SCIENCE CHINA
Life Sciences

•LETTER TO THE EDITOR•

January 2021 Vol.64 No.1: 167–170
https://doi.org/10.1007/s11427-020-1731-9

# FED: a web tool for foreign element detection of genome-edited organism

Qing Liu[1], Xiaozhen Jiao[1], Xiangbing Meng[2], Chun Wang[1], Cao Xu[2], Zhixi Tian[3], Chuanxiao Xie[4], Genying Li[5], Jiayang Li[2,6], Hong Yu[2*] & Kejian Wang[1*]

[1]*State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China;*
[2]*State Key Laboratory of Plant Genomics, and National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China;*
[3]*State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China;*
[4]*National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China;*
[5]*Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan 250100, China;*
[6]*University of Chinese Academy of Sciences, Beijing 100049, China*

Received March 23, 2020; accepted April 20, 2020; published online June 4, 2020

**Citation:** Liu, Q., Jiao, X., Meng, X., Wang, C., Xu, C., Tian, Z., Xie, C., Li, G., Li, J., Yu, H., and Wang, K. (2021). FED: a web tool for foreign element detection of genome-edited organism. Sci China Life Sci 64, 167–170. https://doi.org/10.1007/s11427-020-1731-9

Dear Editor,

Genome editing, especially the newly developed CRISPR technology, is now widely implemented for diverse medical and agricultural applications (Puchta, 2018). However, for genome editing, the DNA cassettes encoding the editing components are usually assembled and delivered into the cells of organisms (Cong et al., 2013). In most cases, these components will be integrated into the organism genome, which enables efficient genome editing and selection of the modified material. However, the integration of these exogenous components becomes unnecessary after genome editing and therefore increases the chance of producing unwanted off-target mutations. In addition, some studies also detected multiple insertions of plasmid vector components on the genome. All those unwanted insertions might lead to safety problems in the genome-edited organism and will be subject to strict regulatory (Li et al., 2019). For example, the genome-edited hornless calves, which had been predicted to be the world's first genome-edited animal approved for marketing, were detected to contain unintended heterozygous integration of the plasmid. This casts a shadow over the previously anticipated industrialization of genome-edited animals (Young et al., 2020).

To ensure the safety of genome-edited organisms, the screening method suited to reliably detect the integration of the foreign elements is of critical importance. At present, PCR or PCR-based methods are the most widely used for detecting these exogenous elements (Dörries et al., 2010). However, the PCR-based method requires the primers designed for the given sequences. If the primed sequences are mutated or the exogenous elements are fragmented, it will be unable to examine them, and therefore lead to a significant false negative rate. Latest research reported that PCR analysis failed to identify these multiple unwanted head-to-tail integration events, which led to a high rate of falsely claimed precisely edited alleles (Skryabin et al., 2020).

*Corresponding authors (Kejian Wang, email: wangkejian@caas.cn; Hong Yu, email: hyu@genetics.ac.cn)

During the past decade, impressive progress has been made in the field of whole genome sequencing (WGS) in terms of speed, read length, and throughput, along with a sharp reduction in per-base cost. It is expected that the whole genome sequencing technique will become a mainstay for sensitively detecting exogenous DNA fragments of genome-edited organism for its high-resolution of transgene structures and editing events, enabling researchers to diagnose both the expected and unexpected outcomes of segregation from these lineages (Huang et al., 2016). Here, we present a web application tool, Foreign Element Detector (FED), which can elucidate the exogenous DNA components using data derived from the whole genome sequencing. The FED can quantitatively analyze exogenous DNA fragments and their integration sites with great sensitivity and accuracy (Figure 1A). It also includes a vector sequence library composed of 26,921 vectors sequences with 46,695 different components, which enables detection of the exogenous DNA fragments in genome-edited organism without vector information. FED is freely available at http://www.hi-tom.net/FED.

The schematic depiction of FED's workflow is illustrated in Figure S1 (in Supporting Information), a complete analysis consisting of several dependent processes. To complete the overall procedure, the user only needs to perform four steps: (1) Choose an appropriate program; (2) create a 'Job title' and upload the sequences, including the raw NGS sequencing data and vector sequence; (3) choose a reference sequence (the webserver hosts a broad category of sequenced organisms including 24 most commonly used plants and thirteen animals, Table S1 in Supporting Information); (4) enter the email address to accept the result and click submit button (Figure S2 in Supporting Information).

