BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 2 2011, pages 286–287
doi:10.1093/bioinformatics/btq639

Genetics and population analysis Advance Access publication November 11, 2010

TE Displayer for post-genomic analysis of transposable elements

Rebecca Rooke[1,2] and Guojun Yang[1,2,*]

[1]Department of Cell and Systems Biology, University of Toronto, Toronto, M5S 1A1 and [2]Department of Biology, University of Toronto at Mississauga, Mississauga, L5L 1C6, Canada

Associate Editor: Jonathan Wren

ABSTRACT

Summary: TE Displayer can be used to retrieve genetic polymorphisms caused by transposable elements (TEs) in large-genomic datasets and present the results on virtual gel images. This enables researchers to compare TE profiles *in silico* and provides reference profiles for experimental analyses.

Availability and implementation: Freely available on the web at http://labs.csb.utoronto.ca/yang/TE_Displayer.

Contact: gage.yang@utoronto.ca

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received and revised on September 28, 2010; accepted on October 31, 2010

1 INTRODUCTION

Transposable elements (TEs) are found in both prokaryotic and eukaryotic organisms. TEs have caused profound structural and functional changes to genomes, where they continue to persist in and shape the evolution of the host. The molecular domestication of TEs in eukaryotic organisms has resulted in viral immunity against murine leukemia virus in *Mus* species (Best *et al*., 1996), the maintenance of telomeres in *Drosophila* (Pardue and DeBaryshe, 2008) and has contributed to site-specific V(D)J recombination in jawed vertebrates [see Sinzelle *et al*. (2009) for review]. Although the majority of TEs are not currently active, their historical activity led to TE-derived polymorphisms. The currently active TEs are a major driving force for genome evolution and revealing TE activity is also essential to understanding genome evolution (Kidwell and Lisch, 1997).

One method to reveal genome-wide insertions caused by TEs is Transposon Display (TD). This technique is a modification of a previously described technique referred to as Amplified Fragment Length Polymorphism (AFLP; den Broeck *et al*., 1998; Vos *et al*., 1995). TD starts with the digestion of genomic DNA with a restriction enzyme. Adapter oligonucleotides are then ligated to the digested genomic DNA fragments. The first round of amplification (pre-amplification) is performed using a primer complementary to the adapter sequence and another primer complementary to the target TE. A second, selective PCR reaction is often performed using the pre-amplification products as templates with a nested primer set. Selective amplification PCR products are analyzed by gel or capillary electrophoresis. If a TE family has a high copy number in a genome, selective nucleotide(s) can be added to the 3′ end of the adapter primer used in the selective amplification, reducing the number of bands per lane (den Broeck *et al*., 1998). The approximate copy number of a TE family in a genome can be estimated from the number of bands. Differences in banding patterns reveal TE insertion polymorphisms between individuals (Schaack *et al.*, 2010; Valizadeh and Crease, 2008). Current TE transposition activity in an individual organism may be inferred from the appearance of new bands on a TD gel (Jiang *et al.*, 2003; Slotkin *et al*., 2009).

Here, we describe a computational algorithm that follows the conceptual principles of TD. The program, called TE Displayer, has a graphical user interface and runs on Windows and Linux operating systems (Supplementary Fig. 1). The output includes a virtual gel image and detailed genomic location information for each band. TE Displayer was tested using TEs in the *Aedes aegypti*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana* and *Oryza sativa* genomes and all of the output from TE Displayer is consistent with the output from manual inspection.

