

RESEARCH

Open Access



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

Fangfang Huang^{1†}, Yingru Jiang^{1†}, Tiantian Chen^{1†}, Haoran Li¹, Mengjia Fu¹, Yazhou Wang¹, Yufang Xu¹, Yang Li¹, Zhengfu Zhou², Lihua Jia³, Yidan Ouyang⁴ and Wen Yao^{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 54,400 users with over 540,000 pageviews. 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 5 years.

Keywords: Rice, *Oryza sativa*, Functionally characterized genes, Functional genomics, Database

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 the world's food security. Rice is also a model organism for plant genomics researches, 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). The data of funRiceGenes database was collected by integrating data extracted from published literature on rice functional genomics deposited in PubMed, the Rice Genome Annotation Project database (Kawahara et al. 2013), the China Rice Data Center (<https://www.ricedata.cn/>) and the Oryzabase database (Yamazaki et al. 2010). Since being released, this database has been continuously updated based on the latest research progress in rice. Up to November 2021, over 4100 functionally characterized rice genes and more than 6000 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.

*Correspondence: yaowen@henau.edu.cn

[†]Fangfang Huang, Yingru Jiang and Tiantian Chen have contributed equally to this work

¹ National Key Laboratory of Wheat and Maize Crop Science, College of Life Sciences, Henan Agricultural University, Zhengzhou 450002, China
Full list of author information is available at the end of the article

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 54,400 users, with more than 142,000 sessions and over 540,000 pageviews (Fig. 1) (Yao et al. 2017). The daily number of users and pageviews of the funRiceGenes database were steadily rising in the last 5 years. An average of 3–4 pages was accessed per session. In 2021, the funRiceGenes database was visited by an average of 1917 users per month, with a maximum of 2519 user visits in July. We observed two significant peaks in the daily pageviews of the funRiceGenes database on December 11, 2017 and February

14, 2018, corresponding to the advance online date and formal publishing date of our previous publication in GigaScience (Fig. 1) (Yao et al. 2017). We also observed other significant peaks, which were probably caused by high-impact publications citing the funRiceGenes database or advertisement of funRiceGenes in plenary talks of prominent experts in the rice community (Wing et al. 2018; Li et al. 2021). The top three most visited pages were the GENE, FAMS and DOCS pages of <https://funricegenes.github.io/>, which listed all the genes, gene families, published papers deposited in the funRiceGenes database, respectively. Apart from the three pages, the top 10 most visited pages were OsPht1, OsHKT1, IPA1, ZYGO1, Gn1a, OsWRKY45, ERF, WRKY, OsTIR1, SD1.

