





OPEN


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

Rasika Rathnayake<sup>1</sup>, Shehan Sahibdeen<sup>1</sup>, Kapila Udawela<sup>2</sup>, Cholani Weebadde<sup>3</sup>, Wijayasiri Weerakoon<sup>4</sup> & Suneth Sooriyapathirana<sup>1,5</sup>  

The development of rice cultivars with desirable traits is essential. The decision-making is a crucial step in rice breeding programs. Breeders can make efficient and pragmatic decisions if an organized pedigree visualization platform is available for the accessions and cultivars in rice breeding germplasm. 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 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 such as the assessment of before organized into Pedimap. Thereby, the worldwide breeding programs can benefit from each other to produce improved varieties to meet global challenges.

Rice is one of the major crops in the world, with an annual production over 700 million metric tons<sup>1</sup>. Half of the world population consumes rice as the staple food<sup>2</sup>. Currently, the demand for rice is rapidly increasing due to the growth of the human population<sup>3</sup>. 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<sup>4</sup>. The rice farming is also a way of living for many people, especially in numerous Asian countries<sup>5</sup>. At present, 1.8 million Sri Lankan families engage in rice farming over 870,000 hectares<sup>6</sup>. 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<sup>7</sup>. Hence, the Sri Lankan government spends about USD 400 million to import rice annually<sup>7,8</sup>.

The rice production is mainly affected by drought and irregular rainfall patterns caused by climate change<sup>9–11</sup>, adverse soil conditions such as salinity<sup>6,12</sup>, and pest and disease attacks<sup>13</sup>. 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<sup>13</sup> due to the unpredictable nature of climate change, soil degradation, variations in pest dynamics, and development of pest resistance<sup>14</sup>. Therefore, breeding is considered as the most successful strategy to produce high yielding and stress-resilient rice varieties<sup>15</sup>. The improved rice genotypes can also contain the traits for higher consumer preference and organic farming<sup>16</sup>. 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.

<sup>1</sup>Department of Molecular Biology and Biotechnology, Faculty of Science, University of Peradeniya, Peradeniya, Sri Lanka. <sup>2</sup>Rice Research and Development Institute, Bathalagoda, Sri Lanka. <sup>3</sup>Department of Plant, Soil and Microbial Sciences, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI 48824, USA. <sup>4</sup>Department of Agriculture, Peradeniya, Sri Lanka. <sup>5</sup>Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka. email: sunethuop@gmail.com

However, the marker-assisted breeding (MAB) is employed in modern breeding programs to introgress valuable genetic loci from landraces and traditional varieties<sup>17,18</sup> and the desirable haplotypes of Quantitative Trait Loci (QTL) to the improved rice varieties<sup>19–21</sup>.

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 and estimated breeding values (EBV) 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<sup>22</sup>. In the decisioning process, initially, the market trends based on consumer and other stakeholder preferences must be recognized<sup>23</sup>. Subsequently, the novelty and the uniqueness of the breeding objective must be assessed before the execution of the breeding program<sup>22</sup>.

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<sup>24</sup>. 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<sup>25,26</sup>. The decision-making process in breeding is entirely based on the available information on phenotypes, genotypes, pedigree, EBVs of key traits, 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<sup>27</sup>. 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 Voorrips et al.<sup>28</sup>, 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, two main rice growing seasons, based on the two different seasons of monsoonal rains. *Yala* season is generally drier<sup>29</sup>), 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, the weight of 1,000 grains, shape of the grain, pericarp color, the weight of a kg, the 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 (Supplementary Table S1 online). The available DNA marker alleles, marker positions in the linkage map, and allelic scores were entered under molecular data<sup>30–33</sup> (Supplementary Table S2 online).

**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 (Supplementary Table S3 online). 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. 1b. 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

	A	B	C	D	E	F	G	H	I	J	K	L	M
<b>a</b>													
1													
2	POPULATION	=	Sri_Lanka_Rice_Germplasm										
3	UNKNOWN	=	-										
4	NULLHOMOZ	=	\$										
5	CONFIRMEDNULL	=	\$\$										
6	PLOIDY	=	2										
7	NALLELES	=	6										
8													
<b>b</b>													
9	PEDIGREE												
10	NAME	PARENT1	PARENT2	Yield	Maturity	Height	Leaf_color	BPH	GM	BL	PDT		
11	Bg94_1	-	-	4.1	105	55	Green	S	S	MS	High		
12	Pokkali	-	-	-	-	-	-	-	-	-	Moderate		
13	Bg300	-	-	5	93	72.2	Green	MR	R	MR	Sensitive		
14	Bg301	-	-	6	93	67	Green	-	-	-	-		
15	At354	Bg94_1	Pokkali	6.5	95	67	Green	MR	MR	S	Sensitive		
16	At401	Bg94_1	Pokkali	5	115	85	Green	MR	R	MS	-		
17													
<b>c</b>													
18	LINKAGEGROUP 12												
19	MAP												
20	RM101	48.2											
21	RM277	62											
22													
23	LOCUS	RM101											
24	ALLELENAMES	110	115	120	125								
25	FOUNDERALLELES	110	110	110	110	115	120						
26													
27	LOCUS	RM277											
28	ALLELENAMES	200	225	250	275	300							
29	FOUNDERALLELES	200	200	200	200	300	275						
30	IBDPOSITIONS	48.2	62										
31													
32	ALLELES	RM323											
33	Bg941	140	0	140	0								
34	Pokkali	160	0	180	0								
35													
36	IBDPOSITION	48.6											
37													
38	Bg941	1	0	0	0	0	0	0	0	0	1		
39	Pokkali	0	0	1	0	0	0	0	0	0	0		
40	At354	0	0	0	0	0	0	0	1	0	0		
41	At401	0	0	0	0	0	0	0	0	0	0		
42													
43	IBDPOSITION	62.0											

**Figure 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.

Example	Trait*					
	Priority <sup>†</sup> 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 <sup>§</sup>	–
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

**Table 1.** The examples used to demonstrate the use of Pedimap in making breeding decisions. \*The trait classes and records are from RRDl records<sup>35</sup>. <sup>†</sup>The traits are given in the order of priority in making breeding decision. <sup>§</sup>Variable grain shapes of the intended varieties to be released.

calculated within Pedimap but can be calculated using other software *e.g.* FlexQTL<sup>34</sup>, which is a software for QTL analysis (<https://www.wur.nl/en/show/FlexQTL.htm>; FlexQTL Version 0.1.0.42). 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.

**Estimation of breeding values (EBV).** The selection of parental cultivars for the examples 1, 2 and 3 (Table 1) to illustrate the procedure of breeding decisioning using Pedimap was verified by calculating EBVs<sup>36</sup> for yield, maturity period, and plant height. The EBVs were calculated for all rice cultivars according to the following formula. The true breeding values (TBV) and the accuracies of EBVs based on the correlation with TBVs were also calculated<sup>37</sup>. The representative heritability ( $H^2$ ) estimates for the traits (0.50 for yield, 0.85 for maturity period, and 0.85 for plant height) were obtained from breeding records available at RRDl and other published sources<sup>38–40</sup>.

