

Directionality bias in T/A cloning

Valeria Dountcheva¹, Athanasios Bubulya², Labib Rouhana^{1§}

¹Department of Biology, University of Massachusetts Boston, Boston, MA, USA

²Department of Biological Sciences, Wright State University, Dayton, OH, USA

§To whom correspondence should be addressed: labib.rouhana@umb.edu

Abstract

T/A cloning is a popular method for generating recombinant DNA plasmids. This method relies on single A:T nucleotide base pairs between PCR product ends and vector. Theoretically, the directionality of insert ligation with relation to the vector is random. However, we have continuously observed directionality bias using the pGEM-T Vector System for T/A cloning in a Course-based Undergraduate Research Experience (CURE). Cloning of over 400 inserts has shown directional bias higher than 74% (p -value < 0.0005) “sense” to the T7 promoter of the vector. Awareness of biased insertion directionality in our applications reduces time and cost in cloning and downstream analyses.

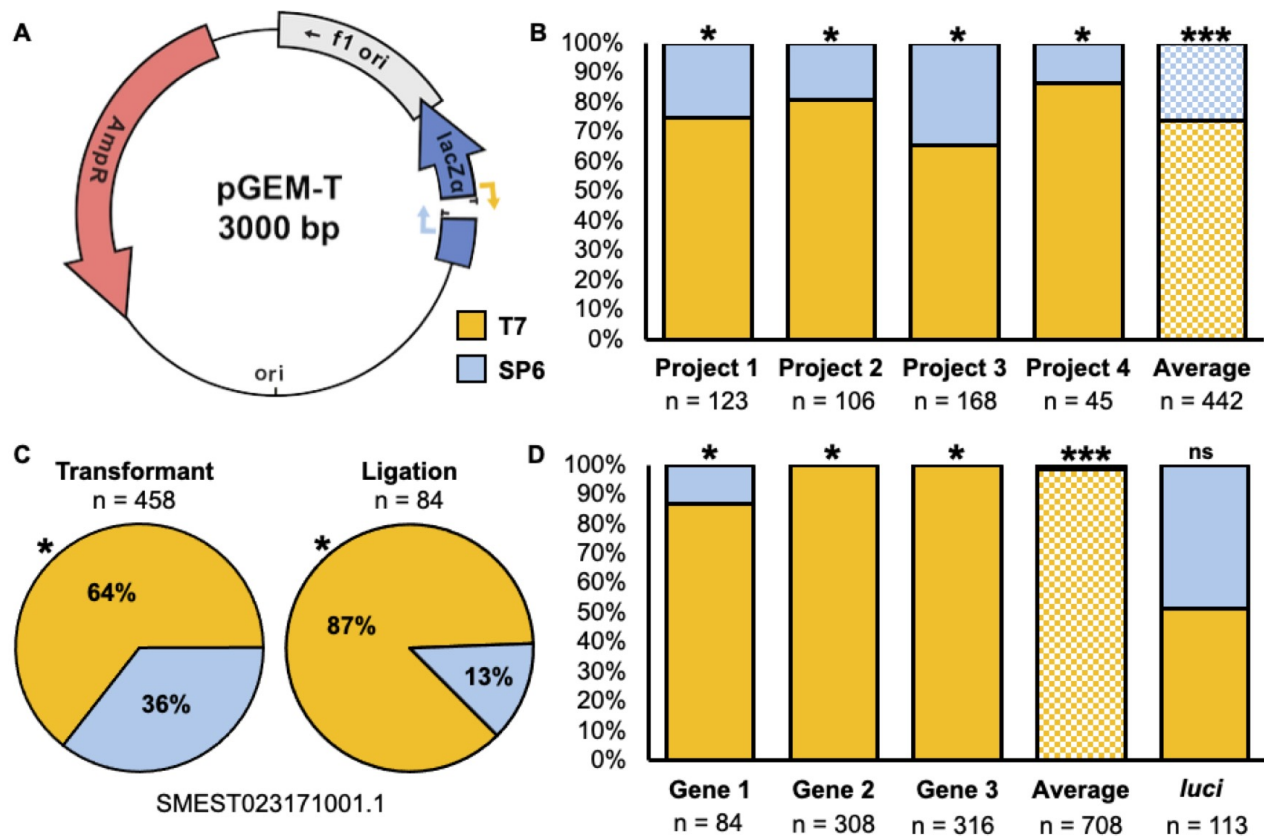


Figure 1. Orientation bias in insertion of DNA fragments during T/A cloning using the pGEM-T vector:

A) Plasmid map of pGEM-T illustrates the position of T-overhangs used in T/A cloning and flanking T7 and SP6 promoters. **B)** Percentage of cDNA insertion events in the sense orientation relative to the T7 promoter (yellow) and the SP6 promoter (blue) in constructs generated over the course of four CURE projects. The average across the projects is shown (checkered bar). **C)** Measured distribution of insert orientation from a cloning attempt of SMEST023171001.1 (*pumilio* homolog) cDNA fragment assessed after the transformant screening selection (left) or ligation (right) steps. **D)** Bar graphs display orientation distribution of inserts after the ligation step from the cDNA fragment in (C; Gene 1) and two additional genes (SMEST004045004.1, Gene 2; and SMEST012773001.1, Gene 3), as well as their average (checkered bar). Orientation of Control Insert DNA (*luciferase*) included in the pGEM-T Vector System after the ligation step is also shown. Asterisks show statistical significance according to the chi-squared test (* = $p < 0.005$) and unpaired Student's *t*-test (***) = $p < 0.0005$).

Description

Taq DNA Polymerase deposits a non-templated adenosine at the 3' end of DNA strands (Clark, 1988; Williams, 1989). This 3' end adenosine "overhang" is exploited for rapid and effective insertion of PCR products into plasmid vectors that carry single thymidine overhangs during "T/A cloning" (Holton and Graham, 1991; Kovalic et al., 1991; Marchuk et al., 1991; Mead et al., 1991). T/A cloning is favorable to blunt-end ligation due to the increased rate of success (Holton and Graham, 1991; Liu et al., 2018). In theory, the directionality of ligation between the insert and vector in T/A cloning is random. However, directional insertion is often desirable during cloning to facilitate downstream applications such as expression of proteins and generation of riboprobes in high throughput analyses. While directional cloning is a time-consuming technique reliant on thoughtful design and processing with multiple restriction enzymes, we have observed a strong directional bias using the pGEM-T Vector System for T/A cloning. In this report, we present evidence obtained from several years of cloning using the pGEM-T vector system in our Course-based Undergraduate Research Experience (CURE) that demonstrates statistically significant bias in the directionality of inserts.

A bias in directionality of partial cDNA sequence inserts in T/A cloning events was observed repeatedly over 6 years of activities in a CURE using the pGEM-T vector (Figure 1A). In each CURE project, partial cDNA amplicons corresponding to approximately 400-600 bps of planarian (*Schmidtea mediterranea*) sequence were inserted into pGEM-T and individual clones were validated through Sanger sequencing. Quantification of insert orientation with respect to the T7 and SP6 promoters flanking the insertion site, revealed that 75% of the insertions generated for Project 1 were positioned in the sense orientation relative to the T7 promoter (Figure 1B). Enrichment of sense orientation relative to the T7 promoter was also observed in subsequent projects as follows: 81% for the Project 2, 65% for Project 3, and 87% for Project 4. Distribution of directionality in each project was found to be statistically deviant from random insertion using the chi-squared test. On average, 74% of the inserts were positioned sense to the T7 promoter in the 442 constructs generated by over 100 separate individuals (Figure 1B). Statistical analysis using Student's *t*-test revealed that the overall deviation in directionality of insertion is significantly deviant from the 50% distribution expected for T/A cloning ($p < 0.0005$).