The bioinformatics pipeline implemented in FED made up of three sections and each consists of several sequential steps that lead from the raw WGS data to the final detailed results (Figure S1 in Supporting Information). In the section of exogenous DNA fragment identification for samples with unknown vector information, a vector sequence library composed of 26,921 vectors' sequences downloaded from National Center of Biotechnology Information (NCBI) and the Addgene Vector Database was established. Then the similar regions between the reference genome and vector library were annotated as the first step and WGS reads were mapped to the vector sequence library by using BWA with default parameters (Li and Durbin, 2009). Reads that can be aligned were preserved, and converted to FASTA format, then aligned against the vector library again using BLAT (Kent, 2002). Finally, the vector sequence information contained in the sample is identified by local perl script.

In the section of identifying the distribution of exogenous DNA fragments with detailed vector sequence information provided by the user, it is basically articulated into four steps: (1) annotating similar regions to exclude false positive due to that exogenous DNA which may carry a similar sequence in the host genome; (2) filtering low quality reads; (3) mapping the WGS reads onto the submitted vector sequence; (4) analyzing the coverage position of the whole vector sequence (Figure S1 in Supporting Information).

In the section of identifying the integration sites of the inserted exogenous DNA fragments (location in the genome flanking the insertion), the analysis of reads is divided into two cases (Figure S1 in Supporting Information). The first case is NGS paired-end (PE) reads spanning the insertion junction and allows for one read maps to the reference genome, and the other maps to the vector sequence. Another case is that a 150 bp single-end read must be partly mapped to the reference genome and vector sequence simultaneously. Finally, the reads with the same split location are summarized for the prediction of exogenous DNA integration sites. In the final output, the FED tool exports multiple information for each integration site (Tables S2 and S3 in Supporting Information).

To evaluate the performance of the FED tool, an artificial transgenic rice genome was generated using sequence of pCambia1300 (PC1300) as exogenous DNA sequence. The artificial transgenic rice genome includes seven transgene integration events with different length from 50 to 1,600 bp (Figure S3 in Supporting Information). First, to evaluate the robustness of FED, a round of simulations with sequencing depth from 0.5× to 30× for the artificial transgenic rice genome using the ART software were performed (Huang et al., 2012). The results of simulated data showed that seven segments can be accurately identified when the coverage depth is higher than 5-fold. What is more, precise identification of the integration sites can be realized when the coverage depths reach 10-fold. (Figure S3 and Table S2 in Supporting Information). Subsequently, the genome-edited rice sequence data from Wang's data with depth from 0.5× to 30× (Wang et al., 2019) and genome-edited wheat with different depths from 1× to 10× were also analyzed by FED, and the results demonstrated that the identification of exogenous DNA fragments and integration sites coincided exactly with the simulated data at appropriate sequencing depths (Figure 1B; Figures S4 and S5, Table S3 in Supporting Information). To verify the accuracy of FED in detecting the position of integration sites, these junction locations detected in genome-edited rice were amplified and Sanger sequencing was performed. The results are consistent with the FED identification results (Figure S6 and Table S4 in Supporting Information). The results of simulated data and actual data showed that FED can efficiently screen out all the exogenous DNA fragments and integration sites with vector sequence.

We also performed another series of tests to evaluate FED reliability in identifying exogenous DNA fragments in different genome-edited plants and animals without vector se-

