

## PCR-Based Strategies to Clone Unknown DNA Regions from Known Foreign Integrants

### *An Overview*

Eric Ka-Wai Hui, Po-Ching Wang, and Szecheng J. Lo

### 1. Introduction

Many foreign DNAs, such as some virus DNAs and almost all transposable elements (transposons), are capable of integrating host genomes, and the effects of integration can be pleiotropic. To investigate the mechanism and biological effect of foreign DNA insertions, characterization of the integration site, called integrant-host junction (IHJ), in the host genome becomes important. Traditional genomic library construction and screening for the cloning and analysis of IHJ are time-consuming, labor-intensive, and tedious. Therefore, a variety of efficient and reliable polymerase chain reaction (PCR)-based techniques have been developed. Application of the PCR to yield enough amounts of DNA for cloning and analysis is highly recommended especially for those specimens that are in a minute amount. Because the amplification process of PCR requires a pair of primers that can anneal to known sites at two end of the target DNA template, it seems that PCR is not applicable to IHJ searching because only one side of the fragment sequence in the integrant is known. A number of PCR-based techniques, however, have been developed to amplify the unknown cellular DNA flanking sequence from the foreign DNA. In this chapter, we introduce the PCR-based methodologies for the rapid acquisition of unknown DNA sequences. Based on the underlying principles, we classified these techniques into five categories: 1) PCR after intramolecular circularization; 2) interspersed repetitive sequence PCR (IRS-PCR); 3) ligation-anchored PCR (LA-PCR); 4) arbitrarily primed PCR (AP-PCR); and 5) reverse transcription PCR (RT-PCR). These techniques include inverse PCR (IPCR), partial IPCR (PI-PCR), long IPCR (LR-iPCR or LI-PCR), novel *Alu*-PCR, long interspersed repetitive element PCR (LINE-PCR), B1-PCR, vectorette-PCR, multipletstep-touchdown vectorette-PCR (MTV-PCR), long-distance vectorette-PCR

From: *Methods in Molecular Biology*, Vol. 192: *PCR Cloning Protocols*, 2nd Edition  
Edited by: B.-Y. Chen and H. W. Janes © Humana Press Inc., Totowa, NJ

(LDV-PCR), splinkerette-PCR, thermal asymmetric interlaced PCR (TAIL-PCR), and retroviral LTR arbitrarily primed PCR (RELAP-PCR); and the capture PCR (C-PCR), which can improve the PCR amplification, is also discussed.

## 2. PCR After Intramolecular Circularization

A PCR technique, which is used to amplify an unknown DNA region adjacent to an integrated sequence after its intramolecular circularization, is called IPCR (inverse or inverted or inside-out PCR).

### 2.1. IPCR

The concept of IPCR first came in the 1980s (*1–5*) and has been in use for many years. The principle of this technique is illustrated in detailed in **Fig. 1**. IPCR begins with the digestion of genomic DNA with a restriction enzyme. Intramolecular self-ligation of restriction enzyme digested DNA fragments created a small monomeric circle. Within this circularized form of DNA, a conventional PCR technique is applied to amplify the IHJ region by using two opposite direction primers on the known integrant sequences called integrant specific primers (ISP). Hence, the primers are designed to anneal to the region of known sequence in IPCR. Generally, this technique has been used to characterize fragments up to 5 kb (*6*). However, a new DNA sequence or restriction sites could be used to start a new round of IPCR to obtain additional information. This strategy may be repeated to “walk” both upstream and downstream of a known DNA sequence (*7*). IPCR has been already applied to identify the integration sites of hepatitis B virus (HBV) (*8,9*), human T-lymphotrophic virus type-1 (HTLV-1) (*10–15*), human immunodeficiency virus type-1 (HIV-1) (*16*), and some transposable elements such as IS30 (*17*), T-DNA, *Ds* element (*18*), *dTph1* (*19*), Tn5 (*20*), Tn55 (*21*), and P element (*22*). The insertion of reticuloendotheliosis virus (REV) on Marek’s disease virus (MDV) genome is also identified by this method (*23*).

### 2.2. PI-PCR and Long-Distance IPCR

Some alternative IPCR methodologies have been published in recent years.

Partial inverse PCR (PI-PCR) (*see Fig. 2*) employs the genomic DNA partially digested by using 4-base recognition restriction enzyme (such as *Sau3A1*). After self-ligation, the circular DNA fragments are used as templates for IPCR (*24*). This is based on the preference of PCR for amplifying relatively smaller fragments. A DNA fragment that is less than 1 kb facilitates amplification via self-ligation and the IPCR process (*9*). A wide range of partial digests should be generated to find one that gives an optimal PCR amplification. Moreover, this approach eliminates the need to have any prior knowledge of restriction enzyme sites surrounding the integrant.

Long-distance IPCR method, designated long-range IPCR (LR-iPCR) (*25*) or long inverse PCR (LI-PCR) (*26*), enables the direct amplification of relatively large size flanking from circularized DNA fragments. The central to the long-distance IPCR is the use of a thermostable polymerase. This technique has been adapted to amplify relatively large-size flanking fragments up to 10 kb by using a highly thermal stable polymerase.

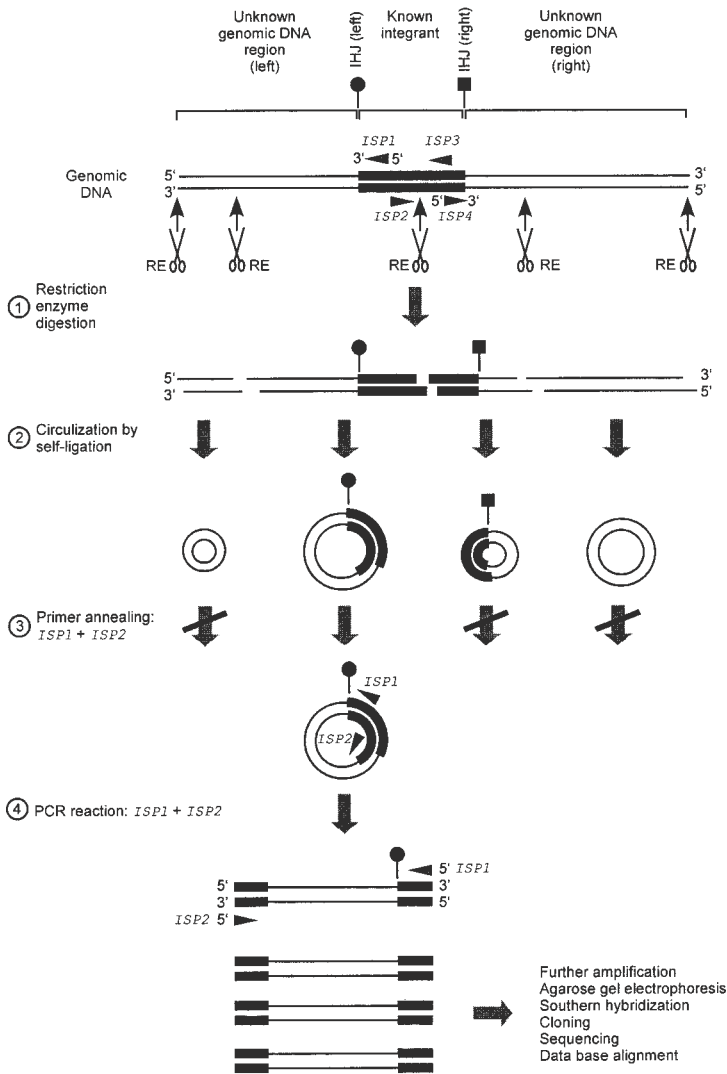


Fig. 1. Schematic flow diagram of IPCR protocol. Two complementary strands of genomic DNA have been shown at the top. The heavy and thin line regions represent the integrant fragment and cellular genomic unknown sequence, respectively. The positions of both left and right IHJs are indicated as closed circle and square signs, respectively. Integrant specific primers *ISP1*, 2, 3, and 4 for I-PCR are shown as arrowheads. In this particular example, the restriction enzyme cutting site (RE, scissors shape) is present in the integrant. *ISP1* and *ISP2* have been applied to amplify the left IHJ. The right IHJ has been amplified by using the other pair of ISP, *ISP3* and *ISP4*, under the same principle. If the RE is not present in the integrant, then any set of primer can be used to amplify both left and right IHJ (71). The slant lines on thick arrows indicate that no primer annealing will occur and no further amplification products. For detail manipulations see (9,17,97).

