Journal of Entomology and Zoology Studies

Available online at www.entomoljournal.com

E-ISSN: 2320-7078
P-ISSN: 2349-6800
www.entomoljournal.com
JEZS 2020; 8(2): 444-449

Received: 10-01-2020
Accepted: 12-02-2020

Samar Jyoti Chutia
M.F.Sc. (Fish Biotechnology From ICAR-CIFE): Ph.D. Scholar, Dept. of Aquaculture, College of Fisheries, Assam Agricultural University, Assam, India

Garima Bora
College of Fisheries, Assam Agricultural University, Assam, India

Mukesh Kumar
M.F.Sc. (Fish Biotechnology From ICAR-CIFE): Ph.D. Scholar, Fish Biotechnology, ICAR-CIFE, Versova, Andheri (W), Mumbai, Maharashtra, India

Rupam Jyoti Nath
Pursuing M.F.Sc., Department of AEM, College of Fisheries, Assam Agricultural University, Assam, India

Yashwanth BS
M.F.Sc. (Fish Biotechnology from ICAR-CIFE)**:** Ph.D. Scholar, Fish Biotechnology, ICAR-CIFE, Versova, Andheri (W), Mumbai, Maharashtra, India

Pranab Dihingia
Pursuing M.F.Sc., Department of Aquaculture, ICAR-CIFE, Versova, Andheri (W), Mumbai, Maharashtra, India

Nishi Sarmah
Pursuing M.F.Sc., Department of AEM, College of Fisheries, Assam Agricultural University, Assam, India

Nevil Pinto
SRF (DBT Funded Project), ICAR-CIFE, Versova, Andheri (W), Mumbai, Maharashtra, India

Prathik MR
Pursuing M.F.Sc., Fish Biotechnology, ICAR-CIFE, Versova, Andheri (W), Mumbai, Maharashtra, India

Dipak Kumar Sarma
Professor, Dept. of Aquaculture, College of Fisheries, Assam Agricultural University, Assam, India

Corresponding Author:
Rupam Jyoti Nath
Pursuing M.F.Sc., Department of AEM, College of Fisheries, Assam Agricultural University, Assam, India

Recent developments in RACE-PCR for the full-length cDNA identification

Samar Jyoti Chutia, Garima Bora, Mukesh Kumar, Rupam Jyoti Nath, Yashwanth BS, Pranab Dihingia, Nishi Sarmah, Nevil Pinto, Prathik MR and Dipak Kumar Sarma

Abstract
Novel strategies have been developed for the identification of full-length cDNA using bioinformatics tools and multiplexed PCR methods. Usually, sequences are either incomplete or have missing UTR-sequences. Researchers still use the Rapid Amplification of cDNA Ends technique to obtain the full-length cDNA sequences. An oligo(dT) anchor-primer with a hairpin structure is designed for amplification of 3′ cDNA ends. Arbitrary degenerate and sequence-specific reverse primers were also developed for the amplification of 5′ cDNA ends, and tail-PCR is performed until the 5′ sequence of multi-assembled fragment reaches the exon-1 region which can be identified by aligning these fragments to reference genome database. Inhibitory and functional adapters are specially designed; adapter with phosphate can attach to full-length mRNAs with cap structure. In prokaryotes, a technique is reported to capture primary transcripts based on capping the 5′ triphosphorylated RNA. For the identification of full-length cDNA, specific-primers need to be designed for further processes.

Keywords: mRNA, RACE, adapter, Tail-PCR, Cap-switching, and arbitrary degenerate (AD)

