Brief Communication

eRice: a refined epigenomic platform for japonica and indica rice

Pingxian Zhang†, Yifan Wang*, Sadaruddin Chachar, Jian Tian and Xiaofeng Gu†
Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China

Received 26 September 2019; revised 30 November 2019; accepted 26 December 2019.
*Correspondence (Tel +86-010-82105326; fax +86-010-82105326; emails guxiaofeng@caas.cn (X.G), tianjian@caas.cn (J.T))
†These authors contributed equally to this work.

Keywords: eRice, rice (Oryza sativa), epigenomic database, DNA methylation, N6-methyldeoxyadenosine (6mA), histone modification.

Epigenetic modifications including histone modifications and DNA methylation influence various biological processes in multicellular organisms. DNA 5-methylcytosine (5mC) is the most widely studied DNA modification mark in eukaryotes and is involved in modulating the activities and functions of developmental signals (Wu and Zhang, 2017). Recent studies have shown that the previously unknown epigenetic mark DNA N6-methyldeoxyadenosine (6mA) is widely distributed throughout the genomes of model animals such as Drosophila (Zhang et al., 2015), and throughout the human genome (Xiao et al., 2018). In land plants, the distribution patterns and potential functions of 6mA sites were largely undiscovered until recent papers reported genome-wide 6mA sites in dicot Arabidopsis and monocot rice (Liang et al., 2018; Zhang et al., 2018). A previous study revealed the 6mA methylome in the two main rice cultivars japonica Nipponbare (Nip) and indica 93-11, which have been at single-nucleotide resolution using single-molecule real-time (SMRT) sequencing (Zhang et al., 2018). Analysis of the genomic distribution of 6mA and its biological functions in rice genomes showed that 6mA is associated with gene expression, plant development and stress responses. Until now, an epigenomic database especially for rice, particularly for DNA methylation has not been available. Here, we describe the species-specific epigenomic annotation database, eRice (http://www.elabcaas.cn/rice/index.html), that will facilitate efficient annotation of epigenomic data for both japonica and indica rice. The eRice database integrates DNA methylation data for 6mA and 5mC at single-base resolution, artificial intelligence (AI)-based 6mA predictions, histone modifications, genomic, transcriptomic resources in the Nip and 93-11 reference cultivars. The epigenomic information about the distributions of epigenetic markers, especially 6mA, across rice genomes will enhance our understanding of the epigenetic regulation of complex biological processes in plant development for future breeding by molecular design.

The eRice database is dedicated to providing efficient and reliable epigenomic and genomic resources for both japonica (Nip) and indica (93-11) rice cultivars. Briefly, we have stored refined, publicly available genomic and epigenomic resources including the latest reference genomes, 6mA, 5mC and transcriptomes of 3-week-old rice seedlings under short day (SD) conditions (Zhang et al., 2018), and various histone modifications of rice young seedlings or leaves under SD conditions (Lu et al., 2015; Zhang et al., 2012). The eRice database also provided 6mA prediction sites based on the deep-learning approaches (Washburn et al., 2019). The schematic structure (Figure 1a) and the home page (Figure 1b) of eRice are shown, and eRice includes gene annotation, DNA methylation, multi-JBrowse and BLAST pages for performing gene information and analysing DNA methylation.

The eRice genomic annotation page provides a flexible interface for efficient retrieval and graphical visualization of genome-wide annotation data and sequence information (Figure 1c). On this page, a keyword-based search engine is provided to allow searches for all relevant genes by first selecting the most closely related reference genome Nip or 93-11 as the defaults and entering a locus identifier (currently only supporting the full-name locus ID) or functional keyword (e.g. demethylase). Links to pages with detailed information on gene location; gene structure; gene annotations; nucleotide and amino acid sequences; and gene expression data are then provided. In this section, details about 6mA and 5mC sites associated with the candidate gene are also shown. A more detailed description of the 6mA and 5mC information available on the eRice DNA methylation page follows.