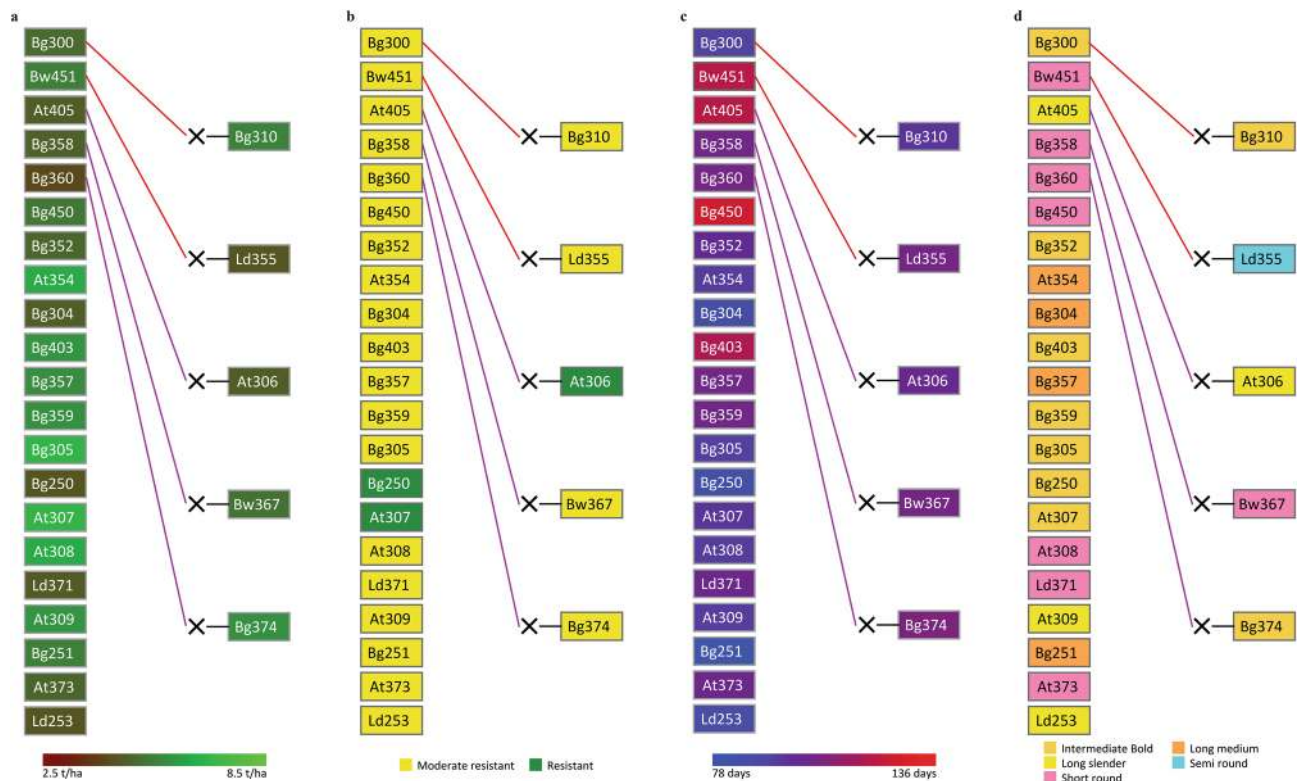
$$EBV = H^2(P - \bar{P}) \quad (\text{Ref}^{37})$$

where  $H^2$  = heritability of the trait;  $P$  = trait values of the individual or cultivar;  $\bar{P}$  = population mean value of the trait;  $P - \bar{P}$  = phenotypic superiority.

## Results

Worldwide plant genetics and breeding programs use Pedimap as the platform for maintaining breeding databases and pedigree visualization. In the RosBREED project<sup>41</sup>, 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<sup>42,43</sup>, HIDRAS project<sup>44</sup>, and visualize of *Arabidopsis thaliana* crosses<sup>45</sup>. Selecting parentage, sketching out crossing schemes, estimating the probability of allelic segregation, and choosing compatible molecular markers for MAB can be achieved using Pedimap<sup>28</sup>. 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 of rice breeding germplasm in Sri Lanka 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 (Supplementary Table S1 online and Supplementary Fig. 1 online). In Example 1, we considered a scheme to select cultivars 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. Figure 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. Figure 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



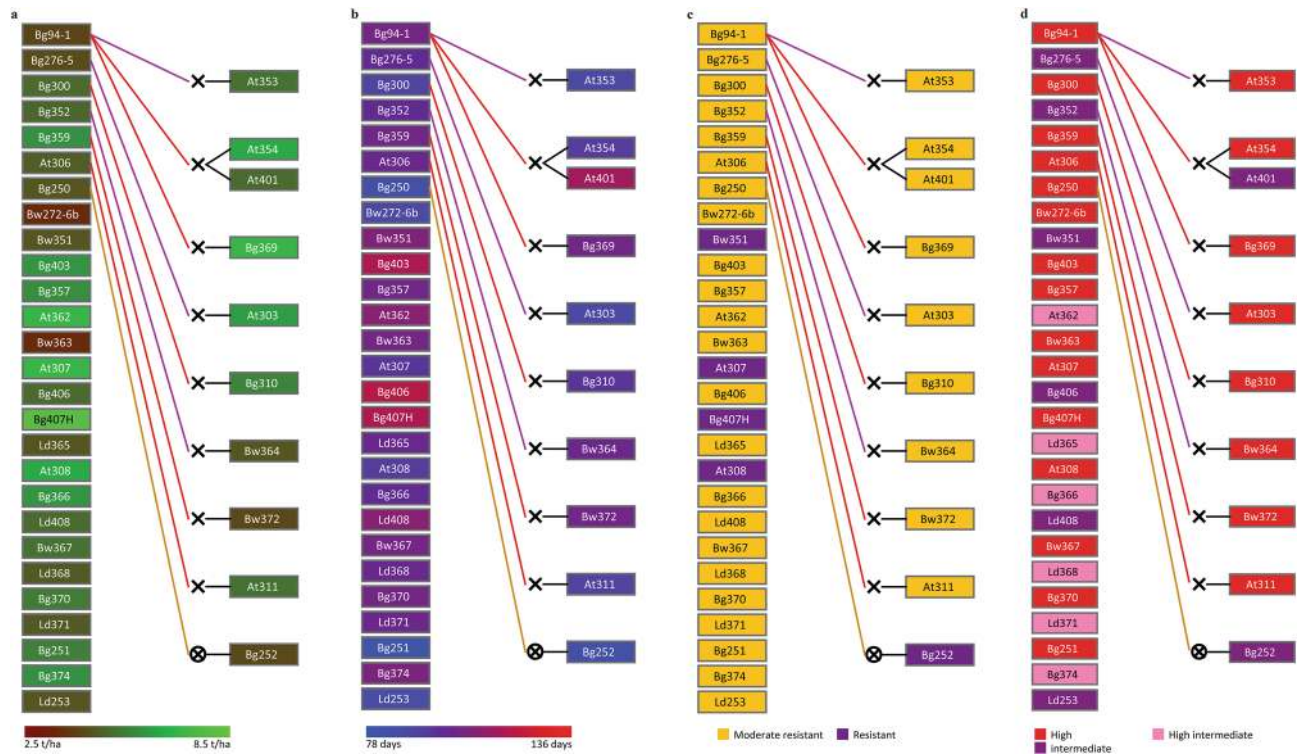
**Figure 2.** The pedigree visualization for Example 1 (Parents with white pericarp, yield  $\geq 3.5$  mt/ha, moderate or complete BPH resistance, maturity period  $\leq 125$  days, and diverse grain shapes). The selected pedigree is colored separately for four traits. (a): Yield; (b): Degree of resistance to brown planthopper (BPH); (c): Maturity period; (d): Grain shape. Female and male parentages are indicated by red and purple lines, respectively. The symbol 'x' 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.

resistance for BPH and short-round grains. However, Bg450 brings the alleles for an extended maturity period (Fig. 2).

