

Repeated Template Switching: Obstacles in cDNA Libraries and Ways to Avoid Them

Zoltan Villanyi^{1,2,*}, Antal Mai¹ and Janos Szabad¹

¹University of Szeged, Faculty of Medicine, Department of Biology, H-6720 Szeged, Somogyi Street 4, Hungary

²University of Szeged, Faculty of Biology, Department of Biochemistry and Molecular Biology, H-6726 Szeged, Középfasor 52, Hungary

Abstract: We screened zebra finch cDNA libraries with the plaque-based competitive hybridization technique for genes differentially expressed in the brains of courting and non-courting males. The cDNA libraries were generated by the SMARTTM cDNA Library Construction Kit.

Sixty-six percent (35/53) of the isolated clones possessed high homology to sequences present in the cDNA libraries of species ranging from *Arabidopsis* to humans. The clones had to be false positives since the corresponding sequences are absent in genomic libraries.

The false positive clones were tandem repeats of different lengths of the SMART IVTM oligonucleotide and originated through repeated template switching during reverse transcription. We report here based on *in silico* data that multiple template switching and subsequent repeated incorporation of primers at 3' end of first strand cDNA is a common event during the generation of cDNA libraries created by the template switching strategy. We demonstrate that the repeat containing clones disturbs hybridization procedures aimed to identify differentially expressed genes. We propose an additional hybridization step with repeat containing clones to avoid the identification of false positives, and recommend the use of switch primers without palindrome sequences to avoid the formation of tandem repeats during cDNA library construction.

Keywords: Artifact sequences, false positives, reverse transcription, template switching, plaque-based competitive hybridization.

INTRODUCTION

Reverse transcriptase (RT) accomplishes two template-switching events during reverse transcription of the retroviral genome to complete DNA synthesis at both ends of the retroviral RNA [1]. Reverse transcriptase frequently undertakes additional template switches that lead to great variability in retroviral populations [2]. Reverse transcriptase switches templates in a homology-dependent manner. One should be aware of RT actions when RT is used as a tool for synthesizing cDNAs. Template switching can also take place between different mRNAs and even between different parts of the same mRNA once it contains internal repeats. The generated cDNA fragments can be mistakenly interpreted as alternative splice isoforms. Although RTs with increased thermostability and lacking RNaseH activity have largely resolved the above problem [3], template switching is frequently made use of, as an advantage, during the construction of full-length cDNA libraries [4]: oligo d(T) primers containing a linker are used to create the first strand cDNA on the basis of an mRNA, after which a few additional nucleotides, primarily deoxycytidines, are

attached to the 3' end of the newly synthesized cDNA strand by the terminal transferase activity of RT. So-called switch primers are also used with an oligo (G) sequence at their 3' ends to anchor the first cDNA strand that contains a few additional (C) nucleotides added by the recommended RT (Superscript II). Then, RT switches templates and extends the cDNA with the complementary sequence of the switch primer. The cDNAs with a switch primer at their 5' ends and a linker at their 3' ends with restriction sites can then be PCR amplified and cloned (Fig. 1A).

We have recently developed a fast and sensitive technique called plaque-based competitive hybridization (PBCH) to identify genes differentially expressed in different samples [5] and made use of it by comparing the transcriptomes of courting and non-courting zebra finch males. Zebra finch (*Taeniopygia guttata*) is an important model species in neurobiological studies and eventually an international consortium has unveiled its genome [6].

While they are caring for their offspring the zebra finch males do not court. The differences in the gene expression patterns of courting and caring zebra finch male brains may reveal the molecular components that underlie altruistic (caring) and selfish behavior (courting). Using PBCH, we compared cDNA libraries, prepared with the SMARTTM cDNA library construction kit, that reflect the transcriptomes

*Address correspondence to this author at the University of Szeged, Faculty of Medicine, Department of Biology, H-6720 Szeged, Somogyi street 4, Hungary; Tel: 62/ 544544; E-mails: villanyi@sb4.szote.u-szeged.hu, villanyi22@gmail.com

