A genome walking strategy for the identification of eukaryotic nucleotide sequences adjacent to known regions

Claudia Leoni[1], Raffaele Gallerani[1,2], and Luigi R. Ceci[2]

[1] Department of Biochemistry and Molecular Biology, University of Bari, Bari and
[2] Institute of Biomembranes and Bioenergetics, Italian National Research Council, Trani, Italy

Determination of nucleotide sequences adjacent to a known region is a recurring need in many genome scale studies. Various methods have been developed based on PCR techniques in order to fulfill this aim and overcome the time-consuming approach of screening genomic libraries. Usually these protocols rely on specific requirements and strategies, such as the presence of suitable nucleotide restriction sites and ligation of specific single- or double-strand linkers, thus limiting their application to a certain extent. In this paper we present an alternative PCR-based protocol, consisting of four main steps: (i) extension of a sequence-specific primer; (ii) 3′-tailing of extended single-strand DNA; (iii) PCR; and (iv) nested PCR amplifications. This method, which appears to be a valid alternative to the other PCR-based protocols, was used for the identification of sequences flanking the cDNA encoding region of the Lhcb1.1 *gene (one member of the multigene family coding for the light harvesting protein Lhcb1) in the spinach genome.*

INTRODUCTION

Isolating and identifying nucleotide sequences flanking known regions is a common requirement in a number of studies of gene and genome characterization. It may be used for identification of regulatory sequences outside cDNA coding regions and gaps in genome sequencing projects, or for mapping of insertional mutagenesis events produced by retroviruses and transposable elements (1,2).

Subsequent to the introduction of the PCR technique, several protocols have been developed to identify nucleotide sequences outside known regions. With the exception of the "inverse PCR" (3), these methods adopt the strategy of linking a small stretch of synthetic DNA to the known sequence, thereby allowing PCR amplifications to be carried out. Certain methods combine preliminary restriction digestions of genomic DNA with ligations of either double-strand DNA cassettes, as in "vectorette PCR" (4), "splinkerette PCR" (5), "capture-PCR" (6), "T-linker PCR" (7) or single-strand oligonucleotides, as in "panhandle"

(8), and "boomerang PCR" (an atypical PCR performed using a single primer, working in both directions) (9). In addition, "inverse PCR" requires restriction digestion of genomic DNA, but restricted fragments are self-ligated and subjected to PCR using reverse oriented primers (3). As an alternative to restriction digestion of genomic DNA, in the ligation-mediated method proposed by Mueller and Wold (10), a double-strand DNA cassette is ligated to blunt-end DNA fragments obtained by primer extension of a specific oligonucleotide carried out on chemically nicked DNA. Other methods perform PCR amplifications using degenerate primers coupled with sequence-specific primers, such as the so-called TAIL (thermal asymmetric interlaced) PCR (11) and the UFW (universal fast walking) method (1), or the method by Levano-Garcia et al. (12), in which consensus-degenerate primers are used. Critical overviews of some of the above-mentioned methods have already been reported (9,13). All of these methods have been successfully used, but with certain limitations due either to the requirements of restriction

sites, the efficiency of ligation reactions or annealing of degenerate primers. Accordingly, there are only a few commercial kits for genome walking, such as the GenomeWalker kit (Clontech, Mountain View, CA, USA) and UVS1 Vectorette Genomic Systems (Sigma-Aldrich, St. Louis, MO, USA). These kits, based on restriction digestions of genomic DNA and ligation of a double-strand DNA cassette, are ready to use only for the human genome (as well as mouse and rat for the GenomeWalker kit), but require isolation and restriction digestions of genomic DNA if intended for use with other organisms. Alternatively, Evrogen (Moscow, Russia) provides a customized service based on a similar approach.

The genome walking approach that we have developed is independent of the presence of specific restriction sites and does not require the use of random primers or ligation of single- or double-strand linkers. Our method is based mainly on a classical 5′-RACE approach, but it is applicable for both 5′ and 3′ genome walking. Whereas a similar protocol was applied with success only for a bacterial genome (*Microcystis aeruginosa*) (14), our method has broader potential for application since it has been developed for the genome of a higher eukaryote.

In this paper we show the application of the genome walking method for the identification of the nucleotide sequences flanking the cDNA coding region of the spinach *Lhcb1.1* gene (one member of the multigene family coding for the light harvesting protein Lhcb1) (15).

