Note

IonBreeders: bioinformatics plugins toward genomics-assisted breeding

Eri Ogiso-Tanaka*, Shiori Yabe and Tsuyoshi Tanaka

Institute of Crop Science, National Agriculture and Food Research Organization (NARO), 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8518, Japan

Polymorphism information generated by next-generation sequencing (NGS) technologies has enabled applications of genome-wide markers assisted breeding. However, handling such large-scale data remains a challenge for experimental researchers and breeders, calling for the urgent development of a flexible and straightforward analysis tool for NGS data. We developed "IonBreeders" as bioinformatics plugins that implement general analysis steps from genotyping to genomic prediction. IonBreeders comprises three plugins, "ABH", "IMPUTATION", and "GENOMIC PREDICTION", for format conversion of genotyping data, preprocessing and imputation of genotyping data, and genomic prediction, respectively. "ABH" converts genotyping data derived from NGS into the ABH format, which is acceptable for our further plugins and with other breeding software tools, R/qtl, MapMaker, and AntMap. "IMPUTATION" filters out non-informative markers and imputes missing marker genotypes. In "GENOMIC PREDICTION", users can use four statistical methods based on their target trait, quantitative trait locus effect, and number of markers, and construct a prediction model for genomic selection. IonBreeders is operated in Torrent Suite, but can also handle genotype data in standard formats, e.g., Variant Call Format (VCF), by format conversion using free software or our provided scripts.

Key Words: next-generation sequencing, plugin, genotyping, genomic prediction.

Introduction

Next-generation sequencing (NGS) data are now widely utilized in studies related to genetics and breeding with advantages of a reduction of sequencing costs and analytical support owing to the development of various analysis platforms (Phan and Sim 2017). For example, genetic analyses now predominantly focus on the identification of insertion-deletion polymorphisms (Indel) and single nucleotide polymorphism (SNP) markers developed by genome resequencing and genotyping by sequencing (GBS) rather than on ordinary simple sequence repeat (SSR) markers (Kim *et al.* 2016, 2017). Identification of genome-wide markers enables detailed quantitative trait locus (QTL) analysis and genome-wide association studies (GWAS) (Wang *et al.* 2016) to isolate genes and identify closely related markers for agronomically important traits, leading to significant progress in the field of breeding science.

