

Author Manuscript

Faculty of Biology and Medicine Publication

This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Published in final edited form as:

Title: Identification of HIV integration sites in infected host genomic DNA.

Authors: Ciuffi A, Barr SD

Journal: Methods (San Diego, Calif.)

Year: 2011 Jan

Volume: 53

Issue: 1

Pages: 39-46

DOI: 10.1016/jymeth.2010.04.004

Title Page:

Title: Identification of HIV integration sites in infected host genomic DNA.

Authors: Angela Ciuffi² and Stephen D. Barr^{1*}

Affiliations:

¹The University of Western Ontario, Department of Microbiology and Immunology, Dental Sciences Building Room 3006B, London, Ontario, Canada.

² Institute of Microbiology, University Hospital Center and University of Lausanne, Bugnon 48, 1011 Lausanne, Switzerland.

***Manuscript Correspondence:**

stephen.barr@uwo.ca, Tel: 519-661-2111 x83438, Fax: 519-661-3499

Acknowledgements: SDB is supported by the Canadian Institutes of Health Research; AC is supported by the Swiss National Science Foundation. Many thanks to our mentor Prof. Frederic Bushman for his guidance and support all these years.

Abstract: (250 words)

The integration of the Human Immunodeficiency Virus (HIV) genetic information into the host genome is fundamental for its replication and long-term persistence in the host. Isolating and characterizing the integration sites is central to understanding the forces dictating HIV integration site selection. The methods outlined in this article describe a highly efficient and precise technique for identifying HIV integration sites in the host genome on a small scale with cloning and standard sequencing, and on a massive scale with 454 pyrosequencing.

Keywords:

HIV, integration site mapping

1. Introduction:

The HIV genome is approximately 10kb in length and is flanked by two long terminal repeat (LTR) sequences. Following entry of HIV into its target cell the viral RNA genome is converted into a double-stranded DNA molecule. The HIV DNA associates with several viral and host proteins to form a multimeric complex that bridges both HIV LTRs in what is referred to as a “preintegration complex” (PIC). Once assembled, the integrase enzyme component of the PIC primes the viral DNA ends in a process called 3’ processing. 3’ processing involves cleavage of the terminal dinucleotide immediately 3’ of a conserved CA dinucleotide motif of each LTR, resulting in two recessed 3’-hydroxyl groups [1-5].

After docking with the host genome, the viral DNA undergoes a strand transfer reaction that involves a nucleophilic attack by each of the two recessed 3’-hydroxyl groups on a 5’-phosphate of the target DNA. The points of joining of each end of the LTRs are staggered by five base pairs across the major groove of the target DNA helix. After joining, the intervening five base pairs are melted, yielding gaps at the junctions of proviral and target DNA. The integration reaction is completed when the two protruding 5’ proviral nucleotides are trimmed and the gaps repaired, most likely by host repair enzymes [6,7]. The end result is the permanent integration of the HIV genome into a specific site on the host genome. The precise mapping of the resulting integration site, to the resolution of a single nucleotide, can be useful for determining the integration frequency near genomic features such as genes, CpG islands, repeat elements, DNA methylation sites, promoter elements, and more.

1.1. Overview of the method

Cells are infected with replication competent or incompetent HIV for a period of approximately 48 hours to allow for integration to take place (Fig. 1). Genomic DNA containing

the integrated proviruses is isolated and digested with restriction enzymes in order to generate genomic fragment sizes suitable for PCR amplification, cloning, and sequencing. Linker DNA is added to the ends of the digested genomic DNA and the ligated DNA is PCR amplified using primers specific for the linker DNA and the HIV LTR. Genomic DNA immediately flanking one of the HIV LTRs is amplified and sequenced to identify the single nucleotide immediately adjacent to the end of the HIV LTR (representing the integration site), and to query the host genome to identify the genomic location and features associated with that particular integration site.

2. Method:

2.1. Production of replication competent or incompetent HIV-based particles

Replication competent HIV particle production must be carried out in a certified Biosafety Level 3 facility. Seed one or more 10cm² dishes with 1x10⁶ 293T cells (obtained from American Type Culture Collection) in Dulbecco's modified Eagle's medium (DMEM; Lonza) supplemented with 10% heat-inactivated fetal calf serum (Hyclone), 100 U/ml of penicillin (Cellgro) and 0.1 mg/ml of streptomycin (Sigma) (D-10). Grow at 37°C until cells reach a confluence of 40-50%. Four hours before transfection, replace the media with 10 ml of fresh pre-warmed media. For production of replication competent HIV, transfect 30μg of a plasmid encoding a full-length provirus of HIV (eg. LAI or R9 [8]) using the standard calcium phosphate transfection method. For replication incompetent HIV, co-transfect: 12μg of pΔR8.91 (coding for Gag, Pol, Tat and Rev), 12μg of pSINcPPTeGFP (containing the LTR sequences and the packaging signal, and coding for a GFP reporter transgene), and 5μg of pMD.G (coding for the VSV-G pseudotyping envelope) [9,10]. After 12-16 hour incubation, aspirate the media, wash the cells gently with 1x Phosphate Buffered Saline (PBS) (137 mM NaCl; 2.7 mM KCl; 4.3 mM

Na₂HPO₄; 1.47 mM KH₂PO₄; adjust to a final pH of 7.4.), add 7 ml of pre-warmed D-10 media and incubate a further 48 hours. Gently swirl the plate of cells and remove the supernatant.

Clarify the supernatant by centrifugation at 500xg for 5 mins and filter through a 0.45µm membrane. Aliquot virus and freeze at -80°C.

