENENHTKLS TELEEKNRQL RQMKGEDLEE
DAM5    SSSSTKDVIE RYNADINGVE KLNNQEIELQ LENENHIKLS KELEEKSRQL RQMKGEDLEG
DAM6    SSSSTEDVIE RYKAHTNDLE KSNKQFLELQ LENENHIKLS KELEEKSRQL RQMKGEDLQG

DAM1    LNFDELQKLE QLVDASLGRV IETK------ ---------- ---------- -------MLS
DAM2    LNFDELQKLE QLVDASLGRV IETEEELIMS EIMALERKGA ELVEANNQ-- -LRQRMVMLS
DAM3    LNFDELQKLE QLVDASLGRV IETKEELRMS EIMALERKGA ELVEANNQ-- -LRQ-TMMLS
DAM4    LDLDELLKLE QLVEATLVRV METKEELIMS DIVALEKKGT ELVEANNQMV MLRERMVMLS
DAM5    LNLDELLKLE QLVEASLGRV METKEELIKS EIMALERKGT ELVEANNQ-- -LRQTMVMLS
DAM6    LNMDELLKLE QLVEASLGRV IETKEELIMS EIMALEKKGA ELVEANNQ-- -LRQKMAMLS

DAM1    RGNIGPALME PERLNNNIGG GGEEEGMSSE SATS---TTC NSAPSLS--L EDDSDD-VTL
DAM2    RGNIGPAPTE PERFVNNIGG GGEE-GMSSE SATNATISSC SSGPSLS--L EDDCSD-VTL
DAM3    GGNTGPTLME PERLSNNIGG GGEEEGMSSE SAIS---TTC NSALSLSPSL GDDSDD-VTL
DAM4    KRNTGPALME P--------- --------SE SATS---TSC NSALSLS--L EDDCSDDVVL
DAM5    GGNTGPALMD PERLNNNIEG GGEEEGMSAE S---AISTTC NSAVSLS--L EDDSSDEVTL
DAM6    GGNTGPAFVE PETLITNVGG GGEEDGMSSE SAIIATSTSC NSAHSLS--L EDDCSD-VTL

DAM1    SLKLGLP--- ---------
DAM2    ALKLGLP--- ---------
DAM3    SLKLGLS--- ---------
DAM4    SLKLGLTVRA GRRPMCLKT
DAM5    SLKLGR---- ---------
DAM6    SLKLGLP--- ---------
```

cloned and sequenced to verify that the products were from the sequenced locus and not spurious amplification of gene family members from other loci in the genome. Five of the confirmed genes in the region (EVG.1, EVG.2, EVG.3, EVG.16, EVG.17) were also expressed in the mutant tree samples and no differences were seen in the size of the amplification products between wild-type and mutant template sources (Table 3). All six of the MADS-box transcription factor genes in the region are not expressed in cDNA samples prepared from RNA collected from mutant trees (Fig. 4). EVG.10, the putative retrotransposon sequence, was also not expressed in the mutant tissues examined (data not shown).

