MiRiQ Database: A Platform for In Silico Rice Mutant Screening

Kubo, Takahiko
Faculty of Agriculture, Kyushu University

Yamagata, Yoshiyuki
Faculty of Agriculture, Kyushu University

Matsusaka, Hiroaki
Faculty of Agriculture, Kyushu University

Toyoda, Atsushi
National Institute of Genetics

他

出版情報：Plant And Cell Physiology. 65 (1), pp.169-174, 2023-11-01. 日本植物生理学会
バージョン：

PCP
PLANT & CELL PHYSIOLOGY

MiRiQ Database: A Platform for In Silico Rice Mutant Screening

Takahiko Kubo[1,†,*], Yoshiyuki Yamagata[1,†], Hiroaki Matsusaka[1], Atsushi Toyoda[2], Yutaka Sato[2] and Toshihiro Kumamaru[1]

[1]Faculty of Agriculture, Kyushu University, 744 Motooka Nishi-ku, Fukuoka, 819-0395 Japan
[2]National Institute of Genetics, Yata 1111, Mishima, Shizuoka, 411-8540 Japan
[†]These authors equally contributed to this work.
*Corresponding author: E-mail, takubo@agr.kyushu-u.ac.jp
(Received 15 September 2023; Accepted 31 October 2023)

Genetic studies using mutant resources have significantly contributed to elucidating plant gene function. Massive mutant libraries sequenced by next-generation sequencing technology facilitate mutant identification and functional analysis of genes of interest. Here, we report the creation and release of an open-access database (https://miriq.agr.kyushu-u.ac.jp/index.php), called Mutation-induced Rice in Kyushu University (MiRiQ), designed for in silico mutant screening based on a whole-genome-sequenced mutant library. This database allows any user to easily find mutants of interest without laborious efforts such as large-scale screening by PCR. The initial version of the MiRiQ database (version 1.0) harbors a total of 1.6 million single-nucleotide variants (SNVs) and InDels of 721 M_1 plants that were mutagenized by *N*-methyl-*N*-nitrosourea treatment of the rice cultivar Nipponbare (*Oryza sativa* ssp. *japonica*). The SNVs were distributed among 87% of all 35,630 annotated protein-coding genes of the Nipponbare genome and were predicted to induce missense and nonsense mutations. The MiRiQ database provides built-in tools, such as a search tool by keywords and JBrowse for mutation searches. Users can request mutant seeds in the M_2 or M_3 generations from a request form linked to this database. We believe that the availability of a wide range of gene mutations in this database will benefit the plant science community and breeders worldwide by accelerating functional genomic research and crop improvement.

Keywords: In silico mutant screening • Next-generation sequencing • *N*-methyl-*N*-nitrosourea (MNU) • *Oryza sativa* • Rice mutant library • Single-nucleotide variant (SNV)

Introduction

In higher plants, gene knockout by mutagenesis is an effective strategy for elucidating the biological function of genes. Numerous forward and reverse genetic studies using artificially induced mutants have been conducted in *Arabidopsis*, rice and maize (Jankowicz-Cieslak and Till 2015). Mutant populations of crop plants are used as breeding materials as well as genetic tools for basic plant science. Rice (*Oryza sativa* L.) is a valuable monocotyledon and crop species model due to its compact and fully sequenced genome. Transformation and genome-editing techniques are well established in rice, and extensive research data and databases are available, including gene annotation databases (MSU7.0, http://rice.uga.edu/cgi-bin/gbrowse/rice/; RAP-DB: https://rapdb.dna.affrc.go.jp/), a transcriptome covering almost all organs/tissues during the entire life cycle (Sato et al. 2013), proteome and metabolome databases (Agrawal and Rakwal 2011, Kusano et al. 2015) and gene/QTL-related information based on empirical evidence (curated genes on RAP-DB; Oryzabase, https://shigen.nig.ac.jp/rice/oryzabase/). Moreover, thousands of cultivars and wild relatives of rice are available as natural variation resources and provide valuable materials for breeding and evolutionary biology research (Jackson 1997, Nonomura et al. 2010, Huang et al. 2012). Nipponbare (*O. sativa* ssp. *japonica*) was the first rice cultivar to have its genome and full-length mRNA sequenced, and the database continues to be updated frequently (Kikuchi et al. 2003, Sasaki 2005, Kawahara et al. 2013). The next challenge for functional genomics is determining the function of all genes encoded by the rice genome, thereby increasing the need for reverse genetic tools. The rice research community has made great efforts to generate several types of mutant libraries, such as T-DNA and Tos17 (Hirochika et al. 2004). However, some of these formerly public mutant libraries have already been closed to the public. More recently, a large-scale CRISPR (Clustered Regularly Interspaced Palindromic Repeats)/Cas9 mutant library was developed in rice (Meng et al. 2017), but it has not been a publicly available resource outside the country of origin.

