Aquaculture Reports 20 (2021) 100691

Contents lists available at ScienceDirect

Aquaculture Reports

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

Application of a novel 50K SNP genotyping array to assess the genetic diversity and linkage disequilibrium in a farmed Pacific white shrimp (*Litopenaeus vannamei*) population

Baltasar F. Garcia [a,b], Álvaro Bonaguro [c], Carolina Araya [b], Roberto Carvalheiro [a,d], José M. Yáñez [b,*]

[a] School of Agricultural and Veterinary Sciences, UNESP - São Paulo State University, Jaboticabal, 14884900, Brazil
[b] Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, 8820000, Chile
[c] OPUMARSA, Santa Elena, 241702, Ecuador
[d] National Council for Scientific and Technological Development (CNPq), Brasília, 71605001, Brazil

ARTICLE INFO

Keywords:
Shrimp
SNP array
Linkage disequilibrium
Effective population size

ABSTRACT

The inclusion of genomic information became a reality in shrimp breeding and it is expected to accelerate the genetic gain over time. The decay of linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs) is an important measure to evaluate the feasibility of implementing genomic selection. The aim of this study was to evaluate the use of a novel 50 K SNP array tool to characterize the genomic diversity, LD and effective population size (N_e) in a farmed shrimp population. A total of 96 animals (40 sires and 56 dams) were genotyped using the novel Illumina AquaArray HD (50 K) vannamei®. Quality control (QC) of genomic data was performed and three different minor allele frequency (MAF) exclusion thresholds were applied: < 0.10 (QC1), < 0.05 (QC2) and < 0.01 (QC3). After QC, 34,425, 39,091 and 42,789 SNPs were retained for QC1, QC2 and QC3, respectively, validating the high informativeness of this SNP array to this particular shrimp breeding population. The population showed a considerable high overall heterozygosity in comparison to other aquaculture species meaning that genetic diversity is stable despite selection. The principal component analysis revealed three genetically distant groups with the first two principal components explaining 27.7 % of total variation. LD decayed rapidly in the first 30Kb of distance between markers from 0.20 to 0.07 and then decreased to 0.02 in the long-range distance. These results suggest a relatively recent incorporation of animals from different populations in the broodstock. N_e size reduced from 7,871 to 301 animals in 827 generations for QC1, 8,253 to 305 animals for QC2 and 8,957 to 315 animals for QC3, both in 899 generations. Contemporary N_e was close to 86 for all QCs. The level of LD estimated suggests that genomic selection and genome-wide association studies are feasible in shrimp by using this SNP array.

1. Introduction

The Pacific white shrimp (*Litopenaeus vannamei*) is the most produced shrimp world-wide, accounting for approximately 83 % of all shrimp and prawn farmed in 2018 (FAO, 2020). In the Americas, *L. vannamei* farming is present mainly in Brazil, Mexico and Ecuador accounting for more than 80 % of all shrimp produced in this continent in 2018 (FAO, 2020). A strategy to increase overall production in shrimp and other aquaculture species is the implementation of breeding programs (Gjedrem, 2005). Several studies showed expressive improvement (around 5–15 % by generation) in economic relevant traits in shrimp through selection, such as, growth (Argue et al., 2002; Campos-Montes et al., 2013) and resistance to diseases (Hernández-Ruíz et al., 2020; Trang et al., 2019).

As in other aquaculture species, most of breeding programs in shrimp farming are based on recording phenotypic and pedigree information to estimate breeding values (Neira, 2010). With the advent of genomic technologies, genomic information has also been included in breeding programs paving the way for which has been called genomic selection (GS) (Meuwissen et al., 2001). GS is feasible when a sufficient number

* Corresponding author.
E-mail address: jmayanez@uchile.cl (J.M. Yáñez).

https://doi.org/10.1016/j.aqrep.2021.100691
Received 13 November 2020; Received in revised form 4 March 2021; Accepted 6 April 2021
Available online 16 April 2021

(thousands) of single nucleotide polymorphism (SNP) markers, distributed all over the genome, are available, allowing to capture genetic variability through linkage disequilibrium (LD) with quantitative trait loci (QTL) (Goddard, 2009).

The use of genomic information may allow higher accuracy in estimating the genetic merit (i.e. genomic estimated breeding values) in comparison to traditional pedigree methods, thus accelerating the rate of genetic gain (Correa et al., 2017; Ødegård et al., 2014; Vallejo et al., 2017; Yoshida et al., 2019b, 2018). GS has special importance for traits that are difficult and/or expensive to measure (e.g. resistance to pathogens and product quality traits), because it explores the relationship between phenotyped and selection candidates more accurately (Calus and Veerkamp, 2011). Another application of genomic information is related to genome-wide association studies (GWAS). This approach may help to unravel the genetic architecture of traits and detect genomic regions associated to genetic variation by using LD information between markers and QTLs. Besides, the LD pattern may be used to understand historical processes and evolutionary forces that affected the population over time (e.g. artificial/natural selection, bottlenecks, cross-breeding) (Heuertz et al., 2006). Thus, the level and extension of LD may help to determine the potential of a specific SNP panel to predict SNP effects in the implementation of GS as well as mapping of QTLs in GWAS and also provide important information about population genetic diversity (Johnson et al., 2010).

Different methods were proposed to evaluate LD in populations, among them, r^2 is the most common measure (Pritchard and Przeworski, 2001). By definition, the LD is the non-independent association between two loci, resulting in frequency of occurrence distinct that the expected by chance. In case of no recombination between these alleles, there is complete LD. LD in farmed populations has been subject of study in livestock as, for example, in dairy and beef cattle (Bohmanova et al., 2010; Espigolan et al., 2013); pigs (Ai et al., 2013; Amaral et al., 2008; Du et al., 2007), sheep (Al-Mamun et al., 2015; Prieur et al., 2017), chicken (Andreescu et al., 2007; Fu et al., 2015) and goats (Berihulay et al., 2019; Brito et al., 2015; Mdladla et al., 2016). LD has also been investigated in aquaculture species as, for example, in Atlantic salmon (Barría et al., 2018; Gutierrez et al., 2015; Hayes et al., 2006; Kijas et al., 2017); coho salmon (Barría et al., 2019); rainbow trout (Rexroad and Vallejo, 2009; Vallejo et al., 2018); Nile tilapia (Yoshida et al., 2019a) and Pacific oyster (Zhong et al., 2017).

For shrimp, Du et al. (2010) and Ciobanu et al. (2010) reported the first results of LD estimation using 1,344 and 1,221 putative SNPs, respectively. Jones et al. (2017) used the first 6.4 K SNP array for shrimp to estimate the mean r^2 of adjacent SNP pairs as 0.15 for distances lower than 1 cM. More recently, Wang et al. (2019), using the 2b-RAD method for discovering and genotyping SNPs in 200 shrimps, estimated LD decay for about 2.6 million of paired SNPs. A novel 50 K SNP array is commercially available for shrimp breeding. Up to date, there are no studies aimed at evaluating the applicability of this tool for the incorporation of genomic information in shrimp breeding populations. In this study, we present the first application of a 50 K SNP to characterize the genetic diversity, estimate the linkage disequilibrium and the effective population size of a breeding population of Pacific white shrimp from Ecuador.

2. Material and methods

2.1. Population and samples

Animals were obtained from the shrimp breeding population owned by Opúsculo del Mar S.A. (Santa Elena, Ecuador). The population was established in 1,999, in which a mass selection scheme has been implemented since then to improve harvest weight for about 20 generations. An average of 140 families are created on each generation by using hierarchical matings (1:2−3 male:female ratio) and the generation interval spans 12 months. A considerable increase in growth rate has been achieved in the population across 20 generations. For instance, the average body weight of females at the age of six months in 1,999 and 2,020 was 29 g and 60 g, respectively. No pedigree information is available for this breeding program. With the objective to evaluate a SNP genotyping tool to incorporate genomics to assist selection, a total of 96 animals (40 sires and 56 females) were randomly sampled in January 2020 and used in this study. Due to unknown information about the families, the relatedness among animals was assessed using genomic relationship through the G matrix (Strandén and Garrick, 2009). The average relatedness was close to 0 (-1.03^{-10} SD = 0.08), with a minimum of -0.09 and maximum of 0.53 (negative relatedness estimate means that the individuals are less related than the average). It is important to mention that these samples have no genetic links with the samples used to develop the 50 K SNP array.

