W708–W713 *Nucleic Acids Research, 2006, Vol. 34, Web Server issue*
doi:10.1093/nar/gkl263

PLOTREP: a web tool for defragmentation and visual analysis of dispersed genomic repeats

Gábor Tóth*, Gábor Deák, Endre Barta and György B. Kiss

Agricultural Biotechnology Center, Gödöllő, Szent-Györgyi Albert u. 4, H-2100, Hungary

Received February 14, 2006; Revised March 1, 2006; Accepted March 31, 2006

ABSTRACT

Identification of dispersed or interspersed repeats, most of which are derived from transposons, retrotransposons or retrovirus-like elements, is an important step in genome annotation. Software tools that compare genomic sequences with precompiled repeat reference libraries using sensitive similarity-based methods provide reliable means of finding the positions of fragments homologous to known repeats. However, their output is often incomplete and fragmented owing to the mutations (nucleotide substitutions, deletions or insertions) that can result in considerable divergence from the reference sequence. Merging these fragments to identify the whole region that represents an ancient copy of a mobile element is challenging, particularly if the element is large and suffered multiple deletions or insertions. Here we report PLOTREP, a tool designed to post-process results obtained by sequence similarity search and merge fragments belonging to the same copy of a repeat. The software allows rapid visual inspection of the results using a dot-plot like graphical output. The web implementation of PLOTREP is available at http://bioinformatics.abc.hu/PLOTREP/.

INTRODUCTION

Repetitive sequences are ubiquitous in eukaryotic genomes (1,2). The fraction of the genome occupied by repetitive elements varies by species and ranges from a few percent in lower eukaryotes to >70% in some plants (3,4). Dispersed or interspersed repeats almost exclusively result from the transposition of mobile genetic elements, which belong to one of the main classes of DNA transposons, retrotransposons or retrovirus-like elements. Proper annotation of a genome involves the identification and classification of transposable elements (TEs). Recognizing repetitive DNA is essential for accurate genome assembly while masking them is a prerequisite for sequence similarity searches aimed at gene prediction and functional annotation. Cataloging and further analysis of TEs promotes our understanding of TE and genome evolution. The best-known tools for systematic annotation of repeat families and subsequent repeat masking are RepeatMasker (A. F. Smit, R. Hubley and P. Green; http://www.repeatmasker.org/) and CENSOR (5,6). Both programs perform similarity searches based on local alignments using precompiled libraries of consensus or representative sequences of repeat families. A profile HMM-based method has been applied with success to find certain groups of TEs in the rice genome (7). BLAST-based searches can also prove useful in TE annotation of genomic sequences (8–10). These approaches offer reliable results in repeat identification. However, the coverage of the precompiled libraries is inevitably patchy for species with incompletely sequenced genomes. The species-specific nature of many TEs requires such TE and repeat databases to be built for each genome sequencing project simultaneously as genome assembly and genome annotation proceeds. Structural features characteristic for particular superfamilies of TEs can be utilized to find superfamily members: the LTR_STRUC program (11) identifies LTR retrotransposons, while the FINDMITE (12) and MAK (13) programs are designed to locate MITEs (miniature inverted repeat TEs). Recently, several more general methods for automated *de novo* repeat identification and classification have been described and implemented in the programs RECON (14), PILER (15) and RepeatScout (16). While these tools perform relatively well in finding repetitive families, their output is often redundant and the quality of the consensus sequences derived is not comparable with that of the entries in manually curated databases.

Another frequently observed problem when searching against a repeat library is that putative TEs in the query sequence appear only partially and fragmented in the output of the program. This phenomenon is usually due to the considerable divergence of the repeat from the reference sequence

*To whom correspondence should be addressed. Tel: +36 28 526224; Fax: +36 28 526101; Email: tothg@abc.hu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

in the database. It is more pronounced when good consensus sequences have not yet been generated for each subfamily of the particular TE. Divergence from the once active founder element, which is supposedly reconstructed in the consensus sequence, is caused by various mutations (nucleotide substitutions, deletions or insertions) in the sequence since the transposition. If the comparison involves a representative copy instead of a consensus, the sequence difference may double. Merging the fragmented hits to identify the whole region that represents an ancient copy of a mobile element is challenging, particularly if the element is large and suffered multiple deletions or insertions.

