# Digestion-Ligation-Amplification (DLA): A simple genome walking method to amplify unknown sequences flanking *Mutator* (*Mu*) transposons and thereby facilitate gene cloning

**Sanzhen Liu, An-Ping Hsia, Patrick S. Schnable**

This method is published in the following book:

Plant Transposable Elements

Methods and Protocols
Series: Methods in Molecular Biology, Vol. 1057
Peterson, Thomas (Ed.)
2013

http://www.springer.com/life+sciences/plant+sciences/book/978-1-62703-567-5

**SUMMARY**

Digestion-Ligation-Amplification (DLA), a novel PCR-based genome walking method, was developed to amplify unknown sequences flanking known sequences of interest. DLA specifically overcomes the problems associated with amplifying genomic sequences flanking high copy number transposons in large genomes. Two DLA-based strategies, *Mu*Clone and DLA-454, were developed to isolate *Mu*-tagged alleles. *Mu*Clone allows for the amplification of DNA flanking subsets of the numerous *Mu* transposons in the genome using unique three-nucleotide tags at the 3'-ends of primers, simplifying the identification of flanking sequences that co-segregate with mutant phenotypes caused by *Mu* insertions. DLA-454, which combines DLA with 454 pyrosequencing, permits the efficient amplification and sequencing of *Mu* flanking regions in a high-throughput manner.

**Key words:** Genome walking, Ligation-mediated PCR, Next-gen sequencing, Gene cloning, Mu transposon, Mutator

## 1. INTRODUCTION

Insertional mutagenesis is widely used in functional genomics. The *Mutator* (*Mu*) transposons of maize have been effectively used for both forward and reverse genetics because of their high copy number and frequent transposition (*1*). Both applications require a method to amplify DNA sequences that flank transposons responsible for mutant phenotypes. For example, after a mutant phenotype has been identified following a forward genetic screen in a *Mu* tagging population, the challenge in cloning the affected gene is to identify the specific genic sequence that flanks the causative insertion. Alternatively, reverse genetic resources can consist of very large numbers of individuals, each of which contains multiple independent *Mu* insertions, all of which must be amplified and sequenced.

Multiple versions of adaptor ligation-mediated PCR have been developed to amplify sequences flanking insertional mutagens (*2-10*). Active *Mu* transposon lines contain 50~200 copies of *Mu* (*11, 12*). Such high copy number improves efficiency of insertional mutagenesis but greatly complicates downstream analyses. To improve amplification of *Mu* flanking sequences, we developed a simple and efficient genome walking method, digestion-ligation-amplification (DLA) (*13*). DLA uses a single-stranded oligonucleotide in place of the adaptor (which is by definition double stranded) used in adaptor ligation-mediated PCR methods. The single-stranded oligo anneals to the overhang created by restriction-enzyme-digestion. The nick between the single-stranded oligo and the digested end of the DNA molecule is repaired using ligase. After ligation, the *Mu* specific primer is used in combination with the single-stranded oligonucleotide primer to specifically amplify *Mu* flanking sequences.. Here, we provide the basic DLA protocol and two DLA-based approaches, *Mu*Clone and DLA-454 (*13, 14*). *Mu*Clone is a simple approach for cloning *Mu*-tagged genes; DLA-454, combines DLA with a high-throughput next generation sequencing technology to analyze *Mu* flanking sequences. For a detailed comparison of these two methods, please refer to Liu et al. (*13*).

## 2. Materials

### 2.1 Enzymes and Kits

1. *Nsp*I (NEB, 10,000U/ml) (For options of restriction enzyme selection, please see Notes 1-3)
2. T4 DNA ligase
3. AmpliTaq Gold DNA Polymerase