Unlike transformation and genetic-editing techniques that must be subjected to environmental risk/safety assessments, chemical mutagenesis is known to be a relatively safe technique that can create novel mutations. *N*-methyl-*N*-nitrosourea

Plant Cell Physiol. 00(00): 1–6 (2023) doi:https://doi.org/10.1093/pcp/pcad134, Advance Access publication on 1 November 2023, available online at https://academic.oup.com/pcp

(MNU) is a highly effective chemical mutagen used in plants and can randomly induce point mutations. Such point mutations can cause a complete loss of function, hypomorphic mutations (partial loss of function) or hypermorphic mutations (increases in gene activity). This variety of mutations can be useful in revealing the function of any gene, including embryonic lethal mutations in essential genes. MNU causes the alkylation of guanine bases, resulting in G/C to A/T transition. Satoh and Omura (1979) developed an effective MNU treatment method for fertilized egg cells just after rice fertilization. Compared with conventional mutagenesis of dry seeds, mutagenesis of fertilized egg cells yields a higher mutation efficiency and reduces the frequency of M_1 chimeric plants and intercellular competition. Using this method, our institute has developed and released >10,000 rice mutant lines (Oryzabase, https://shigen.nig.ac.jp/rice/oryzabase/). These mutant lines have included a broad range of phenotypic abnormalities, leading to the identification of gene functions (Satoh et al. 2010). In the reverse genetic approach, mutation screening steps are required for each target gene. Targeting Induced Local Lesions In Genomes (TILLING) is a PCR-based reverse genetic approach to identify single-nucleotide polymorphisms (SNPs) in a gene of interest from mutant libraries obtained by chemical mutagenesis (Till et al. 2003). In this method, an SNP can be detected as a mismatch cleavage of mutant and wild-type DNA heteroduplexes by the Cel I enzyme. TILLING has identified a large number of gene mutations and has helped to elucidate many gene functions in plants (Kurowska et al. 2011). Although the conventional TILLING method is proven to be a high-throughput technique for obtaining mutants by reverse genetic screening, the method always requires a large-scale PCR screening of mutants for each gene that takes several weeks.

The recent next-generation sequencing (NGS) technique allows the identification of genome-wide variations among multiple samples. Hence, developing an open, online platform publishing mutant libraries whose genomes are sequenced by NGS is expected to significantly accelerate gene function analysis and genomics. Such an online mutant screening (in silico TILLING) strategy has been proposed for over a decade (Wang et al. 2012), but no publicly available platform is currently available. Recently, we reported the whole-genome sequencing (WGS) of a small MNU-mutant population ($N = 261$) and its genomic features (Kubo et al. 2022). This pilot study demonstrated the feasibility of in silico TILLING. Here, we report a recently developed web-based tool for searching mutations that includes an easy interface for users to screen mutations within genes of interest using a search function and JBrowse visualization.