Tips: Relication competent HIV can be concentrated using Amicon low-speed centrifugation devices with a 100 kDa cut-off membrane such as Centricon-Plus 70 or Jumbosep Centrifugal Devices (Pall Corporation). Make sure that the devices fit your Biosafety centrifuge containers. Replication incompetent HIV particles can be concentrated, either as replication competent HIV, or by layering the virus-containing supernatant over a 20% sucrose cushion and centrifuged at 23,000rpm for 2 hours at 4°C in a Beckman SW-28 rotor. Pelleted virus can be resuspended in ~1ml of D-10.

2.1.2 Determination of HIV titer

Replication competent HIV can be titrated in several ways. One relatively easy way is by using GHOST(3)R3/X4/R5 cells [11] (NIH AIDS Research & Reference Reagent Program, Cat.# 3943). These cells contain a green fluorescent protein (GFP) marker under the control of the Tat-responsive HIV-2 long terminal repeat (LTR). Upon infection with HIV, and consequently the introduction of Tat, transcription from the LTR is initiated leading to the synthesis of GFP. Infected cells can be detected by fluorescence activated cell sorting (FACS) to obtain a multiplicity of infection (MOI). Upon infection with replication incompetent HIV, a GFP marker gene is integrated into the target genome via LTR-directed integration. Successful infection and integration can be measured by FACS to obtain the MOI. For integration site mapping, we generally want to use an MOI of around 10 to obtain an infection rate of 30-70% to

maximize recovery of integration sites. However, integration sites can be obtained with lower infection rates. We have obtained integration sites with as low as a 2% infection rate of primary human macrophages [12].

2.2. Infection of cells

Adherent cells- are grown to 50% confluence in a vessel of choice. A 6-well plate is often sufficient for adherent cells. Immediately prior to infection, virus is incubated for 1 hour at 37°C with RNase-free DNaseI (10 U per 100μl of virus) (Roche Applied Science, Mannheim, Germany, Cat.# 04716728001) to digest any plasmid carried over into the virus preparation from the transfection. Media is aspirated from the cells and replaced with media containing sufficient virus to cover the cells (typically, 1 ml for a 6-well) and supplemented with 2mg/ml polybrene (Sigma Cat.#H9268) or 10 μg/ml DEAE-dextran to improve infection efficiency. Always include an uninfected sample. After 1.5 to 4 hours, fresh pre-warmed media is added to the virus and incubated a further 48 hours to allow integration of the vector. Longer incubation times may result in the selection of cells containing integration sites that promote cell survival, and selection against integrations that are detrimental to the cell; thereby possibly introducing a bias. Adjust the incubation time accordingly depending on the question under study. *Suspension cells*- Infect 2×10^6 cells in 1 ml final volume in 24-well and proceed as described above for adherent cells.

2.3. Isolation of genomic DNA containing integration sites

Aspirate media from infected and uninfected cells, wash twice with PBS, and proceed to isolate genomic DNA according to your method of choice. We routinely use the DNeasy Tissue Kit (Qiagen, Cat.# 69506) according to the manufacturer's instructions. Check approximately

one tenth of your sample on a 1% agarose gel containing ethidium bromide or another nucleic acid stain to verify its integrity. You should observe a single high molecular weight band (Fig. 2A). Smearing below this band is indicative of shearing and/or degradation and should be avoided. Quantify your purified DNA sample by spectrophotometry (OD 260).

Tips: Avoid rough handling of the sample to preserve the integrity of the genomic DNA (eg. no vortexing or vigorous pipeting). You will require at least 2 μ g of genomic DNA per infected cell sample.

2.4. DNA digestions

Mapping integration sites using the method outlined in this article requires restriction enzyme digestion, linker ligation, PCR amplification, and cloning of genomic fragments adjacent to the integrated vector. Therefore, we must generate restriction fragments of lengths suitable for cloning and PCR amplification. When used to digest human genomic DNA, the restriction enzyme MseI completely digests the genomic DNA to yield sizes less than 10kb with the majority of the sizes in the range of ~0.1-3kb. Alternatively, a combination of AvrII, SpeI, and NheI can be chosen as previously described [13]. Restriction digestion reagents were purchased from New England Biolabs (Ipswich, MA). Digest your genomic samples as follows:

2 μ g genomic DNA (from uninfected or infected cells)

20 μ l NEB2 (10x stock)

12 μ l MseI [10 U/ μ l]

2 μ l BSA (100x stock)

XX μ l H₂O (total volume of 200 μ l)

As a positive control:

10 µg pSINcPPTeGFP

10 µl NEB2 (10x stock)

6 µl MseI [10 U/µl]

1 µl BSA (100x stock)

XX µl H₂O (total volume of 200µl)

Incubate the digestions overnight at 37°C. The quality of the digestion can be checked by electrophoresis using 5 µl of purified digested DNA (Figure 2A).

Tips: Digestions of the genomic samples will be complete if the genomic samples lack the high molecular weight band seen in the undigested control sample and a visible smear in the range of ~1-10kb with the brightest part of the smear in the range of ~0.5-5kb.

Samples can be left to digest overnight if it is more convenient.

2.5 Purification of digested genomic DNA

Purify the entire sample with the method of your choice. We prefer the Strataprep PCR purification kit (Stratagene, La Jolla, CA, Cat.#400771), however any PCR purification kit should suffice. Purify the samples according to the manufacturer's instructions, except split the digested pSINcPPTeGFP samples into two fractions and purify with two separate spin columns to avoid over-loading the columns. Elute the samples with 50µl of 10mM Tris pH 8.0. Check the quality of the digestion by electrophoresis using 5 µl of purified digested DNA (Figure 2A).

Tips: Samples can be stored at 4°C for up to a week or frozen at -20°C for longer storage.

Prolonged storage of the digested samples may reduce yield due to damage to the single-stranded restriction over-hangs.

