agronomy

MDPI

Communication

RicePedigree: Rice Pedigree Database for Documentation and Assistance in Rice Breeding

Dong-U Woo [1], Yejin Lee [1], Ho-Hwi Jeon [1], Halim Park [1], Jin-Hwa Park [1], Sung-Hoon Choi [1], Chang-Min Lee [2], Youngjun Mo [3] and Yang-Jae Kang [1,4,*]

[1] Division of Bio & Medical Bigdata Department (BK4 Program), Gyeongsang National University, Jinju 52828, Republic of Korea
[2] National Institute of Crop Science, Rural Development Administration, Wanju 55365, Republic of Korea
[3] Department of Crop Science and Biotechnology, Jeonbuk National University, Jeonju 54896, Republic of Korea
[4] Division of Life Science Department at Gyeongsang National University, Jinju 52828, Republic of Korea
* Correspondence: kangyangjae@gnu.ac.kr; Tel.: +82-55-772-1352; Fax: +82-55-772-1349

Abstract: For the purpose of breeding documentation, researchers and breeders kept handwritten records of the breeding history, including parental information and breeding methods. The cultivars were used again as parents for further breeding, and modern cultivars of rice have a wide range of alleles from many generations of parents and ancestors. To understand such a breeding history, it is necessary to ask around for relevant information, which is then usually documented in Excel or Word by multiple breeders or breeding institutes. Here, we constructed RicePedigree, which contains the breeding history of rice based on the documents provided by the Rural Development Administration (RDA) in Korea. We devised a simple method for collecting a breeding history and storing it in a database. RicePedigree is a web-based application on the database that facilitates researchers' and breeders' utilization of the breeding history of rice. Based on the query cultivar name, it will return a hierarchical tree of breeding histories and a list of cultivars and breeding lines that contain query cultivars in their breeding histories. This app would be a good way to review and keep track of information about current and future cultivars.

Keywords: rice; breeding history; database; web application

check for updates

Citation: Woo, D.-U.; Lee, Y.; Jeon, H.-H.; Park, H.; Park, J.-H.; Choi, S.-H.; Lee, C.-M.; Mo, Y.; Kang, Y.-J. RicePedigree: Rice Pedigree Database for Documentation and Assistance in Rice Breeding. *Agronomy* **2023**, *13*, 69. https://doi.org/10.3390/agronomy13010069

Academic Editor: Salvatore Ceccarelli

Received: 4 December 2022
Revised: 21 December 2022
Accepted: 22 December 2022
Published: 25 December 2022

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

It is believed that plant breeding started with agriculture. People chose edible plants that had certain desirable traits, and, over time, these useful traits accumulated. Modern breeding started with the artificial crossing of parental lines based on the understanding of inheritance and genomics [1]. In modern breeding, many alleles from different germplasms are mixed together, and useful wild alleles are added to elite cultivars [2].

Along with the advance of high-throughput genotyping technologies, it is now possible to directly observe the alleles of loci [3]. This enables the understanding of the effects of alleles or genomic regions on certain phenotypes [4]. Moreover, it allows the prediction of phenotypes based on genotypes with a genomic selection scheme [5]. Eventually, understanding the genomic architecture will help us formulate a precise breeding strategy to cope with market needs and global climate changes. However, the number of polymorphic genetic factors in the genome is too high to explain their individual and epistatic effects, especially for the quantitative traits in agriculture. This limits the power of genome-wide association studies to find important quantitative trait loci (QTL) in rice [6].

The pedigree information has been documented by breeders and researchers. As the breeding period of crops such as rice is very long—nearly 10 years—the pedigree documentations of breeding histories are well stored in many document types. The breeding history consisted of the artificial crossing of selected parental lines and their strategy for achieving the desired phenotype. Even though the pedigree documentation may not contain every

circumstance of the breeding history, it is still referred to by breeders to determine the breeding plan for the next year. Moreover, it has been suggested that the pedigree would also be helpful in enhancing genomic selection [7].

