Dear Editor,

In higher eukaryotes, alternative splicing (AS) and alternative polyadenylation (APA) events can produce multiple transcript isoforms in the majority of genes, which significantly increase the protein-coding potential of a genome (Anvar et al., 2018; Pan et al., 2008). Different transcript isoforms might encode proteins with different functions or affect the mRNA stability and translational capacity, in some sense AS and APA events can dramatically increase the complexity and flexibility of the entire transcriptome and proteome (Feng et al., 2015; Li et al., 2017a; Wang et al., 2017a; Yang et al., 2016). Many databases contained AS events and transcripts in animals are available in some public resources such as ASTD and MAASE (Zheng et al., 2015), whereas there is no database containing full-length transcripts and AS events in plants up to now. Next-generation sequencing (NGS) technology has limitation for identifying AS and APA events due to short reads and low accuracy. In recent years, isoform sequencing (iso-Seq) using Pacbio single molecule real-time sequencing (SMRT) platform can generate full-length sequences and provide accurate information about AS and transcriptional start sites (Li et al., 2017a). In this study, we collected the plant iso-Seq data sequenced by Pacbio platform from NCBI database up to the end of 2017, and employed unified pipelines to process all the full-length transcripts in different species. Based on these data, we constructed Plant ISOform sequencing database (PISO, http://cbi.hzau.edu.cn/piso/).

Plant ISOform sequencing database was performed on Linux operation system and Apache web server (http://www.apache.org/). The obtained transcripts, AS events, novel genes and analysis tools were organized and stored in MySQL database (http://www.mysql.com/). Bootstrap framework (https://getbootstrap.com/) and jQuery (https://blog.jquery.com/) were applied for constructing the website, and PHP was used to interact with back-end data and ECharts (http://echarts.baidu.com/Highcharts (http://www.hcharts.cn/)) for data visualization. Currently, there are 19 plant species listed in PISO (i.e. Amborella trichopoda, Arabidopsis thaliana, Beta vulgaris subsp. Vulgaris, Chenopodium quinoa, Coffea arabica, Fragaria vesca, Gossypium barbadense, Hevea brasiliensis, Panax ginseng, Phyllostachys edulis, Sorghum bicolor, Triticum aestivum, Zea mays, Allium sativum, Astragalus membranaceus, Dipterix oleifera, Nepenthes ampullaria, Nepenthes rafflesiana and Salvia miltiorrhiza), of which 13 have reference genomes (nine diploids and four allopolyploids), while the other six species do not have genome sequences. We employed three pipelines for analysing iso-Seq data of the above three types of plant species (Figure 1a). Firstly, standard TAPIS pipeline was utilized for nine diploids with reference genomes (Abdelghany et al., 2016; Wang et al., 2017a); secondly, two extra steps, phasing and adjusting, were added into TAPIS pipeline for studying the four allopolyploid species owing to the large number of homoeologous genes from different sub-genomes (Wang et al., 2017b); thirdly, for analysing the other six species without reference genome, high-quality reads were mapped to a pseudo-genome generated by Cogent software (Li et al., 2017b).

Plant ISOform sequencing database provided millions of high quality transcript isoforms. In total, 1 391 165 transcripts, 50 803 novel gene loci, 878 057 AS and 81 416 APA events were obtained from 19 plant species. Different types of AS events, i.e. intron retention (IntronR), exon skipping (ExonS), alternative donor site (AltD), alternative acceptor site (AltP), alternative position (AltP) and other types of AS were provided for each species. For instance,
the genome of *Amborella trichopoda* in Ensembl Plants database contains 27,313 protein-coding gene loci and only 27,313 transcripts. In PISO, 34,733 transcripts, 15,039 AS events and 3,315 APA events were obtained for the detected 9,060 expressed protein-coding gene loci, and 769 novel gene loci were identified and listed. Therefore, transcripts and novel protein-coding gene loci gained in PISO greatly expanded the current genome annotation. Simultaneously, PISO provided functional annotation and protein sequence for each transcript.