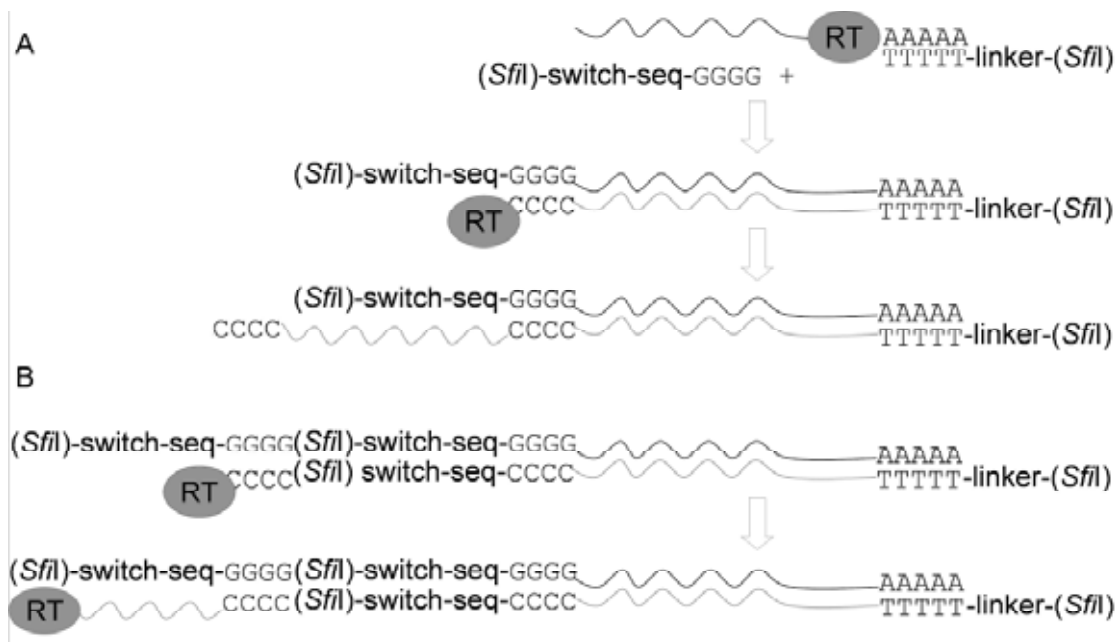


Fig. (1). The origin of tandem repeats during reverse transcription. A: Flow chart showing template switching. **B:** Flow chart showing the mechanism of repeat formation during repeated template switching. Black and gray lines represent mRNA and first strand cDNA, respectively. RT, reverse transcriptase. *SfiI*-switch-seq stands for the switch primer with a *SfiI* restriction site inside.

of courting and non-courting zebra finch males in order to identify the genes that differ in the two types of behavior.

Every cDNA library contains low levels of chimeric and no-insert clones, and other “contaminating” sequences [7]. The artificial clones are usually not abundant and do not interfere with further analyses. However, while screening the zebra finch brain cDNA libraries for clones representing differentially expressed genes between courting and non-courting males, we frequently identified clones that contained numerous repeats of the switch primer. It also turned out that EST databases ranging from *Arabidopsis* to humans contain several clones consisting of or containing switch primer repeats (Supplementary file 1). Since the repetitive sequence is not present in any of the genomic libraries, several of the cDNAs must contain artifact sequences.

The results described here suggest that template switching may take place repeatedly at the 5' end of the cDNAs. After the first switch primer base-pairs with the deoxycytidine stretch, RT completes transcription of the switch primer as an extension of the cDNA. Due to its terminal transferase activity, RT adds a few additional (C) nucleotides to the 3' end of the extended cDNA and another primer can anchor to the clone, thus creating a repeat (Fig. 1B). These are the repeats that can subsequently disturb the hybridization procedures including the widely used suppression subtractive hybridization or microarray hybridization. The long primer repeats in the different clones can hybridize with each other, which results in the development of false-positives, for example on cDNA microarray slides or on dot blot membranes.

The present report describes the origin and nature of the artifact sequences, proposes ways to avoid their formation

and, once there, describes the identification of clones with these sequences while using the PBCH technique. The PBCH technique offers advantages when DNA microarray approaches are not feasible or are not available for the identification of differentially expressed genes. It combines the principles of differential hybridization with subtractive hybridization and is particularly useful when libraries with few differences are to be analyzed [5, 8]. A flowchart of PBCH is presented in the left-hand panel of Fig. (2).