Results and Discussion

Mutant library construction and variant identification

We developed 721 mutant M_1 plants derived from MNU treatment of fertilized egg cells of the rice cultivar Nipponbare to construct an in silico mutant library. WGS of the 721 M_1 plants resulted in an average of 12.0 Gb short reads per individual. To identify the induced mutation variants, the short-read sequences were mapped onto the Nipponbare reference genome (genome size, 373.2 Mb). The average depth of coverage for each mutant genome was 24.5, and the breadth of coverage was 98.0%. By resequencing the mutant genomes, a total of 1,595,263 induced mutations, including single-nucleotide variants (SNVs) (1,588,402) and InDels (6,861), were detected through the variant call procedure reported previously (Kubo et al. 2022). The mutation rate was estimated to be 6.0 SNVs/Mb. The average numbers of SNVs and InDels per individual were 2,203.1 and 9.5, respectively (**Supplementary Tables S1, S2**). Of the 1,588,402 SNVs, 92.0% were G/C to A/T transitions, as expected for MNU mutagenesis (**Supplementary Table S3**). The effects of induced mutations on gene function were annotated using the SnpEff program. The identified SNVs induced nonsense mutations in 10.3% of 35,630 protein-coding genes annotated by IRGSP-1.0, and missense mutations in 85.8% of them. In the MSU7 annotation, nonsense mutations were found in 14.5% of the 55,801 annotated genes, and missense mutations were found in 88.4% of them (**Fig. 1**). The SNVs, characterized as high and moderate impact levels, were found in 87.5% of the RAP-DB protein-coding genes (31,174 of 35,630 genes) (**Supplementary Table S4**). Thus, the current collection of mutations covered around 90% of all rice genes with at least one nonsynonymous mutation. These results show that MNU-induced mutants derived from treating single fertilized egg cells are an effective source for constructing rice in silico mutant libraries. Among these mutants, the occurrence of InDels was extremely minor, affecting only about 0.016% of all genes (573 of 35,630 genes); statistical data for the InDels are shown in **Supplementary Table S5** and **Fig. S1**.

Fig. 1 SNV impacts on rice annotated genes. The distributions of missense, nonsense and silent mutations for 35,630 IRGSP-1.0 protein-coding genes and 55,801 MSU7.0 genes.

Database content

To use the defined mutant library effectively, we constructed a database containing detailed information about mutation sites. This database, named the Mutation-induced Rice in Kyushu University (MiRiQ) database (hereafter called MiRiQ-DB), was developed with PHP/MySQL-containing variant types,

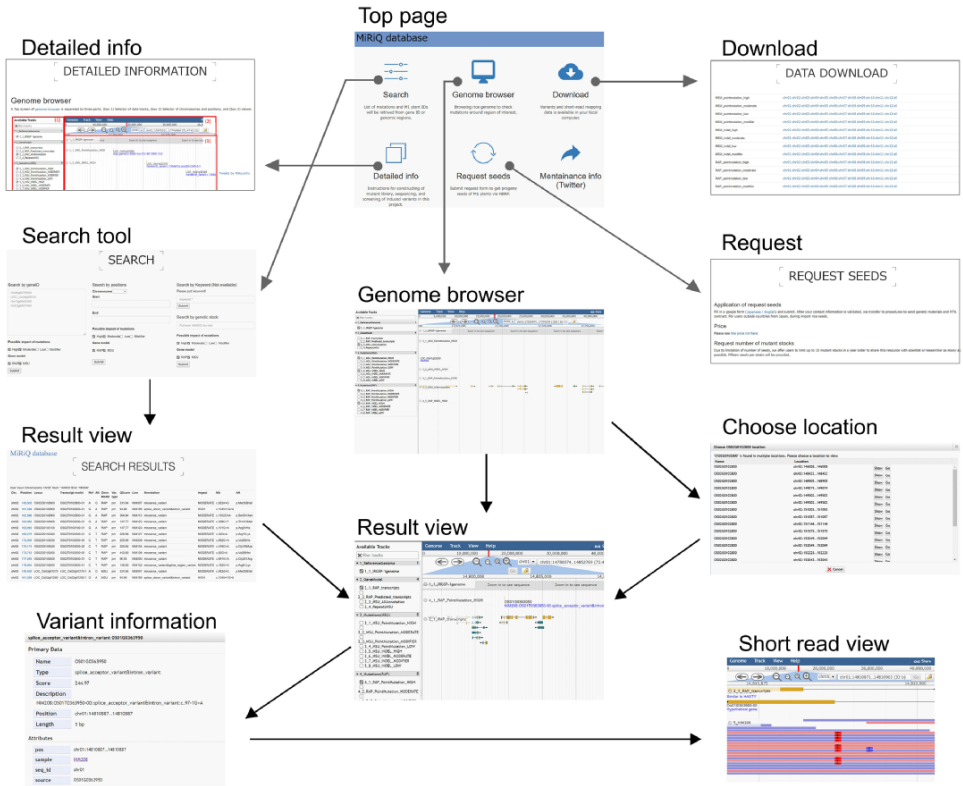

Fig. 2 A sitemap schematic diagram of the MiRiQ-DB. There are six main functions: Top Page, Search Tool, Genome Browser, Download, Detailed Information and Request Seeds. Arrows indicate options to move to the next page.