Here we report PLOTREP, a web tool designed to address the problem of apparent fragmentation of search results observed during repeat annotation of genomic sequences. PLOTREP can identify repeats in BAC-sized genomic regions (up to several 100 kb). First, a sequence similarity search is carried out to detect matches against a library of various reference sequences. The user can compile and upload his/her own set of known repetitive sequences to be used as the reference library or select from the libraries offered by the server. The results of the search are then post-processed by the program and defragmentation of the regions belonging together is carried out. All fragments predicted to be parts of the same copy of a repeat (i.e. the result of a single transposition event) are merged and plotted to show the whole region covered by the element. Positions of large deletions and insertions with respect to the reference sequence are listed in the output beside the positions of the merged repetitive regions. PLOTREP allows rapid visual inspection of the results: a two-dimensional (2D) dot-plot like graphics is generated for each reference sequence showing unmerged and merged hits in the query sequence. The graphical output is particularly useful in the analysis of large, several kilobase-long TEs like retrotransposons and retrovirus-like elements, which are more prone to suffer deletions and insertions if they were present in the genome for long times.

METHODS AND IMPLEMENTATION

The input of the PLOTREP web tool is a genomic query sequence of length up to 1 Mb . The operation of the program can be divided into three main steps: (i) a sequence similarity search is carried out against a reference library of known interspersed repetitive sequences; (ii) the search result is post-processed to find matches that can be merged into one combined region representing a single copy of a repeat and (iii) the results of the first and second step are displayed in both tabular and graphical format. In the first step, significant local alignments between the genomic query and the sequences of the reference library are found by the software CENSOR (5). CENSOR, like RepeatMasker, the other well-known tool for library-based repeat identification, is designed to locate and mask regions in genomic sequences that correspond to known repetitive elements. CENSOR uses the fast and sensitive similarity search program WU-BLAST (W. Gish; http://blast.wustl.edu/). Optionally, the BLASTN or BLASTX programs of the WU-BLAST package can be used directly instead of CENSOR. All three programs allow the relatively rapid identification of fragments homologous to sequences in a repeat library either supplied by the user or chosen from a list offered in the PLOTREP search form. The matching fragments are often not contiguous even if they have been originated from the same transposition event of a TE. Gaps, deletions and insertions of unrelated sequences may disrupt the alignment. Throughout this article, we refer to sequences probably homologous to the reference but not sufficiently similar to appear among the local alignments generated by CENSOR as 'gaps'.

The idea behind the second, defragmentation or merging step is based on the proven usability of the dot-plot method for repeat analysis. 2D dot-plots are often used to check and visually inspect repeats (including duplications and inversions) within sequences or local similarities between two otherwise unrelated sequences. Generation of a full dot-plot with a program like Dotter (17) is very time consuming for large sequences and dot-plot programs do not allow the automatic determination of the borders of matching regions. Applying the dot-plot approach on the results of a relatively fast local similarity detection program combines the advantage of the visual inspection of the matches with the possibility of automated processing of the results if manual intervention is not feasible. A similar approach was proven useful in the BLAST2GENE program designed to convert BLAST output into independent genes and gene fragments (18).

Hereafter the line that represents a matching fragment in the 2D plot will be referred to as a diagonal (Figure 1). The merging step involves diagonals that maintain consistency. Two diagonals are consistent if their order is the same with respect to both the query and the reference sequences. In PLOTREP we use the notion of the offset difference between diagonals. The offset difference is the distance between two parallel or nearly parallel diagonal lines. If the two sequences are drawn starting from the upper left corner of the rectangle then the absolute offset of a positively oriented diagonal is measured from the lower left corner, while the absolute offset of a negatively oriented diagonal is measured from the upper left corner. Diagonals closer to each other than a given maximum offset difference are combined into a group of consistent diagonals. Since deletions and insertions increase the offset difference, they prevent the flanking fragments from being grouped together. Therefore, pairs of neighboring groups are examined whether they can be considered consistent under the assumption that they are separated only by a deletion (i.e. fragments are adjacent in the query sequence but separated in the reference sequence) or an insertion (i.e. fragments are adjacent in the reference sequence but separated in the query sequence). Fragments are accepted as adjacent if they are closer too each other than a given maximum distance in one of the two sequences. Gaps separate two groups on diagonals with no or small offset difference between them. If the offset difference is below a pre-defined threshold, even such groups are combined. Depending on parameter settings, all or most fragments predicted to be parts of the same individual copy of a repeat are merged and boundaries are calculated for the whole region covered by the element.