MATERIALS AND METHODOLOGY

mRNA Isolation

The mRNA samples were isolated from whole brains of adult courting and non-courting zebra finch males using the FastTrack[®] MAG Maxi mRNA Isolation Kit (Invitrogen[™]), according to the manufacturer's instructions.

cDNA AND PLAQUE LIBRARIES

First strand cDNAs were constructed using the SuperScript II Reverse Transcriptase (Invitrogen[™]). The cDNA libraries were prepared with the SMART[™] cDNA Library Construction Kit (Clontech), according to the manufacturer's instructions, using 0.5 µg polyA RNA as the starting material. Plaque libraries were created with the aid of Gigapack[®] III Plus Packaging Extracts (Stratagene[®]).

PLAQUE-BASED COMPETITIVE HYBRIDIZATION

The experiments were performed according to Villanyi et al., (2008) [5] with minor modifications: cDNAs in the courting library were labeled with the Roche PCR DIG Labeling Kit and the signals were developed using the DIG Nucleic Acid Detection Kit (Roche). Petri dishes of 9 cm diameter and compatible Hybond-N+ hybridization

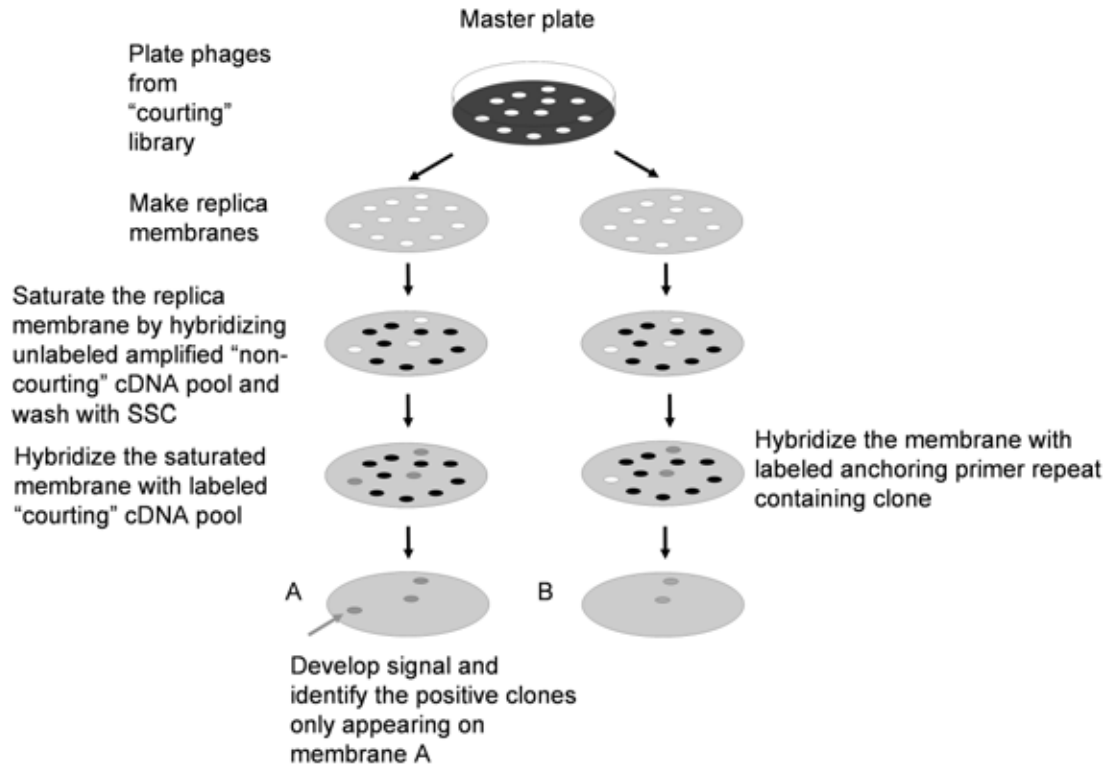


Fig. (2). Flow chart of plaque-based competitive hybridization combined with a library screening procedure to describe a way to avoid the identification of repeat-containing clones.

