BAC Library Development and Clone Characterization for Dormancy-Responsive *DREB4A*, *DAM*, and *FT* from Leafy Spurge (*Euphorbia esula*) Identifies Differential Splicing and Conserved Promoter Motifs

Author(s): David P. Horvath , David Kudrna , Jayson Talag , James V. Anderson , Wun S. Chao , Rod Wing , Michael E. Foley , and Münevver Doğramacı
Source: Weed Science, 61(2):303-309. 2013.
Published By: Weed Science Society of America
DOI: http://dx.doi.org/10.1614/WS-D-12-00175.1
URL: http://www.bioone.org/doi/full/10.1614/WS-D-12-00175.1

Weed Science 2013 61:303–309

BAC Library Development and Clone Characterization for Dormancy-Responsive *DREB4A*, *DAM*, and *FT* from Leafy Spurge (*Euphorbia esula*) Identifies Differential Splicing and Conserved Promoter Motifs

David P. Horvath, David Kudrna, Jayson Talag, James V. Anderson, Wun S. Chao, Rod Wing, Michael E. Foley, and Münevver Doğramacı*

We developed two leafy spurge bacterial artificial chromosome (BAC) libraries that together represent approximately 5× coverage of the leafy spurge genome. The BAC libraries have an average insert size of approximately 143 kb, and copies of the library and filters for hybridization-based screening are publicly available through the Arizona Genomics Institute. These libraries were used to clone full-length genomic copies of an *AP2/ERF* transcription factor of the A4 subfamily of *DEHYDRATION-RESPONSIVE ELEMENT-BINDING PROTEINS (DREB)* known to be differentially expressed in crown buds of leafy spurge during endodormancy, a *DORMANCY ASSOCIATED MADS-BOX* (*DAM*) gene, and several *FLOWERING LOCUS T* (*FT*) genes. Sequencing of these BAC clones revealed the presence of multiple *FT* genes in leafy spurge. Sequencing also provided evidence that two different *DAM* transcripts expressed in crown buds of leafy spurge during endo- and eco-dormancy result from alternate splicing of a single *DAM* gene. Sequence data from the *FT* promoters was used to identify several conserved elements previously recognized in Arabidopsis, as well as potential novel transcription factor binding sites that may regulate *FT*. These leafy spurge BAC libraries represent a new genomics-based tool that complements existing genomics resources for the study of plant growth and development in this model perennial weed. Furthermore, phylogenetic footprinting using genes identified with this resource demonstrate the usefulness of studying weedy species to further our general knowledge of agriculturally important genes.

Nomenclature: Arabidopsis, *Arabidopsis thaliana* L. ARATH; leafy spurge, *Euphorbia esula* L. EPHES.

Key words: Alternate splicing, flowering, phylogenetic footprinting, transcription factors.

Characterization of genomic DNA sequence from an organism is crucial to understanding its biology, evolution, and gene expression regulatory networks. Genomic sequence data can serve as a scaffold for next-generation sequencing that will allow for detailed transcriptome and epigenetic analyses. Whole genome sequencing, in model plants and crops, is invaluable for understanding plant development, physiology, and interactions with the environment. Until recently, whole genome sequencing had been out of reach for most weed biologists and thus weed scientists have not yet fully utilized these tools to answer fundamental questions such as these: How does herbicide resistance evolve and what genes make a given plant invasive or contribute to its weedy characteristics?

Leafy spurge is an auto-allo hexaploid (Schulz-Schaeffer and Gerhardt 1989; Stahevitch et al. 1988), and is one of the most genomically characterized invasive weeds (Horvath 2009a). Sanger sequencing of a normalized cDNA library derived from all plant tissues exposed to a range of environmental conditions resulted in a leafy spurge expression sequence tag (EST) database containing > 23,000 unique sequences representing > 19,000 unigenes (Anderson et al. 2007). The EST database was further leveraged to produce a cDNA microarray that contained all sequenced leafy spurge unigenes. These microarrays have been used in numerous studies to identify differentially expressed genes during dormancy transitions of both underground adventitious buds (crown and root buds) and seeds of leafy spurge, as well as genes involved in drought responses and invasiveness (Chao et al. 2011; Doğramacı et al. 2011; Foley et al. 2010; Horvath et al. 2008). More importantly, bioinformatics analyses of the microarray data identified numerous differentially expressed genes that could be grouped into coordinately regulated gene clusters. These clusters provide suitable starting points for identifying common transcription factor binding sites among the coordinately regulated genes, which could serve as potential targets for future manipulation of plant growth and development (Anderson 2008; Stewart et al. 2009).

To identify and characterize sequences of promoters and other regulatory regions required for gene expression, it is necessary to obtain genomic sequence covering the 3′ and 5′ flanking regions of transcribed leafy spurge genes. Numerous methods have been used to obtain genomic sequence data from plants, including large-scale Sanger shotgun sequencing approaches combined with use of ordered BAC libraries and next-generation sequencing approaches such as 454 FLX and Illumina, capable of generating multiple gigabases of sequence data relatively inexpensively. Although early use of next-generation technologies for genome sequencing was hampered by the short read lengths, advancements have made it possible to generate > 500 bp reads from 454 FLX technologies. Combined with their ability to do paired-end reads from libraries of 1,000 to 10,000 bases in length, next-generation sequencing offers greater capacity and read length than the earlier Sanger technology (Loman et al 2012). Likewise, Illumina technology can now produce upwards of 250 base reads, also with paired-end capabilities. Additionally, new third-generation technologies such as those developed by Pacific Biosciences Inc. (often referred to as PacBio), and nanopore technology can produce reads of over 20,000 bases, albeit with a relatively high error rate (Henson et al. 2012). However, computer programs have been developed to correct these long-read PacBio sequences using the shorter, but more accurate, Illumina or 454 FLX sequences (Koren et al. 2012).

