Review

Current status of genomic resources on wild relatives of rice

Richa Kamboj, Balwant Singh, Tapan Kumar Mondal and Deepak Singh Bisht*

National Institute for Plant Biotechnology, LBS Building, Pusa Campus, New Delhi 110012, India

Rice is a food crop of global importance, cultivated in diverse agro-climatic zones of the world. However, in the process of domestication many beneficial alleles have been eroded from the gene pool of the rice cultivated globally and eventually has made it vulnerable to a plethora of stresses. In contrast, the wild relatives of rice, despite being agronomically inferior, have inherited a potential of surviving in a range of geographical habitats. These adaptations enrich them with novel traits that upon introgression to modern cultivated varieties offer tremendous potential of increasing yield and adaptability. But, due to the unavailability of their genetic as well as genomic resources, identification and characterisation of these novel beneficial alleles has been a challenging task. Nevertheless, with the unprecedented surge in the area of conservation genomics, researchers have now shifted their focus towards these natural repositories of beneficial traits. Presently, there are several generic and specialized databases harboring genome-wide information on wild species of rice, and are acting as a useful resource for identification of novel genes and alleles, designing of molecular markers, comparative analysis and evolutionary biology studies. In this review, we introduce the key features of these databases focusing on their utility in rice breeding programs.

Key Words: database, genomics, Oryza, rice, wild rice.

Introduction

Rice belongs to genus Oryza of family Gramineae that consists of 24 species, 22 wild and 2 cultivated species (Vaughan 1989). Based on their genome, they are classified into 11 groups, six diploid and five tetraploid (AA, BB, CC, BBCC, CCDD, EE, FF, GG, HHJJ, HHHKK and KKLL) (GRiSP 2013). Among these 2 species, two species Oryza sativa (Asian) and Oryza glaberrima (African) are the only cultivated species grown worldwide and in some parts of West and Central Africa, respectively. Further, the O. sativa cultivars have been divided into five major cultivar groups’ aus, indica, aromatic japonica, temperate japonica, and tropical japonica (Garris et al. 2005, Glaszmann 1987). In addition, some less explored rice groups such as Boro, Sali, Rayada etc. are endemic to special geographical regions (Morishima and Sano 1992). The presence of such a vast genetic diversity in rice and its wild relatives enables the plant to adapt to different climatic and environmental conditions (Singh et al. 2018). However, with time, the genetic diversity of Oryza genus has experienced serious genetic erosion, largely due to replacement of indigenous and traditional species with genetically uniform modern high-yielding varieties, habitat destruction and climate change. To deal with this imminent risk of genetic erosion it is important to conserve the available genetic resources of wild relatives of rice. Notably, for rice the genetic diversity conserved in gene banks is impressive. For instance, International Rice Genebank Collection Information System (IRGCIS) of International Rice Research Institute (IRRI) has 130,000 accessions (for both cultivated and wild rice) as on December, 2018 (URL-https://www.irri.org/international-rice-genebank). Similarly, National gene bank of India situated at National Bureau of Plant Genetic Resources (NBPRG) have 86997 indigenous collection (IC) and 11316 exotic collection (EC) accessions of cultivated rice, 365 accessions of O. rufipogon and 777 accessions of O. nivara (as accessed on 11 March 2019).

The availability of the whole genome of the two subspecies of O. sativa i.e. indica and japonica (Nipponbare) in 2005 has provided a fundamental advance in our understanding of rice biology (IRGSP 2005, Yu et al. 2002). Subsequently, the refinement of genomic techniques enabled resequencing of more and more rice cultivars and its underutilized species. Recently, sequencing of many different wild rice genotypes under different projects like rice pan-genome project (sequencing of 3000 divergent accessions), OMAP (Oryza Map Alignment Project) etc. have made genome-wide primary and derivative data publicly available (McNally et al. 2009, Stein et al. 2018, Zhao et al. 2018). The technological advances that are catalyzing...
the ‘omics’ resources are also simultaneously facilitating the development of genomic databases with inventorying of data on genes, gene families and functional annotation enabling access to the researchers with diverse interest. Presently, with a goal of assisting plant geneticists and breeders in rice improvement programs several databases on wild relatives of rice have been developed. In this review, we have surveyed a collection of the databases hosting information on various aspects of wild relatives of rice (Table 1). Despite the unavailability of any exclusive database on wild rice, we have tried to gather information scattered across various other databases and presented it here.

**Foundational Genomic Resources on Wild Relatives of Rice**

Wild relatives of rice are known to harbor untapped reservoir for many useful traits of agronomic importance (Singh et al. 2018). Several of them have been already transferred into the cultivated rice varieties. For instance, landmark introgression from wild *Oryza* genus including resistance to grassy stunt virus (Khush and Ling 1974), bacterial leaf blight resistance gene *Xa-21* (Khush et al. 1990), blast resistance genes *Pi-9* (Amante-Bordeos et al. 1992), *Pi-40(t)* (Jeung et al. 2007), brown plant hopper resistance (*Li et al. 2006, Yan et al. 1997*) and L-myo-inositol 1-phosphate synthase gene for salt tolerance (*Das-Chatterjee et al. 2006*) had a profound impact on rice cultivation at global scale. To further intensify the search of novel genes and allele present in the wild relatives of rice, in 2007, IOMAP (International Oryza Map Alignment Project) was initiated with a goal to generate genomic information on wild relative of rice that can be used as a research platform to study evolution, development, genome organization, polyploidy, domestication, gene regulatory networks and crop improvement (Wing et al. 2005). Since its inception, the project has delivered several foundational resources for 23 species of *Oryza* genus including reference level genome and transcriptome assemblies of all the 23 species, advance mapping population for functional and breeding studies, and collection of diverse *Oryza* species for diversity and evolutionary studies (Jacquemin et al. 2013).

