



HAL
open science

Insertion site preference of Mu, Tn5, and Tn7 transposons.

Brian Green, Christiane Bouchier, Cécile Fairhead, Nancy Craig, Brendan
Cormack

► **To cite this version:**

Brian Green, Christiane Bouchier, Cécile Fairhead, Nancy Craig, Brendan Cormack. Insertion site preference of Mu, Tn5, and Tn7 transposons.. *Mobile DNA*, 2012, 3 (1), pp.3. 10.1186/1759-8753-3-3 . pasteur-00675691

HAL Id: pasteur-00675691

<https://pasteur.hal.science/pasteur-00675691>

Submitted on 1 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SHORT REPORT

Open Access

Insertion site preference of Mu, Tn5, and Tn7 transposons

Brian Green¹, Christiane Bouchier², Cécile Fairhead³, Nancy L Craig⁴ and Brendan P Cormack^{1*}

Abstract

Background: Transposons, segments of DNA that can mobilize to other locations in a genome, are often used for insertion mutagenesis or to generate priming sites for sequencing of large DNA molecules. For both of these uses, a transposon with minimal insertion bias is desired to allow complete coverage with minimal oversampling.

Findings: Three transposons, Mu, Tn5, and Tn7, were used to generate insertions in the same set of fosmids containing *Candida glabrata* genomic DNA. Tn7 demonstrates markedly less insertion bias than either Mu or Tn5, with both Mu and Tn5 biased toward sequences containing guanosine (G) and cytidine (C). This preference of Mu and Tn5 yields less uniform spacing of insertions than for Tn7, in the adenosine (A) and thymidine (T) rich genome of *C. glabrata* (39% GC).

Conclusions: In light of its more uniform distribution of insertions, Tn7 should be considered for applications in which insertion bias is deleterious.

Keywords: Tn7, Mu, Tn5, Mutagenesis, Insertion site, DNA transposon, Mobile element

Background

Transposons, mobile DNA elements that can integrate into target DNA molecules, are useful for insertional mutagenesis, gene tagging, gene transfer, and sequencing applications. A major class of transposable elements used for genome engineering is DNA 'cut and paste' transposons. The transposases for DNA transposons cut the transposon away from the donor DNA by a variety of mechanisms and the excised transposon integrates into a new target site by joining of its 3'OH termini to staggered positions on the top and bottom DNA strands of the target. This staggered joining results in a target site duplication of a defined number of base pairs, which can be used to map precisely the site of integration for the transposon [1].

In most of the applications of transposons to molecular biology, it is important that the transposon insert into target DNA with little to no sequence bias. Limited sequence bias will lead to more complete coverage of a region for a given number of insertion events. However,

most transposons have been shown to exhibit some preference for certain sequences or sequence features [1]. Clearly, insertion site bias may be a confounding factor for large scale transposon mutagenesis projects.

A number of manuscripts reporting insertion motifs for various transposons have been published, but the target DNA, transposition protocol and environment (*in vitro* versus *in vivo*) vary widely, making direct comparisons difficult. For example, individual genes [2], *Escherichia coli* genomic DNA [3], and *Saccharomyces cerevisiae* genomic DNA [4] have been used. In this publication, three transposon systems were evaluated using the same target DNA *in vitro*: Mu, Tn5, and a modified Tn7 [5]. Previous work had identified a CPy (G/C)PuG or similar motif for Mu [6-8], a GPyPyPy(A/T)PuPuPuC motif for Tn5 [9,10] and negligible bias for the modified Tn7 [11]. Since previous publications all used different target DNA, and because our DNA of interest (*C. glabrata* genomic DNA) has a moderately high A/T content (61%, [12]), specificity and distribution of insertion sites for all three transposons was assessed on the same target DNAs.

* Correspondence: bcormack@jhmi.edu

¹Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Hunterian 617, 725 North Wolfe Street, Baltimore, MD 21205-2185, USA