2.6. Linker preparation

Linker sequences are added to the ends of the digested genomic samples in order to provide a primer binding site for PCR amplification of the LTR/genomic DNA junctions. You will need to order two oligos from a company such as Eurofins MWG Operon (www.operon.com):

1.) MseI Linker (+):

5' GTAATACGACTCACTATAAGGCTCCGCTTAAGGGAC 3'

2.) MseI Linker (-):

5' [Phos]TAGTCCCTTAAGCGGAG[AmC7-Q] 3'

The [AmC7-Q] modification of the 3' nucleotide of the MseI Linker (-) is necessary to prevent linker-linker amplification to the detriment of desired LTR-linker amplification.

Prepare the linker by mixing 20 µl MseI Linker (+) [40 µM] + 20 µl MseI Linker (-) [40 µM] in a PCR tube. Denature the mixture at 90°C for 5 mins in a PCR machine and then cool 1°C every 3 mins until the temperature reaches 20°C. Alternatively, turn the PCR machine off and let the heat block cool on its own until its temperature reaches ~room temperature. The annealed linker products will be referred to as “adapter mix” hereafter.

Tips: Store the annealed linker at 4°C for no longer than one week to preserve the integrity of the overhangs. Do not freeze. Freeze-thawing will cause damage to the single-stranded overhangs.

2.7. Linker Ligation

Ligation reagents are from New England Biolabs. Set up the following four ligation reactions:

1) For genomic DNA negative control, mix:

2 μ l Ligase Buffer (10x Stock)

1 μ l T4 DNA ligase (400 U/ μ l)

0 μ l adapter mix

13.5 μ l digested genomic DNA

3.5 μ l H₂O (total volume of 20 μ l)

2) For genomic DNA (infected and uninfected), mix:

2 μ l Ligase Buffer (10x Stock)

1 μ l T4 DNA ligase (400 U/ μ l)

3.5 μ l adapter mix

13.5 μ l of the MseI-digested genomic DNA

0 μ l H₂O (total volume of 20 μ l)

3) For plasmid DNA negative control, mix:

2 μ l Ligase buffer (10x Stock)

1 μ l T4 DNA ligase (400 U/ μ l)

0 μ l adapter mix

10 μ l digested pSINcPPTeGFP (from section 2.4, linearized with MseI) (~2 μ g)

7 μ l H₂O (total volume of 20 μ l)

4) For plasmid DNA positive control, mix:

2 μ l Ligase buffer (10x Stock)

1 μ l T4 DNA ligase (400 U/ μ l)

3.5 μ l adapter mix

10 μ l digested pSINcPPTeGFP (from section 2.4, linearized with MseI) (2 μ g)

3.5 μ l H₂O (total volume of 20 μ l)

Incubate ligations at 21°C for ~14 hours and then heat-inactivate the ligase at 65°C for 15 mins.

Digest the ligation reactions with SacI/DpnI as follows:

20 μ l of the heat-inactivated ligation mix (unpurified)

3 μ l NEB4 (10x stock)

0.5 μ l BSA (100x stock)

22.5 μ l H₂O

2 μ l DpnI [20U/ μ l] (to get rid of plasmid contamination)

2 μ l SacI [20U/ μ l] (to limit amplification of internal fragment)

0 μ l H₂O (total volume of 50 μ l)

Separate the plasmid DNA positive control (reaction 4 of step 2.8) in two: one half digested with SacI and DpnI as control and the other half with SacI only, and that will be used as positive control for subsequent PCR amplifications. Incubate the digestions at 37°C for 4 hours.

Tips: Check 5 μ l of the plasmid ligations on a 1% agarose gel to ensure that ligation of the linkers took place (Fig. 2B). If so, the dominant plasmid form should remain linear and circularized plasmid form should be in minor quantities. DpnI digestion will ensure the digestion of the original plasmid DNA carryover, while SacI digestion will prevent the amplification of the internal HIV genome contaminant (Fig. 1 and Fig. 2B).

2.8. PCR amplification

The next step uses PCR to amplify the junctions between the integrated HIV LTR sequence and the adjacent genomic sequence. Two rounds of PCR are used to increase the

specificity of the amplified fragments for authentic junctions. The first PCR uses primers specific for the HIV LTR and linker sequence, and the second nested PCR uses primers internal to the first set of primers, yet still specific for the HIV LTR and linker sequence. Use the Advantage 2 PCR Polymerase Mix from Clontech (Cat.# 639201). 0.2ml thin-wall tubes were used (BioRad, Cat.# 223-9473). ** NOTE: If you plan on sequencing a small number of integrations sites, proceed with steps 2.8.1 to 2.11; if sequencing a large number of sites, finish step 2.8.1 and then proceed to step 2.12 for the nested PCR2 for high-throughput sequencing.

2.8.1 First round PCR (PCR1):

LTR-specific primer (LTR1)- 5' CTTAAGCCTCAATAAAGCTTGCCTTGAG 3'

Linker-specific primer (MseL1)- 5' GTAATACGACTCACTATAGGGC 3'

Perform the PCR1 reaction on each of the 5 ligation reactions from section 2.7 plus an additional "No DNA" negative control as described below:

0.5 µl MseL1 primer [15 µM]

0.5 µl LTR1 primer [15 µM]

2.5 µl Buffer 2PCR (10x stock)

0.5 µl dNTPs (50x stock) (10mM each)

0.5 µl Advantage 2 PCR Polymerase mix

5 µl Linker ligated DNA sample (from section 2.7)

15.5 µl H₂O

Total : 25 µl

PCR cycling conditions:

1x94°C for 1 min

5x 94 °C for 2 secs; 72 °C for 1 min

20x 94 °C for 2 secs; 67 °C for 1 min

1x 72 °C for 4 mins

Hold at 4°C.

Check 5 µl on a 1% agarose gel. You should expect to see a very faint smear (ranging from ~0.1-2kb) for the infected genomic DNA samples and a distinct band for the plasmid positive control sample (Fig. 2C). No bands should be present in the other samples.