2 RESULTS

2.1 Algorithm and implementation

The required parameters to perform TD experiments include the restriction enzyme, the pre-amplification primer sequences, the selective amplification primer sequences, the size of the adaptor oligonucleotides and the selective base(s). To simulate TD procedure *in silico*, the genomic coordinates of each location harbouring a target TE element are retrieved with BLAST searches from genome sequences using the TE-specific pre-amplification primer as the query sequence [Fig. 1A (i)]. Mismatches between the pre-amplification primer and the genomic sequence are allowed. Because pre-amplification in TD is performed with enzymatically digested genomic DNA, the resulting amplicon is comprised of the sequence between the pre-amplification primer and the first restriction site. Therefore, the next step in the algorithm is to locate the first restriction site from the pre-amplification primer location. This is achieved by retrieving a 5 kb flanking sequence followed by searching for the nearest restriction site [Fig. 1A (ii), (iii)]. The sequence between the pre-amplification primer and the restriction site is scanned for the presence of the TE-specific selective primer [Fig. 1A (iv)]. If a selective primer sequence is found and is in the correct orientation, the size of the conceptual amplicon is calculated as the combined sizes of the selective primer, adaptor and the region between them [Fig. 1A (v)]. Every location harbouring a target TE is processed and a conceptual amplicon is produced.

For each amplicon, the genomic location, amplicon size, selective base and the number of mismatches are recorded in a text field of the graphical interface. In addition, all the conceptual amplicons from a genome are presented as bands in a virtual gel lane (Fig. 1B–E).

*To whom correspondence should be addressed.

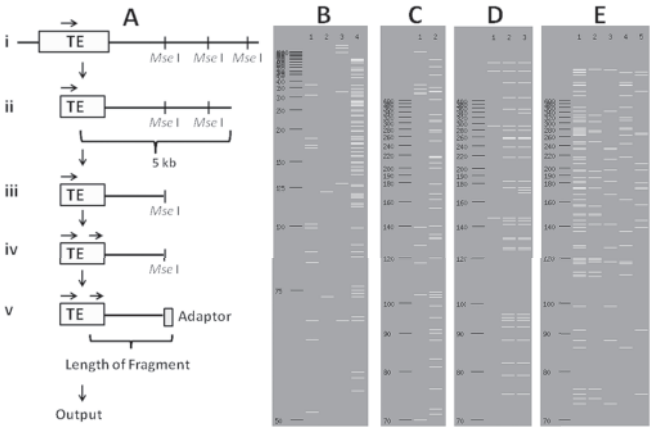

Fig. 1. Algorithm of TE Displayer (**A**) and its virtual gel image output (**B**–**E**). (A) Diagram of TE Displayer workflow (see Algorithm and Implementation). Left arrowhead, pre-amplification primer; right arrowhead, selective primer. (B) *hAT* families in different species. Lane 1: *A.thaliana*; lane 2: *C.elegans*; lane 3: rice; lane 4: *A.aegypti*; lane 5: *D.melanogaster*. (C) *mPing* elements in rice. Lane 1: *O.sativa* var. *indica*; lane 2: *O.sativa* var. *japonica*. (D) TF000720 family in *A.aegypti* with different allowed primer mismatches. Lane 1: no mismatches; lane 2: 1 mismatch; lane 3: 2 mismatches. (E) TF000700 family in *A.aegypti* with different selective bases. Lane 1: no selective base; lane 2: A; lane 3: C; lane 4: T; lane 5: G.

The mobility of a DNA fragment on acrylamide gel is inversely proportional to its size. Therefore, the positions of the bands are calculated using the formula $D1/D2 = S2/S1$, where D is distance and S is size.

2.2 Display TE families in different genomes

To test TE Displayer, we displayed elements of different *hAT* families in five genomes: *A.thaliana, C.elegans, O.sativa, A.aegypti* and *D.melanogaster*. Detailed information on primer sequences and other parameters is provided in the Supplementary Material. Because each TE family differs in sequence, each primer set generates a unique banding pattern on the virtual gel, as expected (Fig. 1B). The copy number of each family, reflected as the number of bands in a virtual gel lane, and the size of each band were consistent with manual inspection of the genomic sequences.

TE Displayer can be used to compare the profiles of a TE family in different but closely related genomes. Virtual gels for *mPing* were generated using two completed rice *(O.sativa)* genome sequences (Fig. 1C). In the *O.sativa* var. *indica* genome, there are only nine *mPing* elements that resolve on the virtual gel, compared with 33 in the *O.sativa* var. *japonica* genome. This is consistent with previous data (Jiang *et al*., 2003), which shows significantly more *mPing* copies in the *japonica* variety than the *indica* variety.