To understand how this bias might be generated, we looked at the structure of the pGEM-T vector. The insertion site for T/A cloning lies within the *LacZα* reporter gene used in blue/white colony screening in the pGEM-T system. *LacZα* is transcribed from a promoter upstream in the same orientation as the SP6 promoter, and in opposite orientation to the T7 promoter on the other side of the insertion site. White colonies are picked in cloning attempts because these represent successful insertion of amplicons within *LacZα* and therefore disruption of a functional *LacZα* product. We considered the possibility that insertions of small portions of ORF in the sense orientation relative to *LacZα* could result in a functional enzyme and therefore generate blue colonies representing false negative events. Therefore, we hypothesized that the bias observed in insert orientation may be an artifact of blue/white screening. To tease this apart, a fragment corresponding to a planarian *Pumilio* homolog (SMEST023171001.1) was amplified from cDNA and ligated into pGEM-T. A portion of the ligation reaction was transformed into *Escherichia coli* JM109 cells, while another portion of the ligation reaction was used directly as template to sequence across the insertion site between the T7 and SP6 promoters using third generation long read sequencing (PCR-EZ, Azenta). 64% of insertion events sequenced from white colonies were sense orientation relative to the T7 promoter in this experiment (458 reads, Figure 1C). Unexpectedly, sequences obtained directly from the ligation reaction also showed a strong bias in sense orientation relative to the T7 promoter (87% of 84 reads; Figure 1C). Ligation reactions were sequenced for two more planarian cDNA amplicons (SMEST004045004.1 and SMEST012773001.1), both of which revealed 100% insertion events in the sense orientation relative to the T7 promoter (624 reads, Figure 1C). Sequencing results from each of the three ligation reactions deviate significantly from the expected random directionality of insertion according to chi-squared test. Together the results from sequencing insertion events prior to transformation into *E. coli* showed a 98% average of insertion of planarian ORF gene fragments in the sense orientation relative to the T7 promoter ($p < 0.0005$, Figure 1D). Finally, to investigate whether insert directionality bias is observed when cloning DNA fragments obtained from other sources, we sequenced a ligation reaction using the control insert provided in the pGEM-T Vector System, which is a 542 bp fragment of the firefly *luciferase* gene. Ligations using the control insert showed no orientation bias, as would be expected from random insertion events (Figure 1D). These results indicate that something about the cDNA fragments selected, or the mechanism by which they are derived, is driving the bias in directionality observed during T/A cloning of planarian cDNA fragments.

Further investigation is needed to determine the mechanism driving the orientation bias observed at the ligation step during T/A cloning using the pGEM-T Vector System. Perhaps, coding sequences in our model, the planarian *S. mediterranea*, are composed of nucleotide combinations that cause preferential structural compatibility with one end of the pGEM-T vector cloning site. For a few inserts (e.g., *Smed-nanos* homolog, SMEST018169001.1), we saw strong orientation bias sense to the SP6 promoter, which is the opposite direction from the vast majority of our clones. Another possibility is that the tool used for designing primers in our cloning efforts (Primer3; Koressar and Remm, 2007; Untergass et al., 2012) integrates an unapparent difference in composition of the forward primer from the reverse primer (in our case, the forward primer is in the T7 orientation). Comparing 96 primer sets revealed that the forward primers had 10% more purines than their reverse counterparts, which is a statistically significant difference ($p = 1.94 \times 10^{-06}$). Other

parameters of primer composition, such as GC content, annealing strength, annealing temperature, and length were all comparable. Regardless of the mechanism, awareness of this bias has allowed us to streamline cloning of fragments utilized for generation of riboprobes in our mid-scale screens. It will be interesting to see if others observe the same phenomenon using sequences from other organisms or in different T/A cloning systems.