2.8.2. Second round nested PCR (PCR2):

LTR-specific primer (LTR2)- 5' AGACCCTTTAGTCAGTGTGGAAAATC 3'

Linker-specific primer (MseL2)- 5' AGGGCTCCGCTTAAGGGAC 3'

Dilute each PCR1 reaction 1:200 and use 1µl of this dilution as your template for PCR2. Also, substitute the primers with MseL2 and LTR2. PCR reagents and cycling conditions are as described for PCR1. If you do not observe a visible smear (ranging from ~0.1-2kb) for the infected genomic DNA samples or a distinct band for the plasmid positive control, repeat PCR2 using a 1:50 dilution of the PCR1 reactions (Figure 2D). Additionally, you may want to try optimizing the annealing conditions of the primers using the plasmid control and your specific thermocycler model.

2.9. Gel purification of PCR products

We use the S.N.A.P.™ UV-free Purification Kit (Invitrogen, Cat.# K2000) to purify the PCR2 DNA. Prepare 50ml of a 1% agarose gel in TAE buffer (40mM Tris base, 20mM acetate, 1mM EDTA) containing 40µl of crystal violet solution (2mg/ml stock) (provided in the kit). This step involves separating your PCR2 sample on an agarose gel in order to purify large and small sized fragments. As a convenient marker to distinguish the large (~3-6kb) from the smaller (~0.1-3kb) sized fragments, we prepare a marker sample consisting of 1µl crystal violet loading dye (6x stock), 2µl gel loading dye (6x stock) (Promega, Cat.# G190A, Madison, WI), and 3µl H₂O. The xylene cyanol present in this sample will migrate at ~4kb, whereas the bromophenol blue will migrate at ~400bp.

To prepare your PCR2 sample, add 4µl of the 6x crystal violet loading dye to 20µl of your PCR2 product. Load and run your marker and PCR2 samples only enough to separate the xylene cyanol and bromophenol blue dyes in your marker sample. This is a very short run (~10mins at 90V) in order to keep the DNA concentration high in a minimal amount of gel for purification. Using the marker sample as a guide, harvest the smaller sized fragments found just below the bromophenol blue dye and the xylene cyanol dye bands (Figure 3). Purify the DNA from each gel slice using the S.N.A.P.™ UV-free Purification Kit according to the manufacturer's instructions, except elute the DNA in 40µl low TE buffer. Check 10µl of the purified product on a 1% agarose gel.

Tips: Use a separate gel and scalpel for each DNA sample to prevent contamination.

2.10. TOPO cloning

The small DNA fragments from section 2.9 are cloned into the TOPO TA cloning kit (Invitrogen, Cat.#K4575) with OneShot TOP10 electrocompetent *Escherichia coli* (Invitrogen, Cat.#K4700). Set up the following reactions:

TOPO TA: 4 μ l of purified small DNA fragments from section 2.9

1 μ l of Salt Solution provided in the kit (diluted 1:4)

1 μ l TOPO TA plasmid

Incubate at 21°C for 30 mins and then place on ice.

Bacterial Transformation: Add 2 μ l of each TOPO reaction into 50 μ l of the TOP10 electrocompetent bacteria and place into a chilled electroporation cuvette (2mm gap). Gently mix and place on ice for 5 mins. Electroporate the samples at 2.5kV (preset bacterial program #2 on the BioRad GenePulser System). Immediately place the electroporated cells 250 μ l (for TOPO TA) of SOC medium in a polypropylene round bottom tube (Falcon, Cat.# 2059, Becton-Dickinson, NJ). Incubate the bacteria at 37°C with shaking (225rpm) for 1 hour. Spread 30 μ l and 120 μ l of the TOPO TA bacteria, onto LB agar plates supplemented with 50 μ g/ml kanamycin. Incubate the plates at 37°C for ~17 hours until the bacterial colonies are the size of a pin head.

2.11. Picking colonies

Add 100 μ l of LB media containing 50 μ g/ml kanamycin into each well of two 96-well plates under sterile conditions. The second plate will serve as a backup. Using a sterile pipet tip, gently touch the tip to one single isolated bacterial colony from the plate. Place the tip into one well of one of the 96-well plates and swirl the tip one or two complete circles. Remove the tip and repeat with the corresponding well on the duplicate plate. Once the plate is filled, cover and place in a 37°C incubator without shaking overnight. Check for cloudiness in the wells (indicating bacterial growth). Add 34 μ l of LB medium containing 60% glycerol to each well

(use the multichannel pipet-quicker) and mix. Seal the plates with polyethylene PCR film adhesive (VWR International, Cat.# 60941-122) and freeze at -80°C. Send one plate for plasmid extraction and sequencing using M13-reverse (5' CAGGAAACAGCTATGAC 3') and M13-forward (5' TGTAAAACGACGGCCAGT 3') primers and keep the duplicate plate as backup.

Tips: One entire 96-well plate (in duplicate) can be seeded by hand in approximately 2-3 minutes. Alternatively, pick colonies from the bacterial plate and put into the well of one 96-well plate containing 150 µl LB-50µg/ml kanamycin in each well and grow over night. The following morning, add 100µl of 38% glycerol to each well, mix up and down and transfer 100µl of the cells into the duplicate back-up plate. Store at -80°C.

2.12. High-throughput (454) sequencing

2.12.1 Second round nested PCR2 for 454 sequencing (PCR2-454)

Linker-specific primer (MseL2-454)-

5' gcctccctcgccatcagAGGGCTCCGCTTAAGGGAC 3'

LTR-specific primer (LTR2-454)-

5' gcctgccagccgctcagN₍₄₋₈₎AGACCCTTTAGTCAGTGTGGAAAATC 3'

Lower case nucleotides indicate the sequence homologous to the primers used for 454 sequencing, Primer A and Primer B respectively. N₍₄₋₈₎ indicates the sequence barcode (4 to 8 nucleotides) that can be introduced when processing multiple samples in the same sequencing run, and is randomly generated by the user. Use barcodes when sample multiplexing is desired. Use a 4 nucleotide barcode when a small number of samples are processed and an 8 nucleotide barcode with larger amounts of samples are processed. Upper case nucleotides indicate the nested sequence primer that anneal in the LTR and linker regions respectively.