MATERIALS AND METHODS

Standard Nucleic Acids Isolation and Manipulation Methods

Genomic DNA was purified from two-week old spinach seedlings (var. America), using the Gene Elute Plant genomic kit (Sigma-Aldrich).

Restriction digestion of DNA, cloning, plasmid DNA isolation, and sequencing were carried out according to standard procedures [(16); pGEM-

Short Technical Reports

T Easy Vector System II manual (Promega, Madison, WI, USA)].

Oligonucleotides

Oligonucleotides specific for spinach DNA were from Operon (Alameda, CA, USA) or Sigma-Aldrich. Oligonucleotide sequences are reported in Table 1. For poly-dG containing primer, we used the 5′-RACE abridged anchor primer (AAP) from Invitrogen (Carlsbad, CA, USA).

Genome Walking

The overall procedure for genome walking is shown in the Results and Discussion section, along with results obtained by changing parameters in the various steps. The schematic protocol is given here, divided into the four main steps.

Extension. Sixty pmoles of a gene-specific primer (GSP) were added to 0.5 µg of genomic DNA and subjected to extension in the presence of 0.4 mM of each dNTP and 1.5 µl (7.5 U) of TripleMaster Polymerase Mix (Eppendorf, Hamburg, Germany) in a total of 100 µl of the specific HighFidelity Buffer supplied by the enzyme vendor (Eppendorf). Extension reaction was carried out in a thermal cycler according to the following program: 5 min at 94°C (1 min at 94°C, 1 min at the proper annealing temperature for extension (TA_E) of the GSP reported in Table 1, and 1 min at 72°C) for 35 times, 7 min at 72°C. The reaction product was purified using GFX MicroSpin columns (GE Healthcare, Chalfont St. Giles, UK).

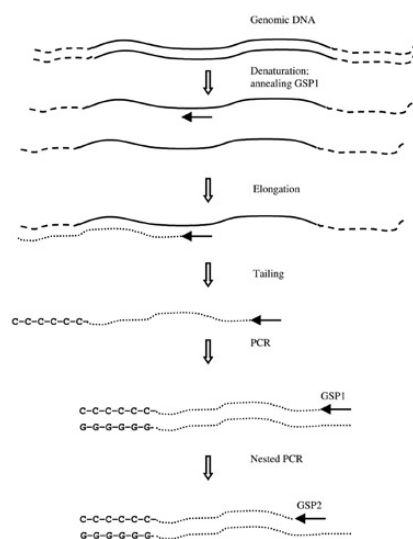**Tailing.** Five µl of purified DNA from the extension reaction were

denatured at 94°C for 3 min in a final volume of 24 µl containing 5 µl of 5× TdT buffer (50 mM Tris-HCl pH 8.4, 125 mM KCl, 7.5 mM MgCl$_2$) and 0.2 mM dCTP. After denaturation the tube was immediately chilled in ice for at least 1 min before addition of 1 µl (15 U) of terminal deoxynucleotidyl transferase (Invitrogen). The reaction was allowed to proceed at 37°C for 20 min. The enzyme was then inactivated at 65°C for 10 min.

PCR. PCR amplifications of tailed single-strand DNA were carried out using 5 µl of the tailing reaction mix, in the presence of 20 pmoles of each oligonucleotide (AAP and the GSP primer used in the extension step), 0.2 mM of each dNTP, and 2.5 U of TripleMaster Polymerase Mix in 50 µl of the same buffered solution used for the extension step. After the first DNA denaturation step at 94°C for 5 min, amplifications were run for 35 cycles consisting of 1 min at 94°C, 1 min at

the specific annealing temperature chosen for the couple of primers (see below), and 1 min at 72°C. A final elongation step was then run at 72°C for 7 min. The couples of primers and annealing temperatures (TA) were as follows:

AAP /GSP_1.1_R, TA, 40°C;

AAP /GSP_1.1_OUT_F, 40°C;

AAP /GSP_1.1_OUT_R, 39°C.

PCR amplification products were purified through GFX MicroSpin columns.