Cloning of deletion present in the *evg* mutant

A hybridization strategy was used to identify the approximate borders of the deletion previously hypothesized to exist in the mutant locus (Bielenberg et al. 2004). Starting from one end of the 132 kb sequenced region, individual *Hin*dIII digestion fragments of the BAC DNA were labeled as probes and hybridized with size-separated, HindIII-digested wild-type and mutant genomic DNA which had been transferred to a nylon membrane. Southern hybridization results were scored for fragment size polymorphisms between the wild-type and mutant DNA. Beginning from the SP6 end of PpN089G02, a single hybridizing band of

Table 4 Ten most similar matches to DAM1 (EVG.8 in Table 2) resulting from a tBLASTx similarity search against the NCBI non-redundant (nr) database (date of search: July 31, 2006)

Rank	E value	Species	Gene	Accession (GenBank)	Rank for DAM2	Rank for DAM3	Rank for DAM4	Rank for DAM5	Rank for DAM6
1	$3\times e^{-54}$	*Populus tomentosa*	PtMADS1	AY501392	1	1	1	1	1
2	$3\times e^{-48}$	*Ipomoea batatas*	IbMADS3	AB054255	7	4	12	3	9
3	$5\times e^{-54}$	*Ipomoea batatas*	IbMADS4	AB054256	3	2	2	2	2
4	$2\times e^{-47}$	*Eucalyptus occidentalis*	SVP-like	AY273873	4	5	13	7	6
5	$2\times e^{-47}$	*Eucalyptus grandis*	SVP-like	AY263809	6	6	16	9	10
6	$9\times e^{-47}$	*Petunia x hybrida*	FBP13	AF335237	2	3	3	4	3
7	$3\times e^{-46}$	*Malus x domestica*	JOINTLESS	DQ402055	18	14	17	15	15
8	$4\times e^{-46}$	*Arabidopsis thaliana*	AGL24	NM _118587	16	18	9	20	18
9	$3\times e^{-45}$	*Physalis pubescens*	MPF2	AY356367	10	8	4	5	7
10	$5\times e^{-45}$	*Arabidopsis thaliana*	SVP	NM_127820	17	17	11	14	17

The e-value, species, gene name, and GenBank accession number are listed for each result. The numerical rank of each of the DAM1 top ten results in a similar search for each of the other MADS-box proteins in the EVG region is shown in the last five columns.

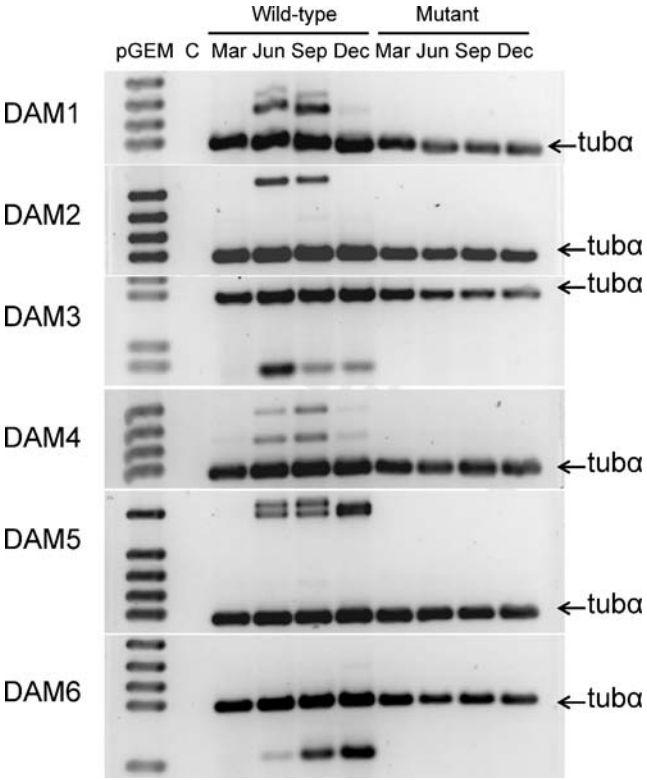

Fig. 4 RT-PCR amplifications of DAM1 (EVG.8), DAM2 (EVG.9), DAM3 (EVG.12), DAM4 (EVG.13), DAM5 (EVG.14), and DAM6 (EVG.15) from wild-type and mutant peach cDNAs from four points during the annual growth cycle. The cDNA used for amplification were the terminal ends of stems containing developing leaves, stems, and the meristem during the growth cycle or stem and bud tissue during dormancy. pGEM marker is used in each gel. Amplification of a peach tubulin-α gene (DY650410) was performed in the same reaction tube as the target gene for a positive control of cDNA isolation and template quantity

equal size was present in the wild-type and the mutant genomic DNAs with each probe until the inverted repeat-containing region at approximately 33,000 to 45,000 bp. Probes from within the repeat region hybridized to multiple fragments in both wild-type and mutant genomes, and there was no interpretable fragment size polymorphisms between wild-type and mutant DNA. Probes from the region flanking the inverted repeat segments (>45,000 bp) produced one clear band in the wild-type and none in the mutant HindIII-digested DNA. Probe 18HB09 (position 87,262–93,415) showed a clear fragment length polymorphism between the wild-type (6,154 bp restriction fragment) and the mutant (4,090 bp restriction fragment).

