ELSEVIER

COMPUTATIONAL
ANDSTRUCTURAL
BIOTECHNOLOGY
JOURNAL

journal homepage: www.elsevier.com/locate/csbj

dQTG.seq: A comprehensive R tool for detecting all types of QTLs using extreme phenotype individuals in bi-parental segregation populations

Pei Li, Liu-Qiong Wei, Yi-Fan Pan, Yuan-Ming Zhang [*]

College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

ARTICLE INFO

Article history:
Received 2 April 2022
Received in revised form 5 May 2022
Accepted 5 May 2022
Available online 14 May 2022

Keywords:
Bulked segregant analysis
Bi-parental segregation populations
Extremely over-dominant QTLs
Small-effect QTLs
R
dQTG-seq

ABSTRACT

Although methodologies and software packages for bulked segregant analysis (BSA) are well established, it is difficult to detect extremely over-dominant and small-effect genes for quantitative traits in F_2 population. To address this issue, we proposed a combinatorial strategy to identify all types of quantitative trait loci (QTLs) using extreme phenotype individuals in F_2. To popularize this strategy, we developed an R software package dQTG.seq v1.0.1. It has some features not found in other BSA software packages: 1) new (dQTG-seq1 and dQTG-seq2) and existing (G', deltaSNP, Euclidean distance (ED), and SmoothLOD) methods are available to identify all types of QTLs in bi-parental segregation populations, one data file with two BSA and three QTL-mapping data formats was inputted, and two *.csv files and one figure were outputted; 2) main smoothing methods (AIC, Window size, and Block) have been incorporated into each of the above-mentioned methods; 3) the threshold value of LOD score for significant QTLs is determined by permutation experiments. To save running time, vroom function was used to read the dataset, and parallel operation was used to estimate parameters. In real data analyses, users should select a suitable initial value of window size, depending on the species, and appropriate smoothing methods to obtain the best result. dQTG-seq2 detects more known loci and genes for rice grain number per panicle than composite interval mapping (CIM) and inclusive CIM, especially extremely over-dominant and small-effect genes. A handbook for our software package (https://cran.r-project.org/web/packages/dQTG.seq/index.html) has been provided in the supplemental materials for the users' convenience.

1. Introduction

Since the establishment of bulked segregant analysis (BSA) by Giovannoni et al. [1] and Michelmore et al. [2] in the early 1990s, it has been widely used to associate molecular markers with a trait of interest at a relatively low cost. More importantly, BSA has been integrated with whole-genome resequencing (BSA-seq) of two extreme pools of F_2 plants from the cross between a mutant and its wild type to identify recessive/dominant mutant genes via software package SHOREmap [3]. With the advancement of sequencing technology and reduction of sequencing costs, BSA methods are more and more widely used [4]. However, it is difficult to detect extremely over-dominant and small-effect genes using existing BSA software packages [5].

Many BSA software packages are available at present. In these software packages, several main statistical indicators have been used to identify loci for objective traits (Table A.1). First, SNP index in one extreme pool [3] and deltaSNP index between two extreme pools [6,7] are available to detect the loci using some BSA software packages, such as NGM [8], QTL-BSA [9], BSAseq [10], and block regression mapping (BRM) [11]. Then, G' [12] is used to detect the loci via the software package BSA4yeast [13]. We also noted that deltaSNP index and G' are simultaneously included in software packages QTLseqr [14] and PyBSASeq [15]. Next, Euclidean distance (ED) [16] is used to detect the loci via the BSA software package SIMM [17]. Finally, SmoothLOD score is adopted to identify the loci using the software package QTG-seq [18]. It should be noted that almost all the statistics in the above-mentioned BSA software packages are based on marker allelic frequencies in extreme pools. This results in the difficulty of detecting extremely over-dominant and small-effect genes. To address this issue, the numbers of mar-

Abbreviations: BSA, bulked segregant analysis; SNP, single nucleotide polymorphism; QTL, quantitative trait locus; QTG, quantitative trait gene; ED, Euclidean distance; BC, backcross; DH, doubled haploid; RIL, recombinant inbred line; CIM, composite interval mapping; GCIM, genome-wide composite interval mapping; ICIM, inclusive composite interval mapping.

[*] Corresponding author.

E-mail address: soyzhang@mail.hzau.edu.cn (Y.-M. Zhang).

https://doi.org/10.1016/j.csbj.2022.05.009
2001-0370/

ker alleles and genotypes in extreme pools have been used to construct a new statistic G_w, and a new combinatorial strategy has been proposed to detect all types of quantitative trait loci (QTLs) using extreme phenotype individuals in F_2 [5].