However, the markers derived from QTL analysis or GWAS still do not have sufficient explanatory power for obtaining the desired phenotypes in marker-assisted breeding on polygenic traits (Bernardo 2008). Alternatively, genomic selection (GS) was developed as a new breeding technology for genome-wide prediction using all available markers (Bhat *et al.* 2016, Meuwissen *et al.* 2001). This technique establishes prediction models in a training population using both phenotype and marker genotype data, resulting in predicted genetic values for individuals based only on marker genotype information. For instance, the genomic prediction has been applied to maize, wheat, rice, soybean, rapeseed, buckwheat, and tomato in plants (Crossa *et al.* 2014, Matei *et al.* 2018, Spindel *et al.* 2015, Würschum *et al.* 2014, Yabe *et al.* 2018, Yamamoto *et al.* 2017). However, the technique is not familiar to breeders.
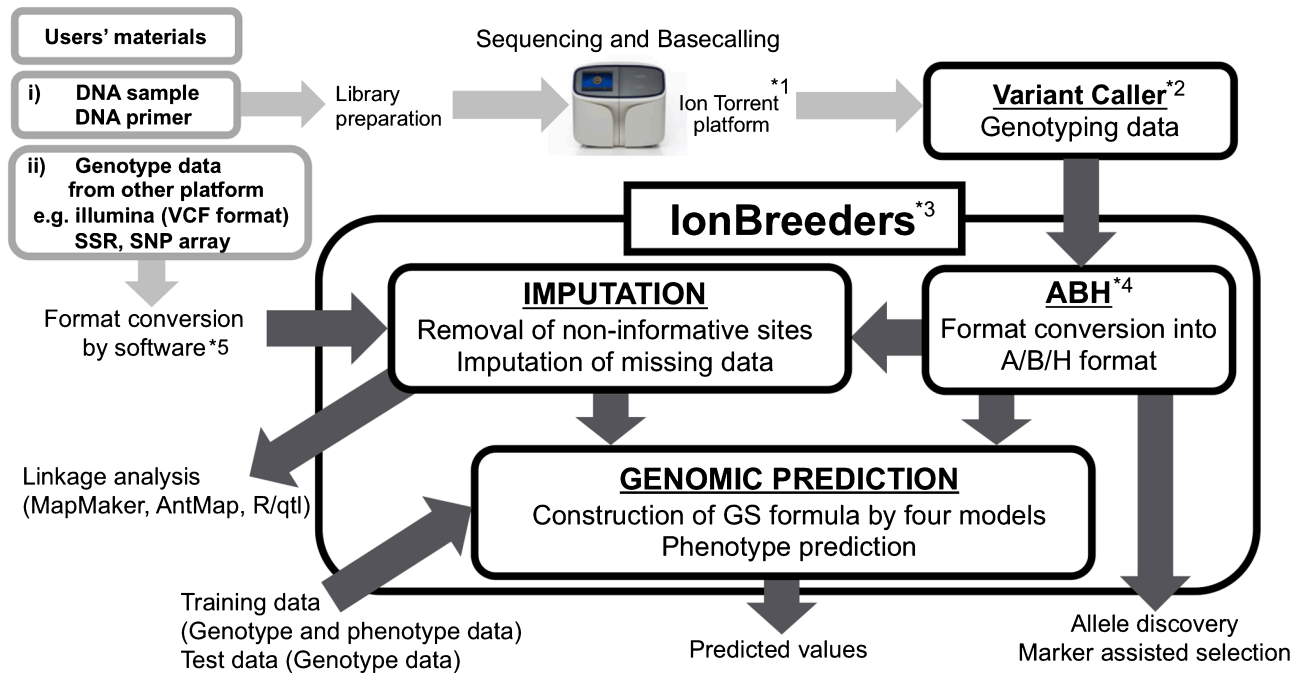
One of the main limitations of the practical application of GS is the lack of a suitable bioinformatics infrastructure for analysis that is easy to use for people who are unfamiliar with informatics. In particular, limited graphical user interface (GUI)-based free software tools are available for handling NGS data. Galaxy is one of the most popular GUI tools for genome data analysis (Afgan *et al.* 2018) and comprises various pipelines for NGS analysis, including prediction of genomic breeding values (Juanillas *et al.* 2019). However, Galaxy is a web-based system and cannot be used locally without introduction of complicated system settings. Moreover, although the representative NGS platforms such as Illumina (e.g., MiSeq and MiniSeq) and

Fig. 1. Workflow of genotyping by purpose using IonBreeders plugin in IonTorrent platform. The plugin names are in bold and underlined. IonBreeders is consist of three plugins, "ABH", "IMPUTATION" and "GENOMIC PREDICTION". The dark grey arrows show the input/output of the plugins.

IonTorrent (e.g., IonPGM, Ion S5) provide output data as short reads, it remains a challenge to convert these large-scale genotyping data to file formats that are acceptable for downstream genetic analysis and interpretation.

To solve this problem, we developed a new analysis plugin named "IonBreeders" to facilitate analyses of genotyping and genome data for genomic prediction. The main goal of IonBreeders is to enable obtaining genotype data and conducting genomics-assisted breeding with only basic knowledge of NGS and genetics, without bioinformatics, machine learning, and more specialized skills. By integrating IonBreeders with another NGS analysis tool, users can convert the data to a genotyping format, impute missing genotyping data, and perform genomic prediction with several available statistical models.

Materials and Methods

IonBreeders was developed as a series of plugins for the Torrent Suite, which is a GUI-based analysis software in the IonTorrent system (maintained by Thermo Fisher Scientific) and is operated on the Ion PGM, GeneStudio S5/S5 Plus/S5 Prime sequence system (Thermo Fisher Scientific), Ion Torrent Server (Thermo Fisher Scientific), or

Linux virtual machine (**Fig. 1**). Torrent Suite software and IonBreeders can be downloaded from Github (**Fig. 1**). After installation of Torrent Suite, users can download IonBreeders and test data from Github at https://github.com/DEMETER298/IonBreeders, and install the program on Torrent Suite. Detailed instructions are provided on the IonBreeders wiki page (**Fig. 1**, **Supplemental Text 1**).

