Genet. Res., Camb. (2003), **82**, *pp.* 85–88. *With* 1 *figure.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672303006359 Printed in the United Kingdom

Target sequences of Tc1, Tc3 and Tc5 transposons of *Caenorhabditis elegans*

VIVIAN PRÉCLIN, EDWIGE MARTIN AND LAURENT SÉGALAT*
CGMC, CNRS-UMR 5534, Université Lyon 1, 43 bld du 11 Novembre, 69622 Villeurbanne, France

(*Received 22 April 2003 and in revised form 24 June 2003*)

Summary

We report here the consensus target sequence of transposons Tc1, Tc3 and Tc5 of *Caenorhabditis elegans*. These sequences were obtained by molecular analysis of 1008 random new insertions which have not been exposed to natural selection. This analysis reveals consensus target sites slightly different from those previously reported, and confirms that the *mariner* elements Tc1 and Tc3 insert in sites which are not preferentially palindromic.

1. Introduction

Tc elements from the *mariner* family have been used extensively for both forward and reverse genetics purposes in *Caenorhabditis elegans* (Anderson, 1995; Plasterk & van Luenen, 1997). We previously reported the molecular characterization of more than a thousand new random insertions of transposable elements of *C. elegans* (Martin *et al.*, 2002). These insertions were obtained by propagating independent strains in the widely used *mut-7* genetic background, which derepresses transposition in the germ-line of *C. elegans* (Ketting *et al.*, 1999). New insertions of the *mariner* family elements Tc1, Tc3 (Plasterk *et al.*, 1999) and of the distantly related Tc5 element (Collins & Anderson, 1994) were localized at the nucleotide level (Martin *et al.*, 2002). The analysis of their chromosomal distribution shows that they are scattered in the whole genome of *C. elegans* (Martin *et al.*, 2002). These new insertions have been submitted to no (or very limited) natural selection since they were analysed 1 to 10 generations after transposition occurred (Martin *et al.*, 2002). Thus, they constitute an unequalled resource to define the consensus target sequence of each transposon.

2. Materials and methods

The insertion's flanking sequences, which constitute the raw data of this study, were obtained from a collection of 1088 new insertions of transposons Tc1, Tc3 and Tc5 of *C. elegans* (Martin *et al.*, 2002). Eighty insertions for which the insertion site was ambiguous were not included in the current study. A total of 1008 sequences (oriented 5′ to 3′ relative to the transposon orientation) were used for the analysis. These sequences can be downloaded from the website http://www.cgmc.univ-lyon1.fr/cgmc_info_celeganstp.php.

Base frequencies were visualized by the Seqlogo program (Schneider & Stephens, 1990). Frequency analysis was done by the chi-square test.

* Corresponding author. Tel: +33 4 72432951. Fax: +33 4 72440555. e-mail: segalat@cgmc.univ-lyon1.fr

3. Results and discussion

(i) *Consensus sequence for Tc1*

Like most transposable elements of the *mariner* family, Tc1 is known to have an absolute TA target site (Anderson, 1995; Plasterk *et al.*, 1999). Analysis of consensus target sequences of Tc1 around the TA site has previously been reported, (1) on data obtained from Tc1-induced mutants (Mori *et al.*, 1988; Eide & Anderson, 1988), (2) from PCR-detected insertions within a single gene (van Luenen & Plasterk, 1994), and (3) from analysis of Tc1 flanking regions in strains carrying high number copies of transposons (Korswagen *et al.*, 1996). Our results, based on the alignment of 588 sequences, show the following consensus sequence: AYA**TA**TRT (Fig. 1, top) (R = G or A; Y = C or T). In this consensus, the A at position −3 and the T at position +3 are the more conserved (75% and 73% respectively). Although the consensus

Table 1. *Absence of bias toward a palindromic sequence for Tc1 insertion sites*

	+1		+2		+3	
Position in sequence	T	C	A	G	T	A
Percentage (and number) of occurrence in all sequences	51% ($n=301$)	22% ($n=132$)	32% ($n=192$)	30% ($n=178$)	73% ($n=430$)	15% ($n=91$)
Percentage of occurrence in sequences where						
A is the mirror base[a]	54% ($n=178$)				75% ($n=335$)	
T is the mirror base			32% ($n=96$)			15% ($n=12$)
G is the mirror base		24% ($n=27$)				
C is the mirror base				31% ($n=43$)		

[a] The mirror base for positions +1, +2, +3 is the base at position −1, −2, −3 respectively.

