

Genome-wide Association Studies Reveal Similar Genetic Architecture with Shared and Unique QTL for Bacterial Cold Water Disease Resistance in Two Rainbow Trout Breeding Populations

Roger L. Vallejo,^{*} Guangtu Gao,^{*} Sixin Liu,^{*} Breno O. Fragomeni,[†] Alvaro G. Hernandez,[‡]
James E. Parsons,[§] Kyle E. Martin,[§] Jason P. Evenhuis,^{*} Timothy J. Welch,^{*} Timothy D.
Leeds,^{*} Gregory D. Wiens,^{*} and Yniv Palti^{*,1}

^{*}National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV 25430

[†]Animal and Dairy Science Department, University of Georgia, Athens, GA 30602

[‡]High-Throughput Sequencing and Genotyping Unit, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[§]Troutlodge, Inc., P.O. Box 1290, Sumner, WA 98390

Running title:

GWAS for BCWD resistance in Rainbow Trout

Key Words:

Aquaculture

Bacterial cold water disease

Genome-wide association study

Quantitative trait loci

Rainbow trout

¹Corresponding author:

Yniv Palti

National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, U.S.

Department of Agriculture, 11861 Leetown Rd., Kearneysville, WV 25430, USA

Phone: 304-724-8340

Email: yniv.palti@ars.usda.gov

ABSTRACT

Bacterial cold water disease (BCWD) causes significant mortality and economic losses in salmonid aquaculture. In previous studies, we identified moderate-large effect QTL for BCWD resistance in rainbow trout (*Oncorhynchus mykiss*). However, the recent availability of a 57K SNP array and a genome physical map have enabled us to conduct genome-wide association studies (GWAS) that overcome several experimental limitations from our previous work. In the current study, we conducted GWAS for BCWD resistance in two rainbow trout breeding populations using two genotyping platforms, the 57K Affymetrix SNP array and restriction-associated DNA (RAD) sequencing. Overall, we identified 14 moderate-large effect QTL that explained up to 60.8% of the genetic variance in one of the two populations and 27.7% in the other. Four of these QTL were found in both populations explaining a substantial proportion of the variance, although major differences were also detected between the two populations. Our results confirm that BCWD resistance is controlled by the oligogenic inheritance of few moderate-large effect loci and a large-unknown number of loci each having a small effect on BCWD resistance. We detected differences in QTL number and genome location between two GWAS models (weighted single-step GBLUP and Bayes B), which highlights the utility of using different models to uncover QTL. The RAD-SNPs detected a greater number of QTL than the 57K SNP array in one population, suggesting that the RAD-SNPs may uncover polymorphisms that are more unique and informative for the specific population in which they were discovered.

INTRODUCTION

Bacterial cold water disease (BCWD) causes significant mortality and economic losses in salmonid aquaculture (NEMATOLLAHI *et al.* 2003; BARNES AND BROWN 2011). The etiological agent of BCWD is a gram-negative bacterium, *Flavobacterium psychrophilum* (*Fp*), and current methods to control BCWD outbreaks are limited. At the National Center for Cool and Cold Water Aquaculture (NCCCWA), we have developed a selective breeding program to increase rainbow trout genetic resistance against BCWD and have shown that BCWD resistance is a moderately heritable trait that responds to selection (LEEDS *et al.* 2010). Furthermore, we revealed complex genetic architecture of BCWD resistance (VALLEJO *et al.* 2010) and identified several moderate-large effect quantitative trait loci (QTL) for this trait in the NCCCWA odd- and even-year rainbow trout selective-breeding populations (WIENS *et al.* 2013; VALLEJO *et al.* 2014a; LIU *et al.* 2015b; PALT *et al.* 2015b). While those loci can be fine mapped to identify positional candidate genes, the complex genetic architecture of BCWD resistance and high genetic variability discovered in past studies (VALLEJO *et al.* 2014a) suggest that whole genome-enabled selection is more effective for improving genetic resistance against BCWD in rainbow trout aquaculture, and we were able to empirically demonstrate that whole genome-enabled selection can double the accuracy of predicted genetic merit of potential breeders compared to traditional family-based selection (VALLEJO *et al.* 2017).

For agricultural livestock species, single nucleotide polymorphism (SNP) chips have been the platform of choice for whole genome genotyping of at least 50K SNPs (MATUKUMALLI *et al.* 2009; RAMOS *et al.* 2009; GROENEN *et al.* 2011); including the recently developed 57K

SNP chip for rainbow trout (PALTI *et al.* 2015a). However, sequencing-by-genotyping methods that do not require *a priori* marker discovery or a reference genome sequence and are capable of simultaneous marker discovery and genotyping in many individuals were developed for genetic analyses (DAVEY *et al.* 2011). One such technique is restriction-site-associated DNA (RAD) sequencing (MILLER *et al.* 2007; BAIRD *et al.* 2008). In recent years, the method of RAD genotyping by sequencing has been widely used in salmonid species for SNP discovery, generating linkage maps, QTL mapping, genome-wide association studies (GWAS) and for evaluating genome-enabled selection (HECHT *et al.* 2012; HOUSTON *et al.* 2012; MILLER *et al.* 2012; HALE *et al.* 2013; HECHT *et al.* 2013; NARUM *et al.* 2013; BRIEUC *et al.* 2014; CAMPBELL *et al.* 2014; GONEN *et al.* 2014; HOUSTON *et al.* 2014; PALTI *et al.* 2014; LIU *et al.* 2015a; LIU *et al.* 2015b; PALTI *et al.* 2015b; VALLEJO *et al.* 2016).

Several moderate-large effect QTL associated with BCWD resistance have been identified on 24 of the 29 rainbow trout chromosomes using linkage analysis mapping (JOHNSON *et al.* 2008; WIENS *et al.* 2013; PALTI *et al.* 2014; QUILLET *et al.* 2014; VALLEJO *et al.* 2014a) and GWAS methods (CAMPBELL *et al.* 2014; LIU *et al.* 2015b; PALTI *et al.* 2015b). However, these previous studies had limitations on QTL detection. First, most of the reported genome-wide association analyses have tested one SNP at a time using single-regression or mixed linear models with a fixed SNP effect along with a random polygenic effect to capture the effects of all other genes. Although those studies have been successful in detecting associations, those associations typically explain only a small fraction of the trait genetic variance (VISSCHER *et al.* 2010). Conversely, in GWAS analysis using whole-genome selection models that simultaneously fit all SNPs as random effects, the SNPs jointly explain a larger proportion of the genetic

variance which highlights the utility of using multiple-regression GWAS models with the SNPs joined into genomic windows for accurate QTL mapping (HAYES *et al.* 2010; FAN *et al.* 2011; ONTERU *et al.* 2011). Second, most of the reported QTL for BCWD resistance were identified using linkage-based methods and GWAS was performed within individual segregating families with relatively small sample size and low detection power. Consequently about half of the reported findings from such QTL mapping studies are expected to represent false positives (HIGGINSON AND MUNAFO 2016). Third, the previous studies used lower density genotyping platforms and did not have a reference genome physical map for accurate prediction of the order and physical proximity of the genetic markers. Furthermore, to our knowledge, in the current study we are using for the first time SNP genotype data with the genome physical map coordinates for GWAS in rainbow trout.

There is ambiguity on the best computational algorithm when using multiple-regression based models in genomic selection (GS) and GWAS experiments. The genetic architecture of the trait and the population structure can have a significant impact on the accuracy of the genomic predictions and estimated marker effects. Therefore, it is important to compare the performance of the best competing algorithms on GS and GWAS when evaluating a trait with complex inheritance for the first time in a population. This will allow effective discovery of QTL underlying the genetic basis of the complex trait and control the type I error rate, which is often high in genome-wide discovery experiments.

In GWAS with models that fit all SNPs simultaneously, the genomic BLUP (GBLUP) method assumes a polygenic architecture of the trait and uses all SNP data to estimate the

genomic relationship (G) matrix. In contrast, the Bayesian variable selection method assumes that the genetic variance is explained by a reduced number of markers with small-moderate or large effects (HABIER *et al.* 2007; HAYES *et al.* 2009; DE LOS CAMPOS *et al.* 2013; FERNANDO AND GARRICK 2013; HOWARD *et al.* 2015). Based on this assumption, GBLUP is expected not to perform as well as Bayesian variable selection models when the trait is controlled by few moderate-to-large effect QTL. The GBLUP method was modified into the single-step GBLUP (ssGBLUP) method which allows combining the pedigree (A) and genomic-derived relationships into an H relationship matrix (AGUILAR *et al.* 2010; LEGARRA *et al.* 2014), and to the weighted single-step GBLUP (wssGBLUP) method which emulates the Bayesian variable selection models by fitting in the multiple regression model selected SNPs that explain moderate-large fraction of the trait genetic variation (WANG *et al.* 2012).

The recent development of the 57K SNP array (PALTÍ *et al.* 2015a), a dense genetic linkage map with 47,939 SNP markers (GONZALEZ-PENA *et al.* 2016), and the release of the improved rainbow trout reference genome (GenBank assembly Accession GCA_002163495) have provided the needed tools for performing GWAS to identify genomic regions associated with BCWD resistance in rainbow trout. The main objectives of this study were to (1) identify and validate QTL associated with BCWD resistance in two commercially-relevant rainbow trout breeding populations; (2) characterize the genetic architecture of rainbow trout resistance to BCWD; (3) compare the QTL mapping efficiency and determine whether the Chip-SNP and RAD-SNP genotyping platforms detect the same QTL; and (4) compare the QTL mapping efficiency of two widely-used multiple-regression GWAS models.

MATERIALS AND METHODS

Ethics statement

Protocols for this study were reviewed and approved by the NCCCWA Institutional Animal Care and Use Committee (Kearneysville, WV).

Rainbow trout rearing and BCWD challenge

Details of the 21-day BCWD challenge have been reported elsewhere (SILVERSTEIN *et al.* 2009; LEEDS *et al.* 2010). Mortalities were removed and recorded daily and fin clipped. Fish that survived the challenge were euthanized in a lethal dose of Tricaine methane sulfonate (Tricaine-S, Western Chemical, Inc., Ferndale, WA) and fin clipped. Fin clips from all fish (mortalities and survivors) were individually kept in 95% ethanol until DNA was extracted using established protocols (PALT *et al.* 2006).

Rainbow trout populations used in GWAS

Fish used in this study were sampled from two populations, and all analyses were performed separately by population. The first sample included fish with genotypes and phenotypes from 10 full-sib (FS) families sampled from a total of 71 pedigreed FS families with phenotype data from year-class (YC) 2005 of the NCCCWA BCWD resistant line (NCCCWA) and it was described in our previous GS study (Vallejo *et al.*, 2016). Briefly, the YC 2005 families represented the base generation of the breeding line, and thus had not previously been selected for BCWD resistance. Each family had $n = 39$ –80 fish evaluated in the laboratory BCWD challenge in one or two tanks per family. The phenotypic dataset included disease resistance phenotypes from $n =$

4,492 fish from 71 FS families (Table 1), and the pedigree file included 4,659 records. In this NCCWA sample, a total of $n = 583$ fish had both genotype and phenotype records. Following pedigree quality control, the original NCCCWA sample of $n = 583$ genotyped fish was reduced into $n = 577$ because four fish were flagged as duplicated or cloned samples by the QC pipeline and two fish did not assign to the expected family based on the pedigree records.

The second sample included 102 pedigreed FS families from YC 2013 of the commercial breeding company Troutlodge, Inc., May-spawning population (TLUM) and it was described in our previous GS study (VALLEJO *et al.* 2017). Briefly, the original study design was to sample $n = 1500$ fish with phenotypes and genotypes from 50 FS families; in practice, a total of $n = 1473$ fish had both phenotype and genotype records from those 50 FS families ($n = 17$ –40 fish per family). The 102 YC 2013 families represented a commercial nucleus breeding population undergoing selection for growth, and thus had not previously been selected for BCWD resistance. The fish were evaluated for BCWD survival in the laboratory challenge in two tanks per family with an initial stocking of 40 fish per tank. The phenotypic dataset included BCWD disease phenotype records from $n = 7893$ fish from 102 FS families (Table 1), and the pedigree file included 32,279 records. A summary of the experimental variables of GWAS conducted with fish sampled from these two rainbow trout populations is presented in Table 1.

BCWD resistance phenotypes

The discrete BCWD resistance phenotype DAYS, the number of days post-challenge until the fish succumbed to the disease, was recorded for all mortalities and survivors were assigned a value of 21. Each fish also had a binary survival STATUS record. The BCWD resistance

phenotype STATUS had two categories: 1 = the fish died during the 21 days post challenge evaluation period; and 2 = the fish survived for the duration of the challenge. The DAYS and STATUS records were analyzed separately using univariate GWAS models described below.

SNP genotyping platforms

The fish sampled from TLUM and NCCCWA populations were genotyped using the Rainbow Trout Affymetrix 57K SNP array (Chip) following previously described procedures (PALTI *et al.* 2015a) and the samples were genotyped by a commercial service provider (Geneseek, Inc., Lincoln, NE) following the Axiom genotyping procedures described by the array manufacturer (Affymetrix). The quality control (QC) bioinformatics pipeline applied to the Chip-SNP genotype data collected in the TLUM (VALLEJO *et al.* 2017) and NCCCWA (VALLEJO *et al.* 2016) populations were already described. After genotype data QC, a total of 41,868 and 49,468 SNPs were included in the TLUM and NCCCWA raw Chip genotype datasets, respectively.