2.2. Genotyping, quality control and genetic diversity

Pleopods of the 96 animals were sampled and used for genomic DNA extraction. Further, genotyping was performed using the novel Illumina AquaArray HD (50 K) vannamei® panel developed by Neogen® (Nebraska, USA) and commercially available through the Center for Aquaculture Technologies (San Diego, USA). This panel uses the assembly ASM378908v1 (GenBank accession GCF_003789085.1; Zhang et al., 2019) of *L. vannamei* as a reference to discovery SNPs.

Quality control of genotypes consisted of four exclusion criteria: call-rate < 0.8 for SNPs and samples, Hardy-Weinberg equilibrium (p-value $< 10^{-5}$) and minor allele frequency (MAF) with different threshold values, forming three quality control sets (QC): < 0.10 (QC1), < 0.05 (QC2) and < 0.01 (QC3). By using different levels of MAF, we investigated the effect of including less frequent alleles in the LD estimation.

The genetic diversity of the population was evaluated by calculating the observed (H_0) and expected (H_E) mean heterozygosity for each QC and using principal component analysis (PCA) on the genomic relationship matrix. The observed heterozygosity for each SNP ($H_{0(SNP)}$) was obtained as: $H_{0(SNP)} = \frac{N_{AB}}{(N_{AA}+N_{AB}+N_{BB})}$, where N_{AA}, N_{AB}, N_{BB} are the number of individuals carrying AA, AB and BB alleles for each SNP, respectively. The expected heterozygosity ($H_{E(SNP)}$) for each SNP was estimated as: $H_{E(SNP)} = 2pq$, where p is the frequency of A allele and q is the frequency of B allele (obtained simply as: $1-p$), respectively. The H_0 and H_E values may be compared to other populations to indicate the level of genetic diversity genome-wide. If H_0 in a breeding population is lower in comparison to a wild population, for instance, there is loss of genetic diversity in the first population and inbreeding may be a cause of such poor genetic variability. On the other hand, a higher value of H_0 indicates high genetic variability due to introduction of meta-populations, for example. A PCA was also performed based on the variance-standardized relationship matrix (Yang et al., 2011) which may be calculated as:

$$A_{jk} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)}$$

where A_{jk} is the genetic relationship between individuals j and k, N is the number of SNPs after quality control, x_{ij} and x_{ik} are the numbers of copies of the reference allele for the ith SNP of the jth individual and kth individual, respectively, and p is the frequency of the reference allele. The building of the genetic relationship allows to investigate the possible existence of subpopulations by plotting the two principal components that explained higher proportions of variation in this matrix. The quality control and genetic diversity parameters described above were evaluated using the PLINK 1.9 software (Purcell et al., 2007).

2.3. LD estimation

Considering two biallelic loci, one with alleles *a* and *A* and other with alleles *b* and *B*, the LD was estimated using the r^2 formula (Hill and Robertson, 1968): $r^2(p_A, p_B, p_{AB}) = \frac{(p_{AB}-p_A p_B)^2}{p_A(1-p_A)p_B(1-p_B)}$, where p_{AB} is the frequency of haplotypes with the allele A at locus 1 and allele B at locus 2, and p_A and p_B are the frequencies of alleles A and B, respectively. The parameters: –ld-window-kb 10,000, –ld-window 99,999 and –ld-window-r2 set to zero were used in PLINK 1.9 to estimate r^2 between all SNPs in each contig for windows of 10Mb maximum size. The minimum and maximum distances between two adjacent SNPs in the LD analysis were 2.21 and 10,000Kb, respectively. Each pair-wise r^2 was included in a determined bin based on the physical distance between syntenic SNPs to generate the LD decay plot.

2.4. Effective population size (N_e)

The historical effective population size (N_e) was estimated using the SNeP v1.1 tool (Barbato et al., 2015). SNeP applies a formula developed by Corbin et al. (2012):

$$N_{et} = \frac{1}{(4f(c_t))}\left(\frac{1}{E\left[r^2_{adj}\middle|c_t\right]} - \alpha\right)$$

where N_{et} is the effective population size t generations ago, c_t is the recombination rate for a specific physical distance between SNPs, r^2_{adj} is the LD adjusted for sample size ($r^2_{adj} = r^2 - (1/sample\ size)$) and α is the adjustment for mutation rate. The c_t component is inferred using mapping functions that translate physical distance into linkage distance. Firstly, the recombination rate between two SNPs is inferred using an approximation: $\delta = kd$, where δ is the physical distance in Mb, k is a constant equals to 10^{-8} and d is the linkage distance in centiMorgan (cM). This approximation is not valid for large values of d due to higher probabilities of recombination and non-linearity between map distance and the recombination rate. To overcome this, we applied a mapping function to modify the c_t developed by Sved (1971):

$$c_t = \frac{1 - \left(\frac{d}{2}\right)}{(1-d)^2}$$

where c_t is the recombination rate between two SNPs and d is the linkage distance previously estimated. We considered $\alpha = 2$, meaning that mutation does occur. This is a theoretical value that assumes mutations occurring at different time points on the same chromosome which has been previously derived by Ohta and Kimura (1971). The N_e was estimated between SNPs within a distance of 0 to 5Mb and the number and size of each bin was 30 and 50 Kb, respectively. Due to the low number of SNPs per contig, the mean N_e was expressed using harmonic means calculated for each bin (Alvarenga et al., 2018). In order to estimate the contemporary N_e, we used the software NeEstimator v2.01 (Do et al., 2014). First, the genotype files in PLINK format were transformed to GENEPOP format (Rousset, 2008) using the R software (R Core Team, 2017) through the radiator package (Gosselin et al., 2020). Then, the LD method was chosen assuming non-random mating and a critical value for removing rare alleles of 0.05 to obtain the contemporary N_e considering each quality control.

3. Results

3.1. SNP array validation and quality control

The initial genotype dataset had 50,811 SNPs from 96 animals. After quality control 6 animals were removed due to low call-rate (<0.80). The number of SNPs that were not present in at least 80 % of animals was only 3,570 showing the high quality of genotyping of samples using the novel 50 K array (Table 1). It is possible to notice that MAF was the filter that eliminated most SNPs (11,075, 7,374 and 3,664, for QC1, QC2 and QC3, respectively). The final numbers of SNPs after each quality control was 34,425, 39,091 and 42,789 for QC1, QC2 and QC3, respectively. There was small variation in MAF levels regarding the quality controls (Fig. 1). Comparing the two most contrasting scenarios (QC1 *vs* QC3), the proportion of SNPs loci ranged from 1.85 to 4.44 % considering all MAF classes. The mean MAF was 0.31 ± 0.12 for QC1, 0.28 ± 0.13 for QC2 and 0.26 ± 0.14 for QC3. In addition, a large proportion of SNPs have MAF > 0.2, showing high level of polymorphism present in the studied population and captured by the SNP array. The number of SNPs that were eliminated by more than one criterion is shown in the supplementary material.

3.2. Genetic diversity

Observed and expected heterozygosity values are showed in Table 1. The observed and expected values were very similar for all QCs and the overall heterozygosity seems to be at an acceptable level despite selection for 20 generations. As expected, a reduction in the H_E is noticed as MAF decreases, i.e. the proportion of homozygous loci increases when SNPs with lower MAF are included. This reduction in the heterozygosity was equal to 6.4 and 13.5 % for QC2 and QC3 in comparison to QC1, respectively. For the PCA, all QCs produced similar results, thus, we presented results of QC1 only. The two most important principal components explained 27.7 % of the genomic relationship variation among animals in this population (Fig. 2). Three clusters may be observed in the PCA: a larger group of 75 animals (top-left), a small group of 7 females (bottom-centre) and a last small group with 8 animals of both sexes (top-right). The first principal component differentiates the three clusters (14.4 % of variation) and the second component differentiates the female group from the other two groups (13.3 % of variation).

3.3. LD and N_e estimation

The estimates of LD for each contig are shown in Table 2. Overall, LD levels varied among contigs ranging from 0.21 to 0.26, considering all adjacent SNPs in each contig and quality controls. The MAF had minor effect on LD estimations within each contig. The LD decay (Fig. 3) revealed a rapid decrease of r^2 as the physical distance between markers increased. The most significant drop was observed in the first 30Kb of distance and then the r^2 level slightly decreased to values close to 0.02. Regarding the different QCs, the inclusion of SNPs with less frequent alleles (QC2 and QC3) seemed to decrease the r^2 in the short-range

Table 1
Summary of number of markers included after different quality controls (QC) performed in the 50 K SNP genotypes in a farmed shrimp population.