Plant ISOform sequencing database has a flexible user interface. A great number of interactive graphs generated by JavaScript were used to display significant results. For example, the doughnut chart

---

Figure 1  The architecture, pipeline and some screenshots in PISO. (a) The architecture and pipeline of PISO. (b) Species homepage. (c) Alternative splicing search. (d) Transcript browser results for one transcript. (e) Transcript browser results for one gene. (f) BLAST results. (g) Full-length match results. (h) GBrowse.
illustrated the distribution of different types of AS events in each species (Figure 1b). Meanwhile, AS statistics, GO statistics and browser of novel genes were integrated in ‘Species’ page. Locations on chromosome and gene names can be used to identify the AS and APA events on ‘Alternative Splicing’ page (Figure 1c). Several bar charts demonstrated the statistics of APA positions and AS events for each gene. Besides, the customized graphs showed the relative positions of different sequences by using tools of ‘Functional Annotation’, BLAST, ‘Full-length Match’ and ‘Transcript Browser’. Protein functional annotation linked to other databases and the sequence of each transcript for a given gene was displayed on ‘Alternative Splicing’ page. At the same time, PISO also supplied annotation corresponding to different transcripts (Figure 1e).

Several convenient tools were provided in PISO, such as Functional Search, BLAST, Full-length Match and GBrowse. In ‘Functional Search’ section, transcripts can be retrieved by InterPro accession, GO ID, Pfam accession, transcript name, gene name, etc., the description of different transcripts can also be exhibited in this page. After selecting sequence and parameters, BLAST would return detailed results (Figure 1f). Moreover, PISO provided a useful tool, Full-Length Match, to search for the best full-length transcript that matched the user-uploaded sequence (Figure 1g). Based on a new sequence, gene name or transcript name can be obtained through Blast or ‘Full-Length Match’. A local GBrowse offered user with manipulating and displaying annotation on genomes. Detailed information of specific regions on the chromosome can be accessed by GBrowse (Figure 1h).

In summary, we constructed a significant bioinformatics platform, PISO, to provide a comprehensive repertory of sequenced Iso-Seq full-length transcripts for 19 plant species. PISO possesses a flexible user interface to display transcripts, novel genes, AS and APA events. Furthermore, it involves the pipeline, guide document and reference links. With the widespread application of Pacbio SMRT-based sequencing, PISO will be continuously updated and provide more valuable information for researchers. PISO is freely available at http://cbi.hzau.edu.cn/piso/.

**Acknowledgements**

This work was supported by the National Key Research and Development Program of China (2018YFD1000101), the National Natural Science Foundation of China (31871269, 31571351) and the Fundamental Research Funds for the Central Universities (2662017PY043). No conflict of interest declared.

**References**

Abdelghany, S.E., Hamilton, M., Jacobi, J.L., Ngam, P., Devitt, N., Schilkey, F., Benhur, A. et al. (2016) A survey of the sorghum transcriptome using single-molecule long reads. Nat. Commun. 7, 11706.

Anvar, S.Y., Allard, G., Tseng, E., Sheynkman, G.M., de Klerk, E., Vermaat, M., Yin, R.H. et al. (2018) Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. Genome Biol. 19, 46.

Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., Yan, S. et al. (2015) SKIP confers osmotic tolerance during salt stress by controlling alternative gene splicing in Arabidopsis. Mol. Plant. 8, 1038–1052.

Li, Y., Dai, C., Hu, C., Liu, Z. and Kang, C. (2017a) Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. Plant J. 90, 164.

Li, J., Haratalee, Y., Denton, M.D., Feng, Q., Rathjen, J.R., Qu, Z. and Adelson, D.L. (2017b) Long read reference genome-free reconstruction of a full-length transcriptome from Astragalus membranaceus reveals transcript variants involved in bioactive compound biosynthesis. Cell Discov. 3, 17031.

Pan, Q., Shao, D., Lee, L.J., Frey, J. and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. 40, 1413–1415.

Wang, T., Wang, H., Cai, D., Gao, Y., Zhang, H., Wang, Y., Lin, C. et al. (2017a) Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (Phyllostachys edulis). Plant J. 91, 684–699.

Wang, M., Wang, P., Liang, F., Ye, Z., Li, J., Shen, C., Pei, L. et al. (2017b) A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. New Phytol. 217, 163.

Yang, X., Coulombe-huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., Sun, S. et al. (2016) Widespread expansion of protein interaction capabilities by alternative splicing. Cell, 164, 805–817.

Zheng, C.L., Kwon, Y.S., Li, H.R., Zhang, K., Coutinho-Mansfield, G., Yang, C., Nair, T.M. et al. (2005) MAASE: an alternative splicing database designed for supporting splicing microarray applications. RNA, 11, 1767.