Introduction
Full-length cDNA sequence information is a prerequisite for the functional characterization of genes. Identification of an unknown gene sequence is done traditionally by degenerate primer PCR targeting evolutionarily conserved regions of the coding sequence. This often results in only partial cDNA sequence information with missing 5' and 3' sequence. RACE (Rapid amplification of cDNA ends) is a conventional PCR based technique, used to obtain the full-length sequence of an RNA transcript (Eyal, Y., Neumann, H., Or, E. and Frydman, A. 1999 and Park, D.J., Park, A.J., Renfree, M.B. and Graves, J.A.M. 2003) [1, 2]. The transcriptomes of the increasing number of species have been sequenced by next-generation sequencing (NGS) technique. However, the transcriptomes are always incomplete in length, especially the ends of genes, and due to the complexity and high cost, the complete genome sequencing is only conducted in some limited species. Therefore, researchers still use the PCR based RACE technique to identify the full-length cDNA sequences (Chen, N., Wang, W.M., and Wang, H.L., 2016) [3]. Also, RACE PCR is the most simple and cost-effective technique when the study involves a few targeted genes. Two general RACE strategies exist: one amplifies 5′ cDNA ends (5′ RACE) and the other captures 3′ cDNA end sequences (3′ RACE). Gene-specific primers are designed within the known cDNA sequence oriented towards the unknown sequence. In 3’ RACE, an oligo dT adaptor primer targeted against the universal polyA tail of mRNAs is used as a reverse primer along with the gene-specific forward primer. In 5' RACE, a known sequence is first incorporated into the 5' end of the mRNA sequence either through homopolymeric tailing or adaptor ligation. Then PCR is performed using a forward primer against the homopolymeric tail or adaptor sequence and a gene-specific reverse primer (Bower and Johnston, 2010 and Fromont-Racine, M., Bertrand, E., Pictet, R. and Grange, T., 1993) [4, 5]. Creating a small known region at the 5' cDNA end is a major issue (Edery, I., Chu, L.L., Sonenberg, N. and Pelletier, J. 1995; Bower and Johnston, 2010) [6, 4]. Homopolymeric tailing by the terminal transferase enzyme is the common practice. Common approaches, such as homopolymeric tailing and ligation anchored tailing requires a lot of enzymatic reaction after completion of reverse transcription (1st strand cDNA synthesis). Each enzymatic reaction can potentially introduce failures and destroy the integrity of cDNA. However, the efficiency of TdT is low, resulting in poor yield (Schaefer, 1995) [7].

Also, homopolymeric tailing does not differentiate between full-length (5' cap intact) and 5' truncated transcripts. Hence, there is a higher representation of 5' partial cDNAs.

Modifications in 5' RACE technique

Nowadays, several RACE techniques including RLM-RACE and Cap-switching RACE have been developed, especially for the 5′ RACE due to the difficulty in operation (Liu and Gorovsky, 1993; Schmidt and Mueller, 1999; Schramm, G., Bruchhaus, I. and Roeder, T., 2000) [8-10]. In the new RACE or RLM-RACE, before conducting the reverse transcription (RT) of mRNA into the first-strand cDNA, an anchored RNA adaptor is ligated to the 5′ ends of mRNA. Shrimp alkaline phosphatase (SAP) is used to remove a phosphate group, and then with the help of T4 RNA ligase, the anchor RNA adaptor is ligated to the 5′ ends. However, the RNA truncated at the 5' end can also be ligated to the RNA adaptor in this method. To address this limitation, an alternate method, oligo-capping, was developed (Maruyama and Sugano, 1994) [11].

The Cap-switching RACE (CapFinder) is performed by Moloney murine leukemia virus (MMLV) reverse transcriptase. The enzyme can add extra 2-4 cytosines to the 3′ ends of newly synthesized first-strand cDNAs upon reaching the 5' cap structure of mRNAs (Schramm, G., Bruchhaus, I. and Roeder, T., 2000) [10]. A template-switching oligonucleotide (TSO) containing a 3' poly(rG) tail is added to base pair with the extra cytosine residues. Thus, introducing a known adaptor to the 3' end of cDNA (TSO serves as a template for reverse transcription). The method has additionally been developed into inverse PCR, T-RACE, and step-out PCR to enhance the specificity and efficiency (Matz, M., Shagin, D., Bogdanova, E., Britanova, O., Lukyanov, S., Diatchenko, L. *et al.*, 1999 and Huang & Chen, 2006) [12, 13].

Recent Developments in RACE-PCR

cDNA made of degraded mRNA or full-length segments are difficult to identify or recognize in conventional RACE technique. Therefore, an improved RACE technique has been developed to reproduce only 5' end of full-length cDNA, exclusively. With the help of two kinds of adapters (inhibitory and functional adapters) all degraded mRNA components and unmethylated cap can be omitted (after the first-strand cDNA synthesis). The inhibitory adapter is designed especially (without phosphate group) to attach all un-wanted mRNA components and unmethylated cap. Suppressor adapter can't attach to full-length mRNAs with cap structure due to the lack of phosphate at the end of the 5' of the adapter. Therefore, degraded or uncapped mRNAs can be easily removed from the reaction. Having phosphate at 5' end of the functional adapter helps binding to mRNA with the cap structure. Thus the fruitful result can be obtained after only one PCR; 1st step PCR is enough to get the result with the minimum background (Purfard, a.m., Nouri, m.z. and Kolagar, a.d., 2017) [14].