**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 Fig. 3a–c, 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.

**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 phosphorus deficiency (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  $\geq 5.0$  mt/ha, maturity period  $\leq 105$  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 recombinant inbred lines (RILs) at advanced



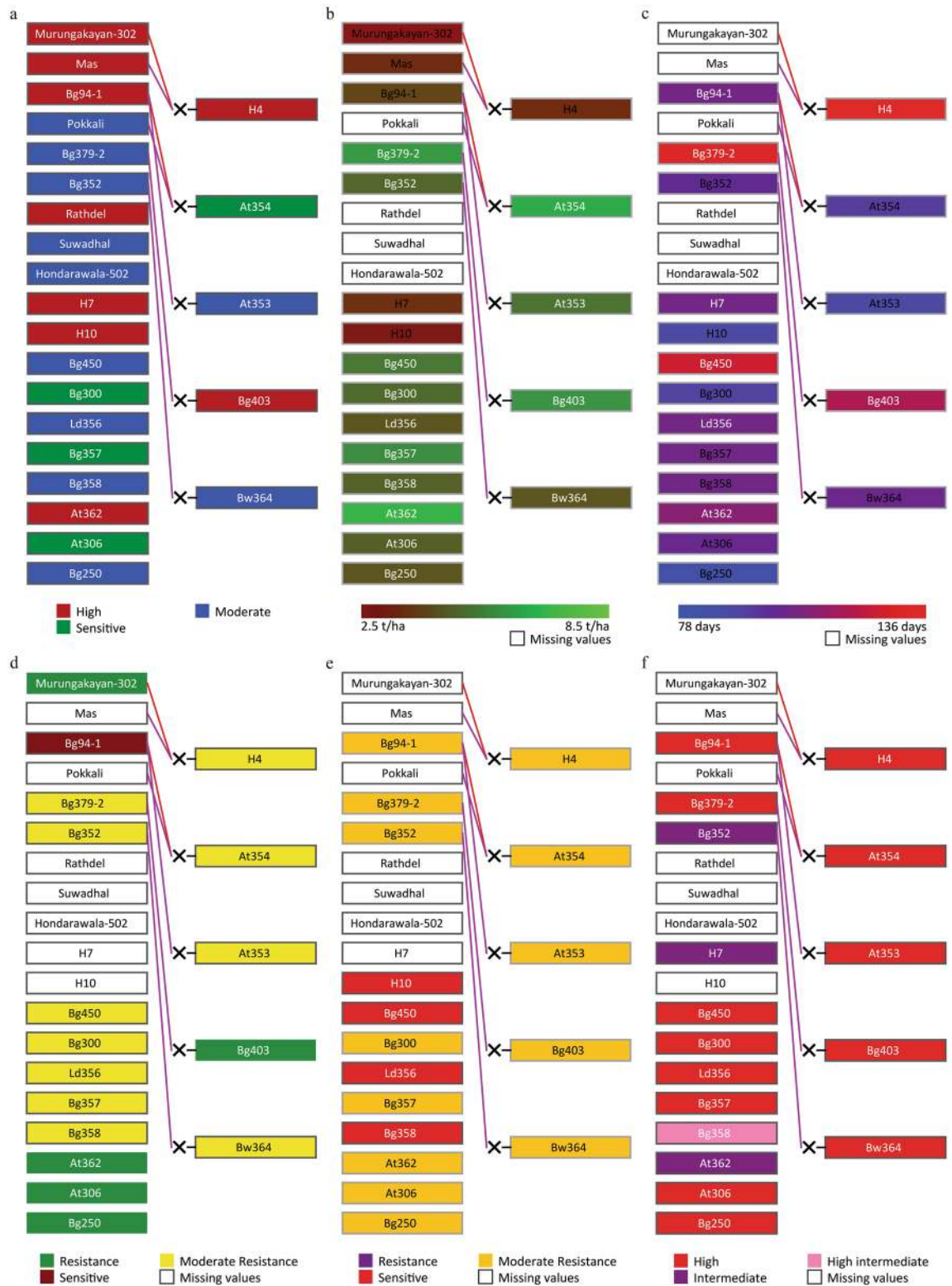
**Figure 3.** The pedigree visualization for Example 2 (parents with high, high-intermediate, and intermediate amylose content, yield  $\geq 3.5$  mt/ha, moderate or complete resistance to rice blast disease and maturity period  $\leq 125$  days). The selected pedigree is colored separately for four 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 ‘x’ indicates the cross between two parents, and ‘x’ inside the circle represents selfing. The background colors of the cultivar-name boxes indicate the trait values, as shown in the colored legends below.

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 Supplementary Fig. 2a,b online, respectively.

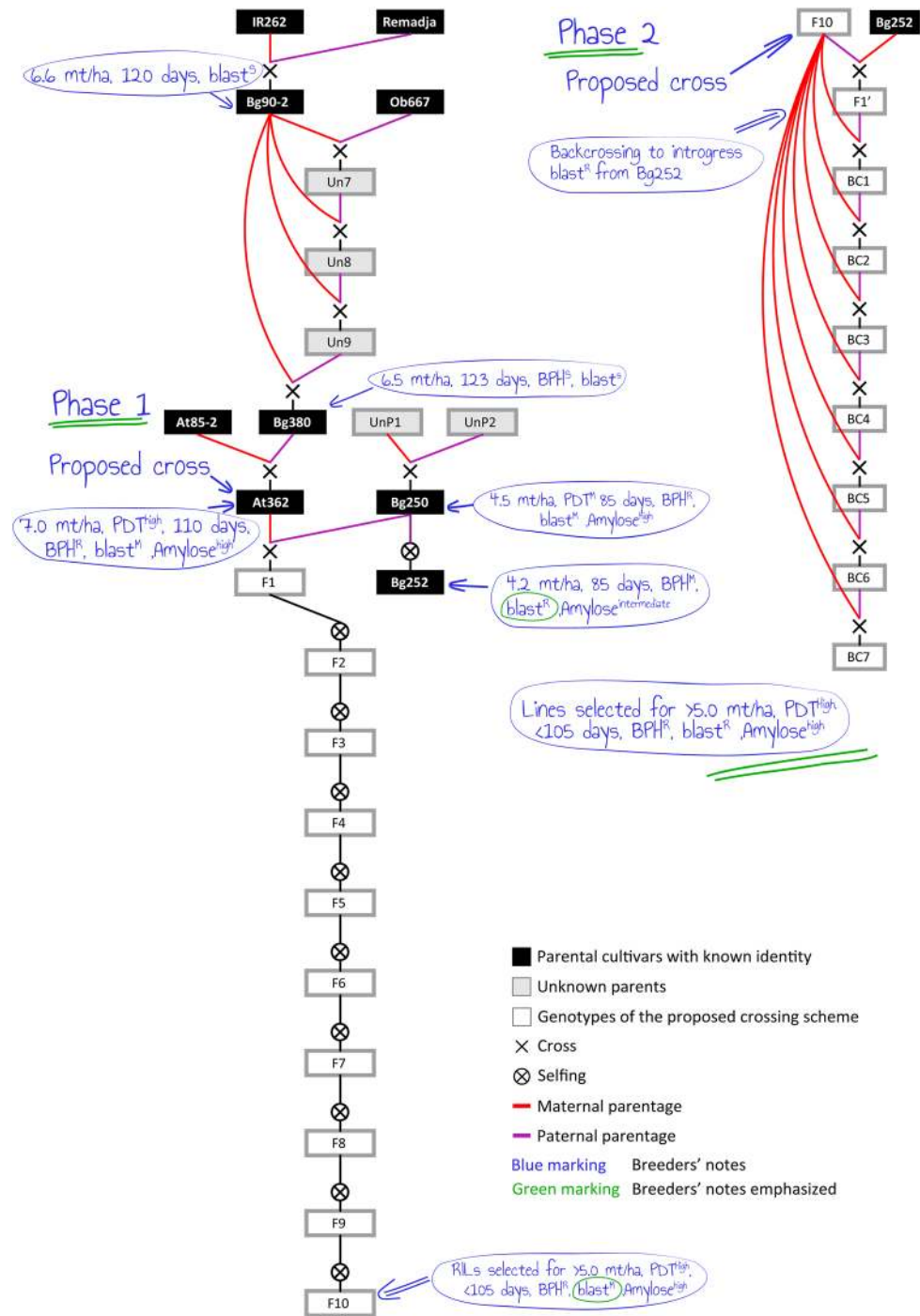
**Estimated breeding values (EBV) of the rice cultivars to support the breeding decisions in examples 1, 2 and 3.** The calculations revealed that the EBV-yield of the cultivar At307 is 1.03, which received the second rank that justifying its selection for the cross selected in example 1. The cultivar Bg450 is also ranked at 29th position in terms of its EBV-yield. Thus, Bg450 also brings a higher genetic effect for yield. However, Bg450 carries genes for extended maturity (EBV-maturity period rank of 62); however, At307 got the rank 20th; hence, the progeny has a chance to receive genes for a shorter maturity period. For plant height, At307 and Bg450 got rank of 50th and 24th positions, respectively, indicating that progeny would have a strong basis for shorter plant height, which is desirable to prevent lodging and increase the fertilizer use efficiently (Table 2; Figs. 6, 7 and Supplementary Table S4 online).

