International Journal of Veterinary Sciences and Animal Husbandry

ISSN: 2456-2912
VET 2023; SP-8(2): 01-05

www.veterinarypaper.com
Received: 01-12-2022
Accepted: 05-01-2023

Komal
Ph.D. Scholar, Department of Animal Genetics and Breeding, LUVAS, Hisar, Haryana, India

Amandeep
Ph.D. Scholar, Department of Livestock Production Management, LUVAS, Hisar, Haryana, India

Ritu
Ph.D. Scholar, Department of Animal Nutrition, LUVAS, Hisar, Haryana, India

Nidhi Sukhija
Ph.D. Scholar, Division of Animal Genetics and Breeding, ICAR-NDRI, Karnal, Haryana, India

Kanaka KK
Ph.D. Scholar, Division of Animal Genetics and Breeding, ICAR-NDRI, Karnal, Haryana, India

Corresponding Author:
Amandeep
Ph.D. Scholar, Department of Livestock Production Management, LUVAS, Hisar, Haryana, India

A review on genome wide association studies for mastitis in livestock species

Komal, Amandeep, Ritu, Nidhi Sukhija and Kanaka KK

Abstract
Genome-wide association studies (GWAS) have been extensively used to identify genetic variants affecting production traits in a variety of livestock species. One of the biggest obstacles in performing GWAS is accurately selecting the most promising single-nucleotide polymorphisms (SNPs). A huge advantage of SNPs as genetic markers is their relatively low mutation rate (1×10^{-9} for SNPs and 1×10^{-5} for microsatellites), even distribution across all chromosomes, ease of implementation in high-throughput genotyping technologies and simple results standardization procedures. The standard disease and quantitative trait association tests can be carried out for linear and logistic regression models allowing for multiple binary or continuous covariates having both the main effects and interactions.Breeding plans can be created to enhance the frequency of advantageous alleles in the population by taking advantage of the identification of genomic areas that have favourable effects on mastitis incidence by virtue of expression of immune system related genes.

Keywords: GWAS, marker, SNPs, logistic, binary, genomic, immune

1. Introduction
Using high throughput genotyping technology, genome-wide association studies (GWAS) link hundreds of SNPs to observable traits and clinical problems (Pearson and Manolio, 2008) [1]. Contrary to gene expression research, where comparatively smaller sample sizes are sufficient, GWA investigations need for a large sample size (Ziegler *et al.,* 2008) [2]. GWAS have been successful in identifying the genetic variants consistently associated with the common complex studies, but explaining the amount of heritability by these risk alleles still faces a limitation, which may be caused by genetic heterogeneity and small sample sizes (Waltoft *et al.*, 2015) [3]. Breeding plans can be created to enhance the frequency of advantageous alleles in the population by taking advantage of the identification of genomic areas that have quantitative effects on a trait. This is especially crucial for qualities with low h^2 that are difficult to routinely document, including disease resistance (Tiezzi *et al.*, 2015) [4]. Although GWAS have benefits such as not requiring prior knowledge of gene function and not requiring assumptions about the type of variant involved (Hirschhorn *et al.*, 2005) [5], they also have significant drawbacks, such as the possibility of false-positive and false-negative results and the lack of SNP variation in the population being studied (Pearson and Manolio, 2008) [1]. Replication of results in independent samples has been an important method to identify true-positives from false-positives (Chanock *et al.*, 2007) [6]. GWAS have been extensively used to identify genetic variants and QTL affecting production traits in a variety of livestock species (Jiang *et al.*, 2010; Bush and Moore, 2012; Li *et al.*, 2011) [7, 8, 9]. GWAS are a potent tool for examining the genetic architecture of complex traits. Due to a high number of loci having minor effects (Visscher *et al.*, 2012; Wood *et al.*, 2014; Fang *et al.*, 2017) [10, 11, 12], classical GWAS have only partially succeeded in illuminating the genetic architecture of complex traits (Deng *et al.*, 2019) [13]. Gene-based or pathway-based GWAS, which combines genetic information for all SNPs in a gene or pathway to enhance the capability to find novel genes and subsequently generate more valuable and informative results (Neale and Sham, 2004; Wang *et al.*, 2010; Xia *et al.*, 2017) [14, 15, 16] proved to be a novel approach to overcome this limitation. However, there are certain problems that make this research more challenging, such as differences in enrichment results between software tools, bias in enrichment analyses caused by pathway membership, and some unidentified relationships between linked genes (Deng *et al.*, 2019) [13].