The potential advances in sequencing technology have led to the completion of several large scale sequencing projects of rice wild relatives (Brozynska et al. 2017). The magnitude of genomic data generated across accessions and the derived inferences have gradually started unravelling numerous interesting facts related to the *Oryza* genus. For instance, sequencing of *O. brachyantha* revealed that reduced activity of long-terminal retro-transposons and internal deletions of ancient long-terminal repeat elements has resulted in the compactness of its genome (Chen et al. 2013). Such findings and many related discoveries set the stage for researchers to explore the structural and evolutionary dynamics of the genome of genus *Oryza*. Presently, genome of 13 *Oryza* species had been sequenced and for the five species it is in progress (Table 2).

Sequencing of organelle genomes like mitochondrial and chloroplast genomes provides an excellent resource to understand the evolutionary divergence and phylogenetic relatedness between different wild rice accessions. Standard DNA barcoding based on chloroplast genome of wild rice accessions has been developed for cataloguing the rice germplasm on the basis of evolutionary divergence (Liu et al. 2016, Nock et al. 2011, Wambugu et al. 2015). For instance, DNA barcoding based phylogenetic comparison of plastome of *O. brachyantha* with 13 other *Oryza* species and two outgroup *Oryza* species revealed that *O. brachyantha* is an early diverging lineage in *Oryza* genus. Likewise, mitochondrial genome sequence information can also be used for phylogenetic and evolutionary studies amongst species (Asaf et al. 2016). A list of sequenced organellar genome of different members of *Oryza* genus is provided in Table 3.

**Databases of Oryza genus**

**RPAN genome browser**

RPAN refers to the Rice Pan-genome browser, created from 3000 rice genome project. Pan-genome is a union of all the gene sets present in rice species and provides new dimension to the genome complexity based on the presence or absence of variations in a genome. It gives information about the genomic sequence, gene annotations, PAVs (presence/absence of variations among the genes present in the genomes) and gene expression data of rice pan-genome. Around 12000 new genes which are absent in the reference genome are also included in the pan-genome. These novel genes are involved in many important functions like freezing response, cold acclimatization etc. It contains ~370 Mbp IRCGP (International Rice Genome Sequencing Project) genome and ~260 Mbp novel sequences. It also provides multiple search (basic and advanced) and visualization functions. In basic search option, information about single gene, single rice accession can be searched whereas in advanced search, multiple genes and multiple rice accession information can be retrieved. The homepage of RPAN shows heatmap of gene PAVs, genome composition in pan-genome and individual genome in the form of a pie chart, graphical distribution of the rice varieties in the form of a phylogenetic tree based on PAVs of 453 high quality accessions and statistical genomic information of pan-genome and individual genomes of rice. The reference pan-genome sequence and annotation can be downloaded from the hyperlink “download” present on the homepage. The database has been used by researchers of varied interest. Wang et al. 2018, analyzed genetic variation, population structure and diversity among 3010 diverse rice genomes by using RPAN genome browser. They identified 29 million SNPs, 2.4 million small indels and over 90,000 structural variations that contribute to within- and between-population variation. Using pan-genome analyses, they identified more
| S. No | Databases/Bioinformatic Tools/Genetic resources | Type of information | Unique information | Reference |
|-------|-----------------------------------------------|---------------------|-------------------|-----------|
| 1     | RPAN Genome Browser http://cgm.sjtu.edu.cn/3kriceb/ | PAV and sequence information (Genomic, Transcriptomic and Proteomic) | PAV | Sun et al. 2016 |
| 2     | Gramene http://www.gramene.org/ | QTL, Metabolic pathway, Sequence information (Genomic, Transcriptomic, Proteomic), comparative maps, Markers (SNP, SSRs and other markers) | – | Tello-Ruiz et al. 2016 |
| 3     | RiTE-Db www.genome.arizona.edu/rite | Repeated sequences and Transposable elements | Repeat elements and Transposable Elements | Copetti et al. 2015 |
| 4     | RAP-Db http://rapdb.dna.affrc.go.jp/ | Sequence information (Genomic Transcriptomic, Proteomic), SNP, ESTs, SNPs, INDELs | INDELs | Sakai et al. 2013 |
| 5     | Rice Genome knowledgebase http://rgkbase.big.ac.cn/RGKbase/ | QTL, Sequence information (Genomic, Transcriptomic, Proteomic) and Epigenetic data | Epigenetic data | Wang et al. 2012 |
| 6     | OMAP http://www.omap.org/ | Accessions and Sequence information (Genomic, Transcriptomic and Proteomic) | – | Jacquemin et al. 2013 |
| 7     | GeneBank project: NARO http://www.gene.affrc.go.jp/index_en.php | Accessions, Landraces, Improved varieties, Wild varieties | Landraces, Improved varieties and Wild varieties | Yamasaki et al. 2017 |
| 8     | IRIS www.iris.irri.org | Sequence information (Genomic, Transcriptomic and Proteomic) | – | McLaren et al. 2005 |
| 9     | Oryzabase https://shigen.nig.ac.jp/rice/oryzabase/about/oryzabase • OryzaGenome http://viewer.shigen.info/oryzagenome/mapview/Top.do • Wild Rice Core Collection https://shigen.nig.ac.jp/rice/oryzabase/strain/wildCore/about | Strain stock information, mutant information, Chromosomal maps Genomic sequence information, SNPs, Variant Distribution | Strain stock information, mutant information, chromosomal maps and Variant distribution | Kurata and Yamazaki 2006 |
| 10    | Rice SNP-Seek Database http://snp-seek.irri.org/ | Sequence information (Genomic, Transcriptomic and Proteomic), SNP | – | Mansueto et al. 2017 |
| 11    | Rice Gene Thresher http://rice.kps.ku.ac.th/Site/index.html | Metabolic pathway, Single Featured Polymorphism, QTL, sequence information (Genomic, Transcriptomic and Proteomic), ESTs | Single Featured Polymorphism | Thongjuea et al. 2009 |
| 12    | Indian wild rice database http://nksingh.nationalprof.in:8080/iwrdb/ | Accessions, SNPs, SSRs, morphological characters and geographical locations | Morphological characters and Geographical locations | Tripathy et al. 2018 |
| 13    | RiceWiki http://ricewiki.big.ac.cn | Gene curation, genomic, transcriptomic, proteomic, metabolic pathway | Gene curation | – |
than 10,000 novel full-length protein-coding genes and a high number of presence–absence variations.