To popularize the above combinatorial strategy, here we developed a new R software package dQTG.seq to identify all types of QTLs for quantitative traits using extreme phenotype individuals in backcross (BC), recombinant inbred line (RIL), doubled haploid (DH) and F_2 populations. In this software package, new (dQTG-seq1 and dQTG-seq2) and existing (SmoothLOD, ED, G', and deltaSNP) BSA methods are available, whereas dQTG-seq2 is used to detect extremely over-dominant and small-effect QTLs in F_2 [5], and main smoothing methods (AIC [16], Window size [12] and Block [19]) are integrated with each of the above-mentioned BSA methods to optimize BSA results. In addition, we discussed the factors that affect the BSA results.

2. Materials and methods

2.1. Genetic mapping population

In F_2, if mixed DNA/RNA samples of each extreme pool are deeply sequenced (75–100-fold coverage) [18], the numbers of marker alleles are observed, and dQTG-seq1, SmoothLOD, G', deltaSNP, and ED are available. If users want to detect extremely over-dominant and small-effect genes, another reserved DNA/RNA sample of each F_2 plant with extreme phenotype is deeply sequenced, the numbers of marker alleles and genotypes are observed, and dQTG-seq2, G', SmoothLOD, deltaSNP, and ED are available. In BC, DH, and RIL populations, SmoothLOD, G', ED, and deltaSNP are available. The work flow diagram of this software package is shown in Fig. 1.

2.2. Rice real dataset for grain number per panicle

Real dataset for rice grain number per panicle in immortalized F_2 (IMF_2) in 1998 [20] was downloaded from https://www.pnas.org/doi/full/10.1073/pnas.1214141109 and re-analyzed using CIM [21] and ICIM [22], implemented by win QTL Cartographer v2.5 and QTL IciMapping v4.1, respectively. In the IMF_2, there were 278 individuals and 1619 bin markers available. All the IMF_2 individuals were sorted according to their phenotypic values, and 20% of extremely high and low individuals were selected to form high and low pools, respectively. These extreme individuals were analyzed using dQTG.seq, while all the IMF_2 individuals were analyzed using CIM and ICIM.

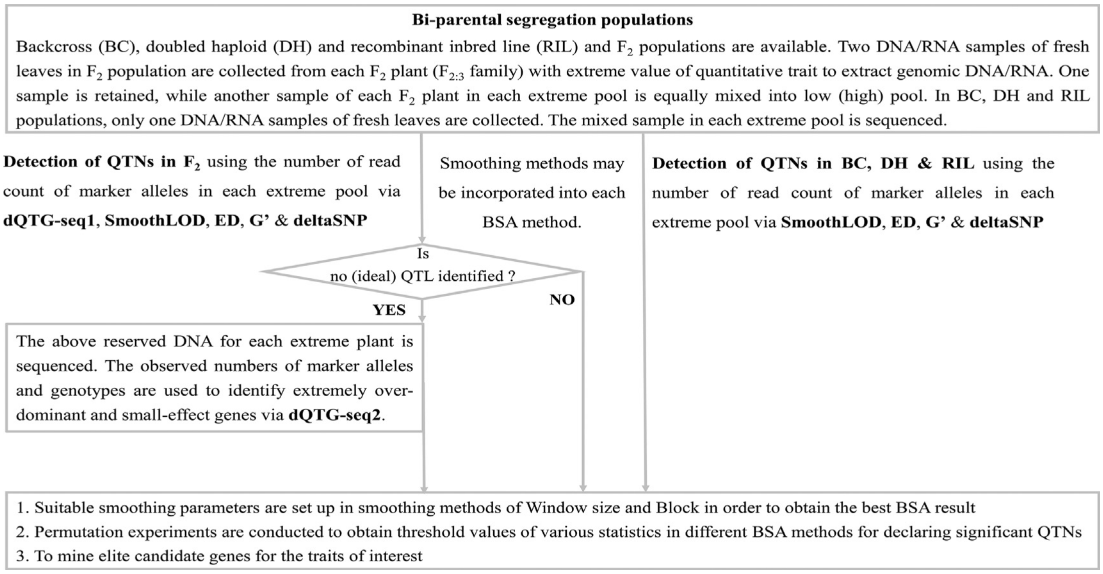
2.3. Development of the dQTG.seq software package

R software package dQTG.seq contains three modules: dataset inputting, parameter settings and plot drawing. Once the dataset inputting and parameter settings are finished, users may run the program and all the results will be saved in the directory set by the user. To reduce the running time, parallel calculation is adopted for parameter estimation, and function vroom is used to read the dataset; Parallel is used to detect the number of CPU cores on the current host and create a set of copies of R running in parallel and communicating over sockets; doParallel is used to register the parallel backend with the foreach package. Once the software package is successfully installed in R environment, users can write R script to analyze the datasets.

