Application of Pedimap — a pedigree visualization tool — to facilitate the decisioning of rice breeding in Sri Lanka

P.G.R.G. Rathnayake[1], S.M. Sahibdeen[1], U.A.K.S. Udawela[2], C.K. Weebadde[3], W.M.W. Weerakoon[4], S.D.S.S. Sooriyapathirana[1,5*]

[1]Department of Molecular Biology and Biotechnology, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka

[2]Rice Research and Development Institute, Bathalagoda, Sri Lanka

[3]Department of Plant, Soil and Microbial Sciences, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI, USA.

[4]Department of Agriculture, Peradeniya, Sri Lanka.

[5]Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka

*Corresponding author:

sunethuop@gmail.com (SDSSS)

ABSTRACT

The development of rice cultivars with desirable traits is essential. The decision-making is a crucial step in rice breeding programs. The breeders can make efficient and pragmatic decisions if an organized pedigree visualization platform is available for the material of the rice breeding germplasm. The staple food in Sri Lanka is rice, and there is a great demand for improved varieties with high yield and other promising traits. In the present study, the available data of all the rice varieties released by Rice Research and Development Institute, Sri Lanka, and the related landraces and genotypes were arranged in Pedimap, a pedigree visualization tool. Pedimap can showcase pedigree relationships, phenotypic, and molecular data. The Identity by Descent (IBD) probabilities were calculated using FlexQTL software and included in the Pedimap database. The parentage selection based on the variations of phenotypic traits, selection of marker alleles for molecular breeding, and detection of the founders of genetic effects can be swiftly conducted using Pedimap. However, the power of harnessing the value of Pedimap for making breeding decisions relies on the availability of data for the traits, markers, and genomic sequences. Thus, it is imperative to characterize the breeding germplasms using standard phenomic and genomic characterization procedures before organized into Pedimap. Thereby, the worldwide breeding programs can benefit from each other to produce improved varieties to meet global challenges.

Keywords: Breeding database, Breeding decisions, Marker assisted breeding, Pedigree visualization, Planning crosses

INTRODUCTION

Rice is one of the major crops in the world, with an annual production over 700 million metric tons [1]. Half of the world population consumes rice as the staple food [2]. Currently, the demand for rice is rapidly increasing due to the growth of the human population [3]. However, the current rice production cannot meet the increasing demand causing severe food security issues. The biotic and abiotic stresses also exert a negative influence on rice production [4]. The rice farming is also a way of living for many people, especially in numerous Asian countries [5]. At present, 1.8 million Sri Lankan families engage in rice farming over 870,000 hectares [6]. The annual rice production in Sri Lanka is approximately 2.3 million metric tons (MT), which is insufficient to fulfill the domestic rice demand of 3.0 million [7]. Hence, the Sri Lankan government spends about USD 400 million to import rice annually [7,8].

The rice production is mainly affected by drought and irregular rainfall patterns caused by climate change [9,10,11], adverse soil conditions such as salinity [6,12], and pest and disease attacks [13]. The biotic and abiotic stresses in rice farming can be controlled using numerous agronomic practices such as irrigation, drainage, fertilization, and the application of pesticides. However, the rate of success of the controlling methods is limited [13] due to the unpredictable nature of climate change, soil degradation, variations in pest dynamics, and development of pest resistance [14]. Therefore, breeding is considered as the most successful strategy to produce high yielding and stress resilient rice varieties [15]. The improved rice genotypes can also contain the traits for higher consumer preference and organic farming [16]. In the past, the rice varietal improvement was conducted with classical breeding techniques, which are tedious, lengthy, and less feasible in cases such as breeding for pest resistance and submergence tolerance. However, the marker-assisted breeding (MAB) is employed in modern breeding programs to introgress valuable genetic loci from landraces and traditional varieties [17,18]

and the desirable haplotypes of Quantitative Trait Loci (QTL) to the improved rice varieties [19-21].

The decision-making process in a breeding program is crucial for successful outcomes. The formulation of decisions before breeding is a multi-step process that consists of the identification of breeding priorities, determination of the genetics of target traits, and employment of pre-breeding methods if required. The economic and technical feasibility, number of parents for crosses, number of selfing and outcrossing cycles, length of the breeding program/cycles, and identification of the selection methods must also be assessed [22]. In the decisioning process, initially, the market trends based on consumer and other stakeholder preferences must be recognized [23]. Subsequently, the novelty and the uniqueness of the breeding objective must be assessed before the execution of the breeding program [22].

The selection of suitable varieties or individual plants as parents and the determination of the selection methods are the two most critical aspects in planning breeding programs [24]. The parental selection depends on the number of prioritized traits for breeding. When multiple characteristics are to be introgressed, the breeders require a prioritized order of parents for stepwise crossing and selection [25,26]. The decision-making process in breeding is entirely based on the available information on phenotypes, genotypes, pedigree, available budget, field and greenhouse space, desired time-to-market etc. Although the data for decision-making for breeding are indispensable, haphazardly collected information would provide less value to the breeders. In many conventional breeding programs, most of the data are recorded in field notebooks and stored in the breeding stations, while very little information is available as computerized databases. If an organized database containing all the essential information for

the rice varieties released and the parental genotypes used in breeding, the decisions can be easily made.

The construction of a database with all the necessary information from varieties and their parents promotes the capacity of data sharing, mining, visualization, and retrieval [27]. Pedimap is a pedigree visualization software. The data needed can be imported to Pedimap from FlexQTL, or with some custom script from any other database program. Pedimap is used by many contemporary plant genetics and breeding programs worldwide. As stated in Voorips et al. (2012) [28], Pedimap can be used to record and utilize breeding history. Pedimap illustrates the available phenotypic and genetic data through pedigrees. All the information, including parentage, qualitative and quantitative data, marker alleles/genotypes, and the calculated identity-by-descent (IBD) probabilities can be presented in Pedimap. Currently, breeders prefer to use pedigree visualization tools like Pedimap since it allows them to access the large pool of genetic and phenotypic data quickly and generate pedigrees that are essential in making breeding decisions.

In Sri Lanka, Rice Research and Development Institute (RRDI) is the sole organization conducting the rice breeding programs for the national needs. Therefore, in the present study, we report an attempt to organize the information of the released varieties and the parental genotypes of RRDI breeding programs as a Pedimap based database which is a valuable step to take accurate breeding decisions and speed up the process of releasing novel varieties.