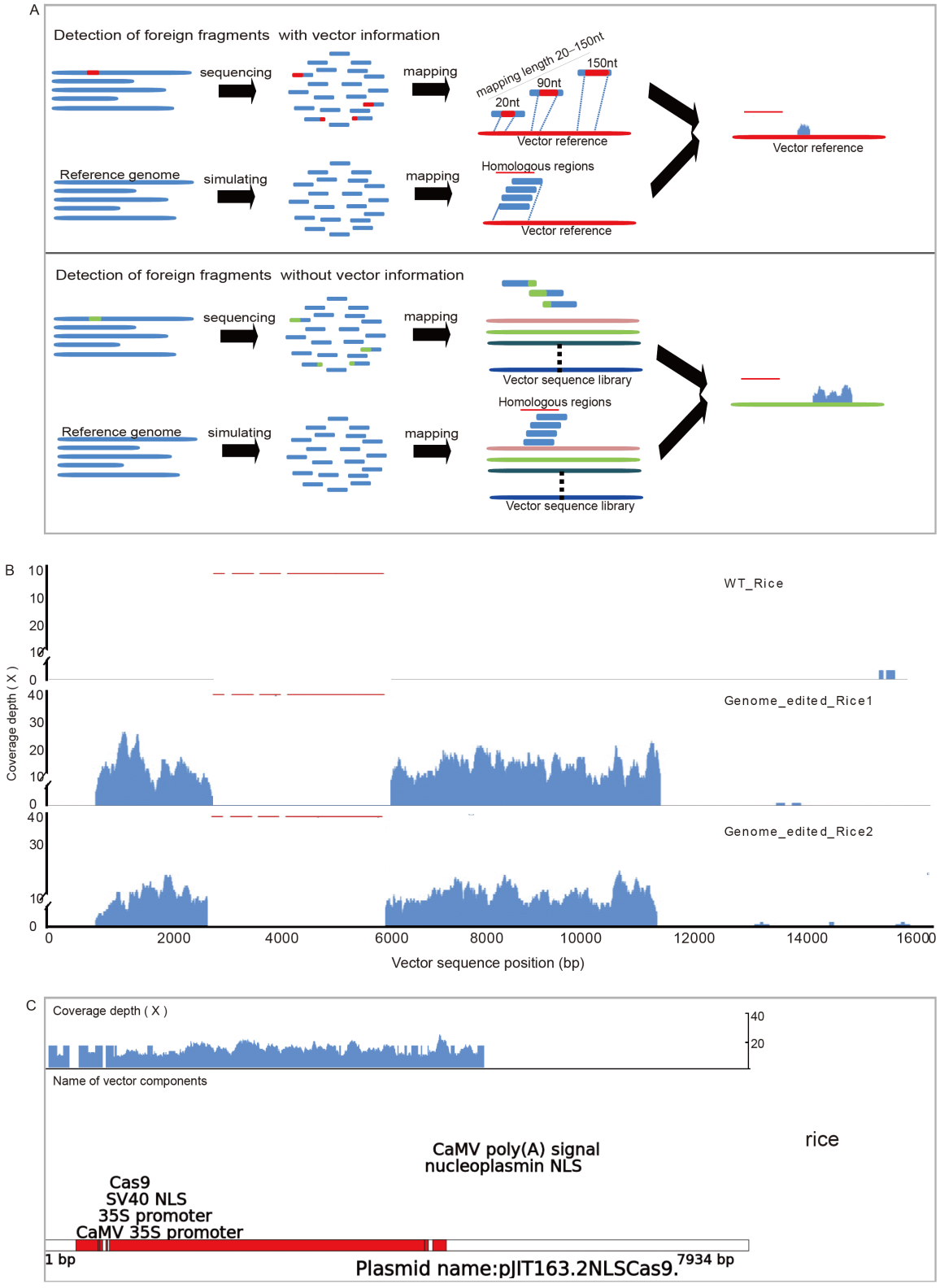


**Figure 1** The design and key features of the Foreign Element Detector. A, A whole genome sample is experimentally fragmented and sequenced. All the reads which align to the vector sequence submitted by users or sequence included in the vector library are recorded. At the same time, next-generation sequencing reads of the reference genome were generated by ART software. Reads were mapped to the vector sequence submitted by users or sequence included in vector library to annotate the similar regions between the reference genome and vector. B, The position and coverage depth of exogenous DNA fragments in genome-edited rice with vector information. The red region represents the homologous region of the vector sequence and the host genome sequence. The vector used is pCambia1300. C, Identification of exogenous DNA fragments of genome-edited rice without vector information. Blue stands for the position and coverage depth of exogenous fragment, and red stands for the position of vector structure.