2.4. Preparation of input file

The software package has five types of input data formats: BSA, Extreme individual, ICIM, CIM, and genome-wide CIM (GCIM). If marker genotypes of each individual in each extreme pool are unknown, use the "BSA" format. If they are known, use the "Extreme individual" format. If marker genotypes of all the individuals in bi-parental segregation populations are known, use one of the "ICIM, CIM, and GCIM" data formats.

BSA format of dataset file. The input dataset file with BSA data format is shown in Fig. 2A and includes three sections of information. First, parameter information with a 10 × 2 matrix is located on the first ten lines. Then, physical map information with a (m + 1) × 3 matrix is under the block of parameter information and composed of marker name, chromosome number and marker physical position (bp) on genome, where m is the number of markers. Finally, the numbers of marker alleles with a (m + 1) × 4 matrix are located to the right of physical map information, where the numbers of marker alleles (A and a) in extremely low and high pools are indicated by AL and aL, and AH and aH, respectively,



Bi-parental segregation populations

Backcross (BC), doubled haploid (DH) and recombinant inbred line (RIL) and F_2 populations are available. Two DNA/RNA samples of fresh leaves in F_2 population are collected from each F_2 plant ($F_{2:3}$ family) with extreme value of quantitative trait to extract genomic DNA/RNA. One sample is retained, while another sample of each F_2 plant in each extreme pool is equally mixed into low (high) pool. In BC, DH and RIL populations, only one DNA/RNA samples of fresh leaves are collected. The mixed sample in each extreme pool is sequenced.

Detection of QTNs in F_2 using the number of read count of marker alleles in each extreme pool via **dQTG-seq1, SmoothLOD, ED, G' & deltaSNP**

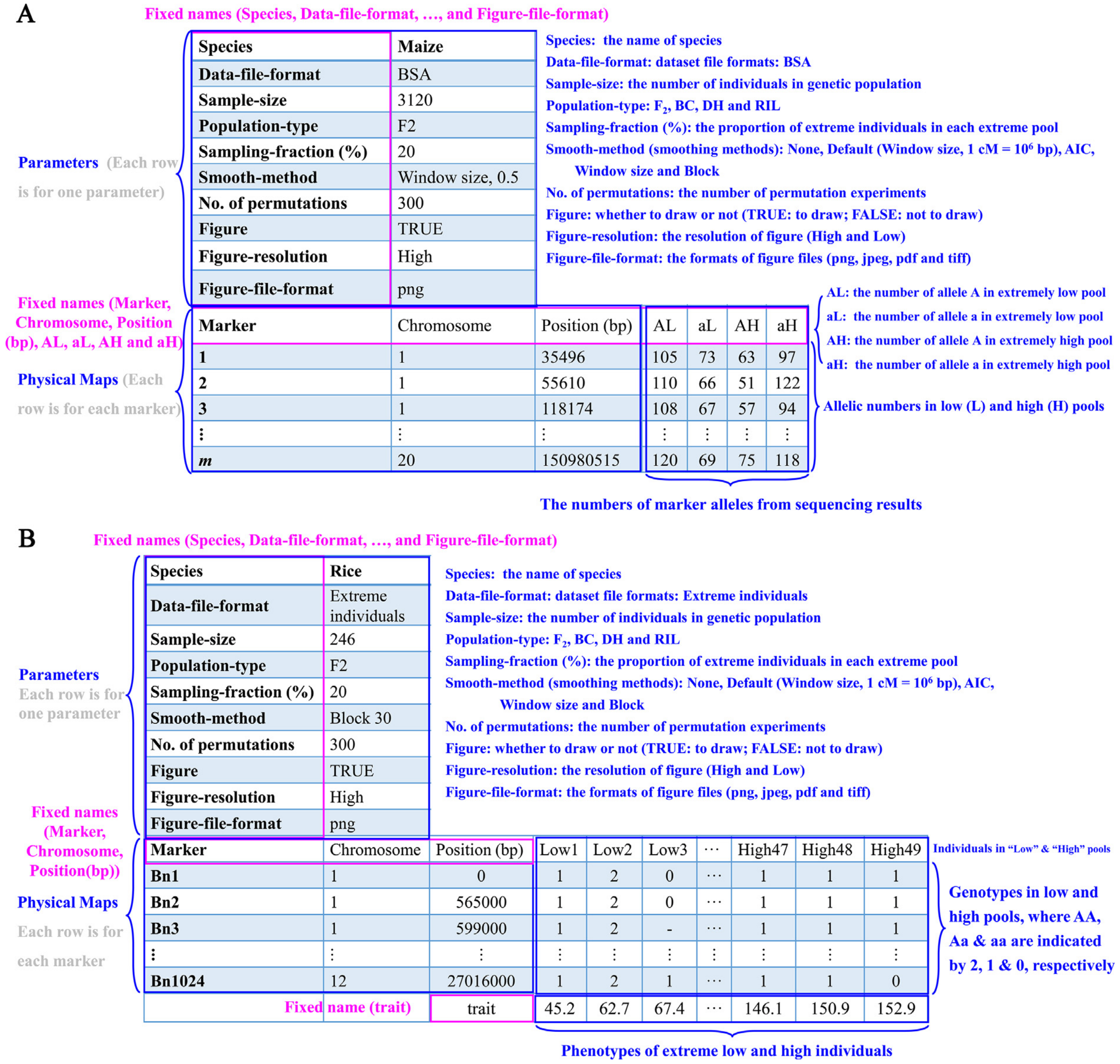
Smoothing methods may be incorporated into each BSA method.