**Gramene Database**

Gramene is an integrated database developed for the comparative functional genomics in various plant species. It was developed as a repository in the year 2000 and has since been updated several times. The hyperlinks present on the Gramene website provide information about the gene ontology, quantitative trait loci (QTLs), molecular markers, metabolic pathways, genetic and physical maps. Gene ontology based pathway projections have been generated by using reference set of rice pathways. Genomic sequence data of 53 plant species and partial assemblies of about a dozen wild rice species is accessible at Gramene database. It contains reference quality genome information of wild rice accessions and also contains SNP data of three *Oryza* species: (i) 5418373 SNPs of *O. sativa ssp. indica* (NCBI dbSNP), (ii) 5512746, SNPs of *O. sativa ssp. japonica* (McNally et al. 2009, Zhao et al. 2010, NCBI dbSNP), and (iii) 7172036 SNPs of *O. glaberrima* (*Oryza* Genome Evolution Project). The presence of different types of data of rice accessions in Gramene database makes it a most suitable database for studies like comparative genomics, phylogenetic analysis and synteny mapping. The current version of Gramene has been updated with the gene models for *O. sativa ssp. japonica*. The regular database curation makes Gramene a reliable and an accurate platform for conducting database search. The data present in the Gramene can be accessed through both graphical as

---

### Table 2. Completed and ongoing sequencing projects on *Oryza* species

| Species                  | Genome type | Genome size (Mb) | GenBank Assembly Id   | Genbank Bioproject | References           |
|--------------------------|-------------|------------------|-----------------------|---------------------|----------------------|
| *O. sativa ssp. indica*  | AA          | ~400             | GCA_000004655.2       | PRJNA361            | Yu et al. 2002       |
| *O. sativa ssp. japonica*| AA          | ~400             | GCA_000005425.2       | PRJNA13141          | IRGSP 2005           |
| *O. rufipogon*           | AA          | ~445             | GCA_000817225.1       | PRJNE4137           | Huang et al. 2012    |
| *O. nivara*              | AA          | ~375             | GCA_000576065.1       | PRJNA48107          | Jacquemin et al. 2013|
| *O. barthii*             | AA          | ~335             | GCA_000182155.2       | PRJNA30379          | Jacquemin et al. 2013|
| *O. glaberrima*          | AA          | ~354             | GCA_000147395.2       | PRJNA13765          | Wang et al. 2014     |
| *O. glumaepatula*        | AA          | ~334             | GCA_000576495.1       | PRJNA48429          | Jacquemin et al. 2013|
| *O. meridionalis*        | AA          | ~340             | GCA_000338895.2       | PRJNA48433          | Jacquemin et al. 2013|
| *O. longistaminata*      | AA          | ~347             | NA                    | PRJNA245492         | Jacquemin et al. 2013|
| *O. punctata*            | BB          | ~423             | GCA_000573905.1       | PRJNA13770          | Jacquemin et al. 2013|
| *O. brachyantha*         | FF          | ~261             | GCA_000231095.2       | PRJNA70533          | Chen et al. 2013     |
| *Leersia perrieri*       | Out group   | ~323             |                       |                     | Wing, USA, draft 2012, completed unpublished |
| *O. officinalis*         | CC          | ~653             |                       |                     | Kurata, Japan, in progress |
| *O. eichingeri*          | CC          | ~650             |                       |                     | Kurata, Japan, in progress |
| *O. rhizomatis*          | CC          | ~650             |                       |                     | Kurata, Japan, in progress |
| *O. australiensis*       | EE          | ~960             |                       |                     | Panaud, France, in progress |
| *O. granulata*           | GG          | ~862             |                       |                     | Gao, China, in progress |

### Table 3. Sequenced organellar genome of *Oryza* species

| Rice variety                  | Genome size and organelle                      | References               |
|-------------------------------|-----------------------------------------------|--------------------------|
| *O. sativa ssp. japonica*     | 134,551 (Chloroplast)/490,520 (Mitochondria)  | Nock et al. 2011, Notsu et al. 2002 |
| *O. sativa ssp. indica*       | 134,496 (Chloroplast)/491,515 (Mitochondria)  | Tang et al. 2004, Tian et al. 2006 |
| *O. nivara*                   | 134,494 Chloroplast                           | Masood et al. 2004       |
| *O. rufipogon*                | 134,557 (Chloroplast)/559,045 (Mitochondria)  | Fujii et al. 2010, Waters et al. 2012 |
| *O. meridionalis*             | 134,551 Chloroplast                           | Nock et al. 2011, Waters et al. 2012 |
| *O. australiensis*            | 134,549 Chloroplast                           | Nock et al. 2011         |
| *O. longistaminata*           | 134,567 Chloroplast                           | Wambugu et al. 2015      |
| *O. barthii*                  | 134,674 Chloroplast                           | Wambugu et al. 2015      |
| *O. glumaepatula*             | 134,583 Chloroplast                           | Wambugu et al. 2015      |
| *O. officinalis*              | 134,911 Chloroplast                           | Wambugu et al. 2015      |
| *O. glaberrima*               | 134,606 Chloroplast                           | Wambugu et al. 2015      |
| *O. minuta*                   | 135,094 (Chloroplast)/515,022 (Mitochondria)  | Asaf et al. 2016, 2017   |
| *O. brachyantha*              | 134,604 Chloroplast                           | Liu et al. 2016          |
well as program based interfaces. Interested researchers can retrieve information on rice polymorphic markers and genetic linkage maps developed using different mapping populations from Gramene database. Gramene database has been used extensively for BLAST analysis, plant reactome studies, expression analysis, visualizing rice metabolic network and hormonal regulation under biotic and abiotic stress (Gupta et al. 2016).