### 2.2. Oligo Sequences

1. MuTIR 32 mer: 5’ AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC
2. Mu53s 19 mer: 5’ GCCTCYATTTCGTCGAATC

3. NspI-5B 22 mer: 5’ CAGAACGTCACAGCATGTCATG
4. NspI-5 20 mer: 5’ GAACGTCACAGCATGTCATG
5. NspI-P 19 mer: 5’ AACGTCACAGCATGTCATG
6. NspI tail primers (N=32) for *Mu*Clone method:

| **primer_name** | **Sequence (5’-3’)**[1] |
|---|---|
| Nsp-16caa | GTCACAGCATGTCATGcaa |
| Nsp-15cac | TCACAGCATGTCATGcac |
| Nsp-15cag | TCACAGCATGTCATGcag |
| Nsp-16cat | GTCACAGCATGTCATGcat |
| Nsp-15cca | TCACAGCATGTCATGcca |
| Nsp-15ccc | TCACAGCATGTCATGccc |
| Nsp-15ccg | TCACAGCATGTCATGccg |
| Nsp-15cct | TCACAGCATGTCATGcct |
| Nsp-15cga | TCACAGCATGTCATGcga |
| Nsp-15cgc | TCACAGCATGTCATGcgc |
| Nsp-15cgg | TCACAGCATGTCATGcgg |
| Nsp-15cgt | TCACAGCATGTCATGcgt |
| Nsp-16cta | GTCACAGCATGTCATGcta |
| Nsp-15ctc | TCACAGCATGTCATGctc |
| Nsp-15ctg | TCACAGCATGTCATGctg |
| Nsp-16ctt | GTCACAGCATGTCATGctt |
| Nsp-16taa | GTCACAGCATGTCATGtaa |
| Nsp-16tac | GTCACAGCATGTCATGtac |
| Nsp-16tag | GTCACAGCATGTCATGtag |
| Nsp-16tat | GTCACAGCATGTCATGtat |
| Nsp-16tca | GTCACAGCATGTCATGtca |
| Nsp-15tcc | TCACAGCATGTCATGtcc |
| Nsp-15tcg | TCACAGCATGTCATGtcg |
| Nsp-16tct | GTCACAGCATGTCATGtct |
| Nsp-16tga | GTCACAGCATGTCATGtga |

| | |
|---|---|
| Nsp-15tgc | TCACAGCATGTCATGtgc |
| Nsp-15tgg | TCACAGCATGTCATGtgg |
| Nsp-16tgt | GTCACAGCATGTCATGtgt |
| Nsp-16tta | GTCACAGCATGTCATGtta |
| Nsp-16ttc | GTCACAGCATGTCATGttc |
| Nsp-16ttg | GTCACAGCATGTCATGttg |
| Nsp-16ttt | GTCACAGCATGTCATGttt |

[1] the terminal bases (lower case) are the PCR selective bases

7. DLA-454 primers

| **Primer** | **Sequence (5'-3')** |
|---|---|
| barcodeMu[1] | GCCTCCCTCGCGCCATCAGxxxxxxGCCTCYATTTCGTCGAA TC |
| BnspI-P[2] | GCCTTGCCAGCCCGCTCAGAACGTCACAGCATGTCATG |

[1] Barcoded composite primers, containing 454 primer A (5' GCCTCCCTCGCGCCATCAG 3'), the barcode (xxxxxx, e.g., atgctg), and the *Mu*-specific primer (5' GCCTCYATTTCGTCGAATC 3').

[2] The barcoded composite primer, containing 454 primer B (5' GCCTTGCCAGCCCGCTCAG 3') and the primer, NspI-P, which matches the single-stranded oligo.

## 3. Methods

In this section, the basic DLA protocol is first described. Subsequently, the protocols for two DLA-related strategies, *Mu*Clone and DLA-454, are provided.

### 3.1 The Digestion-Ligation-Amplification (DLA) method

1. Digest (300 ng) genomic DNA in the following mixture. Incubate for 1.5 h at 37°C

| Component | Volume (µl) |
|---|---|
| $H_2O$ | - |
| 10x NEBbuffer 2 | 4 |
| 100x Bovine Serum Albumin (BSA) | 0.4 |

| | |
|---|---|
| (10 mg/ml) | |
| Genomic DNA | 300 ng |
| *Nsp*I (10,000 U/ml) | 1 |
| **Total** | **40** |

CAUTION: RNase treatment to remove RNA during genomic DNA extraction is recommended. Otherwise, the apparent concentration of gDNA isolated may be overestimated.