Detection of QTNs in BC, DH & RIL using the number of read count of marker alleles in each extreme pool via **SmoothLOD, ED, G' & deltaSNP**

The above reserved DNA for each extreme plant is sequenced. The observed numbers of marker alleles and genotypes are used to identify extremely over-dominant and small-effect genes via **dQTG-seq2**.

1. Suitable smoothing parameters are set up in smoothing methods of Window size and Block in order to obtain the best BSA result
2. Permutation experiments are conducted to obtain threshold values of various statistics in different BSA methods for declaring significant QTNs
3. To mine elite candidate genes for the traits of interest

Fig. 1. A combinatorial strategy of mapping all types of QTLs for quantitative traits in bi-parental segregation populations via combination of BSA and whole-genome sequencing.


A
Fixed names (Species, Data-file-format, …, and Figure-file-format)
Parameters (Each row is for one parameter)
Fixed names (Marker, Chromosome, Position (bp), AL, aL, AH and aH)
Physical Maps (Each row is for each marker)
AL: the number of allele A in extremely low pool
aL: the number of allele a in extremely low pool
AH: the number of allele A in extremely high pool
aH: the number of allele a in extremely high pool
Allelic numbers in low (L) and high (H) pools
The numbers of marker alleles from sequencing results
B
Fixed names (Species, Data-file-format, …, and Figure-file-format)
Parameters Each row is for one parameter
Fixed names (Marker, Chromosome, Position(bp))
Physical Maps Each row is for each marker
Individuals in "Low" & "High" pools
Genotypes in low and high pools, where AA, Aa & aa are indicated by 2, 1 & 0, respectively
Fixed name (trait)
Phenotypes of extreme low and high individuals

Species	Maize
Data-file-format	BSA
Sample-size	3120
Population-type	F2
Sampling-fraction (%)	20
Smooth-method	Window size, 0.5
No. of permutations	300
Figure	TRUE
Figure-resolution	High
Figure-file-format	png

Species: the name of species
Data-file-format: dataset file formats: BSA
Sample-size: the number of individuals in genetic population
Population-type: F_2, BC, DH and RIL
Sampling-fraction (%): the proportion of extreme individuals in each extreme pool
Smooth-method (smoothing methods): None, Default (Window size, 1 cM = 10^6 bp), AIC, Window size and Block
No. of permutations: the number of permutation experiments
Figure: whether to draw or not (TRUE: to draw; FALSE: not to draw)
Figure-resolution: the resolution of figure (High and Low)
Figure-file-format: the formats of figure files (png, jpeg, pdf and tiff)

Marker	Chromosome	Position (bp)	AL	aL	AH	aH
1	1	35496	105	73	63	97
2	1	55610	110	66	51	122
3	1	118174	108	67	57	94
⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	20	150980515	120	69	75	118

Species	Rice
Data-file-format	Extreme individuals
Sample-size	246
Population-type	F2
Sampling-fraction (%)	20
Smooth-method	Block 30
No. of permutations	300
Figure	TRUE
Figure-resolution	High
Figure-file-format	png

Species: the name of species
Data-file-format: dataset file formats: Extreme individuals
Sample-size: the number of individuals in genetic population
Population-type: F_2, BC, DH and RIL
Sampling-fraction (%): the proportion of extreme individuals in each extreme pool
Smooth-method (smoothing methods): None, Default (Window size, 1 cM = 10^6 bp), AIC, Window size and Block
No. of permutations: the number of permutation experiments
Figure: whether to draw or not (TRUE: to draw; FALSE: not to draw)
Figure-resolution: the resolution of figure (High and Low)
Figure-file-format: the formats of figure files (png, jpeg, pdf and tiff)