**RITE-db: Database for Rice Repetitive and Transposable Elements**

A detailed annotation of repeat sequence and transposable elements is critical for understanding the complexity of genome. Several computational algorithms are available for the precise identification and annotation of repetitive sequences from the genome assemblies resulting into the development of large number of generic and specialized repeat databases. For rice, RITE-db is a specialized database offering information on the repeat sequences and transposable elements of about 11 diploid species of genus *Oryza* along with the closely related species *Leersia perrieri* (Copetti et al. 2015). One can retrieve the information about annotated repetitive sequences from different sections present in the RITE-db database, viz. PReDA (Plant Repeat Database), Repeat Explorer libraries and Full-Length Elements. PReDA offers information on repeat and transposable elements that are mostly present in the plants. Repeat explorer library provides information on repeated sequences of rice and its wild relatives identified *de novo* from whole genome sequencing data of 13 rice species. Full-Length Elements are the collection of complete TEs (LTR-RT, TRIM, SINE, Mutator, Na-DNAT, Helitron) of most of the super families isolated from sequence assemblies. In the latest update of RITE-db the false positive LTR-RTs (Long terminal repeat-retrotransposon) sequences were removed from the database. The database allows several functions including browsing and downloading of repetitive sequences and thus serves as an important resource for plant biology community (Choi and Purugganan 2017, Vieira et al. 2016).

**Rice annotation project database**

Rice annotation project database was developed in 2004 after the completion of sequencing of first rice genome with a major goal to provide annotation for all the rice genes. Since then, the database has been updated several times with the addition of other genomic assemblies. The last major update in RAP-DB was of genome assembly, Os-Nipponbare-Reference-IRGSP-1.0 (IRGSP-1.0) released in 2011 (Sakai et al. 2013). Afterwards, several minor updates have been made at the regular intervals in RAP-DB. It provides information on gene structure based on the full length cDNA sequences, ESTs (Expressed Sequence Tags) and protein sequences. Information regarding the gene annotation can be downloaded in the tab-delimited format. Routine curation of RAP-DB with new genome assemblies makes it a one stop solution for all the genomics queries related to rice and its close relatives. RAP-DB is accessible at the URL: http://rapdb.dna.affrc.go.jp/. Sakai et al. 2013, have explained the steps in detail to access the database contents in RAP-DB which includes “Gene map view and gene details view, homology search, batch retrieval, short read assembly browser, plant gene family database, Interactive Database of Cereal Gene Phylogeny (IDCGP) and Maintenance of legacy assemblies”. In addition, it is equipped with user friendly browsers like GBrowse and TASUKE, providing a decent visualization of gene sequence, transcript information, repeat regions, SNP/indels etc. Hyperlinks like OryzaBase, RiceXpro, Q-TARO, SALAD database and Gramene are also available on the homepage of RAP-DB making it more informative and user friendly.

**Rice Genome Knowledgebase**

RGKbase is a rice genome repository that provides information on comparative genomics and evolutionary relationships among different rice species and its close relatives. Presently, it holds information on 4 rice cultivars; Nipponbare (*japonica*), 93-11 (*indica*), PA64s (*indica*), the African rice (*O. glaberrima*) and a wild rice species (*O. brachyantha*). It has three major components: (i) integrated data curation for rice genomics and molecular biology; (ii) User-friendly viewers; (iii) Bioinformatic tools (Wang et al. 2012). It is a user-friendly database as it provides three easy-to-use browsers. First one, the GBrowse can be used for RefSeq mapping, gene structure and repetitive element annotations. GBrowse to GBrowse-syn provides the information about the conserved regions and evolutionary changes in the genome of the rice and its close relatives. Second, Genebrowser provides the structural information and functional annotation of the genes. Third, Circos, provides mapping of the genes and repetitive elements onto the chromosomes. Compositional and synteny analyses, gene family classifications, gene ontology terms and pathways and gene co-expression networks can be done by using the bioinformatics tools (Wang et al. 2012). This database is extensively used for the sequence analysis, SNP mining, comparative and evolutionary genomic studies (Chen et al. 2016, Song et al. 2018, Wang et al. 2014).

**Oryza Map Alignment project (OMAP)**

OMAP is a project developed by collaborative efforts of Arizona Genomics Institute, Cold Spring Harbor Laboratory, Purdue University and National Center of Gene Research, China. It was developed to provide the closed model system to unravel the evolution, physiology and biochemistry of the genus *Oryza*. Information regarding BAC libraries, physical maps, genomic and transcriptomic assembly are also available in OMAP. The database has many user friendly bioinformatics tools which are of extreme utility for research and educational purpose. OMAP has also started the In-situ conservation of wild
Oryza species with a goal to preserve the biodiversity of Oryza genus from the threats related to human activities and climate change (Jacquemin et al. 2013). A large number of AMPs (Advance Mapping Populations), like advance back cross (ABC), chromosome segment substitution lines (CSSLs) and recombinant inbred lines (RILs) in diverse backgrounds have been developed under OMAP. These mapping populations should facilitate the mapping of useful genes and thus can be efficiently utilized in breeding programs. The resources developed by OMAP are made available for the global rice research community upon request.