In example 2, our selection of Bg407H and At307 is firmly validated by the EBV-yield ranks of first and second received by these two cultivars, respectively. However, the EBV-maturity period of Bg407H was ranked 55th, and EBV-plant height was ranked 67th, indicating that Bg407H would bring favorable genes for extended maturity and taller plants. However, At307 got rank 20th for the EBV-maturity period and 15th for plant height, causing decreasing genetic effects on extended maturity period and tallness of the plants (Table 2; Figs. 6, 7 and Supplementary Table S4 online).

In example 3, our selection of At362 is validated by the EBV-yield rank. This cultivar provides the second-best possible genetic effect for yield in the breeding germplasm available in RRD1. In the proposed crossing schemes in Figs. 4 and 5, Bg250 and Bg252 were selected to provide the genetic basis for the shorter maturity period, and the EBV-maturity period values of those cultivars support this selection. Also, Bg250 was previously used by RRD1 as a breeding parent, its accuracies of EBV for maturity period were 0.98 and 0.99 for *Yala* and *Maha* seasons, respectively. The perfect correlations between EBV and TBV of Bg250 regarding the maturity period, indicates that it is an ideal parent to provide the genetic basis for shorter maturity period to the progeny.



**Figure 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 ‘x’ 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. White boxes indicate the cultivars with missing-trait values.



**Figure 5.** The pedigree visualization for planning a crossing scheme. Phase 1: Initial crossing of At362 and Bg250 and pedigree selection to obtain RILs with  $\geq 5.0$  mt/ha of mean yield,  $\leq 105$  days of the 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.

However, Bg250 and Bg252 provide a higher genetic effect for taller plants than that of At362 (Table 2; Figs. 6, 7 and Supplementary Table S4 online).

### Discussion

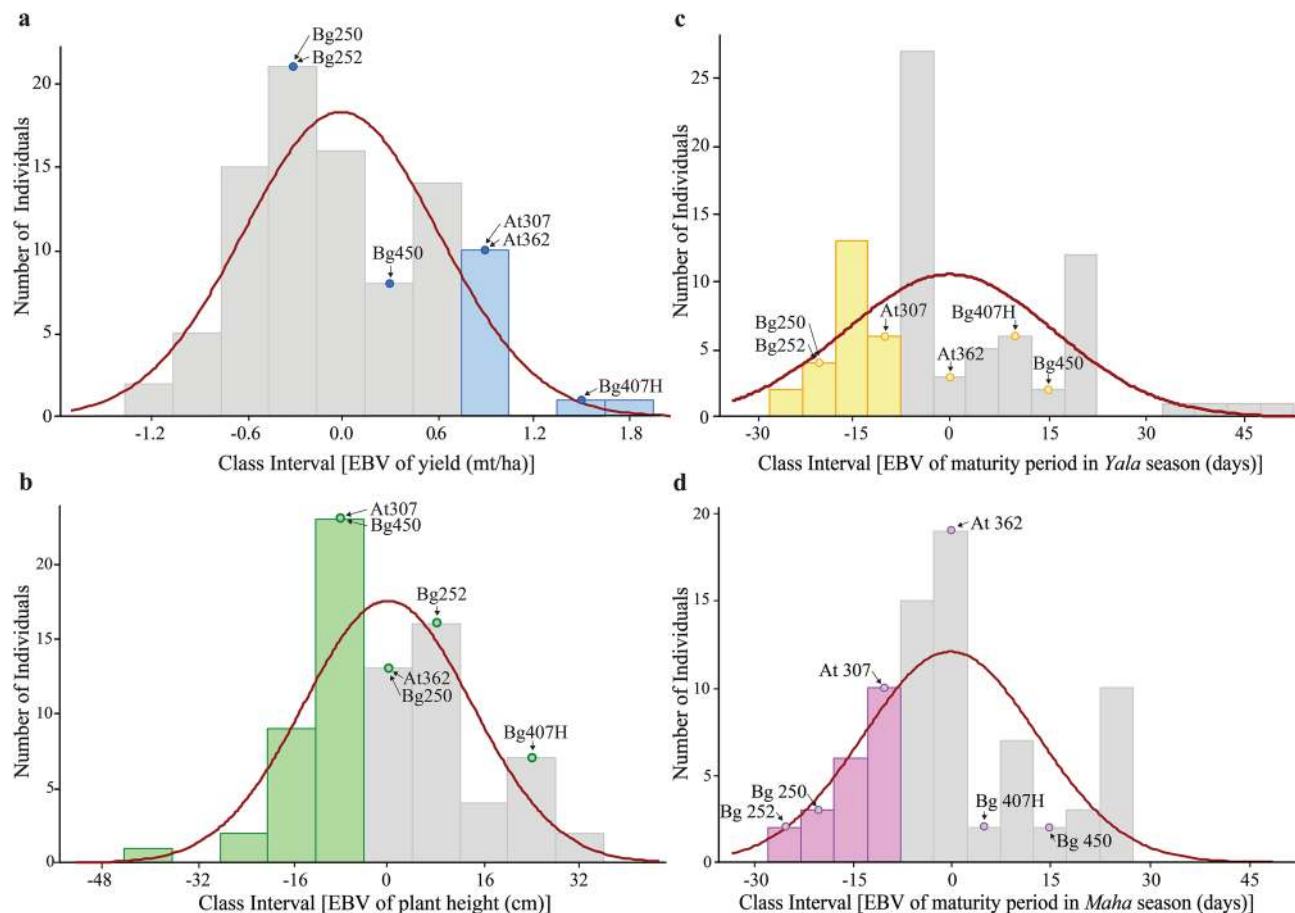
The decision-making process in breeding is a tedious task<sup>22</sup>. The breeding germplasm is complex with large numbers of improved varieties, traditional cultivars, landraces, wild germplasm, and accessions. Also, there can be large mapping populations and unreleased varieties due to various reasons. The numerous cultivars in breeding germplasm may have extensive records on agronomic data, pest and disease resistance, quality traits,