The pedigrees have been documented with handwriting or are stored in PowerPoint image files. Because of this, it is hard to use or share information in a systematic way. The Rural Development Administration (RDA) in Korea created Word files for the documentation of the pedigrees to address this issue. However, the Word file format was not good enough for other breeders and researchers to share and concurrently update. There have been efforts to create pedigree exploration services. As an integrated genomic database for rice, Ricebase (http://ricebase.org, accessed on 21 December 2022) was created and provided interactive Scalable Vector Graphics representations of the pedigree trees [8]. However, the service is currently inaccessible. Pedimap is another software tool for visualizing the provided pedigree input into a tree-like image [9]. The Korean Rice Breeding Information Management System (KRBIMS), a database and management tool that offers integrated management of rice breeding information, was developed but was not intended for use by the public [10]. The China National Research Institute operates a rice variety database (https://www.ricedata.cn/variety/index.htm (accessed on 21 December 2022)), which contains detailed descriptions of the history and characteristics of various rice cultivars, as well as pedigree trees. In this paper, we describe a simple database system that can hold information about RDA pedigrees. We also made a web-based application called RicePedigree that allows you to view pedigrees and search for cultivar names. We made efforts to improve the user interface and visualization of the pedigree in RicePedigree, in order to make it easier for users to use. Additionally, we translated the cultivar names into English in order to reach a wider audience. The web application can be accessed at the web address, http://ricepedigree.plantprofile.net/ (accessed on 21 December 2022). This program will be a good way to evaluate and keep track of cultivars, now and in the future.

2. Materials and Methods

2.1. Breeding History Database and the Web Application Construction

The RDA in South Korea kindly provided the breeding history documents. The documents contained pedigree trees, which had to be manually converted into a tab-separated form for the conversion into a database. The tab-separated table was converted into SQLite (https://sqlite.org/index.html (accessed on 21 December 2022)). The web application was constructed using the Python-Django framework (https://www.djangoproject.com/ (accessed on 21 December 2022)) and Semantic-UI (https://semantic-ui.com/ (accessed on 21 December 2022)). The pedigree of each accession was visualized by D3.js (https://d3js.org/ (accessed on 21 December 2022)).

2.2. Clustering of NCBI-Deposited Accessions

NCBI-deposited sequences of rice accessions were downloaded using the SRA toolkit [11]. The downloaded accessions and their SRR IDs are as follows: Namcheon (SRR12701914), IR24 (SRR19634031), Milyang29 (SRR12701912), Pungsan (SRR12701908), Hanareum (SRR10083950), Dasan (SRR1014726), Hwayeong Byeo (SRR10083951), Samgwang (SRR10083924), Koshihikari (DRR099981), Milky queen (SRR12701969), Haedamssal (SRR12701921), and Sangju Byeo (SRR10083937). The reference genome for the read mapping was Oryza_sativa.IRGSP-1.0 [12]. The reads were mapped to the reference genome with BWA [13]. The variants were called using Samtools and Bcftools [14]. For the clustering, the variants on the genic regions were extracted, clustered, and visualized with Seaborn [15].

3. Results

3.1. Systematic Representation of Pedigree

The pedigree documents (2012–2018) from the RDA in Korea contain images showing the breeding histories (Figure 1A). The documents were written in Korean for the 108 rice cultivars registered from 2012 to 2018. The registered cultivars were described with the

cultivar name, the line name, the registered year, the combination of crosses, the institute, the ecotype, the maturity group, and the heading date. Most importantly, it depicted the breeding history, which shows several generations of parents. Figure 1 shows the registered cultivar, Haepum (해품), and the breeding history shows additional upstream parental lines from the direct parents. After converting the graphical representation of the breeding history of 108 registered cultivars into the node-to-node relationship, the pedigree tree of Haepum was expanded, as shown in Figure 1B, because the breeding histories of other registered cultivars could give more information about the pedigree of upstream parental lines.