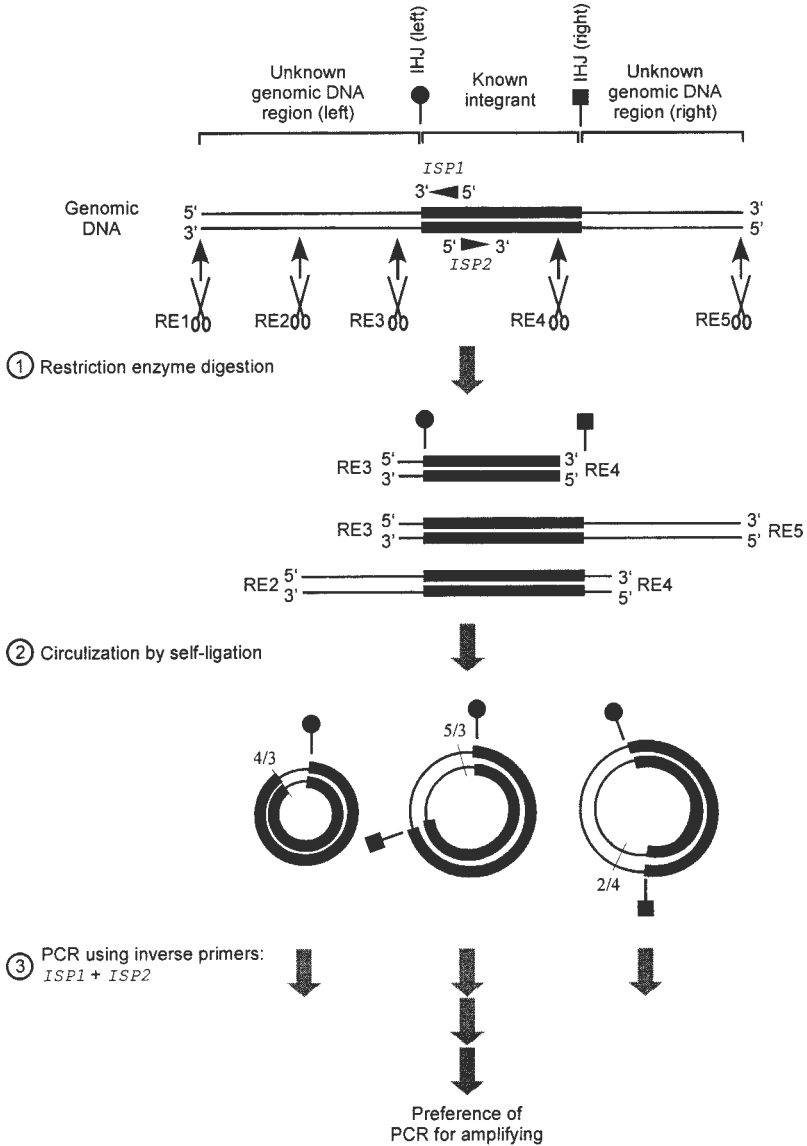


Fig. 2. Description of PI-PCR. Two complementary strands of genomic DNA have been shown at the top. The heavy and thin line regions represent the integrant fragment and cellular genomic unknown sequence, respectively. The positions of both left and right IHJs are indicated as closed circle and square signs, respectively. The different 4-base restriction enzyme recognition sites (RE, scissors shape) have been marked from 1 to 5. The partial digested DNA fragments are self-ligated and can be amplified from a known ISP, *ISP1* and *ISP2*. PCR is preference to amplify smaller fragments even if there are comparable amounts of large and small DNAs. For detail manipulations see (24).

### 2.3. Remarks for IPCR

The major advantage of IPCR is to amplify the flanking unknown sequence by using two known specific primers on the integrant, so-called integrant specific primer (ISP). The intramolecular circularization (self-ligation) of template is a key step for IPCR (*I*). This technique, occasionally, does not produce any product in a particular reaction, presumably because of the ineffective intramolecular ligation. This illustrates that the circularization step in IPCR is a fastidious procedure and not easy to optimize. Wang et al. (*9*) have demonstrated that by choosing the nearest restriction site, which can be determined by conventional PCR, gives a higher successful rate for cloning the IHJ by IPCR. Moreover, the noncircularized, intermolecular ligation, and free viral or transposon DNA fragment may interfere with the PCR. If intermolecular ligation had occurred, multiple PCR products would have been generated. To avoid intermolecular ligation (ligation between two digested fragments), the concentration of DNA has to be decreased and this results in a large volume for ligation. In addition, IPCR has been proved to be less sensitive than the other PCR-based methods (*II*). Therefore, IPCR requires a relatively large size of sample for compensating the low efficiency of ligation.

## 3. IRS-PCR-Based Techniques

The principle of IRS-PCR (interspersed repetitive sequence PCR) is based on the fact that the IRS elements are interspersed in the human genome. In this technique, amplification proceeds with one ISP to the known integrant sequence and the other primer specific to the known cellular interspersed repetitive sequence (IRS), which is distributed among the genome. IRSs are present in a high copy number in most multicellular organisms (*see Table 1*) (reviewed in *ref. 27*). The extension products from these specific primers include a segment containing the region of IHJ.

### 3.1. Novel *Alu*-PCR

Novel *Alu*-PCR (novel *Alu* element-mediated PCR) is an IRS-PCR, which uses a primer to *Alu* element and ISP for the PCR amplification. The overall strategy of novel *Alu*-PCR is outlined in *Fig. 3*. *Alu* elements are the largest family of short interspersed repetitive elements (SINEs) (*see Table 1*). The average density of *Alu* repeats on human genome is at the mean interval of about 3–6 kb, although *Alu* is not uniformly distributed (reviewed in *refs. 28–29*). This technique actually was first applied to amplify human genomes in the background of nonhuman genome and called “*Alu*-PCR” (*30–32*). Hence, extending the applicability of *Alu*-PCR, the inserted foreign sequence can be directly amplified between the known inserted sequence and the *Alu* consensus sequence, and to identify the IHJ (*33*).

Two specific primers are needed in novel *Alu*-PCR: ISP annealing to the known integrant sequence and the other to human *Alu* repeat sequences. In order to avoid illegitimate products, which are amplified from *Alu* sequences itself (*Alu*-*Alu* or inter-*Alu* amplification), two technical skills have been suggested (*33*). First, the primers should be synthesized by deoxyuridine triphosphates (dUTPs). This chemically modified primer can then be destroyed by uracil DNA glycosylase (UDG) after the first

**Table 1**  
**Organization of the Human Genome**

DNA organization	% of total genome	Size of repeat unit	Copy #
Human genome			
I. Mitochondrial genome	0.0005 %		
II. Nuclear genome	99.9995 %		
A. Genes and gene-related sequences	approx 25.0 %		
1. Coding DNA	approx 2.5%		
2. Noncoding	approx 22.5%		
a. Introns, untranslated region, etc.			
b. Pseudogenes			
c. Gene fragments			
B. Extragenic DNA	approx 75.0 %		
1. Unique or low copy no.	approx 45.0 %		
2. Moderate to highly repetitive	approx 30.0 %		
a. Tandemly repeated/clustered repeats	approx 3.7 %		
i. Megasatellite DNA			
ii. Satellite DNA			
iii. Minisatellite DNA			
iv. Microsatellite DNA			
b. IRS approx 26.3 %			
i. SINE class	approx 8.7 %		
- <i>Alu</i> family	approx 7.0 %	Full: 280 bp	approx $0.5-1 \times 10^6$
- MIR families	approx 1.7 %	Average: 130 bp	approx $4 \times 10^5$
ii. LINE class	approx 10.6 %		
- LINE-1 (L1H or <i>Kpn</i> ) family	approx 8.5 %	Full: 6 kb; Average: 0.8 kb	approx $1-5 \times 10^5$
- LINE-2 family	approx 2.1 %	Average: 250 bp	approx $2.7 \times 10^5$
iii. LTR class	approx 4.6 %		
- HERV/RTLV family	approx 1.3%	Average: 1.3 kb	approx $5 \times 10^4$
- THE-1, MER, and other families	approx 3.3%		approx $2 \times 10^5$
iv. DNA transposon	approx 1.6 %		approx $2 \times 10^5$
- <i>mariner</i> family		Average: 250 bp	
- Others			
v. Others	approx 0.8 %		approx $6 \times 10^4$

HERV: human endogenous retroviruses; IRS: interspersed repetitive sequence; LINE: long interspersed nuclear element; LTR: long terminal repeat; MER: medium reiteration frequency; MIR: mammalian-wide interspersed repeat; RTLV: retrovirus-like elements; SINE: short interspersed nuclear element; THE-1: transposable human element.

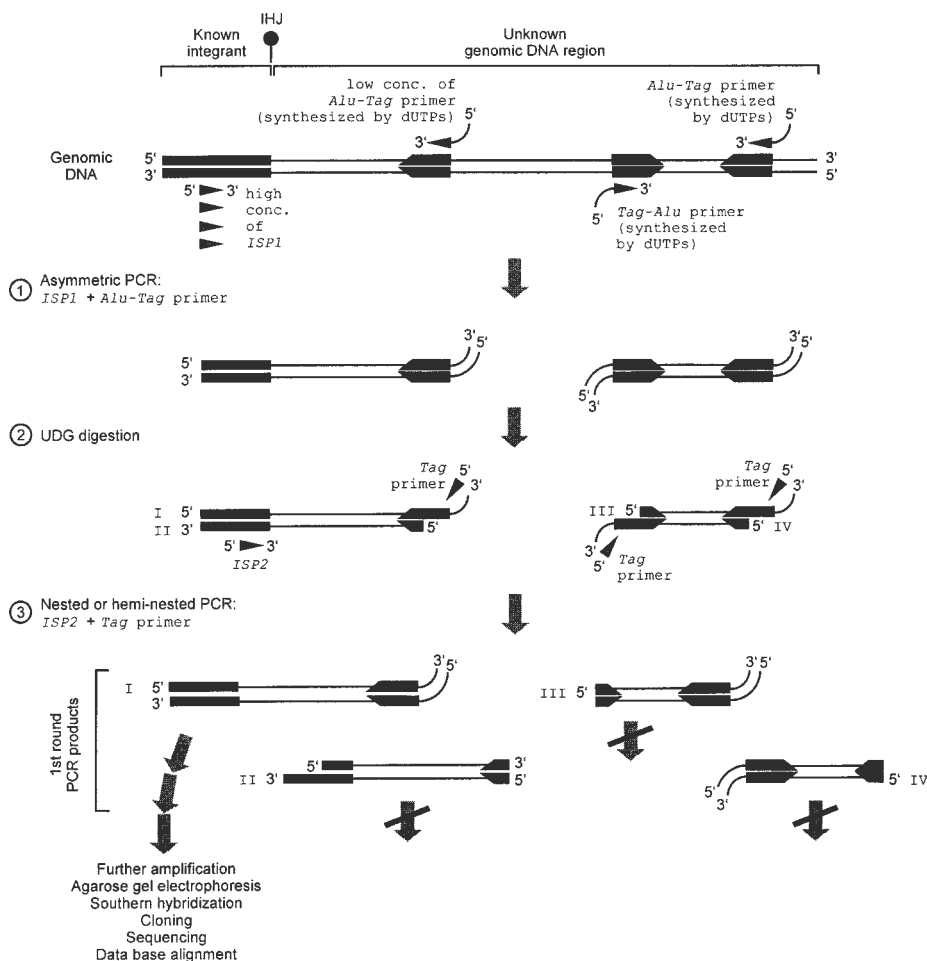


Fig. 3. Schematic diagram of the novel Alu-PCR. Two complementary strands of genomic DNA have been shown at the top. The heavy and thin line regions represent the foreign integrant fragment and cellular genomic unknown sequence, respectively. The position of IHJ is indicated by closed circle. Closed arrow boxes represent Alu elements on the human genome in different orientations. Primers for PCR, *ISP1*, *ISP2*, *Alu-Tag*, and *Tag* primers, are shown as arrowheads. The non-annealing Tag region on *Alu-Tag* primer is shown by curved thin line. Human genomic DNA was amplified by using the first set of primers: *ISP1* and *Alu-Tag* primer (step 1). After an initial 10 cycles of PCR, the *Alu-Tag* primers on new synthesized DNA is destroyed by UDG (step 2). These digest PCR products are further amplified by using an internal primer: *ISP2* and *Tag* primers (step 3; nested PCR). Only DNA product I, which still contains both *ISP2* and *Tag* nested primer target sites, will be amplified further. The cross marks on thick arrows indicate that no primer annealing will occur and no further amplification products. For detail manipulations see (34).