the mutation effects and positional coordinates on the Nipponbare rice reference genome and the Genome Browser JBrowse (Skinner et al. 2009) to display variant positions and short reads on the gene annotation. The MiRiQ-DB portal (https://miriq.agr.kyushu-u.ac.jp/index.php) provides five major functions: Search Tool, Genome Browser, Download, Detailed Information and Request Seeds (**Figs. 2**, **3A**). Users can search mutation variants influencing genes of interest through the Search Tool, Genome Browser and Download functions as detailed below.

Search Tool. Mutated variations can be searched by keywords such as specific gene ID, chromosomal position and mutant line ID. The Search Tool identifies mutation variants by the possible impact of mutations on gene function as proposed by the gene model of IRGSP-1.0/MSU7.0. A single gene ID is acceptable to initiate a keyword search in 'Search by gene ID', but not multiple IDs. Users can search for variations within a specified genomic region using 'Search by positions'. Search by mutant ID generates a list of up to 500 variation sites contained in the searched mutant genome. The hit variations are classified according to the functional impact of the mutations predicted by the SnpEff program against the IRGSP-1.0/MSU7.0 gene model. The functional impact levels contain four levels in order of impact: High, Moderate, Low and Modifier. The High impact variants have a highly disruptive effect on the protein; stop-gain (nonsense) and frame-shift variants are classified into this impact level. Moderate and Low impacts mainly correspond to missense and synonymous mutations, respectively. Modifier describes UTR, intron and upstream/downstream variants and is predicted to change gene expression levels. Details of variant effects are shown on the SnpEff website (https://pcingola.github.io/SnpEff/). The resultant output data (Search results) from the Search Tool is documented as an embedded table containing position, locus ID/Transcript ID, variant type, quality score, mutant line ID and the SnpEff annotation. Nucleotide and amino-acid change type with the site position within gene/protein position are also included in the output list. This table can be saved and edited in a Microsoft Excel format after copying and pasting.

Genome Browser. The MiRiQ-DB uses the JBrowse Genome Browser to visualize the positions of mutant variants within the Nipponbare IRGSP-1.0 genome (**Fig. 3B–D**). The mutation type based on the MSU7.0 and RAP-DB can be selected from the Available Tracks window. If a mutation site [shown by the Genome Browser as the mutant line number (NIMxxx) with a gene ID] is found within the gene of interest, clicking on the mutation site generates a pop-up window showing detailed information about the SNV, including the effect of the SNV on the gene function quality value score and chromosome position. Furthermore, clicking on the sample ID in the 'Attributes'

