Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2020) **18**, pp. 2379–2381 doi: 10.1111/pbi.13425

Brief Communication

RTRIP: a comprehensive profile of transposon insertion polymorphisms in rice

Zhen Liu[1,†], Tingzhang Wang[2,†], Lin Wang[3,†], Han Zhao[4], Erkui Yue[1], Yan Yan[1], Faiza Irshad[1], Ling Zhou[4], Ming-Hua Duan[5] and Jian-Hong Xu[1,*]

[1]*Institute of Crop Science, Zhejiang Key Laboratory of Crop Germplasm, Zhejiang University, Hangzhou, China*
[2]*Zhejiang Institute of Microbiology, Hangzhou, China*
[3]*Systems Biology Division, Zhejiang-California International Nanosystems Institute (ZCNI), Zhejiang University, Hangzhou, China*
[4]*Institute of Biotechnology, Jiangsu Provincial Key Laboratory of Agrobiology, Jiangsu Academy of Agricultural Sciences, Nanjing, China*
[5]*Zhejiang Zhengjingyuan Pharmacy Chain Co., Ltd. & Hangzhou Zhengcaiyuan Pharmaceutical Co., Ltd, Hangzhou, China*

Received 26 July 2019;
revised 2 April 2020;
accepted 19 May 2020.
**Correspondence* (Tel 86-571-88982406; fax 86-571-88982406; email jhxu@zju.edu.cn)
†These authors contributed equally.

Keywords: rice, transposable element, insertion polymorphisms, molecular marker, gene tagging, next-generation sequencing.

Transposable elements (TEs), also known as transposons, a type of mobile genetic elements, are widespread across all investigated eukaryotic organisms and typically constitute the major portion of most genomes, especially in grasses, where they can account for up to 90% of the genome (Vitte *et al*., 2014). They not only are actively involved in altering gene structure and regulating gene expression, but also have played a profound role in reshaping genomic architecture and maintaining genomic stability (Lisch, 2013). Apart from important biological functions, TEs have been widely exploited as gene tagging and molecular markers for gene function and genetic research (Kumar and Hirochika, 2001). Their active transposition can introduce abundant genetic polymorphisms among individuals considering the presence and absence of insertions, which have been shown to contribute to genome evolution and differentiation between populations (Gonzalez *et al*., 2008; Studer *et al*., 2011). A comprehensive profile of transposon insertion polymorphisms (TIPs) is critical to TE family characterization, genetic evolution research as well as molecular marker-assisted breeding. Therefore, a variety of sequencing strategies and bioinformatics algorithms have been developed to efficiently identify TE loci based on next-generation sequencing (NGS) technology, and only few profiles have been constructed in well-studied model organisms, such as *Drosophila melanogaster*, *Caenorhabditis elegans* and *Homo sapiens* (Kofler *et al*., 2012; Laricchia *et al*., 2017; Rishishwar *et al*., 2015). However, it has not been reported in rice and most plants until now.

With this problem in mind, we obtain a comprehensive TIP profile of 60 743 TE loci by analysing the resequenced data from 3000 diverse rice accessions using our developed pipeline (Figure 1a; http://ibi.zju.edu.cn/Rtrip/method.html). About 75% loci are shared by two or more rice accessions and show abundant presence/absence variations, while the remaining are private to a single accession. The average number of TE loci is 6304 for each accession, and shows large difference among accessions, varying from 4898 to 10 155. Moreover, 19 160 TE loci are inserted within or nearby genes (200 bp flanking regions), which may have a potential effect on gene function. To facilitate querying and retrieval of these data, a convenient database named RTRIP (Rice Transposon Insertion Polymorphism; Figure 1b; http://ibi.zju.edu.cn/Rtrip/index.html) has been established, which contains the information of 3000 rice varieties, 60 743 TE loci and genotyping of each variety, and provides versatile searching and browsing functions through intuitive web-based interfaces.

The varieties module includes the information of a core collection of 3000 rice accessions, which represent the genetic diversity of this species to a large extent (The 3000 rice genomes project, 2014). These accessions are from 89 countries/regions and classified into five varietal groups, including indica, aus/boro, basmati/sadri, tropical japonica and temperate japonica. In addition to the country and varietal group designation, sample name, source, variety name, designation, genetic stock accession id, DNA accession id, biosample and SRA sample are also listed for each accession in our database if available (Figure 1c). User can browse the entire list of varieties, and they are also allowed to retrieve a subset of the list by imposing one or more filtering conditions based on the searching function on the interface.

The TE loci module incorporates a total of 60 743 TE loci identified from the rice population, which cover 496 TE families and show abundant insertion polymorphisms among accessions. For each locus entry, we offer detailed information, such as the locus ID, genomic position, insertion orientation, subordinate family and reference length, population frequency, presence/absence status in reference genome and its description (Figure 1d). If a TE locus is detected from both the forward and reverse directions of insertion site, the locus ID corresponding to the other direction is showed under column 'Mates'. To facilitate their application as molecular markers, the 200 bp sequences flanking TE insertions have been extracted and stored under the 'Flanking Sequence'. Users can click on the name of TE family to get detailed information of the family, where the hierarchical classification and TE consensus sequence are presented. Similarly, a searching function is also added to the module for users to extract TE loci for a given chromosome region or TE characteristic. Furthermore, we have developed a dynamic mapping tool to graphically visualize the distribution of filtered locus subset in rice genome (Figure 1d).