QC[1]	Call-rate[2]		MAF[3]	HWE[4]	Heterozygosity[5]	
	Samples	SNPs (%)*	SNPs (%)	SNPs (%)	H_O	H_E
QC1	90	47,241 (93.0)	36,166 (71.2)	35,425 (69.2)	0.4041	0.3994
QC2	90	47,241 (93.0)	39,867 (78.5)	39,091 (76.9)	0.3783	0.3744
QC3	90	47,241 (93.0)	43,577 (85.8)	42,789 (84.2)	0.3496	0.3461

[1] Quality control applied (differences only in MAF filter).
[2] Number of SNPs and samples included after the exclusion criterion call-rate < 0.8.
[3] Number and proportion of SNPs included by minor allele frequency (MAF) threshold: QC1 (MAF > 0.10), QC2 (MAF > 0.05), and QC3 (MAF > 0.01).
[4] Number and proportion of SNPs in Hardy–Weinberg Equilibrium (HWE) ($p > 10^{-5}$).
[5] Heterozygosity observed (H_O) and expected (H_E).
* All percentages are in comparison to the initial number of SNPs (50,811).

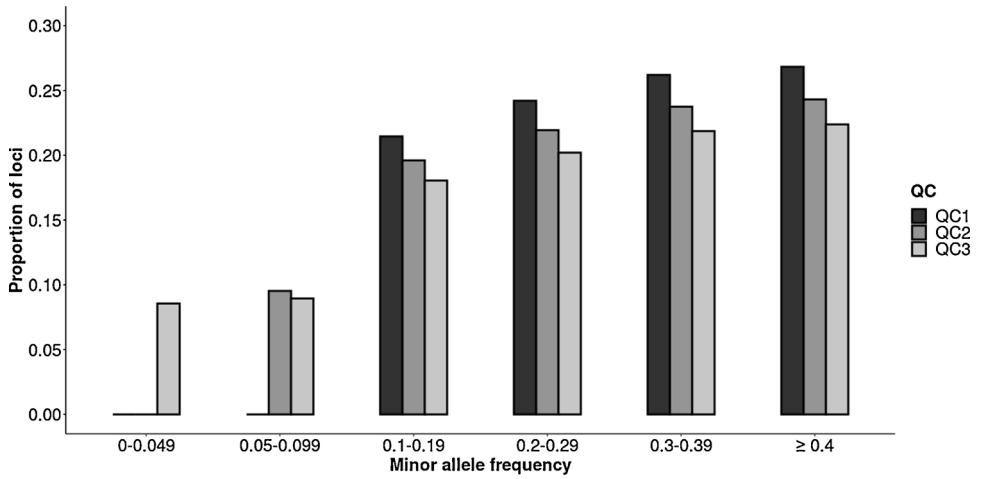

Fig. 1. Proportion of SNPs in different minor allele frequency (MAF) intervals after distinct quality controls (QC) using three MAF exclusion criteria thresholds: QC1 (< 0.10), QC2 (< 0.05) and QC3 (< 0.01).

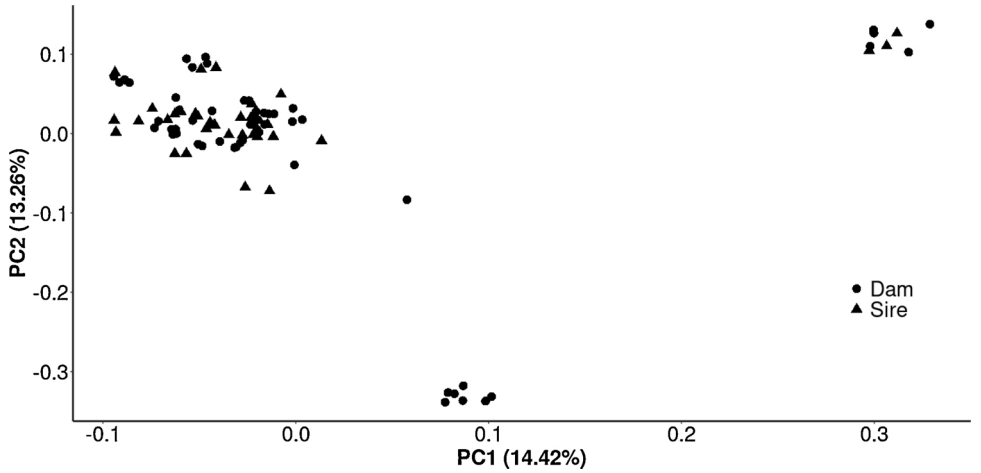

Fig. 2. Genetic structure of a farmed shrimp population based on principal component analysis of the genomic relationship matrix (G). Sires and dams are identified by different shapes. Only SNPs with minor allele frequency ≥ 0.10 were considered to compute G.

distances, compared to QC1, but all had similar pattern of LD decay in the long-range distances.

The historic N_e using the LD method showed a decrease in size of 7,871 to 301 animals in 827 generations for QC1, 8,253–305 animals in 899 generations for QC2 and 8,957 to 315 animals in 899 generations for QC3 (Fig. 4). The most significant decrease was observed in the last 300 generations with approximately 62 % of total reduction in this period. The contemporary N_e estimated was very similar among all quality controls 85.0, 86.6 and 86.7 for QC1, QC2 and QC3, respectively.

4. Discussion

4.1. Genetic diversity

This study shows the first application of a 50 K SNP genotyping array to a farmed shrimp population. The panel presents 50,811 SNPs, which after quality control yielded 69.2, 76.9 and 84.2 % out of the total SNPs for QC1, QC2 and QC3, respectively. Furthermore, the proportion of SNPs with MAF > 0.2 was close to 70 % with a mean value of 0.3 for all QCs. These results show the high adequacy of this novel tool for this particular farmed shrimp population from Ecuador.

The evaluation of genetic diversity is important because it may help to assess the genetic variability of broodstock, genetic "bottlenecks", level of inbreeding and show perspectives for genomic selection (Yáñez et al., 2015). A possible effect of mass selection is the reduction of overall heterozygosity through generations. Our results showed that overall heterozygosity is at a satisfactory level and its reduction overtime was likely slow in this population. Despite genetic diversity has been studied in shrimp using different types of genetic markers (see Benzie, 2000 for a review), few studies were found using SNPs (Ciobanu et al., 2010; Lu et al., 2018; Perez-Enriquez et al., 2018) and none study was found characterizing a farmed population using dense genomic information becoming difficult to compare our results with previous reports in literature. Indeed, the comparison of the results obtained with wild *L. vannamei* or other commercial populations would be more adequate to clearly visualize the effect of selection on genetic diversity. However, expected heterozygosity found in the present study (0.35−0.40) is in the upper level to what has been found in other farmed aquaculture species such as, Tilapia (0.28−0.36) (Yáñez et al., 2020; Yoshida et al., 2019a), sea trout (0.27) (Drywa et al., 2013) and Atlantic salmon (0.22−0.38) (Kijas et al., 2017).

A possible explanation for the considerable overall heterozygosity may be the high genetic variability present in the population. Likely, there was inclusion of animals from different strains in this broodstock recently. This assumption may be verified by the PCA results which display three distinct groups. We expect that these groups represent subpopulations rather than families because hybridization with other wild and farmed populations could not be discarded. Although inbreeding is

Table 2
Estimated average linkage disequilibrium ($r^2 \pm$ standard deviation), number of SNPs and mean distance between adjacent SNPs (Dist. $\pm$ standard deviation in Mb), for a farmed shrimp population. Values are expressed for each contig after distinct quality controls (QC) using different minor allele frequency (MAF) exclusion criteria thresholds: QC1 (< 0.10), QC2 (< 0.05) and QC3 (< 0.01).