Flowering and vegetative reproduction contribute to seed and bud bank capacity and make leafy spurge a particularly

DOI: 10.1614/WS-D-12-00175.1

* First, fourth, fifth, seventh and eight authors: Research Scientists, and Research Leader, Sunflower and Plant Biology Research Unit, Red River Valley Agricultural Research Center, U.S. Department of Agriculture–Agricultural Research Service, Fargo, ND 58102; second, third, and sixth authors: Group Leader, BAC/EST Resource Center Staff, and Director, respectively, Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721. Corresponding author's E-mail: david.horvath@ars.usda.gov

difficult weed to control. Understanding the molecular basis of sexual and asexual reproduction is critical for developing novel practices and compounds needed to control leafy spurge, as well as other tenacious perennial weeds. Dormancy in vegetative shoot buds allows many perennials such as leafy spurge to survive winter and escape conventional control measures. One class of genes that appears to play a significant role in bud dormancy in perennials is the *DAM* genes (Horvath 2009b). It has been demonstrated that *DAM* genes are induced by low temperatures (Horvath et al. 2010), which are likely regulated by AP2/ERF transcription factor subfamily members known as DREB/C-REPEAT BINDING FACTOR (CBF). Previous studies have shown that DREB/CBF transcription factors are involved in cold and ethylene responses (Kahn 2011) and some of these subfamily members can also be regulated or "gated" through the action of circadian regulators such a LONG HYPOCOTYL (LHY) and CIRCADIAN CLOCK–ASSOCIATED1 (CCA1) under short day conditions (Dong et al. 2011). We postulated that induction of *DAM* genes is required to down-regulate the expression of *FT* (Horvath et al. 2008, 2010). In black cottonwood (*Populus trichocarpa* Torr. & Gray), two different *FT* genes separately regulate flowering and seasonal growth cessation and dormancy (Hsu et al. 2011; Rinne et al. 2011). A proposed model for these regulatory circuits is shown in Figure 1. In Arabidopsis, numerous transcription factors likely regulate *FT*, but only a few transcription factor binding sites have been identified (Pin and Nilsson 2012). The availability of genomic sequences from several plant species opens the possibility of utilizing phylogenetic footprinting (looking for short sequences within the promoters of orthologous genes that are conserved across species) to identify important transcription factor binding sites (Blanchette et al. 2002). BAC libraries containing large stretches of contiguous DNA can facilitate identification of promoter sequences of genes of interest as well as linked genes that can provide evidence of microsynteny to assist in identification of orthologous genes.

Here we describe the development and initial characterization of a leafy spurge BAC library, and use of this library to identify and characterize a new *DAM* gene, a novel *TINY-like DREB* transcription factor, and several *FT* genes that may play a role in seasonal dormancy of underground adventitious shoot buds of leafy spurge.

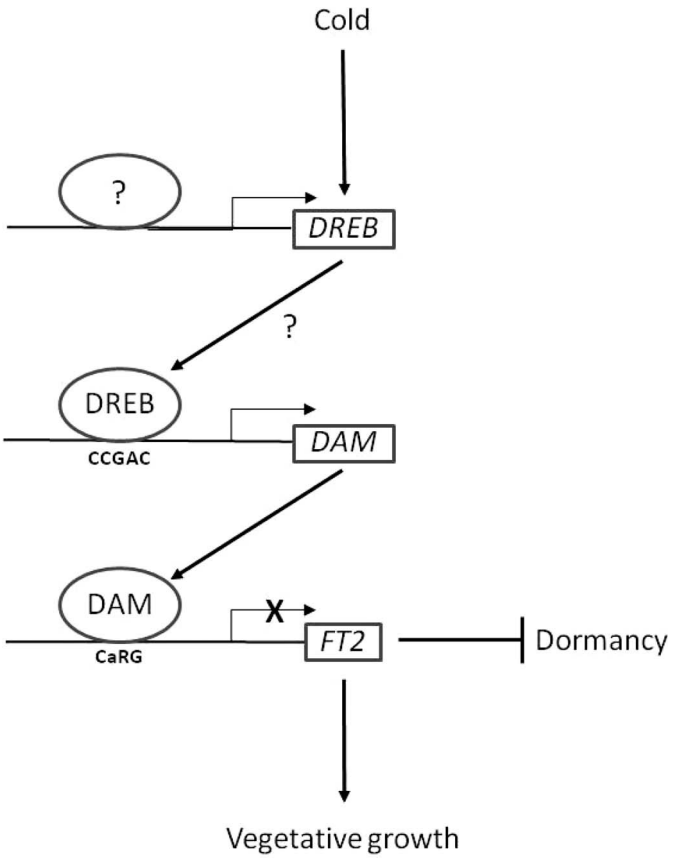

Figure 1. Model for how dormancy is regulated by cold through the action of transcription factors including DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN (DREB) which may positively regulate *DORMANCY ASSOCIATED MADS-BOX* (*DAM*). DAM is supected of binding to and inhibiting the expression of the *FLOWERING LOCUS T2* (*FT2*) gene (Horvath et al. 2010). Expression of *FT2* has been shown to inhibit endodormancy induction and positively regulate vegetative growth (Hsu et al. 2011). Question marks represent suspected factors and interactions requiring further investigation. The X through the arrow represents transcriptional repression.