Marker	Chromosome	Position (bp)	Low1	Low2	Low3	⋯	High47	High48	High49
Bn1	1	0	1	2	0	⋯	1	1	1
Bn2	1	565000	1	2	0	⋯	1	1	1
Bn3	1	599000	1	2	-	⋯	1	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋯	⋮	⋮	⋮
Bn1024	12	27016000	1	2	1	⋯	1	1	0
		trait	45.2	62.7	67.4	⋯	146.1	150.9	152.9

Fig. 2. BSA (A) and Extreme individual (B) formats for input file.

and the alleles of parent P_1 or reference genome are viewed as allele A.

Extreme individual format for dataset file. The input dataset file with Extreme individual data format is shown in Fig. 2B and includes three sections of information. The first two sections of information are the same as those in the BSA data format, except for "Data-file-format" that should be "Extreme individuals". The last one is marker genotypic information with a $(m + 1) \times 2n$ matrix, in which marker genotypes AA, Aa, and aa of each extreme individual are indicated, respectively, by 2, 1, and 0 in a column, "Low" and "High" respectively indicate extremely low and high individuals, and n is the number of individuals in each pool.

The input dataset file with CIM, ICIM and GCIM formats is shown in Fig. A.1 and can also be found in the users' instructions for software packages QTL Cartographer (*.csv), QTL IciMapping (*.xlsx) and QTL.gCIMapping (*.csv) [23], respectively.

2.5. Threshold values of various statistics via permutation experiments

The threshold values of various statistics for significant QTLs at the 0.05 probability level were determined by s permutation experiments [24], and the significant probability level was suggested as 0.10 for dQTGseq1 and 0.05 for other methods in Li et al. [5], where users may set the value of s, such as 1000, and change the probability level. In the permutation experiments, real mapping population may be simulated by the R package "qtl", and sample size and sampling fractions in low and high pools are the same as those in the real mapping population.

2.6. Installation of the software package

This software package can be installed in two ways: online installation and offline installation. For online installation, users can install directly using the command below:

install.packages ("dQTG.seq")

All the add-on software packages and dQTG.seq will be installed automatically.

For offline installation, users first open R GUI, select "Packages" — "Install package(s) from local files…", and then find and install the add-on software packages, which include the software packages: "data.table", "BB", "doParallel", "openxlsx", "qtl", "stringr", "vroom", and "writexl". Finally, users install the dQTG.seq software package, which was downloaded on the computer.

2.7. Implementation of the software package

Once the software package is installed, users can run the software package using two commands:

library (dQTG.seq)

dQTG.seq(dir="D:/users",filegen="D:/users/BSA.csv",chr="all",color = c("blue","red"), CLO = NULL)

If users want to restart this software package, the above two commands can be used as well. In the function *dQTG.seq()*, five parameters must be set up: 1) the path of output files, such as dir="D:/users"; 2) input file and its path, such as filegen="D:/users/BSA.csv"; 3) chromosome, such as chr="all" for all chromosomes and chr="c(n_1,n_2,n_3)" for chromosomes n_1, n_2, and n_3; 4) colors of smoothing lines in adjacent chromosomes, such as color = c("blue","red"); 5) parallel, for which CLO = NULL is the default setup (File A.1).

3. Results

3.1. The description for the result files

After running is finished, two result files (all_result.csv and significant_result.csv) and one plot file with the "png", "jpeg", "pdf", and "tiff" formats will be outputted in the output path.

In the "all_result.csv" file, there are thirteen columns for F_2 population and eleven columns for BC, DH, and RIL populations. In F_2 population, the first three columns "Marker", "Chromosome", and "Position" show marker name, chromosome on which the marker resides, and marker position (bp) on genome, respectively; columns 4 to 8 show the estimates of statistics G_w, LOD, G, deltaSNP, and ED, respectively; columns 9 to 13 show smoothing estimates of five statistics Smooth_G_w, Smooth_LOD, G', Smooth_deltaSNP, and Smooth_ED, respectively. In BC, DH, and RIL populations, eleven columns, except for the G_w and Smooth_G_w columns, are consistent with those in F_2 population.

In the "significant_result.csv" file, there are five sheets; each sheet shows all the significant QTNs identified by one method. In each sheet, the first six columns show "Marker", "Chromosome", "Position (bp)", estimate of statistic, smooth estimate of statistic, and critical value of the statistic for significant QTN, respectively.

In the output plot, there are five sub-plots in png, jpeg, pdf and tiff formats (File A.1). In each sub-plot from one method, users may modify some parameters, such as colors.