MATERIALS AND METHODS

Data Curation

The data were collected from RRDI, Sri Lanka and classified under three main categories, namely pedigree history, phenotypic data, and molecular data on rice varieties/landraces/genotypes (herein after collectively referred to as cultivars). The male and female parents and the order of crosses were taken under pedigree history. The average yield of the rice plants, the maturity period in different growing seasons (*Yala* and *Maha* seasons of Sri Lanka [29]), plant height, basal leaf sheath color and additional color patterns, recommended type of the land, level of phosphorus deficiency tolerance, amount of brown rice recovery, milling recovery, head rice recovery, amylose content, gelatinization temperature, weight of 1000 grains, shape of the grain, pericarp color, weight of a kg, color of the buff coat and resistance/susceptibility to pests and diseases; brown planthopper (BPH), bacterial leaf blight and rice blast disease were recorded under phenotypic data (S1 Table). The available DNA marker alleles, marker positions in the linkage map, and allelic scores were entered under molecular data [30-33] (S2 Table).

Pedimap Procedure

A Pedimap input data file is created in MS Excel (2019), and the data file is exported as a tab-delimited text (.txt) file (S3 Table). The input file contains four main subdivisions; header, pedigree, marker data, and IBD probability section (Fig 1). The header consists of five essential elements and one additional element. The name of the population and symbols for unavailable or missing data, null homozygous alleles, and confirmed null alleles are entered to the pedigree section, as shown in Fig 1A. The name of the cultivar must be a string with text or numerical values without spaces.

Next to the header, the pedigree section is entered, as shown in Fig 1B. The first column denotes the name of the variety or landrace, and second and third columns are reserved for maternity and paternity information, respectively. The numbers and strings can be included to represent the phenotypes in the first three columns. From the fourth column onwards, any desirable quantitative or qualitative trait values can be entered. All the collected phenotypic data are introduced, as shown in Fig 1B. The third section of the input data file is for marker information. The linkage group of the DNA marker and the marker positions in the linkage map are entered, as shown in Fig 1C. If there are more than one linkage group, all the linkage group maps should be defined successively before entering the allelic scores. The detailed data for each DNA marker can be inserted after revealing the map positions. The respective number of columns, according to the ploidy level, should be incorporated to enter allelic scores. The fourth section is for IBD probability values (Fig 1C). The IBD probabilities cannot be calculated within Pedimap but can be calculated using other software *e.g.* FlexQTL [34], which is a software for QTL analysis. FlexQTL can also generate a complete Pedimap input data file.

Demonstration of the Usability of Pedimap

We used the examples 1 and 2 given in Table 1 to show how parental cultivars can be selected for crossing based on diverse breeding objectives and the prioritized traits. The example 3 in Table 1 was used to select parents, indicate the DNA marker allelic representation for MAB, identity by descent calculations, and planning crosses to deduce related details necessary for decision-making for breeding.

Table 1. The examples used to demonstrate the use of Pedimap in making breeding decisions.

Example	Trait*					
	Priority[#] 1	**Priority 2**	**Priority 3**	**Priority 4**	**Priority 5**	**Priority 6**
1	White pericarp	Yield ≥3.5 mt/ha	Resistance or moderate resistance to brown planthopper (BPH)	Maturity period ≤125 days	Grain shape[$]	-
2	High, high-intermediate, and intermediate amylose content	Yield ≥3.5 mt/ha	Maturity period ≤125 days	Resistance or moderate resistance to blast	-	-
3	Phosphorous deficiency tolerance	Yield ≥5.0 mt/ha	Maturity period 90-105 days	Resistance or moderate resistance to BPH	Resistance or moderate resistance to blast	High, high-intermediate, and intermediate amylose content

*The trait classes and records are from RRDI records [35].

[#]The traits are given in the order of priority in making breeding decisions.

[$]Variable grain shapes of the intended varieties to be released.

Fig 1 The input data file structure of the Pedimap; The input file was created as an MS Excel worksheet, contains four main sections. A: Header, B: Pedigree and phenotypic data, C: Genotypic data. A: In the header section, essential elements are highlighted in blue, which contains the population name, ploidy and codes used in the data. (i): abbreviations for missing data (i.e., unknown), possible null alleles, confirmed null alleles; (ii): NALLELES is only necessary if the IBD probabilities are used, and specifies the total number of founder alleles (i.e. the number of founder times the ploidy). B: The Pedigree section contains the pedigree data of all the individuals, and any phenotypic data of the individuals. The pedigree part is highlighted in purple. (iii): founders (initial parents) are entered with missing values for their parents. Phenotypic data are entered in subsequent columns (iv). C: The Genotypic data section (if present) is divided into three parts: one part for each linkage group the genetic map (v), general information per locus (vi) and positions where IBD probabilities are calculated (vii); a part with the observed alleles per locus per individual (viii), and a part with the Identity-by Descent (IBD) probabilities per position per individual (ix). The final file must be saved as a text (.txt) file.

RESULTS AND DISCUSSION

Worldwide plant genetics and breeding programs use Pedimap as the platform for maintaining breeding databases and pedigree visualization. In the RosBREED project [36], the parental and progeny identification, tracing founders, and calculation of allelic representation are conducted using Pedimap. The pedigree display of Pedimap is used to plan crosses in the Rosaceae research community [37, 38], HIDRAS project [39] and visualize of *Arabidopsis thaliana* crosses [40]. Selecting parentage, sketching out crossing schemes, estimating the probability of allelic segregation, and choosing compatible molecular markers for MAB can be achieved using Pedimap [28]. The use of Pedimap as a pedigree visualization tool for the decision-making process in rice breeding is described using three examples (Table 1).

Example 1: Selecting parents for higher yield, BPH tolerance, short duration and white pericarp with diverse grain shapes

The Pedimap database rice breeding gerplasm in has a total of 224 input cultivars. There are 36 intermediate genotypes such as F1 and F2 that were not reported, but we included them to complete the pedigree in Pedimap. Thus, the database has a total of 188 rice cultivars and accessions with known identities with records (S1 Fig, S1 Table). In Example 1, we considered a scheme to select accessions as parents with the parameters given in Table 1 for white pericarp, yield, BPH resistance, maturity period, and the grain shape. These thresholds defined a subpopulation of 26 cultivars (Fig 2). The variation of the yield is given in Fig 2A. According to the color shading given, the breeder can select the required parents for crossing to obtain higher yield levels. However, as shown in Fig 2B, only three cultivars show the complete resistance to BPH. If breeder plans to introgress the complete BPH resistance to the novel varieties, only Bg250, At307, and At306 are available as the sources of resistance. Fig 2C displays the variation for the maturity period. The breeder can choose the parents depending

on his objective for the intended maturity period for the novel varieties. Example 1 was exclusively planned to breed for white pericarp. However, the grain shape is also important as a significant quality trait to become a successful variety in the market. Fig 2D shows the variation for grain shapes for the breeder to carry out the selection. If we consider all the traits and selected At307 as a parent based on the pedigree visualization in Pedimap, At307 can provide the genetic basis for high yield, complete resistance to BPH, approximately three months for maturity and intermediate-bold shaped grains. If Bg450 was selected, the yield is still in the higher range with moderate resistance for BPH and short-round grains. However, Bg450 brings the alleles for an extended maturity period (Fig 2).