Materials and Methods

Plant Material. Three- to four-month-old leafy spurge (cultivar 1984-ND001) plants were placed in a dark growth chamber at approximately 25 C for 7 d, and etiolated shoot tips from them were collected and frozen in liquid nitrogen. Prior to harvest, all plants were grown from shoot cuttings in Sunshine (Fisons Horticulture Inc., Bellevue, WA) mix in 3.8 cm by 15 cm cone-tainers (Stuewe and Sons, Corvallis, OR) under greenhouse conditions (16 h natural light plus mix of halogen and incandescent light at approximately 25 C with daily watering). Etiolated shoot tips were harvested into liquid N_2 and shipped to the Arizona Genomics Institute.

BAC Library Construction and Analysis. BAC libraries were prepared from etiolated shoot tip material using methods previously described (Luo and Wing, 2003). Two separate libraries were constructed (one following partial digestion of the leafy spurge DNA with *Hind*III and another following partial digestion with *Bst*YI). Following size selection, the DNA was ligated into a pAGIBAC1 BAC vector and transformed into *Escherichia coli* DH10B competent cells (Invitrogen, Grand Island, NY). Colonies were assembled into 384-well barcoded plates using automated robotics. Analysis of BAC-end sequences was done to assess the level of chloroplast DNA in each library. The two libraries are publically available from the Arizona Genomics Institute (http://www.genome.arizona.edu/orders/) and are named EEENBa and EEENBb, for the *Hind*III and *Bst*YI, respectively. Positively charged nylon membranes spotted with the BAC DNAs were prepared by the Arizona Genomics Institute and hybridized to P^{32}-labeled cDNAs representing the *FT*, the *DAM1*, and an endodormancy-specific *DREB*-like transcription factor that is highly up-regulated following endodormancy induction in crown buds of leafy spurge (Doğramacı et al. 2010; Horvath et al. 2008). Hybridizing BAC clones were commercially sequenced (Amplicon Express, Pullman, WA). Libraries of 200 and 400 bases as well as a paired end library were sequenced using Illumina technology to between 20 and 100× coverage. The assembly was accomplished using a proprietary procedure. The resulting

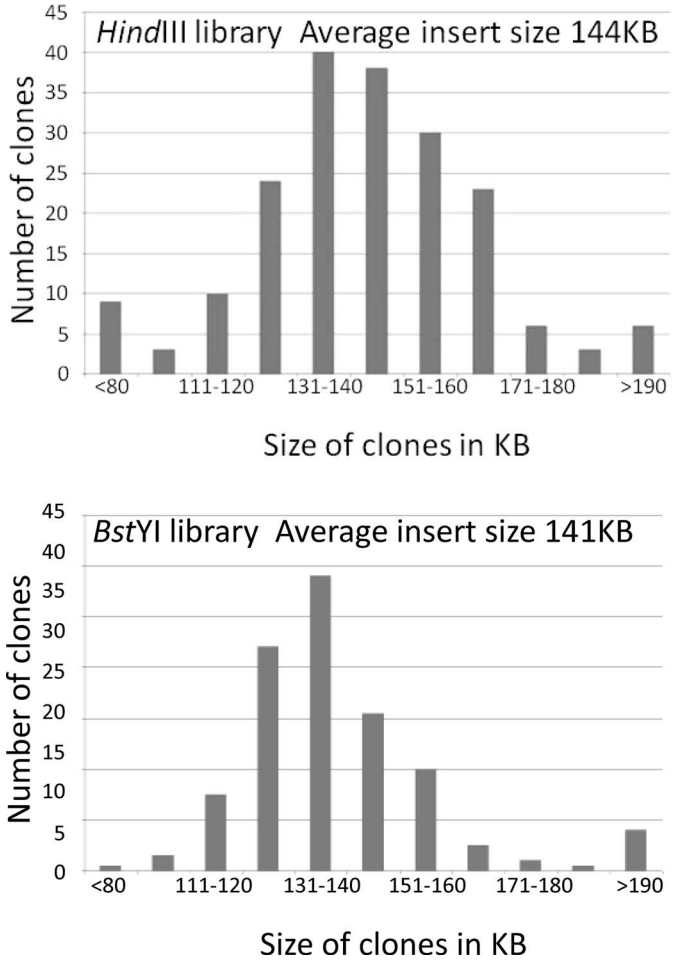

Figure 2. Distribution of insert size following excision of inserts with *Not*I and subsequent size fractionation by pulse field electrophoresis for the *Hind*III library (top) and the *Bst*YI library (bottom).

contigs were examined for *DAM*, *DREB*, and *FT* genes using the DNA Star suite (DNASTAR Inc. Madison, WI) of sequence analysis programs and annotated using previously characterized cDNA clones for these three genes.