3.2. Real data analysis for rice grain number per panicle

The dataset of rice grain number per panicle from Zhou et al. [20] was re-analyzed using the BSA (dQTG-seq2, SmoothLOD, G', deltaSNP, and ED) and QTL mapping (CIM and ICIM) methods. The BSA methods were implemented using the new software package in this study. The results are listed in Tables A.2-A.3. As a result, 42, 12, 8, 26, 25, 13, and 7 significant QTLs were identified by the dQTG-seq2, SmoothLOD, G', deltaSNP, ED, CIM, and ICIM methods, respectively (Fig. 3; Table A.2). Among these QTLs, 21, 5, 5, 12, 12, 5, and 4 known genes were detected by the above-mentioned

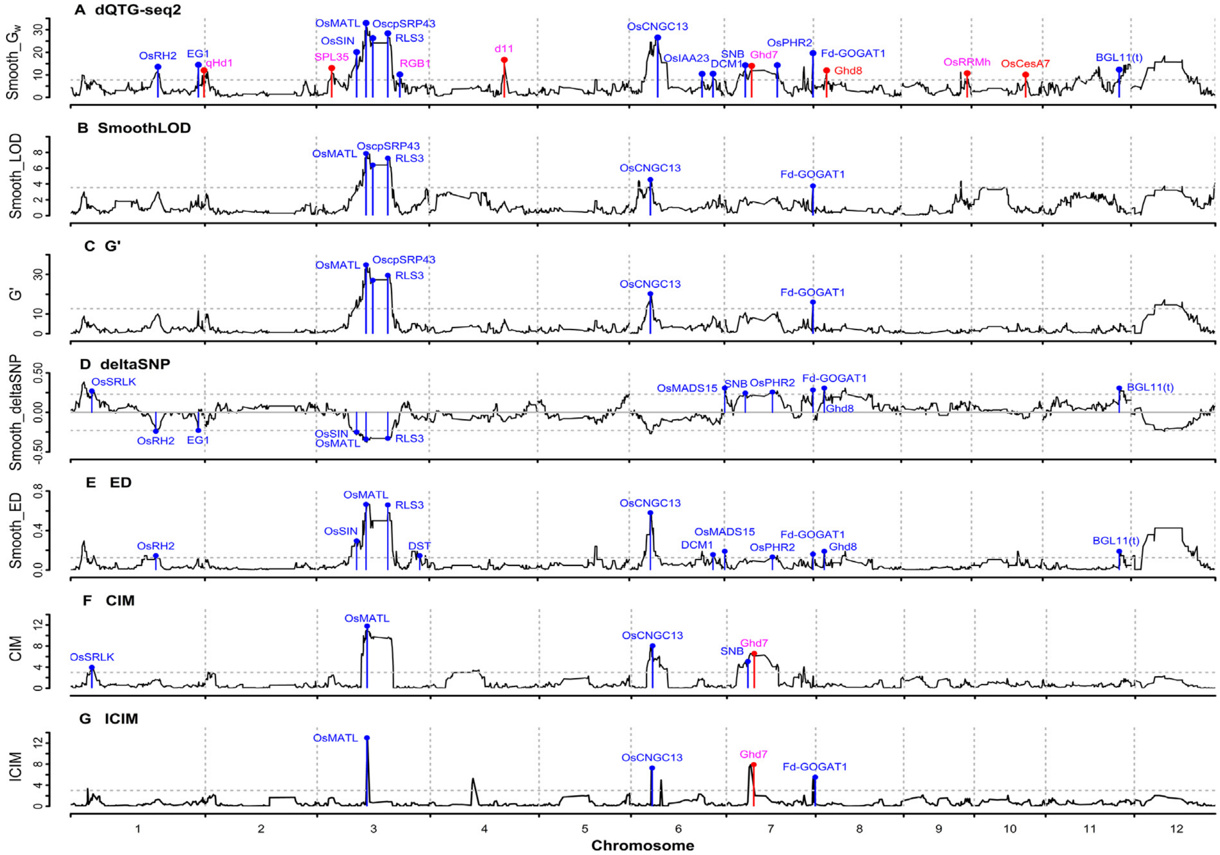


Fig. 3. Previously reported genes for rice grain number per panicle in immortalized F_2 using the dQTG-seq2 (A), SmoothLOD (B), G' (C), deltaSNP (D), ED (E), composite interval mapping (CIM, F) and inclusive CIM (ICIM, G) methods. Horizontal dotted lines indicate thresholds of significant QTLs. Various statistics of genome-wide scanning using new and existing methods are indicated by black curves. The genes with absolute dominant ratio $|d/a| < 2.0$, small-effects, and $|d/a| \geq 2.0$ are indicated by blue, pink, and red colors, respectively. If $|d/a| \geq 2.0$ and its size is small, the gene name is in pink color and its corresponding solid line is in red color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

