

# Rice Information GateWay: A Comprehensive Bioinformatics Platform for *Indica* Rice Genomes

Dear Editor,

*Oryza sativa* subsp. *indica* and *japonica* are two subspecies of Asian cultivated rice, among which *indica* rice is much more widely grown and genetically diverse. Over the past years, the Rice Annotation Project Database (RAP-DB) (Ohyang et al., 2006) and Michigan State University Rice Genome Annotation Project (MSU-RGAP) (Ouyang et al., 2007) are two popular databases that have been developed to manage rice genomic and transcriptomic data based on the unified reference genome of *japonica* cultivar Nipponbare (International Rice Genome Sequencing Project, 2005). Beijing Genomics Institute Rice Information System (BGI-RIS) (Zhao et al., 2004) is an available resource for *indica* rice cultivar 93-11; however, its application was limited due to the lack of high-quality *indica* reference genomes. To fill the gap, we constructed an integrative and comprehensive platform, Rice Information GateWay (RIGW, <http://rice.hzau.edu.cn/>), to provide genomics, transcriptomics, protein-protein interactions (PPIs), metabolic network, metabolites, and computational tools by using our newly obtained map-based reference genomes of *indica* rice Zhenshan 97 (ZS97) and Minghui 63 (MH63) (Zhang et al., 2016). RIGW serves the rice community by making a wealth of genomics and other omics data available through an intuitive web-based interface.

RIGW was implemented in the Linux operation system and Apache Tomcat web server (<http://tomcat.apache.org/>). All the genomic data, annotation, homologs, gene expression, PPIs, metabolites, and literature are organized and stored in a MySQL database (<http://www.mysql.com/>). The architecture, some representative resources, and computational tools in RIGW are shown in Figure 1A. A local GBrowse (<https://github.com/GMOD/GBrowse>) is deployed for visualization of ZS97 and MH63 genomic and transcriptomic data (Figure 1B), and the selected tracks include gene annotation and RNA-sequencing evidence in flag leaf, panicle, and shoot, respectively. In addition, comparative analysis among ZS97, MH63, and Nipponbare genomes is supplied with the Gbrowse\_synteny tool (Figure 1C), and all the syntenic regions with corresponding annotations can be easily displayed in parallel (each genome can be set as a reference).

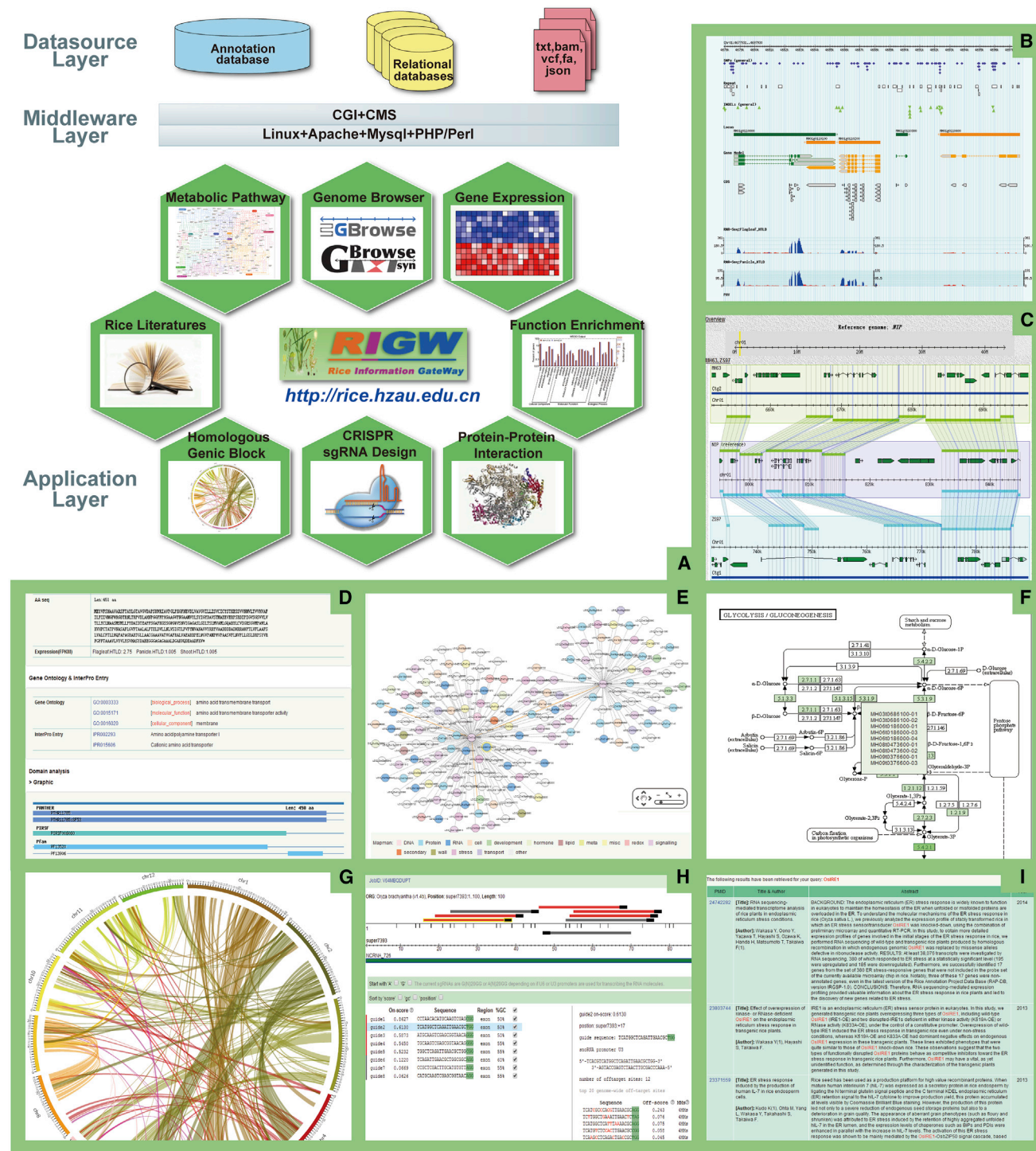
We developed a flexible query interface to retrieve and graphically visualize various data efficiently. For example, a keyword-based search engine is provided to look for all relevant genes by entering keywords (e.g., gene locus, gene function) and linked to detailed pages (e.g., gene location, gene structure, alternative splicing, homologs in other rice cultivars, nucleotide and amino acid sequences, gene expression levels, etc.). In addition, Gene Ontology (Harris et al., 2004), InterPro domain information, predicted subcellular location, and PPI, as well as extra links to external databases, are also listed in the search results if available