Results

IonBreeders is a program comprising three plugins: "ABH", "IMPUTATION", and "GENOMIC PREDICTION" (**Fig. 1**). Since each plugin is performed automatically, users just provide an output file from the previously run plugin as an input for the next plugin based on the analysis step to use IonBreeders. We could confirm that format conversion by the ABH plugin worked properly and the IMPUTATION plugin filtered out and imputed data reasonably (Ishikawa *et al.* 2018, Marubodee *et al.* 2015, Uga *et al.* 2013, Zhao *et al.* 2010). The GENOMIC PREDICTION plugin could generate the predicted values with reasonable accuracy. Furthermore, we have also confirmed the operation of the plugins in genetic analysis of F_2 and RIL population with SSR and amplicon markers to show that

they work in palaeopolyploid soybean (in preparation). IonBreeders can be used in both English and Japanese, and the user can conveniently switch the language of preference on the screen plugin. Each of these plugins is described in detail below.

ABH: format converter

In the Ion Torrent system, variant detection was performed by the Variant Caller plugin (one of the Torrent Suite software) (**Fig. 1**). By specifying hotspot (marker site) information, Variant Caller plugin can detect the presence or absence of variants on the site. The Variant Caller plugin output genotype data in a Variant Call Format (VCF) and as a Torrent Variant Caller (TVC) output. In these files, variants are shown as the same ("0/0" and "Absent" in VCF and TVC output), different ("1/1" and "Homozygous") or heterozygous ("0/1" and "Heterozygous") compared to the reference genome. The ABH plugin uses only the genotype data on the hotspot site. The ABH plugin can convert a file format of polymorphism detection obtained by Variant Caller output with hotspot analysis to formats with either A/B/H or reference/homo/hetero information. When there is no information for polymorphic regions or sites (marker position), a hotspot file can be generated from the TVC output of Variant Caller plugin for subsequent variant detection. If the genotypes (bi-allelic sites) of both parents are known, genotypes of their segregation population are established in the ABH formatted genotype data (A = parent 1 allele, B = parent 2 allele and H = Hetero) based on the input of parent genotypes by ABH plugin (**Supplemental Text 1**). The ABH genotype file is output in the Comma-Separated Value (CSV) format ("csvr" format in R/qtl; http://www.rqtl.org/sampledata/ [Broman *et al.* 2003]). This file can be used as input file for R/qtl software for mapping QTL, "IMPUTATION" and GENOMIC PREDICTION plugins in IonBreeders. Of the three plugins in IonBreeders, the ABH plugin handles only Ion Torrent data. For other two plugins, several types of genotype data are acceptable ("csvr" format in R/qtl from ABH plugin output and other platforms, e.g. SSR, array and other NGS data) as input files (**Fig. 1**, **Supplemental Text 1**). Since the input file for IMPUTATION and GENOMIC PREDICTION plugin is in the CSV format, users can also construct the genotype file using their genotype data, e.g., SSR markers, in the Excel platform (Microsoft Inc.). In case that only genotype file in the Variant Call Format (VCF) from the Genome Analysis toolkit (McKenna *et al.* 2010) and the bcftools (http://github.com/samtools) is available, users need to convert from VCF to ABH format using either public tools, e.g. TASSEL (Bradbury *et al.* 2007) or our Perl scripts (VCF2ABH.pl: https://github.com/DEMETER298/ genotyping_illumina). If the genotypes of both parents are unknown, genotypes of their segregation population are established in ABH formatted genotype data based on the reference genome (**Supplemental Text 1**). Even if we use a population with unrelated individuals (e.g. genetic

resources) instead of segregating populations, the plugin can also be applied to reference genome-based genotyping so that the output file can be used to perform genomic prediction or discover new alleles.

IMPUTATION: data processing and complementation of missing data

After the genotyping using NGS data, preprocessing of the genotypes is indispensable step. For example, missing data and non-informative markers should be excluded from the genotyping dataset in the process of genomic prediction. The step is necessary to improve not only the quality of genotype data but also the accuracy of the further analysis. The plugin "IMPUTATION", that is followed by the former plugin "ABH" in the IonBreeders, either excludes low quality/non-informative sites or imputes genotypes for subsequent genetic analysis. Based on the thresholds, each site is defined as either reliable sites, low quality/non-informative sites and imputed sites in the following strategy. First, heterozygosity rate and missing rates are calculated for each site. Then, from the calculated rates and the threshold users set up, the sites are grouped into either low quality sites or imputed sites. Finally, perfectly linked markers detected from the same NGS reads or within a close genetic distance are integrated as a single polymorphism. The "IMPUTATION" plugin is constructed by nine processed and partially a wrapper of several functions implemented in R/qtl (Broman *et al.* 2003, **Supplemental Text 1**).