The eRice DNA methylation page provides the first genome-wide epigenomic data resource for the distributions of 6mA and 5mC (Figure 1d) in rice. This information has been refined to permit searching by selecting a candidate chromosome and entering a site of interest (from the starting site to the ending site); which then provides links to more detailed pages. On the detailed page for 6mA, detailed methylation parameters such as chromosomal position, DNA strand, fraction score and a 20-bp reference sequence above and below the given methylation site may be viewed. In contrast, the detailed page for 5mC methylation sites of interest shows methylation types (CG, CHG and CHH) but not detailed methylation parameters. Moreover, genomic sequences at a given methylation site that show, for instance, whether the methylation site is within a gene or in an intergenic region are also available.

Deep-learning approaches or AI methods, have been largely responsible for a recent paradigm shift in image and natural language processing. These deep-learning methodologies are being applied to biological problems in agriculture and genetics (Washburn et al., 2019). Here, by leveraging previous deep-learning methods and open-source code (Washburn et al., 2019), we have developed AI models trained with previously released 6mA data for predicting and memorizing on both the motifs and the functional evolutionary history of 6mA in rice (Figure 1e). Briefly, we first retrieved 41-bp sequences (including 20-bp reference sequences up- and downstream from an A base) in the whole rice genome, then encoded the sequences with the one-
The eRice: a refined epigenomic database

(a) The eRice system overview:
- Geno-seq
- SMRT-seq
- BS-seq
- Artificial Intelligence
- ChIP-seq
- RNA-seq

(b) Gene Annotation:
- Input: LOC ID or Keywords
- Output: Sequence information

(c) BLAST:
- Input: Sequence info
- Output: Predicted 64A Aki

(d) Methylation analysis overview:
- C5H H3K27me3
- H3K4me3
- H3K4me1
- H3K36me3
- H3K27ac
- H3K9me3

(e) 6mA predictor overview:
- Type/Potential gene sequence with the fasta format below:

```
AGT
```

(f) Available Tools:
- eRice
- eRice DB
- eRice BLAST

(g) genos:
- Primary Data:
  - Name: LOC_01046100
  - Type: gene
  - Position: start-1593161, stop-1603161 (in genome)
  - Length: 100 bp
- Attributes:
  - id: LOC_01046100
  - strand: 1
  - Source: eRice
  - Region: 5'UTR
hot approach ($A = 1000, G = 0100, T = 0010, C = 0001$). Our prediction models were constructed in Python 2.7 using Keras 2.2.4 with a Tensorflow back end. The final architecture consists of two convolutional layers, with each group of layers followed by a maximum pooling and a dropout layer, and a final prediction layer. A ‘relu’ activation function was used for each layer in the model (except for the final prediction layer, which used a Softmax activation function depending on the model) (Washburn et al., 2019). Using these trained deep-learning models, we tested 6mA in two Nip and 93-11 cultivars for predicting 6mA site from the whole rice genome region (Loss = 0.125, Accuracy = 0.958 and AUC (Area Under Curve) = 0.989, which was calculated with 10-folds cross validation). This architecture will allow rice researchers to efficiently query the eRice database and predict potential 6mA sites in targeted genes or regions.

The eRice database runs JBrowse (Buels et al., 2016), a powerful genomic data tool for visualizing the Nip and 93-11 genomes. Different tracks can be chosen to review genome-wide reference sequences and annotation information, 6mA methylation data and AI-based 6mA prediction data for different genomic regions (Figure 1f and g). Moreover, we have added genome-wide epigenomic resources for single-base-resolution 5mC and various histone modifications to facilitate tracking of associations between this epigenomic information and other ‘omics-based data. Further, the ViroBLAST (Deng et al., 2007) tool allows searches of nucleotide or amino acid sequences for candidate homologs in the reference genome and CDS (coding DNA sequence) data and shows epigenetic modification sites with alignment results in text format. Setting the desired parameters using the BLAST tool is enabled for advanced searches. All of the genomic and epigenomic data in the database may be freely accessed and some useful links to other databases have also been added.

