Genome walking in pea: an approach to clone unknown flanking sequences

Chawla, R. and DeMason, D.

Botany and Plant Sci. Univ. of California, Riverside, CA 92507, USA

The isolation and characterization of unknown DNA sequences flanking known regions are critical, especially for the analysis of upstream and downstream noncoding regions. The traditional approach for 'walking' from regions of known sequence into flanking DNA sequences involves the successive probing of libraries with clones obtained from prior screenings. This method of screening DNA libraries is a relatively time consuming procedure and requires the use of radioactive probes. Advancements in the PCR technique have helped researchers to reduce this time and avoid the use of radioactive probes (1). Genome walking is a relatively fast, reliable and general approach to sequence or clone DNA adjacent to a known region.

Promoters are segments of DNA that regulate the timing and location of gene expression. The promoter sequence is usually located upstream of the transcription start site, but regulatory elements can be present in 5' untranslated regions (UTRs), within introns or in the 3' UTRs of genes. Analysis of promoter sequences in combination with new databases like PLACE (http://www.dna.affrc.go.jp/htdocs/PLACE/fasta.html) and PlantCARE (http://oberon.fvms.ugent.be:8080/PlantCARE/index.html) provide possible insights into the regulation of important plant genes. One very powerful way of modifying the characteristics of plants is to target the expression of introduced genes to specific parts of the plant or at specific stages of the life cycle using promoters with known specificity. Very few pea promoters are currently known. The goal of this paper is to summarize our success with cloning promoter sequences of pea genes using the technique of genome walking. We discuss the method by which it works and the results obtained.

Materials and Methods

Plant materials

The pea genotype used in this study is from the Marx collection, which resides in the USDA Western Regional Plant Introduction Station. It is W6 22593, which is designated as *WT* or *Af St Tl*. Seeds were sown in UC soil mix supplemented with slow release fertilizer in 1 gallon pots and plants were grown under standard greenhouse conditions and natural light regimes. Leaves of 1-month old plants were frozen at -80 C until DNA extraction.

The GenomeWalking technique

The GenomeWalking technique is summarized in Fig. 1. First, genomic DNA is isolated from plant tissues. The quality of DNA is checked for high average molecular weight on a 0.8% agarose gel. The starting DNA must be very clean and have a high average

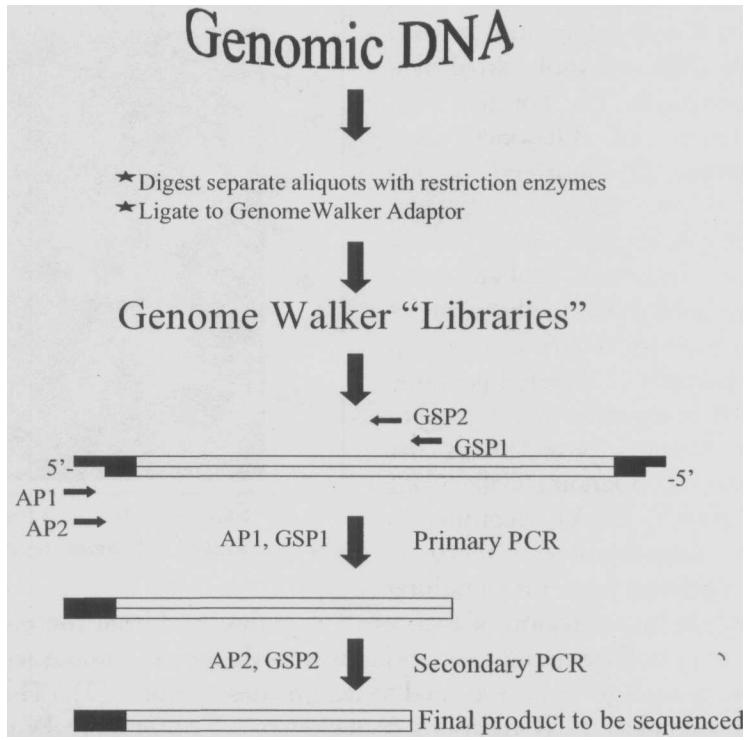

Fig. 1. Flow chart of the GenomeWalker protocol. AP1 and AP2 represent adaptor primers and GSP1 and GSP2 represent gene specific primers.