The fish sampled from the NCCCWA population were also genotyped by sequencing of restriction-site-associated DNA (RAD) tag libraries as we have previously described (VALLEJO *et al.* 2016). After genotype data QC, a total of 24,465 RAD-SNPs were included in the raw RAD genotype dataset. The raw sequence data from the RAD libraries were deposited in the NCBI SRA database (Accession SRP063932).

Before performing GWAS analyses, the raw marker genotype datasets were further QC filtered using algorithms implemented in the software BLUPF90 (MISZTAL *et al.* 2015). For the Chip data, the QC retained SNPs with a genotype calling rate higher than 0.90, minor allele

frequency higher than 0.05, and departures from Hardy-Weinberg equilibrium less than 0.15 (difference between expected and observed frequency of heterozygotes). Parent-progeny pairs were tested for discrepant homozygous SNPs, and SNPs with a conflict rate of more than 1% were discarded from further analysis. For the RAD data, the QC retained SNPs with a genotype calling rate higher than 0.70. Following this final QC step, 33,838 SNPs, 39,112 SNPs and 9,534 SNPs were retained for analyses of the TLUM, NCCCWA (Chip) and NCCCWA (RAD) datasets, respectively. Next, we determined the physical map position (GenBank assembly Accession GCA_002163495) of each of the QC filtered markers and found that a small fraction did not have a physical map location. The numbers of effective genotyped markers and effective genotyped fish that were used with each specific GWAS model and genotyping platform in the evaluated populations are shown in Table 1.

Genome-wide association analyses

This study was conducted to identify chromosomal regions that have the greatest impact on the variation of BCWD resistance because it is difficult to infer individual SNP effects from a multiple regression model that fits markers simultaneously at a genome-wide scale (GARRICK AND FERNANDO 2013). Therefore, instead of using the markers effect to make an inference on a particular locus, we used the markers effect to make an inference about a particular genomic region that encompasses a number of contiguous loci in association with the trait (FAN *et al.* 2011; FERNANDO *et al.* 2014). Thus the GWAS analysis was performed separately for each population and BCWD resistance phenotype using multiple-regression GWAS models that use simultaneously all the SNPs in the association test. Two multiple-regression GWAS methods

were used: Bayesian variable selection BayesB (BayesB) (FERNANDO AND GARRICK 2009) and weighted single-step GBLUP (wssGBLUP) (WANG *et al.* 2012; MISZTAL *et al.* 2015).

Before starting the search for genomic regions associated with BCWD resistance, we tested 0.5 and 1Mb exclusive-windows in GWAS with BayesB model and found that 1Mb windows provided Manhattan plots with less noisy baseline which was in agreement with other GWAS reports (KIZILKAYA *et al.* 2013; SAATCHI *et al.* 2013); so we decided to use 1Mb exclusive-windows in the GWAS performed with BayesB. We also used 1Mb sliding-windows in GWAS performed with wssGBLUP which are more informative than exclusive-windows.

Bayesian variable selection model BayesB

The BayesB model uses only fish that had both genotype and phenotype records. The TLUM population sample included $n = 1473$ fish from 50 YC 2013 families with phenotype and genotype records, and the NCCWA sample included $n = 577$ fish from 10 YC 2005 families with phenotype and genotype records (Table 1).

The GWAS for DAYS was performed using this linear model: $\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\alpha + \mathbf{e}$; where \mathbf{y} is $n \times 1$ vector of phenotypic records; $\mathbf{1}$ is a vector of all ones; μ is overall mean of phenotypic records; \mathbf{Z} is an $n \times k$ matrix of genotype covariates (coded as -10, 0, or 10) for k SNP markers, α is a $k \times 1$ vector of random regression coefficients of k additive marker effects, and \mathbf{e} is a vector of residuals. The genotype and phenotype records were used to estimate markers effect using the Bayesian variable selection model BayesB implemented in the software GENSEL (FERNANDO AND GARRICK 2009) as we have previously described (VALLEJO *et al.* 2016). The GWAS for the

binary phenotype STATUS was performed using the option for categorical analysis implemented also in the software GENSEL (FERNANDO AND GARRICK 2009; FERNANDO AND GARRICK 2013; GARRICK AND FERNANDO 2013).

The BayesB model fits a mixture model to estimate marker effects assuming there are two types of SNP markers: a fraction of non-zero SNP effects ($1 - \pi$) that are drawn from distributions with marker-specific variance (σ_α^2), and other known fraction of markers (π) that have zero effect on the quantitative trait (MEUWISSEN *et al.* 2001). In this study, the mixture parameter π was determined empirically by testing each population dataset listed in Table 1 (TLUM-Chip, NCCCWA-Chip and NCCCWA-RAD) with several π values (0.95, 0.96, 0.97, 0.98 and 0.99). We then decided to use π values of 0.97 and 0.99 with TLUM-Chip and NCCCWA samples, respectively, in the GWAS analysis with BayesB because these π values yielded best accuracy genomic predictions (Results not presented).

The BayesB model uses Gibbs sampling approach in the GWAS analysis (GARRICK AND FERNANDO 2013). In this study, the BCWD phenotypes were analyzed using 270,000 Markov Chain Monte Carlo (MCMC) iterations from which the first 20,000 samples were discarded as burn-in; from the remaining 250,000 samples, we saved one from every 50 samples so the marker effects and variances were estimated as the posterior means of collected 5,000 independent samples. The proper mixing and convergence of the MCMC iterations were assessed with the R package CODA (PLUMMER *et al.* 2006).

Weighted single-step GBLUP model

The GWAS analysis with wssGBLUP, in contrast to BayesB, uses all available information on sampled fish such as pedigree, genotype and phenotype records including those fish that had only phenotypic records (i.e., those with missing genotype data) as long as the sampled fish are pedigree related (AGUILAR *et al.* 2010; CHRISTENSEN AND LUND 2010). The TLUM sample included $n = 7893$ fish from 102 YC 2013 families with phenotype records, and the NCCWA sample included $n = 4492$ BCWD fish from 71 YC 2005 families with phenotype records (Table 1).

In GWAS with wssGBLUP, the weights for each SNP are 1's for the 1st iteration which means that all SNPs have the same weight (i.e., standard single-step GBLUP). For the next iterations (2nd, 3rd, etc.), the weights are SNP specific variances calculated using both the SNP allele-substitution effect estimated in the previous iteration and their corresponding allele frequencies (WANG *et al.* 2012). In this study, we decided to use results from the 2nd iteration because they provide the highest accuracy genomic predictions (VALLEJO *et al.* 2016) and marker effects (WANG *et al.* 2012; IRANO *et al.* 2016; MELO *et al.* 2016).

In GWAS with TLUM population sample, the linear and threshold models for DAYS and STATUS, respectively, included the effects of population mean, random animal, random common environment and random error. The full-sib fish progeny from each family was allocated into two tanks for BCWD challenge evaluation, so the variable tank nested within family was used to model the common environment effect. The linear and threshold models used with the NCCCWA sample were similar to those used with the TLUM sample. The linear model for DAYS and threshold model for STATUS were fitted using computer applications

implemented in the software BLUPF90 (MISZTAL *et al.* 2015). The binary phenotype disease STATUS was analyzed with a threshold model using a Bayesian approach which included a single chain with a total of 270,000 iterations; the first 20,000 iterations were discarded as burn-in iterations; then from the remaining 250,000 samples, one from every 50 samples were saved. Thus, 5000 independent samples were used in the analysis. The proper mixing and convergence of these MCMC iterations were also assessed using the R package CODA (PLUMMER *et al.* 2006).

Criteria to declare QTL associated with BCWD resistance

The results from the GWAS performed with BayesB and wssGBLUP were used to identify genomic windows and QTL associated with BCWD resistance. A two stage approach was used to identify a QTL associated with BCWD resistance. First, the genomic windows with explained genetic variance (EGV) greater than 1% and 2% in the TLUM and NCCCWA populations, respectively, were declared as genomic regions associated with BCWD resistance. The threshold to declare a QTL was raised to $EGV \geq 2\%$ in the relatively small NCCCWA sample ($n = 577$) to control the type I error rate. Second, to determine if neighboring or overlapping windows on the same chromosome belong to the same QTL region we used the following criteria: all windows associated with BCWD resistance that were bounded within a region smaller than 20Mb and were less than 10Mb apart from another associated window were grouped to a single QTL region. The QTL nomenclature we used was based on the chromosome number and QTL region within the chromosome, where the region with the lowest genome assembly position numbers determined to be QTL1 on that chromosome the next QTL2 and so on. For example, the QTL

regions detected on chromosome Omy3 and listed in Table 2 were separated to QTL 3.1, 3.2 and 3.3 based on the physical genome map positions of the SNPs that flanked each QTL region.

The BayesB algorithm estimated the proportion of models where the tested 1Mb-window was included as with non-zero variance ($p > 0$) and therefore accounted for more than 0% of the genetic variance. These $p > 0$ estimates were used to calculate the probability of false positives [$PFP = 1 - (p > 0)$] for each tested window (FERNANDO AND GARRICK 2013). These PFP estimates enabled accounting for multiple testing to control the probability of false positive conclusions across all the undertaken GWAS tests with the BayesB model. Thus by using these EGV thresholds (PETERS *et al.* 2013) and PFP estimates (FERNANDO AND GARRICK 2013), the claims on significant QTL findings was restricted to the strongest associations with BCWD resistance to control the type I error rate in this study.

QTL segregating in both NCCCWA and TLUM populations

In order to identify QTL that might be overlapping between the two populations and between the two genotyping platforms in the NCCCWA population, we assigned the genome physical map positions to all flanking markers of the genomic windows associated with BCWD resistance using the rainbow trout reference genome sequence (GenBank assembly Accession GCA_002163495) and searched for overlapping QTL regions within each chromosome using the flanking markers physical map genome coordinates.

Data availability

The authors state that all data necessary for supporting the conclusions of this research article are included within the article and its Supplemental Material. Figure S1 shows Manhattan plot with GWAS results for survival DAYS in TLUM sample genotyped with 57K Chip-SNP. Figure S2 shows Manhattan plot with GWAS results for survival DAYS in NCCCWA sample genotyped with 57K Chip-SNP. Figure S3 shows Manhattan plot with GWAS results for survival DAYS in NCCCWA sample genotyped with RAD-SNPs. Table S1 contains summary of all genomic windows and QTL detected in this GWAS. Table S2 contains list of QTL associated with survival DAYS found in TLUM population using the 57K Chip-SNP. Table S3 contains list of QTL associated with survival DAYS found in NCCCWA population using the 57K Chip-SNP. Table S4 contains list of QTL associated with survival DAYS found in NCCCWA population using RAD-SNPs genotyping. Table S5 contains list of QTL that are segregating in both NCCCWA and TLUM populations. Table S6 contains list of QTL that are private to either NCCCWA or TLUM population. Table S7 contains summary of QTL for BCWD resistance reported in previous studies. The rainbow trout reference genome sequence is available at GenBank with assembly accession number: GCA_002163495.

RESULTS

Heritability of BCWD resistance

The heritability or proportion of phenotypic variance explained by the markers for survival DAYS and the binary survival STATUS were previously reported (VALLEJO *et al.* 2016; VALLEJO *et al.* 2017). Briefly, here they were moderate and relatively constant with a range of 0.24–0.34 (DAYS) and 0.23–0.35 (STATUS) (Table 1); and the mean heritability for DAYS and

STATUS were similar (0.29). The heritability of STATUS estimated on the underlying scale of liability using a threshold model was transformed to the observed scale of disease survival STATUS using already described procedures (VALLEJO *et al.* 2017). Overall, for both BCWD phenotypes and across genotyping platforms and populations, the mean heritability estimated with wssGBLUP (0.32) was slightly higher than that estimated with BayesB (0.27).

QTL associated with BCWD resistance in the TLUM population (57K SNP Chip)

We have previously shown that the BCWD resistance phenotypes DAYS and STATUS yielded similar results in QTL mapping (VALLEJO *et al.* 2014a; LIU *et al.* 2015b; PALTI *et al.* 2015b) and genomic selection experiments (VALLEJO *et al.* 2016; VALLEJO *et al.* 2017). In this study, we observed also that STATUS and DAYS were affected by similar QTL regions with few exceptions (Overall, DAYS detected three more QTL than STATUS) (Table S1). However, because STATUS resembles the disease resistance trait closer than DAYS; and selection programs for improved disease resistance would most likely favor improvement of resistance over endurance or tolerance (ODEGARD *et al.* 2011), we present the results from the STATUS survival phenotype in the main body of this report. The complete results from the analysis with the DAYS phenotype are presented in the Supplementary Material section.

In the TLUM population, a total of 45 windows with EGV $\geq 1\%$ were detected on chromosomes Omy3, 5, 8, 13 and 25 (Table 2; Figure 1; Table S1 and Figure S1). Fourteen windows were detected with BayesB with EGV up to 57.6% and 31 windows were detected with wssGBLUP with EGV up to 28.7%.

Four QTL (3.2, 8.1, 13.2 and 25.1) were detected by both GWAS models (Table S1). We did not detect any BayesB model specific QTL (i.e., all QTL detected with BayesB were also detected with wssGBLUP); and four QTL (3.1, 3.3, 5.1 and 13.1) were detected only with wssGBLUP.

Overall, we detected eight QTL (3.1, 3.2, 3.3, 5.1, 8.1, 13.1, 13.2 and 25.1) associated with BCWD resistance which jointly explained up to 61% of the genetic variance for BCWD resistance in this TLUM population when accounting only for the largest EGV window in each QTL (Table 2; Table S1 and S2). Among these eight QTL, two significant large-effect QTL were detected on Omy8 (QTL 8.1; $PPF=0.0$) and 25 (QTL 25.1; $PPF=0.01$) with BayesB, each explaining up to 19.3% and 35.4% of the genetic variance for BCWD resistance, respectively (Table 2).