TE Displayer allows mismatches between the pre-amplification primer and the genomic sequence, a source of non-specific bands in TD analysis. When displaying the TF000720 TE family in *A.aegypti*, four bands appeared on the virtual gel image when no primer mismatches were permitted. With one and two primer mismatches, 33 and 37 bands appeared, respectively (Fig. 1D, Supplementary Table 2). For analyses of TE families with high copy numbers, users may choose selective base(s) to reduce the number of bands per lane, resulting in an increased resolution of each band (Fig. 1E).

3 METHODS

The algorithm was implemented with PERL. BLAST search is performed with the standalone program package 2.2.22 with an *E*-value of 10 000. The modules Perl/Tk is used to implement the graphical interface (Supplementary Fig. 2). GD-2.43 module was used to generate the virtual gel images. BioPerl modules including StandAloneBlast and Bio::SearchIO were used to perform BLAST search and parse the output. TE Displayer was tested on Linux and Windows (XP, Vista, 7) standalone systems and the SciNet high performance system (University of Toronto).

4 CONCLUSION AND DISCUSSION

Emerging sequencing technologies promise high efficiency and low cost. Whole-genome deep sequencing is becoming a routine procedure and will benefit studies of TEs. TE Displayer for post-genomic analyses of TEs enables quick retrieval of TE-derived polymorphisms in large datasets, providing a link between *in silico* analysis and experimental testing.

Discrepancies may be found between an *in silico* profile and one obtained from an experimental TD analysis. For example, although the banding patterns of *mPing* between the TD gel and the virtual gel image are similar, several bands differ. This may be caused by transposition activity of the target TE family. Other factors contributing to the discrepancy may include: (i) incomplete genome sequences which may result in fewer bands on a TE Displayer virtual gel image than that obtained from an experimental analysis; (ii) sequencing errors and inaccuracy in genome sequence assembly that may cause changes in band sizes; (iii) the non-specific amplification in experimental analyses caused by priming of non-target sequences. Therefore, in addition to obtaining TE profiles from genome sequences, TE Displayer may also be useful in detecting transposition activity and sequencing or genome assembly errors.

Funding: National Sciences and Engineering Research Council (RGPIN371565 to G.Y.); Canadian Foundation for Innovation (24456 to G.Y.); Ontario Research Fund; University of Toronto.

Conflict of Interest: none declared.

REFERENCES

Best,S. *et al*. (1996) Positional cloning of the mouse retrovirus restriction gene Fvl. *Nature*, **382**, 826–829.

den Broeck,D.V. *et al*. (1998) Transposon display identifies individual transposable elements in high copy number lines. *Plant J.*, **13**, 121–129

Jiang,N. *et al*. (2003) An active DNA transposon in rice. *Nature*, **421**, 163–167.

Kidwell,M.G. and Lisch,D. (1997) Transposable elements as sources of variation in animals and plants. *Proc Natl Acad. Sci. USA*, **94**, 7704–7711.

Pardue,M.L. and DeBaryshe,P.G. (2008) Drosophila telomeres: a variation on the telomerase theme. *Fly*, **2**, 101–110

Schaack,S. *et al.* (2010) DNA transposons and the role of recombination in mutation accumulation in *Daphnia pulex*. *Genome Biol.*, **11**, R46.

Sinzelle,L. *et al*. (2009) Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell. Mol. Life Sci.*, **66**, 1073–1093

Slotkin,R.K. *et al.* (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, **136**, 461–472.

Valizadeh,P. and Crease,T.J. (2008) The association between breeding system and transposable element dynamics in *Daphnia Pulex*. *J. Mol. Evol.*, **66**, 643–654.

Vos,P. *et al*. (1995) AFLP. A new technique for DNA fingerprinting. *Nucleic Acids Res.*, **23**, 4407–4414.