molecular weight, requiring a higher quality preparation than the minimum suitable for Southern blotting or coventional PCR. The DNA is digested by blunt end cutting enzymes. *Dra*I, *Pvu*II, *Eco*RV and *Stu*I are the enzymes provided in the Universal GenomeWalker kit (Clontech, USA). However, other enzymes that leave a blunt end could be also be used, expanding the number of libraries that can be generated. The digested DNA is then purified and ligated overnight at 16 C with the adaptors provided in the kit. These are referred to as GenomeWalker "libraries". The next step is to design a pair of gene specific primers (GSP1 and GSP2). The GSP1 primer can be from the coding region of the gene and GSP2 should be further upstream to give a nested product in the subsequent PCR. These primers should be 26-30 nucleotides in length and have a G/C-content of 40-60%. This ensures that the primers anneal effectively to the template at an annealing and extension temperature of 67 C. The primary PCR reaction uses the outer adaptor primer (AP1) and the outer gene-specific primer (GSP1). The product of the primary walk is then diluted and used as a template for a secondary walk with the nested adaptor (AP2) and nested gene-specific (GSP2) primers. The major PCR products obtained are gel extracted using Ultra DNA kit (Millipore) and sequenced and aligned with the help of the GCG program. Each of the DNA fragments begins with a known sequence at the 5' end of GSP2 and extends into the unknown adjacent genomic DNA. This DNA can be cloned for further analysis. The 5' upstream regions thus obtained can be scanned for regulatory elements using the PLACE (3) and *Plant*CARE (*cis*-acting regulatory elements) (4) programs. This whole sequence can be repeated to obtain additional 5' as needed.

Specific Results

The DNA was isolated from WT pea leaves according to Dellaporta et al. (2) except that the genomic DNA was RNase-treated and extracted with phenol and chloroform, washed with 70% ethanol, dried and dissolved in TE. For the construction of GenomeWalker libraries, 11 blunt-end cutting restriction enzymes (*Eco*RV, *Dra*I, *Pvu*II, *Stu*I, *Sma*I, *Sna*BI, *Hpa*I *Nru*I, *Nae*I, *Ssp*I and *Sca*I) were used individually to digest this genomic DNA completely. Each batch of digested genomic DNA was purified and ligated to the adaptors provided in the Universal GenomeWalker kit (Clontech, USA) according to the manufacturer's protocol. The primers used for obtaining the 5' upstream region of each gene were designed from the coding region of each gene and in cases where multiple walks were required, primers were designed from the sequence obtained from the first set of walking. The Primer3 program was used to design these primers (5). The libraries were then used to perform primary and secondary PCR reactions as necessary. The primary PCR gives multiple bands in each lane with a general background smear (Fig. 2a). The secondary PCR using the internal primers and the diluted primary PCR product as the template, selectively amplifies the desired product in the subsequent nested PCR resulting in a single bright band (Fig. 2b). The kit also provides preconstructed Human GenomeWalker Library as well