In the third step, the output is generated and displayed. The output contains (i) a diagram summarizing all repeats, both unmerged (i.e. the raw CENSOR output) and merged (processed in the second step), found in the

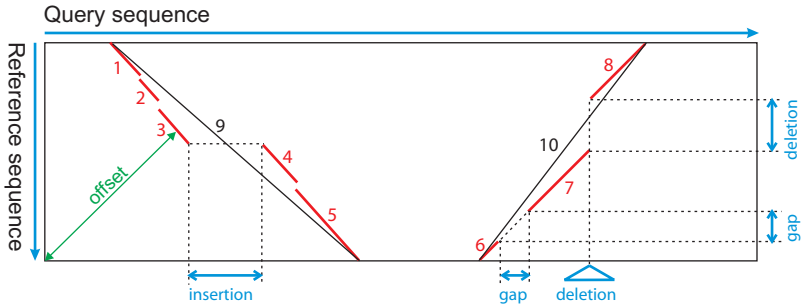

Figure 1. The diagram explains the terms used in the description of the algorithm and provides help to interpret the 2D plot. Matching fragments are shown as red diagonals (1–8). Fragments 1–5 are in positive orientation while fragments 6–8 are in negative orientation. The absolute offset of the diagonals is calculated as indicated for fragment 3 as an example. The offset differences separating fragments 1, 2 and 3 are small, therefore they can be grouped together as the initial step of merging. Similarly, fragments 4 and 5 can also be grouped. A gap is a discontinuity with small offset difference between flanking fragments like 6 and 7. Insertions and deletions are defined with respect to the reference sequence on the vertical axis, and their presence is examined after the groups of 'same-diagonal' fragments are formed. An insertion or a deletion is characterized by large offset difference and adjacency of the flanking fragments in the reference sequence or in the query, respectively. Depending on the parameters, fragments belonging together can be merged as shown by the black lines 9 and 10.

query; (ii) tables with position data for unmerged and merged repeat regions and (iii) 2D dot-plot like representations of sequence similarity between the query sequence and each of those reference sequences that matched it in the first search step. The web interface is programmed in Perl CGI and JavaScript, while the programs performing the fragment merging step and generating the graphical output are written in Perl. The latter scripts are able to process not only the output of CENSOR but also the GFF-format RepeatMasker output or other simple plain text table listing positions of matched fragments. The scripts are available to the academic community upon request.

FEATURES

We designed PLOTREP to be suitable for anyone who wants to annotate BAC-sized or smaller genomic sequences and identify interspersed repeats similar to consensus or reference sequences representing known families of repetitive elements. To meet this requirement, PLOTREP finds matching regions in the genomic sequence and merges them if they are predicted to jointly compose the same individual copy of a repeat, which in turn is presumed to have resulted from a single transposition event. PLOTREP also allows rapid visual inspection of the findings via a dot-plot like 2D graphical output.

On the other hand, PLOTREP can also fulfill the requirements of those who would like to analyze certain TEs or identify novel transposons or retrotransposons. We are interested in plant repetitive elements and it is reflected in the inclusion of the TIGR Plant Repeat Database (http://www.tigr.org/tdb/e2k1/plant.repeats/) (19) among the repeat libraries offered by the server. Repbase Update, the most comprehensive database of repetitive element consensus sequences, is also available for use in the CENSOR or BLASTN search step. Repbase Update is compiled and maintained by the Genetic Information Research Institute (6,20,21). A useful feature of the server is that it also allows searches against user-supplied repeat libraries, a frequent claim when analyzing genomes where sequencing is still under progress and custom-made libraries are preferred because public repeat libraries do not contain TE sequences for the species.