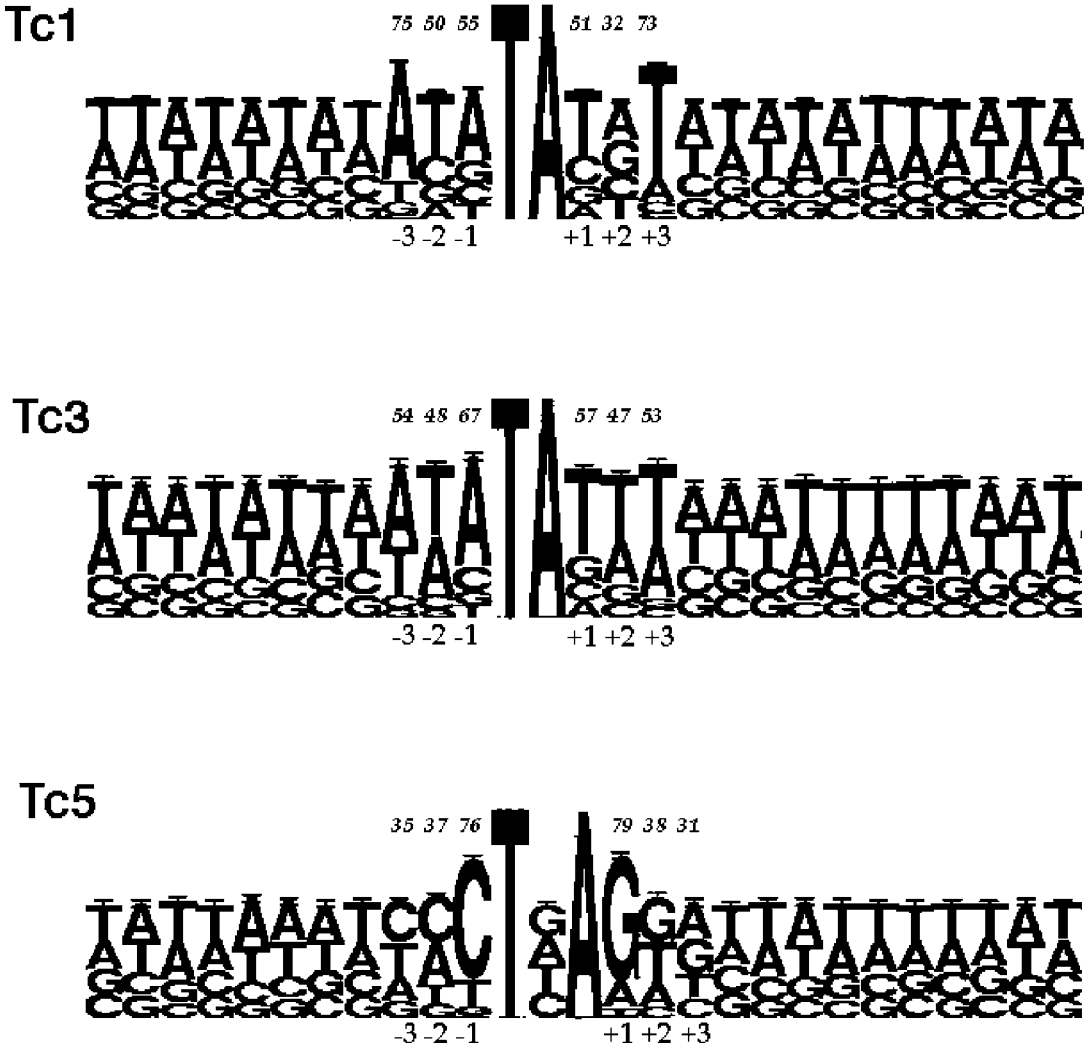

Fig. 1. Consensus target sequence of *C. elegans* transposons Tc1 (top), Tc3 (middle) and Tc5 (bottom) drawn by the Seqlogo program. For each position, the size of the letter represents the frequency of occurrence. Numbers below the sequence represent the base position mentioned in the text. Numbers above the sequence represent the percentage occurrence of the preferred base at positions −3 to +3. The analysis is based on 588, 223 and 197 new independent and distinct insertions, respectively.

https://doi.org/10.1017/S0016672303006359 Published online by Cambridge University Press

sequence might suggest that Tc1 prefers palindromic sites, detailed analysis shows that this is not the case. First, only 38 of the 588 insertions (6%) display a 6-base palindromic sequence around the central TA. Second, careful analysis of the bases at mirror positions indicate that there is no preference for the complementary base. For instance, whereas the occurrence of an A at position +2 is 32% in all Tc1 insertions, it is also 32% among the insertions carrying a T at position −2, indicating that the complementary base is not preferred (Table 1 recapitulates these frequencies for positions 1 to 3). This deviation from the palindromic sequence was already visible in the sequences reported in a previous study, albeit on a smaller sample (van Luenen & Plasterk, 1994). The consensus obtained from our Tc1 insertions differs from the one published by Korswagen *et al*. (1996). In this study, performed on sequences obtained from strains carrying many copies of Tc1, the consensus CAYATATRTG was reported (Korswagen *et al*., 1996). We do not find any base preference at position 4 and following, nor at position −4 and upstream (Korswagen *et al*., 1996). Re-examination of the data presented in this paper suggests that the bias toward a C/G at position −4/+4 is not statistically sustained.