QTL for BCWD resistance in the NCCCWA population (57K SNP Chip)

In the NCCCWA population using the 57K SNP chip, we detected a total of 11 windows associated with BCWD resistance on chromosomes Omy3, 5, 10, 22 and 25 (Table 3; Figure 2; Table S1 and Figure S2). Two and nine QTL windows were detected with BayesB and wssGBLUP, respectively, and each GWAS model explained up to 5.6% and 16.1% of the genetic variance, respectively.

Only one QTL (3.2) was detected by both GWAS models (Table S1). We did not detect any BayesB model specific QTL, and four QTL (5.1, 10.1, 22.1 and 25.1) were detected only with the wssGBLUP method.

Overall, we detected five QTL (3.2, 5.1, 10.1, 22.1 and 25.1) associated with BCWD resistance that explained up to 18.2% of the genetic variance in the NCCCWA population when accounting only for the largest EGV window in each QTL (Table 3 and Table S3).

QTL for BCWD resistance in the NCCCWA population detected with RAD SNPs

In the NCCCWA population using the RAD SNP genotypes, we detected a total of 18 windows associated with BCWD resistance on chromosomes Omy3, 5, 10, 11, 13, 15 and 25 (Table 4; Figure 3; Table S1 and Figure S3). Four windows were detected by BayesB with EGV up to 17.3% and 14 windows were detected by wssGBLUP with EGV up to 26.4%.

Three QTL (3.2, 15.1 and 25.1) were detected by both GWAS models (Table S1). QTL 5.2 was detected only with BayesB, and five QTL (5.1, 10.1, 11.1, 13.1 and 25.2) were detected only with wssGBLUP.

Overall, we detected nine QTL (3.2, 5.1, 5.2, 10.1, 11.1, 13.1, 15.1, 25.1 and 25.2) associated with BCWD resistance which explained up to 31.9% of the genetic variance in this NCCCWA population dataset when accounting only for the largest EGV window in each QTL (Table 4 and Table S4).

DISCUSSION

In GWAS studies, the use of correct statistical models and computer algorithms is paramount to successfully identify the underlying genetic basis of resistance to complex diseases in livestock and aquaculture species. To date there have been several reported GWAS using single-marker association tests in fin fish species (CAMPBELL *et al.* 2014; AYLLON *et al.* 2015; GENG *et al.* 2015; GONEN *et al.* 2015; LIU *et al.* 2015b; PALTI *et al.* 2015b; TSAI *et al.* 2015; TSAI *et al.* 2016) which generally are associated with high type I error rate because single-marker methods do not account for linkage disequilibrium (LD) between physically linked loci in the association test. In this study, we performed GWAS for loci associated with BCWD resistance using multiple-regression models which estimate the effect of all markers simultaneously and consequently do account for LD between neighboring SNPs (FERNANDO AND GARRICK 2013; GARRICK AND FERNANDO 2013; MISZTAL *et al.* 2014).

In the current GWAS, we detected 14 QTL associated with BCWD resistance in two commercially-relevant rainbow trout breeding populations from which 11 were validated QTL from previous studies and three were novel QTL. Here we confirmed that BCWD resistance is controlled by the oligogenic inheritance of few moderate-large effect loci and a large-unknown number of loci with small effects on BCWD resistance. However, despite the similar genetic architecture for this trait in both populations and the detection of overlapping QTL, we still detected major QTL differences between the two populations. We also found that the RAD and Chip genotyping platforms did not detect the same QTL in the NCCCWA population, and overall the RAD platform detected a greater number of QTL than the Chip platform. In addition,

the wssGBLUP and the BayesB multiple-regression GWAS models did not detect the same QTL, which highlights the utility of using different GWAS models to effectively optimize the discovery of QTL.

The genetic architecture of BCWD resistance in rainbow trout

Previously, we predicted that six to 10 QTL explaining 83% to 89% of phenotypic variance with either additive or dominant disease-resistant alleles plus polygenic effects may underlie the genetic architecture of BCWD resistance in the same NCCCWA population using Bayesian complex segregation analysis of phenotype and pedigree records (VALLEJO *et al.* 2010). In the current study we were able to confirm our prediction on the genetic architecture of the trait. With 10 families from that original NCCCWA population, we uncovered 10 moderate-large effect QTL that explained up to 27.7% of the additive genetic variance for BCWD resistance (Table S1). Similarly, in the TLUM odd-year population, we detected 8 moderate-large effect QTL that explained up to 60.8% of the additive genetic variance for BCWD resistance.

Four QTL regions located on chromosomes Omy3, 5, 13 and 25 are segregating in both populations (Table S5). The shared QTL regions explain a substantial proportion of the additive genetic variance for BCWD resistance in the two populations (up to 18% and 38.6% of the genetic variance in NCCCWA and TLUM, respectively); suggesting a common underlying genetic architecture for BCWD resistance in the two populations. However, major differences were also detected between the two populations. Six QTL, which explained up to 9.7% of the genetic variance and are located on chromosomes Omy5, 10, 11, 15, 22 and 25 were found only in the NCCCWA population (Table S6). Conversely, four QTL which explained up to 22.2% of

the genetic variance and are located on chromosomes Omy3, 8 and 13 were only found in the TLUM population. Overall, our GWAS results confirmed the hypothesis that BCWD resistance is controlled by the oligogenic inheritance of several moderate-large effect QTL and many small effect polygenic loci (VALLEJO *et al.* 2010; VALLEJO *et al.* 2014a; LIU *et al.* 2015b; PALTI *et al.* 2015b).

Further fine-mapping of the BCWD-QTL position and eventual identification of putative candidate genes or disease-causal mutations would be advantageous for applying marker-based selection and advancing the understanding of the mechanisms of genetic resistance to BCWD in rainbow trout populations. This can be achieved by genotyping and disease testing a greater number of SNPs from positions within and near the major QTL regions, by re-sequencing highly characterized BCWD resistant and susceptible individuals as was successfully done in the search for the IPNV resistance gene in Atlantic salmon (MOEN *et al.* 2015). In addition, positional and functional candidate genes for the QTL can be generated by interrogating the newest version of the rainbow trout reference genome sequence (GenBank assembly Accession GCA_002163495).

Comparing the two SNP genotyping technologies in the NCCCWA population

Overall, the RAD genotyping technology (18 windows; EGV= 32.8%; Table S1) detected a greater number of windows associated with BCWD resistance than the Chip technology (11 windows; EGV= 18.2%) in the NCCCWA population. From the 10 QTL found in the NCCCWA population, more than half of the detected QTL were genotyping platform specific: One QTL was detected only with the Chip technology (QTL 22.1); and five QTL were detected only with the RAD technology (QTL 5.2, 11.1, 13.1, 15.1 and 25.2). Four QTL were detected by both

SNP genotyping technologies (QTL 3.2, 5.1, 10.1 and 25.1). The overall better performance of RAD-SNPs than Chip-SNPs in detecting QTL associated with BCWD resistance in the NCCCWA population may be due to sample ascertainment bias effects considering that the 57K SNP Chip was developed using a collection of samples from different rainbow trout populations to maximize SNP polymorphism and discovery (PALTI *et al.* 2015a). Therefore, the polymorphic markers in the SNP Chip might be less informative than the SNPs we genotyped with the RAD technology, which were specifically discovered in the sampled families from the NCCCWA population, and are therefore more informative for characterizing genome loci polymorphisms in this dataset.

Comparing BayesB and wssGBLUP models

Overall, we noticed that the wssGBLUP detected higher number of windows (54) associated with BCWD resistance than the BayesB (20) across the three datasets (TLUM-SNP, NCCCWA-SNP and NCCCWA-RAD) we used here (Table S1). We also noticed that the QTL detection power of BayesB was more negatively impacted by sample size reduction than wssGBLUP: BayesB detected 14, 2 and 4 QTL windows and wssGBLUP detected 31, 9 and 14 QTL windows in TLUM-SNP, NCCCWA-SNP and NCCCWA-RAD datasets, respectively. Performing GS with the Chip genotyped SNPs, we have shown that BayesB predicts GEBVs with higher accuracy than wssGBLUP when using a training sample size of $n = 1473$ (VALLEJO *et al.* 2017); however, wssGBLUP outperforms BayesB when using a smaller training sample size of $n = 583$ (VALLEJO *et al.* 2016). So, in agreement with these previous GS results, it seems also that the QTL detection power of GWAS with BayesB is more sensitive to sample size reduction than wssGBLUP. We think that the difference in power robustness to sample size reduction of these

GWAS models is due to their algorithmic differences. The BayesB model does not explicitly use the available pedigree information and includes in the analysis only animals that had both genotype and phenotype records and as a consequence is less power robust to sample size reduction and needs a minimum sample size for optimal performance, i.e. $n \geq 1000$ animals. In contrast, the power of GWAS analysis with wssGBLUP is more robust to sample size reduction because it uses the available pedigree information and includes in the analysis animals that have both phenotype and genotype records, and also animals that have only phenotype records (those with missing genotype data).

In the Chip-SNP genotyped TLUM sample, four QTL (3.2, 8.1, 13.2 and 25.1) were detected with both GS models, and four QTL (3.1, 3.3, 5.1 and 13.1) were detected only with wssGBLUP (Table S1). Similarly, in the Chip-SNP genotyped NCCCWA sample, only the QTL 3.2 was detected with both GS models, and most of the QTL (5.1, 10.1, 22.1 and 25.1) were detected only with wssGBLUP. In contrast, in the RAD-SNP genotyped NCCCWA sample, three QTL (3.2, 15.1 and 25.1) were detected with both GS models, the QTL 5.2 was detected only with BayesB, and five QTL (5.1, 10.1, 11.1, 13.1 and 25.2) were detected only with wssGBLUP. Thus, these results highlight the importance of using at least two different GWAS algorithms to efficiently uncover the underlying genetic basis of resistance against BCWD in the studied populations.

Comparing QTL detected in TLUM and NCCCWA populations

In spite of the smaller sample size of the NCCCWA population in comparison to the TLUM population, we detected 10 QTL in the NCCCWA population and only eight QTL in the TLUM

population (Table S1). We hypothesize that because the TLUM sample size was much larger than the NCCCWA sample size, it is likely that the type I error rate was smaller in the TLUM sample than in the NCCCWA sample. Therefore, we predict that most of the QTL detected in the TLUM population are real, but some of the QTL detected in the NCCCWA population are false positives. Specifically, the NCCCWA population had two unique QTL (5.2 and 15.1) that were also not reported in past studies. Thus, those two NCCCWA-specific QTL may very be false positives.

Comparing our GWAS results with previous studies

Eleven of the 14 QTL we detected in this study were also reported in previous studies in the NCCCWA germplasm and in other populations (Table S7) (JOHNSON *et al.* 2008; WIENS *et al.* 2013; CAMPBELL *et al.* 2014; QUILLET *et al.* 2014; VALLEJO *et al.* 2014a; VALLEJO *et al.* 2014b; LIU *et al.* 2015b; PALT *et al.* 2015b). Campbell *et al.* (CAMPBELL *et al.* 2014) detected RAD SNPs associated with BCWD resistance in another commercial rainbow trout population that were about 1Mb from our QTL Omy8.1 (Table S1 and Table S7); they also reported QTL that overlap or are close to our detected QTL 10.1 and 25.2. Kutyrev *et al.* (KUTYREV *et al.* 2016) measured the expression of immune relevant genes on spleen tissue sampled from BCWD resistant (ARS-Fp-R) and susceptible (ARS-Fp-S) genetic lines after laboratory disease pathogen challenge and detected differential expression between the tested genetic lines for genes *il1r-like-1* and *tnfrsf1a-like-a*. Interestingly, two SNPs for the gene *il1r-like-1* (Affx-88933101 and Affx-88915186) were about 240Kb from our QTL 3.2 which explained up to 5.6% of the genetic variance in the NCCCWA population.

Previous studies also detected QTL for BCWD resistance on chromosomes Omy1 (VALLEJO *et al.* 2014a; PALTÍ *et al.* 2015b), 2 (VALLEJO *et al.* 2014a; LIU *et al.* 2015b), 7 (QUILLET *et al.* 2014; VALLEJO *et al.* 2014a; PALTÍ *et al.* 2015b), 12 (VALLEJO *et al.* 2014a; LIU *et al.* 2015b), 17 (JOHNSON *et al.* 2008; CAMPBELL *et al.* 2014; QUILLET *et al.* 2014), 26 and 28 (LIU *et al.* 2015b) which were not detected in this study. These conflicting results in QTL mapping can be expected due to several reasons including: (1) QTL effects can be population and/or family specific with unique extent/phase of linkage between QTL and marker alleles; and (2) they can also represent false positive results due to limitations and weaknesses of experimental-design and power of analysis as we describe here.

Conclusion

This GWAS is the most comprehensive genome-wide scan for QTL associated with BCWD resistance performed to date in two commercially-relevant rainbow trout breeding populations, using two whole-genome SNP genotyping platforms and two multiple-regression GWAS models. We identified a total of 14 moderate-large effect QTL associated with resistance to BCWD resistance, and four of those QTL were segregating in the two populations. These GWAS results confirmed that the genetic architecture of BCWD resistance is controlled by the oligogenic inheritance of few moderate-large effect genes and many small effect resistance loci. Overall, the wssGBLUP detected higher number of QTL than the BayesB and both GWAS models did not detect the same QTL which highlights the utility of using two different GWAS algorithms to effectively discover QTL. The RAD genotyping platform detected higher number of QTL than the Chip technology and also both genotyping platforms did not detect the same QTL in the NCCCWA population. These GWAS results will advance the biological and

functional analysis of positional candidate genes using the annotation of the new rainbow trout reference genome (GenBank assembly Accession GCA_002163495). They are also likely to be used in the implementation of more efficient selective breeding strategies which will utilize the QTL-flanking SNPs in genome-enabled selection for BCWD resistance in rainbow trout aquaculture.