To conduct the plugin, users prepare either the output file from the former plugin, "ABH" or the CSV-formatted file with ABH formatted genotype data along with genetic distance information (**Supplemental Text 1**), and returns the output file in the same format. Thus, the user can follow and monitor the process in Excel. This plugin can impute missing genotype using neighbor polymorphism information and genetic distances. The output files (".csv" and ".raw" files) of the plugin can also be applied to other conventional widely used genetic analysis tools for breeding applications, such as MapMaker (Sharma and Kaur 2014), AntMap (Iwata and Ninomiya 2006), and R/qtl (Broman *et al.* 2003). Therefore, the plugin mainly supposes populations with highly related individuals such as F_2 and RIL populations in data preprocessing for linkage analysis, QTL analysis, and genomic prediction (**Supplemental Text 1**).

The IMPUTATION plugin is preferable for data analysis of genotyping sequences from specific regions/sites with a sufficient read depth (at least 10X coverage), such as amplicon sequences, including Amplicon-seq with a two-step tailed PCR (Cruaud *et al.* 2017, Ishikawa *et al.* 2018), GT-seq (Campbell *et al.* 2015), MTA-seq (Onda *et al.* 2018), AmpliSeq (in preparation), and GRAS-Di (Hosoya *et al.* 2019). While on the other hand, this plugin does not fully support data analysis of GBS/RAD-seq/ddRAD-seq, which is known for a high rate of missing genotypes and genotyping errors (Andrews *et al.* 2016, Davey *et al.* 2013,

Table 1. Prediction model options in GENOMIC PREDICTION plugin

Option	Recommended situation			Implementation by	Reference
	Number of QTL	QTL effect	Number of markers		
RR-BLUP	Medium–Large	Additive	Medium–Large	R package "rrBLUP"	Endelman 2011
RHKS	Medium–Large	Additive & Nonadditive	Medium–Large	R package "rrBLUP"	Endelman 2011
LASSO	Small–Medium	Additive	Medium–Large	R package "glmnet"	Friedman *et al.* 2010, Tibshirani 1996
LM	Small	Additive	Small	R package "lm"	
LM with interaction	Small	Additive & Epistasis	Small	R package "lm"	

RR-BLUP, ridge regression best linear unbiased prediction; RKHS, reproducing kernel Hilbert spaces regression; LASSO, least absolute shrinkage and selection operator; LM, linear regression based on ordinary least squares.

Schweyen *et al.* 2014). This plugin does not support the function of error correction. Therefore, we recommend conducting error correction and imputation of missing data for GBS/RAD-seq/ddRAD-seq data using tools such as Beagle (Browning *et al.* 2018), LinkImpute (Money *et al.* 2015), ABHgenotypeR (Furuta *et al.* 2017), Genotype-Corrector (Miao *et al.* 2018) before using this plugin.

GENOMIC PREDICTION: construction of a model for phenotype prediction from genotype data

Finally, users can perform the genomic prediction analysis with the "GENOMIC PREDICTION" plugin. The plugin can predict phenotype from marker genotype data using genomic prediction models. Users must prepare marker genotype and phenotype data for training the prediction model (i.e., training data), and marker genotype data of the current breeding population for predicting the genotypic values of the selection candidates (i.e., test data). For the genotyping data, the output files derived from either the ABH or IMPUTATION plugin (AA/BB/AB genotype) and ABH formatted genotype file in a CSV format (R/qtl "csvr" format) are acceptable. Genotype datasets with missing data are also acceptable, and users can then set a threshold proportion of missing data to filter out for each marker. In the GENOMIC PREDICTION plugin, the user has a choice among four kinds of statistical models (**Table 1**) (Endelman 2011, Friedman *et al.* 2010, Tibshirani 1996). Each model has preferable conditions according to the number of QTLs, QTL effects, and number of markers (Desta and Ortiz 2014). Users can select an optimal model based on the prior knowledge about the target trait and the number of markers the input data includes (**Table 1**). For example, users can select 'LASSO' when users expect the target trait is controlled by just a few QTLs, whose positions are unknown, and users can prepare data of a number of markers to select the effective markers. If it is difficult to select an optimal prediction model, we recommend that users try all prediction models and select good genotypes based on the summary of the results. Using the model constructed from the training data, users can then predict the genetic value (i.e., performance explained by genotype) from the marker genotype data of test data. The output file provides predicted genetic values for each genotype in CSV format.