The 4,090 bp mutant genomic fragment was isolated by inverse PCR. HindIII-digested mutant genomic DNA was circularized by self-ligation and used as a template for PCR amplification. The resulting fragment was cloned and sequenced. The amplified product was 2,939 bp in size and contained the breakpoints of a large deletion in the mutant which resulted in a gap of 41,746 bp of sequence relative to the wild-type sequence beginning at base 49,620 and ending with base 91,365 (Fig. 5). The deleted sequence was replaced by four pyrimidine bases connecting the mutation breakpoints (Fig. 5).

Discussion

We used a genetic mapping approach to identify a region of genomic sequence responsible for terminal bud formation in response to winter dormancy-inducing conditions. Identification of genomic sequence responsible for specific traits is rare in trees (Yang et al. 2003; Peace et al. 2005; Rai 2006). Genetic mapping projects in trees are limited primarily by the expense and time required to maintain sufficiently large mapping populations over many years, long time spans before phenotypic traits can be evaluated, and the large size of the mature sporophytes. Genetic limitations to mapping include large genome sizes, self-incompatibility, dioecious life history, and the quantitative nature of many of the key traits of interest.

We took advantage of a perennial dormancy mutant in peach [*P. persica* (L.) Batsch] to identify and sequence a region of the genome responsible for terminal bud formation. Peach has a number of advantageous characteristics for a mapping approach in a tree species: the genome is small (≈ 240 Mb), the generation time is relatively short (2–4 years), peach is diploid ($2n$=16), self-fertile, monoecious individuals can be rapidly multiplied by vegetative propagation, and there are a number of mutants and many cultivated varieties with well-described phenotypic variation (Scorza and Sherman 1996). Additionally, public molecular genetic resources for experiments with peach (and closely related *Prunus* species) are abundant, including >40,000 EST sequences, several BAC genomic librar-

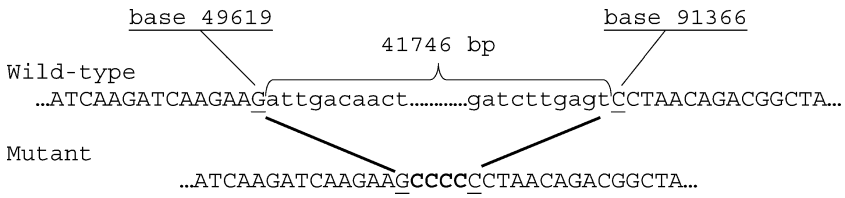

Fig. 5 Sequence bordering the deletion in the *evg* peach mutant. The four cytosine bases in *bold* are not found in the wild-type sequence. *Underlined bases* represent the left and right border of the deletion in the wild-type (*upper sequence*) and mutant (*lower sequence*)

ies, multiple genetic maps anchored to a common reference genetic map, and a nearly complete physical map with anchored genetic markers and ESTs (Jung et al. 2004; Horn et al. 2005).

Repetitive elements

Several putative repetitive elements were identified in the sequenced region. A cluster of inverted repeats located between 3,300 and 6,500 bp indicate possible multiple insertions at a single locus over evolutionary time (Table 2; Fig. 2). There are no protein-coding domains associated with the inverted repeats, and, thus, these are likely non-autonomous transposable elements. Another group of three inverted repeats located between 33,000 bp and 45,000 bp in the sequenced region indicates another site of preference for insertion. In this group, there are two pairs of nearly identical exons, one coding for a transposase and the other coding for a putative DNA-binding domain. We have characterized these sequences as belonging to the Harbinger group of transposable elements based on similarity of the predicted proteins and the presence of the two coding sequences (Jurka et al. 2005).