ACKNOWLEDGEMENTS

This project was supported by funds from the USDA Agricultural Research Service in house project numbers 8082-32000-006 and 8082-31000-012. We would like to acknowledge the following people for providing technical assistance including Roseanna Long, Kristy Shewbridge, and Cassandra Parker for sample preparation and genotyping; Caird Rexroad, Josh Kretzer, Kyle Jenkins, Travis Moreland, Clayton Birkett, Jen and Ryan Lipscomb and Bryce Williams for fish rearing, phenotyping and sampling. The authors are very grateful to Ignacy Misztal, Shogo Tsuruta, Rafael Silva and Daniela Lourenco for insightful discussions on performing single-step GBLUP analysis with software BLUPF90. We also acknowledge Dorian Garrick for helping to perform GWAS analysis using Bayesian methods with software GENSEL. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

REFERENCES

- Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta *et al.*, 2010 Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93: 743-752.
- Ayllon, F., E. Kjaerner-Semb, T. Furmanek, V. Wennevik, M. F. Solberg *et al.*, 2015 The vglI3 Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (*Salmo salar* L.) Males. *Plos Genetics* 11.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver *et al.*, 2008 Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *Plos One*.
- Barnes, M., and M. Brown, 2011 A review of *Flavobacterium psychrophilum* biology, clinical signs, and bacterial cold water disease prevention and treatment. *The Open Fish Science Journal* 4: 40-48.
- Brieuc, M. S. O., C. D. Waters, J. E. Seeb and K. A. Naish, 2014 A Dense Linkage Map for Chinook salmon (*Oncorhynchus tshawytscha*) Reveals Variable Chromosomal Divergence After an Ancestral Whole Genome Duplication Event. *G3-Genes Genom Genet*: 447-460.
- Campbell, N. R., S. E. LaPatra, K. Overturf, R. Towner and S. R. Narum, 2014 Association Mapping of Disease Resistance Traits in Rainbow Trout Using Restriction Site Associated DNA Sequencing. *G3: Genes | Genomes | Genetics* 4: 2473-2481.
- Christensen, O. F., and M. S. Lund, 2010 Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42.
- Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen *et al.*, 2011 Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12: 499-510.

- de los Campos, G., A. I. Vazquez, R. Fernando, Y. C. Klimentidis and D. Sorensen, 2013 Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *Plos Genetics* 9.
- Fan, B., S. K. Onteru, Z. Q. Du, D. J. Garrick, K. J. Stalder *et al.*, 2011 Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. *PLoS One* 6: e14726.
- Fernando, R. L., and D. J. Garrick, 2009 GenSel – User Manual for a Portfolio of Genomic Selection Related Analyses. Animal Breeding and Genetics Iowa State University Ames.
<http://www.biomedcentral.com/content/supplementary/1471-2105-12-186-S1.PDF>, pp.
- Fernando, R. L., and D. J. Garrick, 2013 Bayesian methods applied to GWAS, pp. 237-274 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J. v. d. Werf and B. Hayes. Humana Press, New York.
- Fernando, R. L., A. Toosi, D. J. Garrick and J. C. M. Dekkers, 2014 Application of Whole-Genome Prediction Methods for Genome-Wide Association Studies: Bayesian Approach. Proceedings, 10th World Congress of Genetics Applied to Livestock Production.
- Garrick, D. J., and R. L. Fernando, 2013 Implementing a QTL detection study (GWAS) using genomic prediction methodology, pp. 275-298 in *Genome-Wide Association Studies and Genomic Prediction*, edited by C. Gondro, J. v. d. Werf and B. Hayes. Humana Press, New York.
- Geng, X., J. Sha, S. K. Liu, L. S. Bao, J. R. Zhang *et al.*, 2015 A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *Bmc Genomics* 16.
- Gonen, S., M. Baranski, I. Thorland, A. Norris, H. Grove *et al.*, 2015 Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (*Salmo salar*). *Heredity* 115: 405-414.
- Gonen, S., N. Lowe, T. Cezard, K. Gharbi, S. Bishop *et al.*, 2014 Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* 15: 166.

- Gonzalez-Pena, D., G. Gao, M. Baranski, T. Moen, B. M. Cleveland *et al.*, 2016 Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus mykiss*). *Front Genet* 7: 203.
- Groenen, M., H.-J. Megens, Y. Zare, W. Warren, L. Hillier *et al.*, 2011 The development and characterization of a 60K SNP chip for chicken. *BMC Genomics* 12: 274.
- Habier, D., R. L. Fernando and J. C. Dekkers, 2007 The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177: 2389-2397.
- Hale, M. C., F. P. Thrower, E. A. Berntson, M. R. Miller and K. M. Nichols, 2013 Evaluating Adaptive Divergence Between Migratory and Nonmigratory Ecotypes of a Salmonid Fish, *Oncorhynchus mykiss*. *G3: Genes|Genomes|Genetics* 3: 1273-1285.
- Hayes, B. J., J. Pryce, A. J. Chamberlain, P. J. Bowman and M. E. Goddard, 2010 Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. *Plos Genetics* 6.
- Hayes, B. J., P. M. Visscher and M. E. Goddard, 2009 Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research* 91: 47-60.
- Hecht, B. C., N. R. Campbell, D. E. Holecek and S. R. Narum, 2013 Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology* 22: 3061-3076.
- Hecht, B. C., F. P. Thrower, M. C. Hale, M. R. Miller and K. M. Nichols, 2012 Genetic Architecture of Migration-Related Traits in Rainbow and Steelhead Trout, *Oncorhynchus mykiss*. *G3: Genes|Genomes|Genetics* 2: 1113-1127.
- Higginson, A. D., and M. R. Munafo, 2016 Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. *PLoS Biol* 14: e2000995.

- Houston, R., J. Taggart, T. Cezard, M. Bekaert, N. Lowe *et al.*, 2014 Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* 15: 90.
- Houston, R. D., J. W. Davey, S. C. Bishop, N. R. Lowe, J. C. Mota-Velasco *et al.*, 2012 Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* 13.
- Howard, J. T., S. Jiao, F. Tiezzi, Y. Huang, K. A. Gray *et al.*, 2015 Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. *BMC Genet* 16: 59.
- Irano, N., G. M. de Camargo, R. B. Costa, A. P. Terakado, A. F. Magalhaes *et al.*, 2016 Genome-Wide Association Study for Indicator Traits of Sexual Precocity in Nellore Cattle. *PLoS One* 11: e0159502.
- Johnson, N. A., R. L. Vallejo, J. T. Silverstein, T. J. Welch, G. D. Wiens *et al.*, 2008 Suggestive association of major histocompatibility IB genetic markers with resistance to bacterial cold water disease in rainbow trout (*Oncorhynchus mykiss*). *Marine Biotechnology* 10: 429-437.
- Kizilkaya, K., R. G. Tait, D. J. Garrick, R. L. Fernando and J. M. Reecy, 2013 Genome-wide association study of infectious bovine keratoconjunctivitis in Angus cattle. *BMC Genet* 14: 23.
- Kutyrev, I., B. Cleveland, T. Leeds and G. D. Wiens, 2016 Proinflammatory cytokine and cytokine receptor gene expression kinetics following challenge with *Flavobacterium psychrophilum* in resistant and susceptible lines of rainbow trout (*Oncorhynchus mykiss*). *Fish Shellfish Immunol* 58: 542-553.
- Leeds, T. D., J. T. Silverstein, G. M. Weber, R. L. Vallejo, Y. Palti *et al.*, 2010 Response to selection for bacterial cold water disease resistance in rainbow trout. *Journal of Animal Science* 88: 1936-1946.

- Legarra, A., O. F. Christensen, I. Aguilar and I. Misztal, 2014 Single Step, a general approach for genomic selection. *Livestock Science* 166: 54-65.
- Liu, S., R. Vallejo, G. Gao, Y. Palti, G. Weber *et al.*, 2015a Identification of Single-Nucleotide Polymorphism Markers Associated with Cortisol Response to Crowding in Rainbow Trout. *Marine Biotechnology* 17: 328-337.
- Liu, S., R. L. Vallejo, Y. Palti, G. Gao, D. P. Marancik *et al.*, 2015b Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Frontiers in Genetics* 6: 1-10.
- Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan *et al.*, 2009 Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE* 4: e5350.
- Melo, T. P., L. Takada, F. Baldi, H. N. Oliveira, M. M. Dias *et al.*, 2016 Assessing the value of phenotypic information from non-genotyped animals for QTL mapping of complex traits in real and simulated populations. *BMC Genet* 17: 89.
- Meuwissen, T. H. E., B. J. Hayes and M. E. Goddard, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.
- Miller, M. R., J. P. Brunelli, P. A. Wheeler, S. Liu, C. E. Rexroad *et al.*, 2012 A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology* 21: 237-249.
- Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko and E. A. Johnson, 2007 Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17: 240-248.
- Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra *et al.*, 2015 Manual for BLUPF90 family of programs, pp., University of Georgia, Athens, GA, USA.

- Misztal, I., H. Wang, I. Aguilar, A. Legarra, S. Tsuruta *et al.*, 2014 GWAS using ssGBLUP. Proceedings, 10th World Congress of Genetics Applied to Livestock Production.
- Moen, T., J. Torgersen, N. Santi, W. S. Davidson, M. Baranski *et al.*, 2015 Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon. *Genetics* 200: 1313-1326.
- Narum, S. R., N. R. Campbell, K. A. Meyer, M. R. Miller and R. W. Hardy, 2013 Thermal adaptation and acclimation of ectotherms from differing aquatic climates. *Molecular Ecology* 22: 3090-3097.
- Nematollahi, A., A. Decostere, F. Pasmans and F. Haesebrouck, 2003 *Flavobacterium psychrophilum* infections in salmonid fish. *J Fish Dis* 26: 563-574.
- Odegard, J., T. Gitterle, P. Madsen, T. H. E. Meuwissen, M. H. Yazdi *et al.*, 2011 Quantitative genetics of taura syndrome resistance in pacific white shrimp (*penaeus vannamei*): a cure model approach. *Genetics Selection Evolution* 43.
- Onteru, S. K., B. Fan, M. T. Nikkila, D. J. Garrick, K. J. Stalder *et al.*, 2011 Whole-genome association analyses for lifetime reproductive traits in the pig. *Journal of Animal Science* 89: 988-995.
- Palti, Y., G. Gao, S. Liu, M. P. Kent, S. Lien *et al.*, 2015a The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources* 15: 662-672.
- Palti, Y., G. Gao, M. R. Miller, R. L. Vallejo, P. A. Wheeler *et al.*, 2014 A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Molecular Ecology Resources* 14: 588-596.
- Palti, Y., J. T. Silverstein, H. Wieman, J. G. Phillips, F. T. Barrows *et al.*, 2006 Evaluation of family growth response to fishmeal and gluten-based diets in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* 255: 548-556.

- Palti, Y., R. L. Vallejo, G. Gao, S. Liu, A. G. Hernandez *et al.*, 2015b Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing. Plos One 10: e0138435.
- Peters, S. O., K. Kizilkaya, D. J. Garrick, R. L. Fernando, J. M. Reecy *et al.*, 2013 Heritability and Bayesian genome-wide association study of first service conception and pregnancy in Brangus heifers. J Anim Sci 91: 605-612.
- Plummer, M., N. Best, K. Cowles and K. Vines, 2006 CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6: 7-11.
- Quillet, E., N. Dechamp, C. Hervet, F. Krieg, C. Chantry-Darmon *et al.*, 2014 Genetic Bases of Resistance versus Susceptibility to *Flavobacterium psychrophilum* in Rainbow Trout. Proceedings, 10th World Congress of Genetics Applied to Livestock Production.
- Ramos, A. M., R. P. M. A. Crooijmans, N. A. Affara, A. J. Amaral, A. L. Archibald *et al.*, 2009 Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. PLoS ONE 4: e6524.
- Saatchi, M., D. J. Garrick, R. G. Tait, Jr., M. S. Mayes, M. Drewnoski *et al.*, 2013 Genome-wide association and prediction of direct genomic breeding values for composition of fatty acids in Angus beef cattle. BMC Genomics 14: 730.
- Silverstein, J. T., R. L. Vallejo, Y. Palti, T. D. Leeds, C. E. Rexroad *et al.*, 2009 Rainbow trout resistance to bacterial cold-water disease is moderately heritable and is not adversely correlated with growth. Journal of Animal Science 87: 860-867.
- Tsai, H. Y., A. Hamilton, A. E. Tinch, D. R. Guy, J. E. Bron *et al.*, 2016 Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. Genetics Selection Evolution 48.