Fig 2 The pedigree visualization for Example **1 (**Parents with white pericarp, yield ≥ **3.5** mt/ha, moderate or complete BPH resistance, maturity period ≤ **125** days, and diverse grain shapes). The selected pedigree is colored separately for four traits. A: Yield; B: Degree of resistance to brown planthopper (BPH); C: Maturity period; D: Grain shape. Female and male parentages are indicated by red and purple lines, respectively. The symbol '×' indicates the cross between two parents. The background colors of the cultivar-name boxes indicate the trait values, as shown in the colored legends below.

Example 2: Selecting parents for high/high-intermediate amylose content, higher yield, short duration, and resistance to blast disease

In Example 2, we considered a scheme to select cultivars/accessions as parents with the parameters given in Table 1 for high/high-intermediate amylose content, higher yield, short duration, and resistance to blast disease. These thresholds defined a subpopulation of 37 cultivars/accessions (Fig 3). The breeder can select the high yielding, short-duration, and blast-resistant cultivars as parents from pedigrees visualized in Figs 3A, 3B, and 3C, respectively. The high, high-intermediate, and intermediate amylose contents are depicted in the pedigree given in Fig 3D. Only Bw351, At307, Bg407H, At308, and Bg252 show the complete resistance to blast (Fig 3C). However, At307 is the most promising parent with high yield (Fig 3A), short duration (Fig 3B), and high amylose content (Fig 3D) along with complete resistance

to blast (Fig 3C). Also, Bg407H is the highest yielding (Fig 3A), blast-resistant (Fig 3C), and high in amylose content (Fig 3D). However, Bg407H is a long duration variety compared to At307. Therefore, the breeder may plan to cross At307 and Bg407H to accomplish the breeding objective of Example 2.

Fig 3 The pedigree visualization for Example 2 (parents with high, high-intermediate, and intermediate amylose content, yield ≥ 3.5 mt/ha, moderate or complete resistance to rice blast disease and maturity period ≤ 125 days). The selected pedigree is colored separately for four traits. A: Yield; B: Maturity period; C: Degree of resistance to rice blast disease; D: Amylose content. Female and male parentages are indicated by red and purple lines, respectively. The symbol '×' indicates the cross between two parents, and '×' inside the circle represents selfing. The background colors of the cultivar-name boxes indicate the trait values, as shown in the colored legends below.

Example 3: Selecting parents for phosphorus deficiency tolerance, higher yield, short duration, resistance to both BPH and blast, and high/intermediate-high amylose content

We selected a set of rice cultivars from the Pedimap database based on the availability of ranked scores for phosphorus deficiency tolerance (PDT). Twenty-four cultivars contain the PDT ranks of high, moderate, and sensitive (Fig 4A). The same set was illustrated using Pedimap for yield (Fig 4B), maturity period (Fig 4C), degree of resistance to BPH (Fig 4D) and blast (Fig 4E), and amylose content (Fig 4F). If At362 is considered as a parent, it can bring resistance to PD, and BPH, moderate resistance to blast, high yield, average maturity period, and intermediate-high amylose content. Similarly, if Bg250 is selected, it can bring moderate resistance to PD and blast, resistance to BPH, moderate yield and shortest maturity period, and high amylose content (Fig 4).

A sample crossing scheme is shown in Fig 5 to produce a rice variety with high PDT, mean yield ≥5.0 mt/ha, maturity period ≤105 days, resistant to BPH and blast disease, and higher amylose content. Since there is no reported cultivar for high PDT with complete blast resistance

(Fig 4), the illustrated crossing scheme in Fig 5 is proposed with two phases. In the first phase, the crossing of At362 and Bg250 followed by numerous rounds of selfing and selection of the most beneficial lines among the RILs at advanced generations would accomplish the breeding objective only without complete resistance to blast (i.e., a moderate level of blast resistance is possible). In the second phase, the selected RILs from phase 1 can be backcrossed to Bg252 as the donor parent to introgress the complete resistance to blast. The breeder can come up with diverse crossing schemes like the one given in Fig 5 to make effective decisions for breeding and maximize the resource utilization to release varieties in the shortest possible time. The breeder can select any number of parents that are needed to use as sources of resistance and other traits to start crossing. Also, the marker alleles and the IBD probabilities can be checked as illustrated in S2A Fig and S2B Fig respectively.

Fig 4 The pedigree visualization for Example 3 (parents ranked for phosphorus deficiency tolerance). The selected pedigree is colored separately for six traits. A: PDT; B: Yield; C: Maturity period; D: Degree of resistance to BPH; E: Degree of resistance to BLAST; F: Amylose content. Female and male parentages are indicated by red and purple lines, respectively. The symbol '×' indicates the cross between two parents. The background colors of the cultivar-name boxes indicate the trait values, as shown in the colored legends below. The cultivars with missing-trait values are indicated by white boxes.

Fig 5 The pedigree visualization for planning a crossing scheme. Phase 1: Initial crossing of At362 and Bg250 and pedigree selection to obtain RILs with ≥5.0 mt/ha of mean yield, ≤105 day of maturity period, resistant to BPH, moderately resistant to blast and high level of amylose content. Phase 2: Then backcrossing with Bg252 as the donor parent to introgress the blast resistance.

The decision-making process in breeding is a tedious task [22]. The breeding germplasm is complex with large numbers of improved varieties, traditional cultivars, landraces, wild germplasm and accessions. Also, there can be large mapping populations and unreleased varieties due to various reasons. The numerous genotypes in breeding germplasm may have extensive records on agronomic data, pest and disease resistance, quality traits, availability of samples, geographic locations, and utilization in diverse breeding programs as parents [41, 42].