Discussion

The ability to handle the large amounts of accumulating NGS data remains a significant barrier for most breeders and experimental researchers to apply genome information to the breeding field. Despite the vast array of publicly available genetic analysis tools, these data are difficult to handle as the program input and output files are typically not compatible. Thus, the users need to convert the data files to suitable formats, which requires some prior knowledge on bioinformatics and statistics, necessitating further support from bioinformaticians or other experts. Indeed, GUI-based software has been developed for other research applications, including sequence homology searches and phylogenetic analyses, such as BLAST and Clustal W/X (Johnson *et al.* 2008, Larkin *et al.* 2007), which are now widely used even by non-bioinformatics experts. Thus, IonBreeders extends such user-friendly applications by providing a tool for researchers in the breeding sciences to analyze whole-genome data independently. By integrating the IonBreeders plugin to Torrent Suite in the Ion Torrent system (Thermo Fisher Scientific), users can quickly obtain a large amount of genotype data simultaneously and predict the phenotypic value by genomic prediction.

Author Contribution Statement

Conception or designed the work: EOT. Contributed analysis tool "IonBreeders_ABH" and "IonBreeders_IMPUTATION": EOT. Contributed analysis tool "IonBreeders_GENOMIC PREDICTION": SY. Contributed Perl script "VCF2ABH.pl": TT. Wrote the paper: EOT SY TT.

Acknowledgments

Literature Cited

Afgan, E., D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, B.A. Gruning *et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46: W537–W544.

Andrews, K.R., J.M. Good, M.R. Miller, G. Luikart and P.A. Hohenlohe (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. 17: 81–92.

Bernardo, R. (2008) Molecular markers and selection for complex trait in plants: learning from the last 20 years. Crop Sci. 48: 1649–1664.

Bhat, J.A., S. Ali, R.K. Salgotra, Z.A. Mir, S. Dutta, V. Jadon, A. Tyagi, M. Mushtaq, N. Jain, P.K. Singh *et al.* (2016) Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Front. Genet. 7: 221.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635.

Broman, K.W., H. Wu, S. Sen and G.A. Churchill (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Browning, B.L., Y. Zhou and S.R. Browning (2018) A one-penny imputed genome from next generation reference panels. Am. J. Hum. Genet. 103: 338–348.

Campbell, N.R., S.A. Harmon and S.R. Narum (2015) Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. Mol. Ecol. Resour. 15: 855–867.

Crossa, J., P. Perez, J. Hickey, J. Burgueno, L. Ornella, J. Ceron-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li *et al.* (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity (Edinb) 112: 48–60.

Cruaud, P., J.Y. Rasplus, L.J. Rodriguez and A. Cruaud (2017) High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. Sci. Rep. 7: 41948.

Davey, J.W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi and M.L. Blaxter (2013) Special features of RAD Sequencing data: implications for genotyping. Mol. Ecol. 22: 3151–3164.

Desta, Z.A. and R. Ortiz (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 19: 592–601.

Endelman, J.B. (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4: 250–255.

Friedman, J., T. Hastie and R. Tibshirani (2010) Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33: 1–22.

Furuta, T., M. Ashikari, K.K. Jena, K. Doi and S. Reuscher (2017) Adapting genotyping by sequencing for rice F_2 populations. G3 (Bethesda) 10: 881–893.

Hosoya, S., S. Hirase, K. Kikuchi, K. Nanjyo, Y. Nakamura, H. Kohno and M. Sano (2019) Random PCR-based genotyping by sequencing technology GRAS-Di (genotyping by random ampli-

con sequencing, direct) reveals genetic structure of mangrove fishes. Mol. Ecol. Resour. 19: 1153–1163.

Ishikawa, G., M. Saito, T. Tanaka, Y. Kayayose, H. Kanamori, K. Kurita and T. Nakamura (2018) An efficient approach for the development of genome-specific markers in allohexaploid wheat (*Triticum aestivum* L.) and its application in the construction of high-density linkage maps of the D genome. DNA Res. 25: 317–326.

Iwata, H. and S. Ninomiya (2006) Antmap: constructing genetic linkage maps using an ant colony optimization algorithm. Breed. Sci. 56: 371–377.

Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis and T.L. Madden (2008) NCBI BLAST: a better web interface. Nucleic Acids Res. 36: W5–W9.