- Tsai, H. Y., A. Hamilton, A. E. Tinch, D. R. Guy, K. Gharbi *et al.*, 2015 Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *Bmc Genomics* 16.
- Vallejo, R. L., T. D. Leeds, B. O. Fragomeni, G. Gao, A. G. Hernandez *et al.*, 2016 Evaluation of Genome-Enabled Selection for Bacterial Cold Water Disease Resistance Using Progeny Performance Data in Rainbow Trout: Insights on Genotyping Methods and Genomic Prediction Models. *Frontiers in Genetics* 7: 1-13.
- Vallejo, R. L., T. D. Leeds, G. Gao, J. E. Parsons, K. E. Martin *et al.*, 2017 Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet Sel Evol* 49: 17.
- Vallejo, R. L., Y. Palti, S. Liu, J. P. Evenhuis, G. Gao *et al.*, 2014a Detection of QTL in Rainbow Trout Affecting Survival When Challenged with *Flavobacterium psychrophilum*. *Marine Biotechnology* 16: 349–360.
- Vallejo, R. L., Y. Palti, S. Liu, D. P. Marancik and G. D. Wiens, 2014b Validation of linked QTL for bacterial cold water disease resistance and spleen size on rainbow trout chromosome Omy19. *Aquaculture* 432: 139-143.
- Vallejo, R. L., G. D. Wiens, C. E. Rexroad, T. J. Welch, J. P. Evenhuis *et al.*, 2010 Evidence of major genes affecting resistance to bacterial cold water disease in rainbow trout using Bayesian methods of segregation analysis. *Journal of Animal Science* 88: 3814-3832.
- Visscher, P. M., J. Yang and M. E. Goddard, 2010 A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res Hum Genet* 13: 517-524.
- Wang, H., I. Misztal, I. Aguilar, A. Legarra and W. M. Muir, 2012 Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research* 94: 73-83.

Wiens, G. D., R. L. Vallejo, T. D. Leeds, Y. Palti, S. Hadidi *et al.*, 2013 Assessment of Genetic Correlation between Bacterial Cold Water Disease Resistance and Spleen Index in a Domesticated Population of Rainbow Trout: Identification of QTL on Chromosome Omy19. PLoS ONE 8: e75749.

Table 1 Experimental variables of GWAS conducted in two rainbow trout populations using two SNP genotyping methods

Population ^a	Genotyping platform ^b	GWAS method ^c	BCWD phenotype ^d	1Mb windows	Genotyped SNP ^e	Genotyped fish ^e	Phenotyped fish	Genetic parameter ^f			
								σ_g^2	σ_c^2	σ_e^2	h_M^2
TLUM	Chip	BayesB	DAYS	1,840	31,787	1,473	1,473	13.72	Na ^g	44.56	0.24
TLUM	Chip	BayesB	STATUS	1,840	31,787	1,473	1,473	0.58	Na	1.00	0.23
TLUM	Chip	wssGBLUP	DAYS	1,394	31,787	2,500	7,893	15.55	0.50	32.48	0.32
TLUM	Chip	wssGBLUP	STATUS	1,406	31,787	2,500	7,893	1.24	0.03	1.00	0.35
NCCCWA	Chip	BayesB	DAYS	1,847	36,666	577	577	13.30	Na	35.71	0.27
NCCCWA	Chip	BayesB	STATUS	1,847	36,666	577	577	0.79	Na	1.00	0.28
NCCCWA	Chip	wssGBLUP	DAYS	1,420	36,666	652	4,492	13.09	0.23	30.95	0.30
NCCCWA	Chip	wssGBLUP	STATUS	1,408	36,666	652	4,492	0.83	0.01	1.00	0.28
NCCCWA	RAD	BayesB	DAYS	1,777	7,972	574	574	13.40	Na	34.88	0.28
NCCCWA	RAD	BayesB	STATUS	1,777	7,972	574	574	0.86	Na	1.00	0.29
NCCCWA	RAD	wssGBLUP	DAYS	1,243	7,972	649	4,492	15.09	0.20	29.71	0.34
NCCCWA	RAD	wssGBLUP	STATUS	1,253	7,972	649	4,492	1.06	0.01	1.00	0.32

^aGWAS was performed using fish from Troutlodge US May (TLUM) and NCCCWA rainbow trout populations, separately.

^bThe sampled fish were genotyped with the 57K SNP array (Chip) and with RAD-SNPs (RAD) generated by sequencing of RAD tag libraries.

^cGWAS was performed using Bayesian variable selection model BayesB and weighted single-step GBLUP at iteration 2 (wssGBLUP) methods. The BayesB method used 1Mb exclusive-consecutive windows and the wssGBLUP method used 1Mb moving-sliding windows.

^dBCWD resistance phenotypes: survival days after disease challenge (DAYS) and binary fish survival status (STATUS).

^eThese are effective number of genotyped SNPs and fish after data quality control, respectively, used in the GWAS analyses.

^fGenetic parameter estimates: σ_g^2 is the additive genetic variance; σ_e^2 is the common environment variance; σ_e^2 is the residual error; and h_M^2 is the proportion of phenotypic variance explained by the markers. For the binary phenotype STATUS, the h_M^2 estimated on the underlying scale of liability was transformed to the observed scale.

^gNa indicates a non-available estimate. The BayesB GWAS model did not include the common environment random effect.

Table 2 Summary of QTL associated with BCWD survival STATUS in the Troutlodge US May (TLUM) population^a

Omy	QTL ^b	GWAS method ^c	Genetic variance (%) ^d	Physical Map (bp) ^e		Markers in Window		SNPs per window
				Start	End	Start-SNP	End-SNP	
3	3.1	wssGBLUP	1.8	17,812,341	18,600,963	Affx-88909970	Affx-88904917	22
3	3.2	wssGBLUP	2.0	61,621,949	62,558,467	Affx-88925949	Affx-88919479	18
3	3.3	wssGBLUP	1.2	77,108,538	78,076,592	Affx-88925305	Affx-88929879	19
5	5.1	wssGBLUP	1.2	11,339,155	12,329,117	Affx-88916119	Affx-88936955	27
8	8.1	BayesB	19.3	76,070,399	76,907,400	Affx-88955037	Affx-88906927	17
25	25.1	BayesB	35.4	21,006,787	21,805,909	Affx-88924154	Affx-88936445	30

^aThe fish from TLUM population were genotyped with the 57K SNP array (Chip).

^bFrom each QTL, the window with the highest explained genetic variance is presented in this Table.

^cGWAS was conducted using Bayesian variable selection model BayesB (BayesB) and weighted single-step GBLUP (wssGBLUP) methods. BayesB used 1Mb exclusive-consecutive windows and wssGBLUP used 1Mb moving-sliding windows.

^dExplained genetic variance by tested window (%).

^eSNP positions in base pairs (bp) based on rainbow trout reference genome sequence (GenBank assembly Accession GCA_002163495).

Table 3 Summary of QTL associated with BCWD survival STATUS in NCCCWA population detected using the 57K SNP array

Omy	QTL ^a	GWAS method ^b	Genetic variance (%) ^c	Physical Map (bp) ^d		Markers in Window		SNPs per window
				Start	End	Start-SNP	End-SNP	
3	3.2	BayesB	5.6	55,025,670	55,964,831	Affx-88917670	Affx-88935875	24
5	5.1	wssGBLUP	3.7	11,245,430	12,244,569	Affx-88930371	Affx-88921454	36
10	10.1	wssGBLUP	2.7	31,536,788	32,517,865	Affx-88925834	Affx-88904643	47
25	25.1	wssGBLUP	2.9	28,240,466	29,219,522	Affx-88919589	Affx-88945013	40

^aFrom each QTL, the window with the highest explained genetic variance is presented in this Table.

^bGWAS was conducted using Bayesian variable selection model BayesB (BayesB) and weighted single-step GBLUP (wssGBLUP) methods. BayesB used 1Mb exclusive-consecutive windows and wssGBLUP used 1Mb moving-sliding windows.

^cExplained genetic variance by tested window (%).

^dSNP positions in base pairs (bp) based on rainbow trout reference genome sequence (GenBank assembly Accession GCA_002163495).

Table 4 Summary of QTL associated with BCWD survival STATUS in NCCCWA population detected using RAD-SNPs genotyping

Omy	QTL ^a	GWAS method ^b	Genetic variance (%) ^c	Physical Map (bp) ^d		Markers in Window		SNPs per window
				Start	End	Start-SNP	End-SNP	
3	3.2	BayesB	5.1	55,254,048	55,993,431	BCWD10F04977	BCWD10F00765	3
5	5.1	wssGBLUP	2.8	11,966,155	12,948,479	BCWD10F15578	BCWD10F00357	11
5	5.2	BayesB	3.3	41,094,726	41,686,801	BCWD10F02753	BCWD10F18861	6
13	13.1	wssGBLUP	2.1	11,230,016	11,602,309	BCWD10F05067	BCWD10F24318	3
15	15.1	wssGBLUP	3.7	38,446,758	39,349,228	BCWD10F00773	BCWD10F16617	3
25	25.1	wssGBLUP	6.6	17,496,495	18,444,865	BCWD10F06483	BCWD10F14339	11

^aFrom each QTL, the window with the highest explained genetic variance is presented in this Table.

^bGWAS was conducted using Bayesian variable selection model BayesB (BayesB) and weighted single-step GBLUP (wssGBLUP) methods. BayesB used 1Mb exclusive-consecutive windows and wssGBLUP used 1Mb moving-sliding windows.

^cExplained genetic variance by tested window (%).

^dSNP positions in base pairs (bp) based on rainbow trout reference genome sequence (GenBank assembly Accession GCA_002163495).

FIGURE LEGENDS

Figure 1

Manhattan plot showing the association between SNP genomic windows and BCWD resistance in TLUM sample genotyped with 57K Chip-SNP: (A) GWAS for STATUS performed with BayesB using 1Mb exclusive windows. (B) GWAS for STATUS performed with wssGBLUP using 1Mb sliding windows. (PPTX file)

Figure 2

Manhattan plot showing the association between SNP genomic windows and BCWD resistance in NCCCWA sample genotyped with 57K Chip-SNP: (A) GWAS for STATUS performed with BayesB using 1Mb exclusive windows. (B) GWAS for STATUS performed with wssGBLUP using 1Mb sliding windows. (PPTX file)

Figure 3

Manhattan plot showing the association between SNP genomic windows and BCWD resistance in NCCCWA sample genotyped with the RAD-SNPs: (A) GWAS for STATUS performed with BayesB using 1Mb exclusive windows. (B) GWAS for STATUS performed with wssGBLUP using 1Mb sliding windows. (PPTX file)

SUPPLEMENTAL MATERIAL

Supplemental Figures

Figure S1

Manhattan plot showing the association between SNP genomic windows and BCWD resistance in TLUM sample genotyped with 57K Chip-SNP: (A) GWAS for DAYS performed with BayesB using 1Mb exclusive windows. (B) GWAS for DAYS performed with wssGBLUP using 1Mb sliding windows. (PPTX file)

Figure S2

Manhattan plot showing the association between SNP genomic windows and BCWD resistance in NCCCWA sample genotyped with 57K Chip-SNP: (A) GWAS for DAYS performed with BayesB using 1Mb exclusive windows. (B) GWAS for DAYS performed with wssGBLUP using 1Mb sliding windows. (PPTX file)

Figure S3

Manhattan plot showing the association between SNP genomic windows and BCWD resistance in NCCCWA sample genotyped with the RAD-SNPs: (A) GWAS for DAYS performed with BayesB using 1Mb exclusive windows. (B) GWAS for DAYS performed with wssGBLUP using 1Mb sliding windows. (PPTX file)

Supplemental Tables

Table S1

Summary of QTL associated¹ with BCWD resistance detected using two GWAS methods and two SNP genotyping platforms in NCCCWA and TLUM populations. (XLSX file)

Table S2

Summary of QTL associated with BCWD survival DAYS in the Troutlodge US May (TLUM) population. (DOCX file)

Table S3

Summary of QTL associated with BCWD survival DAYS in NCCCWA population detected using the 57K Chip-SNP. (DOCX file)

Table S4

Summary of QTL associated with BCWD survival DAYS in NCCCWA population detected using RAD-SNPs genotyping. (DOCX file)

Table S5

QTL for BCWD resistance that are shared or segregating in both NCCCWA and TLUM populations. (XLSX file)

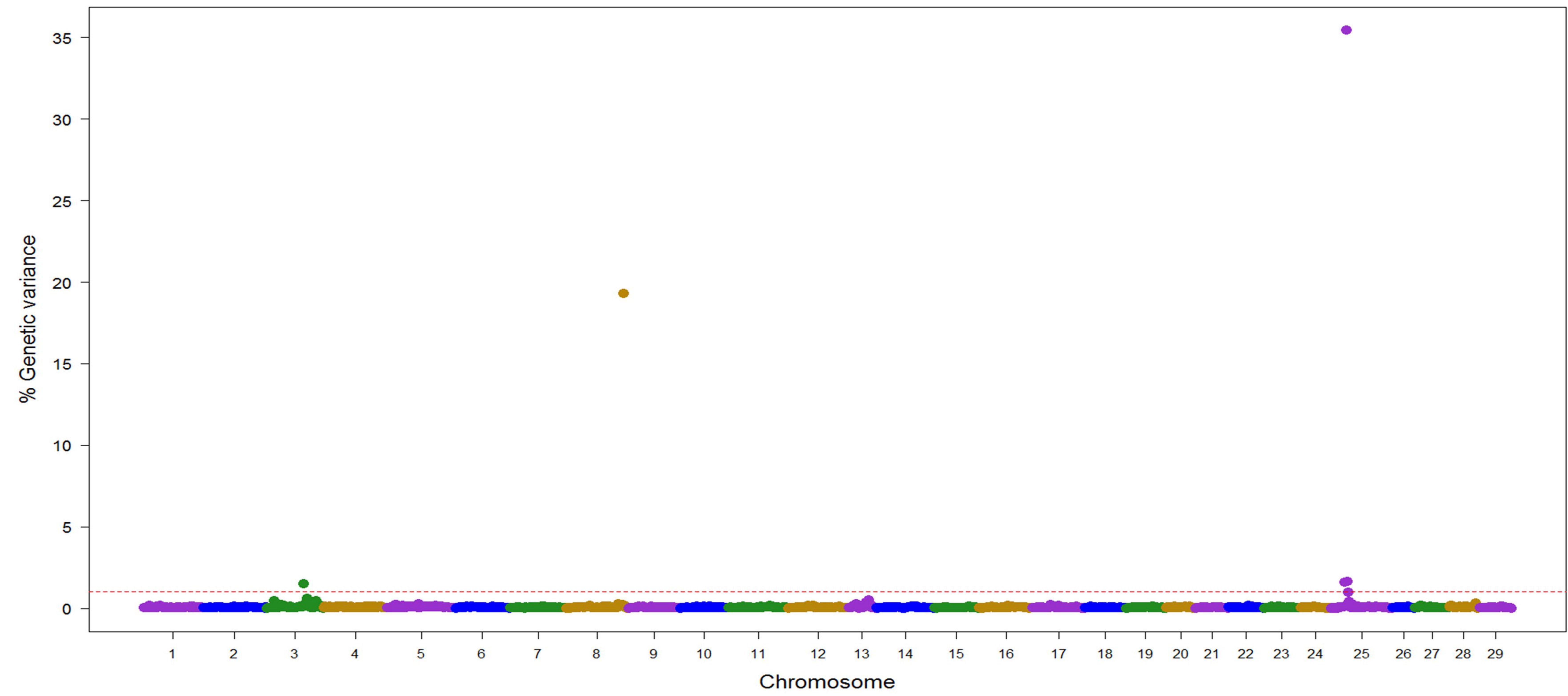
Table S6

QTL for BCWD resistance that are private to either NCCCWA or TLUM population. (XLSX file)