(Figure 1D). The Basic Local Alignment Search Tool (BLAST) is supplied as a sequence-based search engine to find homologs in *indica* rice ZS97, MH63, 93-11, and *japonica* rice Nipponbare by presenting alignment results in graphical and text format.

Characteristics of ZS97 and MH63 genomic features are listed in Supplemental Table 1 and the RIGW homepage. We manually collected >2000 cloned genes in different rice cultivars with related references and >2500 rice metabolites with detailed annotation information. Using the data from CREP (<http://crep.ncpgr.cn/>), we set up a friendly web interface to query and visualize gene expression levels of 39 tissues throughout the life cycle of ZS97, MH63, and their hybrid Shanyou 63 (SY63). For a given gene, its expression profile in all available tissues was visualized in a boxplot, which greatly facilitated the investigation of its expression pattern. As genome-wide PPI networks are very useful to study cellular behavior with a global view, we collected 1 871 563 non-redundant rice PPIs (929 of them are experimentally determined PPIs) from public databases including PRIN (Gu et al., 2011), RiceNet (Lee et al., 2015), and related literature in RIGW. Users can submit one or more gene IDs of ZS97/MH63/Nipponbare on the PPI search page, then the server will return proteins that interact with the query proteins, which can help to reveal the relationships among different kinds of proteins with various functions. The query protein and its interaction partners are visualized with Cytoscape (<http://www.cytoscape.org/>) and different color nodes represent different pathway classifications (Figure 1E). In addition, all the PPIs in Nipponbare, ZS97, and MH63 can be downloaded from the download module.

KEGG metabolic pathway maps are graphical diagrams representing knowledge of reaction networks for metabolism, and each map summarizes experimental evidence in published literature (Kanehisa et al., 2012). Based on KEGG Orthology (KO) groups, we obtained the KEGG orthologs in ZS97 and MH63 genomes and generated their metabolic pathways. KEGG modules in each pathway map can be generated by converting nodes to gene identifiers and are highlighted in green. Metabolic pathways in ZS97 and MH63 contain four categories (i.e., metabolism, genetic information processing, environmental information processing and cellular processes) and each category contains many pathways. When a specific pathway is selected, the enzymes/proteins that have KEGG orthologs in ZS97 and MH63 are indicated in green (Figure 1F).

A series of computational tools are integrated in RIGW for comparative, evolutionary, and functional analysis of rice



**Figure 1. The RIGW Architecture and Some Screenshots of Representative Resources and Computational Tools.**

- (A) The three-layer architecture of RIGW.
- (B) GBrowse.
- (C) GBrowse-synteny.
- (D) Gene information page.
- (E) Predicted PPIs.
- (F) KEGG pathway.
- (G) Segmental duplications in MH63 genome.
- (H) Single guide RNAs designed with CRISPR-P 2.0.
- (I) Literature search results.