Contig	QC1			QC2			QC3		
	$r^2 \pm$ SD	SNPs	Dist. $\pm$ SD	$r^2 \pm$ SD	SNPs	Dist. $\pm$ SD	$r^2 \pm$ SD	SNPs	Dist. $\pm$ SD
1	0.021 ± 0.035	1401	4.88 ± 2.87	0.021 ± 0.035	1508	4.88 ± 2.88	0.020 ± 0.034	1660	4.88 ± 2.88
2	0.021 ± 0.034	1333	4.88 ± 2.88	0.021 ± 0.033	1441	4.88 ± 2.88	0.020 ± 0.032	1557	4.88 ± 2.88
3	0.021 ± 0.033	1173	4.86 ± 2.88	0.021 ± 0.033	1309	4.85 ± 2.88	0.020 ± 0.032	1415	4.85 ± 2.88
4	0.021 ± 0.035	1090	4.85 ± 2.88	0.021 ± 0.034	1198	4.85 ± 2.88	0.021 ± 0.033	1296	4.85 ± 2.88
5	0.021 ± 0.032	1028	4.83 ± 2.87	0.021 ± 0.031	1124	4.84 ± 2.87	0.020 ± 0.031	1212	4.84 ± 2.87
6	0.022 ± 0.033	980	4.82 ± 2.88	0.021 ± 0.033	1080	4.82 ± 2.88	0.021 ± 0.032	1169	4.82 ± 2.88
7	0.022 ± 0.033	948	4.83 ± 2.88	0.021 ± 0.033	1042	4.83 ± 2.88	0.021 ± 0.033	1130	4.83 ± 2.88
8	0.021 ± 0.034	945	4.80 ± 2.88	0.021 ± 0.033	1025	4.80 ± 2.88	0.020 ± 0.032	1093	4.81 ± 2.88
9	0.022 ± 0.033	898	4.81 ± 2.87	0.022 ± 0.033	977	4.81 ± 2.87	0.021 ± 0.032	1065	4.81 ± 2.87
10	0.023 ± 0.035	894	4.80 ± 2.88	0.022 ± 0.034	973	4.80 ± 2.88	0.022 ± 0.034	1064	4.80 ± 2.88
11	0.022 ± 0.036	854	4.80 ± 2.88	0.022 ± 0.035	952	4.81 ± 2.88	0.021 ± 0.034	1022	4.80 ± 2.88
12	0.023 ± 0.040	843	4.82 ± 2.88	0.023 ± 0.039	926	4.81 ± 2.88	0.022 ± 0.037	1013	4.80 ± 2.88
13	0.022 ± 0.034	854	4.80 ± 2.87	0.021 ± 0.034	944	4.80 ± 2.88	0.021 ± 0.033	1021	4.80 ± 2.88
14	0.021 ± 0.034	813	4.80 ± 2.88	0.021 ± 0.033	909	4.79 ± 2.87	0.020 ± 0.032	988	4.80 ± 2.87
15	0.022 ± 0.038	765	4.77 ± 2.87	0.022 ± 0.036	842	4.77 ± 2.87	0.021 ± 0.035	925	4.76 ± 2.87
16	0.025 ± 0.041	705	4.75 ± 2.88	0.024 ± 0.039	784	4.75 ± 2.88	0.023 ± 0.038	856	4.75 ± 2.88
17	0.022 ± 0.032	768	4.77 ± 2.87	0.021 ± 0.032	853	4.76 ± 2.87	0.021 ± 0.031	911	4.76 ± 2.87
18	0.022 ± 0.034	700	4.72 ± 2.87	0.021 ± 0.034	772	4.72 ± 2.87	0.021 ± 0.033	835	4.72 ± 2.87
19	0.021 ± 0.033	736	4.76 ± 2.86	0.021 ± 0.033	810	4.75 ± 2.87	0.021 ± 0.032	865	4.75 ± 2.87
20	0.022 ± 0.035	650	4.74 ± 2.88	0.022 ± 0.034	715	4.72 ± 2.88	0.021 ± 0.033	785	4.72 ± 2.87
21	0.022 ± 0.035	630	4.66 ± 2.87	0.021 ± 0.034	693	4.69 ± 2.86	0.021 ± 0.033	750	4.70 ± 2.87
22	0.023 ± 0.039	637	4.71 ± 2.86	0.022 ± 0.038	694	4.71 ± 2.86	0.021 ± 0.036	763	4.71 ± 2.86
23	0.021 ± 0.038	633	4.72 ± 2.87	0.021 ± 0.036	699	4.71 ± 2.87	0.020 ± 0.035	765	4.71 ± 2.87
24	0.021 ± 0.032	623	4.66 ± 2.86	0.021 ± 0.032	673	4.67 ± 2.86	0.020 ± 0.031	720	4.65 ± 2.86
25	0.021 ± 0.034	597	4.69 ± 2.87	0.021 ± 0.033	654	4.69 ± 2.86	0.021 ± 0.033	702	4.68 ± 2.86
26	0.022 ± 0.037	587	4.71 ± 2.87	0.021 ± 0.036	650	4.72 ± 2.87	0.021 ± 0.035	713	4.71 ± 2.87
27	0.023 ± 0.037	547	4.65 ± 2.86	0.023 ± 0.037	594	4.65 ± 2.86	0.022 ± 0.036	633	4.66 ± 2.86
28	0.022 ± 0.033	484	4.56 ± 2.84	0.022 ± 0.033	528	4.56 ± 2.84	0.021 ± 0.032	573	4.55 ± 2.84
29	0.020 ± 0.033	419	4.52 ± 2.84	0.020 ± 0.032	446	4.52 ± 2.84	0.020 ± 0.032	496	4.52 ± 2.85
30	0.021 ± 0.030	434	4.53 ± 2.84	0.020 ± 0.030	482	4.50 ± 2.84	0.020 ± 0.029	524	4.53 ± 2.85
31	0.020 ± 0.031	424	4.52 ± 2.84	0.020 ± 0.030	468	4.52 ± 2.83	0.019 ± 0.030	504	4.51 ± 2.83
32	0.020 ± 0.031	381	4.43 ± 2.83	0.020 ± 0 .031	429	4.43 ± 2.83	0.019 ± 0.031	467	4.44 ± 2.83
33	0.021 ± 0.031	364	4.37 ± 2.82	0.020 ± 0.030	400	4.38 ± 2.81	0.020 ± 0.030	437	4.39 ± 2.81
34	0.026 ± 0.046	360	4.39 ± 2.81	0.025 ± 0.044	400	4.39 ± 2.81	0.024 ± 0.043	426	4.38 ± 2.81
35	0.021 ± 0.032	339	4.26 ± 2.79	0.020 ± 0.031	363	4.26 ± 2.78	0.020 ± 0.030	393	4.27 ± 2.78
36	0.021 ± 0.035	313	4.22 ± 2.76	0.021 ± 0.034	352	4.24 ± 2.78	0.020 ± 0.033	379	4.23 ± 2.77
37	0.021 ± 0.032	314	4.30 ± 2.80	0.021 ± 0.032	341	4.29 ± 2.79	0.020 ± 0.031	369	4.30 ± 2.79
38	0.022 ± 0.035	328	4.32 ± 2.80	0.021 ± 0.034	353	4.33 ± 2.80	0.020 ± 0.033	388	4.32 ± 2.79
39	0.020 ± 0.033	315	4.27 ± 2.80	0.020 ± 0.032	342	4.28 ± 2.79	0.019 ± 0.031	368	4.27 ± 2.78
40	0.023 ± 0.036	302	4.24 ± 2.79	0.022 ± 0.035	338	4.25 ± 2.80	0.021 ± 0.034	368	4.25 ± 2.79
41	0.021 ± 0.035	270	4.01 ± 2.70	0.021 ± 0.034	294	4.03 ± 2.72	0.020 ± 0.033	318	4.02 ± 2.72
42	0.022 ± 0.041	156	2.57 ± 1.80	0.021 ± 0.039	169	2.58 ± 1.81	0.020 ± 0.037	186	2.60 ± 1.83
43	0.023 ± 0.040	72	1.22 ± 0.85	0.023 ± 0.039	80	1.22 ± 0.86	0.022 ± 0.037	85	1.21 ± 0.84
44	0.026 ± 0.046	60	1.16 ± 0.82	0.025 ± 0.045	65	1.14 ± 0.80	0.024 ± 0.043	69	1.12 ± 0.78

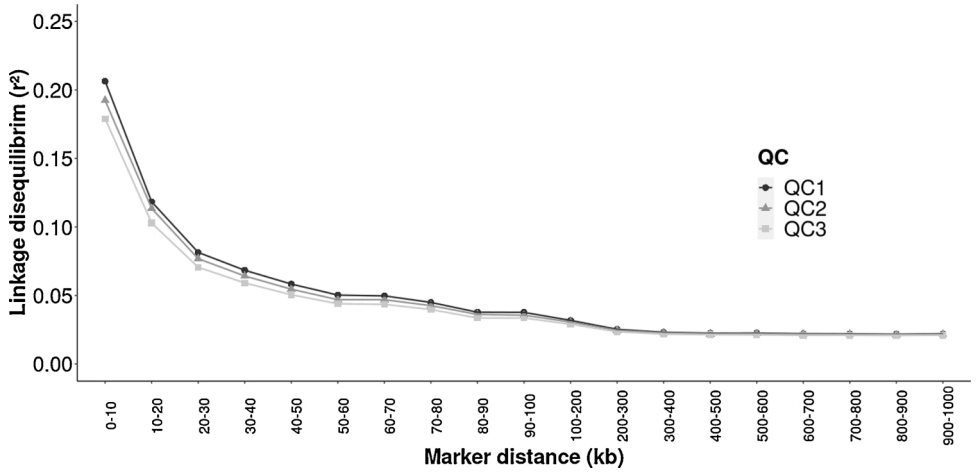

Fig. 3. Estimated average linkage disequilibrium (r^2) over different classes of distance between syntenic markers, for a farmed shrimp population, using different quality controls (QC) with three minor allele frequency exclusion criteria thresholds: QC1 (< 0.10), QC2 (< 0.05) and QC3 (< 0.01).