Genome Wide Association Studies (GWAS)

Since its introduction by Klein and colleagues in 2005 [17], GWAS has been widely used to study a variety of human disease conditions, and its use in the livestock industry is expanding as well (Freebern *et al.*, 2020; Pegolo *et al.*, 2018; Wang *et al.*, 2013; Wray *et al.*, 2008) [18, 19, 20, 21]. Although GWAS have benefits, such as the lack of assumptions about the type of variant involved and the lack of prior knowledge about gene function, they also have significant drawbacks, such as the potential for false-positive and false-negative results and the lack of SNP variation in the population under study (Pearson and Manolio, 2008) [1]. Replication of results in separate samples has been a key tactic in separating true-positives from false-positives. The correct p-value threshold for statistical significance must be determined in order to distinguish between actual false positive and false negative results (Kaler *et al.*, 2019) [22]. Despite the fact that a conventional p-value criterion of 5×10^{-8}has been recommended for more than 1 million SNPs examined, it has not been revised for lower allele frequency spectrum utilised in many recent GWAS and sequencing studies being undertaken globally (Fadista *et al.*, 2016) [23]. The Table 2.1 shows different GWAS conducted for mastitis in various livestock species.

Table 1: GWAS conducted for mastitis

S. No	Species	No. of Samples	Reference(s)
1.	HF cows	103585	Tiezzi *et al.*, 2015 [4]
2.	Chinese Holstein cattle	2093	Wang *et al.*, 2015 [24]
3.	HF cows	40	Siebert *et al.*, 2017 [25]
4.	Nordic Holstein cattle	5147	Cai *et al.*, 2018 [26]
5.	Danish HF cows	993	Welderufael *et al.*, 2018 [27]
6.	Chinese Holstein cows	383	Yang *et al.*, 2019 [28]
7.	HF cows	224	Kurz *et al.*, 2019 [29]
8.	German Black Pied cattle	1062	Meier *et al.*, 2020 [30]
9.	Portuguese Holstein cattle	1338	Silva *et al.*, 2020 [31]
10.	Holstein cows	471	Miles *et al.*, 2021 [32]
11.	Dutch HF	266	Lee *et al.*, 2021 [33]
12.	Local dairy sheep	1813	Sutera *et al.*, 2021 [34]
13.	Assaf sheep	1894	Oner *et al.*, 2021 [35]
14.	Murrah Buffalo	96	Jaglan *et al.*, 2023 [36]

Correction for multiple testing in GWAS

Different statistical techniques, including as the Bonferroni correction, False Discovery Rate, Permutation test, and Bayesian approaches, have been proposed to account for repeated testing in GWAS (Kaler *et al.*, 2019) [22]. Additionally, a different method uses Genomic Control corrected p values, which are based on the idea of genome-wide significance (Bush and Moore, 2012) [8]. The Bonferroni correction is the most conservative one because it assumes that every genetic variant tested is independent of the other genetic variants being tested (Kaler *et al.*, 2019) [22]. All of these techniques reduce type-1 errors (false positives), but to a certain extent they also increase type-2 errors (false negatives).