EVG.10 is a putative polyprotein, although the exon would represent only a fraction of the complete protein-coding region of an autonomous retroelement (Kumar and Bennetzen 1999). The truncated protein and lack of a definable LTR indicates that this element belongs to the non-autonomous class of type I repetitive elements defined as short interspersed repetitive elements or miniature inverted repeat transposable elements (Kumar and Bennetzen 1999).

The close proximity of some of these repetitive elements to the transcription start site of genes such as EVG.8 (DAM1) raises the question of whether they may have a role in the regulation of expression of EVG.8 (and further downstream genes). However, that determination will have to await more detailed functional analysis. It is also tempting to speculate that the apparent activity of mobile elements in or near the deleted region may have a causal connection to the deletion in the mutant genome, particularly the activity of the retroelement (EVG.10), which is missing in the *evg* mutant.

The EVG locus contains multiple MADS-box genes

The sequence of the EVG locus contains six highly similar MADS-box transcription factors in a tandem arrangement (Fig. 2). To our knowledge, this is the largest grouping of tandemly arranged MADS-box genes yet identified in any species. Arabidopsis contains one grouping of four MADS-box genes arranged in tandem (MAF2, MAF3, MAF4, and MAF5) located on chromosome V (Ratcliffe et al. 2003). Several examples of pairs of MADS-box genes in a tandem arrangement have been described, but larger groups appear to be rare, although this may simply be a function of limited genome data (Theissen et al. 1995; Vrebalov et al. 2002; Parenicova et al. 2003; Moore et al. 2005). In evolutionary time, duplicated genes are thought to either lose function, develop new function(s) that diverge from the original gene, or develop regulatory specificity separate from the original gene (Moore et al. 2005; Duarte et al. 2006). Because all six of the MADS-box genes described here are expressed and do not contain sequence alterations that would result in truncated proteins or a frameshift in translation, loss of function in the duplicated genes does not appear to be the case.

All the MADS-box genes in the EVG locus belong to the general clade of MIKC-type MADS-box genes represented by the StMADS11/AGL24/AtSVP group of MADS-box transcription factors. Genes in this group are primarily expressed in vegetative tissues such as roots, stems, leaves and meristems, and regulate vegetative growth or the transition from vegetative to reproductive meristems (Alvarez-Buylla et al. 2000; Garcia-Maroto et al. 2003). Interestingly, one of the genes with the highest similarity to the MADS-box genes we describe here (IbMADS4; Table 3) was demonstrated to promote shoot vegetative growth when ectopically overexpressed in chrysanthemum (Aswath et al. 2004). Additionally, putative orthologues of AtAGL24/AtSVP have been isolated from tree species such as *Paulownia kawakamii* (PkMADS) and *Eucalyptus grandis* (EgSVP; Prakash and Kumar 2002; Brill and Watson 2004). When these genes are ectopically overexpressed in the original species (*P. kawakamii*) or in Arabidopsis (*E. grandis*), precocious growth of lateral meristems is one of the observed phenotypes. This contrasts with the results we present here, where the continuous growth phenotype of *evg* peach appears to result from a loss of expression of the MADS genes in the locus. A more exhaustive survey and functional characterization of vegetative development associated MADS-box genes from peach and other species will be required to determine if these genes are antagonistic to one another in regulating shoot proliferative growth.

Our initial analysis suggests that the six MADS-box genes we have sequenced have seasonally dependent gene expression. None of the six genes are expressed prior to budbreak in the spring (Fig. 4). DAM1, DAM2, DAM3, and DAM4 appear to be expressed primarily in the growing season (Fig. 4). DAM5 and DAM6 have an expression pattern which is strongest in the winter season during induction and maintenance of dormancy. These two distinct patterns suggest that there has been some regulatory divergence of the six genes, despite their striking coding-sequence similarity. Sub- or neo-functionalization of duplicated genes can occur through regulatory divergence and is thought to be one of the evolutionary fates of copies of duplicated genes (Lawton-Rauh 2003; Moore and Purugganan 2005; Duarte

et al. 2006). The duplicated genes at this locus do not appear to have become nonfunctional (as judged by continued expression of mRNA which are predicted to translate to a full-length protein), the most common fate of duplicated genes (Moore and Purugganan 2005). The tandem duplication of multiple MADS-box genes in this region may represent an interesting case study for examining the dynamics of duplicate gene evolution.