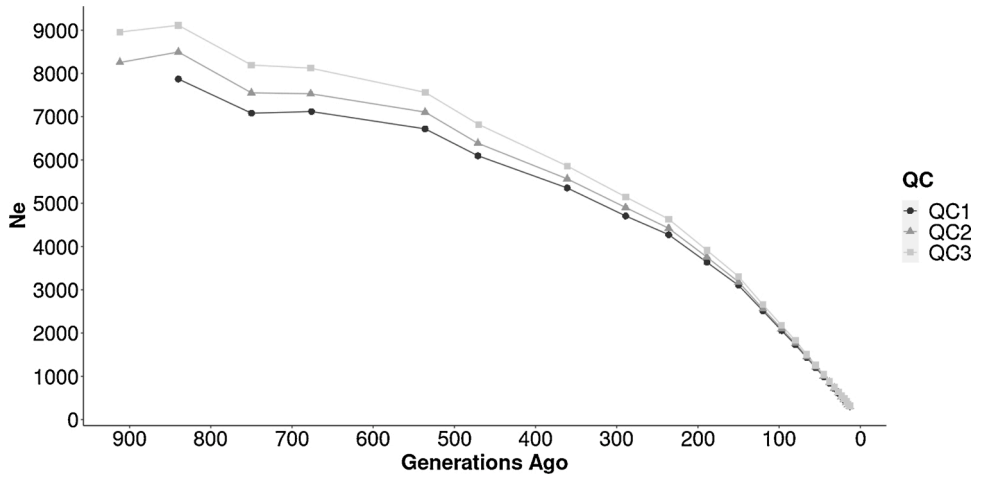

Fig. 4. Effective population size (N_e) estimated in a farmed shrimp population using different quality controls (QC) with three minor allele frequency exclusion criteria thresholds: QC1 (< 0.10), QC2 (< 0.05) and QC3 (< 0.01).

not a direct measure of genetic diversity, the directional selection applied in breeding programs may generate diversity loss and inbreeding depression (i.e. inbred individuals have reduced average performance in comparison to non-inbred) (De Donato et al., 2005). In this population, the recent introduction of animals of different populations or strains allowed a more flexible mating system by broadening options of selection, reducing the chances of reproduction between relatives and controlling inbreeding.

Another aspect that reinforces the consistence of heterozygosity results obtained is associated to the lack of genetic links between the population studied and the population used to discover and build the SNP array. In order to select SNPs to include in the array, a small base population is genome-sequenced and the SNPs are chosen following mainly quality, spacing and frequency as criteria (Malomane et al., 2018). The non-random selection of animals as discovery population and the pre-filter applied on SNPs may generate the ascertainment bias, i.e. the observed allelic frequencies are different than those expected for a random sample (Lachance and Tishkoff, 2013). As a result, measures of genetic variability, such as heterozygosity, in the discovery population may be overestimated (Rogers and Jorde, 1996). Thus, we expect that these parameters were accurately estimated. The inclusion of SNPs by softening the MAF filter (QC2 and QC3) decreased the overall heterozygosity. This result is expected because SNPs with less frequent alleles will increase the proportion of homozygote loci.

4.2. LD and N_e

LD provides important information about previous events that may have changed the population to what is observable on its current status. Our results presented variable r^2 estimations across all 44 contigs (0.21−0.26) with high variation within each contig. This variation most likely happened due to different recombination rates at specific genomic regions (hotspots and coldspots) as a possible result of genetic drift and different genomic structures (Arias et al., 2009; Saunders et al., 2005). Several factors may affect LD, such as admixture, founder effect, inbreeding, mutation and selection (Gaut and Long, 2003). Accuracy of LD estimation depends on demographic and biological factors, as mentioned before, but also on sample size and relationship among animals. Bohmanova et al. (2010) suggested that a satisfactory number of animals to estimate LD is 55 using the r^2 measure. Gutierrez et al. (2015) stated that LD may be overestimated when the animals under study are highly related. The present study used 90 animals and, as explained in the methods section, there is low relationship level among them. Thus, we expect that LD was estimated accurately.

LD was estimated in other shrimp populations (Du et al., 2010; Jones et al., 2017; Wang et al., 2019). However, a direct comparison is inappropriate due to differences in the quality control of genotypes, method of SNP discovery, SNP density and intrinsic characteristics of each population. Most studies of LD in shrimp used low density SNPs arrays (<7 K) which may not represent properly the whole LD considering that the genome of *L. vannamei* has approximately 2,600Mb (Zhang et al., 2019). In addition, the SNP array used in this study offers higher accuracy regarding the SNP-calling process and even distribution of markers across the genome in comparison to genotyping-by-sequencing methods applied in studies that used denser SNPs (Robledo et al., 2018).

The LD decay pattern was different than the one reported by Wang et al. (2019) for a *L. vannamei* population cultivated in China. It is possible to observe that LD at short-range was weaker in our study (~0.2) in comparison to Wang's study (~0.4) and that long-range LD was low (< 0.05) in both studies. The recent admixture of distinct populations is suspected to decrease LD at short-range (Ødegård et al., 2014) also known as admixture linkage disequilibrium (Pfaff et al., 2001). We believe that the level of admixture found within the Ecuadorian population studied here was absent in the Chinese population studied by Wang et al. (2019) generating the difference in the short-range LD. The *L. vannamei* is naturally found in the Pacific coast from Mexico to Northern Peru (Holthuis, 1980) while it was commercially introduced in Asia as an exotic species in 1996 (Biao and Kaijin, 2007; Liao and Chien, 2011). Due to its natural occurrence, the inclusion of animals of different genetic basis or wild strains is recurrent in the Americas (Moss et al., 2012). This "admixture effect" was also described in other species, such as Nile tilapia (Yoshida et al., 2019a) and Salmon (Barría et al., 2018).

Some studies also suggested that the persistency of LD levels through distance (long-range LD) is influenced by admixture as well (Barría et al., 2018; Ødegård et al., 2014; Vallejo et al., 2018). However, the induced LD at long-range was not observed in this population. As suggested by Yoshida et al. (2019a), other biological and demographical factors also play important roles in the LD extension, such as recombination events and effective population size. The MAF filter had minor effect on LD estimation overall. The inclusion of SNPs with less frequent alleles (QC2 and QC3) decreased slightly r^2 at short-range distances. This probably happened because higher MAFs tend to overestimate the level of LD (Espigolan et al., 2013).

The historical N_e decreased considerably overtime with a more pronounced drop in the last 300 generations with a reduction of about 5,000 to 300 in the N_e. However, these results should be interpreted with caution. According to Corbin et al. (2012) the estimation of N_e, in the most recent generations, involves high rates of recombination which may compromise the N_e prediction model derived by Hayes et al.

(2003). In this method, N_e is estimated assuming a relationship between the length of chromosomes segments and number of generations fitting better to lower rates of recombination. Moreover, Santiago et al. (2020) using simulation showed that coalescent methods (such as the one we applied in this study) may produce biased estimations of N_e when populations suffer abrupt reductions or expansions, especially in recent past generations. This likely has occurred in this shrimp population as suggested by the PCA results. This bias could also explain the sharply reduction of N_e size in the more distant past (300 generations ago) rather than when admixture likely occurred (~20 generations ago). Nevertheless, this method is a useful alternative to extract demographic information about unknown events that affected a population and has been widely used for N_e estimation in humans (Tenesa et al., 2007), livestock (Makina et al., 2015) and aquaculture species (Barría et al., 2019, 2018; Yoshida et al., 2019a).