10–15 cycles of amplification. Such modification can break the *Alu-Alu* specific amplification (see **Fig. 3, step 2**). Second, an asymmetric amplification (unequal ratio of two primers) is performed before UDG treatment (see **Fig. 3, step 1**). The primer on the known integrant sequence is added at least 10-fold higher concentration than the primer for the *Alu* sequence (34). In general, during the first 10–20 cycles, dsDNA products are generated. But when the limiting primer is exhausted, ssDNA is produced for the next cycles by primer extension (35,36). No matter the accumulation of dsDNA and ssDNA, the products of integrant-*Alu* amplification are higher than the *Alu-Alu* products, and thus this asymmetric PCR does not favor *Alu-Alu* amplification.

In addition, the design of the primer contains a tag sequence, which can be applied to the other standard PCR protocols such as the nested or hemi-nested PCR (see **Fig. 3, step 3**), to decrease the nonspecific amplification of PCR. Moreover, a single primer control can exclude the false-positive amplification and Southern hybridization has been suggested to facilitate cloning. Some investigators, by using novel *Alu*-PCR, have successfully identified cellular sequences flanking from integrated HBV (34), HIV-1 (37,38), and human papillomavirus type-16 (HPV-16) (39) DNA. The adeno-associated virus (AAV) vector insertion for gene therapy, in addition, was also detected by this way (40).

### 3.2. B1-PCR and LINE-PCR

The *Alu* repeat is primate-specific but other mammals have similar types of sequence such as the B1 family in mouse. Thus, the novel *Alu*-PCR equivalent, B1-PCR has been applied to find the AAV vector integration site from the rat tissues (40).

Based on the same principles, the *Alu* primer can be replaced by the others primers, which can anneal to the other genomic repetitive sequences. In addition to the SINE (such as *Alu* sequence) applied in novel-*Alu* PCR, other IRS has also been used under the same approach such as long interspersed nuclear element (LINE), and so-called LINE-PCR. LINE-PCR has been applied to identify the HPV integration site (39).

### 3.3. Remarks for IRS-PCR

The IRS-PCR offers at least four advantages over IPCR. First, less amount of DNA is required. Second, in contrast to IPCR, an intramolecular ligation reaction is not required in the IRS-PCR. This can overcome the low efficiency of the self-ligation reaction. Third, IRS-PCR is based on only two steps: UDG digestion and conventional PCR procedures, thereby saving most time. Fourth, IRS-PCR avoids the attendant problems in interpretation resulting from episomal contamination.

However, this technique is not suitable for the case of IRS elements within a short distance to the integrant. Unfortunately, many virus genomes tend to insert adjacent to or into repetitive sequence, such as HBV (41–46), simian virus 40 (SV40) (47), murine leukemia virus (MuLV) (48), hamster endogenous retrovirus (49), HIV-1 (16,50–51), HPV-16 (52–54), woodchuck hepatitis virus (WHV) (55), and duck hepatitis B virus (DHBV) (56). Furthermore, the use of IRS-PCR is also limited by the requirement for the adjacent repeat sequences to be in the correct orientation.

## 4. LA-PCR-Based Techniques

LA-PCR (ligation-anchored PCR or cassette ligation-anchored PCR) is based on the ligation of an oligonucleotide cassette unit, called adapter or linker, to the cleaved



genomic fragments (57). The amplification proceeds with ISP to the known integrant sequence and the other primer specific to the known ligated adapter. Indeed, the principle of single-specific-primer PCR (SSP-PCR) (58), rapid amplification of cDNA ends (RACEs) (reviewed in *ref.* 59), and rapid amplification of genomic DNA ends (RAGE) (60) have been developed using the similar concept. In these PCR methods, the ligated unit enables PCR to amplify the DNA fragment between itself and a known primer from known integrant sequence.

#### 4.1. LM-PCR

In ligation mediated-PCR (LM-PCR), genomic DNA is digested by a restriction enzyme and ligated with a primer using T4 DNA ligase. Then, ISP and the ligated primer are used in a classical PCR amplification. This protocol has been applied to the samples from HTLV-I integration genome (11,61–62).

#### 4.2. Vectorette-PCR

The unique feature of vectorette-PCR method is the special secondary structure of the cassette, which termed vectorette unit (*see Fig. 4A*). The vectorette unit contains a central non-complementarity mismatched region resulting in a bubble-shape (*see Fig. 4A*), therefore, the vectorette-PCR is also termed “bubble PCR.” VectorettePCR was first used for rapid isolation of terminal sequences from yeast artificial chromosome (YAC) clones (63), and then applied to the intronic DNA sequence characterization (64). The procedure of vectorette-PCR begins with the digestion of genomic DNA with a restriction enzyme to generate a 5'-overhang, and then ligation with a vectorette unit. The flanking sequences are then amplified by using an ISP of the integrant and the universal vectorette specific primer. The amplification strategy is summarized in *Fig. 4B*. The vectorette primer, which applied to the PCR is, actually, identity sequence to, but not complementary to, the noncomplementarity mismatched region of vectorette unit and therefore it only process PCR extension from the second round of the reaction. This enhances the PCR amplification specific to the IHJ containing genomic fragment. The HIV integration site has been identified by this method (16).

#### 4.3. MTV- and LDV-PCRs

Vectorette-PCR, in addition, has been modified to be a multistep-touchdown vectorette-PCR (MTV-PCR) (65), which is suitable for analysis of the high CG content region. MTV-PCR starts at a hot-start technique and proceeds at touchdown PCR cycle profile. Because the high GC content DNA results in the PCR conflicting secondary structures, the application of touchdown cycling parameters prevent significantly the formation of unspecific DNA fragments.

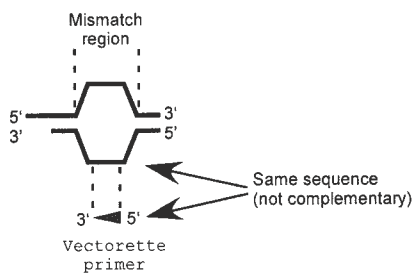
A vectorette-based long-distance PCR has been developed to amplify the fragment up to 5 kb (66), and so-called long-distance vectorette-PCR (LDV-PCR). The use of a mixture of thermostable DNA polymerases is central to this approach.

#### 4.4. Splinkerette-PCR

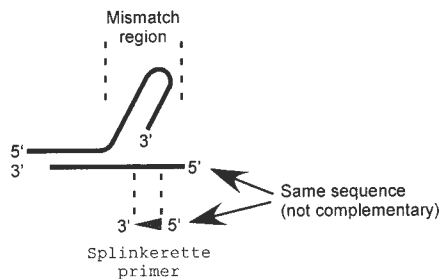
However, undesirable amplifications of nonspecific “end-repair priming” may involve the free cohesive ends of unligated free vectorettes and 5'-overhangs of unknown cellular region (*see Fig. 4C*). These ends are filled during the first cycle of PCR. After the denaturing step, these ends are able to anneal together (as shown in step 4 of