The *evg* mutation is a deletion

We previously characterized the *evg* mutation as a deletion affecting the MADS-box domain of four of the six clustered MADS-box genes in the EVG locus (Bielenberg et al. 2004). This characterization was based on Southern hybridizations of the wild-type and mutant genomic DNA, which clearly showed that only two of the MADS-box domains were retained in the mutant genomic DNA. Here we have cloned and sequenced the apparent break points of the mutant locus. Our data support the initial characterization of the *evg* mutant as a deletion affecting a tandem array of similar MADS-box genes.

While the mutant is only missing four deleted genes (EVG.8, EVG.9, EVG.12, and EVG.13), our expression analysis (Fig. 4) shows that the mutation eliminates expression of all six MADS-box genes. The loss of transcriptional activity in all of the MADS-box genes suggests either that the deletion results in a disruption of the accessibility of the local region to transcriptional machinery or that sequence(s) in the deleted region plays a critical regulatory role in the expression of EVG.14 and EVG.15. The continued expression of the nearby EVG.16 may support the latter hypothesis because a structural impediment to transcription would not likely discriminate on such a fine scale. Additionally, a relatively short distance (<300 bp) separates the transcription stop site of EVG.13 and the transcription start site of EVG.14, which may indicate that additional sequence overlapping or preceding EVG.13 is involved in regulating EVG.14.

Gene candidates for *evg*

The ultimate goal of genetic mapping is to identify gene candidates in the mapped and sequenced region for the trait of interest. One of our flanking markers (pchgms41, Fig. 1) maps within 2 cM of the *evg* mutation; however, we did not identify a flanking marker on the opposite side of *evg* that was closer than 7 cM. In theory, any of the gene sequences found in the 9 cM interval between the two flanking markers (pchgms41 and CPPCT029) may be responsible for the *evg* mutation. Despite what appears to be relatively short physical distances per cM of genetic distance in the region containing pchgms41, pchgms12, and pchgms29 (10 to 35 kb/cM), a sequence walk of >95 kbp across the EVG locus failed to reach the flanking genetic marker CPPCT029 (Figs. 1 and 2). However, pchgms40 is located within the deleted sequence found in the mutant plants and pchgms40 cosegregates with the *evg* mutant phenotype. Therefore, although the full complement of candidate genes for the *evg* mutation is not yet known, we are hypothesizing that the deletion located in the mutant genome which cosegregates with the phenotypic trait is responsible.

The fine-mapping of the EVG locus was initially performed without markers of known genomic position on the *Prunus* reference map (Wang et al. 2002a). Large scale chromosomal rearrangements in a mapping population such as a translocation can also cause positional effects in genes outside of the BACs identified by STS hybridization. These positional effects at the site of translocation may result in loss of function and ultimately be responsible for the mutant phenotype. We tested the possibility that a chromosomal translocation may have occurred in the *evg* germplasm by using genetic markers of known positions on a *Prunus* reference genetic map (Fig. 1; Joobeur et al. 1998; Dirlewanger et al. 2004). The anchored markers from LG1 of *Prunus* span the EVG locus and are in the same relative positioning in our local map as previously found on the reference map; therefore, it is unlikely that there has been a chromosomal rearrangement.