Nested PCR. Five µl of purified first PCR products were subjected to nested PCR using a gene-specific nested primer and the AAP oligonucleotide. Reaction conditions were substantially identical to those of the first PCR, with the exception of the total number of cycles, but were reduced to 30 in the

Table 1. Oligonucleotides Used in the Genome Walking Experiments

Oligonucleotides	Sequences	TA_E	Coordinates
GSP_1.1_R	5′-TC-AAG-CTC-ACG-GTT-CTT-AG-3′	45, 50	358 – 340
GSP_1.1_RA	5′-C-AGC-AGT-GTC-CCA-ACC-GTA-GTC-ACC-3′		312 – 288
GSP_1.1_OUT_F	5′-ATTTGTGAAGCTGGTTCAATTGT-3′	45, 50	867 – 889
GSP_1.1_OUT_FA	5′-GCCTCGGTGTAGTAAATGCTTTC-3′		911 – 933
GSP_1.1_OUT_R	5′-CATTGATCCAGTTCATGTTC-3′	45, 47	-207 – -226
GSP_1.1_OUT_RA	5′-CGCACTCCTGCCAACGCATTTACTAAC-3′		-231 – -257

Coordinates refer to the sequence in Figure 3 and are given for the 5′-3′ orientation of oligonucleotides. TA_E indicates the two annealing temperatures used in the extension experiment.
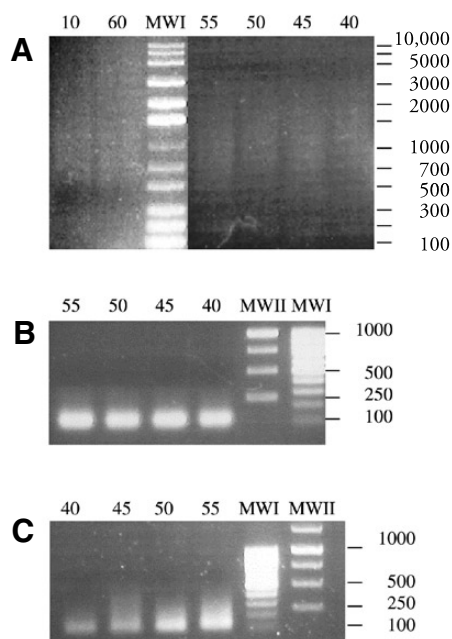
Figure 2. Electrophoretic characterization of genome walking intermediate products. (A) Primer extension. Products obtained using different amounts of primer GSP_1.1_R (10 and 60 pmoles) or different annealing temperatures (TA; 55°, 50°, 45°, and 40°C) are shown. MWI is the MassRuler Express DNA Ladder Mix, Reverse by Fermentas (Burlington, Ontario, Canada). Numbers above the figure indicate the amount of primer (10 and 60) or the TA_E (55°, 50°, 45°, and 40°C). (B) First PCR. PCR amplification products obtained from poly-dC tailed single-strand DNA using abridged anchor primer (AAP) and GSP_1.1_R primers are shown. The reaction was carried out with a TA = 40°C. Numbers above the figure indicate the annealing temperature used in the extension step. MWI is as in (A). MWII is the GeneRuler 1kb DNA Ladder by Fermentas. (C) Nested PCR. PCR amplification products obtained using DNA from the first PCR as substrate and AAP and GSP_1.1_RA as primers are shown. Substrate DNA was from the first PCR amplifications shown in (B). The reaction was carried out with a TA = 60°C.

nested PCR. The couples of primers and annealing temperatures were:

AAP /GSP_1.1_RA, TA, 60°C;

AAP /GSP_1.1_OUT_FA, TA, 60°C ;

AAP /GSP_1.1_OUT_RA, TA, 59°C.

Amplification products from nested PCR were purified through GFX MicroSpin columns and cloned using the pGEM-T Easy Vector System II.

RESULTS AND DISCUSSION

The Genome Walking Method

This study reports a genome walking strategy suitable for the identification in eukaryotic genomes of nucleotide sequences outside of a known region. The method was adapted from classical 5′-RACE protocols, taking into account the main difference in the starting material: an entire eukaryotic genome of about 900 Mb, rather than the relatively small population of RNA molecules. A similar approach was previously reported only for a bacterial genome (14).

A schematic drawing of the whole procedure for the identification of 5′ flanking sequences is reported in Figure 1. The procedure can obviously be used for the identification of unknown 3′ flanking sequences.