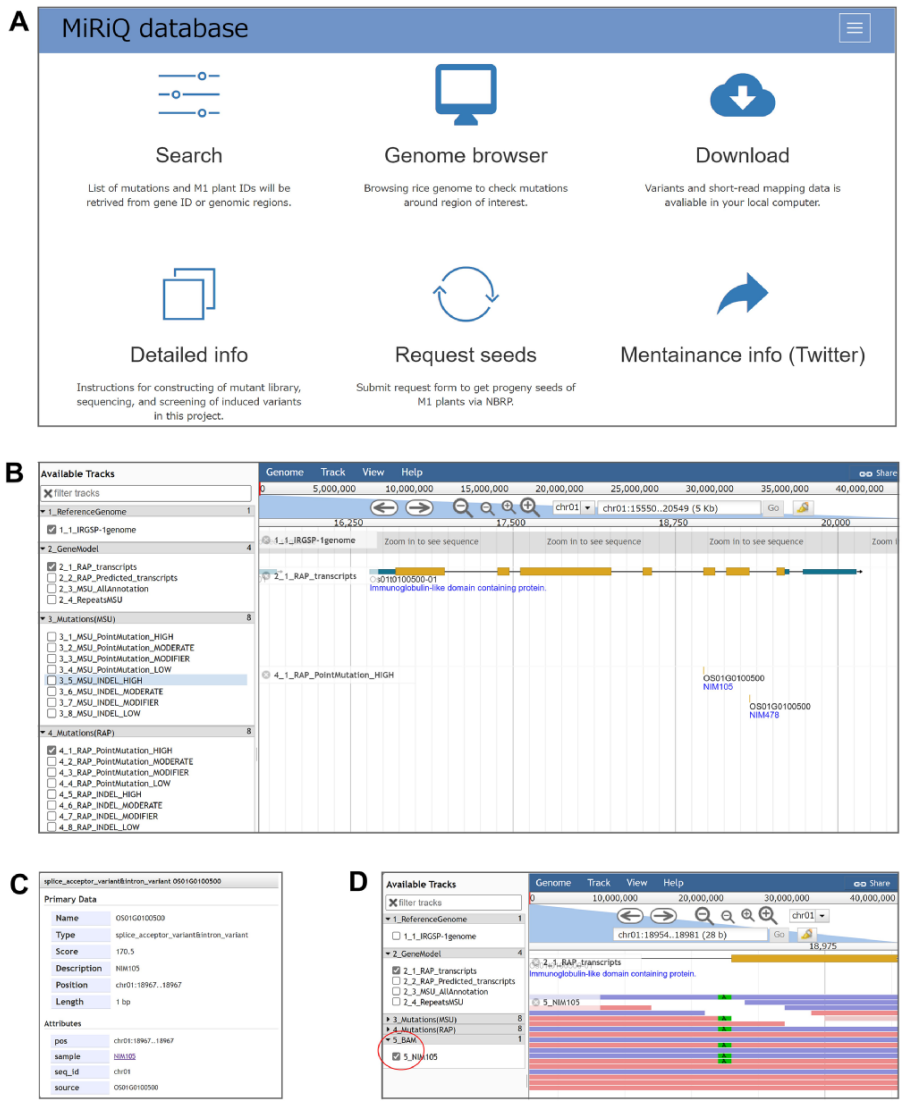

Fig. 3 The MiRiQ-DB web interface. (A) The top page of the MiRiQ-DB. (B) A JBrowse screen displaying the position of induced mutation variants. This screen is shown by clicking the 'Genome Browser' button on the top page. Multiple displays are available at the Available Tracks function, such as gene model and variant type classified by SnpEff impact. (C) Detailed information about the mutation site can be viewed as a pop-up window by clicking the sample ID (NIM line ID) shown in JBrowse. Clicking the sample ID displays the short-read information view (D). (D) The short-read information view. Check the '5_BAM' to display the short reads.

section of the pop-up window generates the short reads around SNVs along the gene. Checking the "5_BAM" in the "Available Tracks" is required to display the short reads. Since M_1 mutant genomes were sequenced, about half of the short reads are expected to contain mutation variants at the variation site.

Download. Batch downloads of the result files are available by selecting 'Download' at the top of the page (**Fig. 3A**). The variant call files and the SnpEff output files can be downloaded in text format, and both file types provide information about the mutant line number (NIMxxx) carrying the mutation. By keeping these result files on the PC, users can screen and filter the genes of interest in text files or Microsoft Excel files without an internet connection. The SnpEff result file is classified based on the annotations model (MSU7.0 and RAP-DB) and the impact level affecting gene function by chromosome.

Detailed Information and Request Seeds. Users can find instructions about using the Genome Browser to find a mutation site in the target gene region from the Detailed Information tab. Also, statistical data of the mutated variations found in the current mutant library can be seen with this function. When interesting mutants are found, users may request the seeds via the 'Request Seeds' function at the top page of the MiRiQ-DB. The mutant seeds in the M_2 or M_3 generation will be distributed from the Institute of Genetic Resources, Kyushu University, Japan. Therefore, users can immediately analyze the morphological and physiological traits of the intended mutants

in an early segregating mutant population after validating the mutated variation by Sanger sequencing. We believe that the MiRiQ-DB will be a valuable tool for reverse genetic research and crop breeding by the plant science community.

