

New Data and New Features of the FunRiceGenes (Functionally Characterized Rice Genes) Database: 2021 Update

Fangfang Huang

Henan Agricultural University

Yingru Jiang

Henan Agricultural University

Tiantian Chen

Henan Agricultural University

Haoran Li

Henan Agricultural University

Mengjia Fu

Henan Agricultural University

Yazhou Wang

Henan Agricultural University

Yufang Xu

Henan Agricultural University

Yang Li

Henan Agricultural University

Zhengfu Zhou

Henan Academy of Crop Molecular Breeding, Henan Academy of Agricultural Sciences

Lihua Jia

Henan Agricultural University

Yidan Ouyang

Huazhong Agricultural University

Wen Yao (✉ yaowen@henau.edu.cn)

Henan Agricultural University

Research Article

Keywords: rice, *Oryza sativa*, functionally characterized genes, functional genomics, database.

Posted Date: December 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1201297/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Rice on April 19th, 2022. See the published version at <https://doi.org/10.1186/s12284-022-00569-1>.

Abstract

As a major food crop and model organism, rice has been mostly studied with the largest number of functionally characterized genes among all crops. We previously built the funRiceGenes database including ~2800 functionally characterized rice genes and ~5000 members of different gene families. Since being published, the funRiceGenes database has been accessed by more than 49,000 users with over 490,000 page views. The funRiceGenes database has been continuously updated with newly cloned rice genes and newly published literature, based on the progress of rice functional genomics studies. Up to Nov 2021, ≥4100 functionally characterized rice genes and ~6000 members of different gene families were collected in funRiceGenes, accounting for 22.3% of the 39,045 annotated protein-coding genes in the rice genome. Here, we summarized the update of the funRiceGenes database with new data and new features in the last five years.

Background

Rice (*Oryza sativa*) is a major food crop for almost half of the world population. Identification of functional genes governing the phenotypes of complex agronomic traits in rice is crucial to safeguard world food security. Rice is also a model organism for plant genomics research, benefiting from its small genome size with accurate genome sequence and a great number of molecular markers, as well as the availability of high-efficiency transgenic systems. Most studies on the functions of rice genes can be directly applied to homologous genes in other cereals and crops. Accordingly, it is important for functional studies and genetic improvement of rice and other crops to collect and build a detailed archive of all functionally characterized rice genes. We previously constructed funRiceGenes, a database that provides the most detailed information of ~2800 functionally characterized rice genes (Yao et al., 2017). Since being released, this database has been continuously updated based on the latest research progress in rice. Up to November 2021, over 4,100 functionally characterized rice genes and more than 6,000 gene family members were collected in the funRiceGenes database, providing an important resource for functional genomics research and molecular marker-assisted breeding in rice. In addition, new features were developed for better user experiences and deeper utilization of the funRiceGenes database.

Results

Application of the funRiceGenes database in rice and other crops

Since being published in December 2017, the funRiceGenes database has been visited by more than 49,000 users, with more than 129,000 sessions and over 490,000 page views (Figure 1) (Yao et al., 2017). An average of 3~4 pages was accessed per session. The number of users and page views of the funRiceGenes database were steadily rising in the last five years. Most of the sessions were from countries including China, the United States, India, Japan, and South Korea. To be noted, users from China, India, Japan, and South Korea accounted for more than 50% of all accessions, consistent with the reality that rice is the main food in Asia. Nevertheless, visits from the United States accounted for 13.3%

of all accessions, ranking second among all countries. An in-depth investigation revealed that the funRiceGenes database was frequently visited by several cities including Beijing, New Delhi, Hangzhou, Wuhan, Seoul, and Shanghai, which was probably attributed to well-known rice research institutions located in these cities, including the Institute of Genetics and Developmental Biology in Chinese Academy of Sciences in Beijing; the China National Rice Research Institute in Hangzhou; the National Key Laboratory of Crop Genetic Improvement in Huazhong Agricultural University in Wuhan; the Institute of Plant Physiology and Ecology in Chinese Academy of Sciences in Shanghai, etc.