Genebank Project, NARO
Genebank project was started in Japan in 1985 as MAFF GeneBank project. In 2001, it was renamed as NIAS GeneBank. Since April 2016, NARO is handling the GeneBank project as central bank of diversity conservation. It is a central coordinating unit in Japan for conservation of plant, microorganism and DNA sequences. In plant section, 224000 accessions are available which includes landraces, improved varieties and wild varieties of 12 groups: rice, wheat and barley, legumes, root and tuber crops, millets and industrial crops, grasses and forage crops, vegetables, ornamental flowers and trees, tea, mulberry, fruit trees, tropical and sub-tropical crops. The DNA resources include 909,000 clones, including rice full length cDNA clones (~30,000 clones) and sequenced PAC/BAC clones of Nipponbare and Kasalath rice varieties. Hyperlinks for the list of clones are available at the website. The detailed information about each of the full length cDNA clone can be retrieved from Rice annotation database. Information about the chromosomal locations of the generated tiling path BAC clones of Kasalath rice variety on the basis of in silico mapping of BAC clones and DNA marker confirmation is also available on the website. Beside, NARO also holds two core collections on rice: NIAS world rice core collection and NIAS core collection of Japanese landraces. The genetic resources are made available to researchers for breeding and educational purpose on the payment of designate fee for the ordered material. Informative manuals are also available on the website of NARO for the optimal handling of genetic resources. NARO Genebank can be accessed at the URL: http://www.gene.affrc.go.jp.

IRIS: International Rice Information System
One of the major challenges in developing a unified repository is the proper cataloguing of all the available information with minimum redundancy, so that the species can be unambiguously identified (Bisht et al. 2018). To fulfil these requirements, Consultative Group for International Agricultural Research (CGIAR) and their collaborative partners have developed an open source database i.e. International Crop Information System for the proper management and integration of global information on genetic resources. ICIS can be accessed from the URL: www.icis.cgiar.org. The international rice information system contains a core database i.e., Genealogy Management System (GMS), which provides information on one and half million rice varieties, breeding lines and accessions. IRIS has been updated with the addition of new modules and harbors the information regarding the data generated by genetic experiments, whole-genome sequencing, transcriptomics, proteomics and functional genomics studies. In addition to this, it also offers free access to the links of specialized software packages and interfaces for the efficient retrieval of data on germplasm and species.

OryzaBase
OryzaBase was developed in 2000 as an integrated rice database sponsored by NBRP (National Bioresource Rice Project) in Japan. It provides information ranging from classical rice genetics to genomics resources, including details of wild type and mutant strains, chromosome maps, gene dictionary and many other featured links. It also harbors information on rice anatomy, its geographical distribution and several other parameters related to growth and development. It provides various user-friendly tools like DNA sequence database, sequence cutter, BLAST, Rice Id checker, Maptools, plant ontology, tool for segregation analysis. The database is accessible at URL: https://shigen.nig.ac.jp/rice/oryzaBase/about/oryzaBase. The two major component of the OryzaBase are ‘Wild Rice Core Collection’ and ‘OryzaGenome’.

Genetic Resources: Wild Rice Core Collection
It is a part of OryzaBase and provides information on 1729 accessions of 18 different wild rice species from 9 genomes AA, BB, CC, BBCC, CCDD, EE, FF, GG and HHJJ. These wild rice genetic stocks collected from different parts of the world are catalogued on parameters like species specific phenotype, distribution in the habitat, comparative strength and availability of seeds. Core collection is divided into 3 Ranks: Rank 1 contains 44 highly admirable representative accessions from 18 species. Rank 2 contains recommended collection of 65 accessions from all species. Rank 3 contains supplementary collection including 173 accessions. The database provides the phenotypic data and growth characterization of the core accessions. Wild core collection is generally used for verification of species (Lam et al. 2019).

OryzaGenome
The OryzaGenome database provides information about the geographical origin, phenotypic traits and genetic resources available for the wild Oryza species. Genome wide variants identified through NGS platforms in various wild rice accessions can also be visualized by using OryzaGenome. Reference sequences of cultivated rice accessions are also available. It also provides information about the SNPs present in the O. rufipogon accessions derived through imputation and deep sequencing. SNP viewer and Variant table are the two modes available for the visualization of genomic variants. Variant distribution can be visualized by SNP viewer whereas the precise location
of the variants can be identified by variant table and downloaded in the portable variant call format file (VCF) or tab-delimited file. The recent version of OryzaGenome i.e. OryzaGenome 2.0 (updated on 22 March, 2018) was released with reference sequence information of 208 additional accessions of 19 wild and 2 cultivated Oryza species. The data generated can be retrieved through DDBJ (DNA Databank of Japan) by using registered DRA (DDBJ Sequence Read Archive). Release 2.1 provides the fully developed SNP table which can be downloaded in the CSV (comma-separated values) format. In future, OryzaGenome has planned to provide the sequence information of 21 wild Oryza species with sequencing of several accessions. Pseudo-molecule and genome-wide variation information for other 20 wild Oryza species will also be provided on the basis of NIG/NBRP (National BioResource Project) collection (200 accessions) in Release 3.0. The SNP information of O. ruffipogon accessions present in this database can be used for marker development and comparative genomic studies. This database can also be used for integrated omic studies (Itoh et al. 2018).