Conclusion

The MiRiQ-DB is a freely available and open-access web tool for searching and browsing mutation sites around genes of interest from the MNU-induced rice mutant library. The current version of this database is based on 721 mutant individuals and contains about 1.6 million variations. This level of variation covers almost all of the annotated genes encoded by the Nipponbare genome with nonsynonymous mutations. Users can easily find gene mutations without a command user interface using this database's Search Tool and the mutant Genome Browser. This easy-to-access information about mutation sites without the hassle of large-scale screening will facilitate efficient gene function analysis and designed breeding with the mutant lines. The mutant seed materials are publicly available from the MiRiQ-DB.

Materials and Methods

Plant materials and mutagenesis

MNU mutagenesis of fertilized egg cells in the Japanese rice cultivar Nipponbare (*O. sativa* L. ssp. *japonica*) was conducted basically according to the method of Suzuki et al. (2008). Briefly, panicles with flowers pollinated on the day of anthesis were dipped in a 1.0-mM MNU solution for 45 min at about 25°C at approximately 18 h after flowering. The M_1 and M_2 plants were grown to obtain leaf samples and seeds in 2017–2020.

WGS and variant discovery analysis.

Total genomic DNA was extracted from leaf samples of the M_1 plants (plant ID NIM 1∼496) and the bulked M_2 plants (NIM 497∼750, $n = 6$–10 for each line) using the CTAB method with some modification (Murray and Thompson 1980) or a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). DNA library construction was carried out following protocols developed by Illumina Co., Ltd (San Diego, CA, USA) and MGI Tech Co. Ltd (Shenzhen, China). Paired-end sequencing was performed using Illumina HiSeq and Novaseq systems and an MGITech DNBSEQ-T7. Paired-end reads preprocessed by Trimmomatic v 0.39 (Bolger et al. 2014) were mapped to Os-Nipponbare-Reference-IRGSP-1.0 pseudomolecules (Kawahara et al. 2013) using bwa-mem software (Li and Durbin 2010). Mutagen-induced variants were identified by methods described in our previous report (Kubo et al. 2022) using the Genome Analysis Tool Kit v 4.1.2.0 best practices (Poplin et al. 2017) with minor modification. Mutation effects on gene function were predicted using SnpEff v4.1 with default parameters (Cingolani et al. 2012) using the gff files (IRGSP1.0.51, 10 March 2021 and MSU release 7, 21 November 2011) obtained from IRGSP-1.0 (http://ftp.ensemblgenomes.org/pub/plants/release-51/gff3/oryza_sativa/) and MSU7.0 (http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/).

Database construction

The MiRiQ-DB was built on the Linux operating system (Ubuntu 18.04.4 Long Term Support). The variant search server provides functionality for users to search for relevant variants based on specified keywords, chromosomes, positional information, mutant line names and SNP/InDel types. The JBrowse server (version 1.4) presents variant positions, short-read data on the reference genome and the RAP-DB and MSU7.0 gene models. Users can effectively explore mutations by adjusting the impact types (High, Moderate, Modifier and Low) as determined by the SnpEff program. Additionally, a separate hosting server has been set up using the Amazon Web Service to facilitate the bulk download of variant information, thereby reducing the strain on network resources caused by downloads. The user interface uses the web application framework Bootstrap (version 3.3.4).

Supplementary Data

Supplementary data are available at *PCP* online.

Data Availability

The sequencing reads of the mutants in this paper have been deposited in the Sequence Read Archive of the DDBJ/GenBank data libraries under the accession numbers DRA014011 and DRA016274.

Funding

National BioResource Project.

Acknowledgments

Computations were partially performed on the NIG supercomputer at the National Institute of Genetics, Research Organization of Information and Systems (ROIS). The rice plants were grown in Ito Plant Experiment Fields and Facilities, Faculty of Agriculture, Kyushu University in the winter season. We thank E. Ikeda, K. Fujita, Y. Matsuo, Y. Nakajo and D. Ishihara (Faculty of Agriculture, Kyushu University, Fukuoka, Japan) for technical assistance and fieldwork.