As of November 2021, the funRiceGenes database and the corresponding publication in *GigaScience* have been cited by 89 times (Yao et al., 2017). funRiceGenes provides the most accurate number of functionally characterized rice genes, which was referenced by many studies (Wing et al., 2018; Yang et al., 2020). The datasets in funRiceGenes were utilized to train models in bioinformatics studies (Gupta et al., 2021). The information of more than 4100 genes collected in funRiceGenes was frequently utilized to annotate the functions of gene sets obtained in many studies in rice (Kim et al., 2020; Qin et al., 2021; Fan Zhang et al., 2021). Furthermore, the funRiceGenes database was also used to disclose the potential functions of homologous genes in non-model plants including wheat, rye, sorghum, barley and pacaya palm (Hosni et al., 2021; Pang et al., 2020; Li et al., 2021; Wittern et al., 2021; Dhaka et al., 2020; Wang et al., 2021).

Features update of the funRiceGenes database

A static website (<https://funricegenes.github.io/>) was developed for browsing and searching functional rice genes deposited in funRiceGenes. Eight menus were deployed in the static website, including HOME, GENE, FAMS, KEYS, NEWS, DOCS, CITE, and LINK. The symbols of more than 4100 functionally characterized rice genes are listed in the GENE menu. Detailed information of a gene can be viewed by clicking on the corresponding gene symbol, including published articles related to the gene, MSU and RAPdb genomic locus of the gene, GenBank accession number, key information on the function of the gene, and information of related genes. The information of more than 6000 gene family members was recorded in the FAMS menu. Information of all members and related publications can be viewed by clicking the name of a specific gene family. Based on the published literature, we extracted more than 400 commonly used keywords concerning various functions of rice genes. All the keywords and the genes related to each keyword can be viewed in the KEYS menu, allowing rapid identification of all the functionally characterized genes related to a specific agronomic trait or keyword. The detailed updating records of the funRiceGenes database since being released in 2014 can be found in the NEWS menu. More than 7000 literatures on rice functional genomics studies are listed in the DOCS menu. To promote the usage of funRiceGenes, publications citing the funRiceGenes database are listed in the CITE menu. The LINK menu provides links to some useful plant databases and bioinformatics applications. Finally, key information including gene list, gene family list, keywords list, and other datasets collected in funRiceGenes can be downloaded from the HOME page of the database. To further facilitate the convenient usage of the static website, in-site searching was enabled with the help of Google and Bing, which were placed on the HOME page of the static website. Real-time visitor statistics of the static

website since Oct 29 2021 recorded by RevolverMaps (<https://www.revolvermaps.com/>) is also displayed on the HOME page.

We further developed an R/Shiny web application for interactive queries of the funRiceGenes database in the previous study (Jia et al., 2021; Yao et al., 2017). The URL of the interactive web application was moved from <http://funricegenes.ncpgr.cn/> to <https://venyao.xyz/funRiceGenes/>. In addition to query by gene symbols or keywords, the web application can be searched by the MSU or RAPdb genomic locus. In this study, we developed a new functionality under the 'Download' menu of the web application, which can be utilized to batch retrieve data from the funRiceGenes database by multiple user-input MSU or RAPdb genomic loci, or a user-input genomic region. The IDConversion functionality of the web application can be used to perform the conversion between MSU and RAPdb genomic locus, as well as the identification of orthologous genes between *indica* and *japonica* rice. To enable searching of the genic sequences, CDS, or protein sequences of all collected functional genes by sequence similarity, a BLAST interface was deployed in the web application, in this study. To further expand the application of the datasets collected in funRiceGenes, we developed a new feature under the 'Annotation' menu of the web application for functional annotation of gene sets obtained in experimental or high-throughput studies in rice. For an input gene set, a word cloud would be created utilizing key messages on the functions of input genes deposited in funRiceGenes.

Data update of the funRiceGenes database

Since being published in December 2017, the funRiceGenes database has grown by the addition of over 1300 newly cloned rice genes. More than 1950 published literature on these genes was deposited in funRiceGenes (Figure 2A). It was found that the keywords including rice, gene, regulates, protein, grain, and tolerance were enriched in the titles of these literature (Figure 2B). Similarly, the keywords including gene, plants, expression, stress, and grain were most frequently used in the abstracts of these literature (Figure 2C). These enriched keywords represent the research focuses on rice functional genomics studies in recent years. A list of representative functionally characterized rice genes is demonstrated in Figure 3 (Yiming Yu et al., 2019).

