**BSA**

**Genomic DNA Library preparation and sequencing**

DNA was extracted from the xxx tissue using CTAB method. The quality of isolated genomic DNA was verified by using these two methods in combination: (1) DNA degradation and contamination were monitored on 1% agarose gels. (2) DNA concentration was measured by ND-2000 (NanoDrop Technologies). Only high-quality DNA sample (OD260/280 = 1.8~2.0, OD260/230 ≥ 2.0) was used to construct sequencing library.

A total amount of 0.5 μg DNA per sample was used as input material for the DNA library preparations. The sequencing library was generated using Truseq Nano DNA HT Sample Prep Kit (Illumina USA) following the manufacturer’s recommendations and index codes were added to each sample. Briefly, the genomic DNA sample was fragmented by sonication to a size of 350 bp. Then DNA fragments were end-polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing, followed by further PCR amplification. After PCR products were purified (AMPure XP system), libraries were analyzed for size distribution by Agilent 2100 Bioanalyzer and quantified by real-time PCR (3nM). Paired-end DNA-seq sequencing library was sequenced with the Illumina NovaSeq system at Shanghai Majorbio Bio-pharm Technology Co.,Ltd.

**Variant Discovery**

Raw reads of low quality (mean phred score < 20), including reads containing adapter contamination and unrecognizable nucleotide (N base > 10) were trimmed or discarded by the software Fastp (Chen et al. 2018).

Reads after trimming were mapped to the reference genome of xxx() using BWA-MEME software (Youngmok and Dongsu 2022) under default mapping parameters.

As the modified GATK Best Bractise (McKenna et al. 2010), the alignment bam files were sorted by samtools (Li et al. 2009) and PCR duplicates marked by MarkDuplicated. After doing the base quality recalibration, the germline variants calling which contained SNPs and InDels across all samples was using the Haplotyper and Gvcftyper programs in Sentieon genomics tools (Freed et al. 2017). Variants were filtered using standard hard filtering parameters according to GATK Best Practices pipeline.

All of the variants were annotated using SnpEff (Cingolani et al. 2012). SNPs and InDels were categorized based on their positions on the chromosome (including intergenic regions, exons, introns, splicing sites, untranslated regions, and 1-kb upstream and downstream regions) and on their effects (including missense, start codon gain or loss, stop codon gain or loss and splicing mutations).

**Bulk Segregation Analysis**

We used six resequencing-based methods to perform BSA analysis, contains the index-slid method (Takagi et al. 2013), the index-loess method, the Euclidean distance (ED) algorithm (Hill et al. 2013), the G′value method (Magwene, Willis, and Kelly 2011) and two other BSA methods DL and K, which were driven by deep learning and least square method (Zhao et al. 2022), respectively.

**index-slid and index-loess**

After Variant Discovery, SNP-index was calculated for all the SNP positions. The sliding window strategy was applied with 1 Mb window size and 10 kb step size to reduce the noise. We calculated the average SNP-index of the SNPs located in the window for further analysis, and windows with <10 SNPs were skipped. Only the window with averaged SNP-index above the threshold of Bootstrap P-value ≤ 0.05 was considered as the candidate region.

Except for sliding windows, we can also reduce the noise in the plot using locally weighted smoothing, or lowess (Locally Weighted Scatterplot Smoothing) as called. Only the candidate region with the loess-fitted values of the markers above the threshold of the 99.9% confidence interval was considered.

**ED**

We chose to measure allele segregation using Euclidean distance (ED), as a metric that does not require parental strain information and is resistant to noise. The sliding window strategy was applied with 1 Mb window size and 10 kb step size to reduce the noise. We calculated the average ED of the SNPs located in the window for further analysis, and windows with <10 SNPs were skipped. Only the window with averaged ED above the threshold of the 99.9% confidence interval was considered as the candidate region.

**G′value**

We calculated G’value for each SNP: the data is four allele counts that can be summarized in a 2\*2 table to perform G-test. The non-overlapped sliding window strategy was applied with 1 Mb window size to reduce the noise. We calculated the average G′value of the SNPs located in the window for further analysis, and windows with <10 SNPs were skipped. Only the window with P-value ≤ 0.05 was considered as the candidate region.

**DeepBSA**

We use DeepBSA software to calculate DL and K for QTL mapping. DeepBSA is compatible with a variable number of bulked pools and performed well with various simulated and real datasets in both animals and plants. The sliding window strategy was applied with 1 Mb window size and 10 kb step size to reduce the noise. We calculated the average DL and K of the SNPs located in the window for further analysis, and windows with <10 SNPs were skipped. Only the window with averaged DL and K above the threshold of the 99.9% confidence interval was considered as the candidate region.

参考文献：

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics (Oxford, England)* 34 (17): i884–90. https://doi.org/10.1093/bioinformatics/bty560.

Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. “A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff.” *Fly* 6 (2): 80–92. https://doi.org/10.4161/fly.19695.

Freed, Donald, Rafael Aldana, Jessica A. Weber, and Jeremy S. Edwards. 2017. “The Sentieon Genomics Tools - A Fast and Accurate Solution to Variant Calling from next-Generation Sequence Data.” bioRxiv. https://doi.org/10.1101/115717.

Hill, Jonathon T., Bradley L. Demarest, Brent W. Bisgrove, Bushra Gorsi, Yi-Chu Su, and H. Joseph Yost. 2013. “MMAPPR: Mutation Mapping Analysis Pipeline for Pooled RNA-Seq.” *Genome Research* 23 (4): 687–97. https://doi.org/10.1101/gr.146936.112.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Magwene, Paul M., John H. Willis, and John K. Kelly. 2011. “The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing.” *PLOS Computational Biology* 7 (11): e1002255. https://doi.org/10.1371/journal.pcbi.1002255.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.” *Genome Research* 20 (9): 1297–1303. https://doi.org/10.1101/gr.107524.110.

Takagi, Hiroki, Akira Abe, Kentaro Yoshida, Shunichi Kosugi, Satoshi Natsume, Chikako Mitsuoka, Aiko Uemura, et al. 2013. “QTL-Seq: Rapid Mapping of Quantitative Trait Loci in Rice by Whole Genome Resequencing of DNA from Two Bulked Populations.” *The Plant Journal* 74 (1): 174–83. https://doi.org/10.1111/tpj.12105.

Youngmok, Jung, and Han Dongsu. 2022. “BWA-MEME: BWA-MEM Emulated with a Machine Learning Approach.” *Bioinformatics (Oxford, England)* 38 (9). https://doi.org/10.1093/bioinformatics/btac137.

Zhao, Li, Chen Xiaoxuan, Shi Shaoqiang, Zhang Hongwei, Wang Xi, Chen Hong, Li Weifu, and Li Lin. 2022. “DeepBSA: A Deep-Learning Algorithm Improves Bulked Segregant Analysis for Dissecting Complex Traits.” *Molecular Plant* 15 (9). https://doi.org/10.1016/j.molp.2022.08.004.