**A**

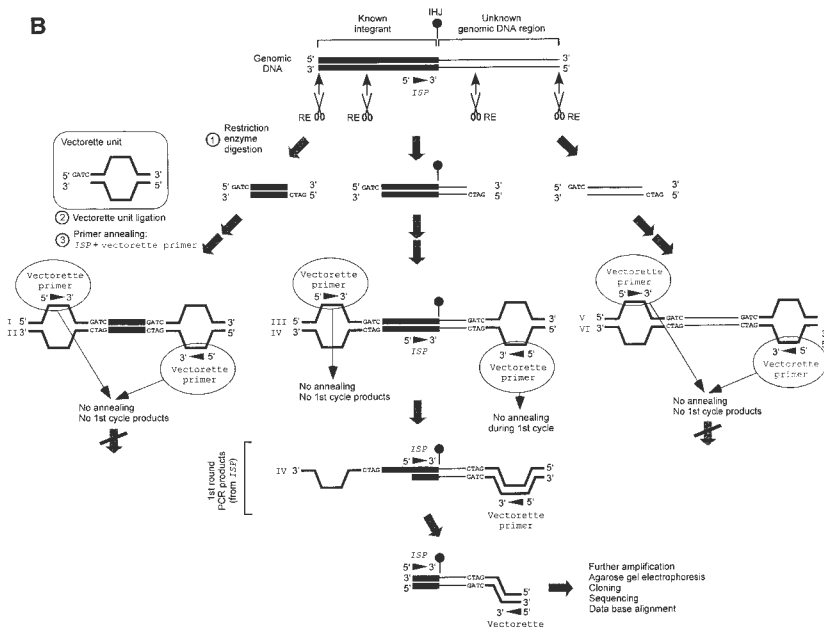
Vectorette unit



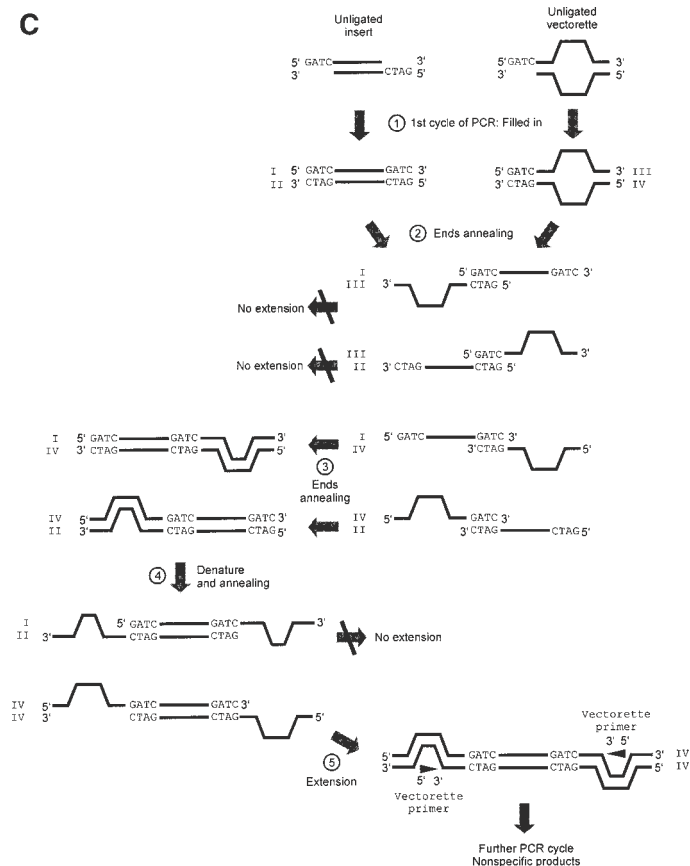
Splinkerette unit



B



**C**



**D**

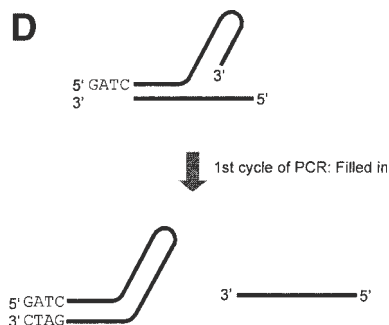


Fig. 4. (A, facing page) Structure of the vectorette and splinkerette units. The primer for vectorette and splinkerette are also shown. (B, facing page) Schematic representation of the principle of vectorette-PCR. Two complementary strands of DNA are shown at the top. The heavy and thin line regions represent the foreign integrant fragment and cellular genomic unknown sequence, respectively. The position of IHJ is indicated by closed circle. *ISP* and vectorette primers for PCR are shown as arrowheads. The slant lines on thick arrows indicate that no primer annealing will occur and no further amplification products. For detail manipulations see (63). (C) A diagram showing the effect of “end-repair priming” in vectorette-PCR. (D) The splinkerette unit do not has “end-repair priming” effect.

**Fig. 4C**). In this procedure, the complementary strand of vectorette primer is generated from unwanted fragments, and this decrease the specificity of vectorettes-PCR. The splinkerette is therefore designed as a hairpin structure on one strand rather than a central DNA mismatch (67), as compared in **Fig. 4A**. The advantage of splinkerette-PCR over vectorette-PCR is the elimination of the end-repair priming phenomena (see **Fig. 4D**). Some researchers have successfully identified flanking regions of transposon *Sleeping Beauty* (*Tc1/mariner* superfamily) by using this method (68).

#### 4.5. Remarks for LA-PCR

LM-PCR has been proved to be more sensitive than IPCR in the detection of integrant (11). Some commercial products (Invitrogen and TaKaRa), which use PCR technology to quickly identify the unknown sequence, are also based on the principle for LA-PCR. However, it requires a proper ligation between oligomer linker to genomic DNA fragments. The ratio of linker DNA and genomic DNA has to be serially diluted to obtain a maximum intermolecular ligation (58).

### 5. AP-PCR-Based Techniques

The principle of AP-PCR (arbitrarily primed PCR) is using nonspecific arbitrary primers for PCR amplification (69). Following this “hemispecific” concept, the targeted gene walking PCR (70; reviewed in 71), single primer reaction (72), differential display PCR (DD-PCR) (73; reviewed in 74,75), and restriction site PCR (RS-PCR) (76; reviewed in 71) have been developed for the amplification of DNA sequences by nonspecific arbitrary primers.

#### 5.1. TAIL-PCR

TAIL-PCR utilizes three nested specific primers on known integrant in successive three rounds of PCRs together with a shorter arbitrary degenerate (AD) primer. The basis for this strategy is thermal asymmetric PCR. Arbitrary priming creates nontarget molecules, because degenerate primers that hybridize randomly in genomic DNA and constitute the bulk of the final unwanted products. The interspersing asymmetric and symmetric PCR cycles are used geometrically to favor amplification of target molecules over nonspecific products.

A schematic diagram of targeted TAIL-PCR is shown in **Fig. 5**. During the high-stringency cycle at first round of PCR (high-stringency PCR program at step 1 in **Fig. 5**) only the long integrant specific primer *ISR1* can efficiently anneal to the DNA template, therefore only specific product (product I in **Fig. 5**) is amplified and little or no nontarget sequence product (which is primed at both ends by *AD* primers; product II in **Fig. 5**) has been formed. In the following single reduced-stringency cycle (low-stringency PCR program at **step 1** in **Fig. 5**), however, both *ISR1* and *AD* primers can anneal to the template DNA. The single-stranded target DNA, which is produced during last high-stringency cycles is replicated to dsDNA and hence providing a several-fold increase of target template for the next round of amplification. In following TAIL-cycling (TAIL programme at **step 1** in **Fig. 5**), the specific product (product I) is

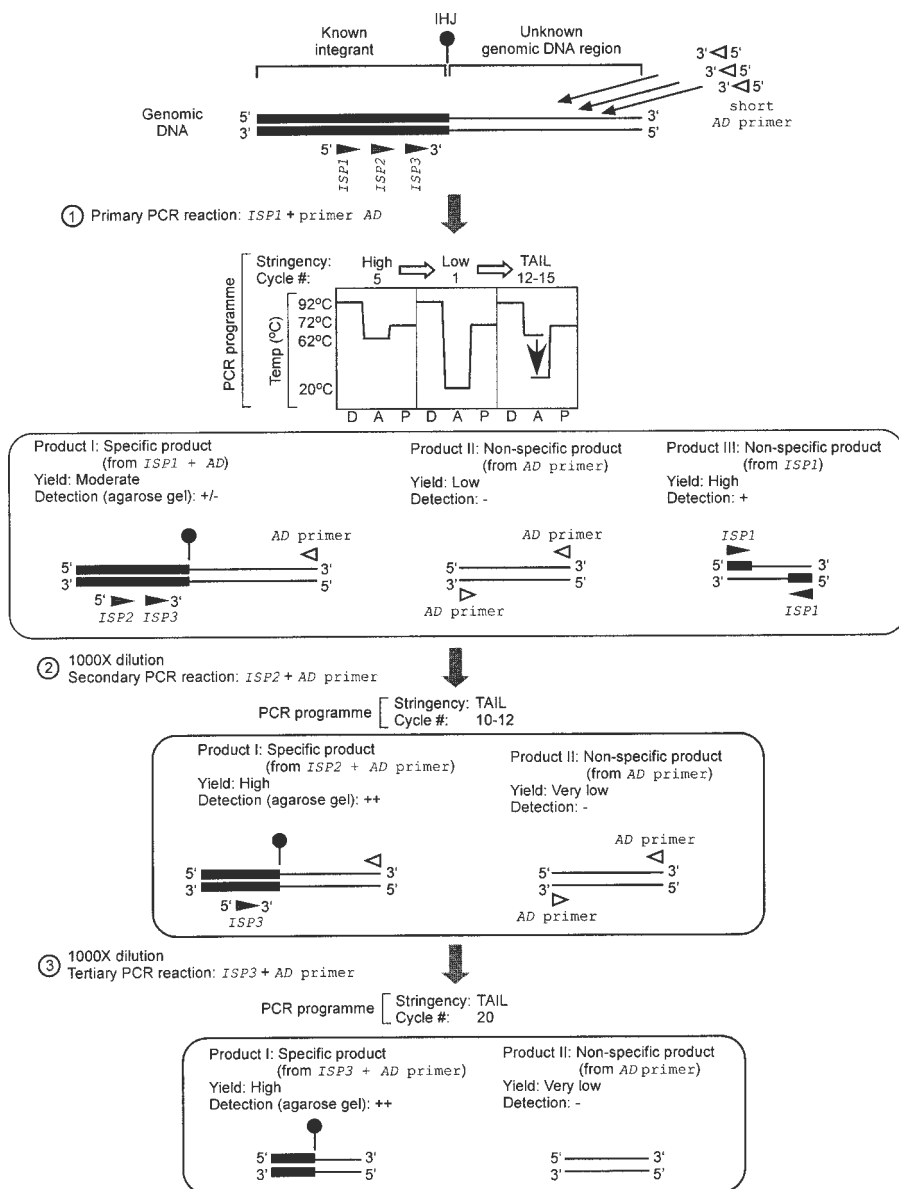


Fig. 5. Summary of the TAIL-PCR procedure. Two complementary strands of DNA have been shown at the top. The heavy and thin line region represent the integrant fragment and cellular genomic unknown sequence, respectively. The position of IHJ is indicated by closed circle. ISP and AD (arbitrary degenerate) primers for PCR are shown as close and open arrow-heads, respectively. The PCR program with different stringencies and cycle number is shown on the top of the box. D, A, and P represents denature, annealing, and polymerization step, respectively, in PCR cycle. For detail manipulation see (84).

possible to be amplified preferentially over nontarget sequence product. But at the same time, nonspecific product (at both ends by the long *ISP1* primers; product III in **Fig. 5**) can also arise efficiently through mispriming. Such undesired products are diluted out, however, in subsequent secondary (**step 2**) and tertiary (**step 3**) round PCR, which using internally nested specific primers *ISP2* and *ISP3*, respectively. The TAIL cycling in both secondary and tertiary is performed in lower background further. In fact, the T-DNA insertion (*77,78*), Ds elements (*79,80*), and *Tto1* introduced (*81*) in *Arabidopsis* have been identified by this technique.