Full list of author information is available at the end of the article

Methods

BACs containing *C. glabrata* genomic DNA were prepared as follows. First, the vector plasmid pBAC-NAT was constructed in two steps. pCR2.1-NAT was constructed by amplifying the NAT cassette using primers ON-5'NAT (CCGCTGCTAGGCGCGCCGTGGAAGTTCTATACTTTCTAGAGAATAGGAACTTCGATCCCCCCATAAAGCACGTGATAGCTTC) and ON-3'NAT (GCAGGGATGCGGCCGCTGACGAAGTTCC-TATTCTCTAGAAAGTATAGGAACTTCAGCTTGA-TATCGAATTCGCAAATTAAGCC) from pCaNAT1 (a gift of Julia Koehler) and cloning that into pCR2.1 using the TA/TOPO kit (Life Technologies, Carlsbad, CA, USA 92008) The NAT cassette was amplified using primers ON3601 (AGTCGCGGCCGCGTTTAAACGGCGCCCCGCTGCTAGGCGCGCCGTG) and ON3602 (AGTCGGCCCCGGCGGCCACGCGTTGACCCGCGGCAGGGATACGGCCGCTGAC), cloned into pCR2.1, sequence verified, and a NotI/SfiI fragment of that was inserted into pBAC [13] cut with NotI/SfiI to yield pBAC-NAT (pB1895).

Next, genomic DNA was inserted into pBAC-NAT. The four plasmids into which transposons were mobilized contain genomic DNA from *C. glabrata* from the indicated ORF to the telomere. The genomic DNA began at *CAGLOA00187g* from the strain BG2 [14] for pB1907 (24,252 bp and 34% GC insert), from *CAGLOC00297g* and strain BG2 for pB1908 (31,757 bp and 34% GC insert), from *CAGLOC05599g* and strain BG2 for pB1909 (25,125 bp and 34% GC insert), and from *CAGLOC05599g* and strain CBS138 [12] for pB1910 (19,423 bp and 31% GC insert). Although pB1909 and pB1910 contain the region from the same gene to the telomere from different strains, they are only homologous for the centromeric (rightmost in figures) approximately 8 kb, after which they diverge completely (data not shown).

Mu transposition reactions were carried out per the manufacturer's recommendations using the Finnzyme Template Generation System Kit (Thermo Fisher Scientific, Waltham, MA, USA 02454), with pB1909 and pB1910 as target DNA sequences. Tn5 transposition reactions were carried out per the manufacturer's recommendations using the Ez-Tn5 kit (Epicentre, Madison, WI, USA 53713) with pB1907, pB1908, pB1909, and pB1910 as targeting sequences. Tn7 reactions were carried out as published, [13] with pB1907, pB1908, pB1909, and pB1910 as targeting sequences.

All sequencing was done using the ABI BigDye Terminator Kit v1.1 (Life Technologies, Carlsbad, CA, USA 92008). The primers used for sequencing the Mu transposon containing clones were SeqE (CGACACACTC-CAATCTTTCC) and SeqW (GGTGGCTGGAGT

TAGACATC). The primers used for sequencing the Tn5 transposon containing clones were KAN-2 FP-1 (ACCTACAACAAAGCTCTCATCAACC) and KAN-2 RP-1 (GCAATGTAACATCAGAGATTTTGAG). The primers used for sequencing the Tn7 transposon containing clones were ON661 (ATAATCCTTAAAACTCATTTCACCCCTCCCAG) and ON662 (GACTT-TATTGTCATAGTTTAGATCTAT TTTGTTTCAG).

BLogo sequence logos were generated using the web form at <http://www.bioinformatics.org/blogo/cgi-bin/Blogo/Blogoform.pl>[15] as type 2 logos with coloring for symbols with $P < 0.001$ (Fisher's exact test) and base representation calculated from the fosmid sequences into which the various transposons were integrated. The background frequencies of A, C, G, and T used for the BLogo sequence logos are given in the figure legends.

Results

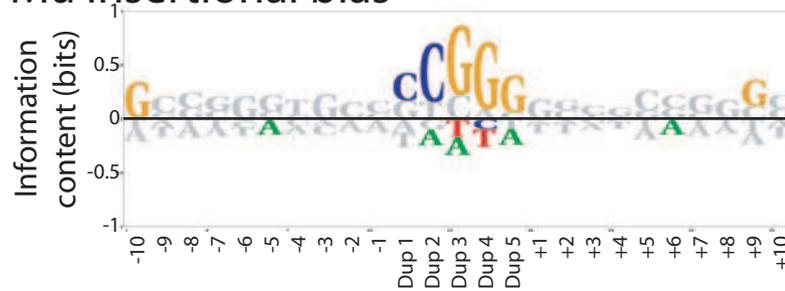
Fosmids containing subtelomeric and telomeric genomic DNA from *C. glabrata* were used as targets for transposon insertion *in vitro* for the transposons Mu, Tn5, and Tn7. Following transformation to select insertions, the resulting clones were individually selected and sequenced from both ends of the transposon. The two reads for each clone were merged to yield the sequence of the ten nucleotides upstream of the transposon mediated duplication, the duplication, and ten nucleotides downstream of the duplication. Table 1 shows the number of these insertion events that could be mapped to locations within the target fosmid.