Results and Discussion

Construction and Assessment of the BAC Libraries. Two leafy spurge BAC libraries were constructed using two restriction enzymes to reduce the amount of unrepresented genomic DNA resulting from restriction site biases. The *Hind*III library (EEEBa) has an average inset size of 144 kb and contains 36,864 clones. BAC-end sequencing used for quality assessment indicated that less than 2% of this library represents chloroplast DNA (Figure 2). The *Bst*YI library (EEEBb) also contains 36,864 clones, has an average insert size of 141 kb, and represents less than 1% chloroplast DNA (Figure 2). Based on the average insert size and number of clones, we estimate that these BAC libraries represent approximately 5× coverage of the 2.1 Gb leafy spurge genome (Chao et al. 2005). Filters prepared from each library are publicly available from the Arizona Genomics Institute and should prove valuable for characterizing genes of interest from leafy spurge. Likewise, BAC clones should be useful for closing gaps and confirming orientation of contigs resulting from ongoing Illumina sequencing of the leafy spurge genome.

Identification of *FT*-, *DAM*-, and *DREB*-Containing BAC Clones. Hybridization of BAC filters to probes prepared from the 3′ or 5′ ends of the *FT* clone identified several strongly hybridizing clones, as well as many that hybridized weakly to the probe. Likewise, between three and five BAC clones per filter hybridized to a probe designed from the first exon of a leafy spurge *DAM* gene. One *DREB*-hybridizing and two *DAM*-hybridizing clones identified from the *Bst*YI library, and three *FT*-hybridizing BAC clones form the *Hind*III library were subjected to southern blot hybridization to confirm presence of the genes of interest. The BAC clones assembled into between 1 and 15 contigs with all but one assembling into less than six contigs—all with greater than 97% coverage. A comparison of the sequence from the cDNAs representing the *FT*, *DAM*, and *DREB* genes to the sequence obtained from the fully sequenced respectively hybridizing BAC clones confirmed the presence of these sequences in all of the hybridizing clones (Table 1).

Characterization of Leafy Spurge *FT* Genes. The BAC sequences were compared to the previously cloned cDNAs to identify and annotate the transcribed portions of these genes. The *FT* genes from BAC_*FT*-2 and BAC_*FT*-4 contain the full gene sequence with greater than 2,000 bases of promoter sequence. The BAC_*FT*-5 clone contains only a partial *FT* gene that covers 978 bases of the second intron through the remainder of the 3′ region of the gene. Phylogenetic analysis was done on equivalent sequences (977 bases from the 3′ end of exon 2 to 200 bases past the stop codon) from all three *FT* genes isolated from the BAC library as well as an additional partial *FT* gene previously isolated from a lambda genomic

Table 1. Clone name, library source, Genbank accession number, indication as to whether the clone is a full-length or partial representation of the gene, and citations where expression analysis or sequence was previously published.

Clone name	Library	Genbank accession no.	Coverage	Citation
BAC_*DAM3*	BAC	JX966351	Full	
Lambda_*DAMa*	Lambda genomic	JX966357	Partial	Horvath et al. 2010
Lambda_*DAMb*	Lambda genomic	EU334633	Partial	Horvath et al. 2010
*DAM1*_cDNA	plasmid cDNA	DV114890	Full	
*DAM2*_cDNA	plasmid cDNA	EU339320	Full	
BAC_*DREB*	BAC	JX966356	Full	
*DREB*_cDNA	plasmid cDNA	DV146983	Partial	Horvath et al. 2008
BAC_*FT2*	BAC	JX966353	Full	
BAC_*FT4*	BAC	JX966355	Full	
BAC_*FT5*	BAC	JX966354	Partial	
Lambda_*FT10*	Lambda genomic	JX966352	Partial	

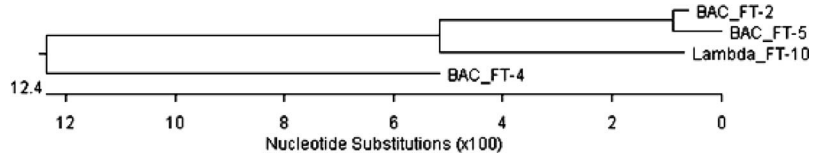

Figure 3. Phylogenetic tree developed following ClustalW alignment of sequences from the four *FT* genes covering 977 bases of intron 2, exon 3, intron 3, exon4, and 200 bases of the 3′ untranslated region (UTR).

library (called Lambda_*FT*-10) (Figure 3). BAC_*FT*-2 and BAC_*FT*-5 are the most similar to each other, and together are most similar to the previously isolated genomic clone called Lambda_*FT*-10. BAC_*FT*-4 is the least similar to the other three clones, but is likely to be a functional *FT* gene based on presence of conserved domains and intron placement (Jeffares et al. 2006). Of the two *FT* genes in black cottonwood, *FT1* appears to be responsible for controlling the transition of the vegetative to floral meristem, whereas *FT2* is responsible for vegetative growth and dormancy (Hsu et al. 2011). These two black cottonwood *FT* genes can be readily distinguished by their expression patterns. *FT1* is induced during extended cold periods when the axillary bud primordia are presumably predetermined to become floral meristems, whereas *FT2* is down-regulated by early season cold resulting in growth cessation and endodormancy induction (Rinne et al. 2011). Research is currently underway to examine the expression patterns of the four *FT* genes from leafy spurge to determine if a similar distribution of expression patterns, and presumably function, occurs.