methods, respectively (Table A.3). Among these known genes, dQTG-seq2 identified 7 known genes with absolute dominance ratio |*d*/*a*| > 2.0 and 6 known genes with allelic frequency difference |AFD| < 0.15, QTL mapping (CIM and ICIM) methods identified one known gene *Ghd7* with large |*d*/*a*| and small |AFD|, and other BSA methods identified no known genes with large |*d*/*a*| and small |AFD| (Table A.3). These results indicate that dQTG-seq2 identifies more significant QTLs and known genes for grain number per panicle than existing BSA and QTL mapping methods, especially the QTLs and known genes with large |*d*/*a*| and small |AFD|. The conclusion is consistent with that of Li et al. [5].

4. Discussion

Significant progress has been made in BSA software package development in this study. First, the new R software package dQTG.seq using the dQTG-seq2 method had higher power in detecting extremely over-dominant and small-effect QTLs and genes for quantitative traits in F_2 than existing BSA software packages [5] (Table A.1). This is mainly due to the utilization of the numbers of marker alleles and genotypes in two extreme pools in the dQTG-seq2 method and the utilization of the numbers of marker alleles in existing BSA software packages. Meanwhile, there are more bi-parental segregation populations (F_2, BC, DH, and RIL) and more BSA methodologies (dQTG-seq1, dQTG-seq2, Smooth-LOD, G', deltaSNP, and ED) available in this new software package as compared with existing BSA software packages frequently used for one method in one segregation population. Then, three main BSA smoothing methods, AIC [16], Window size [12], and Block [19], have been incorporated into each BSA method in order to optimize the results. In previous BSA software packages, each approach has its own specific smoothing method. Finally, permutation experiments are used to determine threshold values of various BSA statistics. This overcomes the subjectivity in the determination of significant QTLs. Thus, the new software package has broad application prospects.

In BSA, it is common for no significant QTLs and genes to be identified. As we know, many factors affect BSA results, such as population type, sample and QTL size, sampling fraction, and sequencing depth. Related discussions can be found in several articles [4,5,25,26], from which suitable sampling plans have been summarized in Table A.4. For example, F_2 is better than BC, DH, and RIL. This is owing to its simple construction and good mapping results [4,25]. However, there are frequently large experimental errors for phenotypic observations of quantitative traits in F_2 plants, especially in maize and cotton. To address this issue, the DNA / RNA samples of F_2 plants are used to obtain their genotypes, and the average of $F_{2:3}$ families is used to measure the phenotype of F_2 plant [27]. Meanwhile, F_2 population is a temporary segregation population. To overcome this issue, immortalized F_2 (IMF_2) population is recommended [28,29]. More importantly, a new statistic G_w has been proposed in F_2 to identify all types of QTLs and genes, especially for extremely over-dominant and small-effect genes [5].

In smoothing methods, parameter of window size for the Window size method [12] and the number of markers in a block for the Block method [19] affect the BSA results [30]. In our software package, window size varies across various species, and the initial value of window size is set up as the ratio of the genome length (Mb) to genetic map length (cM) in more than ten species (File A.1), while the block depends on marker density and the initial value for the number of markers in a block is set up as 10. In application, users may adjust these setups based on the dataset in order to obtain the best result.

The widely-used probability level of significance is 0.05 in statistics. Thus, this level should be adopted in the dQTG-seq2 and existing BSA methods. As we know, the numbers of read counts of marker genotypes in extreme phenotype pools in the dQTG-seq1 method are predicted from the numbers of read counts of marker alleles, and the predicted values have residual error. Thus, the 0.10 level was suggested in Li et al. [5].

Author statement

Y.-M.Z. conceived and designed the study. P.L. and L.-Q.W. wrote the codes. P.L., L.-Q.W., Y.-F.P., and Y.-M.Z. performed the data analyses. Y.-M.Z. and P.L. wrote the draft, and revised the manuscript. All authors reviewed the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32070557 and 31871242), the Fundamental Research Funds for the Central Universities (2662020ZKPY017), and Huazhong Agricultural University Scientific & Technological Self-Innovation Foundation (2014RC020).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.05.009.