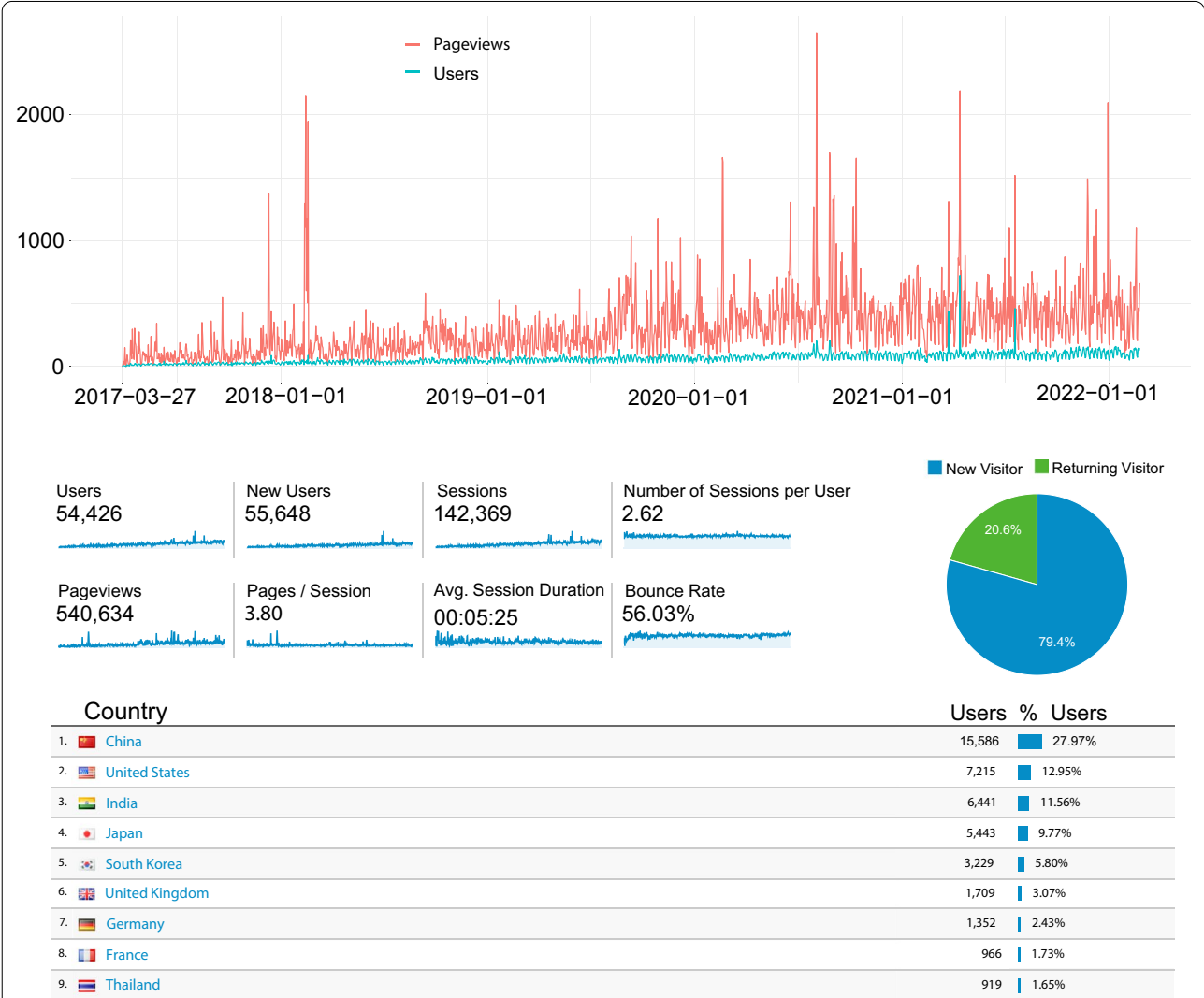


Fig. 1 Statistics of the funRiceGenes database visits reported by Google Analytics. The top panel displays the daily user visits and pageviews of the funRiceGenes database from Mar 27 2017 to Feb 24 2022. The middle panel lists the total user visits, pageviews, and other statistics in the last 5 years. The bottom panel shows the proportion of user visits from different countries in the last 5 years

The pages of ERF and WRKY itemize all the members of the ERF and WRKY gene families, while the other eight pages record the detailed information of eight notable genes in rice (Rice Wrky Working Group 2012; Nakano et al. 2006). *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). 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 February 2022, the funRiceGenes database and the corresponding publication in *GigaScience* have been cited by 102 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; 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

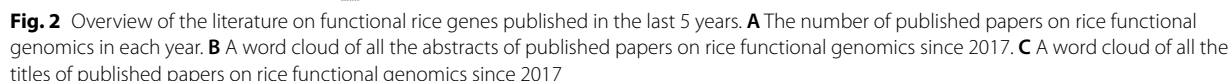
A static website (<https://funricegenes.github.io/>) was developed for browsing and searching functional rice genes deposited in funRiceGenes. Eight menus were deployed in the static website, including HOME, GENE, FAMS, KEYS, NEWS, DOCS, CITE, and LINK. The symbols of more than 4100 functionally characterized 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 (Kawahara et al. 2013; Sakai et al. 2013), 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. Moreover, we utilized Bioschemas to install 590 keywords related to rice functional genomics in the metadata of the static website, aiming to facilitate the findability of the website by search engines (Michel 2018). 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

created utilizing key messages on the functions of input genes deposited in funRiceGenes.

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 (Fig. 2A). It was found that the keywords including protein, grain, development, stress, genes, and tolerance





were enriched in the titles of these literature (Fig. 2B). Similarly, the keywords including grain, protein, mutant, expression, genes, growth, and tolerance were most frequently used in the abstracts of these literature (Fig. 2C).

These enriched keywords represent the research focuses on rice functional genomics studies in recent years. Among the ~4100 genes collected in the funRiceGenes