With the advent of DNA markers and sequencing technologies, a wealth of genomic information is also available [43]. However, one of the recurrent problems in any breeding germplasm in the world is most of the cultivars remain uncharacterized. Thus, they cannot be used directly in breeding activities. Traditionally, breeders keep records in field books. With the development of computer technology, data tabulation is becoming a common practice. However, given the highly complex nature of the datasets in breeding germplasm, data tables have a limited value to the breeders. The tables created with contemporary data managing software cannot graphically display complex pedigrees and variations of qualitative and quantitative traits along with DNA marker information. These database handling platforms do not make use of the pedigree-based capabilities of Pedimap, like selecting related parental varieties/accessions. In this context, Pedimap provides a considerable advantage, as it can visualize pedigree relationships, trait variations, and any other useful information required for decision-making and planning crosses in breeding programs [28]. If all the available details on breeding germplasm are arranged as a database, the breeder can come up with subpopulations based on diverse traits and select the parents for improving multiple traits. However, simple spreadsheets or manually prepared note pages cannot be used to visualize the essential information and complex pedigrees. Breeding programs often suffer a lot when the breeder gets retired or moved to a different position [44-46]. The newly hired breeder cannot practically go through the individual records of the existing breeding germplasm. Thus, there is a strong possibility that valuable breeding germplasm might get lost wasting time, resources, and courage of the retired breeder and his team. However, as a routine practice, if the breeder maintains and updates a Pedimap file for the developing germplasm of breeding materials, the newly hired breeders can go through and identify the value and gaps in the available material for him to plan further. The creation of a Pedimap file is simple, and a novice to informatics can curate and use Pedimap with a little training. Pedimap allows breeders to store data, fetch

and visualize genomic information at any time with less effort and complete accuracy [47]. The straightforward accessibility, direct data interpretation, ability to customize the views in multiple fashions, and editable output file formats are the significant features of Pedimap. The graphic files created can be readily imported to image editing software for further visualizations and illustrations. Pedimap is not an opensource software but can be freely obtained by contacting the developers thus even the breeders in developing countries can benefit from Pedimap [28].

In the current study, we created a Pedimap database for the rice cultivars and accessions prominently used by breeding programs in Sri Lanka. With the available information, significant breeding decisions can be made as we explain in three examples (Figs 2-5). However, it is essential to characterize the cultivars for all the important traits, molecular markers and SNP haplotypes [48], so that breeding decisions can be effectively made [17]. The phenotyping methods must be standard and should follow common procedures across different locations so that the power of the Pedimap database would go up dramatically. Therefore, breeders should always follow the standard, globally acceptable phenomic platforms to characterize the material in breeding germplasm [39, 49].

CONCLUSION

The pedigree visualization with variations of phenotypic and molecular data using Pedimap is a user-friendly tool to plan rice breeding programs with higher accuracy and resource optimization. The present study explains the applicability of Pedimap as a decision-making tool to streamline the rice breeding programs in Sri Lanka. However, it is also important to note that accurate characterization of the breeding germplasm for phenotypic and molecular data is the critical prior step to harness the value of Pedimap for breeding.

354 **REFERENCES**

355 1. Food and Agriculture Organization of the United Nations (FAO). Rice market monitor.
356 2018; 21: 1-3.
357 2. Muthayya S, Sugimoto JD, Montgomery S, Maberly GF. An overview of global rice
358 production, supply, trade, and consumption. Ann. N. Y. Acad. Sci. 2014; 1324(1):7-14.
359 3. Ali AM. Rice to shrimp: Land use/land cover changes and soil degradation in
360 Southwestern Bangladesh. Land Use Policy. 2006; 23(4):421-35.
361 4. Seck PA, Diagne A, Mohanty S, Wopereis MC. Crops that feed the world 7: Rice. Food
362 Secur. 2012; 4(1):7-24.
363 5. Segal R, Le Nguyet M. Unfair Harvest: The state of rice in Asia. Oxford, UK: Oxfam
364 International; 2019.
365 6. Rice Research and Development Institute (RRDI). Rice Cultivation. Sri Lanka:
366 Department of Agriculture; 2019 [cited 2019 Jul 15]. Available from: http://
367 www.doa.gov.lk/rrdi/.
368 7. Central Bank of Sri Lanka. Annual Report of Central Bank of Sri Lanka. 2017 [Cited
369 2019 May 18]. Available from: https://www.cbsl.gov.lk/en/publications/economic-and-
370 financial-reports/annual-reports/annual-report-2017.
371 8. Central Bank of Sri Lanka. Annual Report of Central Bank of Sri Lanka. 2018 [Cited
372 2019 May 18]. Available from: https://www.cbsl.gov.lk/en/publications/economic-and-
373 financial-reports/annual-reports/annual-report-2018.
374 9. Dharmarathna WR, Herath S, Weerakoon SB. Changing the planting date as a climate
375 change adaptation strategy for rice production in Kurunegala district, Sri Lanka.
376 Sustainability science. 2014; 9(1):103-11.
377 10. IPCC. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of
378 Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on

379 Climate Change. In: Parry ML, Canziani OF, Palutikof JP, Van der Linden PJ, Hanson
380 CE, editors. Cambridge, UK: Cambridge University Press; 2007. 976p.
381 11. Mackill DJ, Coffman WR, Garrity DP. Rainfed Lowland Rice Improvement. Los Banos,
382 Philippines: International Rice Research Institute; 1996. 242p.
383 12. Walisinghe BR, Gunaratne LHP, Prasada DVP, Sarthaj IZ. Analysis of the consumer
384 preferences for rice-based noodles and biscuits. Annals of Sri Lanka Department of
385 Agriculture. 2010; 12:201-211.
386 13. Dhanapala MP. Bridging the rice yield gap in Sri Lanka. In Proceedings of the symposium
387 on Expert Consultation on Bridging the Rice Yield Gap in the Asia-Pacific Region,
388 Bangkok, Thailand; 1999 October 5–8. pp. 135–145.
389 14. Shetty PK. Socio-ecological implications of pesticide use in India. Econ. Polit. Wkly.
390 2004; 5261-5257.
391 15. Duvick DN. Progress in conventional plant breeding. In Gene manipulation in plant
392 improvement. Boston, MA: Springer; 1984. Pp. 17-31.
393 16. Calingacion M, Laborte A, Nelson A, Resurreccion A, Concepcion JC, Daygon VD, et
394 al. Diversity of global rice markets and the science required for consumer-targeted rice
395 breeding. Plos One. 2014; 9(1):e85106.
396 17. Jiang GL. Plant marker-assisted breeding and conventional breeding: challenges and
397 perspectives. Adv. Crop Sci. Technol. 2013; 1:e106.
398 18. Sasaki T. Power of conventional rice breeding. Breed. Sci. 2011; 61(4):317.
399 19. Collard BC, Mackill DJ. Marker-assisted selection: an approach for precision plant
400 breeding in the twenty-first century. Philosophical Transactions of the Royal Society B:
401 Biological Sciences. 2007; 363(1491):557-572.