membranes (Amersham Pharmacia Biotech) were used. A 1 ng SMART IVTM oligonucleotide repeat-containing plasmid (previously identified in the courting library, λ phage clone available on request) was used as the template for PCR to DIG-label the probe in order to identify the repeat-containing clones on the plates. The 50 μ l reaction mixture also contained 0.2 μ M insert screening primers provided by Clontech, 1 unit DreamTaqTM polymerase (Fermentas) and 0.2 mM dATP, dCTP, dGTP. The concentration of dTTP was 130 μ M. The PCR program was as follows: 95 °C for 3 min followed by 24 cycles of: 95 °C for 40 s, 60 °C for 30 s, 72 °C for 3 min, and then 72 °C for 5 min (final extension; TC-24/H(b) PCR device; Bioer Technology, Hangzhou, China). The reaction mixture contained 70 μ M DIG-11-dUTP.

***Mwo*I DIGESTION TEST**

Positive plaques were identified and converted to pTriplEx2 plasmids according to the instructions in the SMARTTM cDNA Library Construction Kit (Clontech). The 50 μ l reaction mixture contained 1 ng plasmid DNA as template for PCR, 0.2 μ M insert screening primers provided by Clontech, 1 unit DreamTaqTM polymerase (Fermentas) and 0.2 mM dNTPs. The PCR program was as follows: 95 °C for 3 min followed by 30 cycles of: 95 °C for 40 s, 60 °C for 30 s, 72 °C for 3 min, and then 72 °C for 5 min (final extension). About one μ g of the PCR product was digested with 1 unit *Mwo*I (Fermentas) enzyme in 1 \times yellow buffer (Fermentas) for 1 h. The samples were subjected to electrophoresis in 1.5% agarose gel.

SEQUENCE ANALYSIS

The NCBI's BLAST page at <http://www.ncbi.nlm.nih.gov/Genomes/> was used for the sequence analysis.

RESULTS AND DISCUSSION

The cDNA libraries were prepared from brains of courting and non-courting zebra finch males (the "courting" and the "non-courting" libraries). The "non-courting" cDNA library was PCR amplified and then hybridized to replica membranes that were prepared from a λ phage "courting" library. The "courting" library was represented by about 60,000 plaques. The membrane was washed and the hybridization and washing steps were repeated in order to reduce the number of incomplete saturations. By using two rounds of hybridization the bias caused by occasional mispairing of non-homologous sequences became negligible.

The "courting" library – from which the phage library was prepared – was amplified and DIG-labeled by PCR. After hybridization, the saturated filter with the labeled library, the positive clones (only present in the courting library), were selected. The positive clones were sequenced and among the 53 clones identified 35 contained long repeats consisting of the switch primer (SMART IVTM oligonucleotide). The repeat containing clones always contained a polyA tail and a shorter or longer sequence corresponding to the 3' end of an mRNA. However we cannot take the captured mRNA sequences containing primer repeats as true positives since it is possible that the clones

only showed up as differentially expressed due to their long repetitive switch primer content.

To further demonstrate the abundant presence of the repeat containing clones in cDNA libraries we analyzed the EST database of *Anolis carolinensis*. We chose this EST dataset because it had been developed by one group based on a cDNA library created by the template switching strategy. The possible repeat-containing clones of this dataset may reflect the above described pitfalls of the SMART™ cDNA library generation system. We used the NCBI nucleotide blast page (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and blasted the *Anolis carolinensis* (taxid: 28377) EST database with a 25 anchor primer repeat containing the sequence 25x(5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTA CGGCCGG-3'). We analyzed the resulting blast output file (hit table) with Microsoft Excel software and collected the clones that contained at least three primer repeats (for the GenBank accession numbers, see supplementary file 2).

The repeat-containing clones may originate from multiple template switching during cDNA construction occurring at the 5' end of the cDNAs. Switch primer sequence:

(5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTA CGGCCGG-3').