All insertion events for a given transposon were used to generate BLogo sequence logo plots of position specific sequence bias, with positions colored if significant at $P < 0.001$ (Figure 1). BLogo sequence plots are a position specific log based measure of the overrepresentation (above the line) or underrepresentation (below the line) of each base at each position around the insertion sites. The plot for Mu insertions (Figure 1A) shows a strong bias for a CGG motif central to the 5 bp duplicated region, which has been previously reported [8]. The Tn5 insertions are also biased (Figure 1B), with a strong preference for G at the first bp of the duplication and a general bias for G and C across the analyzed

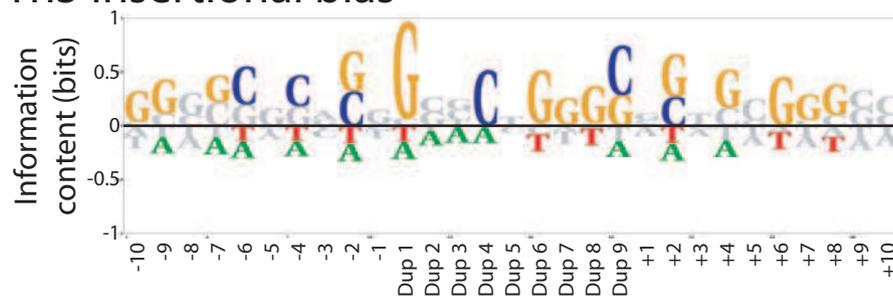
Table 1 Number of insertions mapped

Fosmid	Mu	Tn5	Tn7
pB1907	n/a	97	121
pB1908	n/a	58	113
pB1909	139	73	139
pB1910	113	48	106
Total	252	276	479

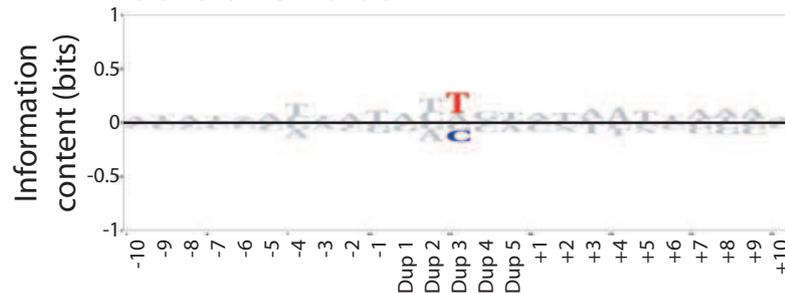
A) Mu insertional bias



B) Tn5 insertional bias



C) Tn7 insertional bias



D) Mu normalized insertional bias

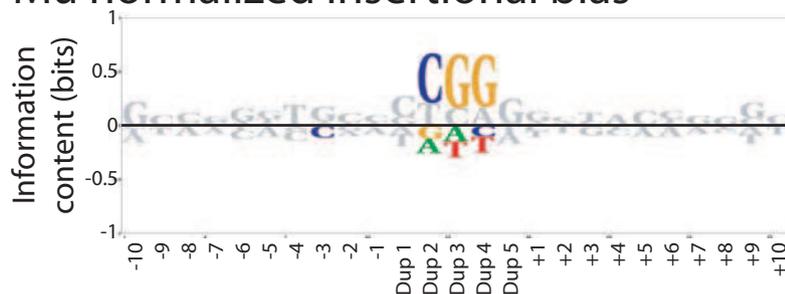


Figure 1 Insertion biases of Mu, Tn5, and Tn7. BLogo sequence logos are shown for all insertion sites for Mu (A), Tn5 (B), and Tn7 (C) transposons into *Candida glabrata* subtelomeric DNA. Ten bases on either side of the transposon mediated duplication of five (Mu and Tn7) or nine (Tn5) bases were used to generate the sequence logo, and bases significant at $P < 0.001$ are shown in color. Mu insertion events were used to create a BLogo sequence logo with background base frequencies obtained from selecting all 25 mers within the target fosmids containing a central CGG (D). The background frequencies of A, C, G, and T used for the BLogo sequence logos are respectively: 0.32, 0.19, 0.18, 0.31 (A,B,C); 0.275, 0.217, 0.224, 0.284 (D).

region. This motif is similar, but not identical, to that previously reported for Tn5 mutagenesis [9]. In contrast, the Tn7 has only a very weak bias, toward a T in the middle position of the duplication site (Figure 1C).