Characterization of Leafy Spurge *DAM* and *DREB* Genes. Both *DAM*-containing BACs appear to be from identical genomic regions and cover the entire *DAM* gene sequence including more than 5,000 bases of promoter sequence. Thus, we only examined one clone, which we designated as BAC_*DAM3*. Two previously cloned *DAM* gene fragments were isolated from the leafy spurge lambda genomic library designated as lambda_*DAMa* and lambda_*DAMb*. Sequence differences in noncoding regions between these clones clearly indicate that BAC_*DAM3* is different from lambda_*DAMa* and lambda_*DAMb*. However, phylogenetic analysis using 2,869 bases from the longest common sequences between the three clones comprising the promoters of lambda_*DAMa,* lambda_*DAMb*, and BAC_*DAM3* indicates that BAC_*DAM3* is most similar to lambda_*DAMb* (Figure 4). Previously published promoter analyses of the two lambda clones indicated that there were deletions that were unique for each promoter (Horvath et al. 2010). The promoter from BAC_*DAM3* appears to have neither deletion, suggesting that perhaps it is a phylogenetically older version of these genes.

A cDNA, called *DAM1*_cDNA, was previously sequenced from the EST database. Probes from *DAM1*_cDNA hybridized to two different leafy spurge transcripts with northern analysis (Horvath et al. 2008; Figure 5). *DAM1*_cDNA contains a consensus MADS-BOX binding domain, but interestingly, the coding region is significantly smaller than most similar *DAM* genes from other species. Of the two transcripts that hybridized to *DAM1*_cDNA, the smallest (designated as *DAM1*) was up-regulated only during endodormancy in leafy spurge crown buds and based on size (Horvath et al. 2008, 2010) appears to be the transcript from which the *DAM1*_cDNA was cloned. A larger transcript was designated as *DAM2*, which accumulated during late endodormancy and was expressed throughout eco-dormancy (Doğramacı et al. 2010; Horvath et al., 2008, 2010). A full-length cDNA designated as *DAM2*_cDNA, presumably representing the larger transcript, was also cloned from the cDNA library from which the ESTs were generated. Sequencing of BAC_*DAM3* indicates that *DAM1* and *DAM2* likely represent spliced variants of the same gene since sequences from the *DAM1* and *DAM2* cDNAs are both contained in the BAC_*DAM3* clone (Figure 5). Indeed, both the *DAM1*_cDNA and *DAM2*_cDNA coding sequences exactly matched sequences in BAC_*DAM3*, and when noncontiguous, were separated by sequences that met the GT-AG rule of most eukaryotic introns. Likewise, the coding sequences identified in Lambda_*DAMb* are identical to those of BAC_*DAM3*. Thus it is currently impossible to determine if *DAM1*_cDNA and *DAM2*_cDNA are derived from BAC_*DAM3* or Lambda_*DAMb*.

The BAC_*DREB* clone, although 96.7% identical to the cDNA sequenced in the EST database, has two additional codons and several single-base differences suggesting that the BAC clone contains another similar *DREB A-4* subfamily member. Due to the sequence similarity, previous microarray and quantitative reverse transcriptase–polymerase chain reaction analyses (Doğramacı et al. 2010; Horvath et al. 2008) used to follow the expression of these dormancy-associated *DREB* genes would likely not have distinguished between these two transcripts. Since the *DREB A-4* subfamily is known to contain at least 17 members in Arabidopsis (The Arabidopsis Information Resource (TAIR) 2013), further research will be required to determine if one or more of these subfamily members is up-regulated during endodormancy.

Phylogenetic Footprinting Identifies Conserved Elements in the Promoters of *FT* Genes. Approximately 2,000 bp of 5′ sequence was recovered from BAC_*FT*-2 and BAC_*FT*-4, which represents the promoter and 5′ untranslated region (UTR) sequences. These sequences were compared to similar regions from the black cottonwood *FT1* (floral regulator) and *FT2* (vegetative growth regulator) genes using the program MEME (Bailey and Elkan 1994) and manually to identify

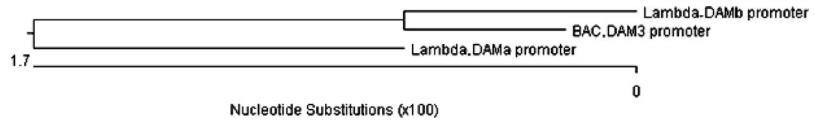

Figure 4. Phylogenetic tree developed following ClustalW alignment of 2,869 bases 5′ to the translation start site from the three *DAM* genes.