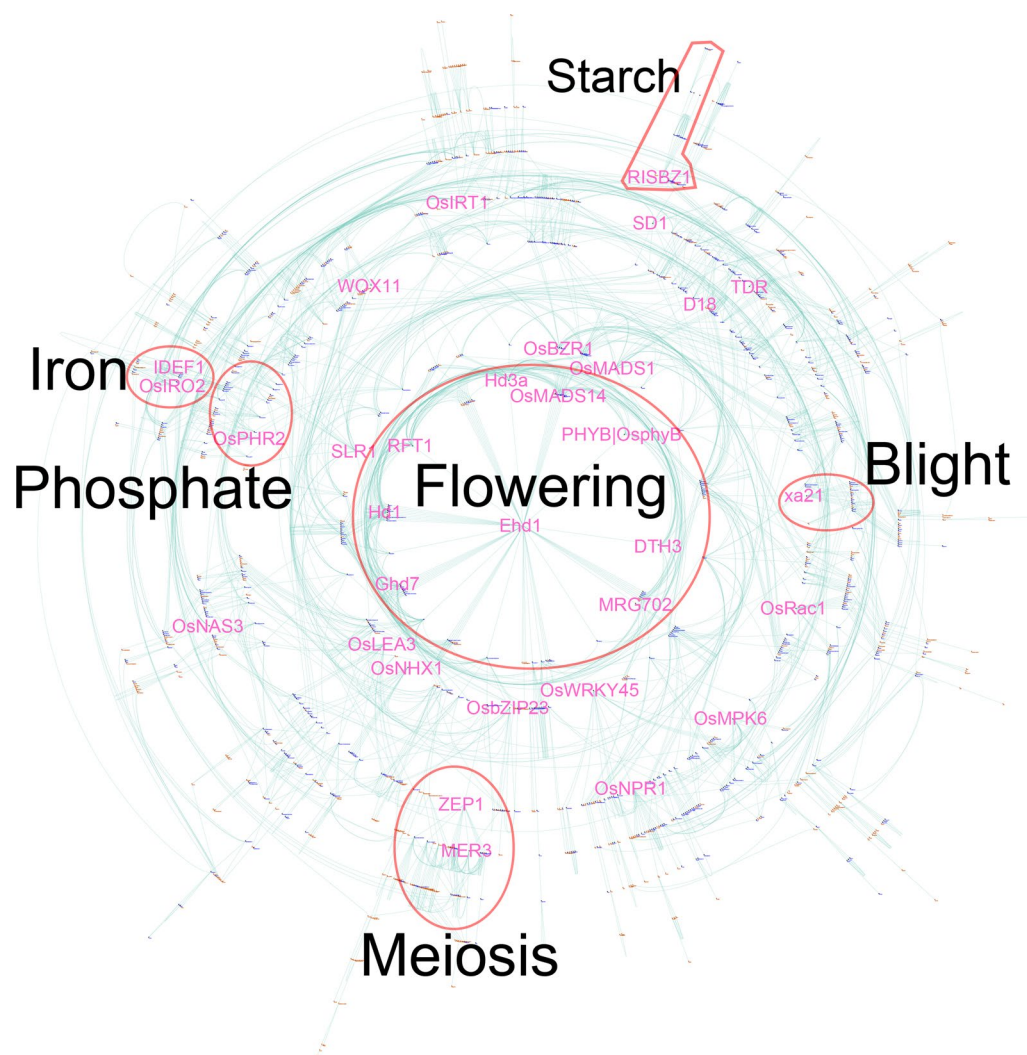


Fig. 4 A comprehensive interaction network comprising 1281 rice genes. Each functionally characterized rice gene is represented as a node with the corresponding gene symbol marked beside. Interacting gene pairs are connected by green lines. Representative well-known genes are indicated with larger font sizes. A total of 762 genes in the largest interaction network built in the previous study are indicated in blue, while 519 new genes involved in the largest network built in the present study are in saddle brown. Genes involved in the same biological pathways are highlighted by red ellipses or polygons

database, 4021 genes can be anchored on the 12 chromosomes (Fig. 3) (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, 68 genes were involved in blast disease resistance, 58 genes participated in blight disease resistance, while 9 genes were related to both diseases.

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 (DECELERATOR 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 5 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. Benzobicyclon (BBC) is a β -triketone herbicide extensively used 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 in the previous study to 1281 in the present study (Fig. 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 5 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, severing the functional genomics researches in rice and other plants.

Abbreviations

ABA: Absciscic 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.

Acknowledgements

Not applicable.

Author contributions

W.Y. conceived the study and revised the manuscript. F.H. and Y.J. carried out the experiment and wrote the manuscript with help from T.C., W.Y., Y.X., L.J., Y.O. and Y.L. H.L., M.F., Y.W., Y.X. and L.J. collected and analyzed the data. W.Y., F.H., Y.J. and T.C. wrote scripts for the database. W.Y., Y.O., Y.L. and Z.Z. revised the manuscript, and evaluated the database. All authors read and approved the final manuscript.

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).