Understanding the genetic mechanisms underlying the size and weight of grains is critical to the improvement of rice yield. Up to November 2021, 162 genes regulating grain size and weight in rice were collected in funRiceGenes, including *Grain Size and Abiotic stress tolerance 1* (*GSA1*), which is a positive regulator of grain size (Dong et al., 2020). Overexpression of *GSA1* led to enhanced grain size and weight, as well as improved resistance to abiotic stresses, including high salt, drought, and high temperature. Unlike abiotic stress resistance genes, disease resistance genes usually lead to reduced yield in rice. Among all the 185 disease related genes collected in funRiceGenes, *Ideal Plant Architecture 1* (*IPA1*) is the first gene identified in rice that can promote yield and disease resistance at the same time by maintaining the balance between growth and disease resistance (Wang et al., 2018). *IPA1* binds to the promoter of the pathogen defense gene *WRKY45* after being infected by the fungus *Magnaporthe oryzae*, leading to enhanced blast disease resistance (Haitao Zhang et al., 2016). Within 48 hours after infection,

IPA1 binds to the promoter of *DENSE AND ERECT PANICLE1* (*DEP1*), a yield-related gene, to support the growth needed for high yield (Huang et al., 2009). In addition to blast disease, blight disease caused by *Xanthomonas oryzae* pv. *oryzae* has also put rice production under threat. A recent study reported that ALEX1, a long noncoding RNA whose expression is highly induced by *Xoo* infection, activated the JA signaling pathway and contributed to blight disease resistance in rice (Yang Yu et al., 2020).

Plant hormones are biochemicals playing critical roles in the regulation of all aspects of plant growth and development, including auxin, gibberellins (GA), abscisic acid (ABA), cytokinins (CK), salicylic acid (SA), ethylene (ET), jasmonates (JA), brassinosteroids (BR), and strigolactones. A total of 623 hormone related genes were collected in funRiceGenes, including 171 auxin related genes, 142 GA related genes, 126 ethylene related genes, 105 SA related genes, 101 JA related genes, 89 BR related genes, 68 CK related genes, and 21 strigolactone related genes. Among all the 623 genes, ACE1 (ACCELERATOR OF INTERNODE ELONGATION 1) and DEC1 (DECCELERATOR OF INTERNODE ELONGATION 1) are GA responsive genes identified in recent years, which act antagonistically in the regulation of internode stem elongation in rice (Nagai et al., 2020).

In the last five years, the funRiceGenes database has also witnessed notable progress in the identification of genes related to various phenotypic traits in rice, including asexual reproduction and herbicide resistance. BABY BOOM1 (BBM1) is an AP2/ERF transcription factor expressed in sperm cells, which can induce parthenogenesis in rice (Khanday et al., 2019). An apomixis system can be established by ectopic expression of BBM1 in the egg cell to achieve asexual reproduction of rice seeds, by replacing meiosis with mitosis. As homologous *BBM*-like genes and *MiMe* (substitute mitosis for meiosis) genes were also present in other cereal crops including maize (Lowe et al., 2016; Horstman et al., 2014), this method can be directly applied to other cereal crops for maintaining heterosis through seed propagation. Benzobicyclon (BBC) is a β-triketone herbicide extensively used in weed control. However, some crops are also susceptible to BBC, which discourages its usage and diminishes its value in weed control. A recently cloned gene *HIS1* (*HPPD INHIBITOR SENSITIVE 1*) encodes an oxidase that detoxifies BBC herbicides by catalyzing their hydroxylation, resulting in enhanced resistance against BBC and other β-triketone herbicides in rice, which would be useful in herbicide-resistant plant breeding (Maeda et al., 2019).