Methods

cDNA amplification and cloning

RNA was extracted from *Schmidtea mediterranea* sexually mature hermaphrodites using TRIzol™ Reagent (Thermo Fisher Scientific, Waltham, Massachusetts). cDNA was generated using the GoTaq 2-Step RT-qPCR System (Promega, Madison, Wisconsin) with 1 μg of total RNA as template and a mix of random and oligo(dT) primers following the manufacturer's instructions. Gene-specific primers targeting 400–600 bps of cDNA coding regions (whenever reference sequence made this length range possible) were designed using the Primer3 web tool (<https://primer3.ut.ee/>; (Koressaar and Remm, 2007; Untergasser et al., 2012). The following parameters were specified during primer design: T_m : 61–63 °C (optimal 62 °C); GC content: 35–65% (optimal 50%); primer size: 18–26 nts (optimal 22 nts); GC clamp: 1 nt. The remaining parameters were used at the default settings. cDNA amplicons were generated using GoTaq® Long PCR Master Mix (Promega, Madison, Wisconsin) with an initial denaturation at 95 °C for 2 minutes, followed by 40 cycles of denaturation at 95 °C for 30 seconds, annealing at 56 °C for 30 seconds, and extension at 72 °C for 1 minute, and a final extension at 72 °C for 5 min. Amplicons were then purified using the DNA Clean & Concentrator-5 kit as per manufacturer instructions (Zymo Research, Tustin, California) and eluted in 20 μL of nuclease-free water. A volume of 1 μL of clean amplicon was ligated into the pGEM-T vector using the Standard Reaction volume as per manufacturer's instructions (Promega, Madison, Wisconsin) and 3 μL of ligation reaction were transformed into 30 μL of JM109 cells. Plasmids were purified from liquid Luria Broth cultures supplemented with ampicillin [100 μg/mL] inoculated with a single colony (unless otherwise noted) using the GeneJET Plasmid Miniprep Kit as per the manufacturer (Thermo Fisher Scientific, Waltham, Massachusetts).

Sequencing of plasmid constructs for identity verification and assessment of directionality

The direction and identity of inserts in individual constructs were verified via Sanger sequencing of purified plasmid DNA using M13F, T7, and/or M13-40FOR primers (GENEWIZ from Azenta Life Sciences, Waltham, Massachusetts). To analyze proportions of insert directionality during ligation and within transformant populations, three planarian genes were amplified with their respective primers (SMEST023171001.1: 5'-GTTCACTGGCAGTTTGATTGG-3' and 5'-CATTTCTCTTGGCTTGATTGG-3'; SMEST004045004.1: 5'-GCGAACTAACGGGAACAAAC-3' and 5'-GGCCATGTGCTGGAATAATG-3'; and SMEST012773001.1: 5'-AGGTAACTGGTGATGCTACTG-3' and 5'-CCTCCTTTCTAGCTCTGTCTAATTC-3') and amplicons were processed as specified above. Then, 1 μL of either the ligation reaction or plasmids purified from a mixture of all the white colonies from the transformation were used as template for PCR amplification using primers flanking the multiple cloning site of pGEM-T (5'-GCGGAATACCTCACTAAGTATACGACTCACTATAGG-3' and 5' CGCGCGCTAATACGACTCATATGATTAGGTGACACTATAG-3'). Amplicons were purified using DNA Clean & Concentrator-5 kit (Zymo Research, Tustin, California), eluted in 15 μL of nuclease-free water, and sent for PCR-EZ Sequencing (GENEWIZ, Azenta Life Sciences, Waltham, Massachusetts). FASTQ files from PCR-EZ sequencing were mapped to their respective predicted insertion sequence if inserted sense to the T7 promoter using minimap2 and the command “minimap2 -ax splice -k14 --MD /path/to/reference.fasta path/to/sequenced-results.fastq -o path/to/aligned-sequence.sam” (Li, 2018). The orientation of reads was then quantified using samtools to identify reads sense to T7 promoter with the command “samtools view -c -F 16” and reads sense to the SP6 promoter “samtools view -c -f 16” (Danecek et al., 2021).

Statistical Analysis

To test for statistical significance in observed distributions of insert orientation for pGEM-T constructs generated in the CUREs, the fraction of constructs with inserts in the sense orientation relative to the T7 promoter were calculated and compared to those generated with sense orientation relative to the SP6 promoter (Figure 1B, 1D). Standard unpaired, two-tailed, Student's *t*-tests were used to measure statistical significance between the fraction of constructs obtained across the four projects (Figure 1B) and three genes (Figure 1D) analyzed. Distribution of orientation for individual projects and cloning events following ligation vs. bacterial selection were tested for statistical significance using chi-squared tests (Figure 1B, 1C, 1D).