One approach to identify a TE which belongs to a new TE family is based on the detection of largish insertions into regions of known or predictable (i.e. conserved) sequence structure. Nested insertions of TEs into each other are frequently observed, particularly in large plant genomes (3,22,23). Thus, transposition of a sequence of unknown identity into a repetitive element that belongs to a known family or subfamily can be easily noticed. Since PLOTREP can detect insertions into nearly full-length or even partial elements, and many of such insertions result from (retro)transpositions, the program can help to identify unknown families of mobile genetic elements and determination of the relative ages of element families or subfamilies.

When searching against a library of repetitive elements, a gap observed between two consecutive hits may be caused either by an unrelated sequence or, more probably, by extensive divergence of a homologous region from the repeat consensus/reference, which prevents detection of the remote similarity by the search program. By examining the offset difference between the diagonals on which the hits flanking the gap lie, PLOTREP can predict whether a region not detected by the similarity search may belong to a TE or not. However, small insertions (e.g. MITEs) or recombination may result in sequences interpreted by PLOTREP as a 'gap', therefore the origin of gaps reported by PLOTREP should be checked manually.

Visual inspection of the graphical output, especially in the dot-plot like 2D form, can reveal even very complex patterns of nested insertions and element duplications. A unique feature of PLOTREP when using a user-supplied reference library is that long terminal repeat (LTR) retrotransposons are treated in a specific way. The LTR sequence and the internal sequence must be in two separate sequence entries and special rules apply for naming the two sequences (see the online Tutorial for details). In this case, PLOTREP attempts to merge fragments for the whole retrotransposon of the structure LTR–internal–LTR, and plots this combination on the 2D dot-plot like image. However, PLOTREP may be unable to resolve complex nested insertions or tandem LTR retroelements resulting from recombination between LTRs belonging to two different elements. The 2D diagram can greatly help the user interpret the results in such cases. Manual editing may

also be required if an element harbors homologies with two closely related but differently named elements. Consequently, the 2D graphical representation of sequence comparison often provides information not conspicuous from the summary figure and the tables, and this surplus is more pronounced when one inspects matches to large repetitive elements including retrotransposons but less so when looking at small repeats. This feature is particularly useful in analyzing plant genomes, since plant LTR retroelements longer than 10 kb are not uncommon (10,24–27).

INPUT AND OPTIONS

Input sequences, search method and defragmentation options can be specified in the Search page. One of the three search programs, CENSOR, BLASTN or BLASTX can be selected by the user, with CENSOR being the default. The query sequence of up to 1 Mb must be in 'FASTA' format. One can either paste it into the input field or upload from a file. Only a single sequence entry is allowed, multisequence files are rejected by the program.

A reference sequence library consisting of known repetitive elements to be searched against has to be specified. There are essentially two options to supply the library, except for BLASTX for which only the first option is available. First, the user can paste reference sequences into an input field or upload a sequence file. Second, the user can select a library from those stored on the server and offered in a list. In the user-supplied reference library, sequences must be in FASTA format and multiple sequence entries are allowed. LTR retroelements are handled in a specific manner (see above and the online Tutorial). Selectable server-based libraries currently include various sections of the Repbase Update database (6) and the TIGR Plant Repeat Database (19). More databases are planned for later addition.

Five parameters affecting the second, processing and defragmentation step of the search, can be modified by the user. Maximum insertion length is the maximum length of an insertion allowed between two consecutive hits to merge them. An insertion longer than this will keep the fragments separated. Maximum deletion length is the maximum length of a deletion allowed between two adjacent hit fragments to merge them. A deletion longer than this will keep the fragments separated. Maximum gap length is the maximum length of a same-diagonal gap allowed between two consecutive hits to merge them. The default value is zero when there is no limit and merging is guided only by the offset difference between the diagonals on which the two hits lie. Minimum coverage to merge is the minimum total coverage of merged fragments in percentage of the reference sequence length to accept the merging of the fragments. Maximum relative offset difference is the maximum relative difference in offset with respect to total repeat length. Two consecutive fragments are merged only if the relative offset difference is smaller than this value.