Rice cultivar <sup>§</sup>	Mean yield				Mean maturity period in <i>Yala</i> season				Mean maturity period in <i>Maha</i> season				Mean plant height			
	Mean (mt/ha)	EBV (mt/ha)	Rank for EBV <sup>†</sup>	Accuracy of EBV <sup>@</sup>	Mean (days)	EBV (days)	Rank for EBV	Accuracy of EBV	Mean (days)	EBV (days)	Rank for EBV	Accuracy of EBV	Mean (cm)	EBV (cm)	Rank for EBV	Accuracy of EBV
Bg407H*	8.00	1.53	1	–	118.00	8.81	55	–	118.00	6.62	55	–	104.00	27.16	67	–
At307*	7.00	1.03	2	–	97.00	–9.04	20	–	97.00	–11.23	20	–	59.00	–11.09	15	–
Bg369	7.00	1.03	2	–	104.00	–3.09	32	–	104.00	–5.28	32	–	59.00	–11.09	15	–
At362*	7.00	1.03	2	–	110.00	2.01	51	–	110.00	–0.18	51	–	70.00	–1.74	37	–
Bw453	7.00	1.03	2	–	135.00	23.26	66	–	135.00	21.07	66	–	78.00	5.06	50	–
Bg305	6.90	0.98	6	–	94.00	–11.59	16	–	94.00	–13.78	16	–	70.00	–1.74	37	–
At402	6.70	0.88	7	–0.27	115.00	6.26	53	0.61	115.00	4.07	53	0.48	85.00	11.01	57	–0.36
Bg90-2	6.50	0.78	8	0.21	120.00	10.51	57	0.35	120.00	8.32	57	0.21	60.00	–10.24	18	–0.11
Bg380	6.50	0.78	8	0.00	123.00	13.06	61	0.84	123.00	10.87	61	–0.39	63.00	–7.69	22	–0.84
At354	6.50	0.78	8	–	95.00	–10.74	17	–	95.00	–12.93	17	–	67.00	–4.29	30	–
At308	6.50	0.78	8	–	95.00	–10.74	17	–	95.00	–12.93	17	–	77.00	4.21	46	–
At303	6.20	0.63	12	–	90.00	–14.99	7	–	90.00	–17.18	7	–	74.00	1.66	44	–
Bg379-2	6.10	0.58	13	0.52	132.00	20.71	64	0.86	132.00	18.52	64	0.04	58.00	–11.94	13	0.55
Bg34-8	6.10	0.58	13	0.75	100.00	–6.49	23	0.95	100.00	–8.68	23	0.98	61.00	–9.39	20	0.59
At309	6.00	0.53	15	–	95.00	–10.74	17	–	95.00	–12.93	17	–	52.80	–16.36	6	–
Bg403	6.00	0.53	15	–	118.00	8.81	55	–	118.00	6.62	55	–	60.00	–10.24	18	–
Bg301	6.00	0.53	15	–	93.00	–12.44	12	–	93.00	–14.63	12	–	67.00	–4.29	30	–
Bg366	6.00	0.53	15	–	99.00	–7.34	22	–	99.00	–9.53	22	–	73.60	1.32	43	–
Bg455	6.00	0.53	15	–	130.00	19.01	63	–	130.00	16.82	63	–	77.00	4.21	46	–
Bg359	5.90	0.48	20	–0.99	104.00	–3.09	32	0.91	104.00	–5.28	32	0.97	64.00	–6.84	24	–0.31
Bg374	5.90	0.48	20	–	107.00	–0.54	49	–	107.00	–2.73	49	–	68.00	–3.44	35	–
Bg400-1	5.90	0.48	20	0.52	135.00	23.26	66	0.99	135.00	21.07	66	0.99	73.00	0.81	42	0.15
Bg357	5.80	0.43	23	–	104.00	–3.09	32	–	104.00	–5.28	32	–	56.00	–13.64	11	–
Bg310	5.60	0.33	24	–	97.00	–9.04	20	–	97.00	–11.23	20	–	65.00	–5.99	26	–
Bg251	5.50	0.28	25	–	80.00	–23.49	2	–	80.00	–25.68	2	–	87.00	12.71	60	–
Bg34-6	5.50	0.28	25	–	105.00	–2.24	37	–	105.00	–4.43	37	–	48.00	–20.44	3	–
Bw451	5.50	0.28	25	–0.84	120.00	10.51	57	–0.24	120.00	8.32	57	–0.52	66.00	–5.14	27	0.36
Bg370	5.40	0.23	28	–	105.00	–2.24	37	–	105.00	–4.43	37	–	85.00	11.01	57	–
Bg450*	5.30	0.18	29	–	125.00	14.76	62	–	125.00	12.57	62	–	64.00	–6.84	24	–
At353	5.20	0.13	30	–	90.00	–14.99	7	–	90.00	–17.18	7	–	80.00	6.76	51	–
At311	5.20	0.13	30	–	93.00	–12.44	12	–	93.00	–14.63	12	–	67.00	–4.29	30	–
Bw367	5.20	0.13	30	–	105.00	–2.24	37	–	3,105.00	–4.43	37	–	97.00	21.21	63	–
Bg250	4.50	–0.22	45	0.45	85.00	–19.24	3	0.98	85.00	–21.43	3	0.99	75.00	2.51	45	0.32
Bg252*	4.20	–0.37	57	–	85.00	–19.24	3	–	85.00	–21.43	3	–	81.00	7.61	53	–

**Table 2.** Estimated breeding values (EBV) and the ranks of the cultivars based on EBV for yield, maturity period (*Yala* and *Maha* seasons) and plant height. <sup>§</sup>According to the EBV of yield, top ranked 30 rice cultivars and, Bg250 and Bg252 (as they were selected as parents with short maturity periods in example 3) are shown. The entire results of EBV calculations are given in Supplementary Table S4 online. \*The selected parental cultivars in examples 1, 2, and 3 (Table 1; Figs. 2, 3, 4 and 5). <sup>†</sup>The rice cultivars are arranged according to the ranks concerning EBVs of the mean yield. <sup>@</sup>The accuracy of EBVs are only shown for the cultivars used as breeding parents at RRDI.