Perform the PCR2-454 as described for PCR1 (section 2.8.1), except dilute each PCR1 reaction 1:200 and use 1 μ l of this dilution as your template for PCR2 and substitute the primers with MseL2-454 and LTR2-454. If you do not observe a visible smear (ranging from ~0.1-2kb) for the infected genomic DNA samples or a distinct band for the plasmid positive control, repeat PCR2-454 using a 1:50 dilution of the PCR1 reactions (Figure 2D). Additionally, you may want to try optimizing the annealing conditions of the primers using the plasmid control and your specific thermocycler model.

2.12.2 Gel purification of PCR products

When processing large amounts of integration sites or multiple conditions, high-throughput sequencing, such as 454 pyrosequencing, will provide a speedy alternative. Since 454 sequencing sequences the PCR2-454 products directly, there is no need for TOPO cloning. Purify your PCR2-454 products as described in section 2.9, except harvest the DNA fragments ranging from 100-600bp, alternatively purify your PCR2-454 products using a Qiaquick PCR purification column (Qiagen). Use ~150ng for the 454 high-throughput sequencing. Consult your 454 pyrosequencing facility for choice of kits to use. Most facilities will process your PCR2-454 samples for you.

2.12.3 High-throughput pyrosequencing

Pyrosequencing technology, as implemented by the 454 Life Sciences [14], is now regularly used for high-throughput sequencing of host genomic regions flanking HIV integration sites [15-23]. For this, purified PCR2-454 products (from section 2.12.2) are mixed with beads pre-coated with oligonucleotides complementary to primers A and B in a dilute solution, to ensure that only one DNA strand is bound per bead. These beads are then used to perform an emulsion PCR using primers A and B that were previously introduced during the nested PCR

(lower case nucleotides) (section 2.12.1). Beads with bound DNA are then distributed on a plate and DNA fragments are sequenced using 454 technology, allowing the sequencing of thousands of DNA molecules to be processed at the same time [24].

2.13. Processing sequences

Once the sequences are received, they need to be processed which involves identifying the integration site nucleotide adjacent to the HIV LTR and trimming away the linker and HIV LTR sequences so that: 1) the sequence can be compared with other sequences to delete duplicate sequences, and 2) the genomic fragment can be queried against the host cell genome using a BLAT server such as the one provided through the UCSC Genome Browser (<http://genome.ucsc.edu/>). Several hundred sequences can be processed by hand in Microsoft Word or Excel making use of the “find and replace” function. Alternatively, a text editor program such as Ultraedit (<http://www.ultraedit.com/>) can be used to speed up the processing, however a basic understanding of Perl or computer programming is required and can be self-taught through the Help menu of the program.

Typically, the criteria for a useable sequence is: (1) a perfect match to the vector terminus (ie. the LTR), (2) a $\geq 98\%$ match to genomic DNA, (3) a unique best hit to the host genome, and (4) a match to host cell DNA beginning within 3 bp of the end of the vector DNA. Several hundred sequences can be queried against the host genome using the BLAT Search Genome feature of the UCSC Genome Browser. Large scale queries ($>100,000$) are best performed using a MySQL database and local BLAT server, such as the one used by Rick Bushman’s lab (<http://bushman-lab.pbworks.com/Intro-to-Insipid>).

3. Concluding Remarks:

The above method describes the isolation of HIV integration sites in the host genome. This method can be adapted for the isolation of any insertional element in any genome. Once the integration sites have been isolated, the downstream applications are numerous and ever-expanding, some of which are described in [25]. Additional considerations are warranted depending on the biological questions proposed. For example, the method described herein involves the isolation and amplification of genomic DNA isolated after cleavage with a restriction enzyme. If one wishes to determine whether there is a significant number of integration sites in genomic features, measures must be taken to account for restriction site bias in recovery such as using random controls that have the same potential bias [13].

References

- [1] F. D. Bushman, T. Fujiwara and R. Craigie, *Science* 249 (1990) 1555-1558.
- [2] R. Craigie, T. Fujiwara and F. Bushman, *Cell* 62 (1990) 829-837.
- [3] R. A. Katz, G. Merkel, J. Kulkosky, J. Leis and A. M. Skalka, *Cell* 63 (1990) 87-95.
- [4] M. Katzman, R. A. Katz, A. M. Skalka and J. Leis, *J.Virol.* 63 (1989) 5319-5327.
- [5] P. A. Sherman and J. A. Fyfe, *Proc.Natl.Acad.Sci.USA* 87 (1990) 5119-5123.
- [6] R. Daniel, J. G. Greger, R. A. Katz, K. D. Taganov, X. Wu, J. C. Kappes and A. M. Skalka, *J.Virol.* 78 (2004) 8573-8581.
- [7] K. Yoder and F. D. Bushman, *J.Virol.* 74 (2000) 11191-11200.
- [8] S. D. Barr, J. R. Smiley and F. D. Bushman, *PLoS Pathog.* 4 (2008) e1000007.