CENSOR can be run in three different sensitivity modes. The WU-BLAST parameter settings corresponding to these modes are listed in the online Tutorial. Certain parameters (word size, *E*-value threshold, gap penalties) of the direct WU-BLAST searches can also be adjusted by the user (see the online Tutorial for details).

OUTPUT

The output consists of three main parts (Figure 2): (i) a summary figure; (ii) two alternative tables of position data, one for the original unmerged hits and another for the merged repeats and (iii) 2D dot-plot like figures. On the top of the Result page, there is a diagram summarizing the regions occupied by all repeats which have been found in the query. Both the unmerged fragments (i.e. raw CENSOR or BLAST output) and the merged ones (defragmented in the second, post-processing step) are indicated by two rows of horizontal color bars below a scaled line representing the query sequence. The diagram can be enlarged to view more details if fragments seem to overlap. Below it, a table with the merged repeats is shown by default but the user can also select and view another table displaying raw results of the CENSOR or BLAST search step. Columns of the table for merged fragments include (i) repeat name; (ii) position of the combined repeat region in the query; (iii) positions of insertions (with respect to the query sequence) if any; (iv) positions of deletions (with respect to the reference sequence) if any and (v) positions of gaps (with respect to the query sequence) if any. Columns of the table for raw CENSOR (or BLAST) output are (i) repeat name; (ii) position of the hit in the query sequence; (iii) length of the fragment in the query sequence; (iv) position of the matching fragment in the reference sequence; (v) sequence similarity; (vi) direction of the repeat in the query sequence. Both tables can be downloaded as plain text files. Sequences of either the merged repeat regions or the insertions can also be downloaded using links in the table of merged fragments. Sequences of the matching query fragments and the original CENSOR alignments can be downloaded using the appropriate links in the table containing the raw output. The 2D plot of sequence comparison between the query and a matching individual reference sequence can be viewed by clicking on the repeat name in either tables. The query sequence is drawn horizontally and a single reference sequence is drawn vertically. If the plot was accessed from the table of merged fragments, a black line indicating the merged region appears beside the red lines representing the original fragments. The 2D plots can be zoomed in and out at the user's convenience. A similar diagram displaying the query sequence compared with itself, thus helping the user to recognize direct and inverted repeats, can be opened by clicking on the 'Show DotPlot' button below the summary figure. The online Tutorial explains the output options in more details.

CONCLUSIONS

The defragmentation and visualization tool PLOTREP facilitates detection and further studies of repetitive elements in eukaryotic genomes. This software supports the identification of full-length elements even if they are fragmented and disrupted by insertions. Further analysis of sequences causing insertions larger than a few dozen base pairs may reveal previously unknown families of mobile genetic elements. Visual inspection of the 2D representation of fragments matching between the query sequence and a TE reference assists the user to grasp the repeat organization of the genomic region of interest.