References

[1] Giovannoni JJ, Wing RA, Ganal MW, et al. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res 1991;19(23):6553–8.
[2] Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis, a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A 1991;88(21):9828–32.
[3] Schneeberger K, Ossowski S, Lanz C, et al. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods 2009;6:550–1.
[4] Li Z, Xu Y. Bulk segregation analysis in NGS era: a review for its teenage years. Plant J 2022;109:1355–74.
[5] Li P, Li G, Zhang YW, et al. A combinatorial strategy to identify various types of QTLs for quantitative traits using extreme phenotype individuals in F_2. Plant Commun 2022;3(3): https://doi.org/10.1016/j.xplc.2022.100319100319.
[6] Fekih R, Takagi H, Tamiru M, et al. MutMap+: genetic mapping and mutant identification without crossing in rice. PLoS ONE 2013;8:e68529.
[7] Takagi H, Abe A, Yoshida K, et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 2013;74:174–83.
[8] Austin RS, Vidaurre D, Stamatiou G, et al. Next-generation mapping of *Arabidopsis* genes. Plant J 2011;67:15–725.
[9] Wu S, Qiu J, Gao Q. QTL-BSA: A bulked segregant analysis and visualization pipeline for QTL-seq. Interdiscip Sci 2019;11:730–7.
[10] Wang L, Lu Z, Regulski M, et al. BSAseq: an interactive and integrated web-based workflow for identification of causal mutations in bulked F_2 populations. Bioinformatics 2021;37:382–7.
[11] Huang L, Tang W, Bu S, et al. BRM: a statistical method for QTL mapping based on bulked segregant analysis by deep sequencing. Bioinformatics 2020;36:2150–6.
[12] Magwene PM, Willis JH, Kelly JK. The statistics of bulk segregant analysis using next generation sequencing. PLoS Comput Biol 2011;7:e1002255.
[13] Zhang Z, Jung PP, Groues V, et al. BSA4Yeast: Web-based quantitative trait locus linkage analysis and bulk segregant analysis of yeast sequencing data. GigaScience 2019;8:giz060.
[14] Mansfeld BN, Grumet R. QTLseqr: An R Package for bulk segregant analysis with next-generation sequencing. Plant Genome 2018;11: doi: 10.3835/plantgenome2018.01.0006.
[15] Zhang J, Panthee DR. PyBSASeq: a simple and effective algorithm for bulked segregant analysis with whole-genome sequencing data. BMC Bioinformatics 2020;21:99.

[16] Hill JT, Demarest BL, Bisgrove BW, et al. MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. Genome Res 2013;23:687–97.

[17] Yan W, Chen Z, Lu J, et al. Simultaneous identification of multiple causal mutations in rice. Front Plant Sci 2017;7:2055.

[18] Zhang H, Wang X, Pan Q, et al. QTG-Seq accelerates QTL fine mapping through QTL partitioning and whole-genome sequencing of bulked segregant samples. Mol Plant 2019;12:426–37.

[19] Pool JE. Genetic mapping by bulk segregant analysis in *Drosophila*: experimental design and simulation-based inference. Genetics 2016;204:1295–306.

[20] Zhou G, Chen Y, Yao W, et al. Genetic composition of yield heterosis in an elite rice hybrid. Proc Natl Acad Sci U S A 2012;109:15847–52.

[21] Zeng ZB. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci U S A 1993;90(23):10972–6.

[22] Li H, Ye G, Wang J. A modified algorithm for the improvement of composite interval mapping. Genetics 2007;175:361–74.

[23] Zhang YW, Wen YJ, Dunwell JM, et al. QTL.gCIMapping.GUI v2.0: An R software for detecting small-effect and linked QTLs for quantitative traits in bi-parental segregation populations. Comput Struct Biotechnol J 2020;18:59–65.

[24] Pitman EJG. Significance tests which may be applied to samples from any populations. Supplement J Royal Stat Society 1937;4:119–30.

[25] Zou C, Wang P, Xu Y. Bulked sample analysis in genetics, genomics and crop improvement. Plant Biotechnol J 2016;14:1941–55.

[26] James GV, Patel V, Nordstrom KJV, et al. User guide for mapping-by-sequencing in *Arabidopsis*. Genome Biol 2013;14(6):R61.

[27] Zhang YM, Xu S. Mapping quantitative trait loci in F_2 incorporating phenotypes of F_3 progeny. Genetics 2004;166:1981–93.

[28] Gardiner JM, Coe EH, Melia-Hancock S, Hoisington DA, Chao S. Development of a core RFLP map in maize using an immortalized F_2 population. Genetics 1993;134:917–30.

[29] Hua J, Xing Y, Wu W, Xu C, Sun X, Yu S, et al. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci U S A 2003;100:2574–9.

[30] Shen F, Huang ZY, Zhang BG, et al. Mapping gene markers for apple fruit ring rot disease resistance using a multi-omics approach. G3 Genes|Genomes|Genetics 2019;G3(9):1663–78.