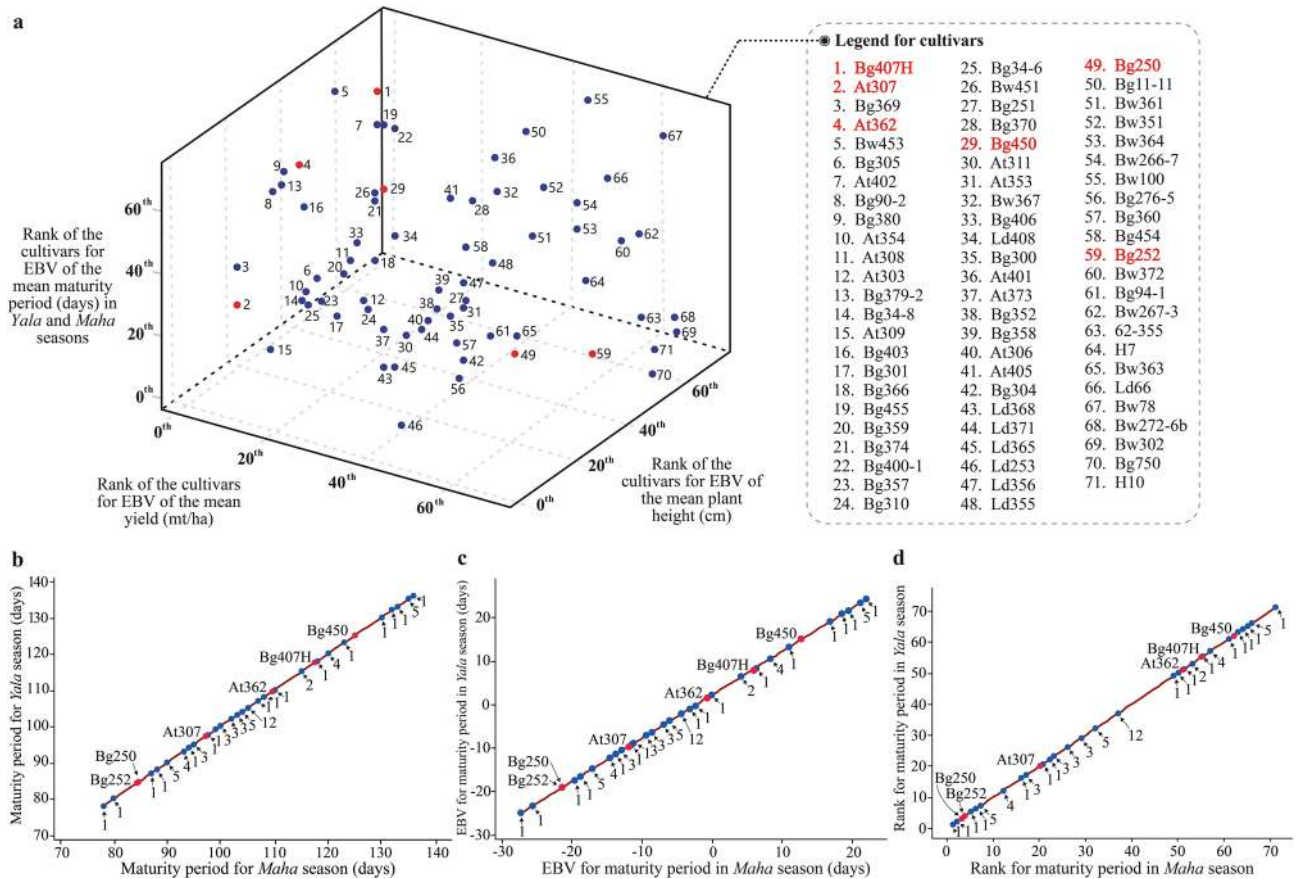
availability of samples, geographic locations, and utilization in diverse breeding programs as parents<sup>46,47</sup>. With the advent of DNA markers and sequencing technologies, a wealth of genomic information is also available<sup>48</sup>. However, one of the recurrent problems in any breeding germplasm 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<sup>28</sup>. 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



**Figure 6.** The distribution of the estimated breeding values (EBV) for rice cultivars assessed. (a): yield; (b): plant height; (c): maturity period in the *Yala* season; (d): maturity period in *Maha* season. The positions for the selected parental cultivars in examples 1, 2, and 3 (Table 1; Figs. 2, 3, 4 and 5) are marked within the histograms. The colored bars in the histograms show the cultivars with desirable EBVs (Table 2; Supplementary Table S4 online).

visualize the essential information and complex pedigrees. Breeding programs often suffer a lot when the breeder gets retired or moved to a different position<sup>49–51</sup>. 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<sup>52</sup>. 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<sup>28</sup>.

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, 3, 4 and 5). However, it is essential to characterize the cultivars for all the important traits, molecular markers, and SNP haplotypes<sup>53</sup>, so that breeding decisions can be effectively made<sup>17</sup>. The EBVs for the parental cultivars and progenies further can be consolidated with the pedigrees to intensify the reliability of the breeding decisioning<sup>54</sup>. 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<sup>44,55</sup>. The novel Agri-tech practices such as vertical farming, artificial intelligence-powered technology, and re-energizing the plant microbiome would improve conventional breeding, leading to second green-revolution<sup>36,56</sup>. Therefore, in addition to genotyping technologies, including whole genome sequencing and DNA marker-assisted selection techniques, high-throughput phenotyping tools/phenome platforms are also essential to develop further breeding systems.



**Figure 7.** The comparative visualization of the ranks of rice cultivars based on the estimated breeding values (EBV) for yield, maturity period and plant height were considered as desirable in ranking. (a): 3-D scatter plot depicting the cultivar-positions with respect to EBV ranks of yield, maturity period and plant height; (b): The linear relationship of maturity periods of the rice cultivars reported in *Yala* and *Maha* seasons; (c): The linear relationship of the EBVs of the maturity periods of rice cultivars reported in *Yala* and *Maha* seasons; (d): The linear relationship of the ranks of EBVs of the maturity periods of rice cultivars reported in *Yala* and *Maha* seasons. The positions for the selected parental cultivars in examples 1, 2, and 3 (Table 1; Figs. 2, 3, 4 and 5) are marked on the curves of the figure-panels (a), (b) and (c). In some instances, more than one cultivar is represented by the ‘dot’ positions of the curve, therefore, numbers with pointing arrows are indicated to show the number of cultivars represented by each dot in the figure panels (b), (c) and (d). Because of the patterns observed in (b), (c) and (d), only one axis was used to represent the maturity period in the figure panel (a).

The application of EBV in breeding is a common practice to decide the additive genetic effect that each parent can bring to the progeny<sup>57</sup>. There are only three quantitative parameters (yield, maturity period for *Yala* and *Maha* seasons and plant height) available in the breeding germplasm at RRDI (Supplementary Table S1 online). We calculated EBV for these three parameters (Supplementary Table S4 online). The first 30 top ranked cultivars for EBV, together with two other important cultivars used in example 3, are given in Table 2. Also, the accuracy/reliability of EBVs are given in Table 2 and Supplementary Table S4 online for the cultivars that were used by RRDI as breeding parents. It is evident from the trait data, EBVs and accuracy of EBVs given in Table 2 that the breeding germplasm at RRDI got elite cultivars that can be used as breeding parents in the future. Interestingly, all these are newly improved and released rice varieties. It is evident that the crossing schemes must always plan using these elite cultivars as parents while carefully adding other landraces of exotic types as resistant sources to avoid linkage drags. The EBVs of yield (Fig. 6a), plant height (Fig. 6b), and maturity periods in *Yala* and *Maha* seasons (Fig. 6c,d) show continuous distribution. For yield and maturity period, RRDI germplasm has promising rice cultivars. However, for plant height, the high yielding parents tend to have increasing genetic effects. The lodging is a frequent problem in rice farming in Sri Lanka; thus, current EBV-plant height estimates imply the necessity of boarding the breeding germplasm with the parental cultivars that can provide a genetic basis for short plants.