Author Contributions

T.Kub., Y.Y. and Y.S. conceived and designed the experiments. T.Kub., H.M. and T.Kum. developed the mutant population. T.Kub., Y.Y. and A.T. performed the experiments and analyzed data. T.Kub. wrote the paper, with input from Y.Y., Y.S. and T.Kum. All the authors read and approved the final manuscript.

Disclosures

The authors have no conflicts of interest to declare.

References

Agrawal, G.K. and Rakwal, R. (2011) Rice proteomics: a move toward expanded proteome coverage to comparative and functional proteomics uncovers the mysteries of rice and plant biology. *Proteomics* 11: 1630–1649.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6: 80–92.

Hirochika, H., Guiderdoni, E., An, G., Hsing, Y.-I., Eun, M.Y., Han, C.-D., et al. (2004) Rice mutant resources for gene discovery. *Plant Mol. Biol.* 54: 325–334.

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490: 497–501.

Jackson, M.T. (1997) Conservation of rice genetic resources: the role of the International Rice Genebank at IRRI. *Plant Mol. Biol.* 35: 61–67.

Jankowicz-Cieslak, J. and Till, B.J. (2015) Forward and reverse genetics in crop breeding. *In Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools*. Edited by Al-Khayri, J.M., Jain S.M., and Johnson D.V. pp. 215–240. Springer International Switzerland: Switzerland.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., et al. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6: 4.

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., et al. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* 301: 376–379.

Kubo, T., Yamagata, Y., Matsusaka, H., Toyoda, A., Sato, Y. and Kumamaru, T. (2022) Whole-genome sequencing of rice mutant library members induced by *N*-methyl-*N*-nitrosourea mutagenesis of fertilized egg cells. *Rice* 15: 38.

Kurowska, M., Daszkowska-Golec, A., Gruszka, D., Marzec, M., Szurman, M., Szarejko, I., et al. (2011) TILLING—a shortcut in functional genomics. *J. Appl. Genet.* 52: 371–390.

Kusano, M., Yang, Z., Okazaki, Y., Nakabayashi, R., Fukushima, A. and Saito, K. (2015) Using metabolomic approaches to explore chemical diversity in rice. *Mol. Plant* 8: 58–67.

Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26: 589–595.

Meng, X., Yu, H., Zhang, Y., Zhuang, F., Song, X., Gao, S., et al. (2017) Construction of a genome-wide mutant library in rice using CRISPR/Cas9. *Mol. Plant* 10: 1238–1241.

Murray, M.G. and Thompson, W.F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic. Acids Res.* 8: 4321–4325.

Nonomura, K.-I., Morishima, H., Miyabayashi, T., Yamaki, S., Eiguchi, M., Kubo, T., et al. (2010) The wild *Oryza* collection in National BioResource Project (NBRP) of Japan: history, biodiversity and utility. *Breed. Sci.* 60: 502–508.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., et al. (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*. 201178.

Sasaki, T., International Rice GenomeSequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436: 793–800.

Satoh, H., Matsusaka, H. and Kumamaru, T. (2010) Use of *N*-methyl-*N*-nitrosourea treatment of fertilized egg cells for saturation mutagenesis of rice. *Breed. Sci.* 60: 475–485.

Satoh, H. and Omura, T. (1979) Induction of mutation by the treatment of fertilized egg cell with N-methyl-IV-nitrosourea in rice. *J. Fac. Agric.* 24: 165–174.

Sato, Y., Takehisa, H., Kamatsuki, K., Minami, H., Namiki, N., Ikawa, H., et al. (2013) RiceXPro Version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic. Acids Res.* 41: D1206–D1213.

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J. and Holmes, I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res.* 19: 1630–1638.

Suzuki, T., Eiguchi, M., Kumamaru, T., Satoh, H., Matsusaka, H., Moriguchi, K., et al. (2008) MNU-induced mutant pools and high performance TILLING enable finding of any gene mutation in rice. *Mol. Genet. Genom.* 279: 213–223.

Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., et al. (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res.* 13: 524–530.

Wang, T.L., Uauy, C., Robson, F. and Till, B. (2012) TILLING in extremis. *Plant Biotechnol. J.* 10: 761–772.