Updated interaction networks of functionally characterized rice genes

Based on the >4100 functionally characterized rice genes, a total of 219 interaction networks comprising 1825 genes were constructed using the approach proposed in the previous study (Yao et al., 2017). In total, 2819 connections between genes supported by 6041 pieces of evidence were extracted from the titles and abstracts of published literature. The number of genes in the largest network increased from 762 to 1281 (Figure 4). Accordingly, genes involved in the same biological pathway or with similar functions were found to be clustered together. We also found that genes associated with flowering are still the key component of the largest interaction network. These gene interaction networks were built by inspecting the concurrence of the symbols of two or more genes in the same sentence of published

literature. The approach and the built interaction networks would shed light on future studies on gene functions in other crops.

Conclusion

With the rapid advance of rice functional genomics studies, the funRiceGenes database was updated with more than 1300 newly cloned genes and more than 1000 members of gene families in the last five years. Up to November 2021, a total of ~4100 functionally characterized rice genes and more than 6100 gene family members were deposited in funRiceGenes, providing a valuable resource for functional dissection of genes in rice and other plants. To facilitate the utilization of funRiceGenes, we also implemented new features, including in-site searching in the static website and a BLAST interface in the interactive web application. In this study, we summarized the newly added genes and new features of funRiceGenes. The funRiceGenes database will be continuously updated with new genes and new data, serving the functional genomics researches in rice and other plants.

Abbreviations

ABA: Abscisic acid

ACE1: ACCELERATOR OF INTERNODE ELONGATION 1

BBC: Benzobicyclon

BBM1: BABY BOOM1

BR: Brassinosteroids

CK: Cytokinins

DEC1: DECELERATOR OF INTERNODE ELONGATION 1

DEP1: DENSE AND ERECT PANICLE1

ET: Ethylene

GA: Gibberellins

GSA1: Grain Size and Abiotic stress tolerance 1

IPA1: Ideal Plant Architecture 1

JA: Jasmonates

SA: Salicylic acid

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of Data and Materials

The funRiceGenes database is freely available at <https://venyao.xyz/funRiceGenes/> and <https://funricegenes.github.io/>.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the National Natural Science Foundation of China (31900451), the project of Henan Provincial Department of Education (20A180011), the Scientific and Technological Research Project of Henan Province (202102110015, 212102110243), and the research start-up fund to topnotch talents of Henan Agricultural University (30500581).

Authors' contributions

WY conceived the study and revised the manuscript. FH and YJ carried out the experiment and wrote the manuscript with help from TC, WY, YX, LJ, YO and YL. HL, MF, YW, YX and LJ collected and analyzed the data. WY, FH, YJ and TC wrote scripts for the database. WY, YO, YL and ZZ revised the manuscript, and evaluated the database. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Dhaka, N., Krishnan, K., Kandpal, M., Vashisht, I., Pal, M., Sharma, M. K., et al. (2020). Transcriptional trajectories of anther development provide candidates for engineering male fertility in sorghum. *Scientific Reports*, 10(1), 897. doi:10.1038/s41598-020-57717-0.
2. Dong, N.-Q., Sun, Y., Guo, T., Shi, C.-L., Zhang, Y.-M., Kan, Y., et al. (2020). UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nature Communications*, 11(1), 2629. doi:10.1038/s41467-020-16403-5.