We hypothesize that one or more of the MADS genes found in the sequenced region are responsible for the non-dormant phenotype of the *evg* mutant. The genetic interval to which the *evg* phenotype was mapped contains a conspicuous genomic deletion in the mutant. This deletion eliminates expression of six transcription factors of a class known to regulate vegetative development. Expression of other genes in the sequenced region is not eliminated in the mutant, although these genes may have indels in the mutant that alter their function. Until functional studies such as reverse genetics experiments are conducted, a definitive association of these six genes with the *evg* phenotype cannot be established. Studies of the tissue specificity and seasonal regulation of the six MADS-box candidate genes should provide further evidence for or against their involvement in the *evg* phenotype.

Acknowledgements The authors would like to thank Dr. M. Staton for valuable assistance with the bioinformatics and Dr. W.V. Baird for valuable comments on the manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Alvarez-Buylla ER, Liljegren SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, Vergara-Silva F, Yanofsky MF (2000) MADS-box gene evolution

beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. Plant J 24:457–466

Anderson JV, Chao WS, Horvath DP (2001) A current review on the regulation of dormancy in vegetative buds. Weed Sci 49:581–589

Anderson JV, Gesch RW, Jia Y, Chao WS, Horvath DP (2005) Seasonal shifts in dormancy status, carbohydrate metabolism, and related gene expression in crown buds of leafy spurge. Plant Cell Environ 28:1567–1578

Aranzana MJ, Garcia-Mas J, Carbo J, Arus P (2002) Development and variability analysis of microsatellite markers in peach. Plant Breed 121:87–92

Aswath CR, Mo SY, Kim SH, Kim DH (2004) IbMADS4 regulates the vegetative shoot development in transgenic chrysanthemum (Dendrathema grandiflora (Ramat.) Kitamura). Plant Sci 166:847–854

Barak S, Tobin EM, Andronis C, Sugano S, Green RM (2000) All in good time: the Arabidopsis circadian clock. Trends Plant Sci 5:517–522

Besemer J, Borodovsky M (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33:W451–W454

Bielenberg DG, Wang Y, Fan S, Reighard GL, Scorza R, Abbott AG (2004) A deletion affecting several gene candidates is present in the Evergrowing peach mutant. J Heredity 95:436–444

Bohlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O (2006) CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. Science 312:1040–1043

Brill EM, Watson JM (2004) Ectopic expression of a Eucalyptus grandis SVP orthologue alters the flowering time of Arabidopsis thaliana. Funct Plant Biol 3::217–224

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. J Mol Biol 268:78–94

Cantini C, Iezzoni AF, Lamboy WF, Boritzki M, Struss D (2001) DNA fingerprinting of tetraploid cherry germplasm using simple sequence repeats. J Am Soc Hortic Sci 126:205–209

Cipriani G, Lot G, Huang WG, Marrazzo MT, Peterlunger E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L) Batsch]: isolation, characterisation and cross-species amplification in Prunus. Theor Appl Genet 99:65–72

Dennis FG (1996) A physiological compasion of seed and bud dormancy. In: Lang GA (ed) Plant Dormancy: Physiology, Biochemistry, and Molecular Biology. CAB, New York, pp 47–56

Dirlewanger E, Cosson P, Tavaud M, Aranzana MJ, Poizat C, Zanetto A, Arus P, Laigret F (2002) Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (Prunus avium L.). Theor Appl Genet 105:127–138

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Nat Acad Sci USA 101:9891–9896

Duarte JM, Cui LY, Wall PK, Zhang Q, Zhang XH, Leebens-Mack J, Ma H, Altman N, dePamphilis CW (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol 23:469–478