This is a modified RACE method to separate the 5' end of the cDNA without using specific enzymes. The modified RACE has several advantages, such as requires less time and lower costs related to the commercially existing procedures; other advantages are:

1. Fast and simple and requires a minimum amount of total mRNA (About 1 µg).
2. Only one PCR is enough to get the result with the minimum background.
3. Removal of degraded mRNA or other nucleotide strings without cap structure.
4. High repeatability for amplification of the desired region.
5. Having phosphate at 5' end of the functional adapter helps binding to mRNA with cap structure.
6. This binding is maintained during PCR and provides a binding site of the primers at the 5' end of mRNA.

Using this RACE technique, 5' end of the *Aeluropus littoralis* SOS1 gene was isolated and sequenced by (Purfard, a.m., Nouri, m.z. and Kolagar, a.d., 2017) [14]. Specific primers were designed to amplify 5' end of the gene (Table 1).

Table 1: Primers sequence used in the new 5' RACE method

Name	Sequences (5′ → 3′)	Modification
P1	ACCTCGGCCG	-
P2	AGCGTGGTCGCGGCCGAGGT	-
P3	AAGGAGTAGTTT	-
P4	AAACTACTCCTTCAGTCCATGTCAGTGTCCTCGTGCTCCA	5′ phosphate
FP1	CTGGAGCACGAGGACACTG	-
FP2	CTGACATGGACTGAAGGAGTA	-
GSP1	TGCCATGTCTCTCAAAATGG	-
GSP2	CAATGTAGGCAACCATTTCCC	-

* P: Primer, FP: Functional primer, GSP: Gene-specific primer

P4 is the functional adapter, having phosphate at 5' end helps binding to mRNA with cap structure. This binding is maintained during PCR and provides a binding site of the primers at the 5' end of mRNA. The alignment of the sequence with the NCBI database confirmed the efficacy of the modified technique (Table 2).

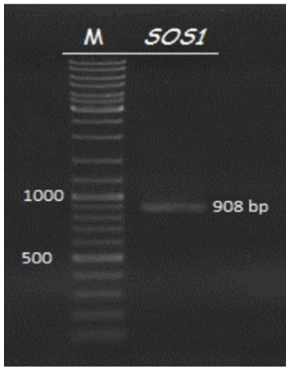

Fig 1: Amplified 5' end of SOS1 gene from *A. littolaris*

Table 2. Three items of the sequences in database of NCBI, which most closely resembled the 5' SOS1 sequence of the *Aeluropus littoralis*

Description	Max score	Total score	Query cover	E Value	Ident.	Accession
Aeluropus littoralis plasma membrane Na+/H+ antiporter (SOS1) mRNA, complete cds.	1639	1639	97%	0.0	100%	JN936862.1
Aeluropus littoralis plasma membrane Na+/H+ antiporter mRNA, complete cds. Distichlis spicata plasma membrane	1520	1520	95%	0.0	98%	HQ329792.2
Distichlis spicata plasma membrane Na+/H+ antiporter (SOS1-1) mRNA, complete cds.	1360	1360	95%	0.0	95%	FJ865581.1

Multiplex PCR based RACE

A novel strategy for amplification full-length cDNA and promoter sequences has been developed using bioinformatics technology and multiplexed PCR methods. The amplification of 3′ ends of cDNA is performed with oligo(dT)-anchor primer using a hairpin structure. For the amplification of 5′ ends of cDNA, 2 or 3-round of TAIL-PCR using arbitrary degenerate (AD) and sequence-specific reverse (SPR) primers is performed until the 5′ sequence of multi-assembled fragment reaches the exon1 region. It can be identified by aligning all the fragments to the reference genome database. By using genomic DNA as a template, another TAIL-PCR or touch-down PCR can be conducted to obtain the remaining 5′ and promoter sequences. The 5′ end sites of cDNA are predicted by aligning finally assembled fragment to homologous reference genes of other species and screening the relative locations of common characteristic cis-elements in silico on the promoter. The assumed 5′ ends are further confirmed by primers corresponding to these predicted sites in cDNAs (Chen, N., Wang, W.M., and Wang, H.L., 2016) [3].