3. Gupta, C., Ramegowda, V., Basu, S., & Pereira, A. (2021). Using Network-Based Machine Learning to Predict Transcription Factors Involved in Drought Resistance. *Frontiers in Genetics*, 12(943). doi:10.3389/fgene.2021.652189.
4. Horstman, A., Willemsen, V., Boutilier, K., & Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: hubs in a plethora of networks. *Trends in Plant Science*, 19(3), 146-157. doi:10.1016/j.tplants.2013.10.010.
5. Hosni, H., Diallo, A., Morcillo, F., Vaissayre, V., Collin, M., Tranchant-Dubreuil, C., et al. (2021). Redox-related gene expression and sugar accumulation patterns are altered in the edible inflorescence produced by the cultivated form of pacaya palm (*Chamaedorea tepejilote*). *Annals of Botany*, 128(2), 231-240. doi:10.1093/aob/mcab060.
6. Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., et al. (2009). Natural variation at the *DEP1* locus enhances grain yield in rice. *Nature Genetics*, 41(4), 494-497. doi:10.1038/ng.352.
7. Jia, L., Yao, W., Jiang, Y., Li, Y., Wang, Z., Li, H., et al. (2021). Development of interactive biological web applications with R/Shiny. *Briefings in Bioinformatics*, bbab415. doi:10.1093/bib/bbab415.
8. Khanday, I., Skinner, D., Yang, B., Mercier, R., & Sundaresan, V. (2019). A male-expressed rice embryogenic trigger redirected for asexual propagation through seeds. *Nature*, 565(7737), 91-95. doi:10.1038/s41586-018-0785-8.
9. Kim, S., Kim, C.-Y., Park, S.-Y., Kim, K.-T., Jeon, J., Chung, H., et al. (2020). Two nuclear effectors of the rice blast fungus modulate host immunity via transcriptional reprogramming. *Nature Communications*, 11(1), 5845. doi:10.1038/s41467-020-19624-w.
10. Li, G., Wang, L., Yang, J., He, H., Jin, H., Li, X., et al. (2021). A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nature Genetics*, 53(4), 574-584. doi:10.1038/s41588-021-00808-z.
11. Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M. J., et al. (2016). Morphogenic Regulators Baby boom and Wuschel Improve Monocot Transformation. *Plant Cell*, 28(9), 1998-2015. doi:10.1105/tpc.16.00124.
12. Maeda, H., Murata, K., Sakuma, N., Takei, S., Yamazaki, A., Karim, M. R., et al. (2019). A rice gene that confers broad-spectrum resistance to β-triketone herbicides. *Science*, 365(6451), 393-396. doi:10.1126/science.aax0379.
13. Nagai, K., Mori, Y., Ishikawa, S., Furuta, T., Gamuyao, R., Niimi, Y., et al. (2020). Antagonistic regulation of the gibberellic acid response during stem growth in rice. *Nature*, 584(7819), 109-114. doi:10.1038/s41586-020-2501-8.
14. Pang, Y., Liu, C., Wang, D., St Amand, P., Bernardo, A., Li, W., et al. (2020). High-Resolution Genome-wide Association Study Identifies Genomic Regions and Candidate Genes for Important Agronomic Traits in Wheat. *Mol Plant*, 13(9), 1311-1327. doi:10.1016/j.molp.2020.07.008.
15. Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell*, 184(13), 3542-3558.e16. doi:10.1016/j.cell.2021.04.046.

16. Wang, J., Wu, X., Yue, W., Zhao, C., Yang, J., & Zhou, M. (2021). Identification of QTL for barley grain size. *PeerJ*, 9, e11287. doi:10.7717/peerj.11287.
17. Wang, J., Zhou, L., Shi, H., Chern, M., Yu, H., Yi, H., et al. (2018). A single transcription factor promotes both yield and immunity in rice. *Science*, 361(6406), 1026-1028. doi:10.1126/science.aat7675.
18. Wing, R. A., Purugganan, M. D., & Zhang, Q. (2018). The rice genome revolution: from an ancient grain to Green Super Rice. *Nat Rev Genet*, 19(8), 505-517. doi:10.1038/s41576-018-0024-z.
19. Wittern, L., Steed, G., Taylor, L. J., Gardner, K., Greenland, A., Hannah, M. A., et al. (2021). Wheat *EARLY FLOWERING3* is a dawn-expressed circadian oscillator component that regulates heading date. *bioRxiv*, 2021.09.03.458922. doi:10.1101/2021.09.03.458922.
20. Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Mol Plant*, 13(2), 187-214. doi:10.1016/j.molp.2020.01.008.
21. Yao, W., Li, G., Yu, Y., & Ouyang, Y. (2017). funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *GigaScience*, 7(1). doi:10.1093/gigascience/gix119.
22. Yu, Y., Yao, W., Wang, Y., & Huang, F. (2019). shinyChromosome: An R/Shiny Application for Interactive Creation of Non-circular Plots of Whole Genomes. *Genomics, Proteomics & Bioinformatics*, 17(5), 535-539. doi:<https://doi.org/10.1016/j.gpb.2019.07.003>.
23. Yu, Y., Zhou, Y., Feng, Y., He, H., Lian, J., Yang, Y., et al. (2020). Transcriptional landscape of pathogen-responsive lncRNAs in rice unveils the role of ALEX1 in jasmonate pathway and disease resistance. *Plant Biotechnology Journal*, 18(3), 679-690. doi:10.1111/pbi.13234.
24. Zhang, F., Hu, Z., Wu, Z., Lu, J., Shi, Y., Xu, J., et al. (2021). Reciprocal adaptation of rice and *Xanthomonas oryzae* pv. *oryzae*: cross-species 2D GWAS reveals the underlying genetics. *The Plant Cell*, 33(8), 2538-2561. doi:10.1093/plcell/koab146.
25. Zhang, H., Tao, Z., Hong, H., Chen, Z., Wu, C., Li, X., et al. (2016). Transposon-derived small RNA is responsible for modified function of WRKY45 locus. *Nature Plants*, 2(3), 16016. doi:10.1038/nplants.2016.16.