The G/C content of the fosmids mutagenized is not uniform across their length (see below), so some of the apparent bias in the flanking base pairs might be simply due to a sampling bias due to a strongly biased central core. In fact, if the base percentages used in BLogo are calculated from all 25mers in the fosmids containing a central CGG core, no other positions show significant bias for Mu transposition (Figure 1D). The results from a similar analysis for Tn5 were ambiguous, but since many of the Tn5 insertions were in the *C. glabrata* telomeric repeats (which are very G/C rich), this could explain the observed flanking bias.

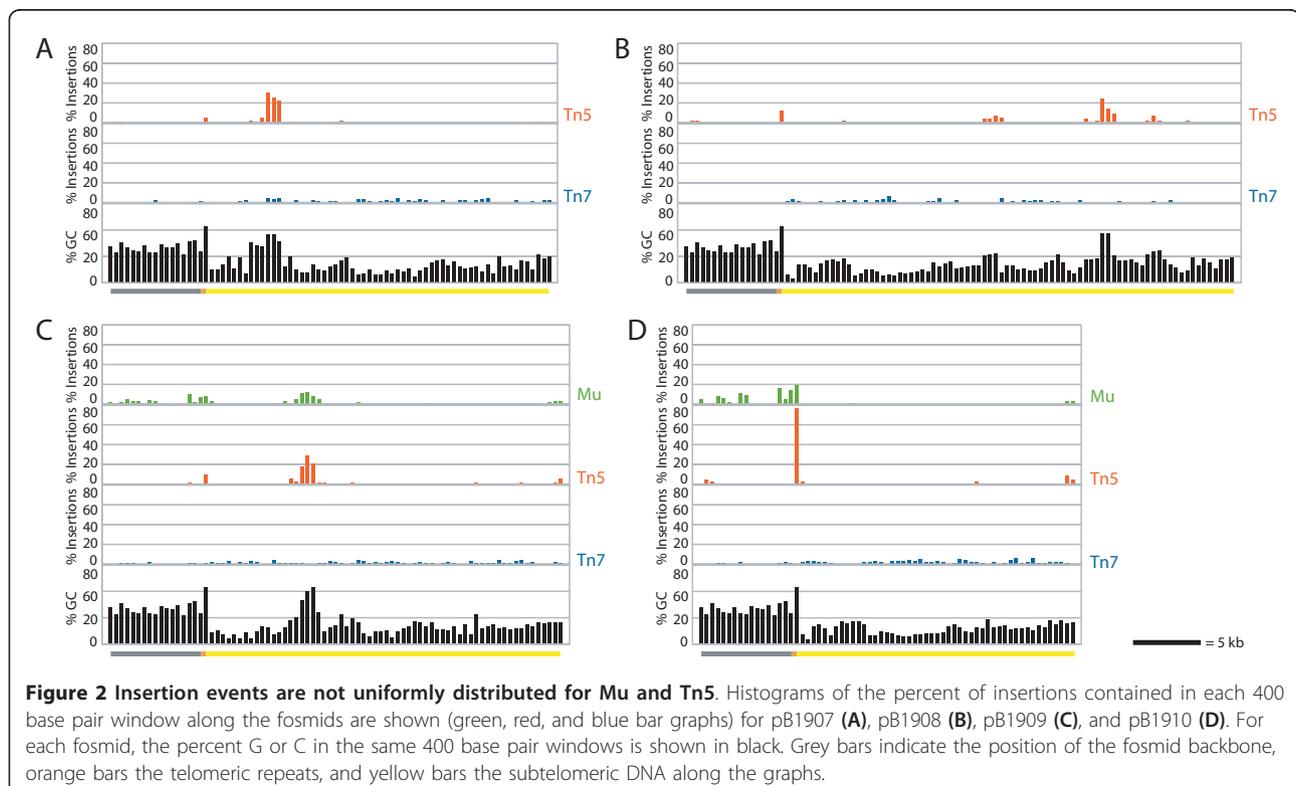
Depending on the nucleotide composition of the target fragment, insertion site bias would be expected to result in non-random spacing of insertions in targets with variable G/C content. Histograms showing the percent of insertions in each 400 bp window spanning the fosmid demonstrate that Mu and Tn5 had strongly clustered insertions (Figure 2). For Tn5, 10% of the 400 bp windows accounted for 92.4% of the insertions; for Mu, it was 72.6%. In contrast, the spacing of Tn7 insertions is much more uniform (Figure 2), with the top 10% of 400 bp windows containing 32.8% of the insertions. The

sequence motifs for both Mu and Tn5 are G/C rich; analysis of the percent G/C for each 400 bp window shows that the strongest peaks in the frequency of Mu and Tn5 insertions occur in regions of relatively elevated G/C content (Figure 2).