## 5.2. RELAP-PCR

The combination of a long ISP designed to detect retroviral long terminal repeat (LTR) and a short arbitrary primer (AD primer) from the AD primer set in different lengths binding in a random fashion under a low-stringency condition are used (*see Fig. 6*, upper panel). Therefore, AP-PCR had been adapted to allow the amplification of LTR-containing retrovirus integration site. This is called RELAP-PCR (retroviral LTR-arbitrarily primed PCR) and has been applied to identify the integration site of mouse mammary tumor virus (MMTV) (*82*).

Hot spot-combined PCR (HS-cPCR), modified AP-PCR for retrovirus, is based on previous finding regarding the host spot of retroviral LTR integration sites. In this technique, the primers have been designed to target on both known retroviral LTR and nonintegrant region in different combination (*83*). It is possible to design primers, which border the “suspected” fragment.

## 5.3. Remarks for AP-PCR

These AP-PCR methods allow rapid detection without any DNA manipulation before PCR, such as restriction enzyme digestion or ligation. Amplification occurs either upstream or downstream from a known sequence. In TAIL-PCR, a set of nested long primer and a short arbitrary primer are important (*77,84*). Besides the primer design, the stringency in the primer-template interaction is an important parameter of this class of PCR (*85*). Specificity of the amplification reaction has been further confirmed by Southern blotting of the PCR products. The single primer control is always necessary and important to exclude the false positive results.

In the case of RELAP-PCR, a series of walking primers have been designed to increase the incidence of positive results. In addition, a series of walking reaction are usually done in parallel. This can be laborious and time-consuming. Moreover, this strategy is only suitable for the LTR-containing integrant.

## 6. RT-PCR-Based Technique

In the case of retrovirus, the promoter activation within 3' untranslated LTR initiates chimeric mRNA transcripts, which consist the viral LTR and the cellular gene fragment in the same transcriptional orientation (*86,87*). Based on this property of retrovirus, the poly(A)-tail containing mRNA is purified and cDNA is synthesized by reverse transcription using an oligo(dT)-adaptor primer which primers on the poly(A)

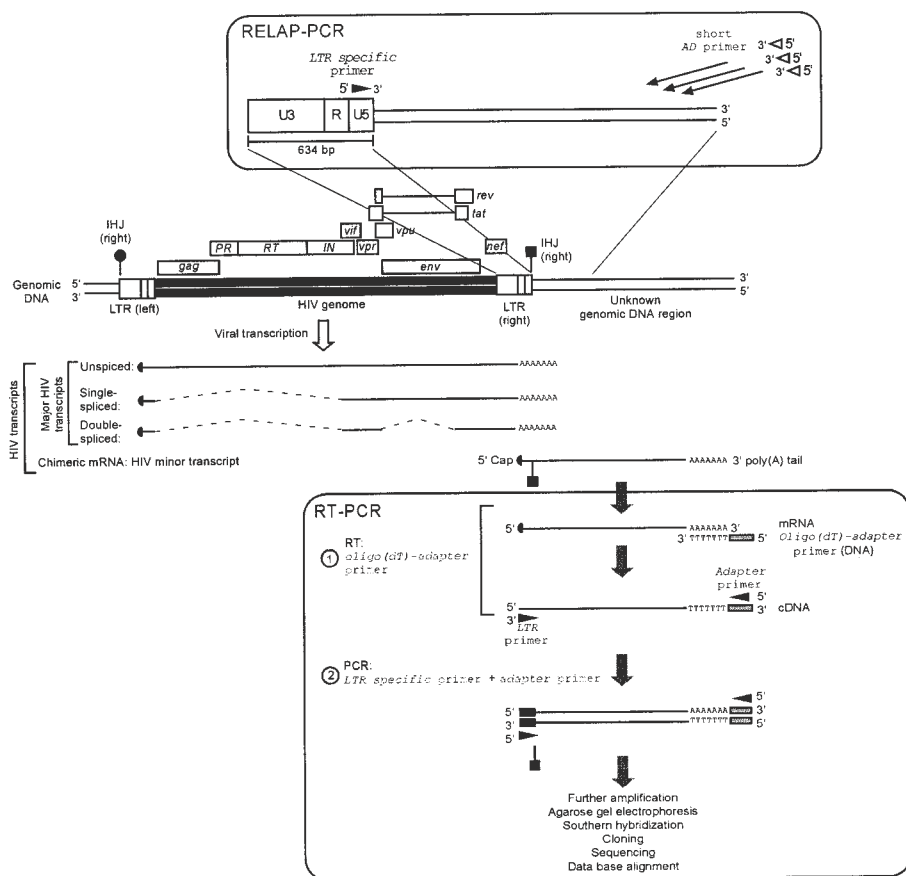
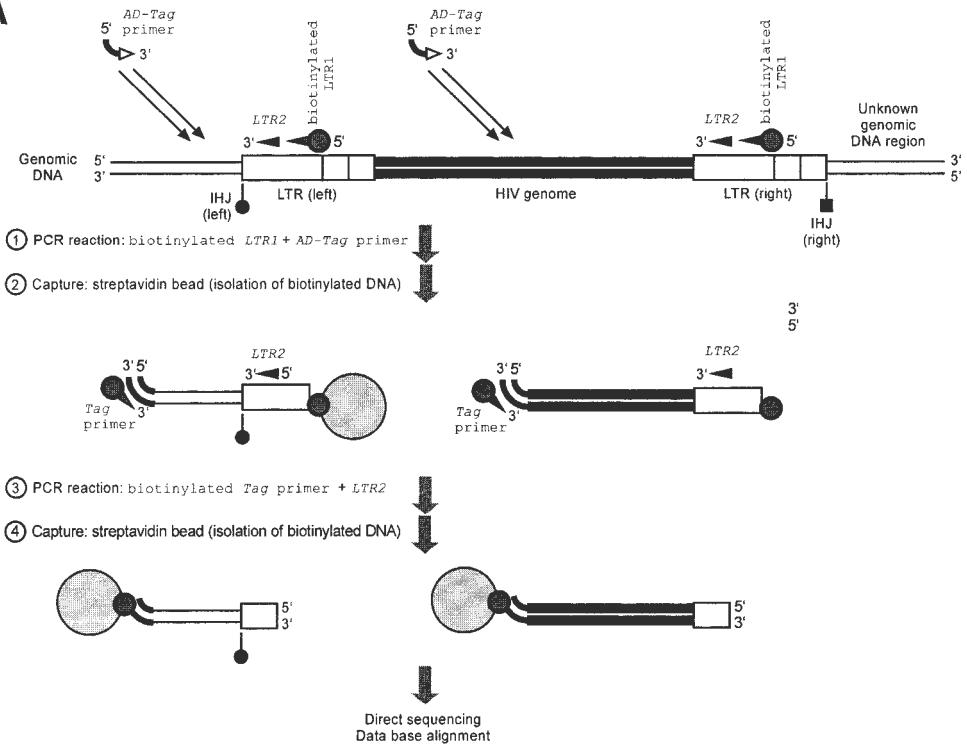


Fig. 6. Summary of the procedure for RELAP- and RT-PCR in the studying on retrovirus insertion. Two complementary strands of HIV DNA have been shown at the middle. The heavy and thin line region represent the integrated HIV fragment and cellular genomic unknown sequence, respectively. The position of left and right IHJ is indicated by closed circle and square, respectively. Boxes with HIV open reading frames (ORFs) are shown. The right and left LTR region of HIV are shown as an open box (the size is not to scale). Primers for PCR are shown as arrowheads. Upper panel is the schematic diagram for RELAY-PCR. For detail manipulation for *see* (82). Lower panel is the schematic flow diagram for RT-PCR based protocol. The major and minor viral transcripts from HIV viral transcription are shown. Only the cDNA from chimeric mRNA contains LTR primer target site. For detail manipulation *see* (90).

tails. Using the adapter primer and an LTR-specific primer, the chimeric mRNA containing the retroviral insertion site is amplified by PCR. The overall RT-PCR based method to isolate these chimeric cDNAs is schematically shown in the lower panel in **Fig. 6**. This principle is similar as anchored PCR (A-PCR) (88), one-sided PCR (89), and RACE. This RT-PCR based method is rapid and simple, and successful to identify the retrovirus integration site (90). However, this method only works on the virus,

**A**

which contains a cis-acting promoter activity sequence (like LTR) and synthesizes the chimeric mRNA.

## 7. Capture PCR Improvement

The capture PCR (C-PCR) is an alternative protocol to enrich the interested DNA fragments by a streptavidin-coated support for the PCR (91). Indeed, both AP- and LM-PCR have been improved by this protocol.