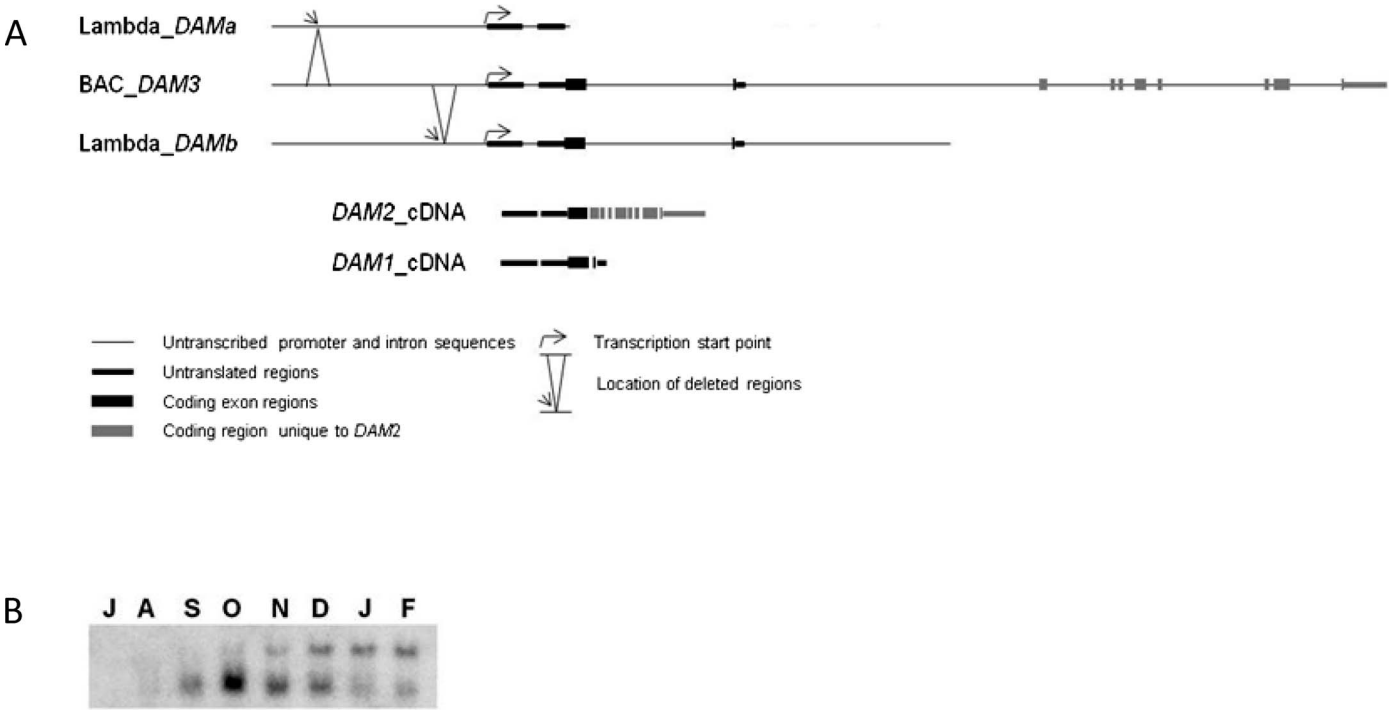

Figure 5. (A) Graphic depiction of the various *DAM* genomic and cDNA clones. The figure depicts which of the 10 coding exons from the full-length *DAM* gene, represented by BAC_*DAM3*, are differentially spliced to produce the two different *DAM* transcripts represented by *DAM1* and *DAM2*_cDNA clones. The promoter and intron regions are noted with a fine line. Medium-thick boxes indicate untranslated regions of the mRNA, and thick boxes indicate coding exons. The exons that are unique to *DAM2*_cDNA are noted in a lighter shade of grey. Transcription start sites are noted with an L-shaped arrow. The relative size and placement of deletions within the promoters of the three genomic copies of the gene are noted with lines connecting the three genomic clones with an arrow pointing out the deleted region. (B) A northern blot showing the expression patterns of *DAM2* (top band) and *DAM1* (bottom band) through a monthly time course (July [J] through February [F]) is shown.

known and suspected conserved transcription factor binding sites (Figure 6). The most striking sequence was GCCA-GATTC, which was present in all four promoters and within 400 bases of the start of translation. This element is predicted to bind an ARABIDOPSIS RESPONSE REGULATOR (ARR). These ARR proteins are involved in many processes, most notable in regulating circadian-responsive and cytokinin-responsive genes (Salome et al. 2006; Schaller et al. 2002). All four genes also have a sequence similar to the GTGT(N2-3)ATG element that binds CONSTANS (Tiwari et al. 2010), but with the sequence GTGT(N3)(A/T)GT. This sequence regulates *FT* in response to day length, and could be involved in down-regulation of *FT2* during endodormancy induction (Bohlenius et al. 2006). Other identified conserved elements included a putative ABA binding element with the sequence GCCATCA. A motif of unknown function with the sequence TTTCATG(A/C)TGGGTT was present in both leafy spurge genes and in black cottonwood *FT1*. Several other less complex elements show conservation only between these three genes (data not shown), possibly suggesting that both of the full-length leafy spurge clones represent floral regulatory versions of *FT*.

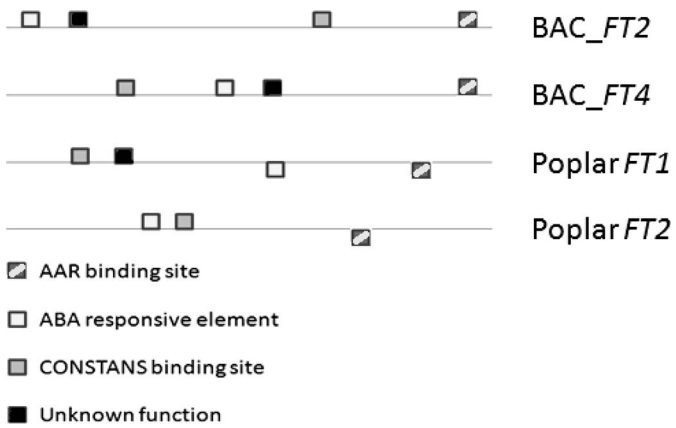

Figure 6. Relative location of phylogenetically conserved sequences in the promoters of the various *FT* genes from leafy spurge (BAC_*FT2* and BAC_*FT4*) and poplar (*FT1* and *FT2*). Boxes below the line indicate the sequence was in the reverse orientation.