The relative rankings of rice cultivars for EBEs calculated for yield, maturity period, and plant height are given in Fig. 7a. The even distribution of rice cultivars in the 3-d sphere (i.e., box based on ranks) highlights the broad genetic diversity of Sri Lankan rice breeding germplasm. However, it has to be completely characterized, and the EBVs must be calculated to understand the complex-multidimensional diversity structures to carry out the Pedimap decision procedures in designing crosses efficiently. The EBV estimates for maturity period imply that

except five cultivars, on average other cultivars do not show any significant differences in *Yala* and *Maha* seasons. The lack of seasonal variations for the maturity period is an advantage for breeding programs as two seasons of selection are possible in every year to fast track the variety development process. (Fig. 7b–d).

In the present study, we only used phenotypic data available for traits to calculate the EBVs. However, for efficient genomic selection, high throughput genomic data such as marker alleles, sequence polymorphisms, and haplotype variants are needed. Thereby EBVs can be translated into more robust genome EBVs (GEBVs). The GEBVs would facilitate the efficient introgression of desirable traits to new varieties through MAB with efficient background and foreground selection schemes<sup>58,59</sup>.

## 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 and the calculated EBVs highly supports to the validity of decisioning based on Pedimap. 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.

## Data availability

All the data of the manuscript are available in Supplemental Material.

Received: 18 March 2020; Accepted: 21 July 2020

Published online: 31 August 2020

## References

- Food and Agriculture Organization of the United Nations (FAO). *Rice Market Monitor* Vol. 21, 1–3 (2018).
- Muthayya, S., Sugimoto, J. D., Montgomery, S. & Maberly, G. F. An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.* **1324**, 7–14. <https://doi.org/10.1111/nyas.12540> (2014).
- Ali, A. M. Rice to shrimp: land use/land cover changes and soil degradation in Southwestern Bangladesh. *Land Use Policy* **23**, 421–435. <https://doi.org/10.1016/j.landusepol.2005.02.001> (2006).
- Seck, P. A., Diagne, A., Mohanty, S. & Wopereis, M. C. Crops that feed the world 7: rice. *Food Secur.* **4**, 7–24. <https://doi.org/10.1007/s12571-012-0168-1> (2012).
- Segal, R. & Le Nguyen, M. *Unfair Harvest: The State of Rice in Asia* (Oxfam International, Oxford, 2019).
- Rice Cultivation. Sri Lanka. <https://www.doa.gov.lk/rrdi/>. Accessed 15 July 2019.
- Annual Report of Central Bank of Sri Lanka. <https://www.cbsl.gov.lk/en/publications/economic-and-financial-reports/annual-reports/annual-report-2017>. Accessed 18 May 2019 (2017).
- Annual Report of Central Bank of Sri Lanka. <https://www.cbsl.gov.lk/en/publications/economic-and-financial-reports/annual-reports/annual-report-2018>. Accessed 18 May 2019 (2018).
- Dharmarathna, W. R., Herath, S. & Weerakoon, S. B. Changing the planting date as a climate change adaptation strategy for rice production in Kurunegala district Sri Lanka. *Sustain. Sci.* **9**, 103–111. <https://doi.org/10.1007/s11625-012-0192-2> (2012).
- IPCC. In *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Parry, M. L., Canziani, O. F., Palutikof, J. P., Van der Linden, P. J. & Hanson, C. E.) 976 (Cambridge University Press, 2007).
- Mackill, D. J., Coffman, W. R. & Garrity, D. P. *Rainfed Lowland Rice Improvement* Vol. 242 (International Rice Research Institute, Los Baños, 1996).
- Walisinghe, B. R., Gunaratne, L. H. P., Prasada, D. V. P. & Sarthaj, I. Z. Analysis of the consumer preferences for rice-based noodles and biscuits. *Ann. Sri Lanka Dep. Agric.* **12**, 201–211 (2010).
- Dhanapala, M. P. Bridging the rice yield gap in Sri Lanka. In *Proceedings of the Symposium on Expert Consultation on Bridging the Rice Yield Gap in the Asia-Pacific Region, Bangkok, Thailand*, 135–145 (1999).
- Shetty, P. K. Socio-ecological implications of pesticide use in India. *Econ. Polit. Wkly.* **39**, 5261–5257 (2004).
- Duvick, D. N. Progress in conventional plant breeding. In *Gene Manipulation in Plant Improvement* (ed. Gustafson, J. P.) 17–31 (Springer, Boston, 1984).
- Calingacion, M. *et al.* Diversity of global rice markets and the science required for consumer-targeted rice breeding. *PLoS ONE* **9**, e85106. <https://doi.org/10.1371/journal.pone.0085106> (2014).
- Jiang, G. L. Plant marker-assisted breeding and conventional breeding: challenges and perspectives. *Adv. Crop Sci. Technol.* **1**, e106. <https://doi.org/10.4172/2329-88631000e106> (2013).
- Sasaki, T. Power of conventional rice breeding. *Breed. Sci.* **61**, 317. <https://doi.org/10.1270/jsbbs.61.317> (2011).
- Collard, B. C. & Mackill, D. J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. B.* **363**, 557–572. <https://doi.org/10.1098/rstb.2007.2170> (2007).
- Tanksley, S. D. & Nelson, J. C. 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.* **92**, 191–203. <https://doi.org/10.1007/BF00223376> (1996).
- Thomson, M. J., Ismail, A. M., McCouch, S. R. & Mackill, D. J. Marker assisted breeding. In *Abiotic Stress Adaptation in Plants* (eds Sopory, S. K. *et al.*) 451–469 (Springer, Dordrecht, 2009).
- Acquaah, G. *Principles of Plant Genetics and Breeding* 2nd edn, 281–334 (Wiley-Blackwell, Oxford, 2012).
- Clark, B. H. Marketing performance measures: history and interrelationships. *J. Mark. Manag.* **15**, 711–732. <https://doi.org/10.1362/026725799784772594> (1999).
- 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. <https://www.rtb.cgiar.org/gender-breeding-initiative> (2008).
- Kariuki, C. M., van Arendonk, J. A., Kahi, A. K. & Komen, H. Multiple criteria decision-making process to derive consensus desired genetic gains for a dairy cattle breeding objective for diverse production systems. *J. Dairy Sci.* **100**, 4671–4682. <https://doi.org/10.3168/jds.2016-11454> (2017).
- Velasquez, M. & Hester, P. T. An analysis of multi-criteria decision-making methods. *Int. J. Oper. Res.* **10**, 56–66 (2013).
- Yu, J. *et al.* CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Res.* **42**, D1229–D1236. <https://doi.org/10.1093/nar/gkt1064> (2013).
- Voorrips, R. E., Bink, M. C. & van de Weg, W. E. Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. *J. Hered.* **103**, 903–907. <https://doi.org/10.1093/jhered/ess060> (2012).