In order to gain as much sequence information as possible by a single experiment, different conditions were tested. In particular, efforts were devoted to the extension step, as the other steps are more commonly used.

First, in order to establish the proper amount of elongation primer to use, starting with a fixed amount of genomic DNA, two experiments were run using different amounts of primer (10 and 60 pmoles for extension with 0.5 µg of DNA). Figure 2A shows the result of elongation reactions carried out with different primer amounts. They appear as a continuous smear of fragments in the 300–3000 bp range of the molecular weight marker.

Using equal volumes of the two elongation products, and running the rest of the procedure as described in the Materials and Methods section, we obtained more positive colonies when the nested PCR products derived from the elongation performed with 60 pmoles of primers were cloned. We estimate an increase of about 20%. More interestingly, amplification fragments inside positive colonies were generally longer in the case of elongation carried out with 60 pmoles.

Indeed, in this case, approximately 50% of clones show fragments longer than 350 bp (the cut-off value of final PCR products we chose), while the percentage is <25 when 10 pmoles of primer were used in elongation. This result may simply be explained by assuming that the higher amount of extension products obtainable with 60 pmoles of primer (see Figure 2A) makes the detection of relatively "long" fragments more probable, when taking into account the successive steps of the whole procedure (column purification, tailing, PCR, and nested PCR).

The extension step was also improved by using different annealing temperatures. The evaluation of the more appropriate annealing temperature in the extension step (TA_E) was necessary since there is more than one formula to predict T_m (melting temperature) values for an oligonucleotide (16) and, therefore, more than one annealing temperature to use in the extension step can be deduced.

We calculated primer T_m according to several methods (17–19), using the oligo T_m calculator available at the Promega web site (www.promega.com/biomath). Using these methods, T_m values may be highly variable, differing also for >8°C. The primer GSP_1.1_R, for example, has T_m values going from 43°C, according to the formula by Bolton and McCarthy (18) and assuming monovalent ion concentration of 0.05 M, to 52°C if calculated taking into account variations of thermodynamic parameters. Therefore in pilot experiments, the extension step was run using several annealing temperatures. Figure 2A shows the products of extension reactions obtained using the primer GSP_1.1_R at four different annealing temperatures (55°, 50°, 45°, and 40°C). Apparently the smear originated by the elongation products shows a slight increase in the amount of fragments and their molecular weight as temperature decreases. All four products were used in the successive steps.

After synthesis, reaction products, purified through GFX MicroSpin columns, were subjected to a tailing reaction in the presence of terminal deoxynucleotidyl transferase and dCTP.

Following the tailing reaction, the first PCR amplification, in which the same GSP of the extension step is used, was carried out at four different annealing temperatures (for GSP_1.1_R: 55°, 50°, 45°, and 40°C). Amplification products show no difference when electrophoretically analyzed. In addition, they seem to be made of low molecular weight fragments. Figure 2B shows the products of the first PCR amplification carried out with an annealing temperature of 40°C. When the first PCR fragments obtained with the couple of primers AAP/GSP_1.1_R were directly cloned and sequenced, they showed no sequences homologous to *Lhcb1.1* (results not shown). Notwithstanding this apparently negative result, amplification fragments from the first PCR were used for nested PCR.

We performed some tests for this step as well, based on different annealing temperatures. For example, for the primer GSP_1.1_RA (for which T_m goes from 57.5° to 62°C), we tested 55° and 60°C. Figure 2C shows products of nested reactions carried out at the annealing temperature (TA) of 60°C on the products of the first PCR shown in Figure 2B. Unlike the products obtained in the first PCR amplifications, those from nested PCRs give different electrophoretic patterns. Amplification products were cloned in pGem-T Easy Vector and analyzed by sequencing. Independent colonies showed sequences corresponding to the *Lhcb1.1* cDNA in more than 95% of cases. Not all colonies were analyzed by sequencing; only those that showed a plasmid insert longer than 350 bp (about 50%). Indeed, fragments below this value would not give valuable sequence information. Most of the selected clones (80%) showed inserts of about 500 bp, while the longest fragments of about 750 bp were found in about 10% of the clones.