Phylogenetic Footprinting Identifies Conserved Elements in the Promoters of *DAM and DREB* Genes. Phylogenetic footprinting analysis of the two previously cloned *DAM* gene promoters has already been reported (Horvath et al. 2010). As noted above, the promoter from BAC_*DAM3* is highly similar to the previously characterized *DAM* genes. Thus, BAC_*DAM3* also has the putative DREB binding sites found in lambda_*DAMa* and *b* as well as the probable circadian-regulatory EVENING elements (Alabadi et al. 2001). It also contains a large conserved sequence of unknown function found in both previously characterized leafy spurge *DAM* gene promoters as well as in the promoters of three likely *DAM* genes from black cottonwood (Horvath et al. 2010). However, due to sequence conservation in the coding regions, it is unclear if some or all of these promoters are active.

Several putative transcription factor binding sites were observed when the promoter for the BAC_*DREB* gene was scanned using the program PlantCARE's web tool (Lescot et al. 2002). These putative binding sites included possible EVENING elements, ABA and ethylene responsive factor

(ERF) binding sites, and DREB binding sites. However, none of these elements were well conserved between the BAC_-*DREB* clone and the promoters of the most similar genes from Arabidopsis and black cottonwood. This lack of phylogenetically conserved sequences may be due to the fact that this *DREB A-4* gene is part of a large gene family (Nakano et al. 2006), and it is possible that the most similar members from other species were not orthologous and thus might be regulated differently. However, it is noteworthy that a large section of sequence surrounding the putative ERF binding site (more than 10 additional bases) was conserved between the leafy spurge BAC_*DREB* gene and one of the three most similar genes, AT4G16750 from Arabidopsis (data not shown). The conservation of the surrounding sequence makes it very likely that this site is functionally conserved.

In conclusion, we have developed a BAC library for the model invasive perennial weed leafy spurge with 5× coverage of the genome. This library has proven useful for cloning and characterizing full-length genomic clones from two gene families (*DAM* and *FT*) and a full-length genomic clone encoding a *DREB A-4* transcription factor associated with bud dormancy processes (Kahn 2011). We have leveraged this information to enhance our knowledge of potential mechanisms regulating leafy spurge and black cottonwood *FT* and *DAM* genes. Additionally, this new genomics resource should prove useful for closing gaps and confirming contig orientation in the ongoing efforts towards full genome sequencing of leafy spurge. These novel genomics tools should enhance our ability to answer fundamental questions concerning the evolution of weediness traits and the genomics of invasiveness.

Acknowledgments

We would like to thank Laura Kelly (U.S. Department of Agriculture) for technical assistance, Raju Jetty (Arizona Genomics Insitute) for helpful suggestions on strategy, and Dr. Douglas Bielenberg (Clemson University) and Drs. Rongmei Wu and Marion Wood (Plant and Food, New Zealand) for critical review of this manuscript.

Literature Cited

Alabadi, D., T. Oyama, M. J. Yanovsky, F. G. Harmon, P. Mas, and S. A. Kay. 2001. Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. Science 293:880–883.

Anderson, J. V. 2008. Emerging technologies: an opportunity for weed biology research. Weed Sci. 56:281–282.

Anderson, J. V., D. P. Horvath, W. S. Chao, M. E. Foley, A. G. Hernandez, J. Thimmapuram, L. Liu, G. L. Gong, M. Band, R. Kim, and M. A. Mikel. 2007. Characterization of an EST database for the perennial weed leafy spurge: an important resource for weed biology research. Weed Sci. 55:193–203.

Bailey, T. L. and C. Elkan. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Pp. 28–36 *in* Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. Menlo Park, CA: AAAI Press.

Blanchette, M., B. Schwikowski, and M. Tompa. 2002. Discovery of regulatory elements by a computational method for phylogenetic footprinting . J. Comp. Biol. 9:211–223.

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

Chao, W. S., M. E. Foley, M. Doğramacı, J. V. Anderson, and D. P. Horvath. 2011. Alternating temperature breaks dormancy in leafy spurge seeds and impacts signaling networks associated with HY5. Funct. Integr. Genomics 11:637–649.

Chao, W. S., D. P. Horvath, J. V. Anderson, and M. E. Foley. 2005. Potential model weeds to study genomics, ecology, and physiology in the 21st century. Weed Sci. 53:929–937.

Doğramacı, M., D. P. Horvath, W. S. Chao, M. E. Foley, M. J. Christoffers, and J. V. Anderson. 2010. Extended low temperature impacts dormancy status, flowering competence, and transcript profiles in crown buds of leafy spurge. Plant Mol. Biol. 73:207–226.

Doğramacı, M., D. P. Horvath, M. J. Christoffers, and J. V. Anderson. 2011. Dehydration and vernalization treatments identify overlapping molecular networks impacting endodormancy maintenance in leafy spurge crown buds. Funct. Integr. Genomics 11:611–626.

Dong, M. A., E. M. Farré, and M. F. Thomashow. 2011. CIRCADIAN CLOCK-ASSOCIATED 1 and LATE ELONGATED HYPOCOTYL regulate expression of the C-REPEAT BINDING FACTOR (CBF) pathway in Arabidopsis. Proc. Nat. Acad. Sci. U. S. A. 108:7241–7246.

Foley, M. E., J. V. Anderson, W. S. Chao, M. Doğramacı, and D. P. Horvath. 2010. Initial changes in the transcriptome of *Euphorbia esula* seeds induced to germinate with a combination of constant and diurnal alternating temperatures. Plant Mol. Biol. 73:131–142.