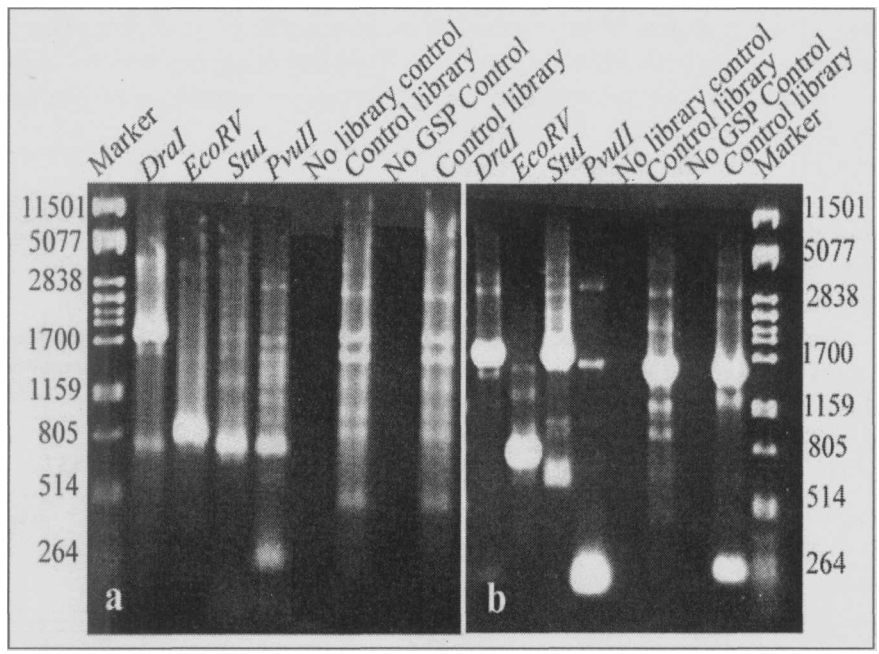

Fig. 2. *Representative gels from primary walk (a) and secondary walk (b) for* PsArgonaute2. *upstream region. Markers are* PstI *cut lambda DNA.*

as specific primers (PCP1 and PCP2 for the plasminogen activator gene) as a positive control for PCR (control library) which generates a major product of 1.5 kb (Fig. 2a, b). For negative controls a reaction lacking a GenomeWalker library or lacking the GSP1 was used. The PCR fragments were then purified using the Ultra DNA kit (Millipore) as per the manufacturer's instructions. In this case the product selected was the 1500 bp *Dra*I product (Fig. 2b, first lane) because it was the largest, major band. To verify the authenticity, this product was sequenced and compared for the 5'-end overlap.

We have identified upstream regions for 5 pea genes using this technique. Two *of* the genes are PsPINl (AY222857) and PsPK2 (M69031) which are orthologs of Arabidopsis PIN1 and PINOID genes. These genes are involved in auxin transport. With two sets of walks, we were able to obtain 1500 bp of PsPINl and 2800 bp of PsPK2 upstream regions. Analysis of their sequences obtained using the PlantCARE and PLACE programs revealed the presence of multiple presumed auxin-responsive elements as well, as those responsive to other plant hormones. We also obtained a 2405 bp of the Unifoliata (AF035163) gene promoter, which is the ortholog of Arabidopsis LEAFY (M91208). Further, promoters of two recently cloned genes involved in pea development, PsArgonautel (589 bp) and PsArgonaute2 (858 bp) were also obtained (unpublished). These promoter sequences were further analyzed in silico using the programs mentioned above, which revealed potentially important aspects about their regulation and are presently under experimental investigation.

Conclusion

The Universal Genome Walking system enables researchers to create uncloned libraries for walking by PCR with any genomic DNA. These libraries can be used over and over to clone additional DNA sequences as necessary. In less than a week, the method provides access to the genomic DNA sequences adjacent to a known DNA sequence. Although we have focused on obtaining promoters, GenomeWalker DNA walking can also be used to map intron/exon junctions and to walk bidirectionally from any sequence-tagged site (STS) or expressed sequence tag (EST). Multiple steps can be strung together to create longer walks. Consequently, this method is useful for filling in gaps in genome maps, particularly when the missing clones have been difficult to obtain by conventional library screening methods.

To summarize, the technique of Genome Walking has been standardized in pea and can now be exploited to dissect out promoters of important genes for future research involving the understanding of transcriptional regulation and directed expression in this economically important legume.

Acknowledgements: The authors thank Janet Giles for performing genome walking for *PsArgonaute1*, *PsArgonaute2* and *Unifoliata* and Fang Bai for *PsPK2*. This work was supported by a grant from USDA/CSREES 2001-35304-10958 to D.A.D.

1. Brown, A.J.H., Perry, S.J., Saunders, S.E. and Burke, J.F. 1999. BioTechniques 26: 804-806.
2. Dellaporta, S.L., Wood, J.and Hicks, J.B. 1983. Plant Mol. Biol. Rep. 1: 19-21.
3. Higo, K., Ugawa, Y., Iwamoto, M.and Korenaga, T. 1999. Nucleic Acids Res. 27: 297-300.
4. Lescot, M., Dehais, P.,Thijis, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P. and Rombauts, S. 2002. Nucleic Acids Res. *30:* 325-327.
5. Rozen, S. and Skaletsky, H.J. 2000. In: Krawetz S. and Misener, S (eds.) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386.