Sequencing results from the products obtained after the whole procedure allow us to suggest some rules of thumb to follow for an efficient genome walking approach. For the extension step, it is advisable to run two different reactions, using two TA_Es comprised between the extreme T_m values calculated from the methods

```
(-668)
AGATAATATAAGCTGCATGTTTCTAGTATTGAGTATAGATTCCGGTGAATATTGTTTGTT -609
GTTTGATTAATTGTTTCGATGGATTTTGGTTTTTCTATCTTCCTTTTTTTTAAAAGAAAAT -549
TACAGAGGTATAAAATTTTCATGACTGATTTCCAGACACTAGTTCTAGTAAAGTTTCAAT -489
TTCAGTCCATTTACCTACCTTAAATTTTATTCCAGAGCTCCTTACCAAACACTCGTACTT -429
ATGTAAGGTGACTTAAATTATTAGCGCCTAATCGTTTAAGGTGACTGAAATTGATTGCCA -369
AACAGAGCCTTAAGCAAATATACAAATCTGCAAATATTTAAGATGAAAATATTAGGCAAA -309
AATGTAAACGGCTCATATTTTTTGGGGGTCTTTGATTATCCGCCAACCCTTGTTAGTAAA -249

TGCGTTGGCAGGAGTGCGACATGAACATGAACTGGATCAATGCAATCCCAATATCCCCAC -189
    GSP_1.1_OUT_RA              GSP_1.1_OUT_R
ATGTCCTCTCCACATTGGTCAATAACTAAACTTTAGAAAACATCCGACTCCATGGCTTAG -129
AGCTTGCCACCTAGGATGGTACCAAAAATCTTCTCAAGCTAATAAGAAAGCAGATGAAGA - 69
TATCCACTTAGATAAGAGGATACCTTCTTTCCTCTATTTATAATAACCAATCTTGTACAT -  9
         1
TTTTCTTCATCACTTGAAAAATTATTTAACTCACTACAAATTACTTTCAATTCCTCTGTT   52
TTTCCCAAATGGCTTCCTCAACAATGGCTCTTTCTTCCCCGGCCTTTGGCGGTAAGGCGGT  112
TAAGCTTAACCCAAGCACCTCCGACATGATTGGTGAGGGAAGGATCAGCATGAGGAAGTC  172
AGCTGGGAAGCCCAAGAATGTATCATCAGGAAGCCCATGGTACGGTCCCGACCGTGTTAA  232
GTACTTGGGCCCATTCTCCGGTGAGGCTCCATCCTACCTTACTGGTGAGTTCCCAGGTGA  292

CTACGGTTGGGACACTGCTGGGCTTTCTGCTGACCCAGAAACCTTCTCTAAGAACCGTGA  352
        GSP_1.1_RA                           GSP_1.1_R
GCTTGAGGTGATCCACTGCAGATGGGCTATGTTGGGAGCTTTGGGCTGTGTTTTCCCAGA  412

GCTTTTGGCCCGTAATGGTGTTAAGTTTGGTGAAGCTGTTTGGTTTAAGGCTGGATCACA  472
AATCTTTAGCGAAGGTGGACTTGACTACTTGGGTAACCCAAGTTTGGTTCATGCACAAAG  532
TATTCTTGCTATTTGGGCTTGCCAAGTTATCTTGATGGGTGCTGTTGAAGGTTACCGTGT  592
TGCTGGTGGGCCACTTGGTGAGGTTGTTGACCCACTTTACCCTGGTGGTAGCTTTGACCC  652
ATTGGGTTTGGCTGAAGACCCAGAGGCTTTTGCCGAATTGAAGGTAAAGGAGATCAAGAA  712
TGGTAGATTAGCAATGTTTTCAATGTTCGGTTTCTTTGTTCAAGCTATTGTTACCGGAAA  772
GGGCCCATTGGAGAACCTTGCTGACCATCTTGCCGACCCGGTTAACAACAACGCTTGGAA  832
CTTTGCCACCAACTTCGTTCCCGGCAACTGAAAAATTTGTGAAGCTGGTTCAATTGTGTC  892
                           GSP_1.1_OUT_F
ACAGTTTCAGCTTTTAATGCCTCGGTGTAGTAAATGCTTTCATTTTTTTTAATGGATGAG  952
           GSP_1.1_OUT_FA
CTTCTTTTAAAATGCCAATTTTCTGTGCTAATCATTTGGCTCATTGCTGAGAAATGGTGA 1012
CTTGTTACATAATCTCTTCTTTTGCCCAAAAAGATATATAAAAGAAATTTCAGTAATAGT 1072
AATCTACAATTAGGAAAGTAAGAAATAATTACACTGTCATTCTACCTGAACAAATATAAT 1132
CTCATACTGAAATTTTATACTAACAAAAATAGGCAGTACATACTTCAGTAACTTGTAATT 1192
ATAGTGTGATGTAAAGTGGACAATTCCATAAACTAAACCGTTTTTCTTGTATGAAAACTG 1252
TAGTGCAATTTGAAAAATTCACCAAACCATAGTTTTATTTTAAGTGCGTCTCCAAAAGTT 1312
ATAAAAAGCCAACCAACAGTTTAGTGCAATATTATTCAATCAAAAAAAAATATTATTCAA 1372
TCAACAAATTGTTATACCGTTTTAAACTATTTTTTTAAACCACTTACAATATTTA      1427
```