- [9] A. Follenzi, L. E. Ailes, S. Bakovic, M. Gueuna and L. Naldini, Nat.Genetics 25 (2000) 217-222.
- [10] L. Naldini, U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma and D. Trono, Science 272 (1996) 263-267.
- [11] A. Morner, A. Bjorndal, J. Albert, V. N. Kewalramani, D. R. Littman, R. Inoue, R. Thorstensson, E. M. Fenyo and E. Bjorling, J.Virol. 73 (1999) 2343-2349.
- [12] S. D. Barr, A. Ciuffi, J. Leipzig, P. Shinn, J. R. Ecker and F. D. Bushman, Mol.Ther. 14 (2006) 218-225.
- [13] A. Ciuffi, K. Ronen, T. Brady, N. Malani, G. Wang, C. C. Berry and F. D. Bushman, Methods 47 (2009) 261-268.
- [14] M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley and J. M. Rothberg, Nature 437 (2005) 376-380.

- [15] F. D. Bushman, C. Hoffmann, K. Ronen, N. Malani, N. Minkah, H. M. Rose, P. Tebas and G. P. Wang, AIDS 22 (2008) 1411-1415.
- [16] G. P. Wang, A. Ciuffi, J. Leipzig, C. C. Berry and F. D. Bushman, Genome Res. 17 (2007) 1186-1194.
- [17] H. M. Marshall, K. Ronen, C. Berry, M. Llano, H. Sutherland, D. Saenz, W. Bickmore, E. Poeschla and F. D. Bushman, PLoS One 2 (2007) e1340.
- [18] P. McKenna, C. Hoffmann, N. Minkah, P. P. Aye, A. Lackner, Z. Liu, C. A. Lozupone, M. Hamady, R. Knight and F. D. Bushman, PLoS Pathog. 4 (2008) e20.
- [19] G. P. Wang, A. Garrigue, A. Ciuffi, K. Ronen, J. Leipzig, C. Berry, C. Lagresle-Peyrou, F. Benjelloun, S. Hacein-Bey-Abina, A. Fischer, M. Cavazzana-Calvo and F. D. Bushman, Nucleic Acids Res. 36 (2008) e49.
- [20] S. Hacein-Bey-Abina, A. Garrigue, G. P. Wang, J. Soulier, A. Lim, E. Morillon, E. Clappier, L. Caccavelli, E. Delabesse, K. Beldjord, V. Asnafi, E. MacIntyre, L. Dal Cortivo, I. Radford, N. Brousse, F. Sigaux, D. Moshous, J. Hauer, A. Borkhardt, B. H. Belohradsky, U. Wintergerst, M. C. Velez, L. Leiva, R. Sorensen, N. Wulffraat, S. Blanche, F. D. Bushman, A. Fischer and M. Cavazzana-Calvo, J.Clin.Invest. 118 (2008) 3132-3142.
- [21] G. P. Wang, B. L. Levine, G. K. Binder, C. C. Berry, N. Malani, G. McGarrity, P. Tebas, C. H. June and F. D. Bushman, Mol.Ther. 17 (2009) 844-850.
- [22] T. Brady, L. M. Agosto, N. Malani, C. C. Berry, U. O'Doherty and F. Bushman, AIDS 23 (2009) 1461-1471.

[23] T. Brady, Y. N. Lee, K. Ronen, N. Malani, C. C. Berry, P. D. Bieniasz and F. D. Bushman, Genes Dev. 23 (2009) 633-642.

[24] C. Hoffmann, N. Minkah, J. Leipzig, G. Wang, M. Q. Arens, P. Tebas and F. D. Bushman, Nucleic Acids Res. 35 (2007) e91.

[25] F. Bushman, M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli and C. Hoffmann, Nat.Rev.Microbiol. 3 (2005) 848-858.

Figure Legends:

Fig.1: Method overview. After infection, genomic DNA (black lines) containing proviruses (red line; red boxes correspond to LTRs) is extracted from the cells and digested with MseI (black arrows). A compatible linker (blue squares) is ligated to the digested DNA sites. To avoid amplification of 5' viral LTR, SacI digestion (blue arrow) is first performed, thereby preventing a successful PCR amplification. Primers (black arrows) annealing in the LTR and in the linker sequence are used for PCR amplification. Nested primers (green arrows) are then used for a second PCR. Upon gel purification, the PCR products are cloned into TOPO TA vectors and sequenced. Alternatively, the nested primers may contain, in addition to the nested primer sequence, a sequencing primer A and B, and a 4 to 8 nucleotide barcode (green arrow with black floating tail). Upon gel purification, the PCR products are processed by 454 high-throughput sequencing.

Fig. 2: Analysis of DNA products by agarose gel electrophoresis and ethidium bromide staining. A) Non digested and MseI-digested samples of genomic DNA or plasmid control. **B)**

Ligation of 0 or 3.5 µl adapter mix to the MseI-digested plasmid, followed by SacI and DpnI digestion. **C)** First PCR amplification of 3'-LTR-host DNA-adapter region using primers annealing to the LTR and to the linker. **D)** Nested PCR amplification using nested primers in the LTR and linker regions. L1000 : 1kb DNA ladder, L100 : 100 bp DNA ladder, S: genomic DNA sample, P: plasmid pSINcPPTeGFP control.

Fig. 3: Isolation of the digested/PCR-amplified genomic samples for TOPO cloning. A marker consisting of xylene cyanol and bromophenol blue is separated on a 1% agarose gel in parallel with the genomic sample. After a very short run (~10mins), the gel is placed on a clean surface and then a clean scalpel is used to isolate the small digested fragments from the gel.

Appendix A: List of materials required for the isolation of HIV integration sites.