quence information. First the genetically edited rice sequencing data were used to detect exogenous DNA fragments without vector information (Wang et al., 2019). Results of the analysis showed that the genome-edited rice contains exogenous DNA fragments most similar to the plasmid pJIT163.2NLSCas9 including the common vector component sequence of Cas9, 35S promoter and nucleoplasmin NLS, but no exogenous DNA fragment was detected in the wild type (Figure 1B and C). The genome-edited *Arabidopsis thaliana* was sequenced with a depth of 15-fold, and analysis results showed that the genome-edited *Arabidopsis* contains exogenous DNA fragments most similar to the plasmid including common vector component sequence of Cas9 (D10A), CaMV 35S promoter and HygR (Figure S7 in Supporting Information). At the same time, the genome-edited wheat, maize and tomato were also sequenced with a depth of 10-fold, and exogenous DNA fragments most similar to the plasmid pBUE411, pRGEB32.BAR and pK7WGF2::hCas9 in the vector library, respectively (Figure S7 in Supporting Information). To test the ability of FED in detecting exogenous DNA fragments without vector information in genome-edited animals, the whole genome sequencing data from genome-edited hornless bull was downloaded and analyzed. Analysis results showed that the genome-edited bull contains exogenous DNA fragments most similar to the plasmid including common vector component sequence of T7 promoter, NeoR/KanR and Ampicillin (Figure S7 in Supporting Information). The result is consistent with previous report (Young et al., 2020).

In summary, we develop a new web application tool that is a user-friendly, highly sensitive and open source tool designed to identify the exogenous DNA fragments using WGS data. By taking the advantage of NGS technologies, FED can be used for the detection of tens of thousands of exogenous components, which ensures the safety of genome-edited products in agriculture.

**Compliance and ethics** *The authors filed a patent application based on the results reported in this paper.*

**Acknowledgements** *We thank Caixia Gao and Huawei Zhang for providing the Arabidopsis materials. This work was supported by the National Transgenic Science and Technology Program (2019ZX08010-003), the National Key Research and Development Program of China (2017YFD0102002), the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences, and the National Natural Science Foundation of China (31901523).*

## References

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

Dörries, H.H., Remus, I., Grönewald, A., Grönewald, C., and Berghof-Jäger, K. (2010). Development of a qualitative, multiplex real-time PCR kit for screening of genetically modified organisms (GMOs). Anal Bioanal Chem 396, 2043–2054.

Huang, S., Weigel, D., Beachy, R.N., and Li, J. (2016). A proposed regulatory framework for genome-edited crops. Nat Genet 48, 109–111.

Huang, W., Li, L., Myers, J.R., and Marth, G.T. (2012). ART: a next-generation sequencing read simulator. Bioinformatics 28, 593–594.

Kent, W.J. (2002). BLAT—The BLAST-like alignment tool. Genome Res 12, 656–664.

Li, G., Liu, Y.G., and Chen, Y. (2019). Genome-editing technologies: the gap between application and policy. Sci China Life Sci 62, 1534–1538.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Puchta, H. (2018). Broadening the applicability of CRISPR/Cas9 in plants. Sci China Life Sci 61, 126–127.

Skryabin, B.V., Kummerfeld, D.M., Gubar, L., Seeger, B., Kaiser, H., Stegemann, A., Roth, J., Meuth, S.G., Pavenstädt, H., Sherwood, J., et al. (2020). Pervasive head-to-tail insertions of DNA templates mask desired CRISPR-Cas9–mediated genome editing events. Sci Adv 6, eaax2941.

Wang, C., Liu, Q., Shen, Y., Hua, Y., Wang, J., Lin, J., Wu, M., Sun, T., Cheng, Z., Mercier, R., et al. (2019). Clonal seeds from hybrid rice by simultaneous genome engineering of meiosis and fertilization genes. Nat Biotechnol 37, 283–286.

Young, A.E., Mansour, T.A., McNabb, B.R., Owen, J.R., Trott, J.F., Brown, C.T., and Van Eenennaam, A.L. (2020). Genomic and phenotypic analyses of six offspring of a genome-edited hornless bull. Nat Biotechnol 38, 225–232.

## SUPPORTING INFORMATION

The supporting information is available online at https://doi.org/10.1007/s11427-020-1731-9. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.