Table S7

Summary of QTL for BCWD resistance reported in previous studies in rainbow trout populations. (XLSX file)

(A)



(B)

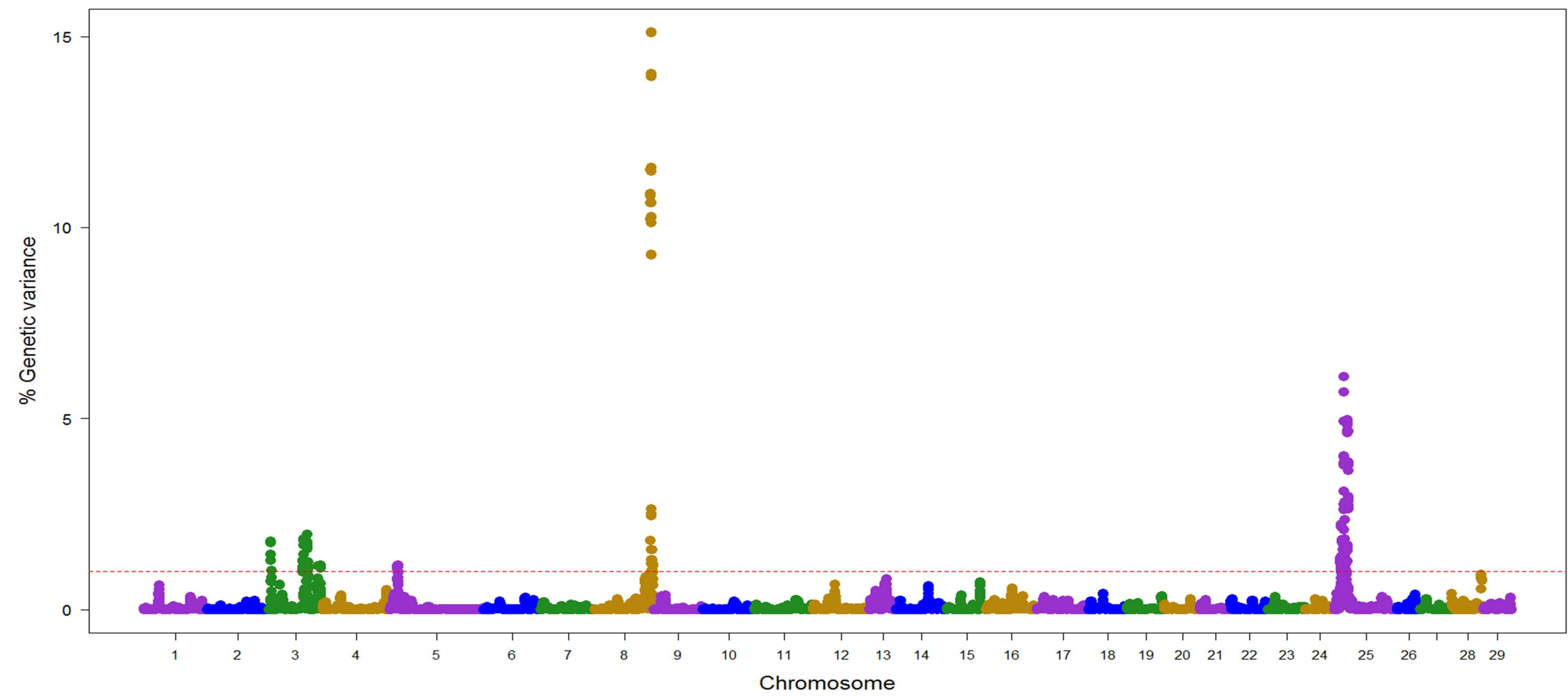
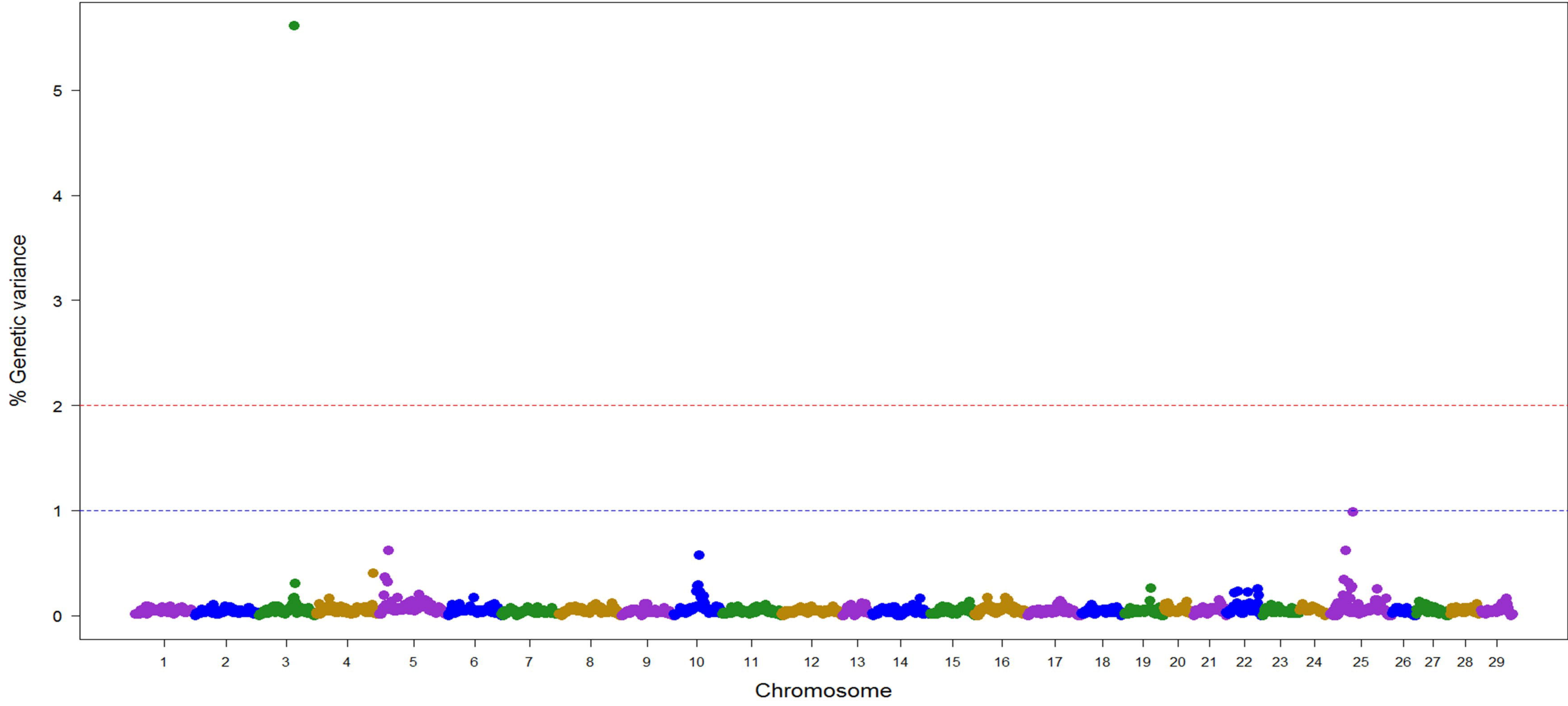
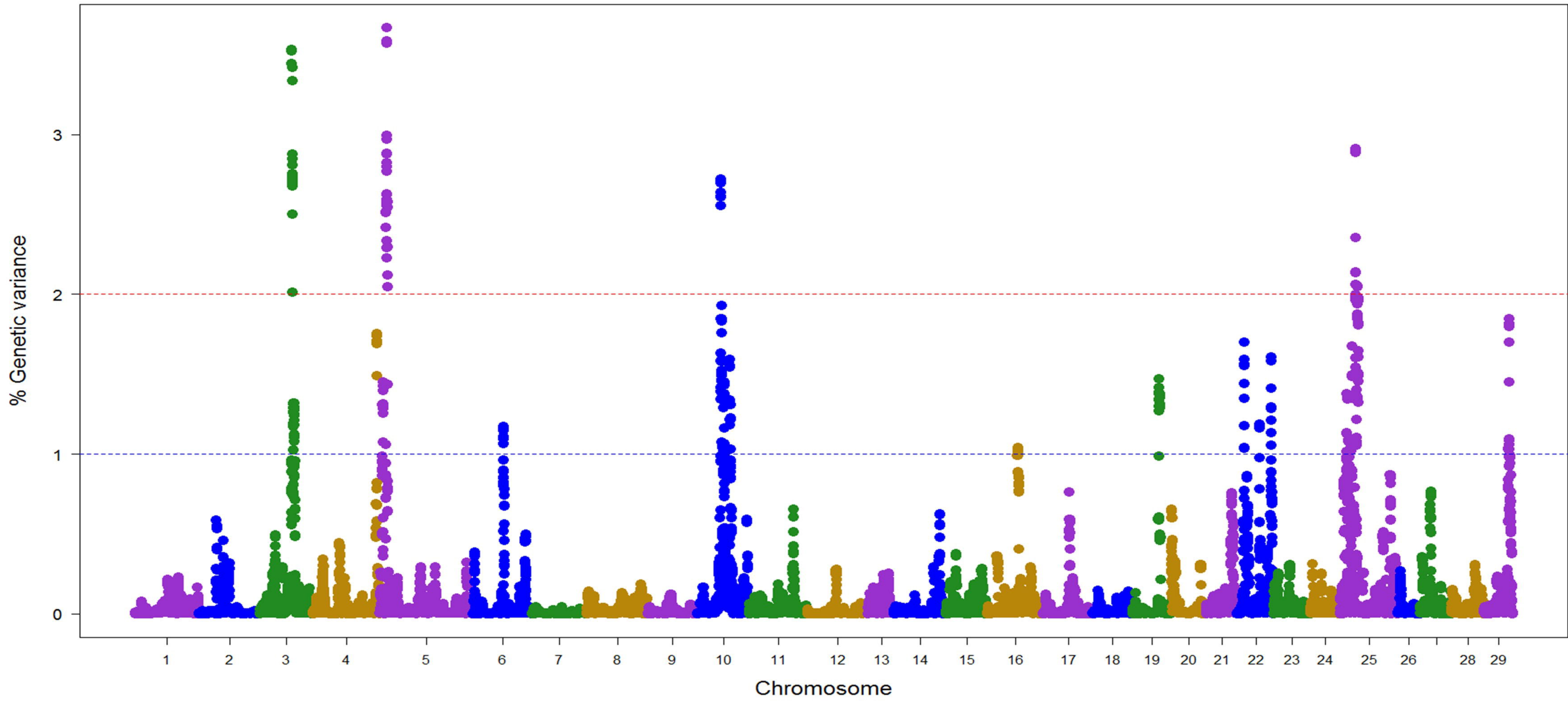


Fig. 1

(A)