Availability of Data and Materials

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

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹National Key Laboratory of Wheat and Maize Crop Science, College of Life Sciences, Henan Agricultural University, Zhengzhou 450002, China. ²Henan Academy of Crop Molecular Breeding, Henan Academy of Agricultural Sciences, Zhengzhou 450002, China. ³National Key Laboratory of Wheat and Maize Crop Science, College of Agronomy, Henan Agricultural University, Zhengzhou 450002, China. ⁴National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, China.

Received: 24 December 2021 Accepted: 6 April 2022
Published online: 19 April 2022

References

- Dhaka N, Krishnan K, Kandpal M, Vashisht I, Pal M, Sharma MK et al (2020) Transcriptional trajectories of anther development provide candidates for engineering male fertility in sorghum. *Sci Rep* 10(1):897. <https://doi.org/10.1038/s41598-020-57717-0>
- Dong N-Q, Sun Y, Guo T, Shi C-L, Zhang Y-M, Kan Y et al (2020) UDP-glucosyl-transferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nat Commun* 11(1):2629. <https://doi.org/10.1038/s41467-020-16403-5>
- Gupta C, Ramegowda V, Basu S, Pereira A (2021) Using network-based machine learning to predict transcription factors involved in drought resistance. *Front Genet*. <https://doi.org/10.3389/fgene.2021.652189>
- 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*). *Ann Bot* 128(2):231–240. <https://doi.org/10.1093/aob/mcab060>
- Jia L, Yao W, Jiang Y, Li Y, Wang Z, Li H et al (2021) Development of interactive biological web applications with R/Shiny. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbab415>
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6(1):4. <https://doi.org/10.1186/1939-8433-6-4>
- 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. <https://doi.org/10.1038/s41586-018-0785-8>
- 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. *Nat Commun* 11(1):5845. <https://doi.org/10.1038/s41467-020-19624-w>
- 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. *Nat Genet* 53(4):574–584. <https://doi.org/10.1038/s41588-021-00808-z>
- Maeda H, Murata K, Sakuma N, Takei S, Yamazaki A, Karim MR et al (2019) A rice gene that confers broad-spectrum resistance to β -triketone herbicides. *Science* 365(6451):393–396. <https://doi.org/10.1126/science.aax0379>
- Michel F (2018) Bioschemas & Schema.org: a lightweight semantic layer for life sciences websites. *Biodivers Inf Sci Stand* 2:e25836
- 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. <https://doi.org/10.1038/s41586-020-2501-8>
- Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol* 140(2):411–432. <https://doi.org/10.1104/pp.105.073783>
- 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. <https://doi.org/10.1016/j.molp.2020.07.008>
- 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. <https://doi.org/10.1016/j.cell.2021.04.046>
- Rice Wrky Working Group (2012) Nomenclature report on rice WRKY's—conflict regarding gene names and its solution. *Rice* 5(1):3. <https://doi.org/10.1186/1939-8433-5-3>
- Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y et al (2013) Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol* 54(2):e6–e6. <https://doi.org/10.1093/pcp/pcs183>
- 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. <https://doi.org/10.1126/science.aat7675>
- Wang J, Wu X, Yue W, Zhao C, Yang J, Zhou M (2021) Identification of QTL for barley grain size. *PeerJ* 9:e11287. <https://doi.org/10.7717/peerj.11287>
- Wing RA, Purugganan MD, Zhang Q (2018) The rice genome revolution: from an ancient grain to Green Super Rice. *Nat Rev Genet* 19(8):505–517. <https://doi.org/10.1038/s41576-018-0024-z>
- Wittern L, Steed G, Taylor LJ, Gardner K, Greenland A, Hannah MA et al (2021) Wheat *EARLY FLOWERING3* is a dawn-expressed circadian oscillator component that regulates heading date. *bioRxiv*. <https://doi.org/10.1101/2021.09.03.458922>
- Yamazaki Y, Sakaniwa S, Tsuchiya R, Nonomura K-I, Kurata N (2010) Oryzabase: an integrated information resource for rice science. *Breed Sci* 60(5):544–548. <https://doi.org/10.1270/jsbbs.60.544>
- Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD et al (2020) Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol Plant* 13(2):187–214. <https://doi.org/10.1016/j.molp.2020.01.008>
- Yao W, Li G, Yu Y, Ouyang Y (2017) funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *GigaScience*. <https://doi.org/10.1093/gigascience/gix119>
- Yu Y, Yao W, Wang Y, Huang F (2019) shinyChromosome: An R/Shiny application for interactive creation of non-circular plots of whole genomes. *Genom Proteom Bioinform* 17(5):535–539. <https://doi.org/10.1016/j.gpb.2019.07.003>
- 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. *Plant Cell* 33(8):2538–2561. <https://doi.org/10.1093/plcell/koab146>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)