We identified 3385 individual EST sequences that contained different numbers, but at least three repeats (Table 1). The *Anolis carolinensis* EST database currently contains 156,736 sequences and the 3385 repeat-containing clones make up 2.15% of the total library, indicating that the occurrence of repeat-containing clones can be relatively high in cDNA libraries created by the template switching strategy.

Blast searches with other products containing switch primer repeat sequences (In-Fusion® SMARTer™ cDNA library construction kit:

25x(5'-AAGCAGTGGTATCAACGCAGAGTACNNNNN-3'); EasyClone cDNA library construction kit: 25x(5'-AAGCAGTGGTATCAACGCAGAGTACGGGGG-3') revealed that repeated template switching takes place in a sequence-independent manner, as several ESTs appeared to contain repeats of the above switch primers (data not shown).

The cDNA library construction methods without a PCR amplification step and adapter ligation do not create repeats. For example, cDNA cloning by the biotinylated CAP trapper technique contains an oligo(dG) tailing step performed by terminal nucleotidyl transferase [9]. The oligo(dC) second-strand cDNA primer is introduced after the oligo(dG) tailing, thus the cloning site containing primers can anchor to the tailed cDNAs only once.

The EST databases based on cDNA libraries created by products that use adapter ligation for 5' primer addition to the cDNAs may also contain repetitive artifacts originating from multiple adapter ligations. For example, the CloneMiner™ II cDNA Library Construction Kit applies 5' primer ligation after RT has finished transcribing the first strand cDNA [10]. In this case, the presence of the Attb1 adapter.

(5'-TCGTCGGGGACAACCTTTGTACAAAAAAGTT GG-3') repeat-containing clones may originate from multiple adapter ligations occurring at the 5' end of the cDNAs. The

Table 1. Distribution of ESTs Containing Different Numbers of Switch Primer Repeat Sequences from the EST Database of *Anolis carolinensis*

Number of Repeats	Clones
21	2
19	1
18	2
17	4
16	8
15	9
14	13
13	14
12	24
11	27
10	28
9	42
8	61
7	125
6	187
5	341
4	855
3	1642
Σ	3385

(There are 156,736 ESTs in the database.)

Ascaris suum (taxid: 6253, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) EST database was generated based on the cDNA library created by the Invitrogen's CloneMiner cDNA Library Construction Kit. The library contains 56,118 sequences. We found 162 sequences that contain at least three repeats of the adapter, which represents 0.28% of the total library (Table 2, supplementary file 3).

Table 2. Distribution of ESTs Containing Different Numbers of 5' Adapter Repeat Sequences from the EST Database of *Ascaris suum*

Number of Repeats	Clones
7	3
6	4
5	2
4	19
3	134
Σ	162

(The database contains 56,118 sequences.)

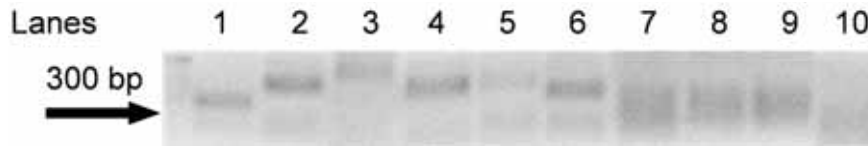


Fig. (3). Electrophoretogram of *MwoI*-digested clones from zebra finch brain cDNA libraries. Differentially expressed clones identified with plaque-based competitive hybridization technique appear as discrete bands after *MwoI* digestion (lanes 1-6). Clones with switch primer repeats inside are digested to a smear (lanes 7-10).

The repeat-containing clones may originate from multiple adapter ligations during cDNA construction occurring at the 5' end of the cDNAs. Adapter sequence:

(5'-TCGTCGGGGACAACCTTTGTACAAAAAAGTTGG-3').

We identified 35 artifact clones in the zebra finch library from screening 60,000 plaques, which represented less than 0.05% of the courting zebra finch library. However, since the repeat-containing clones can be identified as false positives with hybridization techniques aimed at identifying differentially expressed genes, even a relatively low abundance of the repeat-containing clones can disturb the procedure (in our work, 66% of the identified clones were artifacts).