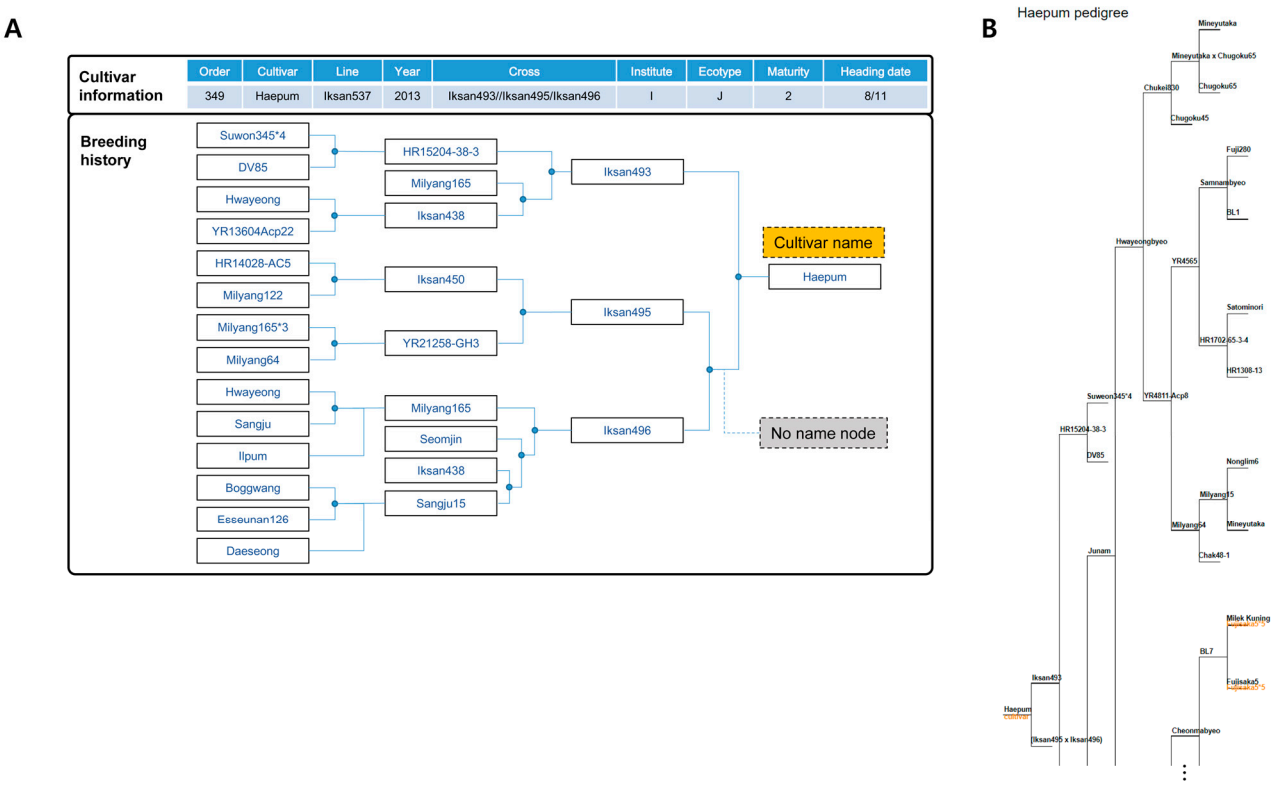


Figure 1. Documentation of cultivar "Haepum" (**A**) Image-based documentation from RDA; the words in the boxes are the translation of Korean into English. The yellow boxes show the registered cultivar names. The gray box indicates the example node without an explicit name or ID. (**B**) Hierarchical tree of "Haepum" from the RicePedigree database.

3.2. Data Processing of Pedigree Information

The relationship between parents and children as well as the breeding strategy, such as the number of backcrosses, are included in the pedigree information (Figure 2A). Numerous parental cultivars or breeding lines, as well as occasionally the breeding method, are shown in the pedigree documents' graphical representations of the 108 cultivars. We manually turned the images into a table with columns for "Cultivar", and "Parent" (Figure 2B). Breeding methods such as backcrossing and mutagen treatments (Sodium azide, Ethyl methanesulfonate, N-Nitroso-N-methyl urea, and gamma-rays) were indicated in the images, and we also added the "Method" column in the table. The "Method" details are shown on the pedigree tree's corresponding edges (Figure 2D). In the case of the nodes without names or IDs (Figure 1A), we named them according to the parental lines (eg. "Iksan495 x Iksan496"). Moreover, Korean cultivar names were consistently Romanized according to the Korean Seed & Variety Service (KSVS) database (http://www.seed.go.kr/seed/199/subview.do/ (accessed on 21 December 2022)). A total of 1153 cultivar-parent relationships were coded into a single table and transferred to the SQLite database engine (Table S1). In order to provide quick access to the pedigree tree, we pre-calculated the tree

based on all record "Cultivar" and "Parent" relationships. The computed pedigree tree is once more kept in a different table with "Cultivar" and "JSON" columns. The pedigree tree of the cultivar "Palbangmi" is described in JSON format in Figure 2C. The website would immediately display the pre-calculated JSON data (Figure 2D).