The MAF had major effect in the long-term N_e and QC3 presented higher N_e estimates at early generations in comparison to QC1 and QC2. The N_e is estimated using LD information and the different LD levels in each quality control as well as distinct recombination rates likely caused distinct patterns of early N_e. The contemporary N_e was close to 86 showing that N_e reduced 214 in the last 13 years (assuming one year of generation interval). The contemporary N_e observed is similar to the reported in other shrimp populations using microsatellites markers ranging from 50 to 173 (Cruz et al., 2004; De Donato et al., 2005; Ren et al., 2018; Vela Avitúa et al., 2013). The number estimated in this population is greater than 50 which was recommended by Ponzoni et al. (2010) to keep inbreeding rate limit close to 1% in breeding programs. Thus, our results showed that despite mass selection applied in this population, genetic variability and effective population size are at acceptable levels most likely as a result of recent incorporation of different populations.

4.3. Practical applications

The results of LD obtained here are relevant for the prospect of genomic tools applied to shrimp breeding. The extent of LD may affect GS and GWAS basically in two ways: for GS, it has implications on the accuracy of genomic estimated breeding values, and for GWAS, on the mapping power and resolution. For GS, the predictive ability of QTL effects may be totally dependent of LD when QTLs are not represented in the panel (Kizilkaya et al., 2010). Meuwissen (2009), suggested that, for unrelated individuals, high accuracy of genomic breeding values may be obtained using $2N_eL$ number of animals and $10N_eL$ number of markers, where N_e is the effective population size and L is the length of genome in Morgans. For this study, assuming N_e equals to 85 and that the genome size of *L. vannamei* has approximately 45.325 Morgans (Jones et al., 2017), 7,706 training records and 38,527 SNPs would be necessary to achieve accuracy of 0.88 – 0.93. Although, these formulas are empirical and were obtained using simulation of small populations, the marker density required is close to the obtained after QC2 and QC3 showing the feasibility of GS using this SNP array.

The level of LD observed in the present study shows opportunity to conduct GWAS using this SNP array. However, the QTL mapping power depends on other factors rather than SNP density, such as genetic architecture, heritability of the trait, sample size and statistical method which should be evaluated before performing association analysis.

5. Conclusion

The 50 K SNP array panel applied in this study showed a high proportion of polymorphic loci being a suitable tool to estimate genetic variability, LD and N_e in a farmed shrimp population. The heterozygosity level and PCA results showed recent incorporation of animals from different populations or strains in this broodstock. This study revealed the potential use of a novel SNP array technology to GS and GWAS in shrimp breeding.

Funding

This study was partially funded by OPUMARSA (Santa Elena, Ecuador) and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

Author's contribution

BFG: Conceptualization, Methodology, Formal analysis, Writing - Original Draft and Writing - Review & Editing. AB: Funding acquisition and Conceptualization and Writing - Review & Editing. CA: Methodology and Writing - Review & Editing. RC: Conceptualization, Methodology, Writing - Review & Editing. JMY: Conceptualization, Methodology, Writing - Review & Editing, Supervision and Project administration. All authors contributed to the design of the study, discussion of results and review of the manuscript. All authors read and approved the final version of the manuscript.

Declaration of Competing Interest

AB was hired by OPUMARSA during the course of the study.

Acknowledgments

We would like to thank OPUMARSA for providing genotyping data used in the present study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aqrep.2021.100691.

References

Ai, H., Huang, L., Ren, J., 2013. Genetic diversity, linkage disequilibrium and selection signatures in Chinese and western pigs revealed by genome-wide SNP markers. PLoS One 8, e56001. https://doi.org/10.1371/journal.pone.0056001.

Al-Mamun, H.A., A Clark, S., Kwan, P., Gondro, C., 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47, 90. https://doi.org/10.1186/s12711-015-0169-6.

Alvarenga, A.B., Rovadoscki, G.A., Petrini, J., Coutinho, L.L., Morota, G., Spangler, M.L., Pinto, L.F.B., Carvalho, G.G.P., Mourão, G.B., 2018. Linkage disequilibrium in Brazilian Santa Inês breed, Ovis aries. Sci. Rep. 8, 1–11. https://doi.org/10.1038/s41598-018-27259-7.

Amaral, A.J., Megens, H.J., Crooijmans, R.P.M.A., Heuven, H.C.M., Groenen, M.A.M., 2008. Linkage disequilibrium decay and haplotype block structure in the pig. Genetics 179, 569–579. https://doi.org/10.1534/genetics.107.084277.

Andreescu, C., Avendano, S., Brown, S.R., Hassen, A., Lamont, S.J., Dekkers, J.C.M., 2007. Linkage disequilibrium in related breeding lines of chickens. Genetics 177, 2161–2169. https://doi.org/10.1534/genetics.107.082206.

Argue, B.J., Arce, S.M., Lotz, J.M., Moss, S.M., 2002. Selective breeding of Pacific white shrimp (Litopenaeus vannamei) for growth and resistance to Taura Syndrome Virus. Aquaculture 204, 447–460. https://doi.org/10.1016/S0044-8486(01)00830-4.

Arias, J.A., Keehan, M., Fisher, P., Coppieters, W., Spelman, R., 2009. A high density linkage map of the bovine genome. BMC Genet. 10, 1–12. https://doi.org/10.1186/1471-2156-10-18.

Barbato, M., Orozco-terWengel, P., Tapio, M., Bruford, M.W., 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front. Genet. 6, 109. https://doi.org/10.3389/fgene.2015.00109.

Barría, A., López, M.E., Yoshida, G., Carvalheiro, R., Lhorente, J.P., Yáñez, J.M., 2018. Population genomic structure and genome-wide linkage disequilibrium in farmed atlantic salmon (Salmo salar L.) using dense SNP genotypes. Front. Genet. 9, 649. https://doi.org/10.3389/fgene.2018.00649.

Barría, A., Christensen, K.A., Yoshida, G., Jedlicki, A., Leong, J.S., Rondeau, E.B., Lhorente, J.P., Koop, B.F., Davidson, W.S., Yáñez, J.M., 2019. Whole genome linkage disequilibrium and effective population size in a coho salmon (Oncorhynchus kisutch) breeding population using a high-density SNP array. Front. Genet. 10, 498. https://doi.org/10.3389/fgene.2019.00498.

Benzie, J.A.H., 2000. Population genetic structure in penaeid prawns. Aquac. Res. 31, 95–119. https://doi.org/10.1046/j.1365-2109.2000.00412.x.

Berihulay, H., Islam, R., Jiang, L., Ma, Y., 2019. Genome-wide linkage disequilibrium and the extent of effective population sizes in six chinese goat populations using a 50K single nucleotide polymorphism panel. Animals 9, 350. https://doi.org/10.3390/ani9060350.