**Rice SNP-Seek Database**

Rice SNP-Seek Database is a part of International Rice Informatics Consortium (IRIC) which provides the central access to the information regarding rice research data. It also offers computational tools for the discovery of novel genes or traits which can be further utilized for crop improvement programs. This database includes the phenotypic, genotypic, varietal information of rice and SNP genotyping data from the 3000 rice genome sequencing project. Phenotypic and passport data for 3000 rice varieties is provided by International Rice Genebank Collection Information System (IRGCIS). The Rice SNP-Seek Database can be accessed through URL: http://snp-seek.irri.org/. The homepage of Rice SNP-Seek Database contains hyperlinks named as ‘Genotype’ (for SNP queries among 3000 genomes), ‘Varieties’ (for variety passport and phenotype queries), ‘JBrowse’ (Rice Genome Browser), ‘GWAS’ (for GWAS results in the form of manhattan plot, allele and subpopulation distribution in the form of histogram, varieties in the form of lists) and ‘Help’ (from which information regarding the database usage and other documentation can be retrieved). Phylogenetic trees and MDS plots for varities which provide a comparative data on the basis of the various traits are also available. One can also order seeds from the IRRI Genebank collection by using the hyperlink “order seeds” on the homepage of the database. Morpho-agronomic, SNPs and structural variant data can also be downloaded from the database (by using URL: http://snp-seek.irri.org/download.zul) in different file formats like tabular format, tab-delimited text, flajack, hapmap, plink and MS excel format. SNP-Seek database offers real-time visualization of millions of SNPs in three thousands rice accessions which makes it a unique tool for allele mining (Mansueto et al. 2017). It can be used to explore the genetic diversity among large collection of germplasm accessions (Leung et al. 2015).

**Rice Gene Thresher**

Rice Gene Thresher is a bioinformatic tool developed by Rice Gene Discovery Unit and funded by National Center for Genetic Engineering and Biotechnology, Thailand. It was developed after the completion of the rice genome sequencing project with the aim to determine the number, location, regulation and the interaction of genes in QTL intervals. This helps the researchers in the discovery of probable candidate genes based on the information present in the database on different features: sequence, structure, expression and pathway. It also provides information about the genetic markers, genome annotation, expressed sequence tags (ESTs), protein domains, gene ontology, plant-stress responsive genes, metabolic pathways and predicted protein-protein interactions. A new web-based tool has been recently included in Rice Gene Thresher for Single-feature polymorphism (SFP) analysis by using rice Affymetrix GeneChip.

**Indian wild Rice Database**

Indian wild rice (IWR) database is developed by (NRCPB) National Research Center on Plant biotechnology, New Delhi. It hosts information on 556 accessions of wild rice collected from different agro-climatic zones of India. The information on the geographical locations, morphological characters and molecular characterization of each of the accessions has been provided in the database. The general information on the wild rice species can be obtained from the hyperlink ‘Wild Rice’ provided on the homepage of Indian Wild rice database. Whereas, specific information like passport data, population structure, plant morphology, leaf information, flower information, culm information, seed information, SNP score, and SSR score can be easily retrieved from the hyperlink Indian ORSC (Oryza ruffipogon Griff. Species Complex). Information regarding salinity and drought tolerance phenotype along with the molecular markers is also available in IWR database which can help geneticists to find and utilize agronomically useful genes in genetic crossing programs. Seeds of the accessions are also available for sharing with rice geneticists and breeders according to the prevailing IPR and Biodiversity rules and guidelines (Tripathy et al. 2018).

**RiceWiki**

RiceWiki is a wiki-based database for community curation of the rice genes. It has been developed by Professor Zhang and his team at Beijing Institute of Genomics, Chinese Academy of Sciences (BIG) in association with research partners at Huazhong Agricultural University, Beijing Institute of Technology, and Chinese Academy of Forestry. It is a publicly editable and open-content platform which can be accessible at the URL: http://ricewiki.big.ac.cn. Ricewiki has data of *O. sativa indica* 93-11 and
O. sativa japonica Nipponbare and covers 66,000 rice genes. As rice genes are the main component of Ricewiki, each gene is having its corresponding wiki page which contains ‘Annotated Information’, ‘Structured Information’, ‘Labs Working on This Gene’ and ‘References’, as well as summary for gene description. The information of genes present in RiceWiki was initially seeded from NCBI RefSeq, Ensembl, RAP-DB and MSU Rice Genome Annotation Project (http://rice.plantbiology.msu.edu). The salient features of Ricewiki includes: each content page is associated with a discussion page (where users can discuss content or leave a comment), a history page (where revision as well as its contributor can be recognized) and category terms (that increase the usability for information management). RiceWiki aims to exploit the full potential of the researchers for the curation of rice genes by using their collective intelligence and provides explicit authorship by quantifying users’ contributions in each curated gene. It also has a potential to build a rice encyclopedia based on community curation.

**Conclusion and future perspective**

Rapid advancements in DNA sequencing technologies and concurrent decline in sequencing cost have made whole genome sequencing a routine across labs. This democratization of sequencing technologies has set the stage for biologist to explore the information stored in DNA and its subsequent use in the crop improvement programs. In crops like rice, large number of different accessions of cultivated as well as wild relatives have already been sequenced, and enormous sequence and the derived secondary information is available across different databases. This vast trove of information has now started unravelling may salient features on domestication and evolution of Oryza genus (Stein et al. 2018). Furthermore, the sequence information is a key resource for designing of molecular markers and identification of novel genes and QTLs. The role of a comprehensive, structured and user friendly databases is seminal in dissemination of results of the large scale sequencing projects. The quantity and quality of data along with routine curation determines the applicability of a database. The databases that we have discussed here are the most widely used databases on wild relatives of rice. Apart from the generic databases having variety of information, some specialized databases like RiTE-db harboring information on some selective traits or genomic features are also discussed. Further, to assist researchers in choosing a suitable database, a concluding depiction on the databases of wild relatives of rice along with the different types of information stored in them is presented in Fig. 1.