Figure 3. Nucleotide sequence of *Lhcb1.1* cDNA coding sequence and flanking regions. Nucleotides identified by the genome walking method are those with negative numbering for the 5′ flanking region, and those starting from nucleotide 957 for the 3′ flanking region. Arrows indicate the locations of oligonucleotides; their precise position is reported in Table 1. CAAT and TATA motifs are boxed. The circadian expression element CAATGCAATC, GATA motifs, and the I-box GATAAG are in bold. The start and stop codons are underlined. The nucleotide sequence was deposited at the EBI Nucleic Acids Database (accession no. AM2312860; www.ebi.ac.uk).

reported above (see Table 1 for the TA_Es we used). Hereafter, the two products can be processed separately, but in the same way. After tailing, the first PCR amplification must be carried out using an annealing temperature below the lowest calculated T_m. Finally, nested PCR can be carried out using an annealing temperature placed between the calculated T_m values. Good results are usually obtained for both extension reactions.

The genome walking protocol that we describe for the identification of unknown nucleotide sequences in eukaryotes represents an improvement over the method reported by Rudi et al. (14) for bacterial genomes. We believe that the main reasons for this improvement can be found in the procedures for the purification of DNA molecules (we used both silica and glass-fiber columns, instead of the phenol/chloroform method) and in the elongation step. In the latter case, besides the optimization of the amounts of oligonucleotide and its annealing temperature described above, other significant changes introduced in the protocol were the DNA polymerase (we used a polymerase mix with high amplification efficiency for genomic DNA templates instead of a unique DNA polymerase) and the number of cycles (35 rather than 25).

Identification of Genomic Sequences Flanking the *Lhcb1.1* Transcript Region

We have applied the genome walking method for the identification of nucleotide sequences flanking the region encoding in the spinach genome the full-length *Lhcb1.1* cDNA (15) (Figure 3). The reliability of this method is demonstrated by the identification of both 5′ and 3′ flanking sequences, along with a two-round application for the 5′ flanking region. Analysis of 5′ flanking sequences allowed for the identification of regulatory elements commonly found for genes coding for light harvesting proteins, such as GATA and I-box motifs (20) and the circadian expression element CAANNNNATC (21). Canonical CAAT and TATA boxes can also be found.

It must be noted that, while the couples of consecutive oligonucleotides GSP_1.1_OUT_F/ GSP_1.1_OUT_FA and GSP_1.1_OUT_R/GSP_1.1_OUT_RA are located outside the *Lhcb1.1* open reading frame (ORF), the couple of primers GSP_1.1_R/GSP_1.1_RA are located inside the ORF, in regions that are perfectly conserved in the three identified genes of the spinach *Lhcb1* gene family (15). Interestingly, when these primers were used, sequences specific of the other two genes of the family were also obtained (data not shown). This result suggests another relevant application of the genome walking method we propose, as it can be usefully applied in the course of studies for the characterization of multigene families. Results regarding the analysis of the spinach *Lhcb1* gene family will be published separately.

In conclusion, we have developed a strategy for the identification of eukaryotic genomic sequences outside of a known region, which relies on a relatively simple and direct genome walking protocol. It can be considered as a valid alternative to already known genome walking protocols.