20. Tanksley SD, Nelson JC. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 1996; 92(2):191-203.
21. Thomson MJ, Ismail AM, McCouch SR, Mackill DJ. Marker assisted breeding. In Abiotic Stress Adaptation in Plants. Dordrecht: Springer; 2009. pp. 451-469.
22. Acquaah G. Principles of plant genetics and breeding. 2nd ed. Oxford, UK: Wiley-Blackwell; 2012. pp. 281-334.
23. Clark BH. Marketing performance measures: History and interrelationships. Journal of marketing management. 1999; 15(8):711-32.
24. Ragot M, Bonierbale M, Weltzien E. From Market Demand to Breeding Decisions: A Framework. Lima (Peru). CGIAR Gender and Breeding Initiative. GBI Working Paper. No. 2 [Internet]. 2008 [cited 2019 Jul 25]. Available from: https://www.rtb.cgiar.org/gender-breeding-initiative.
25. Kariuki CM, van Arendonk JA, Kahi AK, Komen H. Multiple criteria decision-making process to derive consensus desired genetic gains for a dairy cattle breeding objective for diverse production systems. Journal of dairy science. 2017; 100(6):4671-82.
26. Velasquez M, Hester PT. An analysis of multi-criteria decision-making methods. International Journal of Operations Research. 2013; 10(2):56-66.
27. Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, et al. CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Res. 2013; 42(D1):D1229-36.
28. Voorrips RE, Bink MC, van de Weg WE. Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. J. Hered. 2012; 103(6):903-907.

29. Aluwihare YC, Ishan M, Chamikara MD, Weebadde CK, Sirisena DN, Samarasinghe WL, et al. Characterization and selection of phosphorus deficiency tolerant rice genotypes in Sri Lanka. Rice Science. 2016; 23(4):184-95.

30. Aluwihare YC, Chamikara MD, Dissanayake DR, Karannagoda NN, Sirisena DN, Samarasinghe WL, et al. Validation of *K46*, a *Pup1*-linked marker, using a selection of Sri Lankan rice (*Oryza sativa* L.) germplasm for marker assisted selection towards phosphorous deficiency tolerance. Ceylon J. Sci. (Bio. Sci.). 2015; 44(2):45-54.

31. Jayarathne HSM, Rathnayake PGRG, Ranaweera LT, Nizam, R., Rathnayake RMSK, Kannangara SK, et al. Assessment of the variation of performance indicators under p-starved conditions in a core-set of rice cultivars and their diversity in *Pup1*-linked DNA marker-haplotypes. J. Natl. Sci. Found. Sri. 2019; In review.

32. Nakandala ND, Karunarathne SI, Ranaweera LT, Wijesundara WW, Senevirathna RW, Rathnayake RM, et al. Assessment of the association between yield parameters and polymorphic sites of the *Ghd7* locus in a core-set of rice cultivars grown in Sri Lanka. Ceylon J. Sci. 2019; 48(2):121-32.

33. Rathnayake RMSK, Rathnayake PGRG, Jayarathne HSM, Nizam R, Ranaweera LT, Kannangara SK, et al. Assessment of the Sequence-Based Haplotype-Variants in Selected DNA Marker Loci for the Molecular Breeding of Resistant Rice Varieties to Brown Plant Hopper. Sri Lankan Journal of Agriculture and Ecosystems. 2019; 1:2-15.

34. Bink MC, Boer MP, Ter Braak CJ, Jansen J, Voorrips RE, Van de Weg WE. Bayesian analysis of complex traits in pedigreed plant populations. Euphytica. 2008; 161 (1-2):85-96.

35. RRDI (2018). Recommended Rice Varieties in Sri Lanka (1958-2016). Bathalagoda, Ibbagamuwa, Sri Lanka: Rice Research and Development Institute, Department of Agriculture; 2018.

450 36. Weebadde C, Crisosto C, Fazio G, Gallardo K, Main D, Peace C, et al. Use of PediMap
451 in RosBREED for Visualizing Genetic and Phenotypic Data across Pedigrees in
452 Rosaceous Crops. Hort. Science. 2010; 45, S55-S55.
453 37. Peace CP, Luby JJ, Van de Weg WE, Bink MC, Iezzoni AF. A strategy for developing
454 representative germplasm sets for systematic QTL validation, demonstrated for apple,
455 peach, and sweet cherry. Tree Genet. Genomes. 2014; 10(6):1679-1694.
456 38. Rosyara UR, Bink MC, van de Weg E, Zhang G, Wang D, Sebolt A, et al. Fruit size QTL
457 identification and the prediction of parental QTL genotypes and breeding values in
458 multiple pedigreed populations of sweet cherry. Mol. Breeding. 2013; 32(4):875-87.
459 39. Evans K, Guan Y, Luby J, Clark M, Schmitz C, Brown S, et al. Large-scale standardized
460 phenotyping of apple in RosBREED. Acta Hortic. 2012; 945:233-238.
461 40. Paulo MJ, Boer M, Huang X, Koornneef M, Van Eeuwijk F. A mixed model QTL analysis
462 for a complex cross population consisting of a half diallel of two-way hybrids in
463 *Arabidopsis thaliana*: analysis of simulated data. Euphytica. 2008; 161(1-2):107-14.
464 41. Davenport G, Ellis N, Ambrose M, Dicks J. Using bioinformatics to analyse germplasm
465 collections. Euphytica. 2004; 137(1):39-54.
466 42. Peeters JP, Galwey NW. Germplasm collections and breeding needs in Europe. Econ.
467 Bot. 1988; 42(4):503-521.
468 43. Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, et al. DNA
469 molecular markers in plant breeding: current status and recent advancements in genomic
470 selection and genome editing. Biotechnol. Equip. 2018; 32(2):261-85.
471 44. Sooriyapathirana SDSS, Ranaweera LT, Perera HACR, Weebadde CK, Finn CE, Bassil
472 NV, el al. Using SNP/INDEL diversity patterns to identify a core group of genotypes
473 from FVC_{11}, a superior hybrid family of *Fragaria virginiana* Miller and *F. chiloensis* (L.)
474 Miller. Genet. Resour. Crop Ev. 2019; 66(8);1691-1698.

475 45. USDA. USDA Roadmap for Plant Breeding March 11, 2015: Final Report. United States
476 Department of Agriculture: Office of the Chief Scientist Research, Education, and
477 Economics Mission Area [Internet]. 2015 [cited 2019 Jul 25]. Available from:
478 http://www.usda.gov/wps/portal/usda/usdahome?navid=OCS.
479 46. Hmielowski T. Challenges to developing cultivars in the public sector. Crops, Soils,
480 Agronomy News. 2017; 62(10):4-5.
481 47. Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, et al. CottonFGD: an integrated
482 functional genomics database for cotton. BMC Plant Biol. 2017; 17(1):101.
483 48. Vanderzande S, Howard NP, Cai L, Linge CD, Antanaviciute L, Bink MC, et al. High-
484 quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid
485 outbreeding species apple, peach, and sweet cherry through a common workflow. Plos
486 One. 2019; 14, e0210928.
487 49. Fukai S, Fischer KS. Field phenotyping strategies and breeding for adaptation of rice to
488 drought. Frontiers in Physiology. 2012; 3:282.
489

490 Supporting information

491 **S1 Table** Varietal data

492 **S2 Table** Marker data

493 **S3 Table** Pedimap input data file

494 **S1 Fig** Visualization of the entire pedigree of the rice cultivars in the rice breeding germplasm
495 of Sri Lanka. Female and male parentages are indicated by red and purple lines, respectively.
496 The symbol '×' indicates the cross between two parents, and '×' inside the circle represents
497 selfing.