Single Nucleotide Polymorphisms (SNPs)

SNP markers are DNA sequences that only contain one base changed, with a typical alternative of two nucleotides at a certain place. The least frequent allele must have a frequency of 1% or more in order for such a base location with sequence alternatives in genomic DNA to be regarded as an SNP (Vignal *et al.*, 2002) [37]. SNPs are often bi-allelic systems, which indicates that a population typically has only two alleles. As a result, compared to multiallelic microsatellite markers, the information value per SNP marker is smaller (Beuzen *et al.*, 2000) [38]. They make up for their lower information carriage relative to other markers by being widely and densely dispersed across genomes. Additionally, SNPs have an advantage over microsatellites in terms of genetic markers due to their relatively low mutation rate (1×10^{-9} for SNPs vs $1 \times 10^{-5}$for microsatellites), even chromosomal distribution, ease of implementation for high-throughput genotyping technologies, and straightforward result standardisation procedures (Gurgul *et al.*, 2019) [39]. Correctly identifying the most promising single-nucleotide polymorphisms (SNPs) for follow-up is one of the greatest challenges of conducting genome-wide association studies (GWAS). Several association tests are conducted simultaneously across the genome and the resultant *p*-values or test statistics compared and usually ranked to obtain the most promising SNPs (Tabangin *et al.*, 2009) [40]. Conserved non-coding regions (CNCs) present in introns and intergenic regions are related to regulation of gene function (Yang *et al.*, 2018) [41], however, the variations found in these regions do not directly alter the amino acid sequence, but is influencing the regulation of gene expression and indirectly has some effects on biological functions and disease occurrence in mammals (Patrushev and Kovalenko, 2014) [42].

Genome Wide SNPs Identification

The complete genome must be sequenced in order to identify SNPs across the genome, which will reveal a great deal of repetitive, meaningless information. Additionally, the bioinformatics assembly is highly difficult and expensive. Technically, whole genome sequencing is not required for SNP genotyping because genetic markers within a gene or genomic region might have linkage disequilibrium levels as high as 95% to 100%. Instead, tagged SNPs will be chosen for data analysis. Jiang *et al.* (2012) [43] when using an appropriate NGS technology platform, RAD-seq approaches for SNP identification can be 35 times less expensive than whole genome sequencing methods (Davey *et al.*, 2011) [44].

Genomic Sub-Sampling

Genomic Subsampling refers to the study of a portion of a particular organism's entire genome. To choose a portion of a genome for second-generation sequencing analysis, two basic strategies have been devised. The first method enriches the re-sequencing target for particular regions of interest that are chosen either by PCR amplification or by hybridization to complementary oligonucleotides (Albert *et al.*, 2007; Okou *et al.*, 2007) [45, 46]. However, the oligonucleotide libraries

significantly increase the overall costs of this method. For species with limited accessible genome sequence, a different strategy is based on deep sequencing of libraries of reduced complexity built from pooled DNA samples that reflect populations of interest (Van Tassell *et al.,* 2008) [47]. This strategy effectively develops highly reliable SNP sets for species with limited available genome sequence.

Association analysis in PLINK

PLINK is a publicly available and widely used software for genomic data manipulation and analysis (Toth *et al.,* 2021) [48] and is the most comprehensive of all freely available analytic toolkits when it comes to quality control and analytic modules. It fulfills two analytic needs which are – aids in process of performing quality control (QC) on large data sets and secondly, provides basic statistical tools for analyzing the variants in genetic models (Slifer *et al.,* 2018) [49]. For data management in PLINK, a compact binary file format is there to which represents SNP data. The PLINK file format consists of white-space (tab)-delimited, set of *ped* and *map* files. Within the "*.ped*" file, the first column corresponds to the population ID, the second to the individual ID. The "*.map*" file contains four columns, the first one is the chromosome number, second one is SNP ID, the third is the genetic position of the SNP, and the fourth one denotes the physical position (bp) within the RAD loci. For the association analysis to be performed in PLINK, firstly the .ped and .map files have to be converted into the binary format files which includes three files namely - fam, bam and bed files. There are tools as well to transform the binary format to standard text-based formats (Purcell *et al.,* 2007) [50].