and other plants. OrthoMCL (Li et al., 2003) was used to identify homologs in plant genomes, including *Arabidopsis*, *Brachypodium*, maize, grapevine, and sorghum. We performed OrthoMCL (e-value:  $1e-5$ ) to identify putative orthology and inparalogy relationships and generated disjoint clusters of closely related proteins in rice and the above plants. A total of 48 515 putative orthologous groups including inparalogs were identified and stored in RIGW, and can also be obtained from the download module. We defined homologous gene pairs with MCScanX (Wang et al., 2012) among chromosomes (e-value  $< 1e-10$ ) and determined the homologous genic blocks to show the segmental duplicated regions in ZS97 and MH63 genomes (Figure 1G). We also provide a gene ID conversion tool to convert the orthologous gene IDs among ZS97, MH63, 93-11, and Nipponbare. Furthermore, KEGG/GO enrichment, GO slim classification tools are supplied to perform functional enrichment analysis. For the convenience of genome editing in different rice cultivars, we integrated CRISPR-P 2.0 (Liu et al., 2017) to design guide RNA sequences for various Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas systems, and an example result is shown in Figure 1H. Lastly, a text mining tool is available in RIGW, which allows a user to search papers by gene names or keywords in 27 831 rice-related articles obtained from PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>) (Figure 1I).

In summary, we have established a comprehensive bioinformatics platform, RIGW, to provide a GBrowse-based view of ZS97 and MH63 genomic and other omics data. RIGW offers homologs among *indica* and *japonica* rice, and other plant species. We also provide user-friendly web interfaces to show the predicted PPIs in rice, the metabolic pathways in ZS97/MH63, a CRISPR-Cas single guide RNA design tool, and GO enrichment in RIGW. All the genomic sequences and annotation can be freely accessed, and useful links to other public databases are offered. In the near future, we will integrate more available resources and extend its functionality with new tools to make RIGW as a comprehensive bioinformatics platform for rice community. RIGW is freely available at <http://rice.hzau.edu.cn/>.

## SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

## FUNDING

This work was supported by the National Key Research and Development Program of China (2016YFD0100904), the National Natural Science Foundation of China (31571351) and the National Science Foundation of Hubei Province (2015CFA044).

## AUTHOR CONTRIBUTIONS

J.-M.S. and Y.L. constructed the website; J.-M.S., C.-C.S., Y.D., F.X., H.L., J.W., and W.X. collected the data; L.-L.C. and J.Z. designed the project and wrote the paper.

## ACKNOWLEDGMENTS

No conflict of interest declared.

Received: July 31, 2017

Revised: September 19, 2017

Accepted: October 8, 2017

Published: October 14, 2017

Jia-Ming Song<sup>1,2</sup>, Yang Lei<sup>1,2</sup>,  
Cheng-Cheng Shu<sup>1,2</sup>, Yudian Ding<sup>1,2</sup>,  
Feng Xing<sup>1,2</sup>, Hao Liu<sup>1</sup>, Jia Wang<sup>2</sup>,  
Weibo Xie<sup>1,2</sup>, Jianwei Zhang<sup>1,\*</sup> and  
Ling-Ling Chen<sup>1,2,\*</sup>

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, P. R. China

<sup>2</sup>Agricultural Bioinformatics Key Laboratory of Hubei Province, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

\*Correspondence: Jianwei Zhang ([jzhang@mail.hzau.edu.cn](mailto:jzhang@mail.hzau.edu.cn)), Ling-Ling Chen ([llchen@mail.hzau.edu.cn](mailto:llchen@mail.hzau.edu.cn))

<https://doi.org/10.1016/j.molp.2017.10.003>

## REFERENCES

- Gu, H., Zhu, P., Jiao, Y., Meng, Y., and Chen, M. (2011). PRIN: a predicted rice interactome network. *BMC Bioinformatics* **12**:161.
- Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**:D258–D261.
- International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature* **436**:793–800.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**:D109–D114.
- Lee, T., Oh, T., Yang, S., Shin, J., Hwang, S., Kim, C.Y., Kim, H., Shim, H., Shim, J.E., Ronald, P.C., et al. (2015). RiceNet v2: an improved network prioritization server for rice genes. *Nucleic Acids Res.* **43**:W122–W127.
- Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**:2178–2189.
- Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., and Chen, L.L. (2017). CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Mol. Plant* **10**:530–532.
- Ohyanagi, H., Tanaka, T., Sakai, H., Shigemoto, Y., Yamaguchi, K., Habara, T., Fujii, Y., Antonio, B.A., Nagamura, Y., Imanishi, T., et al. (2006). The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.* **34**:D741–D744.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., et al. (2007). The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* **35**:D846–D851.
- Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**:e49.
- Zhang, J., Chen, L.L., Xing, F., Kudrna, D.A., Yao, W., Copetti, D., Mu, T., Li, W., Song, J.M., Xie, W., et al. (2016). Extensive sequence divergence between the reference genomes of two elite *indica* rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl. Acad. Sci. USA* **113**:E5163–E5171.
- Zhao, W., Wang, J., He, X., Huang, X., Jiao, Y., Dai, M., Wei, S., Fu, J., Chen, Y., Ren, X., et al. (2004). BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res.* **32**:D377–D382.