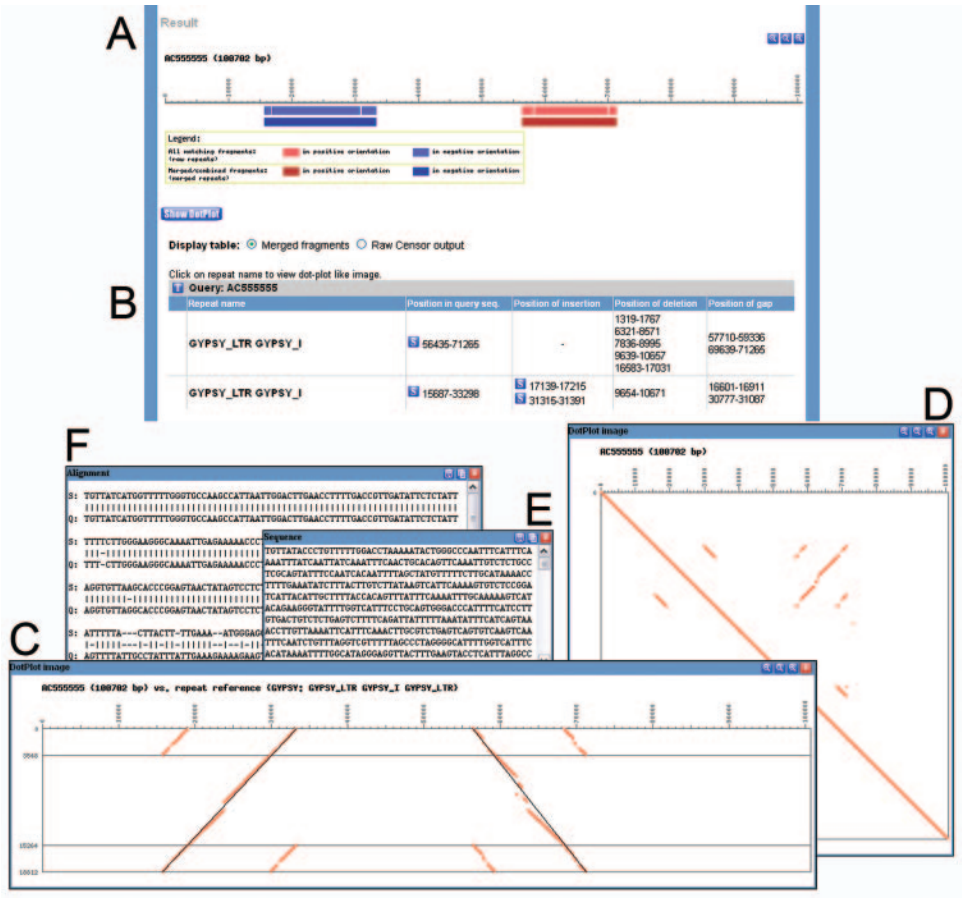

Figure 2. An example of PLOTREP results. A genomic query sequence was searched against a small user-supplied library containing LTR and internal sequences of an LTR retrotransposon. (**A**) A diagram summarizing all matching hits and those merged by PLOTREP, providing an overall picture of repeat positions. (**B**) A table listing positions of merged regions along with the positions of insertions, deletions and virtual gaps within these regions. (**C**) A 2D dot-plot like diagram displaying the comparison between the query (on the horizontal axis) and a library reference sequence (here in combined LTR–internal–LTR structure on the vertical axis). All matching fragments are shown as red lines and the merged regions are depicted as black lines. (**D**) A dot-plot like diagram displaying the query sequence compared with itself. (**E**) The sequence of a merged region or a region covered by an insertion can be downloaded by clicking on the 'S' button. (**F**) Local alignments generated by CENSOR can be viewed by clicking on the 'A' button in the table listing the raw CENSOR output of fragment positions (this alternatively displayed table is not shown here).

ACKNOWLEDGEMENTS

The authors are grateful to Jerzy Jurka and the Genetic Information Research Institute (GIRI) for providing the Repbase Update database and the Censor program. The Censor program is being developed and maintained by Oleksiy Kohany at GIRI. The authors also thank the TIGR Plant Repeat Database Team for making their databases freely available. The authors would like to thank Ferenc Marincs for critical reading of the manuscript, and two anonymous referees for their helpful suggestions. This study was supported by European grant Grain Legumes (grant no. FOOD-CD-2004-506 223), NKFP (Hungarian Science and Development Program) Medicago Biotechnology (grant no. 4/031/2004), and OTKA (Hungarian Scientific Research Fund) T038211, OTKA T046645. G.T. was partly supported by a Bolyai Postdoctoral Fellowship from the Hungarian Academy of Sciences. Funding to pay the Open Access publication charges for this article was provided by the Hungarian Science and Development Program (NKFP grant 4/031/2004).

Conflict of interest statement. None declared.