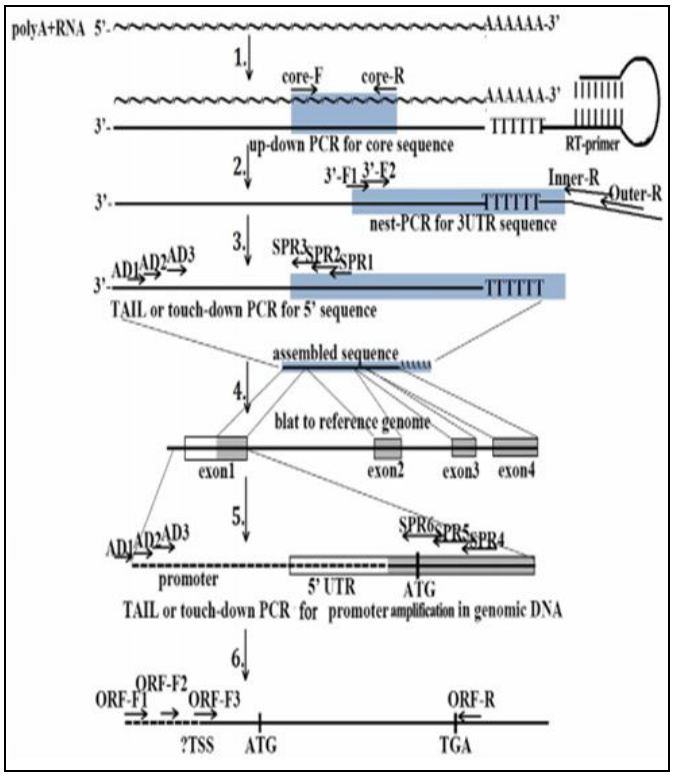

Fig 2: Schematic framework of the 3′ and 5′ RACE approaches

Step 1: First-strand cDNAs synthesis using a special hairpin structure oligo (dT)-anchor primer.

Step 2: Sequence-specific forward primers are designed together with Outer-R and Inner-R primers for 3′ end sequences of cDNA amplification.

Step 3: To amplify 5′ sequences of cDNA by TAIL or touch-down PCR, sequence-specific reverse primers (SPRs) referring to the assembled fragments of 3′ end and core sequences are designed and combined with arbitrary degenerate primers (ADs).

Step 4: After re-assembling, the fragments together with homologous reference gene sequences are aligned to reference genome database for exon distribution studies.

Step 5: Genomic DNA is used as a template for further analysis.

The PCRs in cDNA will not continue until the multiassembled fragments reach the exon1 region, then other SPRs denoting to the sequence located in exon1 are designed. And one more TAIL or touch-down PCR is performed to obtain the promoter sequences.

Step 6: The 5′ end sites of finally assembled sequence are in silico predicted and validated in cDNAs.

The key feature of the TAIL-PCR is thermal asymmetry {Liu and Whittier (1995)} [15]. In tail-PCR protocol, interlaced cycles of high and low stringency take advantage of the thermal asymmetry of the primers to amplify the preferred target product. The arbitrary degenerate primers (AD primers), which are shorter and, therefore, anneal at lower temperatures. Depending on their level of degeneracy, AD primers can hybridize at random to many sites in the genome. And the insertion specific primers with high annealing temperatures favor hybridizing of the long insertion specific primers. SPR primers are designed to read outwards from the known sequence into the unknown flanking sequence. The primary round in the TAIL-PCR protocol starts with 5 cycles with high annealing temperature (i.e., high stringency). In the second step, a cycle with low annealing temperature (i.e., low stringency) allows attachment and extension of nonspecific AD primers. Because of the lower temperature, mismatch pairing is allowed, thus generating AD primer-specific target sites for the next round. TAIL cycling consists of 15 super cycles, in which each super cycle consists of two high stringency cycles and one low stringency cycle.

In the subsequent TAIL-cycling rounds, the preferred target molecule, *i.e.*, type I product (primed by sequence-specific reverse primer and AD primer) is amplified together with nonspecific type II product (primed on both sides by the insertion-specific primer), and another class of nonspecific products, termed as type III (primed on both sides by nonspecific AD primers. and type III products).

Dilution of the primary TAIL-PCR followed by several super-cycles of TAIL-cycling (two high stringency cycles interlaced with one low stringency cycle) with nested SPR, i.e., internal to 1st SPR in the second round ensures the amplification of specific type I products is increased over non-specific products. The secondary round of TAIL-amplification is

followed by a third round using another set of nested insertion-specific primers. Again dilution of the secondary PCR products is essential before employ 3rd round of TAIL-PCR cycling. A limited number of high stringency cycles (20) ensures that all undesired products fail to amplify.