Reagents

Reagent	Source	Cat No/Ref
DNA Clean & Concentrator-5	Zymo Research	D4004
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific	K0503
GoTaq 2-Step RT-qPCR System	Promega	A6010
GoTaq® Long PCR Master Mix	Promega	M4021
pGEM®-T Vector Systems	Promega	A3600
TRIzol™ Reagent	Thermo Fisher Scientific	15596026
<i>Schmidtea mediterranea</i> , hermaphrodite strain	Phillip Newmark Laboratory, Univ. of Illinois at Urbana-Champaign (currently at the Univ. of Wisconsin–Madison)/Howard Hughes Medical Institute	(Zayas et al., 2005)

Acknowledgements: The authors would like to thank the hundreds of undergraduate students who participated in the Molecular Biology laboratory CUREs at UMass Boston (BIOL 370) and Wright State University (BIO 3140).

References

- Clark JM. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* 16(20): 9677-86. PubMed ID: [2460825](#)
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al., Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10(2): 10.1093/gigascience/giab008. PubMed ID: [33590861](#)
- Holton TA, Graham MW. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Res* 19(5): 1156. PubMed ID: [2020554](#)
- Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23: 1289-1291. DOI: [10.1093/bioinformatics/btm091](#)
- Kovalic D, Kwak JH, Weisblum B. 1991. General method for direct cloning of DNA fragments generated by the polymerase chain reaction. *Nucleic Acids Research* 19: 4560-4560. DOI: [10.1093/nar/19.16.4560](#)
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094-3100. DOI: [10.1093/bioinformatics/bty191](#)
- Liu Q, Dang HJ, Wu YH, Li M, Chen YH, Niu XL, Li KM, Luo LJ. 2018. pXST, a novel vector for TA cloning and blunt-end cloning. *BMC Biotechnology* 18: 10.1186/s12896-018-0456-8. DOI: [10.1186/s12896-018-0456-8](#)
- Marchuk D, Drumm M, Saulino A, Collins FS. 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research* 19: 1154-1154. DOI: [10.1093/nar/19.5.1154](#)
- Mead DA, Pey NK, Herrstadt C, Marcil RA, Smith LM. 1991. A Universal Method for the Direct Cloning of PCR Amplified Nucleic Acid. *Nature Biotechnology* 9: 657-663. DOI: [10.1038/nbt0791-657](#)
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40: e115-e115. DOI: [10.1093/nar/gks596](#)
- Williams JF. 1989. Optimization strategies for the polymerase chain reaction. *Biotechniques* 7(7): 762-9.

Zayas RM, Hernández A, Habermann B, Wang Y, Stary JM, Newmark PA. 2005. The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: Analysis of ESTs from the hermaphroditic strain. Proceedings of the National Academy of Sciences 102: 18491-18496. DOI: [10.1073/pnas.0509507102](https://doi.org/10.1073/pnas.0509507102)

Funding: V.D. was supported in part by Graduate Teaching Assistantships from the University of Massachusetts Boston. VD was additionally supported by a College of Science and Mathematics Dean's Doctoral Research Fellowship with funding from Oracle, project ID R20000000025727.

Supported by Eunice Kennedy Shriver National Institute of Child Health and Human Development (United States) R15HD082754 to Labib Rouhana.

Conflicts of Interest: The authors declare that there are no conflicts of interest present.

Author Contributions: Valeria Dountcheva: formal analysis, data curation, investigation, writing - original draft. Athanasios Bubulya: writing - review editing, investigation. Labib Rouhana: conceptualization, funding acquisition, writing - review editing, investigation.

Reviewed By: Anonymous

History: Received February 11, 2026 **Revision Received** April 5, 2026 **Accepted** April 2, 2026 **Published Online** April 6, 2026 **Indexed** April 20, 2026

Copyright: © 2026 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Dountcheva V, Bubulya A, Rouhana L. 2026. Directionality bias in T/A cloning. microPublication Biology. [10.17912/micropub.biology.002065](https://doi.org/10.17912/micropub.biology.002065)