Under the concept of AP-PCR, the biotinylated integrant specific primer and a partly degenerate arbitrary primer are applied for the PCR. The amplified DNA fragment is then isolated by streptavidin-coated magnetic beads (92). The application of this approach into AP-PCR is shown in Fig. 7A. This method has been used for the isolation of the integrated retroviral provirus (92). Under the concept of LM-PCR, after initial ligation of oligonucleotide adapter to all restriction ends, the biotinylated specific primers to known sequence are used for an extension reaction. These biotin-labeled extension products are immobilized on streptavidin-coated beads and then used as templates in a PCR. This technique is also called amplification of insertion mutagenised sites (AIMS) (93). The application of this approach into LM-PCR is shown in Fig. 7B. This method has been applied to the detection of *Bx1* gene in maize by transposon tagging *Mutator* (93,94). This improvement of alternative AP-PCR is





cellular genes that are involved in cell proliferation and differentiation (95,96). However, genomic cloning method requires the establishment of genomic DNA libraries, which is time-consuming and laborious. Therefore, many PCR-based techniques have been developed for the elucidation of unknown flanking DNA sequence adjacent to a region of known integrant sequence. The genomic sequences flanking foreign integrant can then be determined rapidly with these techniques (within 1 wk). PCR-based techniques offer an inexpensive and flexible alternative to IHJ searching, and can be performed in any laboratory equipped with basic molecular biology. The limitation of PCR is the need for the sequence of two target specific primers that flank the region that is intended for amplification. The problem here is how to allow the direct amplification of DNA without a prior knowledge of sequence information. Several strategies have resolved this limitation. IPCR can amplify flanking region directed away from the core region of known integrant sequence after DNA self-ligation (circularization). IRS-PCR depends on the distribution of IRS on the genome. Many techniques for amplifying flanking unknown regions of DNA are based on the creation of new primer binding sites on the potential PCR template by ligating oligonucleotide linkers or cassette unit of known sequences to the ends of DNA fragments, such as LA-PCR. And some other techniques allow primer binding in a random fashion under low-stringency condition, such as AP-PCR. All methods described in this chapter are compared in **Table 2**.

Beside the IHJ searching, these techniques can also be applied to the determination for YAC end points (31,63), cDNA ends (97), genomic breakpoints (deletion or translocation) (66,98), intron-exon junctions (99), gene rearrangements (6), promoter sequence sequences (100,101), mating-type gene switching (102), and gene-targeting vector construction (103). Although, some PCR-based techniques such as RAGE (104), RS-PCR (76; reviewed in 71), panhandle PCR (105,106; reviewed in 107), multiplex RS-PCR (108), gene walking PCR (109), homo-oligomeric tailing based PCR (110), novel step-down PCR (111), and nonspecifically primed suppression PCR (NSPS-PCR) (112) have not yet been used to IHJ study, these principles are also applicable in the integration site searching.

Among all techniques introduced in this chapter, the IRS-PCR- and TAIL-PCR-based methods are highly recommended to apply in the integrant seeking. One reason is the high sensitivity of these two methods. Second, genomic DNA manipulation, such as DNA digestion and ligation reaction, is not required. Third, only the simple straightforward technique of conventional PCR protocol is needed. In fact, each of the approaches to identify IHJ is useful, and the experimental context is the critical feature that determines success. If an experiment is poorly designed (especially the primer set sequence) or the sample is contaminated, the result is a large number of bands after PCR, that are difficult and time-consuming to analyze. Any inefficiencies, mispriming, or incomplete reaction in the PCR or restriction enzyme digestion steps can result in artifacts that are misleading. Some improvements, therefore, are applied in the PCR reaction to increase the specificity of PCR amplification. These are hot start PCR (in any PCR techniques), nested PCR (in IRS-PCR, TAIL-PCR), touch-

**Table 2**  
**Methods Described in this Chapter**

Principle Base on	Methods	DNA digestion	DNA ligation	Amount of DNA used	Primers on	Step to optimize	Source of nonspecific products	Sensitivity
Inverse	IPCR ( <b>Fig. 1</b> ) PI-PCR ( <b>Fig. 2</b> ) LR-iPCR LI-PCR	+	+(self-ligation)	D: 0.2–10 µg L: 0.2–5 ng/µL	Integrand	1. Relative high DNA sample amount 2. DNA concentration of self-ligation	1. Unintegrated DNA 2. Unligated DNA	Low
IRS-PCR	Novel <i>Alu</i> -PCR ( <b>Fig. 3</b> ) LINE-PCR B1-PCR	–	–	P: 10–106 ng	1. Integrand 2. IRS (eg. <i>Alu</i> , LINE, B1)	1. IRS orientation 2. Distance to IRS	Inter-IRS amplification	High
LA-PCR	LM-PCR Vectorette-PCR ( <b>Fig. 4B</b> ) MTV-PCR LDV-PCR Splinkerette-PCR Capture ( <b>Fig. 7B</b> )	+	+	D: 0.5–2 µg L: 0.5–2 µg	1. Integrand 2. Adapter unit	1. Ligation efficiency 2. Adapter unit design ( <b>Fig. 4C</b> )	End-repair priming	Medium
AP-PCR	TAIL-PCR ( <b>Fig. 5</b> ) RELAP-PCR ( <b>Fig. 6</b> ) Capture ( <b>Fig. 7A</b> )	–	–	P: 20–150 ng	1. Integrand 2. Arbitrary primer	Primer set design	Arbitrary priming	High
RT-PCR	RT-PCR ( <b>Fig. 6</b> )	–	+	R: 3 µg poly(A) <sup>+</sup> RNA	1. LTR (integrand) 2. Adapter	cDNA synthesis		Medium

D: DNA digestion reaction; L: ligation reaction; P: PCR reaction; R: RT reaction.

down PCR (in IRS-PCR, MTV-PCR, and TAIL-PCR), specific primer cassette structure (in vectorette- and splinkerette-PCR), asymmetric (or unequal) ratio of the two amplification primers is used (in IRS-PCR and TAIL-PCR), primer synthesis by dUTP and then digest by UDG (in IRS-PCR), isolated biotinylated products (in C-PCR), and -NH<sub>2</sub> and -PO<sub>4</sub> groups modification on the adaptors to prevent nonspecific 3' end elongation during PCR reaction (in novel step down PCR). These modifications are pivotal to successful amplification. It might also be helpful to optimize the PCR conditions with respect to magnesium and dimethyl-sulfoxide (DMSO) concentrations according to standard protocols. Although false positive results could happen, both single primer control and Southern blotting would minimize this problem and confirm the specificity of the amplified products.

## References

1. Collins, F. S. and Weissman, S. M. (1984) Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. *Proc. Natl. Acad. Sci. USA* **81**, 6812–6816.
2. Ochman H., Gerber, A. S., and Hartl, D. L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623.
3. Ochman, H., Ajioka, J. W., Garza, D., and Hartl, D. L. (1989) Amplification of flanking sequences by inverse PCR, in *PCR Technology* Ehrlich, H. A., ed., Stockton, New York, pp. 105–111.
4. Triglia, T., Peterson, M. G., and Kemp, D. J. (1988) A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequence. *Nucl. Acid Res.* **16**, 8186.
5. Silver, J. and Keerikatte, V. (1989) Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. *J. Virol.* **63**, 1924–1928.
6. Willis, T. G., Jadayel, D. M., Coignet, L. J. A., Abdul-Rauf, M., Treleaven, T. J., Catorsky, D., and Byer, M. J. S. (1997) Rapid molecular cloning of rearrangements of the IGHJ locus using long-distance inverse polymerase chain reaction. *Blood* **90**, 2456–2462.
7. Garces, J. A. and Gavin, R. H. (2001) Using an inverse PCR strategy to clone large contiguous genomic DNA fragments. *Meth. Mol. Biol.* **161**, 3–8.
8. Tsuei, D.-J., Chen, P.-J., Lai, M.-Y., Chen, D.S., Yang, C.-S., Chen, J.-Y., and Hsu, T.-Y. (1994) Inverse polymerase chain reaction for cloning cellular sequences adjacent to integrated hepatitis B virus DNA in hepatocellular carcinomas. *J. Virol. Meth.* **49**, 269–284.
9. Wang, P.-C., Hui, E. K.-W., Chiu, J.-H., and Lo, S. J. (2001) Analysis of integrated hepatitis B virus DNA and flanking cellular sequence by inverse polymerase chain reaction. *J. Virol. Meth.* **92**, 83–90.
10. Takemoto, S., Matsuoka, M., Yamaguchi, K., and Takatsuki, K. (1994) A novel diagnostic method of adult T-cell leukemia: monoclonal integration of human T-cell lymphotropic virus type I provirus DNA detected by inverse polymerase chain reaction. *Blood* **84**, 3080–3085.
11. Cavois, M., Wain-Hobson, S., and Wattel, E. (1995) Stochastic events in the amplification of HTLV-I integration sites by linker-mediated PCR. *Res. Virol.* **146**, 179–184.
12. Cavois, M., Wain-Hobson, S., Gessain, A., Plumelle, Y., and Wattel, E. (1996) Adult T-cell leukemia/lymphoma on a background of clonally expanding HTLV-1 positive cells. *Blood* **88**, 4646–4650.
13. Ohshima, K., Suzumiya, J., Kato, A., Tashiro, K., and Kikuchi, M. (1997) Clonal HTLV-I-infected CD4<sup>+</sup> T-lymphocytes and non-clonal non-HTLV-infected giant cells in incipient ATLL with Hodgkin-like histologic features. *Int. J. Cancer* **72**, 592–598.