2. After 1.5 h incubation, add the following mixture to the genomic DNA digestion reaction. Total volume should now be 60 µl. Incubate for 3 h at 16°C.

| Component | Volume (µl) |
|---|---|
| $H_2O$ | 14 |
| oligo NspI-5B (100 µM) | 2 |
| 10x NEB ligase buffer | 2 |
| T4 DNA ligase | 2 |
| **Total** | **20** |

3. Follow the protocol provided for PCR purification with the Qiaquick PCR purification kit to purify the ligated product. Dissolve purified DNA in 30 µl EB buffer. Measure the concentration of each purified PCR product using a Nanodrop instrument.
4. Conduct first PCR with the *Mu* primer and the NspI-5 primer. Add ~50 ng purified ligation product to the following PCR mixture. This PCR program consists of 94°C for 10 min; 15 cycles of 94°C for 30 s, 60°C 45 s, 72°C 2.5 min; and a final extension at 72°C for 10 min in a 20 µl volume.

| Component | Volume (µl) |
|---|---|
| Water | - |
| GeneAmp 10X PCR Buffer II | 2 |
| dNTP (2 mM) | 2 |
| $MgCl_2$ (25 mM) | 1.2 |

| | |
|---|---|
| MuTIR primer (5 µM) | 1.2 |
| NspI-5 primer (5 µM) | 1.2 |
| ligation product | ~50 ng |
| AmpliTaq Gold (5 U/µl) | 0.2 |
| **Total** | **20** |

5. The first PCR product was diluted 10 times and 2 µl of diluted product was used for second PCR with nested primers, a *Mu* primer (e.g., Mu53s) and NspI-P, using AmpliTaq Gold® DNA Polymerase. This PCR program consists of 94°C for 10 min; suitable numbers of cycles (35 and 20 cycles were employed for *Mu*Clone and DLA-454, respectively) of 94°C for 30s, 60°C 45s, 72°C 2.5 min; and a final extension at 72°C for 10 min.

| Component | Volume (µl) |
|---|---|
| Water | 16 |
| GeneAmp 10X PCR Buffer II | 3 |
| 2mM dNTP | 4 |
| $MgCl_2$ (25 mM) | 1.8 |
| Mu53s (5 µM) | 1.5 |
| NspI-P (5 µM) | 1.5 |
| 10X diluted first PCR product | 2 |
| AmpliTaq Gold (5 U/µl) | 0.2 |
| **Total** | **30** |

### 3.2 *Mu*Clone protocol for co-segregation analysis

DLA was adapted to facilitate co-segregation analysis that enables the cloning of *Mu* tagged mutants. *Mu*Clone, a cost-efficient strategy, adds unique three-nucleotide tags to the 3' ends of a common primer based on the single-stranded oligo so that subsets of high-copy *Mu* transposons can be separately amplified in a manner analogous to AFLP technology (*15*). This reduces the

complexity of the amplification products and facilitates the identification of the transposon that co-segregates with the mutant allele in a segregating family.

1. For the co-segregation analysis, DNA samples from multiple pairs of genetically related plants need to be compared. In each pair, one DNA sample carries the mutant allele; the other does not carry the mutant allele. Ideally, the genetic background of each pair should be highly similar as is the case for Bulk Segregant Analysis (*16*). For example, DNAs collected from families carrying the mutant allele and from sibling families without the mutant allele would be suitable for the co-segregation analysis. Due to the high copy number of *Mu* elements in the genome, multiple "mutant-specific" amplicons may be identified from the initial analysis. Once the mutant-specific *Mu* insertion(s) is identified, another round of PCR screen should be conducted using a *Mu* primer and a candidate gene-specific primer targeting that *Mu* insertion with individuals in a larger population for confirmation. It is also highly recommend that the cloning of the candidate gene be confirmed with additional mutant alleles.
2. Follow section 3.1 (steps 1-4) to obtain first PCR product for each DNA sample.
3. The first PCR product is diluted 10 times and 2 µl of diluted product is used in the following second PCR with nested primers, a *Mu* primer and one of the 32 *Nsp*I tail primers, using AmpliTaq Gold® DNA Polymerase. Thus there will be a total of 32 PCR reactions.

| Component | Amount per well (µl) |
|---|---|
| Water | 16 |
| GeneAmp 10X PCR Buffer II | 3 |
| 2mM dNTP | 4 |
| $MgCl_2$ (25 mM) | 1.8 |
| MuTIR or Mu53s (5 µM) | 1.5 |
| NspI tail primer* (5 µM) | 1.5 |
| 10X diluted first PCR product | 2 |
| AmpliTaq Gold (5 U/µl) | 0.2 |

| | |
|---|---|
| **Total** | **30** |

*It is likely that these primers could be further optimized by alternative designs.