(ii) *Consensus sequence for Tc3*

Like Tc1, the Tc3 transposon is known to insert at TA sites (Collins *et al*., 1989; Anderson, 1995). A consensus target sequence for this transposon has been proposed based on 166 insertions occurring at 29 sites within a 1 kb region of the *gpa-2* gene (van Luenen & Plasterk, 1994). The consensus sequence was TAA/C**TA**TTT/A. Our results, based on 223 independent and distinct Tc3 insertions, reveal the consensus ATA**TA**TTT (Fig. 1, middle). No base preference is seen outside of the consensus. This result confirms that the Tc3 preferred target site is A/T-rich and non-palindromic. The discrepancy with the results obtained on *gpa-2* insertions might be due to the sample size. Bases upstream and downstream of the 8-base consensus sequence ATA**TA**TTT (positions −13−4 and +4+13) do not show any particular motif. They are more A/T-rich (65%) than the average genome composition, as expected around an 8 basepair A/T stretch.

(iii) *Consensus sequence for Tc5*

Tc5 was described as having a CTNAG target site, with T and A present at all times (Collins & Anderson, 1994). No large-scale analysis of Tc5 insertion sites has been published. Our study, based on 197 independent and distinct Tc5 insertions, produces the consensus: C/AC**T**N**A**GG/T (Fig. 1, bottom). N shows an even distribution of all four bases. Preferred bases at positions −1 and +1 are C and G respectively (76% and 79%). In the remaining cases, there is a strong bias for the other pyrimidine/purine, so that the consensus could also be written C/AY**T**N**A**RG/T. Extension of the analysis to positions −3/+3 reveals a slight preference for C/T at position −3 and A/G at position +3 ($P<0{\cdot}001$ and $P<0{\cdot}02$ respectively). Contrary to the Tc1 and Tc3 target sites, Tc5 prefers palindromic sequences on each side of the TNA. For instance, a G is detected at position +1 in 79% of sequences overall, but in 85% of sequences where C is the mirror base, and in only 70% of sequences where T is the mirror base. This tendency, however, does not apply to positions 2 and 3, for which the presence of a base on one side does not favour the presence of the matching base on the other side. Overall, 19% of Tc5 insertions are palindromic around the TNA site.

In summary, this short paper better defines the insertion sites of *C. elegans* transposons Tc1, Tc3 and Tc5. In the case of Tc1 and Tc3, which are members of the *mariner* family of transposons, consensus sites are not palindromic.

The authors are grateful to L. Duret and C. Biemont for helpful discussions. This work was supported by the Centre National de la Recherche Scientifique (CNRS), the Rhône-Alpes district, and the Association Française pour la Recherche contre le Cancer (ARC).

References

Anderson, P. (1995). Mutagenesis. In *Caenorhabditis elegans* (ed. H. F. Epstein & D. C. Shakes), pp. 31–58. San Diego: Academic Press.

Collins, J. & Anderson, P. (1994). The Tc5 family of transposable elements in *Caenorhabditis elegans*. *Genetics* **137**, 771–781.

Collins, J., Forbes, E. & Anderson, P. (1989). The Tc3 family of transposable genetic elements in *Caenorhabditis elegans*. *Genetics* **121**, 47–55.

Eide, D. & Anderson, P. (1988). Insertion and excision of *Caenorhabditis elegans* transposable element Tc1. *Molecular and Cellular Biology* **8**, 737–746.

Ketting, R., Haverkamp, T., van Luenen, H. & Plasterk, R. (1999). Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99**, 133–141.

Korswagen, H., Durbin, R., Smits, M. & Plasterk, R. (1996). Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. *Proceedings of the National Academy of Sciences of the USA* **93**, 14680–14685.

Martin, E., Laloux, H., Couette, G., Alvarez, T., Bessou, C., Sookhareea, S., Labouesse, M. & Ségalat, L. (2002). Identification of 1088 new transposon insertions of *C. elegans*, a pilot study towards large-scale screens. *Genetics* **162**, 521–524.

Mori, I., Benian, G., Moerman, D. & Waterston, R. (1988). Transposable element Tc1 of *Caenorhabditis elegans* recognizes specific target sequences for integration. *Proceedings of the National Academy of Sciences of the USA* **85**, 861–864.

Plasterk, R. & van Luenen, H. (1997). Transposons. In *C. elegans II* (ed. D. Riddle, T. Blumenthal, B. Meyer & J. Priess), pp. 97–116. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Plasterk, R., Izsvak, Z. & Ivics, Z. (1999). Resident aliens. *Trends in Genetics* **15**, 326–332.

Schneider, T. & Stephens, R. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* **18**, 6097–6100.

van Luenen, H. & Plasterk, R. (1994). Target site choice of the related transposable elements Tc1 and Tc3 of *Caenorhabditis elegans*. *Nucleic Acids Research* **22**, 262–269.