498 **S2 Fig** Visualization of selected marker genotypes and Identity by Descent (IBD). A: Marker
499 alleles. The alleles of the DNA markers K29-N, K41, K48, and K5-N are given in vertical

order.; B: IBD probabilities of four Pup1 linked markers. The same pedigree as in the S2A Fig is shown with precalculated IBD data for part of the Pup1 linked region on chromosome 12 at about 55 cM. Since the cultivar linkage maps are not available, we assumed a 0.1 cM gap between adjacent markers for the representation of IBD values. Eachcolor represents a different founder, haplotype. Each rectangle represents one copy of the selected chromosome in an individual. The chromosomal position of the alleles represents the vertical bars, and the width of a color bar indicates the IBD probability of the corresponding founder alleles.

A

POPULATION	=	Sri_Lanka_Rice_Germplasm	
UNKNOWN	=	-	
NULLHOMOZ	=	$	i
CONFIRMEDNULL	=	$$	
PLOIDY	=	2	
NALLELES	=	6	ii

B

PEDIGREE

NAME	PARENT1	PARENT2	Yiled	Maturity	Height	Leaf_color	BPH	GM	BL	PDT
Bg94_1	-	-	4.1	105	55	Green	S	S	MS	High
Pokkali	-	-	-	-	-	-	-	-	-	Moderate
Bg300	-	-	5	93	72.2	Green	MR	R	MR	Sensitive
Bg301	-	-	6	93	67	Green	-	-	-	-
At354	Bg94_1	Pokkali	6'5	95	67	Green	MR	MR	S	Sensitive
At401	Bg94_1	Pokkali	5	115	85	Green	MR	R	MS	-

iii

iv

C

LINKAGEGROUP 12

MAP

RM101	48.2										
RM277	62										
LOCUS	RM101										
ALLELENAMES	110	115	120	125							
FOUNDERALLELES	110	110	110	110	115	120					
LOCUS	RM277										
ALLELENAMES	200	225	250	275	300						
FOUNDERALLELES	200	200	200	200	300	275					
IBDPOSITIONS	48.2	62									
ALLELES	RM323										
Bg941	140	0	140	0							
Pokkali	160	0	180	0							
IBDPOSITION	48.6										
Bg941	1	0	0	0		0		0	0	1	
Pokkali	0	0	1	0		0		0	0	0	
At354	0	0	0	0		0		0	1	0	
At401	0	0	0	0		0		0	0	0	
IBDPOSITION	62.0										

v

vi

vii

viii

ix

Define Linkage group

Figure 1

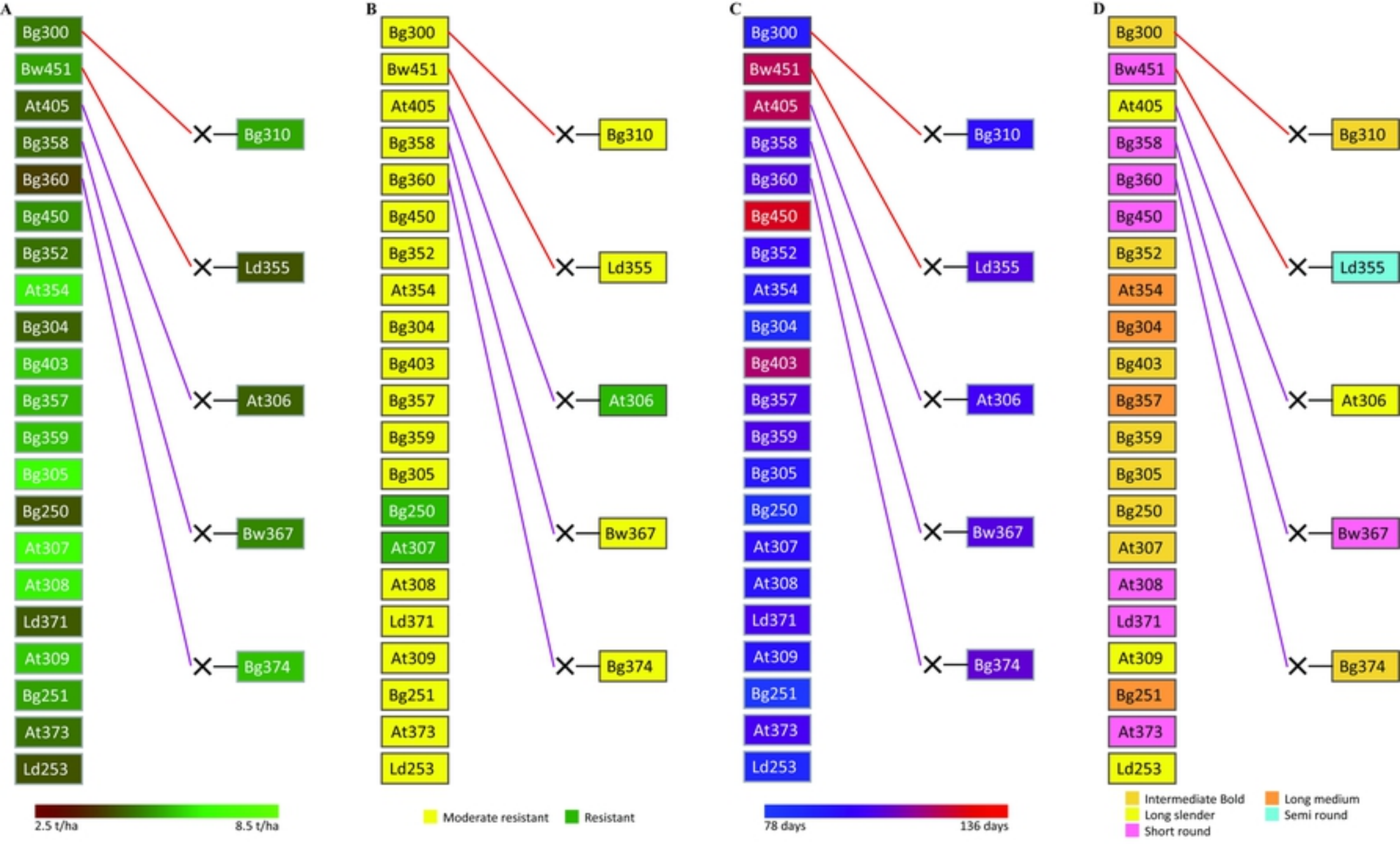
A
Bg300
Bw451
At405
Bg358
Bg360
Bg450
Bg352
At354
Bg304
Bg403
Bg357
Bg359
Bg305
Bg250
At307
At308
Ld371
At309
Bg251
At373
Ld253
Bg310
Ld355
At306
Bw367
Bg374
2.5 t/ha
8.5 t/ha
B
Bg300
Bw451
At405
Bg358
Bg360
Bg450
Bg352
At354
Bg304
Bg403
Bg357
Bg359
Bg305
Bg250
At307
At308
Ld371
At309
Bg251
At373
Ld253
Bg310
Ld355
At306
Bw367
Bg374
Moderate resistant
Resistant
C
Bg300
Bw451
At405
Bg358
Bg360
Bg450
Bg352
At354
Bg304
Bg403
Bg357
Bg359
Bg305
Bg250
At307
At308
Ld371
At309
Bg251
At373
Ld253
Bg310
Ld355
At306
Bw367
Bg374
78 days
136 days
D
Bg300
Bw451
At405
Bg358
Bg360
Bg450
Bg352
At354
Bg304
Bg403
Bg357
Bg359
Bg305
Bg250
At307
At308
Ld371
At309
Bg251
At373
Ld253
Bg310
Ld355
At306
Bw367
Bg374
Intermediate Bold
Long slender
Short round
Long medium
Semi round