- **Reagents**

- 293T cells (obtained from American Type Culture Collection)
- GHOST(3)R3/X4/R5 cells; (NIH AIDS Research & Reference Reagent Program, Cat.# 3943)
- Dulbecco's modified Eagle's medium (DMEM; Lonza)
- Heat-inactivated fetal calf serum (Hyclone)
- Penicillin 100U/ml (Cellgro)
- Streptomycin 0.1mg/ml (Sigma)
- Kanamycin (Sigma, Cat# 60615)
- Polybrene (Sigma Cat.#H9268)
- Agarose (Sigma)

- 1kb DNA ladder (Invitrogen)
- 100bp DNA ladder (Invitrogen)
- Gel loading dye (6x stock) (Promega, Cat.# G190A, Madison, WI)
- Replication competent or incompetent HIV (~MOI=10 to obtain 30-70% infected cells)
- pSINcPPTeGFP (Bushman Lab); See references {Naldini, L., et al. 1996; Zufferey, R., et al. 1998}
- Standard 6-well, 24-well, and 96-well plates (Falcon)
- Standard 0.2ml thin-wall tubes (BioRad, Cat.# 223-9473)
- Polyethylene PCR film adhesive (VWR International, Cat.# 60941-122)
- Standard 10cm bacterial plates (Falcon)
- Polypropylene round bottom tube (Falcon, Cat.# 2059, Becton-Dickinson, NJ)

- **Kits**

- DNeasy Tissue Kit (Qiagen, Cat.# 69506)
- Strataprep PCR purification kit (Stratagene, La Jolla, CA, Cat.#400771)
- Advantage 2 PCR Polymerase Mix from Clontech (Cat.# 639201)
- S.N.A.P.TM UV-free Purification Kit (Invitrogen, Cat.# K2000)
- TOPO TA cloning kit (Invitrogen, Cat.#K4575)
- TOPO XL cloning kit with OneShot TOP10 electrocompetent *Escherichia coli* (Invitrogen, Cat.#K4700)

- **Equipment/Facilities**

- Biosafety Level 3 facility for handling replication-competent HIV
- Standard cell culture 37°C incubator

- Standard Biosafety cabinet
- FACS machine for determining titer of virus
- Standard Spectrophotometer (OD 260)
- Standard thermocycler for PCR
- Standard horizontal gel electrophoresis equipment
- Standard table-top centrifuge (14,000 rpm max)
- BioRad GenePulser System for electroporating bacteria
- 37°C shaking incubator
- Sequencing facilities to harvest plasmids from bacteria in 96-well plates
- 454 pyrosequencing facility

- **Enzymes/Buffers (New England Biolabs)**

- RNase-free DNaseI (Roche Applied Science, Mannheim, Germany; Cat# 04716728001)
- MseI
- NEB2 Buffer (10x stock) (supplied with MseI)
- BSA Buffer (100x stock) (supplied with MseI)
- T4 DNA ligase (400 U/ μ l)
- T4 DNA ligase buffer (400 U/ μ l) (supplied with T4 DNA Ligase)
- DpnI
- SacI
- NEB4 Buffer (10x stock) (supplied with SacI)

- **Oligos**

- MseI Linker (+):

5' GTAATACGACTCACTATAAGGGCTCCGCTTAAGGGAC 3'

- MseI Linker (-): 5' [Phos]TAGTCCCTTAAGCGGAG[AmC7-Q] 3'
- LTR-specific primer (LTR1)- 5' CTTAAGCCTCAATAAAGCTTGCGCTTGAG 3'
- Linker-specific primer (MseL1)- 5' GTAATACGACTCACTATAAGGGC 3'
- LTR-specific primer (LTR2)- 5' AGACCCTTTAGTCAGTGTGGAAAATC 3'
- Linker-specific primer (MseL2)- 5' AGGGCTCCGCTTAAGGGAC 3'
- LTR-specific primer (LTR2-454)-
5' gccttgccagccgctcagN₍₄₋₈₎AGACCCTTTAGTCAGTGTGGAAAATC 3'
- Linker-specific primer (MseL2-454)-
5' gcctccctcgccatcagAGGGCTCCGCTTAAGGGAC 3'

Figure 1

[Click here to download high resolution image](#)