29. Aluwihare, Y. C. *et al.* Characterization and selection of phosphorus deficiency tolerant rice genotypes in Sri Lanka. *Rice Sci.* **23**, 184–195. <https://doi.org/10.1016/j.rsci.2015.10.001> (2016).
30. Aluwihare, Y. C. *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 phosphorus deficiency tolerance. *Ceylon J. Sci. (Bio. Sci.)* **44**, 45–54 (2015).
31. Jayarathne, H. S. M. *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.* In review (2020).
32. Nakandala, N. D. *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.* **48**, 121–132. <https://doi.org/10.4038/cjs.v48i2.7616> (2019).
33. Rathnayake, R. M. S. K. *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 J. Agric. Ecosyst.* **1**, 2–15 (2019).
34. Bink, M. C. *et al.* Bayesian analysis of complex traits in pedigreed plant populations. *Euphytica* **161**, 85–96. <https://doi.org/10.1007/s10681-007-9516-1> (2008).
35. RRD1 (2018). *Recommended Rice Varieties in Sri Lanka (1958–2016)*. (Research and Development Institute, Department of Agriculture, Bathalagoda, Ibbagamuwa, Sri Lanka, 2018).
36. Quddus, C. *et al.* Estimating pedigree-based breeding values and stability parameters of elite rice breeding lines for yield under salt stress during the Boro season in Bangladesh. *Plant Breed. Biotechnol.* **7**, 257–271. <https://doi.org/10.9787/PBB.2019.7.3.257> (2019).
37. Oldenbroek, K. & van der Waaij, L. Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources, The Netherlands and Animal Breeding and Genomics Centre, 2015. Groen Kennisnet: <https://wiki.groenkennisnet.nl/display/TAB/> (2015).
38. Seyoum, M., Alamerew, S. & Bantte, K. Genetic variability, heritability, correlation coefficient and path analysis for yield and yield related traits in upland rice (*Oryza sativa* L.). *J. Plant Sci.* **7**, 13–22. <https://doi.org/10.3923/jps.2012> (2012).
39. Sanghera, G. S., Kashyap, S. C. & Parry, G. A. Genetic variation for grain yield and related traits in temperate red rice (*Oryza sativa* L.) ecotypes. *Not. Sci. Biol.* **5**, 400–406. <https://doi.org/10.15835/nsb539088> (2013).
40. Perera, A. L. T. *et al.* Quantitative genetics of rice. I. Evidence of unexploited genetical variation for yield and other quantitative characters in modern Indica cultivars. *Field Crops Res.* **55**, 245–256. [https://doi.org/10.1016/S0378-4290\(97\)00089-0](https://doi.org/10.1016/S0378-4290(97)00089-0) (1998).
41. Weebadde, C. *et al.* Use of PediMap in RosBREED for visualizing genetic and phenotypic data across pedigrees in rosaceous crops. *Hort. Sci.* **45**, S55–S55 (2010).
42. Peace, C. P., Luby, J. J., Van de Weg, W. E., Bink, M. C. & Iezzoni, A. F. A strategy for developing representative germplasm sets for systematic QTL validation, demonstrated for apple, peach, and sweet cherry. *Tree Genet. Genomes* **10**, 1679–1694. <https://doi.org/10.1007/s11295-014-0788-z> (2014).
43. Rosyara, U. R. *et al.* Fruit size QTL identification and the prediction of parental QTL genotypes and breeding values in multiple pedigreed populations of sweet cherry. *Mol. Breed.* **32**, 875–887. <https://doi.org/10.1007/s11032-013-9916-y> (2013).
44. Evans, K. *et al.* Large-scale standardized phenotyping of apple in RosBREED. *Acta. Hortic.* **945**, 233–238. <https://doi.org/10.17660/ActaHortic.2012.945.31> (2012).
45. Paulo, M. J., Boer, M., Huang, X., Koornneef, M. & Van Eeuwijk, F. A mixed model QTL analysis for a complex cross population consisting of a half diallel of two-way hybrids in *Arabidopsis thaliana*: analysis of simulated data. *Euphytica* **161**, 107–114. <https://doi.org/10.1007/s10681-008-9665-x> (2008).
46. Davenport, G., Ellis, N., Ambrose, M. & Dicks, J. Using bioinformatics to analyse germplasm collections. *Euphytica* **137**, 39–54. <https://doi.org/10.1023/B:EUPH.0000040501.06112.ca> (2004).
47. Peeters, J. P. & Galwey, N. W. Germplasm collections and breeding needs in Europe. *Econ. Bot.* **42**, 503–521. <https://doi.org/10.1007/BF02862795> (1988).
48. Nadeem, M. A. *et al.* DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Equip.* **32**, 261–285. <https://doi.org/10.1080/13102818.2017.1400401> (2018).
49. Sooriyapathirana, S. D. S. S. *et al.* Using SNP/INDEL diversity patterns to identify a core group of genotypes from FVC11, a superior hybrid family of *Fragaria virginiana* Miller and *F. chiloensis* (L.) Miller. *Genet. Resour. Crop Evol.* **66**, 1691–1698. <https://doi.org/10.1007/s10722-019-00819-0> (2019).
50. USDA Roadmap for Plant Breeding March 11, 2015: Final Report. United States Department of Agriculture: Office of the Chief Scientist Research, Education, and Economics Mission Area. <https://www.usda.gov/wps/portal/usda/usdahome?navid=OCS>. Accessed 25 July 2019 (2015).
51. Hmielowski, T. Challenges to developing cultivars in the public sector. *Crops Soils Agron. News* **62**, 4–5. <https://doi.org/10.2134/csa2017.62.1001> (2017).
52. Zhu, T. *et al.* CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol.* **17**, 101. <https://doi.org/10.1186/s12870-017-1039-x> (2017).
53. Vanderzande, S. *et al.* High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. *PLoS ONE* **14**, e0210928. <https://doi.org/10.1371/journal.pone.0210928> (2019).
54. Fukai, S. & Fischer, K. S. Field phenotyping strategies and breeding for adaptation of rice to drought. *Front. Physiol.* **3**, 282. <https://doi.org/10.3389/fphys.2012.00282> (2012).
55. Raina, A. *et al.* Role of mutation breeding in crop improvement-past, present and future. *Asian Res. J. Agric.* <https://doi.org/10.9734/ARJA/2016/29334> (2016).
56. Marvin, D. R. The Second Green Revolution Will Bring Agri-Tech Breakthroughs to Growers. *Ind. Biotechnol.* **14**, 120–122. <https://doi.org/10.1089/ind.2018.29129.drm> (2018).
57. Cobb, J. N. *et al.* Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theor. Appl. Genet.* **132**, 627–645. <https://doi.org/10.1007/s00122-019-03317-0> (2019).
58. Spindel, J. *et al.* 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.* **11**, e1004982. <https://doi.org/10.1371/journal.pgen.1004982> (2015).
59. Collard, B. C. Transforming rice breeding: re-designing the irrigated breeding pipeline at the International Rice Research Institute (IRRI). *Crop Breed. Genet. Genom.* **1**, e190008. <https://doi.org/10.20900/cbpg20190008> (2019).

## Acknowledgements

All the breeders and the staff of the Rice Research and Development Institute, Sri Lanka. Dr. Roeland Voorrips and Dr. Eric van de Weg at Wageningen University and Research Center Nijmegen Area, Netherlands are acknowledged for kindly providing licenses for the Pedimap and FlexQTL analysis. Dr. Roeland Voorrips is also acknowledged for critically reviewing the manuscript.

### Author contributions

Su.S., R.R., K.U., C.W. and W.W. developed the concept. K.U., W.W., C.W., R.R. and Su.S. provided the resources. R.R., Su.S. and Sh.S. performed the experiments and tabulated data. R.R., Su.S. and C.W. wrote the manuscript and analyzed the experiments. Su.S., R.R., K.U., C.W., W.W. and Sh.S. contributed to the finalization and approval of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-71260-y>.

**Correspondence** and requests for materials should be addressed to S.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020