Adenine methylation on the sixth position of the purine ring has been regarded as a sixth base that could be important in plant development and environmental responses (Liang et al., 2018; Zhang et al., 2018). Our eRice database has been established as an extensive bioinformatics platform providing epigenomic resources (especially for 6mA) and genomic annotation of the reference cultivars, Nip and 93-11, respectively. Genomic annotations and epigenomic resources can be downloaded at the eRice download page. Owing to deep-learning methods, which are being widely applied to questions in agricultural and genetic sciences (Washburn et al., 2019), we could intelligently predict genome-wide distribution of 6mA sites in rice genomes. With the multiple epigenomic, other ‘omics-based data and AI-prediction resources of eRice available through a user-friendly website, we would like eRice to serve as an efficient tool for epigenetic design of rice traits involving developmental cues (such as heading date, yield and quality) and responses to environmental stimuli (such as drought response, ambient temperature and salt stress) in rice. To keep up with the latest advances and offer more analysis tools for rice plants, we will continue to supplement the eRice database with additional epigenetic data including for RNA methylation, non-coding RNA and various histone modifications, as well as other ‘omics-based data, to extend the functionality of the eRice database and make it a convenient electronic platform for the community of rice researchers.

Acknowledgments

This work was supported by National Transgenic Major Program of China (2019ZX08010-002), National Natural Science Foundation of China (31871606) and Fundamental Research funds for Central non-profit Scientific Institution (1610392017001) to X.G.

Conflict of interest

No conflict of interest declared.

Author contributions

X.G. and J.T. conceived and designed the study. J.T., Y.W. and P.Z. constructed the website. S.C. contributed to the writing. X.G., J.T. and P.Z. wrote the paper.

References

Buels, R., Yao, E., Densh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G. et al. (2016) JBrowse: A dynamic web platform for genome visualization and analysis. Genome Biol. 17, 66.

Deng, W., Nickle, D.C., Learn, G.H., Maust, B. and Mullins, J.J. (2007) ViroBLAST: A stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. Bioinformatics, 23, 2334–2336.

Liang, Z., Shen, L., Cui, X., Yu, G., Geng, Y., Bao, S. et al. (2018) DNA N6-adenine methylation in Arabidopsis thaliana. Dev. Cell 45, 406–416.

Lu, L., Chen, X., Sanders, D., Qian, S. and Zhong, X. (2015) High-resolution mapping of H4K16 and H3K23 acetylation reveals conserved and unique distribution patterns in Arabidopsis and rice. Epigenetics, 10, 1044–1053.

Washburn, J.D., Mejia-Guerra, M.K., Ramstein, G., Kreming, K.A., Valluru, R., Bucker, E.S. and Wang, H. (2019) Evolutionarily informed deep learning methods for predicting relative transcript abundance from DNA sequence. Proc. Natl. Acad. Sci. USA 116, 5542–5549.

Wu, X. and Zhang, Y. (2017) TET-mediated active DNA demethylation: Mechanism, function and beyond. Nat. Rev. Genet. 18, 517–534.

Xiao, C., Zhu, S., He, M., Chen, D., Zhang, Q., Chen, Y. et al. (2018) N6-methyladenine DNA modification in the human genome. Mol. Cell 71, 306–318.

Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Feehling, M., Crawford, G.E. and Jiang, J. (2012) High-resolution mapping of open chromatin in the rice genome. Genome Res. 22, 151–162.

Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W. et al. (2015) N6-methyladenine DNA modification in Drosophila. Cell, 161, 893–906.

Zhang, Q., Jiang, Z., Cui, X., Ji, C., Li, Y., Zhang, P. et al. (2018) N6-methyladenine DNA methylation in japonica and indica rice genomes and its association with gene expression, plant development, and stress responses. Mol. Plant 11, 1492–1508.