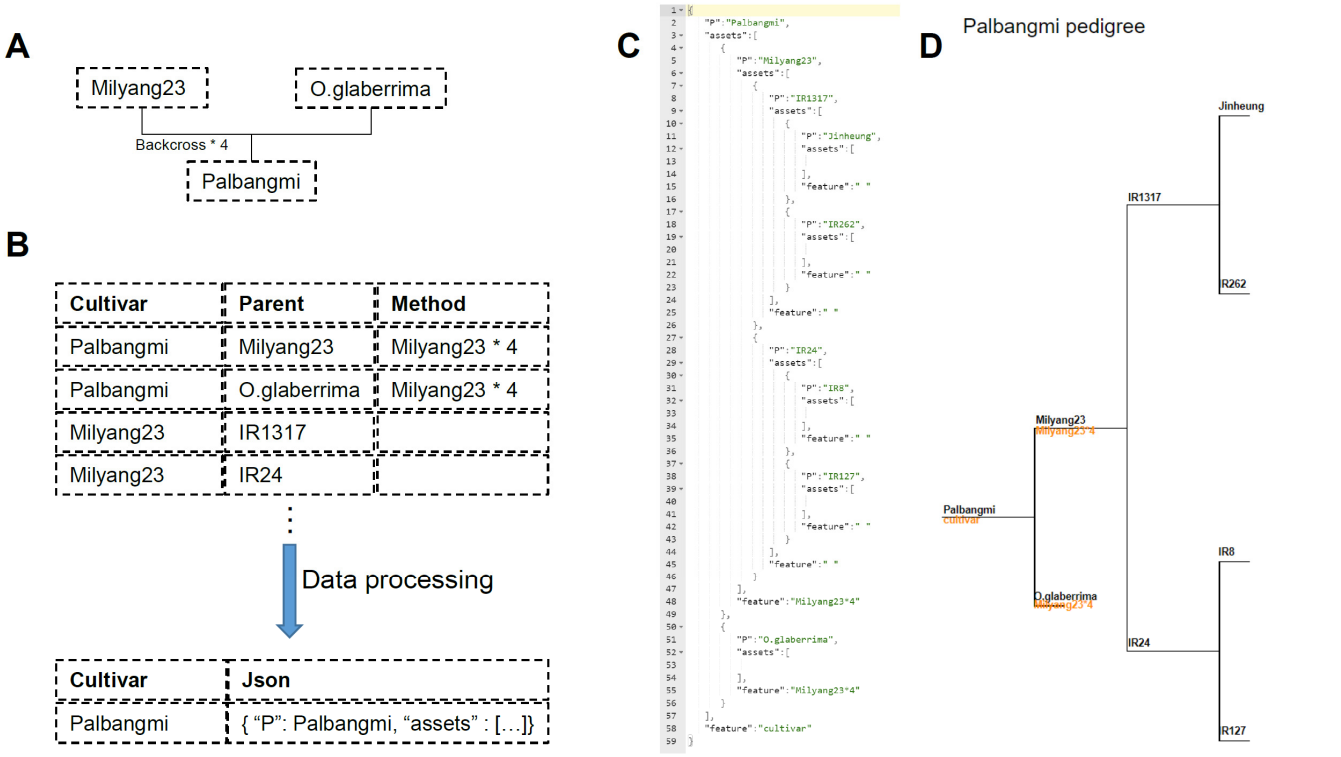


Figure 2. Example of the table coding scheme. (**A**) Parent-cultivar relationship in the pedigree image. (**B**) The parent-cultivar relationship is coded into two rows containing maternal and paternal lines. The method column contains the breeding methods, such as the number of backcrosses or mutagen treatments. The table is converted into a parent-child tree for each cultivar with a data processing function. (**C**) Converted JSON format describing the pedigree of "Palbangmi". (**D**) Hierarchical tree based on consolidated parent-cultivar relationships of "Palbangmi". The breeding method information is annotated with orange letters.

3.3. Database and Web Application Development

We used the SQLite database engine to host the curated table (https://www.sqlite.org/index.html (accessed on 21 December 2022)). To interact with the SQLite database, an object-relational mapping layer (ORM) of the Django framework was used (https://djangoproject.com/ (accessed on 21 December 2022)). After importing the processed cultivar-parent table into the SQLite database, we used the Django framework to build a web application (RicePedigree) that interacts with the database. The database and web application are served by open-source software Nginx (https://www.nginx.com/ (accessed on 21 December 2022)).