Faust M, Erez A, Rowland LJ, Wang SY, Norman HA (1997) Bud dormancy in perennial fruit trees: physiological basis for dormancy induction, maintenance, and release. Hortscience 32:623–629

Fuchigami LH, Wisniewski M (1997) Quantifying bud dormancy: physiological approaches. Hortscience 32:618–623

Garcia-Maroto F, Carmona MJ, Garrido JA, Vilches-Ferron M, Rodriguez-Ruiz J, Alonso DL (2003) New roles for MADS-box genes in higher plants. Biologia Plantarum 46:321–330

Georgi LL, Wang Y, Yvergniaux D, Ormsbee T, Inigo M, Reighard G, Abbott AG (2002) Construction of a BAC library and its application to the identification of simple sequence repeats in peach *Prunus persica* (L.) Batsch. Theor Appl Genet 105:1151–1158

Gevaudant F, Samson I, Guilliot A, Petel G (1999) An improved method for isolating polyphenol-free RNA from woody plant tissues. J Trace Microprobe Tech 17:445–450

Gevaudant F, Petel G, Guilliot A (2001) Differential expression of four members of the H + -ATPase gene family during dormancy of vegetative buds of peach trees. Planta 212:619–626

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. Nature Genet 3:266–272

Horn R (2005) Candidate gene database and transcript map for peach, a model species for fruit trees. Theor Appl Genet 110:1419–1428

Horvath DP, Anderson JV (2002) A molecular approach to understanding root bud dormancy in leafy spurge. Weed Sci 50:227–231

Horvath DP, Anderson JV, Chao WS, Foley ME (2003) Knowing when to grow: signals regulating bud dormancy. Trends Plant Sci 8:534–540

Joobeur T, Viruel MA, de Vicente MC, Jauregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, Quarta R, Dirlewanger E, Arus P (1998) Construction of a saturated linkage map for Prunus using an almond x peach F-2 progeny. Theor Appl Genet 97:1034–1041

Jung S, Jesudurai C, Staton M, Du ZD, Ficklin S, Cho IH, Abbott A, Tomkins J, Main D (2004) GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. BMC Bioinformatics 5: Art. No. 130

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110::462–467

Klee H, Estelle M (1991) Molecular genetic approaches to plant hormone biology. Ann Rev Plant Physiol Plant Mol Biol 42:529–551

Koornneef M, Bentsink L, Hilhorst H (2002) Seed dormancy and germination. Curr Opin Plant Biol 5:33–36

Kovarík A, Matzke MA, Matzke AJM, Koukalová B (2001) Transposition of IS10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts. Mol Genet Genom 266:216–222

Kozlowski TT, Pallardy SG (1997) In: Physiology of woody plants. 2nd edn. Academic, San Diego

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Ann Rev Genet 33:479–532

Lawton-Rauh A (2003) Evolutionary dynamics of duplicated genes in plants. Mol Phylogenet Evol 29:396–409

Majoros WH, Pertea M, Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics 20:2878–2879

Messina R, Lain O, Marrazzo MT, Cipriani G, Testolin R (2004) New set of microsatellite loci isolated in apricot. Mol Ecol Notes 4:432–434

Michaels SD, Amasino RM (2000) Memories of winter: vernalization and the competence to flower. Plant Cell Environ 23:1145–1153

Molmann JA, Berhanu AT, Stormo SK, Ernstsen A, Junttila O, Olsen JE (2003) Metabolism of gibberellin A(19) is under photoperiodic control in Populus, Salix and Betula, but not in daylength-insensitive Populus overexpressing phytochrome A. Physiologia Plantarum 119:278–286

Molmann JA, Asante DKA, Jensen JB, Krane MN, Ernstsen A, Junttila O, Olsen JE (2005) Low night temperature and inhibition of gibberellin biosynthesis override phytochrome action and induce bud set and cold acclimation, but not dormancy in PHYA overexpressors and wild-type of hybrid aspen. Plant Cell Environ 28:1579–1588

Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol 8:122–128

Moore RC, Grant SR, Purugganan MD (2005) Molecular population genetics of redundant floral-regulatory genes in Arabidopsis thaliana. Mol Biol Evol 22:91–103