14. Ohshima, K., Mukai, Y., Shiraki, H., Suzumiya, J., Tashiro, K., and Kikuchi, M. (1997) Clonal integration and expression of human T-cell lymphotropic virus type I in carriers detected by polymerase chain reaction and inverse PCR. *Am. J. Hematol.* **54**, 306–312.
15. Leclercq, I., Cavois, M., Mortreux, F., Hermine, O., and Gessain, A. (1998) Oligoclonal proliferation of human T-cell leukemia virus type 1 bearing T cells in adult T-cell leukemia/lymphoma without deletion of the 3' provirus integration sites. *J. Haematol.* **101**, 500–506.
16. Carteau, S., Hoffmann, C., and Bushman, F. (1998) Chromosome structure and human immunodeficiency virus type 1 cDNA integration: centromeric alphoid repeats are a disfavored target. *J. Virol.* **72**, 4005–4014.
17. Ochman, H., Ayala, F. J., and Hartl, D. L. (1993) Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. *Meth. Enzymol.* **218**, 309–321.
18. Knapp, S., Larondelle, Y., Robberg, M., Furtek, D., and Theres, K. (1994) Transgenic tomato lines containing Ds elements at defined genomic positions as tools for targeted transposon tagging. *Mol. Gen. Genet.* **243**, 666–673.
19. Souer, E., Quattrocchio, F., de Vetten, N., Mol, J., and Koes, R. (1995) A general method to isolate genes tagged by a high copy number transposable element. *Plant J.* **7**, 677–685.
20. Martin, V. J. and Mohn, W. W. (1999) An alternative inverse PCR (IPCR) method to amplify DNA sequence flanking Tn5 transposon insertions. *J. Microbiol. Meth.* **35**, 163–166.
21. De Lencastre, H., Wu, S. W., Pinho, M. G., Ludovice, A. M., Filipe, S., Gardete, S., et al. (1999) Antibiotic resistance as a stress response: complete sequencing of a large number of chromosomal loci in *Staphylococcus aureus* strain COL that impact on the expression of resistance to methicillin. *Microbiol. Drug Res.* **5**, 163–175.
22. Liao, G.-C., Rehm, E. J., and Rubin, G. M. (2000) Insertion site preferences of the P transposable element in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **97**, 3347–3351.
23. Isfort, R., Jones, D., Kost, R., Witter, R., and Kung, H.-J. (1992) Retrovirus insertion into herpesvirus in vitro and in vivo. *Proc. Natl. Acad. Sci. USA* **89**, 991–995.
24. Pang, K. M. and Knecht, D. A. (1997) Partial inverse PCR: a technique for cloning flanking sequences. *BioTechniques* **22**, 1046–1048.
25. Mathur, J., Szabados, L., Schaefer, S., Grunenberg, B., Lossow, A., Jonas-Straube, E., et al. (1998) Gene identification with sequenced T-DNA tags generated by transformation of *Arabidopsis* cell suspension. *Plant J.* **13**, 707–716.
26. Raponi, M., Dawes, I. W., and Arndt, G. M. (2000) Characterization of flanking sequences using long inverse PCR. *BioTechniques* **28**, 839–843.
27. Smit, A. F. (1996) The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6**, 743–748.
28. Mighell, A. J., Markham, A. F., and Robinson, P. A. (1997) Alu sequence. *FEBS Lett.* **417**, 1–5.
29. Szmulewicz, M. N., Novick, G. E., and Herrera, R. J. (1998) Effects of Alu insertions on gene function. *Electrophoresis* **19**, 1260–1264.
30. Nelson, D. L., Ledbetter, S. A., Corbo, L., Victoria, M. F., Ramirez-Solis, R., Webster, T. D., et al. (1989) Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA* **86**, 6686–6690.
31. Nelson, D. L., Ballabio, A., Victoria, M. F., Pieretti, M., Bies, R. D., Gibbs, R. A., et al. (1991) Alu-primer polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: identification of clones associated with a disease locus. *Proc. Natl. Acad. Sci. USA* **88**, 6157–6161.
32. Ledbetter, S. A., Nelson, D. L., Warren, S. T., and Ledbetter, D. H. (1990) Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* **6**, 475–481.

33. Puskas, L. G., Fartmann B., and Bottka S. (1994) Restricted PCR: amplification of an individual sequence flanked by a highly repetitive element from total human DNA. *Nucl. Acids Res.* **22**, 3251–3252.
34. Minami M., Poussin K., Brechot C., and Paterlini P. (1995) A novel PCR technique using Alu-specific primers to identify unknown flanking sequences from the human genome. *Genomics* **29**, 403–408.
35. Gyllensten, U. B. and Erlich, H. A. (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. USA* **85**, 7652–7656.
36. Gyllensten, U. B. and Allen, M. (1993) Sequencing of in vitro amplified DNA. *Meth. Enzymol.* **218**, 1–16.
37. Courcoul, M., Patience, C., Rey F., Blanc, D., Harmache, A., Sire, J., et al. (1995) Peripheral blood mononuclear cells produce normal amounts of detectable Vif- human immunodeficiency virus type 1 particles which are restricted for the preroetrotranscription steps. *J. Virol.* **69**, 2068–2074.
38. Sonza, S., Maerz, A., Deacon, N., Meanger, J., Mills, J., and Crowe, S. (1996) Human immunodeficiency virus type 1 replication is blocked prior to reverse transcription and integration in freshly isolated peripheral blood monocytes. *J. Virol.* **70**, 3863–3869.
39. Carmody, M. W., Jones, M., Tarraza H., and Vary, C.P. (1996) Use of the polymerase chain reaction to specifically amplify integrated HPV-16 DNA by virtue of its linkage to interspersed repetitive DNA. *Mol. Cell. Probes* **10**, 107–116.
40. Wu, P., Phillips, M.I., Bui, J., and Terwilliger, E.F. (1998) Adeno-associated virus vector-mediated transgene integration into neurons and other nondividing cell targets. *J. Virol.* **72**, 5919–5926.
41. Choo, K. B., Liu, M. S., Chang, P. C., Wu, S. M., Su, M. W., Pan C. C., and Han, S. H. (1986) Analysis of six distinct integrated hepatitis B virus sequences cloned from the cellular DNA of a human hepatocellular carcinoma. *Virology* **154**, 405–408.
42. Shaul, Y., Garcia, P. D., Schonberg, S., and Rutter, W. J. (1986) Integration of hepatitis B virus DNA in chromosome-specific satellite sequences. *J. Virol.* **59**, 731–734.
43. Nagaya, T., Nakamura, T., Tokino, T., Tsurimoto, T., Imai, M., Mayumi, T., et al. (1987) The mode of hepatitis B virus DNA integration in chromosomes of human hepatocellular carcinoma. *Genes Deve.* **1**, 773–782.
44. Matsumoto H., Yoneyama T., Mitamura K., Osuga T., Shimojo H., and Miyamura, T. (1988) Analysis of integrated hepatitis B virus DNA and cellular flanking sequences cloned from a hepatocellular carcinoma. *Int. J. Cancer* **42**, 1–6.
45. Quada, K., Saldanha, J., Thomas, H., and Monjardino, J. (1992) Integration of hepatitis B virus DNA through a mutational hot spot within the cohesive region in a case of hepatocellular carcinoma. *J. Gen. Virol.* **73**, 179–183.
46. Chen J. Y., Harrison, T. J., Tsuei, D. J., Hsu, T. Z., Zuckerman, A. J., Chan, T. S., and Yang, C. S. (1994) Analysis of integrated hepatitis B virus DNA and flanking cellular sequences in the hepatocellular carcinoma cell line HCC36. *Intervirology* **37**, 41–46.
47. Dhruva, B. R., Shenk, T., and Subramanian, K. W. (1980) Integration in vivo into simian virus 40 DNA of a sequence that resembles a certain family of genomic interspersed repeated sequences. *Proc. Natl. Acad. Sci. USA* **77**, 4514–4518.
48. Ou, C.-Y., Boone, L. R., and Yang, W. K. (1983) A novel sequence segment and other nucleotide structural features in the long terminal repeat of a BALB/c mouse genomic leukemia virus-related DNA clone. *Nucl. Acids Res.* **11**, 5603–5620.

49. Taruscio, D. and Maneulidis, L. (1991) Integration site preferences of endogenous retroviruses. *Chromosoma* **101**, 141–156.
50. Stevens, S. W. and Griffith, J. D. (1994) Human immunodeficiency virus type 1 may preferentially integrate into chromatin occupied by L1Hs repetitive elements. *Proc. Natl. Acad. Sci. USA* **91**, 5557–5561.
51. Stevens, S. W. and Griffith, J. D. (1996) Sequence analysis of the human DNA flanking sites of human immunodeficiency virus type 1 integration. *J. Virol.* **70**, 6459–6462.
52. Awady, M. K., Kaplan, J. B., O'Brien, S. T., and Burd, R. D. (1987) Molecular analysis of integrated human HPV16 sequences in the cervical cancer cell line SiHa. *Virology* **159**, 389–398.
53. Baker, C. C., Phelps, W. C., Lindgren, V., Braun, M. J., Gonda, M. A., and Howley, P. M. (1987) Structural and transcriptional analysis of HPV16 sequences in cervical carcinoma cell lines. *J. Virol.* **61**, 962–971.
54. Wagatsuma, M., Hashimoto, K., and Matsukura, T. (1990) Analysis of integrated human HPV16 DNA in cervical cancers, amplification of viral sequences together with cellular flanking sequences. *J. Virol.* **64**, 813–821.
55. Bruni, R., Argentini, C., D'Ugo, E., Giuseppetti, R., and Rapicetta, M. (1997) Woodchuck hepatitis virus DNA integration in a common chromosomal region of the woodchuck genome in two independent hepatocellular carcinomas. *Arch. Virol.* **142**, 499–509.
56. Gong, S. S., Jensen, A. D., Chang, C. J., and Rogler, C. E. (1999) Double-stranded linear duck hepatitis virus virus (DHBV) stably integrates at higher frequency than wild-type DHBV in LMH chicken hepatoma cells. *J. Virol.* **73**, 1492–1502.
57. Mueller, P. R. and Wold, B. (1989) In vivo footprinting of a muscle specific enhancer by ligated mediate PCR. *Science* **246**, 780–786.
58. Shyamala, V. and Ames, G. F.-L. (1989) Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR. *Gene* **84**, 1–8.
59. Zhang, Y. and Frohman, M. A. (1997) Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs, in *cDNA Library Protocols*, Cowell, I. G. and Austin, C. A., eds., Humana, Totowa, NJ, pp.61–87.
60. Mizobuchi, M. and Frohman, L.A. (1993) Rapid amplification of genomic DNA ends. *BioTechniques* **15**, 214–216.
61. Wattel, E., Vartanian, J. P., and Wain-Hobson, S. (1995) Clonal expansion of HTLV-I infected cells in asymptomatic and symptomatic carries without malignancy. *J. Virol.* **69**, 2863–2868.
62. Cavois, M., Gessain, A., Wain-Hobson, S., and Wattel, E. (1996) Proliferation of HTLV-I infected circulating cells in vivo in all asymptomatic carries and patients with TSP/HAM. *Oncogene* **12**, 2419–2423.
63. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell S., et al. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucl. Acids Res.* **18**, 2887–2890.
64. Arnold, C. and Hodgson, I. J. (1991) Vectorette PCR: a novel approach to genomic walking. *PCR Meth. Appl.* **1**, 39–42.
65. Rubie, C., Schulze-Bahr, E., Wedekind, H., Borggreffe, M., Haverkamp, W., and Breithardt, G. (1999) Multistep-touchdown vectorette-PCR-a rapid technique for the identification of IVS in genes. *BioTechniques* **27**, 414–418.
66. Proffitt, J., Fenton, J., Pratt, G., Yates, Z., and Morgan, G. (1999) Isolation and characterization of recombination events involving immunoglobulin heavy chain switch regions in