This PCR program consists of 94°C for 10 min; 35 cycles of 94°C for 30 s, 60°C 45 s, 72°C 2.5 min; and a final extension at 72°C for 10 min.

4. Because multiple species of amplicons will be produced, PCR reactions are subject to electrophoresis to resolve various amplicons. Analyze all or at least 20 μl of each PCR reaction via agarose gel (2%) electrophoresis. Bands that are only present in mutant pool (mutant-specific) but not in control pool can be cut from the gel and separately purified with the Qiagen gel extraction kit (Qiagen, cat# 28704). Each resulting product can be sequenced directly, or sequenced following TOPO cloning per TOPO TA Cloning kit protocol (Invitrogen).

### 3.3 DLA-454

DLA-454 combines DLA with 454 pyrosequencing to amplify and sequence *Mu* flanking regions in a high-throughput manner. To enable subsequent 454 sequencing the *Mu*-specific primers and primers based on the single-stranded oligo are concatenated with the 454 sequencing primer to generate various composite primers. DNA barcodes (*17*) are inserted between the 454 sequencing primer and the *Mu*-specific primer to allow different input samples that will be pooled in the same 454 run to be distinguished after sequencing (*see* **Section 2**).

1. Collect DNA samples of interest and quantify each DNA sample.
2. Follow section 3.1 (steps 1-4) to obtain first PCR product for each DNA sample.
3. Dilute the first PCR product 10 times and use 2 μl of diluted product in the following second PCR with composite primers using AmpliTaq Gold® DNA polymerase.

| Component | Amount per well (μl) |
|---|---|
| Water | 16 |
| GeneAmp 10X PCR Buffer II | 3 |

| | |
|---|---|
| 2mM dNTP | 4 |
| $MgCl_2$ (25 mM) | 1.8 |
| barcodeMu (5 μM) | 1.5 |
| BnspI-P (5 μM) | 1.5 |
| 10X diluted first PCR product | 2 |
| AmpliTaq Gold (5 U/μl) | 0.2 |
| **Total** | **30** |

This PCR program consisted of 94°C for 10 min; 20 cycles of 94°C for 30 s, 60°CC 45s, 72° 2.5 min; and a final extension at 72°C for 10 min.

4. After PCR, purify the PCR product following the protocol provided for PCR purification with the Qiaquick PCR purification kit. Dissolve the purified product in 30 μl EB buffer. Measure each purified PCR product via Nanodrop.
5. Add the same amount of purified PCR product in a pool. This pool is the DNA library that includes the 454 sequencing primer for 454 sequencing. To ensure that the library was prepared correctly, the pooled DNA can be TOPO cloned per the protocol of TOPO TA Cloning kit and sequenced (~10 DNA molecules) with Sanger sequencing technology. If most of the sampled amplicons exhibit the expected sequence, the DNA library is ready for 454 pyrosequencing.

### 3.4 Other Adaptations

Next-generation sequencing (NGS) is a fast-evolving technology. Several other NGS platforms can also be used in conjunction with DLA. Compared with 454 pyrosequencing, Illumina sequencing promises higher amount of output data at a lower cost; Ion Torrent sequencing has a shorter running time at a relatively lower price. These two platforms can be easily integrated with DLA by modifying amplification primers. It is likely that future improvements in NGS

technologies will broaden the applications of DLA and further reduce the cost of sequencing DLA products.