Figures

● Users
1,000

Mar 27, 2017 - Nov 19, 2021

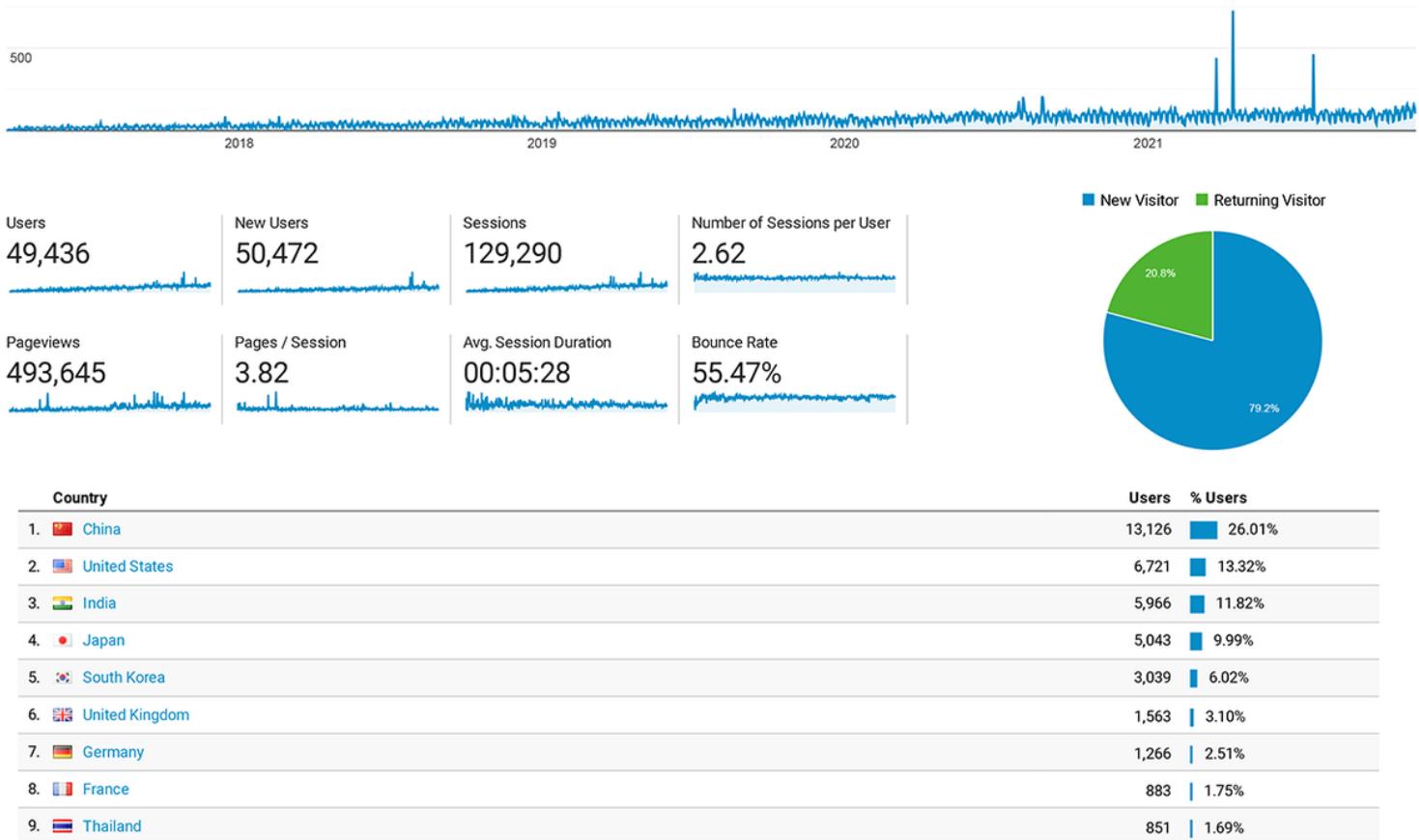


Figure 1

Statistics of the funRiceGenes database visits reported by Google Analytics.

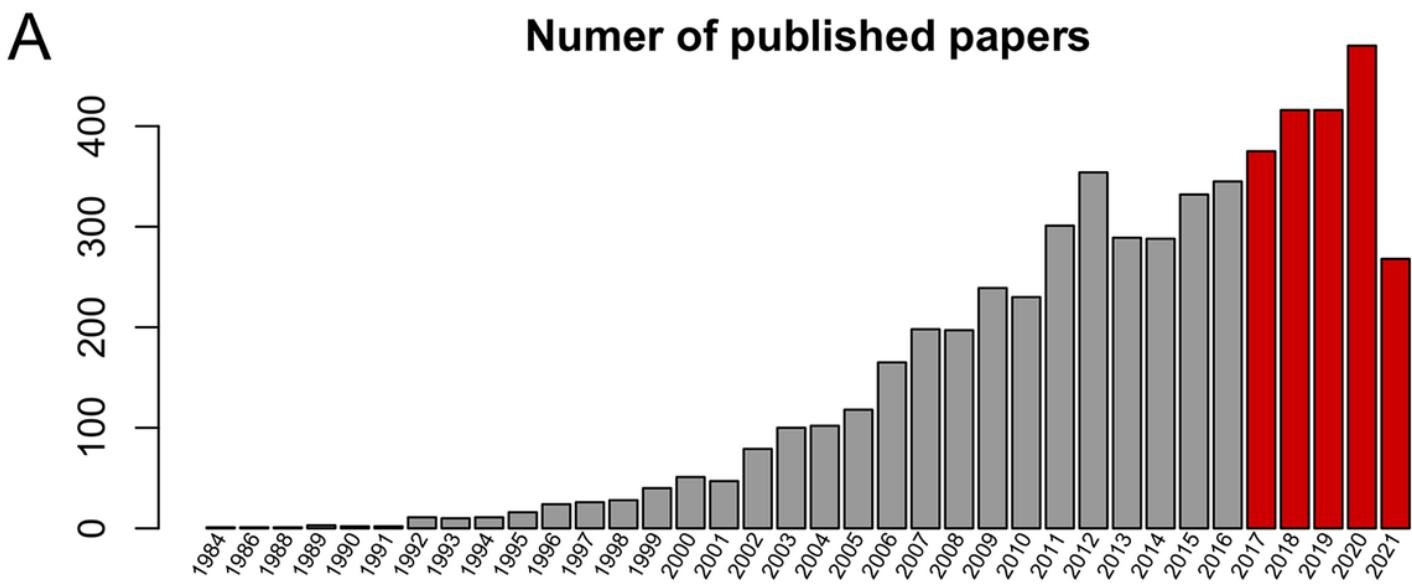


Figure 2

Overview of the literature on functional rice genes published in the last five years. (A) The number of published papers on rice functional genomics in each year. (B) A word cloud of all the abstracts of published papers on rice functional genomics since 2017. (C) A word cloud of all the titles of published papers on rice functional genomics since 2017.



Figure 3

The genomic distribution of 674 representative rice genes cloned in the last five years.



Figure 4

A comprehensive interaction network comprising 1281 rice genes.

Each functionally characterized rice gene is represented as a black node with the corresponding gene symbol marked beside in blue color. Interacting gene pairs are connected by green lines. Representative well-known genes are indicated with larger font sizes.