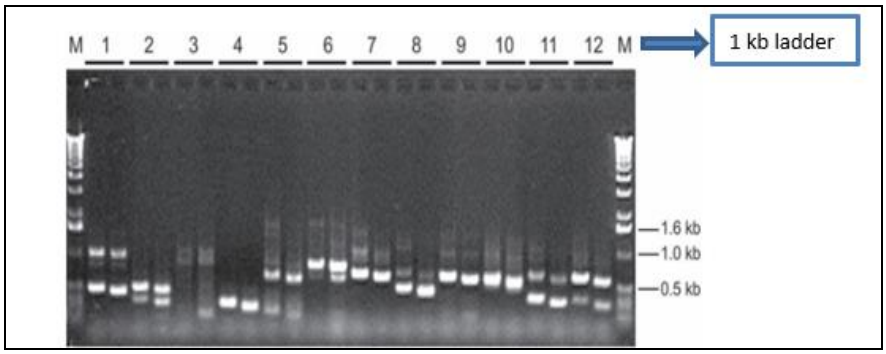

Fig 3: 1% agarose gel with EtBr staining.

Twelve samples of secondary and tertiary TAIL-PCR products are loaded next to each other. Note the slight decrease in size of tertiary TAIL-PCR products.

Through visual inspection on agarose gels of secondary and tertiary PCR products loaded next to each other, specific products are verified through a decrease in product size according to the position of the nested primers (Singer and Burke, 2003) [16].

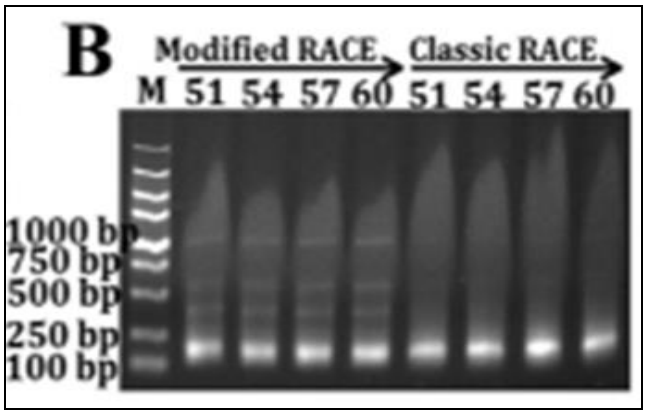

Fig 4: Results comparison of Modified RACE and Classic RACE

The results revealed that the 3′ end of HSF2 was only obtained in cDNAs reversed by hairpin structure (modified RACE) and classic 3′ RACE RT primers, and cDNAs reversed by hairpin structure RT primer showed more clear bands than classic 3′ RACE RT primer in a wider range of denaturation temperature.

Chen, N., Wang, W.M., and Wang, H.L. (2016) [3] compared the efficiency and specificity of the oligo(dT)-anchor primer with a special hairpin structure by amplifying the 3′ ends of low expressed heat shock transcription factor 2 (HSF2) gene of *M. amblycephala*.

Capping-RACE: For Prokaryotes

Prokaryotic RNA generally lack the 5' cap structure, which is essential for the application of CapFinder RACE. By using Vaccinia capping enzyme (VCE), the addition of a 7-methylguanylate cap structure to the 5' end of triphosphorylated RNA was performed (Liu, F., Zheng, K., Chen, H.C. and Liu, Z.F., 2018) [17]. The VCE enzyme has all the necessary components to add 7-methylguanylate cap structures to the 5′end of RNA (Shuman, 1990) [18].

Vaccinia capping enzyme is composed of two main subunits, and those are D1 and D12. The D1 subunit carries three enzymatic activities: a) RNA triphosphatase, b) Guanylyl-transferase, and c) Guanine methyl-transferase (Guo and Moss, 1990 and Mao and Shuman, 1994) [19, 20].

In the incidence of the capping enzyme, reaction buffer, GTP, and the methyl donor, SAM; in-vitro transcripts can be capped in less than 1 hour. Capping is almost 100% efficient, and all capped structures are added in the proper orientation (Grudzien, E.W.A., Stepinski, J., Jankowska-Anyszka, M., Stolarski, R., Darzynkiewicz, E., and Rhoads, R.E., 2004) [21].