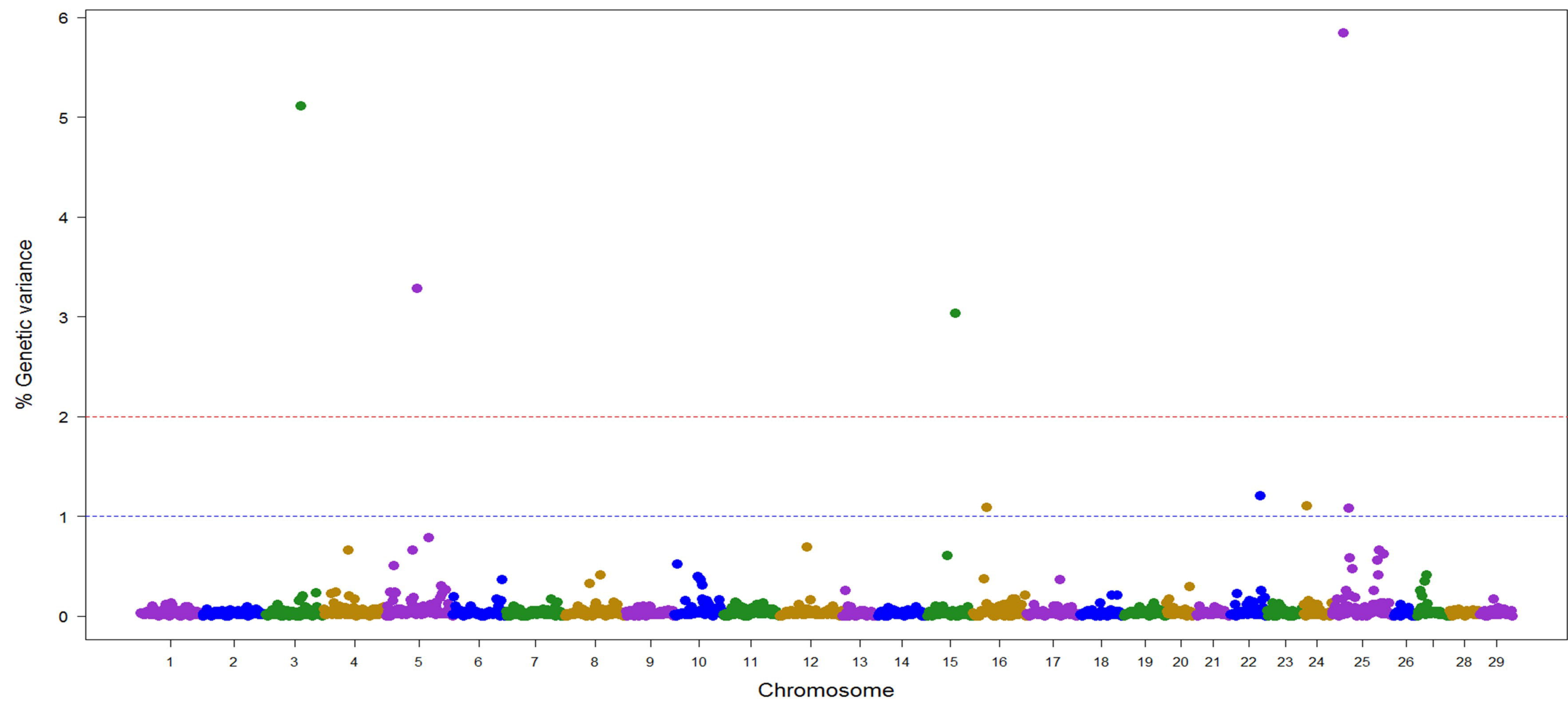
(B)



bioRxiv preprint doi: <https://doi.org/10.1101/162954>; this version posted July 17, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 2

(A)



(B)

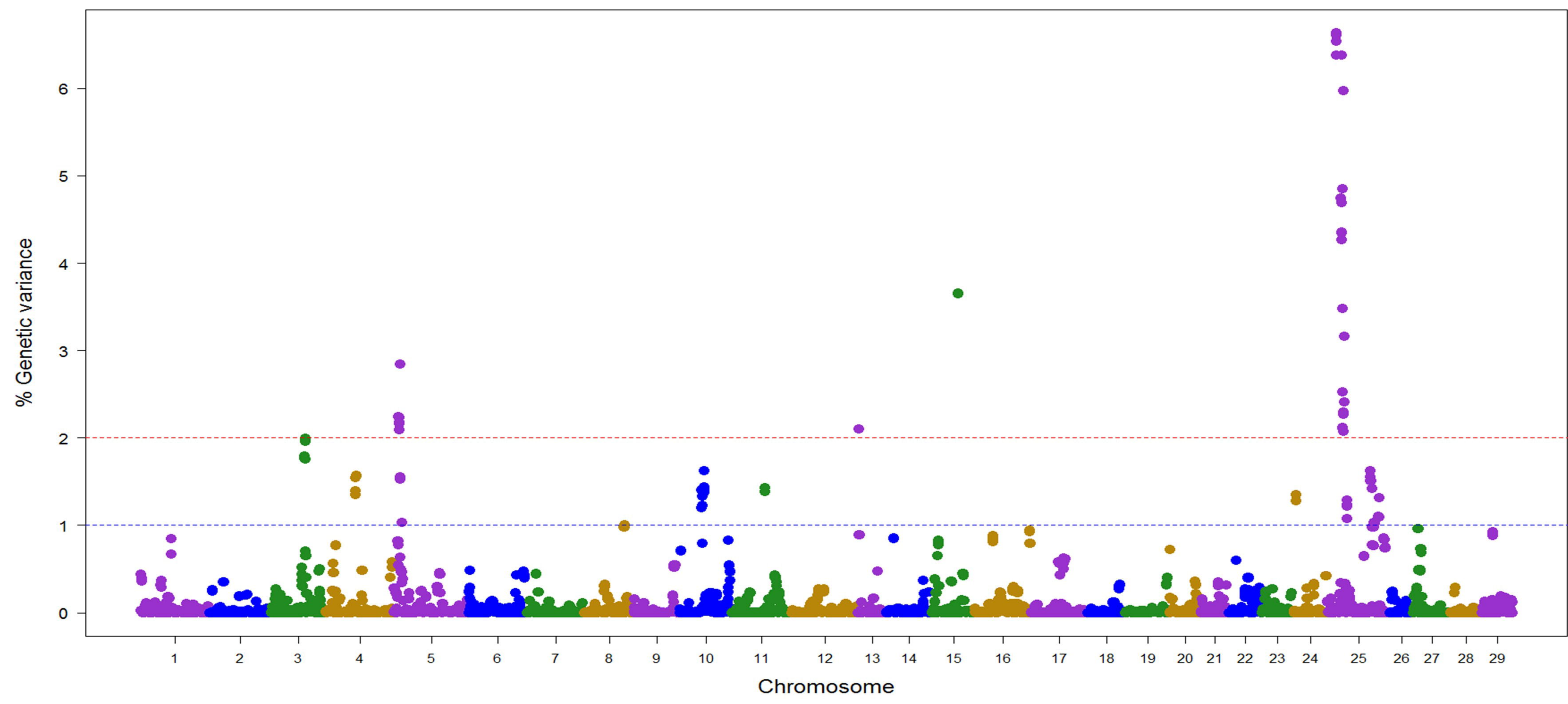
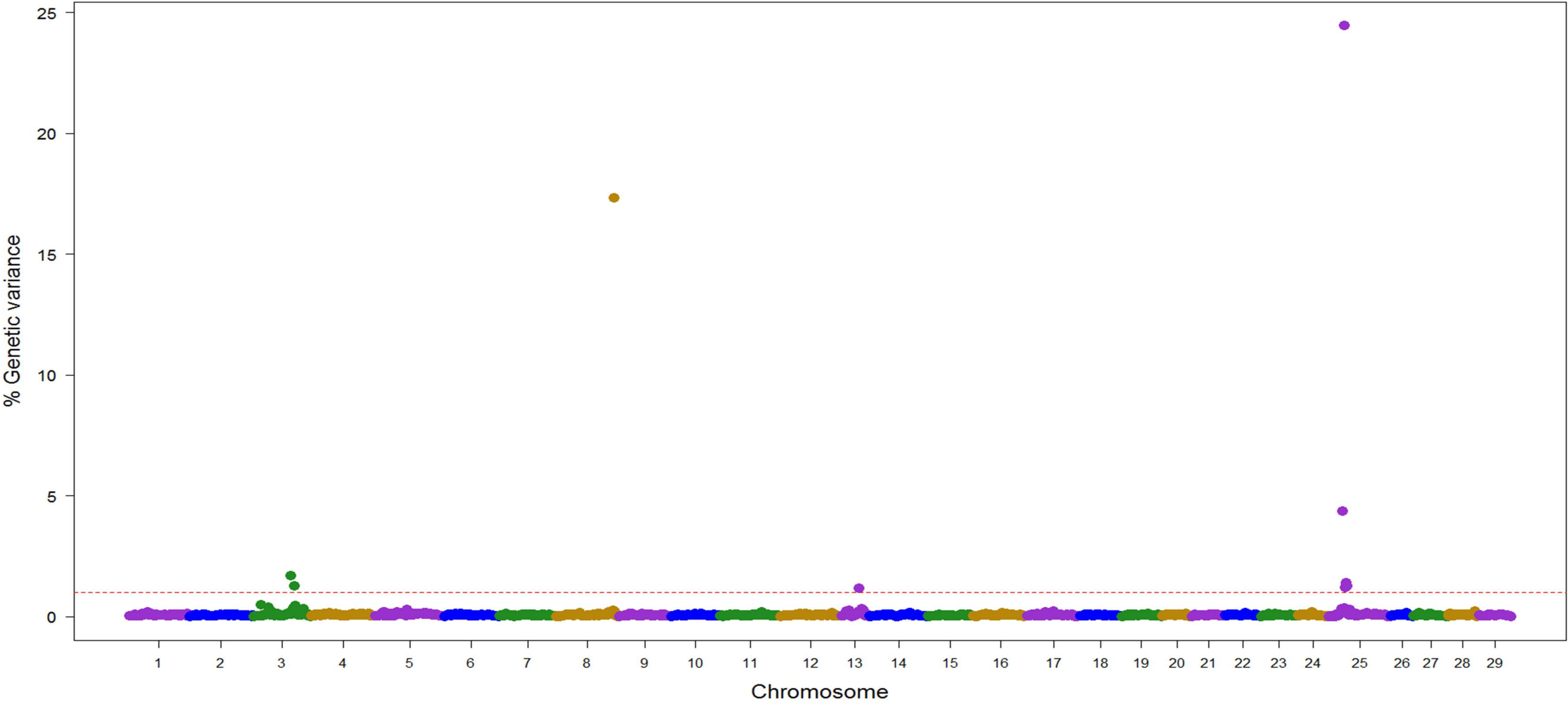


Fig. 3

(A)



(B)