REFERENCES

1. Kazazian,H.H.,Jr (2004) Mobile elements: drivers of genome evolution. *Science*, **303**, 1626–1632.
2. Feschotte,C., Jiang,N. and Wessler,S.R. (2002) Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.*, **3**, 329–341.
3. SanMiguel,P., Tikhonov,A., Jin,Y.K., Motchoulskaia,N., Zakharov,D., Melake-Berhan,A., Springer,P.S., Edwards,K.J., Lee,M., Avramova,Z. *et al.* (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.
4. Vicient,C.M., Suoniemi,A., Anamthawat-Jonsson,K., Tanskanen,J., Beharav,A., Nevo,E. and Schulman,A.H. (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus hordeum. *Plant Cell*, **11**, 1769–1784.
5. Jurka,J., Klonowski,P., Dagman,V. and Pelton,P. (1996) CENSOR–a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.*, **20**, 119–121.
6. Jurka,J., Kapitonov,V.V., Pavlicek,A., Klonowski,P., Kohany,O. and Walichiewicz,J. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.*, **110**, 462–467.
7. Juretic,N., Bureau,T.E. and Bruskiewich,R.M. (2004) Transposable element annotation of the rice genome. *Bioinformatics*, **20**, 155–160.
8. Wicker,T., Guyot,R., Yahiaoui,N. and Keller,B. (2003) CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. *Plant Physiol.*, **132**, 52–63.

9. Chantret,N., Salse,J., Sabot,F., Rahman,S., Bellec,A., Laubin,B., Dubois,I., Dossat,C., Sourdille,P., Joudrier,P. *et al*. (2005) Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell*, **17**, 1033–1045.
10. Sabot,F., Guyot,R., Wicker,T., Chantret,N., Laubin,B., Chalhoub,B., Leroy,P., Sourdille,P. and Bernard,M. (2005) Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. *Mol. Genet. Genomics*, **274**, 119–130.
11. McCarthy,E.M. and McDonald,J.F. (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics*, **19**, 362–367.
12. Tu,Z. (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc. Natl Acad. Sci. USA*, **98**, 1699–1704.
13. Yang,G. and Hall,T.C. (2003) MAK, a computational tool kit for automated MITE analysis. *Nucleic Acids Res*., **31**, 3659–3665.
14. Bao,Z. and Eddy,S.R. (2002) Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res*., **12**, 1269–1276.
15. Edgar,R.C. and Myers,E.W. (2005) PILER: identification and classification of genomic repeats. *Bioinformatics*, **21**, i152–i158.
16. Price,A.L., Jones,N.C. and Pevzner,P.A. (2005) *De novo* identification of repeat families in large genomes. *Bioinformatics*, **21**, i351–i358.
17. Sonnhammer,E.L. and Durbin,R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, **167**, GC1–10.
18. Suyama,M., Torrents,D. and Bork,P. (2004) BLAST2GENE: a comprehensive conversion of BLAST output into independent genes and gene fragments. *Bioinformatics*, **20**, 1968–1970.
19. Ouyang,S. and Buell,C.R. (2004) The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Res*., **32**, D360–363.
20. Jurka,J. (1998) Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struct. Biol*., **8**, 333–337.
21. Jurka,J. (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet*., **16**, 418–420.
22. Rostoks,N., Park,Y.J., Ramakrishna,W., Ma,J., Druka,A., Shiloff,B.A., SanMiguel,P.J., Jiang,Z., Brueggeman,R., Sandhu,D. *et al*. (2002) Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. *Funct. Integr. Genomics*, **2**, 51–59.
23. Gu,Y.Q., Coleman-Derr,D., Kong,X. and Anderson,O.D. (2004) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four triticeae genomes. *Plant Physiol*., **135**, 459–470.
24. Kumar,A. and Bennetzen,J.L. (1999) Plant retrotransposons. *Annu. Rev. Genet*., **33**, 479–432.
25. Vicient,C.M., Kalendar,R. and Schulman,A.H. (2001) Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Res*., **11**, 2041–2049.
26. Wright,D.A. and Voytas,D.F. (2002) Athila4 of *Arabidopsis* and Calypso of soybean define a lineage of endogenous plant retroviruses. *Genome Res*., **12**, 122–131.
27. Neumann,P., Pozarkova,D. and Macas,J. (2003) Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. *Plant. Mol. Biol*., **53**, 399–410.