Olsen JE (2003) Molecular and Physiological Mechanisms of Bud Dormancy Regulation. In: Tanino KK et al (ed) XXVI International Horticultural Congress-Environmental Stress, Vol 618. ISHS, Leuven, pp 437–453

Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell 15:1538–1551

Peace CP, Crisosto CH, Gradziel TM (2005) Endopolygalacturonase: a candidate gene for freestone and melting flesh in peach. Mol Breed 16:21–31

Prakash AP, Kumar PP (2002) PkMADS1 is a novel MADS box gene regulating adventitious shoot induction and vegetative shoot development in Paulownia kawakamii. Plant J 29:141–151

Rai M (2006) Refinement of the *Citrus tristeza* virus resistance gene (Ctv) positional map in *Poncirus trifoliata* and generation of transgenic grapefruit (*Citrus paradisi*) plant lines with candidate resistance genes in this region. Plant Mol Biol 61:399–414

Ratcliffe OJ, Kumimoto RW, Wong BJ, Riechmann JL (2003) Analysis of the Arabidopsis MADS affecting flowering gene family: MAF2 prevents vernalization by short periods of cold. Plant Cell 15:1159–1169

Rodriguez J, Sherman WB, Scorza R, Wisniewski M, Okie WR (1994) Evergreen peach, its inheritance and dormant behavior. J Am Soc Hortic Sci 119:789–792

Rohde A, Bhalerao RP (2007) Plant dormancy in a perennial context. Trends Plant Sci 12:217–223

Rohde A, Howe GT, Olsen JE, Moritz T, Van Montagu M, Junttila O, Boerjan W (2000) Molecular aspects of bud dormancy in trees. In: Jain SM, Minocha SC (eds) Molecular Biology of Woody Plants, Vol 1. Kluwer, Dordrecht, pp 89–134

Rowland LJ, Ogden EL, Arora R, Lim CC, Lehman JS, Levi A, Panta GR (1999) Use of blueberry to study genetic control of chilling requirement and cold hardiness in woody perennials. Hortscience 34:1185–1191

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana, Totowa, pp 365–386

Ruonala R, Rinne PLH, Baghour M, Moritz T, Tuominen H, Kangasjarvi J (2006) Transitions in the functioning of the shoot apical meristem in birch (*Betula pendula*) involve ethylene. Plant J 46:628–640

Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res 10:516–522

Sambrook J, Russell DW (2002) In: Molecular cloning: a laboratory manual Vol 2. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Scorza R, Sherman WB (1996) Peaches. In: Janick J, Moore JN (eds) Fruit Breeding, Vol 1 Tree and tropical fruits. Wiley, New York, pp 325–440

Theissen G, Strater T, Fischer A, Saedler H (1995) Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of agamous-like MADS-box genes from maize. Gene 156:155–166

Thompson MM, Smith DC, Burgess JE (1985) Non-dormant mutants in a temperate tree species, *Corylus avellana* L. Theor Appl Genet 70:687–692

van der Linden CG, Vosman B, Smulders MJM (2002) Cloning and characterization of four apple MADS-box genes isolated from vegetative tissue. J Exp Bot 53:1025–1036

Van Ooijen JW, Voorrips RE (2001) JoinMap® 3.0, Software for the calculation of genetic linkage maps. Plant Research, Wageningen

Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J (2002) A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (Rin) locus. Science 296:343–346

Wang Y, Georgi LL, Reighard GL, Scorza R, Abbott AG (2002a) Genetic mapping of the evergrowing gene in peach *Prunus persica* (L.) Batsch. J Heredity 93:352–358

Wang Y, Georgi LL, Zhebentyayeva TN, Reighard GL, Scorza R, Abbott AG (2002b) High-throughput targeted SSR marker development in peach (*Prunus persica*). Genome 45:319–328

Werner DJ, Okie WR (1998) A history and description of the *Prunus persica* plant introduction collection. Hortscience 33:787–793

Yang ZN, Ye XR, Molina J, Roose ML, Mirkov TE (2003) Sequence analysis of a 282-kilobase region surrounding the Citrus tristeza virus resistance gene (Ctv) locus in *Poncirus trifoliata* L. Raf. Plant Physiology 131:482–492