Moreover, as the wild relatives of rice are having wide geographical distribution, collaborative international efforts should be undertaken to develop a unified repository of curated dataset facilitating the access of information to the global researchers. Such kind of databases along with the

---

**Fig. 1.** A pictorial representation of types of information stored in database related to wild relatives of rice.
recent toolbox of next generation plant breeding technique will surely pave a way ahead for developing rice varieties suitable for cultivation in diverse agro-climatic conditions.

**Author Contribution Statement**

R.K. wrote the manuscript with support from B.S. and T.K.M. provided critical feedback and helped to shape the manuscript. D.S.B. compiled all the information and contributed to the final manuscript.

**Acknowledgments**

The authors are grateful for the support given by Indian Council of Agricultural Research.

**Literature Cited**

Amante-Bordeos, A., L.A. Sitch, R. Nelson, R.D. Dalmacio, N.P. Oliva, H. Aswidinnoor and H. Leung (1992) Transfer of bacterial blight and blast resistance from the tetraploid wild rice *Oryza minuta* to cultivated rice, *Oryza sativa*. Theor. Appl. Genet. 84: 345–354.

Asaf, S., A.L. Khan, A.R. Khan, M. Waqas, S.M. Kang, M.A. Khan, R. Shahzad, C.W. Seo, J.H. Shin and I.J. Lee (2016) Mitochondrial genome analysis of wild rice (*Oryza minuta*) and its comparison with other related species. PLoS ONE 11: e0152937.

Asaf, S., M. Waqas, A.L. Khan, M.A. Khan, S.M. Kang, Q.M. Imran, R. Shahzad, S. Bilal, B.W. Yun and I.J. Lee (2017) The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. Front. Plant Sci. 8: 304.

Bisht, D.S., A.U. Solanke and T.K. Mondal (2018) Informatics of Wild Relatives of Rice. In: Mondal, T.K. and R.J. Henry (eds.) The Wild Oryza Genomes, Compendium of Plant Genomes, Springer, Cham, pp. 27–40.

Brozynska, M., D. Copetti, A. Furtado, R.A. Wing, D. Crayn, G. Fox, R. Ishikawa and R.J. Henry (2017) Sequencing of Australian wild rice genomes reveals ancestral relationships with domesticated rice. Plant Biotechnol. J. 15: 765–774.

Chen, J., Q. Huang, D. Gao, J. Wang, Y. Lang, T. Liu, B. Li, Z. Bai, G.J. Luis, C. Liang et al. (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. Nat. Commun. 4: 1595.

Chen, M., J. Xu, D. Devis, J. Shi, K. Ren, I. Searle and D. Zhang (2016) Origin and functional prediction of pollen allergens in plants. Plant Physiol. 172: 341–357.

Choi, J.Y. and M.D. Purugganan (2017) Evolutionary epigenomics of retrotransposon-mediated methylation spreading in rice. Mol. Biol. Evol. 35: 365–382.

Copetti, D., J. Zhang, M.E. Baidouri, D. Gao, J. Wang, E. Barghini, R.M. Cossu, A. Angelova, L.C.E. Maldonado, S. Roffler et al. (2015) RITE database: a resource database for genus-wide rice genomics and evolutionary biology. BMC Genomics 16: 538.

Das-Chatterjee, A., L. Goswami, S. Maity, K.G. Dastidar, S. Ray and A.L. Majumder (2006) Intrageneseon of a novel salt-tolerant L-myo-inositol 1-phosphate synthase from *Porteresia coarctata* (Roxb.) Tateoka (PoeniO1) confers salt tolerance to evolutionary diverse organisms. FEBS Lett. 580: 3980–3988.

Fujii, S., T. Kazama, M. Yamada and K. Toriyama (2010) Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. BMC Genomics 11: 209.

Garris, A.J., T.H. Tai, J. Coburn, S. Kresovich and S. McCouch (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169: 1631–1638.

Glassmann, J.C. (1987) Isozymes and classification of Asian rice varieties. Theor. Appl. Genet. 74: 21–30.

GRiSP (Global Rice Science Partnership) (2013) Rice Alm, 4th ed. Los Baños (Philippines): International Rice Research Institute. p. 283.

Gupta, P., S. Naithani, M.K. Tello-Ruiz, K. Chougule, P. D’Eustachio, A. Fabregat, Y. Jiao, M. Keays, Y.K. Lee, S. Kumari et al. (2016) Gramene Database: navigating plant comparative genomics resources. Curr. Plant Biol. 7: 10–15.

Huang, X., N. Kurata, X. Wei, Z.X. Wang, A. Wang, Q. Zhao, Y. Zhao, K. Liu, H. Lu, W. Li et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490: 497–501.

IRGSP (2005) The map-based sequence of the rice genome. Nature 436: 793–800.

Itoh, H., K.C. Wada, H. Sakai, K. Shibasaki, S. Fukuoka, J. Wu, J. Yonemaru, M. Yano and T. Izawa (2018) Genomic adaptation of flowering-time genes during the expansion of rice cultivation area. Plant J. 94: 895–909.

Jacquemin, J., D. Bhatai, K. Singh and R.A. Wing (2013) The International Oryza Map Alignment Project: development of a genome-wide comparative genomics platform to help solve the 9 billion-people question. Curr. Opin. Plant Biol. 16: 147–156.