To test Capping-RACE in prokaryotes, Liu, F., Zheng, K., Chen, H.C. and Liu, Z.F., 2018) [17] selected three protein-coding genes, ompA, sodB, and shiA, of the model organism *E. coli* MG1655 as targets.

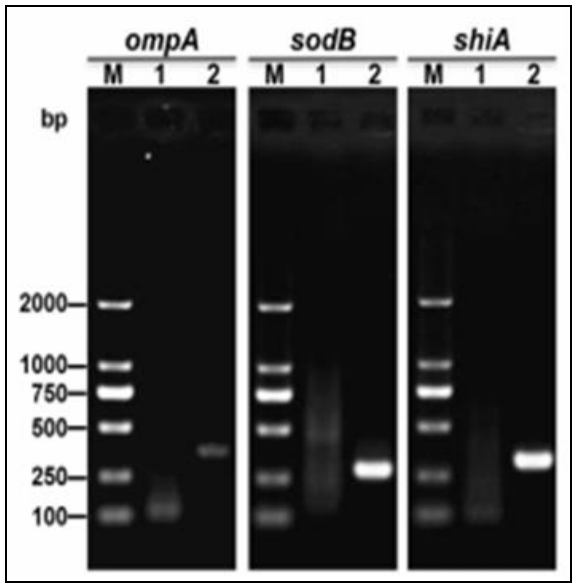

Fig 5: 5' RACE mapping of *E. coli* MG1655 protein-coding genes

Lane M: DL2000 DNA Marker (Takara)
Lane 1: first-round PCR products.
Lane 2: second-round PCR products

Conclusion

Often occurring problems in RACE-PCR are multiple bands and non-specific bands. Masamha and Todd (2018) [22] suggested a few PCR cycling conditions to solve some problems in RACE-PCR. Those are:

1. Increasing the initial denaturation length (up to 3 mins)

and temperature (98 °C) breaks any secondary structures and helps yield the single specific band because the T7 oligo dT primer generally forms weak secondary structures.
2. The long annealing time (5 mins) and extension or elongation step (10-15 mins) allow complete synthesis of incomplete amplicons, enabling the full extension of the initial and final amplification products.

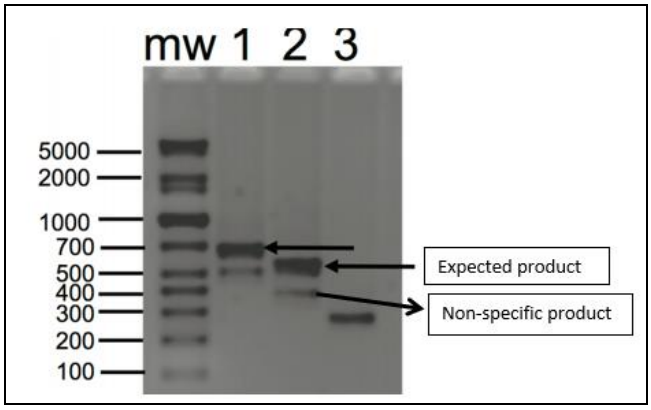

Fig 6: Gel pic showing the complete synthesis of incomplete amplicons by giving long annealing, extension time and higher PCR cycles

3. The non-specific bands are mostly lighter than the expected bands. Suggesting that it can appear at higher PCR cycles (eg., up to 35 cycles). Therefore, limiting the PCR cycles to only 20 reduces the amplification of the non-specific band.
4. Reverse transcription without the MMLV-RT enzyme can be used as a negative control. The presence of a band, in the negative control, signifies the potential contamination with the genomic DNA.

The major significance of this study is to minimize the limitations of different RACE-PCR techniques like Creating a small known region at the 5' cDNA end is a major issue; many a time, the sequences are either incomplete and have missing UTR sequences even after using several Bioinformatics tools and multiplexed PCR methods, and high-throughput revolutionized RNAseq. Therefore this study will help many researchers working in this field to select a proper RACE-kit or develop a cost-effective RACE-kit/chemicals which always enhances reliability for the identification of full-length cDNA of a gene. Thus for developing a kit researcher may focus on the following perspectives:

1. Proper isolation of RNA should be done as RNA isolation is the paramount objective.
2. The kit should be suitable and efficient so that the RNA isolation will increase and the minimum amount of RNA is required for rapid-reaction.
3. The chemicals present are the cornerstone of RACE-kit. They can be used for both 3' and 5' RACE-PCR.
4. The process must be very cost-effective as other process of whole-genome sequencing is extortionate.
5. The process enhances the reliability and success rate.