Figure 2

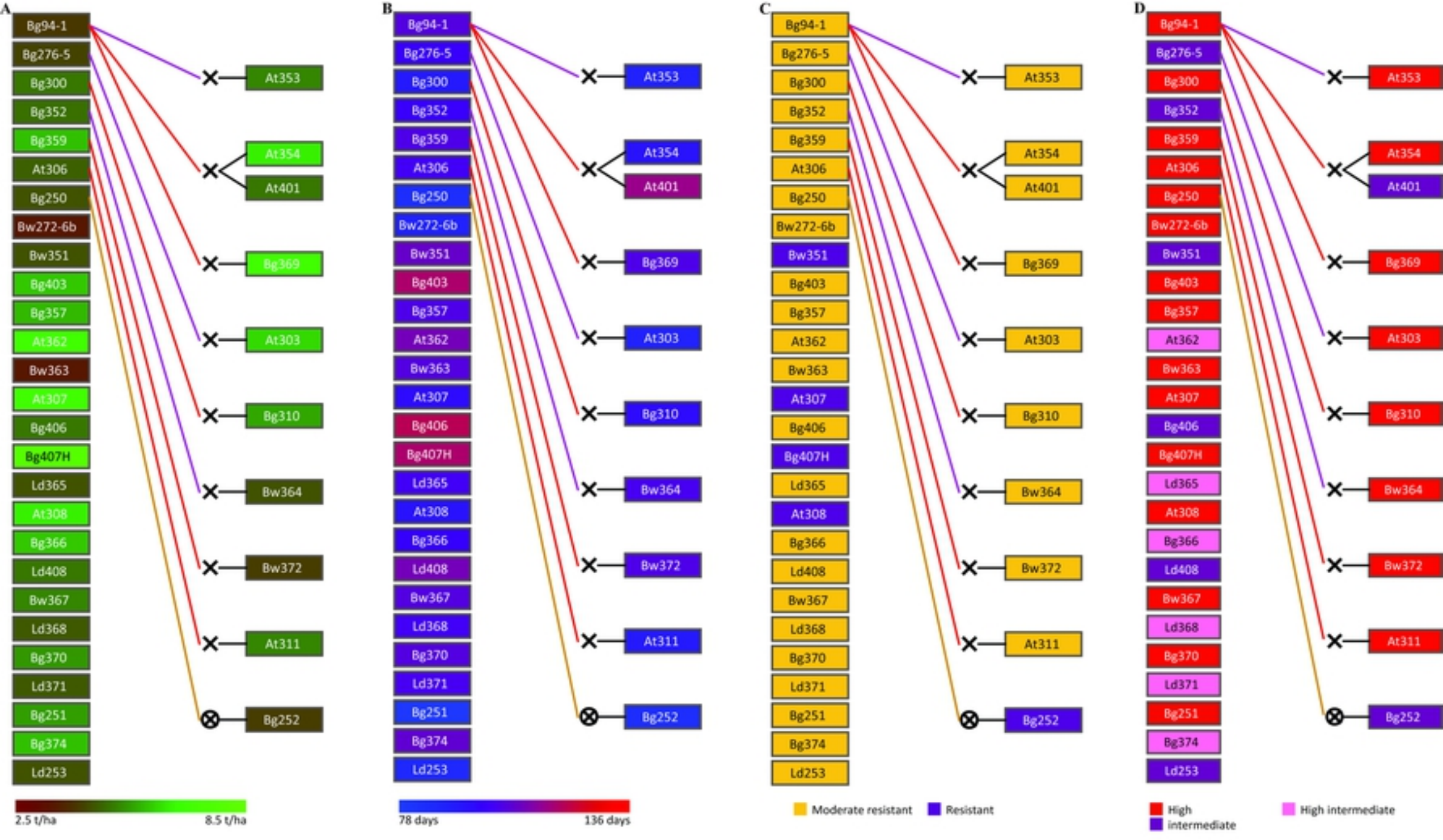
A
Bg94-1
Bg276-5
Bg300
Bg352
Bg359
At306
Bg250
Bw272-6b
Bw351
Bg403
Bg357
At362
Bw363
At307
Bg406
Bg407H
Ld365
At308
Bg366
Ld408
Bw367
Ld368
Bg370
Ld371
Bg251
Bg374
Ld253
At353
At354
At401
Bg369
At303
Bg310
Bw364
Bw372
At311
Bg252
2.5 t/ha
8.5 t/ha
B
78 days
136 days
C
Moderate resistant
Resistant
D
High
intermediate
High intermediate