Among 60 743 TE loci, 19 160 are located within or 200 bp up- and down-stream of annotated genes, which may affect the function of corresponding gene by insertion mutation or epigenetic regulation (Lisch, 2009). Given their potential importance in

Figure 1 The pipeline for data analysis and screenshots of representative resources in RTRIP. (a) The general schematic view of the procedure to identify TE insertion polymorphism. (b) The home page of RTRIP. (c) The varieties module for detailed information of rice varieties. (d) The TE loci module showing information of TE loci identified in this study. (e) An example of annotated genes carrying TE loci. (f) The TE genotyping module providing the presence/absence status of identified TE loci in rice population. (g) The genome browser page for integrating TE variations with other omics data. (h) The BLAST search page.

functional genomics, these loci associated with genes have been separately deposited in the 'TE in Genes' submodule. Here the user can find whether the gene of interest carries a TE insertion by entering keywords. In the table, the entry IDs are clickable and will direct users to the information page of genes, which contains comprehensive annotation of genes and graphical expression data (Figure 1e).

The TE genotyping module hosts the allele information of TE loci in each resequenced accession and consists of three submodules, Presence/Absence Matrix, Search by variety and Search by gene, which enable users to access the information in diverse ways. First, 'Presence/Absence Matrix' provides a glance over the entire data set and allows users to browse and query the presence/absence status of large sets of loci in rice population. The allele information is shown in a table format, where the row and column represent TE locus and rice accession, respectively, while the different number means the presence or absence of TE insertion (Figure 1f). Due to the large amount of data

transmission, we specially developed a grid strategy to load the information block by block in order to improve the responsiveness of web page. Likewise, two sets of searching options for TE locus and variety have been set for users to obtain a subset of their interests.

Second, 'Search by varieties' has been designed to facilitate their application as molecular markers for gene mapping and molecular-assisted breeding. Users can submit two or more varieties and specify a chromosome region on the search interface, and the server will return the allele information of TE loci located in corresponding regions. Besides, the detailed information of TE loci has also been friendly integrated to the output results, which provide a direct opportunity for users to select desired loci. Considering that short fragments are more likely to be successfully amplified by PCR reaction, the list of the candidate loci can be further optimized by specifying the threshold of TE length.

Third, 'Search by Genes' submodule is customized for TE loci associated with genes and will be applied to the study of gene function. For a given gene information, if the corresponding gene (s) carries TE insertions in our data set, the search results will present the genotypes of the loci in selected accessions. Moreover, the detailed information of TE loci and genes has also been friendly integrated to the returned tables, which will provide convenience for users to view the relevant information in a single interface.

Some popular bioinformatics tools are also available in RTRIP for browsing, searching and downloading. A generic genome browser (GBrowse) has been embodied as a platform for integrating TE variations and other omics data (Figure 1g). Here, users can visualize more intuitively the distribution of TE loci in rice genome and their relationship with annotated genes in the reference genome. The BLAST search tool has been deployed to determine whether the query sequences submitted by users encompass identified TE loci (Figure 1h). On the results page of BLAST search, each hit will be linked to the TE locus interface, with their coordinate information automatically filled into the corresponding blanks in new page. In this way, the user can know which TE locus belongs to the matched segment of query sequences. In the meantime, we also provide the download and help modules to facilitate users to obtain data in batches and to familiarize them with the database as soon as possible.

In conclusion, we have established a comprehensive bioinformatics platform for TE variation data from rice population. As far as we know, it is the first public database dedicated to share genetic variations introduced by TEs. These polymorphic TE loci, as molecular markers and gene tags, will serve as a valuable resource for genetic mapping and gene function researches, and potentially assist the process of rice breeding. In addition, this resource will also contribute to the investigation of rice TEs. RTRIP will be updated with more TE variation data as new high-quality resequenced data are generated for more rice varieties, and other omics resources, such as epigenomics and miRNAs, will also be integrated into our database when available.

Acknowledgements

We are very grateful to Mr. Minghua Duan and Mr. Zihua Wang from Zhejiang Zhengjingyuan Pharmacy Chain Co., Ltd. for their funds supporting (H20151699 and H20151788 to JHX).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

J.H.X. and Z.L. conceived and designed the experiments. Z.L., T.W. and L.W. performed the research. Z.L., T.W., L.W., H.Z., E.Y., Y.Y., F.I. L.Z., M.H.D. and J.H.X. analysed the data. Z.L. and J.H.X. wrote the paper.

References

Gonzalez, J., Lenkov, K., Lipatov, M., Macpherson, J.M. and Petrov, D.A. (2008) High rate of recent transposable element-induced adaptation in Drosophila melanogaster. *PLoS Biol.* **6**, e251.

Kofler, R., Betancourt, A. J. and Schlötterer, C. (2012) Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. *PLoS Genet.* **8**, e1002487.

Kumar, A. and Hirochika, H. (2001) Applications of retrotransposons as genetic tools in plant biology. *Trends Plant Sci.* **6**, 127–134.

Laricchia, K.M., Zdraljevic, S., Cook, D.E. and Andersen, E.C. (2017) Natural variation in the distribution and abundance of transposable elements across the Caenorhabditis elegans species. *Mol. Biol. Evol.* **34**, 2187–2202.

Lisch, D. (2009) Epigenetic regulation of transposable elements in plants. *Annual Review Plant Biol.* **60**, 43–66.

Lisch, D. (2013) How important are transposons for plant evolution? Nature reviews. *Genetics*, **14**, 49–61.

Rishishwar, L., Tellez Villa, C.E. and Jordan, I.K. (2015) Transposable element polymorphisms recapitulate human evolution. *Mobile DNA*, **6**, 21.

Studer, A., Zhao, Q., Ross-Ibarra, J. and Doebley, J. (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* **43**, 1160–1163.

The 3000 rice genomes project (2014) The 3,000 rice genomes project. *GigaScience*, **3**, 7.

Vitte, C., Fustier, M.A., Alix, K. and Tenaillon, M.I. (2014) The bright side of transposons in crop evolution. *Briefings Funct. Genom.* **13**, 276–295.