References

1. Eyal Y, Neumann H, Or E, Frydman A. Inverse single-strand RACE: an adapter-independent method of 5′ RACE. BioTechniques. 1999; 27:656-658.
2. Park DJ, Park AJ, Renfree MB, Graves JAM. 3′ RACE walking along a large cDNA employing tiered suppression PCR. BioTechniques. 2003; 34:750-756.
3. Chen N, Wang WM, Wang HL. An efficient full-length cDNA amplification strategy based on bioinformatics technology and multiplexed PCR methods. Scientific Reports. 2016; 6:19420.
4. Bower NI, Johnston IA. Targeted rapid amplification of cDNA ends (T-RACE)—an improved RACE reaction through degradation of non-target sequences. Nucleic Acids Research. 2010; 38(21):e194-e194.
5. Fromont-Racine M, Bertrand E, Pictet R, Grange T. A highly sensitive method for mapping the 5' termini of mRNAs. Nucleic Acids Res. 1993; 21:1683-1684.
6. Edery I, Chu LL, Sonenberg N, Pelletier J. An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). Molecular and cellular biology, 1995; 15(6):3363-3371.
7. Schaefer BC. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Analytical biochemistry. 1995; 227(2):255-273.
8. Liu X, Gorovsky MA. Mapping the 5 and 3 ends of tetrahymena thermophila mrnas using rna ligase mediated amplification of cdna ends (rlm-race). Nucleic Acids Res. 1993; 21:4954-4960.
9. Schmidt WM, Mueller MW. Capselect: a highly sensitive method for 5 cap-dependent enrichment of full-length cdna in pcrmediated analysis of mrnas. Nucleic Acids Res. 1999; 27:e31.
10. Schramm G, Bruchhaus I, Roeder T. A simple and reliable 5 -race approach. Nucleic Acids Res. 2000; 28:E96.
11. Maruyama K, Sugano S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mrnas with oligoribonucleotides. Gene. 1994; 138:171-174.
12. Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L *et al*. Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic acids research. 1999; 27(6):1558-1560.
13. Huang JC, Chen F. Simultaneous amplification of 5′ and 3′ cDNA ends based on template-switching effect and inverse PCR. Biotechniques. 2006; 40:187-189
14. Purfard AM, Nouri MZ, Kolagar AD. Improved race technique for the full sequence identification and amplification of mrna ends. Plant cell biotechnology and molecular biology, 2017, 520-526.
15. Liu YG, Whittier RF. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics. 1995; 25:674-681.
16. Singer T, Burke E. High-throughput TAIL-PCR as a tool to identify DNA flanking insertions. In Plant Functional Genomics (pp. 241-271). Humana Press, 2003.
17. Liu F, Zheng K, Chen HC, Liu ZF. Capping-RACE: a simple, accurate, and sensitive 5′ RACE method for use in prokaryotes. Nucleic acids research. 2018; 46(21):e129-e129.
18. Shuman S. The catalytic activity of vaccinia mRNA capping enzyme subunits coexpressed in Escherichia coli. Journal of Biological Chemistry. 1990; 265(20):11960-11966.
19. Guo PX, Moss B. Interaction and mutual stabilization of the two subunits of vaccinia virus mRNA capping enzyme coexpressed in Escherichia coli. Proceedings of

the National Academy of Sciences. 1990; 87(11):4023-4027.
20. Mao X, Shuman S. Intrinsic RNA (guanine-7) methyltransferase activity of the vaccinia virus capping enzyme D1 subunit is stimulated by the D12 subunit. Identification of amino acid residues in the D1 protein required for subunit association and methyl group transfer. Journal of Biological Chemistry. 1994; 269(39):24472-24479.
21. Grudzien EWA, Stepinski J, Jankowska-Anyszka M, Stolarski R, Darzynkiewicz E, Rhoads RE. Novel cap analogs for *in vitro* synthesis of mRNAs with high translational efficiency. Rna. 2004; 10(9):1479-1487.
22. Masamha CP, Todd Z. Adapting 3'Rapid Amplification of cDNA Ends to Map Transcripts in Cancer. JoVE (Journal of Visualized Experiments). 2018; (133):e57318.