Figure 3

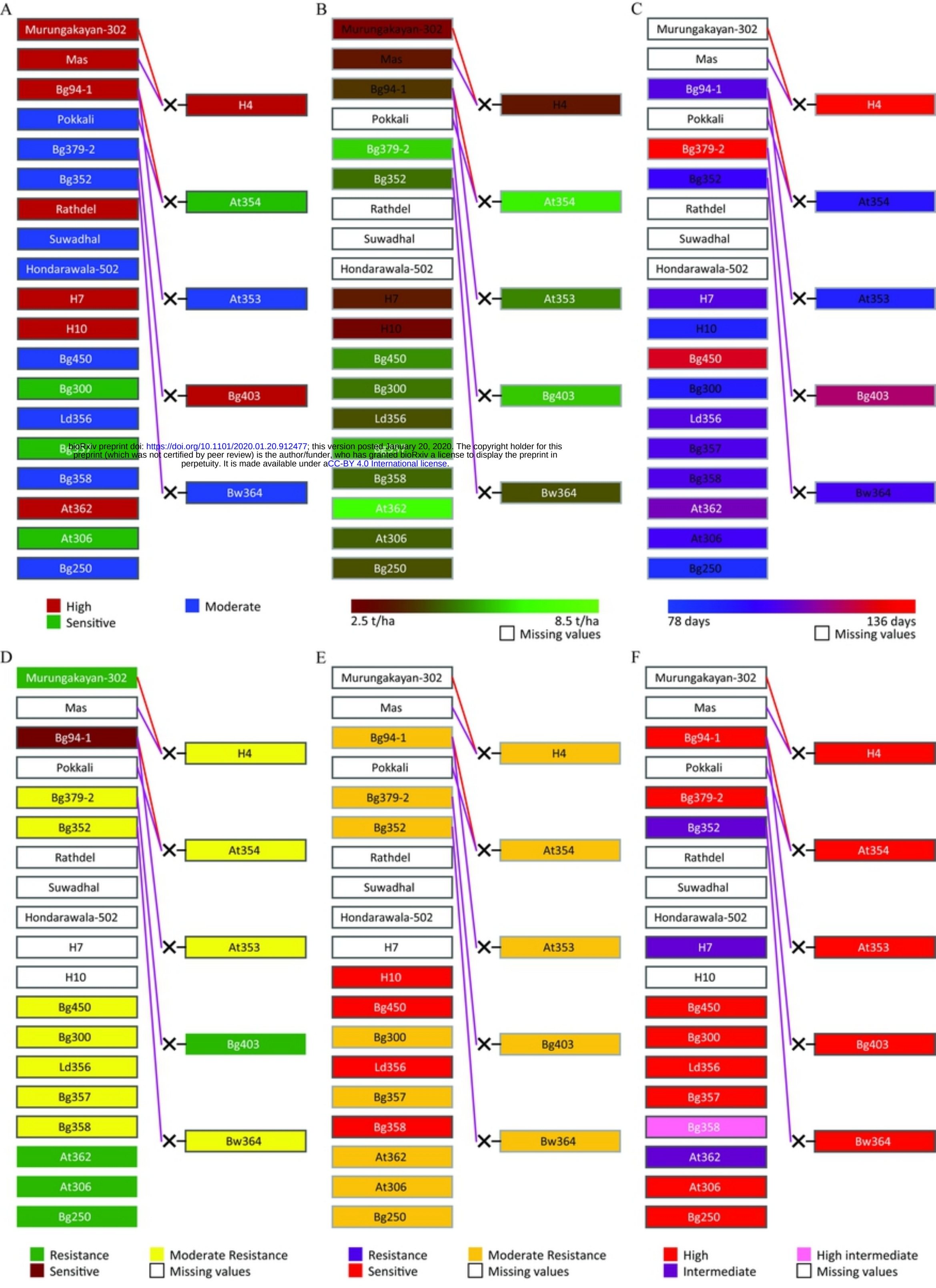

Figure 4

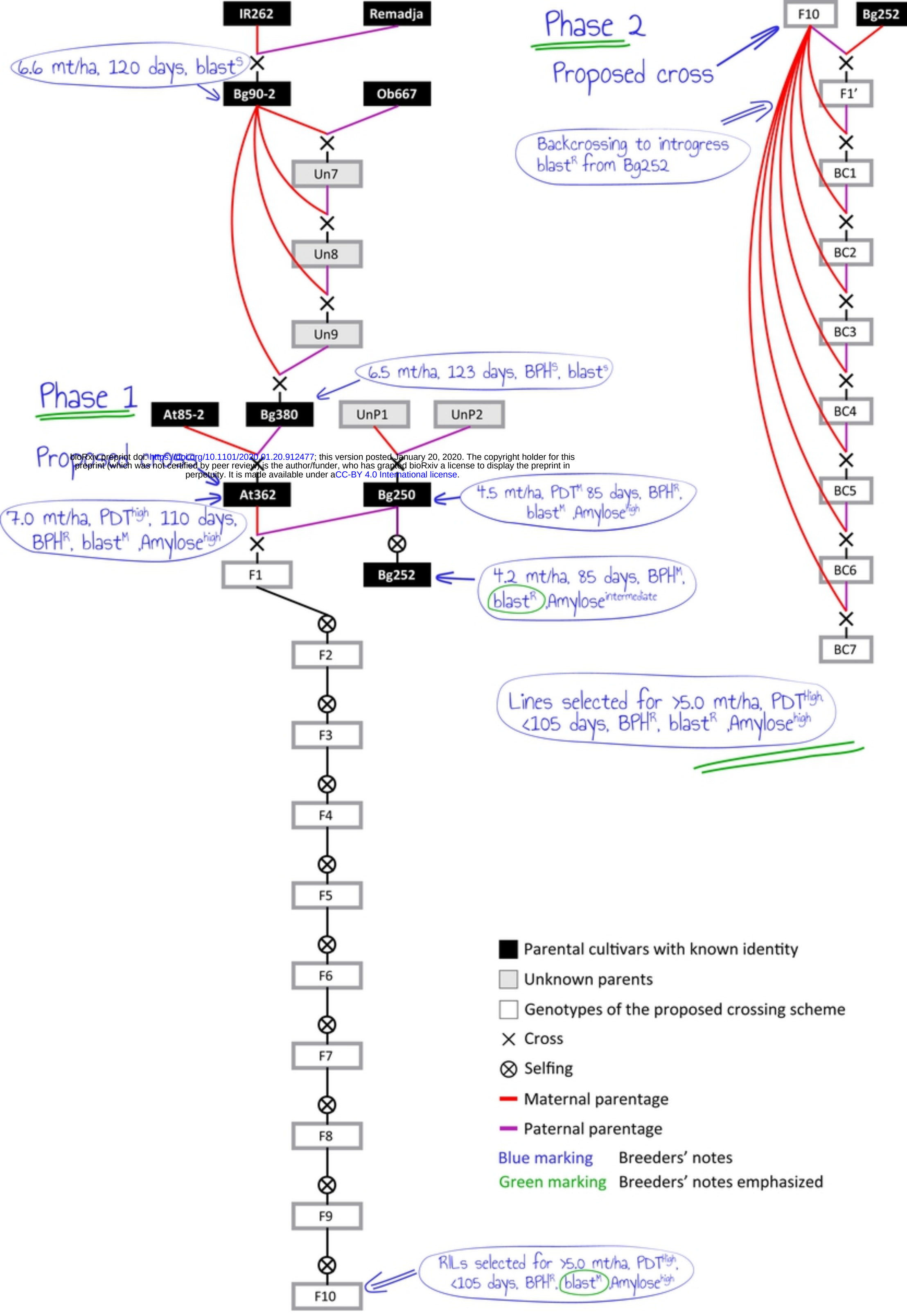

Figure 5