Juanillas, V., A. Dereeper, N. Beaume, G. Droc, J. Dizon, J.R. Mendozo, J.P. Perdon, L. Mansueto, L. Triplett, J. Lang *et al.* (2019) Rice Galaxy: an open resource for plant science. Gigascience 8: giz028.

Kim, C., H. Guo, W. Kong, R. Chandnani, L.S. Shuang and A.H. Paterson (2016) Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Sci. 242: 14–22.

Kim, J., D.S. Ki, E.S. Lee, Y.K. Ah, W.B. Chae and S.S. Lee (2017) The construction of a Chinese cabbage marker-assisted backcrossing system using high-throughput genotyping technology. Hortic. Sci. Technol. 35: 232–242.

Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez *et al.* (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

Marubodee, R., E. Ogiso-Tanaka, T. Isemura, S. Chankaew, A. Kaga, K. Nalito, H. Ehara and N. Tomooka (2015) Construction of an SSR and RAD-marker based molecular linkage map of *Vigna vexillata* (L.) A. Rich. PLoS ONE 10: e0138942.

Matei, G., L.G. Woyann, A.S. Milioli, I. de Bem Oliveira, A.D. Zdziarski, R. Zanella, A.S.G. Coelho, T. Finatto and G. Benin (2018) Genomic selection in soybean: accuracy and time gain in relation to phenotypic selection. Mol. Breed. v38: 117.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Meuwissen, T., B. Hayes and M. Goddard (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Miao, C., J. Fang, D. Li, P. Liang, X. Zhang, J. Yang, J.C. Schnable and H. Tang (2018) Genotype-Corrector: Improved genotype calls for genetic mapping in F2 and RIL populations. Sci. Rep. 8: 10088.

Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G.Y. Zhong and S. Myles (2015) LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. G3 (Bethesda) 5: 2383–2390.

Onda, Y., K. Takahagi, M. Shimizu, K. Inoue and K. Mochida (2018) Multiplex PCR targeted amplicon sequencing (MTA-Seq): simple, flexible, and versatile SNP genotyping by highly multiplexed PCR amplicon sequencing. Front. Plant Sci. 9: 201.

Phan, N.T. and S.C. Sim (2017) Genomic tools and their implications for vegetable breeding. Hortic. Sci. Technol. 35: 149–164.

Schweyen, H., A. Rozenberg and F. Leese (2014) Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing

adapters. Biol. Bull. 227: 146–160.

Sharma, N. and R. Kaur (2014) MAPMAKER/EXP Version 3.0b: Though used since time immemorial still difficult for beginners to start with map construction for the genetic studies. J. Transl. Proteom. Res. 1: 17–27.

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, G. Atlin, J.L. Jannink and S.R. McCouch (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect oftrait 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: e1005350.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Series B Stat. Methodol. 58: 267–288.

Uga, Y., K. Sugimoto, S. Ogawa, J. Rane, M. Ishitani, N. Hara, Y. Kitomi, Y. Inukai, K. Ono, N. Kanno *et al.* (2013) Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. Nat. Genet. 45: 1097–1102.

Wang, H., X. Xu, F.G. Vieira, Y. Xiao, Z. Li, J. Wang, R. Nielsen and C. Chu (2016) The power of inbreeding: NGS-based GWAS of

rice reveals convergent evolution during rice domestication. Mol. Plant 9: 975–985.

Würschum, T., S. Abel and Y. Zhao (2014) Potential of genomic selection in rapeseed (*Brassica napus* L.) breeding. Plant Breed. 133: 45–51.

Yabe, S., T. Hara, M. Ueno, H. Enoki, T. Kimura, S. Nishimura, Y. Yasui, R. Ohsawa and H. Iwata (2018) Potential of genomic selection in mass selection breeding of an allogamous crop: an empirical study to increase yield of common buckwheat. Front. Plant Sci. 9: 276.

Yamamoto, E., H. Matsunaga, A. Onogi, K. Ohyama, H. Miyatake, T. Yamaguchi, H. Nunome, H. Iwata and H. Fukuoka (2017) Efficiency of genomic selection for breeding population design and phenotype prediction in tomato. Heredity (Edinb) 118: 202–209.

Zhao, K., M. Wright, J. Kimball, G. Eizenga, A. McClung, M. Kovach, W. Tyagi, M.L. Ali, C.W. Tung, A. Reynolds *et al.* (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS ONE 5: e10780.