As part of a larger sequencing effort, and due to the minimal sequence bias discussed above, Tn7 was then used to mutagenize 49 fosmids containing *C. glabrata* subtelomeric genomic DNA. A total of 6,700 insertion events were used to generate a BLogo sequence motif (Figure 3). In contrast to the smaller set of Tn7 insertion events discussed above, and due to the larger number of sequences used, the subtle biases exhibited by Tn7 were significant at $P < 0.001$. Although the relative bias is very weak, Tn7 does exhibit a bias toward a central TTG core. There is also a slight palindromic bias in the flanking regions; positions plus 1 to plus 6 are overrepresented for ATGATT and positions minus 6 to minus 1 for AATCAT. Overall, the slight bias seen for Tn7 is for A/T rich sequences and is markedly less pronounced than the G/C biases of Mu and Tn5.

Conclusions

Uniform distribution of transposon insertion is important for their use in many molecular biological applications. In the analysis here, Tn7 demonstrates a far more



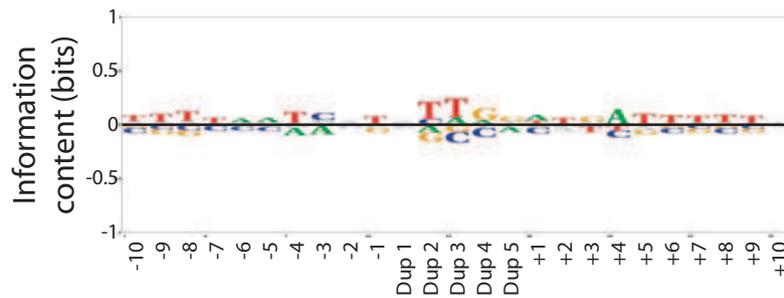


Figure 3 Tn7 does exhibit a very modest bias toward A/T rich sequences. A BLogo sequence logo was generated as in Figure 1, using 6,700 Tn7 insertion events from 49 fosmids. The background frequencies of A, C, G, and T used for the BLogo sequence logos are 0.317, 0.196, 0.186, 0.301, respectively.

random insertion profile than either Mu or Tn5. For Tn5 and Mu, the consensus sequences derived are consistent with extensive published analyses [5-10]. Transposon insertion site preference is complex. For example, previous studies of Mu insertion have shown that particular dinucleotides, or base steps, contribute to site preference; predicted structural features of DNA also may play a role [6-8]. However, even these comprehensive analyses of insertion site preference derived consensus sites that are both G/C rich and consistent with those we derived here. We suggest that the preference for G/C rich sequences exhibited by Mu and Tn5 is of particular importance for investigators working with A/T rich genomes such as *S. cerevisiae* (approximately 62% A/T) and many other fungi, *Caenorhabditis elegans* (approximately 65% A/T), and human (approximately 60% A/T) [16]. Tn7 has less target bias, with a weak preference for A/T rich sequences. For the large genomic inserts analyzed here, this resulted in a broad distribution of insertion sites, with a bias away from the plasmid backbone (which is more G/C rich relative to the cloned genomic DNA). The minimal sequence bias exhibited by Tn7 suggests that Tn7 should be considered for generating random insertions in cloned A/T-rich genomes.

Abbreviations

bp: base pair.

Acknowledgements

We thank Julia Koehler for the gift of pCaNAT. This work was supported by grants 1F32AI080094 (BMG), 2R01AI046223 (BPC), and R01GM076425 (NLC). NLC is an Investigator of the Howard Hughes Medical Institute.