- multiple myeloma using long distance vectorette PCR (LDV-PCR) *Leukemia* **13**, 1100–1107.
67. Devon, R. S., Porteous, D. J., and Brookes, A. J. (1995) Splinkerettes - improved vectorettes for greater efficiency in PCR walking. *Nucl. Acids Res.* **23**, 1644–1645.
68. Ivics, Z., Hackett, P. B., Plasterk, R. H., and Izsvak, Z. (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510.
69. Williams, J. G. K., Kubelik, A. R., Livak, K. J., and Rofalski, J. A. (1990) DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* **18**, 6531–6535.
70. Parker, J. D., Rabinovitch, P. S., and Burmer, G. C. (1991) Targeted gene walking polymerase chain reaction. *Nucl. Acid Res.* **19**, 3055–3060.
71. Hui, E. K.-W., Wang, P.-C., and Lo, S. J. (1998) Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. *Cell. Mol. Life Sci.* **54**, 1403–1411.
72. Parks, C. L., Chang, L.-S., and Sheuk, T. (1991) A polymerase chain reaction mediated by a single primer: cloning of genomic sequences adjacent to a serotonin receptor protein coding region. *Nucl. Acids Res.* **19**, 7155–7160.
73. Liang, P. and Pardee, A. B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967–971.
74. Carulli, J. P., Artinger, M., Swain, P. M., Root, C. D., Chee, L., Tulig, C., et al. (1998) High throughput analysis of differential gene expression. *J. Cell. Biochem. Suppl.* **30/31**, 286–296.
75. Matz, M. V. and Lukyanov, S. A. (1998) Different strategies of differential display: area of application. *Nucl. Acids Res.* **26**, 5537–5543.
76. Sarkar, G., Turner, R. T., and Bolander, M. E. (1993) Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *PCR Meth. Appl.* **2**, 318–322.
77. Liu, Y.-G., Mitsukawa, N., Oosumi, T., and Whittier, R. F. (1995) Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457–463.
78. Campisi, L., Yang, Y., Heiling, E., Herman, B., Carsista, A. J., Allen, D. W., et al. (1999) Generation of enhancer trap lines in Arabidopsis and characterization of expression patterns in the inflorescence. *Plant J.* **17**, 699–707.
79. Smith, D., Yanai, Y., Liu, Y.-G., Ishiguro, S., Okada, K., Shibata, D., et al. (1996) Characterisation and mapping of Ds-Gus-T-DNA lines for targeted insertional mutagenesis. *Plant J.* **10**, 721–732.
80. Parinov, S., Sevugan, M., Ye, D., Yang, W.-C., Kumaran, M., and Sundaresan, V. (2000) Analysis of flanking sequences from Dissociation insertion lines: a database for reverse genetics in Arabidopsis. *Plant Cell* **11**, 2263–2270.
81. Okamoto, H. and Hirochika, H. (2000) Efficient insertion mutagenesis of Arabidopsis by tissue culture-induced activation of the tobacco retrotransposon Tto1. *Plant J.* **23**, 291–304.
82. Casper, C., Leib-Mosch, C., Salmons, B., Gunzburg, W. H., Baumann, G., Hofler, H., et al. (1998) Mapping of mouse mammary tumor virus integration site by retroviral LTR-arbitrary polymerase chain reaction. *Virus Res.* **54**, 207–215.
83. Borenshtein, R. and Davidson, I. (1999) Development of the hot spot-combined PCR assay for detection of retroviral insertions into Marek's disease virus. *J. Virol. Meth.* **82**, 119–127.



84. Liu, Y.-G. and Whittier, R. F. (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. *Genomics* **25**, 674–681.
85. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.
86. Morishita, K., Parker, D. S., Mucenski, M. L., Jenkins, N. A., Copeland, N. G., and Ihle, J. N. (1988) Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell* **54**, 831–840.
87. Askew, D. S., Bartholomew, C., Buchbeg, A. M., Valentine, M. B., Jenkins, N. A., Copeland, N. G., and Ihle, J. N. (1991) His-1 and His-2: identification and chromosomal mapping of two commonly rearranged sites of viral integration in a myeloid leukemia. *Oncogene* **6**, 2041–2047.
88. Loh, E. Y., Elliott, J. F., Cwirla, S., Lanier, L. L., and Davis, M. M. (1989) Polymerase chain reaction with single-sided specificity: analysis of T cell receptor  $\delta$  chain. *Science* **243**, 217–220.
89. Ohara, O., Dorit, R. L., and Gilbert, W. (1989) One-sided polymerase chain reaction: the amplification of cDNA. *Proc. Natl. Acad. Sci. USA* **86**, 5673–5677.
90. Valk, P. J. M., Joosten, M., Vankan, Y., Lowenberg, B., and Delwel, R. (1997) A rapid RT-PCR based method to isolate complementary DNA fragments flanking retrovirus integration sites. *Nucl. Acids Res.* **25**, 4419–4421.
91. Lagerstrom, M., Parik, J., Malmgren, H., Stewart, J., Pettersson, U., and Landegren, U. (1991) Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *PCR Meth. Appl.* **1**, 111–119.
92. Sorensen, A. B., Duch, M., Jorgensen, P., and Petersen, F. S. (1993) Amplification and sequence analysis of DNA flanking integrated proviruses by a simple two-step polymerase chain reaction method. *J. Virol.* **67**, 7118–7124.
93. Frey, M., Settner, C., and Gierl, A. (1998) A general method for gene isolation in tagging approaches: Amplification of insertion mutagenised sites (AIMS) *Plant J.* **13**, 717–721.
94. Frey, M., Chomet, P., Glawischnig, E., Stettner, C., Grun, S., Winklmeier, A., et al. (1997) Analysis of a chemical defense mechanism in grasses. *Science* **277**, 696–699.
95. Peters, G. (1990) Oncogenes at viral integration sites. *Cell Growth Diff.* **1**, 503–510.
96. Jonkers, J. and Berns, A. (1996) Retroviral insertional mutagenesis as a strategy to identify cancer genes. *Biochim. Biophys. Acta* **1287**, 29–57.
97. Huang, S.-H. (1997) Inverse PCR approach to cloning cDNA ends, in *cDNA Library Protocols* (Cowill, J. G. and Austin, C. A., eds., Humana, Totowa, NJ, pp.89–96.
98. Hodzic, D., Frey, B., Marechal, D., Scarcez, T., Grooteclaes, M., and Winkler, R. (1999) Cloning of breakpoints in and downstream the IGF2 gene that are associated with overexpression of IGF2 transcripts in colorectal tumours. *Oncogene* **18**, 4710–4717.
99. Moynihan, T. P., Markham, A. F., and Robinson, P. A. (1996) Genomic analysis of human multigene families using chromosome-specific vectorette PCR. *Nucl. Acids Res.* **24**, 4094–4095.
100. Triglia, T. (2000) Inverse PCR (IPCR) for obtaining promoter sequence. *Meth. Mol. Biol.* **130**, 79–83.
101. Terauchi, R. and Kahl, G. (2000) Rapid isolation of promoter sequences by TAIL-PCR: the 5'-flanking regions of Pal and Pgi genes from yams (*Dioscorea*) *Mol. Gen. Genet.* **263**, 554–560.

102. Witthuhn, R. C., Harrington, T. C., Wingfield, B. D., Steimel, J. P., and Wingfield, M. J. (2000) Deletion of the MAT-2 mating-type gene during uni-directional mating-type switching in *Ceratomyxozoon*. *Curr. Genet.* **38**, 48–52.
103. Akiyama, K., Watanabe, H., Tsukada, S., and Sasai, H. (2000) A novel method for constructing gene-targeting vectors. *Nucl. Acids Res.* **28**, E77.
104. Cormack, R. S. and Somssich, I. E. (1997) Rapid amplification of genomic ends (RAGE) as a simple method to clone flanking genomic DNA. *Gene* **194**, 273–276.
105. Jones, D. H. and Winistorfer, S. C. (1992) Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. *Nucl. Acids Res.* **20**, 595–600.
106. Jones, D. H. and Winistorfer, S. C. (1993) Genome walking with 2– to 4–kb steps using panhandle PCR. *PCR Meth. Appl.* **2**, 197–203.
107. Jones, D. H. (1995) Panhandle PCR, in *PCR Primer: A Laboratory Manual*, Dieffenbach, C. W. and Dveksler, G. S., eds., Cold Spring Harbor Laboratory Press, NY, pp.411–419.
108. Weber, K. L., Bolander, M. E., and Sarkab, G. (1998) Rapid acquisition of unknown DNA sequence adjacent to a known segment by multiplex restriction site PCR. *BioTechniques* **25**, 415–419.
109. Baury, B., Masson, D., Lustenberge, P., and Denis, M. G. (1999) Gene walking by PCR amplification of short fragments from Taq DNA polymerase - modified P1 plasmid DNA and TA cloning. *BioTechniques* **27**, 1118–1122.
110. Rudi, K., Fossheim, T., and Jakobsen, K. S. (1999) Restriction cutting independent method for cloning genomic DNA segments outside the boundaries of known sequences. *BioTechniques* **27**, 1170–1177.
111. Zhang, Z. and Gurr, S. J. (2000) Walking into the unknown: a “step down” PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. *Gene* **253**, 145–150.
112. Tamme, R., Camp, E., Kortschak, R. D., and Lardelli, M. (2000) Nonspecific, nested suppression PCR method for isolation of unknown flanking DNA. *BioTechniques* **28**, 895–902.