The interface of the web application includes "Search", "Pedigree visualization", and "Pedigree search result" panels (Figure 3). The "Search" panel accepts the query cultivar name and recommends similar cultivar names based on the first few inputs (Figure 3A). The "Pedigree tree visualization" panel shows the pedigree tree of the query cultivar or breeding line. It combines all parent-cultivar relationships from the database and returns one hierarchical tree visualized using the Data-Driven Documents (d3.js) JavaScript library (https://d3js.org/ (accessed on 21 December 2022)) (Figure 3B). Hence, the resulting pedigree tree is sometimes larger than the original breeding history image in the RDA documentation (Figure 1B) and the process of adding new relationships is also simplified.



Figure 3. The interface of the RicePedigree web page. (**A**) Search panel recommends possible accession names based on the first few input letters. (**B**) Pedigree tree visualization panel shows parental lines from the database. (**C**) Pedigree search result panel shows a list of cultivars or breeding lines that contain the query in their breeding histories.

Figure 3C shows the "Pedigree search result" panel, which shows a list of cultivars or breeding lines that have the query cultivar in their breeding history. This function allows the breeders to find cultivars or breeding lines that are related to the cultivars in the query based on pedigree trees. This, along with genomic information, will be useful when breeders select parental lines for their breeding plans.

3.4. Case Study: Supporting Evidence for Genomic Similarity

Using the whole genome sequencing data for those rice accessions that are stored in the NCBI-SRA, the SNP profile can be used to estimate the genetic distance between those accessions. In most cases of a concatenated SNP matrix of rice breeds, the types of rice (Indica and Japonica) are clearly grouped together. If the breeding histories between the species are given, the genomic closeness of neighboring cultivars within each cluster can be explained more effectively. To test RicePedigree in this scenario, we downloaded 12 WGS datasets from NCBI-SRA [16], created an SNP matrix, and displayed a sample cluster (Figure 4A). The tight cluster included Japonica-type accessions, such as: "Hwayeong", "Samgwang", "Koshihikari", "Milky queen", "Haedamssal", and "Sangju", while the other cluster included Indica-type accessions: "Namcheon", "IR24", "Milyang29", "Pungsan", "Hanareum", and "Dasan". When "Dasan" is searched for in the RicePedigree database, its pedigree tree is drawn, and the "Pedigree search result" panel displays additional accessions with "Dasan" in their pedigree (Figure 4B). It's notable that "Hanareum" is listed under "Pedigree search result" supporting its neighborhood with "Dasan" based on the sample clustering. The pedigree tree of "Hanareum" revealed "Namcheon", "Dasan", "IR24", "Milyang29", and "Pungsan", which are also closely clustered in the sample clustering result (Figure 4C).