Author details

¹Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Hunterian 617, 725 North Wolfe Street, Baltimore, MD 21205-2185, USA. ²Institut Pasteur, Plate-forme Génomique, Département Génomiques et Génétique, Rue du Dr. Roux, F-75015 Paris, France. ³Institut de Génétique et Microbiologie, Université Paris Sud 11, CNRS UMR8621, 15 rue Georges Clémenceau, Orsay F-91405, France. ⁴Department of Molecular Biology and Genetics, The Howard Hughes Medical Institute, 725 North Wolfe

Street Johns Hopkins University School of Medicine, Baltimore, MD 21205-2185, USA.

Authors' contributions

BG carried out all cloning, Tn7 mutagenesis and all sequence analysis, and wrote the paper. CF and CB carried out the Mu and Tn5 mutagenesis and sequencing. NC and BC conceived of the idea and supervised the project. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 5 October 2011 Accepted: 7 February 2012

Published: 7 February 2012

References

1. Craig NL: *Mobile DNA II* Washington, DC: ASM Press; 2002.
2. Franssen M, Vastiau I, Brees C, Brys V, Mannaerts GP, Van Veldhoven PP: Analysis of human Pex19p's domain structure by pentapeptide scanning mutagenesis. *J Mol Biol* 2005, **346**:1275-1286.
3. Puttamreddy S, Cornick NA, Minion FC: Genome-wide transposon mutagenesis reveals a role for pO157 genes in biofilm development in *Escherichia coli* O157:H7 EDL933. *Infect Immun* 2010, **78**:2377-2384.
4. Xu T, Bharucha N, Kumar A: Genome-wide transposon mutagenesis in *Saccharomyces cerevisiae* and *Candida albicans*. *Methods Mol Biol* 2011, **765**:207-224.
5. Biery MC, Stewart FJ, Stellwagen AE, Raleigh EA, Craig NL: A simple in vitro Tn7-based transposition system with low target site selectivity for genome and gene analysis. *Nucleic Acids Res* 2000, **28**:1067-1077.
6. Butterfield YS, Marra MA, Asano JK, Chan SY, Guin R, Krzywinski MI, Lee SS, MacDonald KW, Mathewson CA, Olson TE, Pandoh PK, Prabhu AL, Schnerch A, Skalska U, Smailus DE, Stott JM, Tsai MI, Yang GS, Zuyderduyn SD, Schein JE, Jones SJ: An efficient strategy for large-scale high-throughput transposon-mediated sequencing of cDNA clones. *Nucleic Acids Res* 2002, **30**:2460-2468.
7. Haapa-Paananen S, Rita H, Savilahti H: DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. *J Biol Chem* 2002, **277**:2843-2851.
8. Manna D, Deng S, Breier AM, Higgins NP: Bacteriophage Mu targets the trinucleotide sequence CGG. *J Bacteriol* 2005, **187**:3586-3588.
9. Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS: Tn5/IS50 target recognition. *Proc Natl Acad Sci USA* 1998, **95**:10716-10721.
10. Shevchenko Y, Bouffard GG, Butterfield YS, Blakesley RW, Hartley JL, Young AC, Marra MA, Jones SJ, Touchman JW, Green ED: Systematic sequencing of cDNA clones using the transposon Tn5. *Nucleic Acids Res* 2002, **30**:2469-2477.
11. Seringhaus M, Kumar A, Hartigan J, Snyder M, Gerstein M: Genomic analysis of insertion behavior and target specificity of mini-Tn7 and Tn3 transposons in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2006, **34**:e57.
12. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, Goffard N, Frangeul L, Aigle M,

Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisramé A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, *et al*: **Genome evolution in yeasts**. *Nature* 2004, **430**:35-44.

13. Castano I, Kaur R, Pan S, Cregg R, Penas Ade L, Guo N, Biery MC, Craig NL, Cormack BP: **Tn7-based genome-wide random insertional mutagenesis of *Candida glabrata***. *Genome Res* 2003, **13**:905-915.
14. Cormack BP, Falkow S: **Efficient homologous and illegitimate recombination in the opportunistic yeast pathogen *Candida glabrata***. *Genetics* 1999, **151**:979-987.
15. Li W, Yang B, Liang S, Wang Y, Whiteley C, Cao Y, Wang X: **BLogo: a tool for visualization of bias in biological sequences**. *Bioinformatics* 2008, **24**:2254-2255.
16. Herold J, Kurtz S, Giegerich R: **Efficient computation of absent words in genomic sequences**. *BMC Bioinformatics* 2008, **9**:167.

doi:10.1186/1759-8753-3-3

Cite this article as: Green *et al.*: Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mobile DNA* 2012 3:3.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