Standard summary measures are available in PLINK including allele and genotype frequencies, genotyping rates, Hardy-Weinberg equilibrium tests using asymptotic and exact procedures as well as single-SNP Mendelian error summaries for family data (Purcell *et al.,* 2007) [50]. The standard disease and quantitative trait association tests can be carried out for linear and logistic regression models allowing for multiple binary or continuous covariates having both the main effects and interactions. PLINK offers a powerful, user-friendly tool for performing many common analyses with whole-genome data (Purcell *et al.,* 2007) [50] as well as for ddRAD data (Cilingir *et al.,* 2021; Lang *et al.,* 2020; Zhou *et al.,* 2019) [51, 52, 53].

Visualizing GWAS results

Common methods for visualizing GWAS results are the two plots namely- "Manhattan plot" and "Q-Q plot". A Manhattan plot is a plot in which $-log_{10}$ (*P*-value) of the association statistic are plotted on the y-axis while the chromosomal positions of the SNPs on the x-axis (Turner, 2014) [54] and each dot on the Manhattan plot denotes a SNP. As the strongest associations have the smallest p-values, their negative logarithms will be the greatest. Hence, regions with many highly associated SNPs in the linkage equilibrium are plotted as "skyscrappers" along the plot. The different colors of each block usually show the extent of each chromosome.

Q-Q (quantile-quantile) plot displays the observed association p-value for all SNPs on y-axis while the expected uniformly distributed p-values under the null hypothesis of no association are plotted on x-axis. Any deviation of the SNPs from the diagonal at upper right end of the plot signifies the strong association of those SNPs, while systematic deviation from the diagonal may indicate population stratification or cryptic relatedness problems in the data (Turner, 2014) [54]. A Q-Q plot is a probability plot, a graphical method for comparing two probability distributions by plotting their quantiles against each other.

Conclusion

Mastitis incidence can be reduced with the help of genetic marker-assisted selection for mastitis features because it produces greater uniformity and phenotypic discrimination than traditional selection. Genetic markers for complex traits like mastitis can be found through genome-wide association studies, which allow for the genotyping of a large number of putative genetic markers, such as single nucleotide polymorphisms (SNPs), throughout the genome. GWAS are now more manageable thanks to recent improvements in our knowledge of genomic variation and the technology that assess it.

References

1. Pearson TA, Manolio TA. How to interpret a genome-wide association study. Jama. 2008;299(11):1335-1344.
2. Ziegler A, König IR, Thompson JR. Biostatistical aspects of genome-wide association studies. Biometrical Journal: Journal of Mathematical Methods in Biosciences. 2008;50(1):8-28.
3. Waltoft BL, Pedersen CB, Nyegaard M, Hobolth A. The importance of distinguishing between the odds ratio and the incidence rate ratio in GWAS. BMC Medical Genetics. 2015;16(1):1-0.
4. Tiezzi F, Parker-Gaddis KL, Cole JB, Clay JS, Maltecca C. A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. PLoS One. 2015;10(2):e0114919.
5. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nature reviews genetics. 2005;6(2):95-108.
6. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, *et al*. Replicating genotype-phenotype associations. Nature. 2007;447:655-660.
7. Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q. Genome wide association studies for milk production traits in Chinese Holstein population. PloS one. 2010;5(10):e13661.
8. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS computational biology. 2012;8(12):e1002822.
9. Li H. Improving SNP discovery by base alignment quality. Bioinformatics. 2011;27(8):1157-1158.
10. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. The American Journal of Human Genetics. 2012;90(1):7-24.
11. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, *et al*. Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics. 2014;46(11):1173-1186.
12. Fang L, Sahana G, Su G, Yu Y, Zhang S, Lund MS, *et al*. Integrating sequence-based GWAS and RNA-Seq provides novel insights into the genetic basis of mastitis and milk production in dairy cattle. Scientific reports. 2017;7(1):1-6.
13. Deng T, Liang A, Liang S, Ma X, Lu X, Duan A, *et al*. Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. Frontiers in genetics. 2019;10:36.

14. Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. The American Journal of Human Genetics. 2004;75(3):353-362.
15. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nature Reviews Genetics. 2010;11(12):843-854.
16. Xia J, Fan H, Chang T, Xu L, Zhang W, Song Y, *et al*. Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle. Scientific reports. 2017;7(1):1-9.
17. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, *et al*. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308(5720):385-389.
18. Freebern E, Santos DJ, Fang L, Jiang J, Parker Gaddis KL, Liu GE, *et al*. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC genomics. 2020;21:1-1.
19. Pegolo S, Mach N, Ramayo-Caldas Y, Schiavon S, Bittante G, Cecchinato A. Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. Scientific reports. 2018;8(1):566.
20. Wang J, Wang H, Chi R, Guo J, Wu Y. Progresses in research of genome-wide association studies in livestock and poultry. Sci. Agric. Sin. 2013;46:819-829.
21. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk of complex disease. Current opinion in genetics & development. 2008;18(3):257-263.
22. Kaler AS, Purcell LC. Estimation of a significance threshold for genome-wide association studies. BMC genomics. 2019; 20:1-8.
23. Fadista J, Manning AK, Florez JC, Groop L. The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. European Journal of Human Genetics. 2016;24(8):1202-1205.
24. Wang X, Ma P, Liu J, Zhang Q, Zhang Y, Ding X, *et al*. Genome-wide association study in Chinese Holstein cows reveal two candidate genes for somatic cell score as an indicator for mastitis susceptibility. BMC genetics. 2015;16(1):1-9.
25. Siebert L, Staton ME, Oliver SP, Pighetti GM. Investigation of Bos taurus CXCR1 gene identifies loci with potential mastitis implications using GWAS and linkage analyses. Identifying genome associations with unique mastitis phenotypes in response to intramammary in response to intramammary Streptococcus uberis Streptococcus uberis challenge challenge, 2017, 86.
26. Cai Z, Guldbrandtsen B, Lund MS, Sahana G. Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. BMC genomics. 2018;19(1):1-1.
27. Welderufael BG, Løvendahl P, De Koning DJ, Janss LL, Fikse WF. Genome-wide association study for susceptibility to and recoverability from mastitis in Danish Holstein cows. Frontiers in genetics. 2018;9:141.
28. Yang F, Chen F, Li L, Yan L, Badri T, Lv C *et al*. Three novel players: PTK2B, SYK, and TNFRSF21 were identified to be involved in the regulation of bovine mastitis susceptibility via GWAS and post-transcriptional analysis. Frontiers in immunology. 2019;10:1579.
29. Kurz JP, Yang Z, Weiss RB, Wilson DJ, Rood KA, Liu GE *et al*. A genome-wide association study for mastitis resistance in phenotypically well-characterized Holstein dairy cattle using a selective genotyping approach. Immunogenetics. 2019;71:35-47.
30. Meier S, Arends D, Korkuć P, Neumann GB, Brockmann GA. A genome-wide association study for clinical mastitis in the dual-purpose German Black Pied cattle breed. Journal of dairy science. 2020;103(11):10289-10298.
31. Silva AA, Silva DA, Silva FF, Costa CN, Silva HT, Lopes PS, *et al*. GWAS and gene networks for milk-related traits from test-day multiple lactations in Portuguese Holstein cattle. Journal of Applied Genetics. 2020; 61(3):465-476.
32. Miles AM, Posbergh CJ, Huson HJ. Direct phenotyping and principal component analysis of type traits implicate novel QTL in bovine mastitis through genome-wide association. Animals. 2021;11(4):1147.
33. Lee YL, Takeda H, Costa Monteiro Moreira G, Karim L, Mullaart E, Coppieters W *et al*. A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle. PLoS Genetics. 2021;17(7):e1009331.
34. Sutera AM, Moscarelli A, Mastrangelo S, Sardina MT, Di Gerlando R, Portolano B *et al*. Genome-wide association study identifies new candidate markers for somatic cells score in a local dairy sheep. Frontiers in Genetics. 2021;12:643531.
35. Öner Y, Serrano M, Sarto P, Iguácel LP, Piquer-Sabanza M, Estrada O *et al*. Genome-wide association studies of somatic cell count in the Assaf breed. Animals. 2021;11(6):1531.
36. Jaglan K, Ravikumar D, Sukhija N, George L, Alex R, Vohra V, *et al*. Genomic clues of association between clinical mastitis and SNPs identified by ddRAD sequencing in Murrah buffaloes. Animal Biotechnology; c2023. p. 1-9.
37. Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics selection evolution. 2002;34(3):275-305.
38. Beuzen ND, Stear MJ, Chang KC. Molecular markers and their use in animal breeding. The Veterinary Journal. 2000;160(1):42-52.
39. Gurgul A, Miksza-Cybulska A, Szmatoła T, Jasielczuk I, Piestrzyńska-Kajtoch A, Fornal A *et al*. Genotyping-by-sequencing performance in selected livestock species. Genomics. 2019;111(2):186-195.
40. Tabangin ME, Woo JG, Martin LJ. The effect of minor allele frequency on the likelihood of obtaining false positives. InBMC proceedings. BioMed Central. 2009;3:1-4.
41. Yang F, Chen F, Li L, Yan L, Badri T, Lv C, *et al*. GWAS using 2b-RAD sequencing identified three mastitis important SNPs via two-stage association analysis in Chinese Holstein cows. bioRxiv. 2018, 434340.
42. Patrushev LI, Kovalenko TF. Functions of noncoding sequences in mammalian genomes. Biochemistry (Moscow). 2014;79:1442-69.
43. Jiang Q, Wang Z, Moore SS, Yang RC. Genome-wide analysis of zygotic linkage disequilibrium and its components in crossbred cattle. BMC genetics. 2012;13:1-3.
44. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature

Reviews Genetics. 2011;12(7):499-510.
45. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, *et al*. Direct selection of human genomic loci by microarray hybridization. Nature methods. 2007;4(11):903-905.
46. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. Microarray-based genomic selection for high-throughput resequencing. Nature methods. 2007;4(11):907-909.
47. Van Tassell CP, Smith TP, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT *et al*. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature methods. 2008;5(3):247-252.
48. Tóth EG, Köbölkuti ZA, Cseke K, Kámpel JD, Takács R, Tomov VT, *et al*. A genomic dataset of single-nucleotide polymorphisms generated by ddRAD tag sequencing in Q. petraea (Matt.) Liebl. populations from Central-Eastern Europe and Balkan Peninsula. Annals of Forest Science. 2021;78(2):43.
49. Slifer SH. PLINK: key functions for data analysis. Current protocols in human genetics. 2018;97(1):e59.
50. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics. 2007;81(3):559-575.
51. Çilingir FG, Hansen D, Ozgul A, Grossen C. Design of SNP markers for Aldabra giant tortoises using low coverage ddRAD-seq. Conservation Genetics Resources. 2021;13:409-412.
52. Lang PL, Weiß CL, Kersten S, Latorre SM, Nagel S, Nickel B, *et al*. Hybridization ddRAD-sequencing for population genomics of nonmodel plants using highly degraded historical specimen DNA. Molecular Ecology Resources. 2020;20(5):1228-1247.
53. Zhou Z, Han K, Wu Y, Bai H, Ke Q, Pu F, *et al*. Genome-wide association study of growth and body-shape-related traits in large yellow croaker (*Larimichthys crocea*) using ddRAD sequencing. Marine Biotechnology. 2019;21:655-670.
54. Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. Biorxiv. 2014 May 14:005165.