ACKNOWLEDGMENTS

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Myrick, K.V. and W.M. Gelbart.** 2007. A modified universal fast walking method for single-tube transposon mapping. Nat. Protocols 2:1556-1563.
2. **Yin, B. and D. Largaespada.** 2007. PCR-based procedures to isolate insertion sites of DNA elements. BioTechniques 43:79-84.
3. **Triglia, T., M.G. Peterson, and D.J. Kemp.** 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. 16:8186.
4. **Riley, J., R. Butler, D. Ogilvie, R. Finniear, D. Jenner, S. Powell, R. Anand, J.C. Smith, and A.F. Markham.** 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Res. 18:2887-2890.
5. **Devon, R.S., D.J. Porteous, and A.J. Brookes.** 1995. Splinkerettes—improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res. 23:1644-1645.
6. **Schmidt, M., G. Hoffmann, M. Wissler, N. Lemke, A. Mussig, H. Glimm, D.A. Williams, S. Ragg, et al.** 2001. Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. Hum. Gene Ther. 12:743-749.
7. **Yuanxin, Y., A. Chengcai, L. Li, G. Jiayu, T. Guihong, and C. Zhangliang.** 2003. T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. Nucleic Acids Res. 31:e68.
8. **Jones, D.H. and S.C. Winistorfer.** 1992. Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. Nucleic Acids Res. 20:595-600.
9. **Hengen, P.N.** 1995. Vectorette, splinkerette and boomerang DNA amplification. Trends Biochem. Sci. 20:372-373.
10. **Mueller, P.R. and B. Wold.** 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246:780-786.
11. **Liu, Y.G., N. Mitsukawa, and R.F. Whittier.** 1993. Rapid sequencing of unpurified PCR products by thermal asymmetric PCR cycle sequencing using unlabeled sequencing primers. Nucleic Acids Res. 21:3333-3334.
12. **Levano-Garcia, J., S. Verjovski-Almeida, and A.C. da Silva.** 2005. Mapping transposon insertion sites by touchdown PCR and hybrid degenerate primers. BioTechniques 38:225-229.
13. **Hui, E.K., P.C. Wang, and S.J. Lo.** 1998. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. Cell. Mol. Life Sci. 54:1403-1411.
14. **Rudi, K., T. Fossheim, and K.S. Jakobsen.** 1999. Restriction cutting independent method for cloning genomic DNA segments outside the boundaries of known sequences. BioTechniques 27:1170-1172, 1176-1177.
15. **Rea, G., M. Volpicella, F. De Leo, L. Zolla, R. Gallerani, and L.R. Ceci.** 2007. Characterization of three members of the multigene family coding for isoforms of the chlorophyll-a/b-binding protein Lhcb1 in spinach. Physiol. Plant. 130:167-176.
16. **Maniatis, T., E.F. Fritsch, and J. Sambrook.** 1989. Molecular Cloning: A Laboratory Manual. CSH Laboratory Press, Cold Spring Harbor, NY.
17. **Wallace, R.B., J. Shaffer, R.F. Murphy, J. Bonner, T. Hirose, and K. Itakura.** 1979. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. Nucleic Acids Res. 6:3543-3557.
18. **Bolton, E.T. and C.B. McCarthy.** 1962. A general method for the isolation of RNA complementary to DNA. Proc. Natl. Acad. Sci. USA 48:1390-1397.
19. **Breslauer, K.J., R. Frank, H. Blocker, and L.A. Marky.** 1986. Predicting DNA duplex stability from the base sequence. Proc. Natl. Acad. Sci. USA 83:3746-3750.
20. **Alexander, L., D. Falconet, B.W. Fristensky, M.J. White, J.C. Watson, B.A. Roe, and W.F. Thompson.** 1991. Nucleotide sequence of Cab-8, a new type I gene encoding a chlorophyll a/b-binding protein of LHC II in Pisum. Plant Mol. Biol. 17:523-526.
21. **Piechulla, B., N. Merforth, and B. Rudolph.** 1998. Identification of tomato Lhc promoter regions necessary for circadian expression. Plant Mol. Biol. 38:655-662.

Address correspondence to Luigi R. Ceci, Institute of Biomembranes and Bioenergetics, Italian National Research Council, Via Amendola 165/A, 70126—BARI, Italy. e-mail: l.ceci@ibbe.cnr.it