The following procedure was developed to avoid identifying clones with long repeats of the switch primer. We prepared two replica membranes from the same master plate of the courting library. The PBCH experiment was performed with one of the filters according to Villanyi *et al.* (2008) [5]. The other filter was saturated with the unlabeled non-courting library (to reduce background effects) and hybridized with labeled clones containing ten tandem repeats of the switch primer (ten times the SMART IV™ oligonucleotide). This additional hybridization revealed all of the repeat-containing clones in the library. As shown earlier, the fidelity and sensitivity of PBCH enables the identification of one specific clone in a library sample of 15,000 [5]. By comparing the results of the PBCH experiment with the results of the above proposed additional hybridization, the repeat containing false-positive clones can be excluded. All further analyses were conducted with the positive clones that only showed up in the PBCH filter (Fig. 2).

Compared to the SMART IV™ oligonucleotide primer repeats, there were single nucleotide polymorphisms and deletions in the repetitive sequences in the various EST databases near or inside the *SfiI* recognition sites (Supplementary file 1). The *SfiI* restriction site serves as a cloning site for λ phage vectors during library construction. It appears that the *SfiI* palindrome sequence inside the switch primer frequently loops out during template switching and therefore reverse transcription skips the looping sites. Primers without palindrome(s) would most likely be correctly transcribed by RT during template switching. The incorrect reverse transcription is probably the reason why *SfiI* cannot digest the long repetitive sequences before cloning. However, there are *MwoI* restriction sites in the

switch primer repeats, and thus they can be digested (even though the *SfiI* sites are mutated and thus the disturbing repeats cannot be removed by *SfiI*.) The *MwoI* restriction site thus offers a convenient way to test whether the identified clones contain primer repeats or not. Since *MwoI* digests the long switch primer repeats to 20-39 bp fragments, a smear below 300 bp appears following *MwoI* digestion on electrophoretograms from clones that contain primer repeats (Fig. 3). The *MwoI* digestion thus reduces the chance of identifying primer repeat-containing clones as false-positives almost to zero.

CONFLICT OF INTEREST

Declared none.

ACKNOWLEDGEMENTS

This research was supported by a Genetic Basis of Cooperation Grant (GEBACO No. 28696).

FOOTNOTES

The work carried out in relation to the research presented in this article complied with the current laws of the country where it was performed.

SUPPORTING INFORMATION

Supporting information is available on the publishers Web site along with the published article.

REFERENCES

- [1] Gilboa E, Mitra SW, Goff S, Baltimore D. A detailed model of reverse transcription and tests of crucial aspects. *Cell* 1979; 18: 93-100.
- [2] Temin HM. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci USA* 1993; 90: 6900-3.
- [3] Cocquet J, Chong A, Zhang G, Veitia RA. Reverse transcriptase template switching and false alternative transcripts. *Genomics* 2006; 88: 127-31.
- [4] Chenchik A, Zhu YY, Diachenko L, Li R, Hill J, Siebert PD. In: Siebert, Larrick JW, Eds. *Gene cloning and analysis by RT-PCR, generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR*. P. D. Massachusetts, BioTechniques Books 1998; pp. 305-319.
- [5] Villanyi Z, Gyurjan I, Steger V, Orosz L. Plaque-based competitive hybridization. *J Biomol Screen* 2008; 13: 80-4.
- [6] Pinaud R. Genome of a songbird unveiled. *J Biol* 2010; 9: 19.
- [7] Clifton SW, Mitreva M. Strategies for undertaking expressed sequence tag (EST) projects. *Methods Mol Biol* 2009; 553: 13-32.
- [8] Stéger V, Molnár A, Borsy A, *et al.* Antler development and coupled osteoporosis in the skeleton of red deer *Cervus elaphus*: expression dynamics for regulatory and effector genes. *Mol Genet Genomics* 2010; 284: 273-87.

- [9] Carninci P, Kvam C, Kitamura A, *et al.* High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* 1996; 37: 327-36.
- [10] Maruyama K, Sugano S. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* 1994; 138: 171-4.

Received: December 07, 2011

Revised: January 17, 2012

Accepted: January 18, 2012

© Villanyi *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.