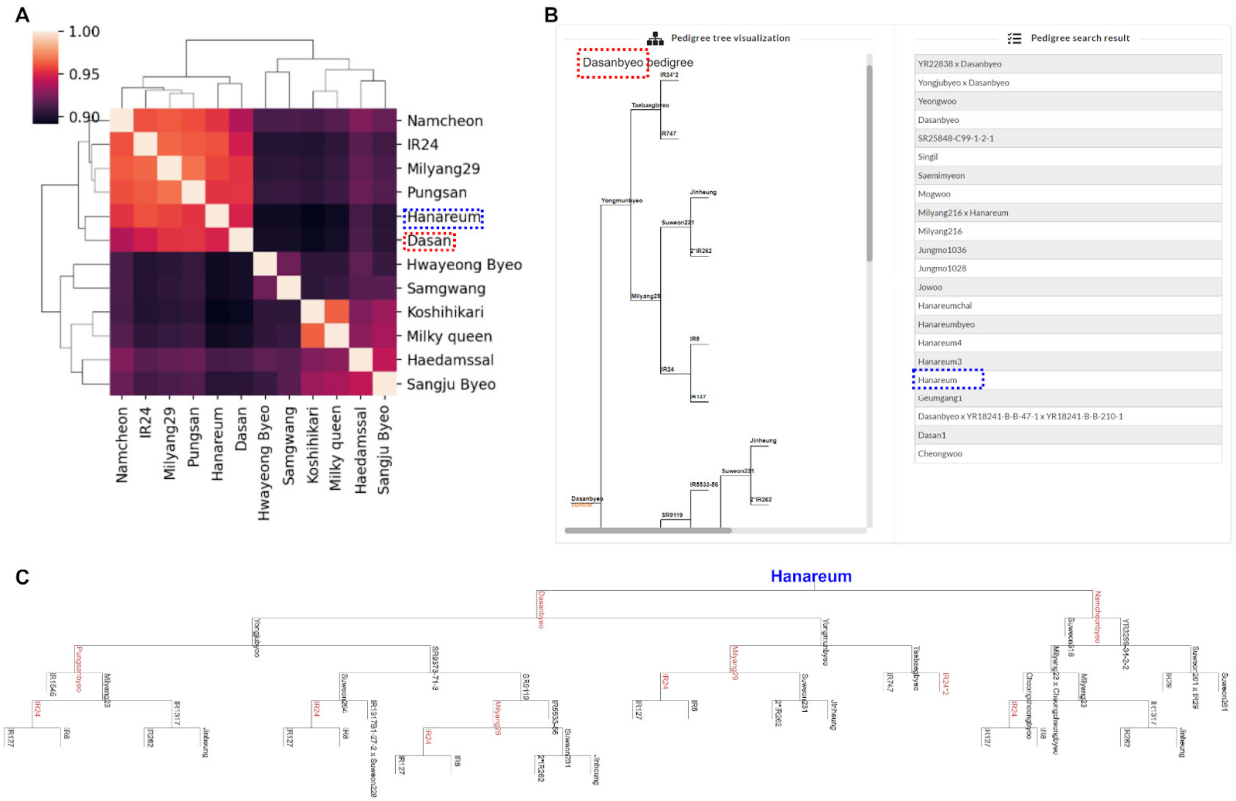


Figure 4. A case study using the RicePedigree database to translate the genomic clustering results based on resequencing data. (**A**) Expected grouping for japonica- and indica-type cultivars, based on genomic SNP analysis. (**B**) RicePedigree search result on query "Dasanbyeo", showing its pedigree tree and list of cultivars and germplasms that contain "Dasanbyeo" in their pedigree. (**C**) Pedigree of "Hanareum", showing upstream parental lines in breeding history.

4. Discussion

RicePedigree is a simple web application that serves pedigree information on rice cultivars and breeding lines. Previously, managing breeding history data with a word processor made it difficult to share and update pedigrees. Due to the document's limited space, the pedigree would only be written up to a few precedents. To trace back to earlier generations, researchers would need to find another page. The fact that the pedigree documents were kept in personal storage and could not be routinely updated made it difficult to share them with other researchers. When describing new cultivars, the pedigree images should have been written in a redundant manner, including information about the most recent generations that have already been provided on other pages.

With RicePedigree, it is straightforward to update, share, and explore the breeding history of queries. A new cultivar or breeding line can be added to the database, using the simple database structure, by including additional breeding combinations. Also, because this is on the web, anyone can look up the breeding history by using the cultivar name. The "Pedigree search result panel", which displays cultivars with query cultivars in their breeding histories, provides breeders with additional information for the parent selection step. In breeding programs, it helps when the breeders and researchers choose the lines of the parents to widen or narrow the combination of alleles.

Additionally, the pedigree data can be used to analyze rice germplasm whole genome sequencing data, to determine the causes of the close genotypic relationships between the germplasms. At the National Center for Biotechnology Information (NCBI), researchers can find more than 400 sets of resequencing data on rice germplasms. Researchers can make assumptions about how the alleles may have been traded or passed down along

with the pedigrees. Also, the pedigree tree and whole genome data can be used to find the important alleles or sets of alleles for breeding.

Here, we demonstrated how a simple database design and web front-end make it easier to manage and use rice pedigree data. In the future, by including a genomic profile for each cultivar and accession, this database can serve as the hub for managing rice resources, because the breeding history of the accessions is closely related to their genomic profiles. RicePedigree will be updated yearly with the breeding history of the registered cultivars provided by the RDA in Korea. This database and web application will be a good way to evaluate and keep track of cultivars, now and in the future.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy13010069/s1. Table S1. Processed pedigree information in single table format.

Author Contributions: D.-U.W., Y.L. and Y.-J.K. determined the data processing scheme, developed the database, built the web-application and wrote the manuscript. H.P., H.-H.J., S.-H.C. and J.-H.P. manually processed the pedigree image data. Y.M. and C.-M.L. provided the pedigree information and curated the processed pedigree table. All authors have read and approved the published version of the manuscript.