Jeung, J.U., B.R. Kim, Y.C. Cho, S.S. Han, H.P. Moon, Y.T. Lee and K.K. Jena (2007) A novel gene, P40(t), linked to the DNA markers derived from NBS-LRR motifs confers broad spectrum of blast resistance in rice. Theor. Appl. Genet. 115: 1163–1177.

Khush, G.S. and K.C. Ling (1974) Inheritance of resistance to grassy stunt virus and its vector in rice. J. Hered. 65: 135–136.

Khush, G.S., E. Bacalango and T. Ogawa (1990) A new gene for resistance to bacterial blight from *O. longistaminata*. Rice Genet. NewsL. 7: 121–122.

Kurata, N. and Y. Yamazaki (2006) Oryzabase. An integrated biological and genome information database for rice. Plant Physiol. 140: 12–17.

Lam, D.T., B.C. Buu, N.T. Lang, K. Toriyama, I. Nakamura and R. Ishikawa (2019) Genetic diversity among perennial wild rice *Oryza rufipogon* Griff., in the Mekong Delta. Ecol. Evol. 9: 2964–2977.

Leung, H., C. Raghavan, B. Zhou, R. Oliva, I.R. Choi, V. Lacorte, M.L. Jubay, C.V. Cruz, G. Gregorio, R.K. Singh et al. (2015) Allele mining and enhanced genetic recombination for rice breeding. Rice (N Y) 8: 34.

Li, J., M. Xia, H. Qi, G. He, B. Wan and Z. Zha (2006) Marker-assisted selection for brown plant hopper (*Nilaparvata lugens*) resistance genes *Bph14* and *Bph15* in rice. Zhongguo Nong Ye Kexue 23: 2123–2137.

Liu, F., L.R. Tembrock, C. Sun, G. Han, C. Guo and Z. Wu (2016) The complete plastid genome of the wild rice species *Oryza brachyantha* (Poaceae). Mitochondrial DNA B Resour. 1: 218–219.

Mansuetto, L., R.R. Fuentes, F.N. Borja, J. Detras, J.M. Abriol-Santos, D. Chebotarov, M. Sanciangco, K. Palis, D. Copetti, A. Poliakov et al. (2017) Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res. 45: D1075–D1081.
Masood, M.S., T. Nishikawa, S.I. Fukuoka, P.K. Njenga, T. Tsudzuki and K.I. Kadowaki (2004) The complete nucleotide sequence of wild rice (Oryza nivara) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. Gene 340: 133–139.

McLaren, C.G., R.M. Bruskiewich, A.M. Portugal and A.B. Cosico (2005) The International Rice Information System: A platform for meta-analysis of rice crop data. Plant Physiol. 139: 637–642.

McNally, K.L., K.L. Childs, R. Bohnert, R.M. Davidson, K. Zhao, V.J. Ulut, G. Zeller, R.M. Clark, D.R. Hoen, T.E. Bureau et al. (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc. Natl. Acad. Sci. USA 106: 12273–12278.

Morishima, H. and Y. Sano (1992) Evolutionary studies in cultivated rice. Oxf. Surv. Evol. Biol. 8: 135.

Nock, C.J., D.L.E. Waters, M.A. Edwards, S.G. Bowen, N. Rice, G.M. Cordeiro and R.J. Henry (2011) Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol. J. 9: 328–333.

Notsu, Y., S. Masood, T. Nishikawa, N. Kubo, G. Akiduki, M. Nakazono, A. Hirai and K. Kadowaki (2002) The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol. Genet. Genomics 268: 434–445.

Sakai, H., S.S. Lee, T. Tanaka, H. Numa, J. Kim, Y. Kawahara, H. Wakimoto, C.C. Yang, M. Iwamoto, T. Abe et al. (2013) Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. Plant Cell Physiol. 54: e6.

Singh, B., N. Singh, S. Mishra, K. Tripathi, B.P. Singh, V. Rai, A.K. Singh and N.K. Singh (2018) Morphological and molecular data reveal three distinct populations of Indian wild rice Oryza rufipogon Griff. species complex. Front. Plant Sci. 9: 123.

Song, S., D. Tian, Z. Zhang, S. Hu and J. Yu (2018) Rice genomics: over the past two decades and into the future. Genomics Proteomics Bioinformatics 16: 397–404.

Stein, J.C., Y. Yu, D. Copetti, D.J. Zwickl, L. Zhang, C. Zhang, K. Chougule, D. Gao, A. Iwata, J.L. Goicoechea et al. (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza. Nat. Genet. 50: 285–296.

Su, C., Z. Hu, T. Zheng, K. Lu, Y. Zhao, W. Wang, J. Shi, C. Wang, J. Lu, D. Zhang et al. (2016) RPAN: rice pan-genome browser for ~3000 rice genomes. Nucleic Acids Res. 45: 597–605.

Tang, J., H. Xia, M. Cao, X. Zhang, W. Zeng, S. Hu, W. Tong, J. Wang, J. Wang, J. Yu et al. (2004) A comparison of rice chloroplast genomes. Plant Physiol. 135: 412–420.

Tello-Ruiz, M.K., J. Stein, S. Wei, J. Preece, A. Olson, S. Naithani, V. Amarasinghe, P. Dharmawardhana, Y. Jiao, J. Mulvany et al. (2016) Gramene 2016: comparative plant genomics and pathway resources. Nucleic Acids Res. 44: D1133–D1140.

Thongjuea, S., V. Ruanjaichon, R. Bruskiewich and A. Vanavichit (2009) Rice Gene Thresher: a web-based application for mining genes underlying QTLS in rice genome. Nucleic Acids Res. 37: D996–D1000.