Henson, J., G. Tischler, and Z. Ning. 2012. Next-generation sequencing and large genome assemblies. Pharmacogenomics 13:901–915.

Horvath, D. P. 2009a. Genomics for weed science. Curr. Genomics 11:47–51.

Horvath, D. P. 2009b. Common mechanisms regulate flowering and dormancy. Plant Sci. 177:523–531.

Horvath, D. P., W. S. Chao, J. C. Suttle, J. Thimmapuram, and J. V. Anderson. 2008. Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (*Euphorbia esula* L.). BMC Genomics. 9:536.

Horvath, D. P., S. Sung, D. Kim, W. S. Chao, and J. V. Anderson. 2010. Cloning, characterization, and regulation of *DORMANCY-ASSOCIATED MADS-BOX* genes from leafy spurge. Plant Mol. Biol. 73:169–179.

Hsu, C.-Y., J. P. Adams, H. Klim, K. No, C. Ma, S. H. Strauss, J. Drnevich, L. Vandervelde, J. D. Ellis, M. D. Rice, N. Wickett, L. E. Gunter, G. A. Tuskan, A. M. Brunner, G. P. Page, A. Barakat, J. E. Carlson, C. W. dePamphilis, D. S. Luthe, and C. Yuceer. 2011. *FLOWERING LOCUS T* duplication coordinates reproductive and vegetative growth in perennial poplar. Proc. Nat. Acad. Sci. U. S. A. 108:10756–10761.

Jeffares, D. C., T. Mourier, and D. Penny. 2006. The biology of intron gain and loss. Trends Genet. 22:16–22.

Kahn, M. S. 2011. The role of DREB transcription factors in abiotic stress tolerance of plants. Agric. Environ. Biotech. 25:2433–2442.

Koren, S., M. C. Schatz, B. P. Walenz, J. Martin, J. T. Howard, G. Ganapathy, Z. Wang, D. A. Rasko, W. R. McCombie, E. D. Jarvis, and A. M. Phillippy. 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat. Biotech. 30:693–700.

Lescot, M., P. Déhais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, P. Rouzé, and S. Rombauts. 2002. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 30:325–327.

Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain, and M. Pallen. 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotech. 30:434–439.

Luo, M. and R. Wing. 2003. An improved method for plant BAC library construction. Methods Mol. Biol. 236:3–20.

Nakano, T., K. Suzuki, T. Fujimura, and H. Shinshi. 2006. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol. 140:411–432.

Pin, P. A. and Nilsson, O. 2012. The multifaceted roles of FLOWERING LOCUS T in plant development. Plant Cell Environ. 35:1742–1755.

Rinne, P.L.H., A. Welling, J. Vahala, L. Ripel, R. Ruonala, J. Kangasjarvi, and C. van der Schoot. 2011. Chilling of dormant buds hyperinduces *FLOWERING LOCUS T* and recruits GA-inducible 1,3-beta-glucanases to reopen signal conduits and release dormancy in *Populus*. Plant Cell 23:130–146.

Salome, P. A., J.P.C. To, J. J. Kieber, and C. R. McClung. 2006. Arabidopsis response regulators ARR3 and ARR4 play cytokinin-independent roles in the control of circadian period. Plant Cell 18:55–69.

Schaller, G. E., D. E. Mathews, M. Gribskov, and J. C. Walker. 2002. Two-component signaling elements and histidyl-aspartyl phosphorelays. *In* C. R. Somerville and E. M. Meyerowitz, eds. The Arabidopsis Book. Rockville, MD: American Society of Plant Biologists, doi/10.1199/tab.0044. http://www.aspb.org/publications/arabidopsis/. Accessed January 3, 2013.

Schulz-Schaeffer, J. and S. Gerhardt. 1989. Cytotaxonomic analysis of the *Euphorbia* spp. (leafy spurge) complex. II. Comparative study of the chromosome morphology. Biol. Zentralbl. 108:64–76.

Stahevitch, A. E., C. W. Crompton, and W. A. Wojtas. 1988. Cytogenetic studies of leafy spurge, *Euphorbia esula*, and its allies (Euphorbiaceae). Can. J. Bot. 66:2247–2257.

Stewart, C. N., Jr., P. J. Tranel, D. P. Horvath, J. V. Anderson, L. H. Rieseberg, J. H. Westwood, C. A. Mallory-Smith, M. L. Zapiola, and K. M. Dlugosch. 2009. Evolution of weediness and invasiveness: charting the course for weed genomics. Weed Sci. 57:451–462.

The Arabidopsis Information Resource (TAIR), http://www.arabidopsis.org/servlets/Search?type=general&search_action=detail&method=1&show_obsolete=F&name=DREB+subfamily+A-4&sub_type=gene&SEARCH_EXACT=4&SEARCH_CONTAINS=1, on 1/3/2013.

Tiwari, S. B., Y. Shen, H.-C. Chang, Y. Hou, A. Harris, S. F. Ma, M. McPartland, G. J. Hymus, L. Adam, C. Marion, A. Belachew, P. P. Repetti, T. L. Reuber, and O. J. Ratcliffe. 2010. The flowering time regulator CONSTANS is recruited to the *FLOWERING LOCUS T* promoter via a unique *cis*-element. New Phytol. 187:57–66.

Received October 29, 2012, and approved December 16, 2012.