Funding: This work was carried out with the support of "BioGreen21 Agri-Tech Innovation Program (Project No. PJ016494)" Rural Development Administration, Republic of Korea.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Prohens, J. Plant Breeding: A Success Story to be Continued Thanks to the Advances in Genomics. *Front. Plant Sci.* **2011**, *2*, 51. [CrossRef] [PubMed]
2. Breseghello, F.; Coelho, A.S.G. Traditional and Modern Plant Breeding Methods with Examples in Rice (*Oryza sativa* L.). *J. Agric. Food Chem.* **2013**, *61*, 8277–8286. [CrossRef] [PubMed]
3. Kang, Y.J.; Lee, T.; Lee, J.; Shim, S.; Jeong, H.; Satyawan, D.; Kim, M.Y.; Lee, S. Translational genomics for plant breeding with the genome sequence explosion. *Plant Biotechnol. J.* **2016**, *14*, 1057–1069. [CrossRef] [PubMed]
4. Verdeprado, H.; Kretzschmar, T.; Begum, H.; Raghavan, C.; Joyce, P.; Lakshmanan, P.; Cobb, J.N.; Collard, B.C. Association mapping in rice: Basic concepts and perspectives for molecular breeding. *Plant Prod. Sci.* **2018**, *21*, 159–176. [CrossRef]
5. Spindel, J.; Begum, H.; Akdemir, D.; Virk, P.; Collard, B.; Redoña, E.; Atlin, G.; Jannink, J.-L.; McCouch, S.R. Genomic Selection and Association Mapping in Rice (Oryza Sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. *PLoS Genet.* **2015**, *11*, e1004982.
6. Zhou, X.; Huang, X. Genome-Wide Association Studies in Rice: How to Solve the Low Power Problems? *Mol. Plant* **2019**, *12*, 10–12. [CrossRef] [PubMed]
7. Ankamah-Yeboah, T.; Janss, L.L.; Jensen, J.D.; Hjortshøj, R.L.; Rasmussen, S.K. Genomic Selection Using Pedigree and Marker-by-Environment Interaction for Barley Seed Quality Traits From Two Commercial Breeding Programs. *Front. Plant Sci.* **2020**, *11*, 539. [CrossRef] [PubMed]
8. Edwards, J.D.; Baldo, A.M.; Mueller, L.A. Ricebase: A Breeding and Genetics Platform for Rice, Integrating Individual Molecular Markers, Pedigrees and Whole-Genome-Based Data. *Database* **2016**, *2016*, 1–6. [CrossRef] [PubMed]
9. Voorrips, R.E.; Bink, M.C.A.M.; Van De Weg, W.E. Pedimap: Software for the Visualization of Genetic and Phenotypic Data in Pedigrees. *J. Hered.* **2012**, *103*, 903–907. [CrossRef] [PubMed]
10. Song, M.-T.; Lee, J.-K.; Yang, S.-J.; Choi, H.-C.; Hwang, H.-G.; Kim, H.-Y.; Park, K.-G.; Cho, Y.-S.; Moon, H.-P.; Han, W.-S.; et al. KRBIMS (Korean Rice Breeding Information Management System): A Database for Rice Breeding Information Management. *Korea J. Breed. Sci.* **2002**, *34*, 111–115.
11. Leinonen, R.; Sugawara, H.; Shumway, M. International Nucleotide Sequence Database Collaboration The Sequence Read Archive. *Nucleic Acids Res.* **2011**, *39*, D19–D21. [CrossRef] [PubMed]
12. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the Oryza Sativa Nipponbare Reference Genome Using next Generation Sequence and Optical Map Data. *Rice* **2013**, *6*, 4. [CrossRef] [PubMed]
13. Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]

14. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
15. Waskom, M.; Botvinnik, O.; O'Kane, D.; Hobson, P.; Ostblom, J.; Lukauskas, S.; Gemperline, D.C.; Augspurger, T.; Halchenko, Y.; Cole, J.B.; et al. seaborn: Statistical data visualization. *J. Open Source Softw.* **2018**, *6*, 3021. [CrossRef]
16. Kim, T.-S.; He, Q.; Kim, K.-W.; Yoon, M.-Y.; Ra, W.-H.; Li, F.P.; Tong, W.; Yu, J.; Oo, W.H.; Choi, B.; et al. Genome-wide resequencing of KRICE_CORE reveals their potential for future breeding, as well as functional and evolutionary studies in the post-genomic era. *BMC Genom.* **2016**, *17*, 408. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.