Biao, X., Kaijin, Y., 2007. Shrimp farming in China: operating characteristics, environmental impact and perspectives. Ocean Coast. Manag. 50, 538–550. https://doi.org/10.1016/j.ocecoaman.2007.02.006.
Bohmanova, J., Sargolzaei, M., Schenkel, F.S., 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics 11, 1–11. https://doi.org/10.1186/1471-2164-11-421.
Brito, L.F., Jafarikia, M., Grossi, D.A., Kijas, J.W., Porto-Neto, L.R., Ventura, R.V., Salgorzaei, M., Schenkel, F.S., 2015. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. BMC Genet. 16, 1–15. https://doi.org/10.1186/s12863-015-0220-1.
Calus, M.P., Veerkamp, R.F., 2011. Accuracy of multi-trait genomic selection using different methods. Genet. Sel. Evol. 43, 26. https://doi.org/10.1186/1297-9686-43-26.
Campos-Montes, G.R., Montaldo, H.H., Martínez-Ortega, A., Jiménez, A.M., Castillo-Juárez, H., 2013. Genetic parameters for growth and survival traits in Pacific white shrimp Penaeus (Litopenaeus) vannamei from a nucleus population undergoing a two-stage selection program. Aquac. Int. 21, 299–310. https://doi.org/10.1007/s10499-012-9553-1.
Ciobanu, D.C., Bastiaansen, J.W.M., Magrin, J., Rocha, J.L., Jiang, D.-H., Yu, N., Geiger, B., Deeb, N., Rocha, D., Gong, H., Kinghorn, B.P., Plastow, G.S., van der Steen, H.A.M., Mileham, A.J., 2010. A major SNP resource for dissection of phenotypic and genetic variation in Pacific white shrimp (*Litopenaeus vannamei*). Anim. Genet. 41, 39–47. https://doi.org/10.1111/j.1365-2052.2009.01961.x.
Corbin, L.J., Liu, A.Y.H., Bishop, S.C., Woolliams, J.A., 2012. Estimation of historical effective population size using linkage disequilibria with marker data. J. Anim. Breed. Genet. 129, 257–270. https://doi.org/10.1111/j.1439-0388.2012.01003.x.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Correa, K., Bangera, R., Figueroa, R., Lhorente, J.P., Yáñez, J.M., 2017. The use of genomic information increases the accuracy of breeding value predictions for sea louse (Caligus rogercresseyi) resistance in Atlantic salmon (Salmo salar). Genet. Sel. Evol. 49, 1–5. https://doi.org/10.1186/s12711-017-0291-8.
Cruz, P., Ibarra, A.M., Mejia-Ruiz, H., Gaffney, P.M., Pérez-Enríquez, R., 2004. Genetic variability assessed by microsatellites in a breeding program of pacific white shrimp (Litopenaeus vannamei). Mar. Biotechnol. 6, 157–164. https://doi.org/10.1007/s10126-003-0017-5.
De Donato, M., Manrique, R., Ramirez, R., Mayer, L., Howell, C., 2005. Mass selection and inbreeding effects on a cultivated strain of Penaeus (Litopenaeus) vannamei in Venezuela. Aquaculture. Elsevier, pp. 159–167. https://doi.org/10.1016/j.aquaculture.2005.02.005.
Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Resour. 14, 209–214. https://doi.org/10.1111/1755-0998.12157.
Drywa, A., Poćwierz-Kotus, A., Was, A., Dobosz, S., Kent, M.P., Lien, S., Bernaś, R., Wenne, R., 2013. Genotyping of two populations of Southern Baltic Sea trout Salmo trutta m. Trutta using an Atlantic salmon derived SNP-array. Mar. Genom. 9, 25–32. https://doi.org/10.1016/j.margen.2012.08.001.
Du, F.X., Clutter, A.C., Lohuis, M.M., 2007. Characterizing linkage disequilibrium in pig populations. Int. J. Biol. Sci. https://doi.org/10.7150/ijbs.3.166.
Du, Z.-Q., Ciobanu, D.C., Onteru, S.K., Gorbach, D., Mileham, A.J., Jaramillo, G., Rothschild, M.F., 2010. A gene-based SNP linkage map for pacific white shrimp, Litopenaeus vannamei. Anim. Genet. 41, 286–294. https://doi.org/10.1111/j.1365-2052.2009.02002.x.
Espigolan, R., Baldi, F., Boligon, A.A., Souza, F.R.P., Gordo, D.G.M., Tonussi, R.L., Cardoso, D.F., Oliveira, H.N., Tonhati, H., Sargolzaei, M., Schenkel, F.S., Carvalheiro, R., Ferro, J.A., Albuquerque, L.G., 2013. Study of whole genome linkage disequilibrium in Nellore cattle. BMC Genomics 14, 1–8. https://doi.org/10.1186/1471-2164-14-305.
FAO, 2020. FAO Fisheries & Aquaculture - Statistics [WWW Document] (accessed 12.30.19). http://www.fao.org/fishery/statistics/en.
Fu, W., Dekkers, J.C.M., Lee, W.R., Abasht, B., 2015. Linkage disequilibrium in crossbred and pure line chickens. Genet. Sel. Evol. 47, 11. https://doi.org/10.1186/s12711-015-0098-4.
Gaut, B.S., Long, A.D., 2003. The lowdown on linkage disequilibrium. Plant Cell. https://doi.org/10.1105/tpc.150730.
Gjedrem, T., 2005. Breeding Programs. Springer, Dordrecht, The Netherlands.
Goddard, M., 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245–257. https://doi.org/10.1007/s10709-008-9308-0.
Gosselin, T., Lamothe, M., Devloo-Delva, F., Grewe, P., 2020. Radiator: RADseq data exploration. Manipulation and Visualization using R. https://doi.org/10.5281/zenodo.3687060.
Gutierrez, A.P., Yáñez, J.M., Fukui, S., Swift, B., Davidson, W.S., 2015. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic Salmon (Salmo salar). PLoS One 10, e0119730. https://doi.org/10.1371/journal.pone.0119730.
Hayes, B.J., Visscher, P.M., McPartlan, H.C., Goddard, M.E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13, 635–643. https://doi.org/10.1101/gr.387103.
Hayes, B.J., Gjuvsland, A., Omholt, S., 2006. Power of QTL mapping experiments in commercial Atlantic salmon populations, exploiting linkage and linkage disequilibrium and effect of limited recombination in males. Heredity (Edinb.) 97, 19–26. https://doi.org/10.1038/sj.hdy.6800827.
Hernández-Ruíz, H., Montaldo, H.H., Bustos-Martínez, J., Campos-Montes, G.R., Castillo-Juárez, H., 2020. Heritability and genetic correlations for infectious hypodermal and hematopoietic necrosis virus load, body weight at harvest, and survival rate in Pacific white shrimp (*Litopenaeus vannamei*). J. World Aquac. Soc. 51, 312–323. https://doi.org/10.1111/jwas.12664.
Heuertz, M., De Paoli, E., Källman, T., Larsson, H., Jurman, I., Morgante, M., Lascoux, M., Gyllenstrand, N., 2006. Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [Picea abies (L.) Karst]. Genetics 174, 2095–2105. https://doi.org/10.1534/genetics.106.065102.
Hill, W.G., Robertson, A., 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38, 226–231. https://doi.org/10.1007/BF01245622.
Holthuis, L.B., 1980. FAO Species Catalogue: Shrimps and Prawns of the World, 1st ed. Food and Agriculture Organization of the United Nations, Rome, Italy.
Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C. A., O'Brien, S.J., 2010. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics 11, 1–6. https://doi.org/10.1186/1471-2164-11-724.
Jones, D.B., Jerry, D.R., Khatkar, M.S., Raadsma, H.W., Steen, H. Van Der, Prochaska, J., Forêt, S., Zenger, K.R., 2017. A comparative integrated gene-based linkage and locus ordering by linkage disequilibrium map for the Pacific white shrimp, Litopenaeus vannamei. Sci. Rep. 7, 1–16. https://doi.org/10.1038/s41598-017-10515-7.
Kijas, J., Elliot, N., Kube, P., Evans, B., Botwright, N., King, H., Primmer, C.R., Verbyla, K., 2017. Diversity and linkage disequilibrium in farmed Tasmanian Atlantic salmon. Anim. Genet. 48, 237–241. https://doi.org/10.1111/age.12513.
Kizilkaya, K., Fernando, R.L., Garrick, D.J., 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes1. J. Anim. Sci. 88, 544–551. https://doi.org/10.2527/jas.2009-2064.
Lachance, J., Tishkoff, S.A., 2013. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. BioEssays 35, 780–786. https://doi.org/10.1002/bies.201300014.
Liao, I.C., Chien, Y.-H., 2011. The Pacific White Shrimp, Litopenaeus vannamei, in Asia: the world's most widely cultured alien crustacean. The Wrong Place - Alien Marine Crustaceans: Distribution, Biology and Impacts. Springer, Netherlands, pp. 489–519. https://doi.org/10.1007/978-94-007-0591-3_17.
Lu, X., Kong, J., Meng, X., Cao, B., Luo, K., Dai, P., Luan, S., 2018. Identification of SNP markers associated with tolerance to ammonia toxicity by selective genotyping from de novo assembled transcriptome in Litopenaeus vannamei. Fish Shellfish Immunol. 73, 158–166. https://doi.org/10.1016/j.fsi.2017.12.005.
Makina, S.O., Taylor, J.F., van Marle-Köster, E., Muchadeyi, F.C., Makgahlela, M.L., MacNeil, M.D., Maiwashe, A., 2015. Extent of linkage disequilibrium and effective population size in four South African sanga cattle breeds. Front. Genet. 6, 337. https://doi.org/10.3389/fgene.2015.00337.
Malomane, D.K., Reimer, C., Weigend, S., Weigend, A., Sharifi, A.R., Simianer, H., 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. BMC Genomics 19, 22. https://doi.org/10.1186/s12864-017-4416-9.
Mdladla, K., Dzomba, E.F., Huson, H.J., Muchadeyi, F.C., 2016. Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. Anim. Genet. 47, 471–482. https://doi.org/10.1111/age.12442.
Meuwissen, T.H., 2009. Accuracy of breeding values of "unrelated" individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41, 1–9. https://doi.org/10.1186/1297-9686-41-35.
Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.
Moss, S.M., Moss, D.R., Arce, S.M., Lightner, D.V., Lotz, J.M., 2012. The role of selective breeding and biosecurity in the prevention of disease in penaeid shrimp aquaculture. J. Invertebr. Pathol. https://doi.org/10.1016/j.jip.2012.01.013.
Neira, R., 2010. Breeding in aquaculture species: genetic improvement programs in developing countries. In: 9th World Congress on Genetics Applied to Livestock Production. Leipzig, Germany, p. 8.
Ødegård, J., Moen, T., Santi, N., Korsvoll, S.A., Kjøglum, S., Meuwissen, T.H.E., 2014. Genomic prediction in an admixed population of Atlantic salmon (Salmo salar). Front. Genet. 5, 402. https://doi.org/10.3389/fgene.2014.00402.
Ohta, T., Kimura, M., 1971. Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. Genetics 68.
Perez-Enriquez, R., Robledo, D., Houston, R.D., Llera-Herrera, R., 2018. SNP markers for the genetic characterization of Mexican shrimp broodstocks. Genomics 110, 423–429. https://doi.org/10.1016/j.ygeno.2018.10.001.
Pfaff, C.L., Parra, E.J., Bonilla, C., Hiester, K., McKeigue, P.M., Kamboh, M.I., Hutchinson, R.G., Ferrell, R.E., Boerwinkle, E., Shriver, M.D., 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am. J. Hum. Genet. 68, 198–207. https://doi.org/10.1086/316935.
Ponzoni, R.W., Khaw, H.L., Nguyen, N.H., Hamzah, A., 2010. Inbreeding and effective population size in the Malaysian nucleus of the GIFT strain of Nile tilapia (Oreochromis niloticus). Aquaculture 302, 42–48. https://doi.org/10.1016/j.aquaculture.2010.02.009.
Prieur, V., Clarke, S.M., Brito, L.F., McEwan, J.C., Lee, M.A., Brauning, R., Dodds, K.G., Auvray, B., 2017. Estimation of linkage disequilibrium and effective population size in New Zealand sheep using three different methods to create genetic maps. BMC Genet. 18, 1–19. https://doi.org/10.1186/s12863-017-0534-2.
Pritchard, J.K., Przeworski, M., 2001. Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. https://doi.org/10.1086/321275.
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool

set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.
Ren, S., Mather, P.B., Tang, B., Hurwood, D.A., 2018. Levels of genetic diversity and inferred origins of Penaeus vannamei culture resources in China: implications for the production of a broad synthetic base population for genetic improvement. Aquaculture 491, 221–231. https://doi.org/10.1016/j.aquaculture.2018.03.036.
Rexroad, C.E., Vallejo, R.L., 2009. Estimates of linkage disequilibrium and effective population size in rainbow trout. BMC Genet. 10, 1–12. https://doi.org/10.1186/1471-2156-10-83.
Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2018. Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev. Aquac. 10, 670–682. https://doi.org/10.1111/raq.12193.
Rogers, A.R., Jorde, L.B., 1996. Ascertainment bias in estimates of average heterozygosity. Am. J. Hum. Genet. 58, 1033–1041.
Rousset, F., 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x.
Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J., Caballero, A., 2020. Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. Mol. Biol. Evol. 37, 3642–3653. https://doi.org/10.1093/molbev/msaa169.
Saunders, M.A., Slatkin, M., Garner, C., Hammer, M.F., Nachman, M.W., 2005. The extent of linkage disequilibrium caused by selection on G6PD in humans. Genetics 171, 1219–1229. https://doi.org/10.1534/genetics.105.048140.
Strandén, I., Garrick, D.J., 2009. Technical note: derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. J. Dairy Sci. 92, 2971–2975. https://doi.org/10.3168/JDS.2008-1929.
Sved, J.A., 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Popul. Biol. 2, 125–141. https://doi.org/10.1016/0040-5809(71)90011-6.
Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E., Visscher, P. M., 2007. Recent human effective population size estimated from linkage disequilibrium. Genome Res. 17, 520–526. https://doi.org/10.1101/gr.6023607.
Trang, T.T., Hung, N.H., Ninh, N.H., Knibb, W., Nguyen, N.H., 2019. Genetic variation in disease resistance against white spot syndrome virus (WSSV) in Liptopenaeus vannamei. Front. Genet. 10, 264. https://doi.org/10.3389/fgene.2019.00264.
Vallejo, R.L., Leeds, T.D., Gao, G., Parsons, J.E., Martin, K.E., Evenhuis, J.P., Fragomeni, B.O., Wiens, G.D., Palti, Y., 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, 1–13. https://doi.org/10.1186/s12711-017-0293-6.
Vallejo, R.L., Silva, R.M.O., Evenhuis, J.P., Gao, G., Liu, S., Parsons, J.E., Martin, K.E., Wiens, G.D., Lourenco, D.A.L., Leeds, T.D., Palti, Y., 2018. Accurate genomic predictions for BCWD resistance in rainbow trout are achieved using low-density SNP panels: evidence that long-range LD is a major contributing factor. J. Anim. Breed. Genet. 135, 263–274. https://doi.org/10.1111/jbg.12335.
Vela Avitúa, S., Montaldo, H.H., Márquez Valdelamar, L., Campos Montes, G.R., Castillo Juárez, H., 2013. Decline of genetic variability in a captive population of Pacific white shrimp Penaeus (Litopenaeus) vannamei using microsatellite and pedigree information. Electron. J. Biotechnol. 16, 9. https://doi.org/10.2225/vol16-issue4-fulltext-11.
Wang, Q., Yu, Y., Zhang, Q., Zhang, X., Yuan, J., Huang, H., Xiang, J., Li, F., 2019. A novel candidate gene associated with body weight in the Pacific White Shrimp Litopenaeus vannamei. Front. Genet. 10, 520. https://doi.org/10.3389/fgene.2019.00520.
Yáñez, J.M., Newman, S., Houston, R.D., 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress. Front. Genet. 6, 1–3. https://doi.org/10.3389/fgene.2015.00128.
Yáñez, J.M., Yoshida, G., Barria, A., Palma-Véjares, R., Travisany, D., Díaz, D., Cáceres, G., Cádiz, M.I., López, M.E., Lhorente, J.P., Jedlicki, A., Soto, J., Salas, D., Maass, A., 2020. High-throughput single nucleotide polymorphism (SNP) discovery and validation through whole-genome resequencing in Nile Tilapia (Oreochromis niloticus). Mar. Biotechnol. 22, 109–117. https://doi.org/10.1007/s10126-019-09935-5.
Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
Yoshida, G.M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J.P., Yáñez, J.M., 2018. Genomic prediction accuracy for resistance against Piscirickettsia salmonis in farmed rainbow trout. G3 Genes, Genomes, Genet. 8, 719–726. https://doi.org/10.1534/g3.117.300499.
Yoshida, G.M., Barria, A., Correa, K., Cáceres, G., Jedlicki, A., Cadiz, M.I., Lhorente, J.P., Yáñez, J.M., 2019a. Genome-wide patterns of population structure and linkage disequilibrium in farmed nile Tilapia (Oreochromis niloticus). Front. Genet. 10, 745. https://doi.org/10.3389/fgene.2019.00745.
Yoshida, G.M., Carvalheiro, R., Rodríguez, F.H., Lhorente, J.P., Yáñez, J.M., 2019b. Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. Genomics 111, 127–132. https://doi.org/10.1016/j.ygeno.2018.01.008.
Zhang, Xiaojun, Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., Liu, C., Wang, Q., Lv, X., Zhang, Xiaoxi, Ma, K.Y., Wang, X., Lin, W., Wang, Long, Zhu, X., Zhang, C., Zhang, J., Jin, S., Yu, K., Kong, J., Xu, P., Chen, J., Zhang, H., Sorgeloos, P., Sagi, A., Alcivar-Warren, A., Liu, Z., Wang, Lei, Ruan, J., Chu, K.H., Liu, B., Li, F., Xiang, J., 2019. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. Nat. Commun. 10, 1–14. https://doi.org/10.1038/s41467-018-08197-4.
Zhong, X., Li, Q., Kong, L., Yu, H., 2017. Estimates of linkage disequilibrium and effective population size in wild and selected populations of the pacific oyster using single-nucleotide polymorphism markers. J. World Aquac. Soc. 48, 791–801. https://doi.org/10.1111/JWAS.12393@10.1111/(ISSN)1749-7345.HARAPAC_2018.