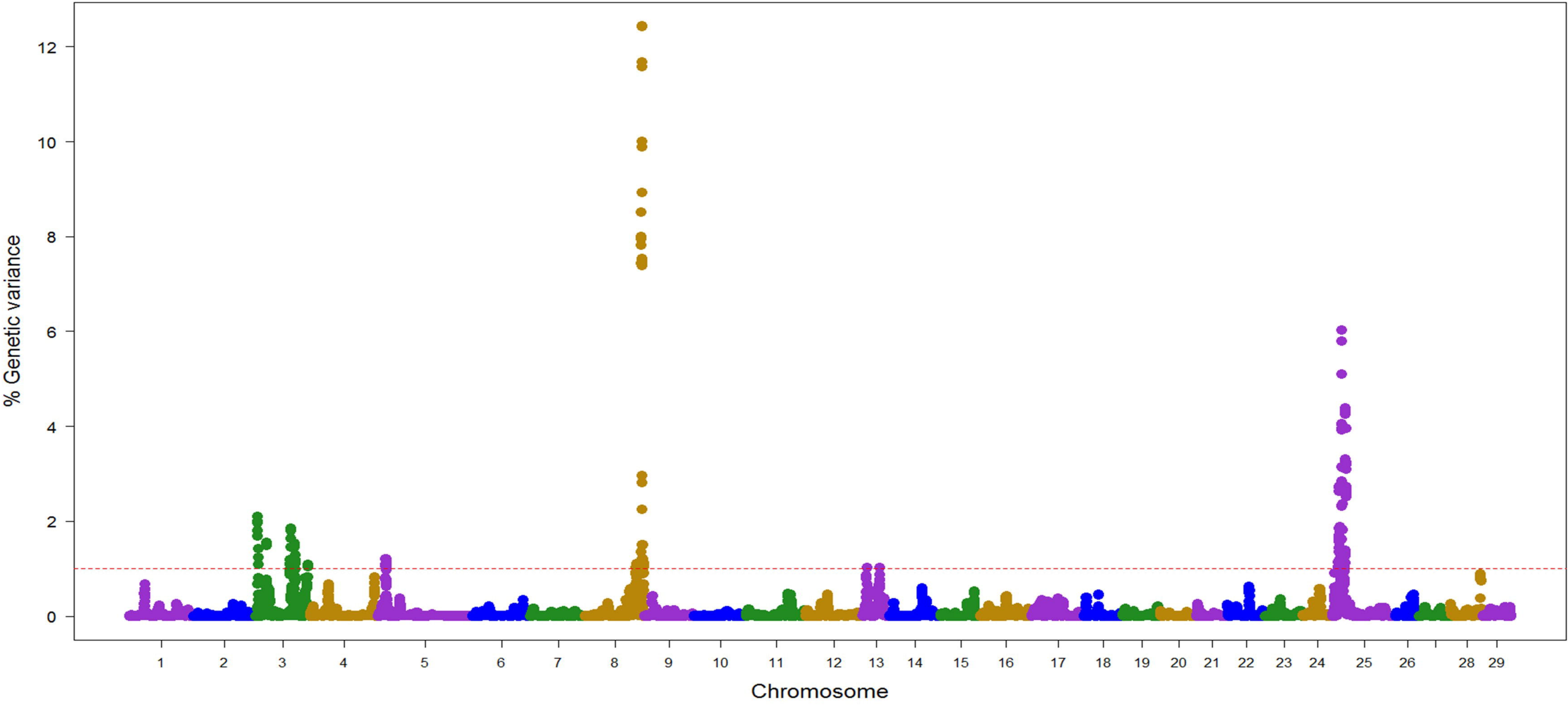
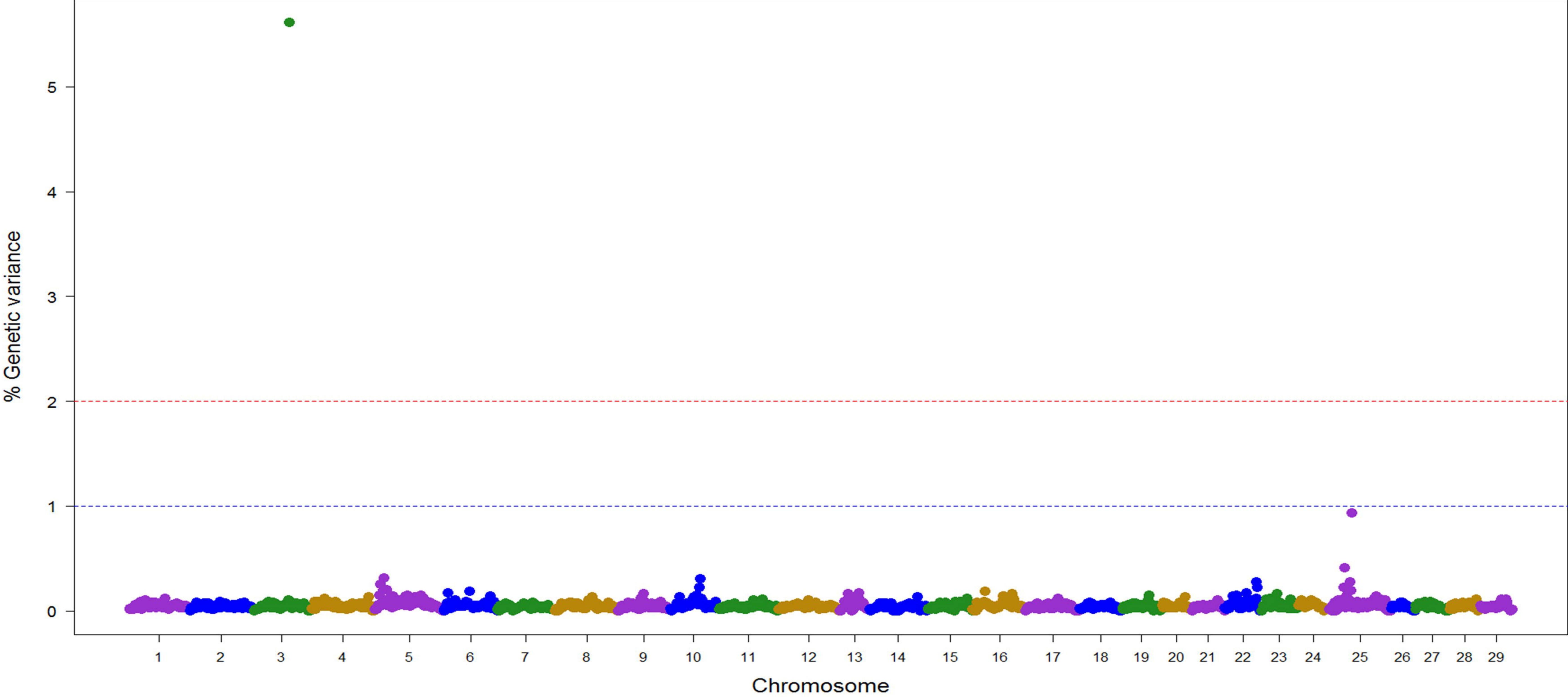
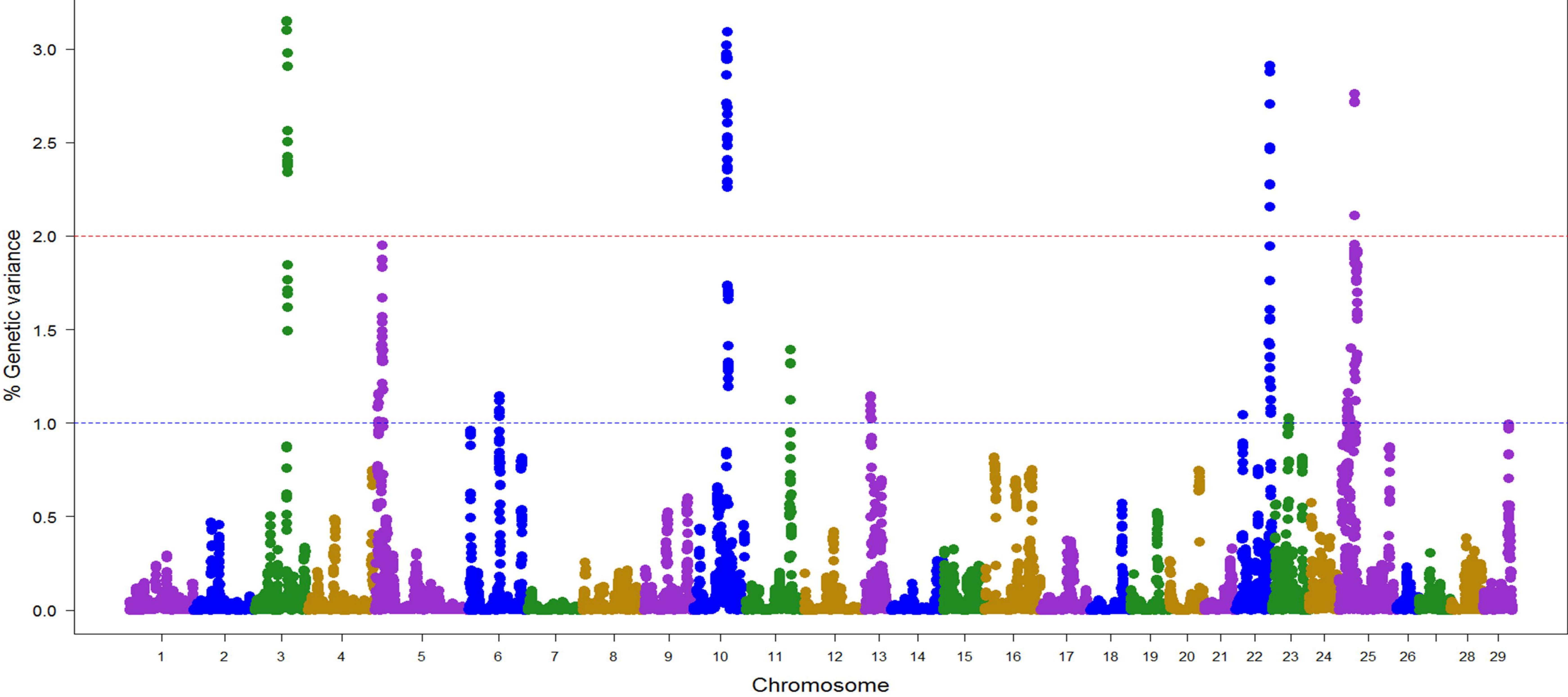


Fig. S1

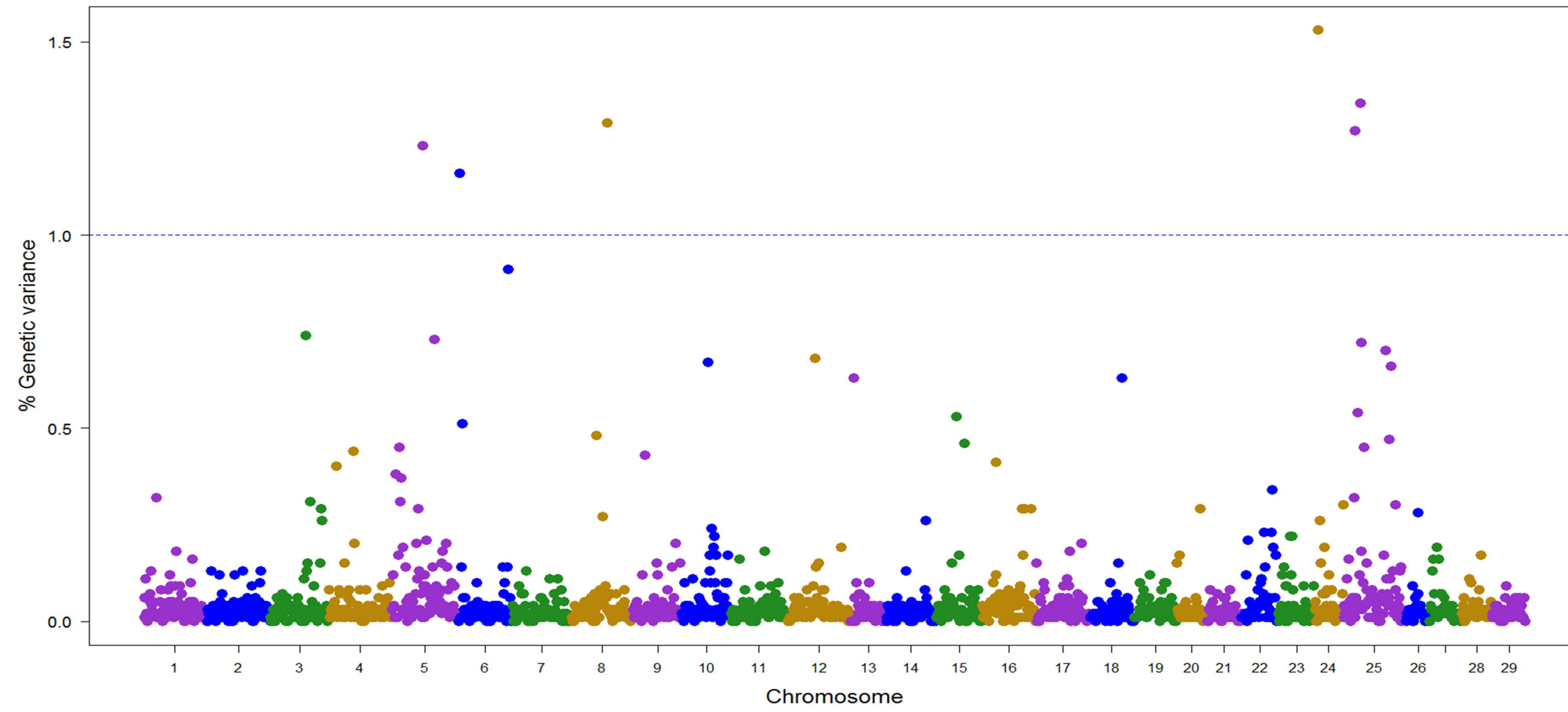
(A)



(B)



(A)



(B)

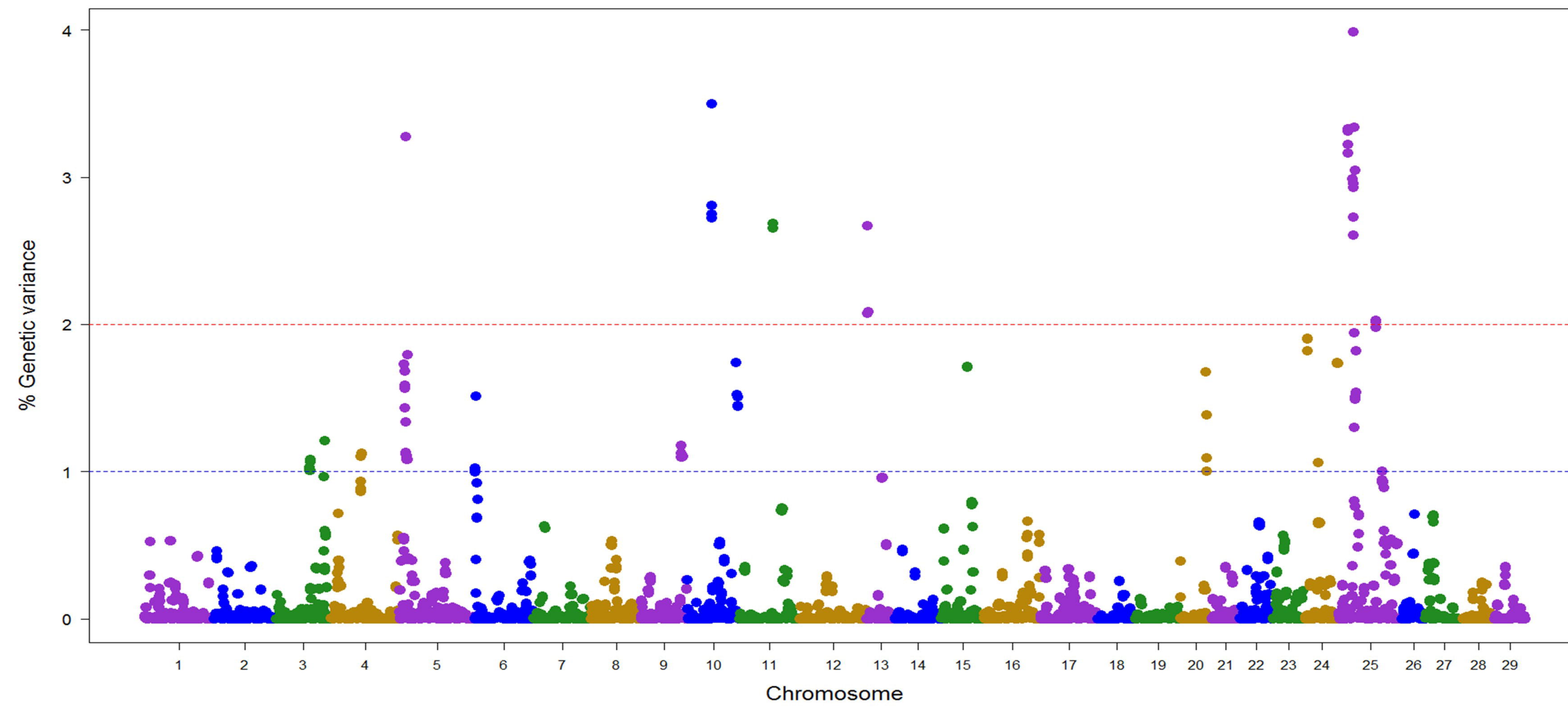


Fig. S3