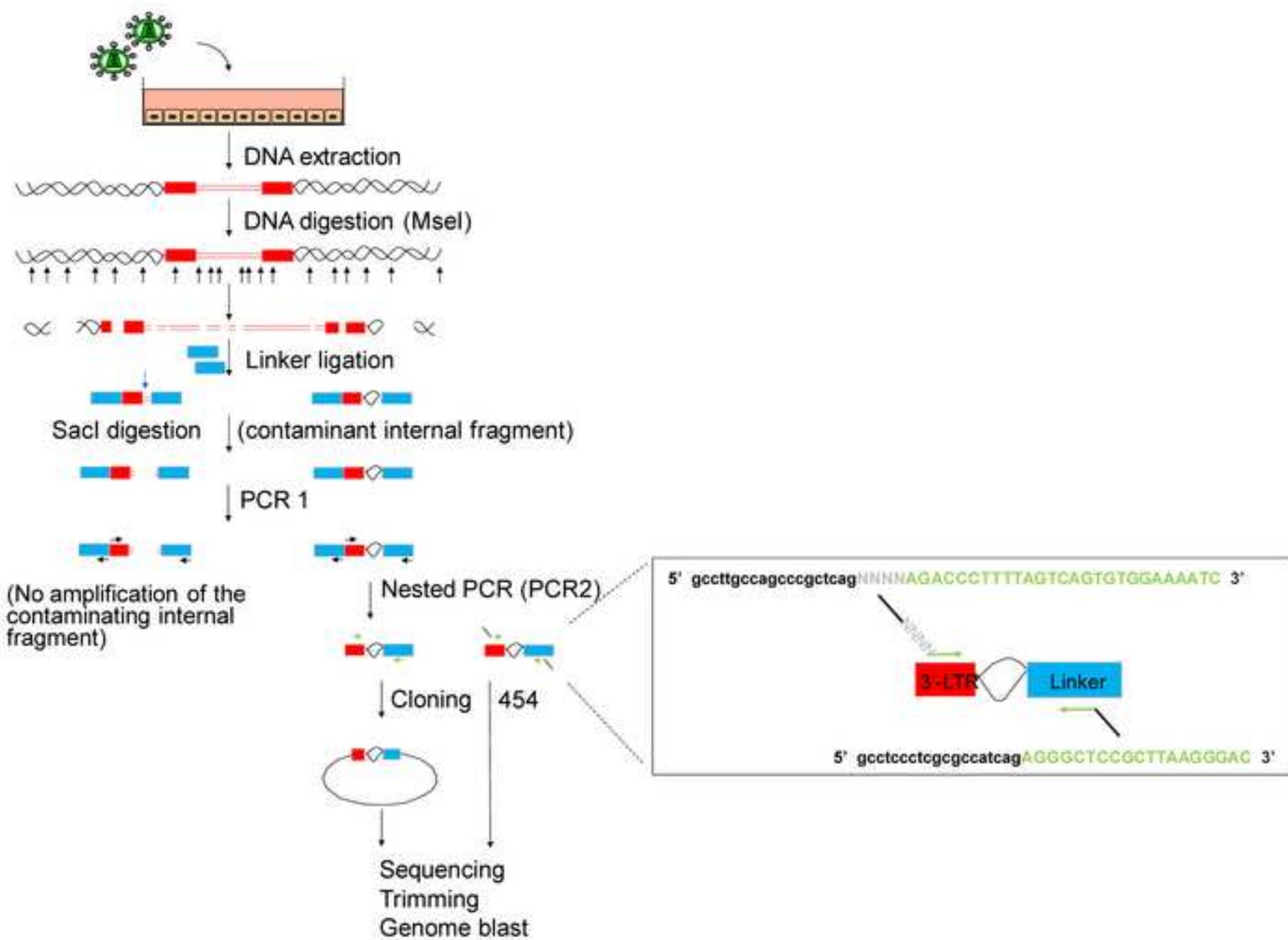


Figure 2

[Click here to download high resolution image](#)

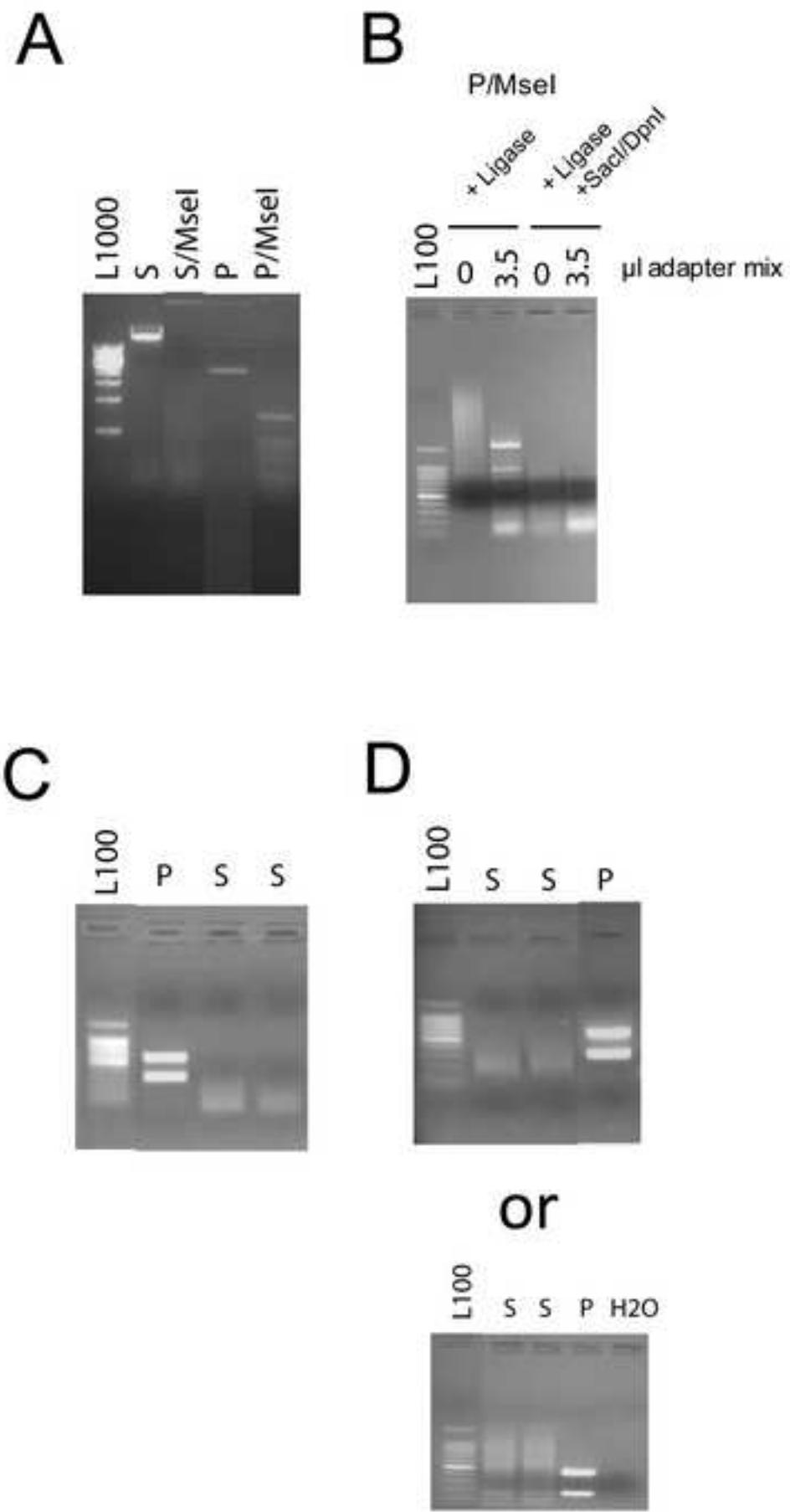


Figure 3

[Click here to download high resolution image](#)