**4. Notes**

1. In the protocol described here, we use *Nsp*I to digest genomic DNA. This enzyme, which exhibits high efficiency and fidelity, has a 6-base degenerate recognition site (5' RCATGY 3') and therefore generates DNA fragments with a 4-base 3' overhang and an average size of ~500 bp. The single-stranded oligo was designed to match this 3' overhang for the ligation. No modifications are required during oligo synthesis.
2. Other enzymes that generate 3' overhangs could also be used, including those that generate overhangs smaller than the 4-base overhang employed here. In fact, the Guoying Wang lab at the Chinese Academy of Agricultural Science successfully adapted DLA for an enzyme, *Hha*I, that generates a 2-base 3' overhang (personal communication, Jiankun Li, Jun Zheng and Guoying Wang).
3. One concern with using enzymes that generate 3' overhangs is that they could potentially lead to the generation of undesirable PCR products via the ligation of the single-stranded oligo to both ends of a genomic DNA fragment which could then be PCR amplified (*13*)**.** Although these amplified artifacts would not be sequenced in the current protocol, in an effort to eliminate them we have explored the use of enzymes that generate 5' overhangs. A 5' phosphorylated single-stranded oligo is used for the ligation. The 5' digestion overhang can not be extended, avoiding the introductions of DNA artifacts. We have tested several enzymes that generate 5' overhangs, such as *Bfu*CI, sufficiently to demonstrate that this strategy has potential, but it still requires further testing.

**References**

1. Benito, M. I., and Walbot, V. (1997) Characterization of the maize Mutator transposable element MURA transposase as a DNA-binding protein, *Mol Cell Biol 17*, 5165-5175.
2. Devon, R. S., Porteous, D. J., and Brookes, A. J. (1995) Splinkerettes--improved vectorettes for greater efficiency in PCR walking, *Nucleic Acids Res 23*, 1644-1645.
3. Edwards, D., Coghill, J., Batley, J., Holdsworth, M., and Edwards, K. J. (2002) Amplification and detection of transposon insertion flanking sequences using fluorescent muAFLP, *Biotechniques 32*, 1090-1092, 1094, 1096-1097.
4. Frey, M., Stettner, C., and Gierl, A. (1998) A general method for gene isolation in tagging approaches: amplification of insertion mutagenised sites (AIMS), *Plant J 13*, 717-821.
5. Hengen, P. N. (1995) Vectorette, splinkerette and boomerang DNA amplification, *Trends Biochem Sci 20*, 372-373.
6. Jones, D. H., and Winistorfer, S. C. (1992) Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA, *Nucleic Acids Res 20*, 595-600.
7. Kilstrup, M., and Kristiansen, K. N. (2000) Rapid genome walking: a simplified oligo-cassette mediated polymerase chain reaction using a single genome-specific primer, *Nucleic Acids Res 28*, E55.
8. O'Malley, R. C., Alonso, J. M., Kim, C. J., Leisse, T. J., and Ecker, J. R. (2007) An adapter ligation-mediated PCR method for high-throughput mapping of T-DNA inserts in the Arabidopsis genome, *Nat Protoc 2*, 2910-2917.
9. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C., and Markham, A. F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones, *Nucleic Acids Res 18*, 2887-2890.
10. Uren, A. G., Mikkers, H., Kool, J., van der Weyden, L., Lund, A. H., Wilson, C. H., Rance, R., Jonkers, J., van Lohuizen, M., Berns, A., and Adams, D. J. (2009) A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites, *Nat Protoc 4*, 789-798.
11. Settles, A. M., Latshaw, S., and McCarty, D. R. (2004) Molecular analysis of high-copy insertion sites in maize, *Nucleic Acids Res 32*, e54.
12. Walbot, V., and Warren, C. (1988) Regulation of Mu element copy number in maize lines with an active or inactive Mutator transposable element system, *Mol Gen Genet 211*, 27-34.
13. Liu, S., Dietrich, C. R., and Schnable, P. S. (2009) DLA-based strategies for cloning insertion mutants: cloning the gl4 locus of maize using Mu transposon tagged alleles, *Genetics 183*, 1215-1225.
14. Liu, S., Yeh, C. T., Ji, T., Ying, K., Wu, H., Tang, H. M., Fu, Y., Nettleton, D., and Schnable, P. S. (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome, *PLoS Genet 5*, e1000733.
15. YUNIS I, SALAZAR M, YUNIS EJ (1991) HLA-DR generic typing by AFLP. Tissue Antigens 38: 78–88

16. Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations, *Proc Natl Acad Sci U S A 88*, 9828-9832.
17. Qiu, F., Guo, L., Wen, T. J., Liu, F., Ashlock, D. A., and Schnable, P. S. (2003) DNA sequence-based "bar codes" for tracking the origins of expressed sequence tags from a maize cDNA library constructed using multiple mRNA sources, *Plant Physiol 133*, 475-481.