SNP markers for the genetic characterization of Mexican shrimp broodstocks

SNP markers for the genetic characterization of Mexican shrimp *Litopenaeus vannamei* broodstocks

Ricardo Perez-Enriquez[1,*], Diego Robledo[2], Ross D. Houston[2], Raúl Llera-Herrera[3,4,*]

[1] Aquaculture Genetics & Breeding Lab, Centro de Investigaciones Biológicas del Noroeste, S.C. La Paz, Baja California Sur, Mexico 23096

[2] The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, EH25 9RG, UK

[3] CONACyT - Centro de Investigaciones en Alimentación y Desarrollo A.C. Sábalo-Cerritos s.n. Mazatlán, Sinaloa, Mexico, 82112

[4] Unidad Académica Mazatlán, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México. PO Box 811, CP 82040. Mazatlán, Sinaloa, Mexico.

* Corresponding authors: rperez@cibnor.mx; raul.llera@gmail.com

Abstract

Selective breeding of shrimp has major potential to enhance production traits, including growth and disease resistance. Genetic characterization of broodstock populations is a key element of breeding programs, as it enables decisions on inbreeding restrictions, family structure, and the potential use of genomic selection. Single Nucleotide Polymorphisms (SNPs) are suitable genetic

markers for this purpose. A set of SNPs was developed to characterize commercial breeding stocks in Mexico. Individuals from local and imported lines were selected for sequencing using the nextRAD technique, resulting in the identification of 2,619 SNPs. Genetic structure analysis showed three to five genetic groups of Ecuadorian and Mexican origins. A subset of 1,231 SNPs has potential for stock identification and management. Further, three SNPs were identified as candidate sex-linked markers. The role of SNPs possibly associated with genes related to traits of importance to shrimp farming, such as growth and immune response, should be further investigated.

Keywords: Single Nucleotide Polymorphisms; stock identification; genetic diversity; selective breeding; disease resistance

1. Introduction

Pacific whiteleg shrimp (*Litopenaeus vannamei*) is one of the most important aquaculture species globally, with more than three million tonnes produced annually (FAO, 2017). Mexico is one of the largest producers of whiteleg shrimp in Latin America, but the industry has been severely hampered by a high incidence of infectious disease outbreaks, particularly White Spot Syndrome Virus (WSSV; Esparza-Leal et al. 2012) and the bacterial Acute Hepatopancreatic Necrosis Disease (AHPND; Soto-Rodriguez et al. 2015). Up until the year of peak shrimp production (2009), most shrimp hatcheries relied on a breeding line called ‘Melagos’, originally imported from Venezuela in 1997 (Perez-Enriquez et al., 2009), that supported an annual production volume increase of ca. 9 % (CONAPESCA, 2013). While this growth performance was encouraging, the line was evidently highly susceptible to WSSV, resulting in a production volume decrease of more than 11 % per year in 2010 – 2012 (CONAPESCA, 2013).

As a strategy to help mitigate WSSV outbreaks, Mexican shrimp breeding companies, either independently or through the National Association of Shrimp Larvae Producers (ANPLAC), initiated the introduction of new broodstock in 2013. These stocks were mainly of Ecuadorian origin as the shrimp production in this country had recovered from catastrophic outbreaks of WSSV that started at the beginning of 2000 (Santana-Navarro, 2015), thought to be due to genetic resistance arising from the natural selection of survivors from disease-affected ponds (Lucien-Brun, 2017). Mexican official reports indicate imports from Ecuador in 2013 and 2015, the USA in 2013 and 2015 [which are of Ecuadorian origin (Gervais, 2014)], Nicaragua in 2014, and Colombia in 2015 [Sistema de Información Arancelaria Via Internet SIAVI (http://187.191.71.239/); Tariff fractions 03.06.26.01 and 03.06.27.01].

Genetic characterization of whiteleg shrimp *Litopenaeus vannamei* breeding lines in Mexico has previously been performed using microsatellite and mtDNA markers (Perez-Enriquez et al., 2009; Mendoza-Cano et al., 2013; Vela-Avitúa et al., 2013). However, these evaluations were focused on the 'Melagos' breeding line. Due to the multiple import events and subsequent mixing of broodstock, the genetic structure of the currently farmed Mexican shrimp populations is not well understood as no recent genetic profiling has been performed.

While some shrimp farming is based on the aforementioned selected broodstock lines, selective breeding programs for continuous genetic improvement of shrimp are at a formative stage in Mexico. A key element for these selective breeding programs is the characterization of genetic diversity and composition of stocks, as it will enable decisions on the composition of base populations, inbreeding restrictions, family structure, and the potential use of genomic selection (FAO, 1993). The genetic characterization will also aid in stock identification for traceability purposes.

Single Nucleotide Polymorphisms (SNPs) have been demonstrated to be suitable markers for the genetic characterization of broodstock populations, parentage assignment, and genome selection in several farmed aquaculture species [see review in Robledo et al. (2017)]. In *L. vannamei*, Perez-Enriquez and Max-Aguilar (2016) reported a SNP-based genotyping service for parentage assignment but the identity of the SNPs is not known. A commercially-available Illumina SNP-chip has been created (Jones et al., 2017). However, the SNPs on the array were discovered in a single-origin broodstock, and it might not represent the genetic diversity in broodstock populations derived from different sources. Genotyping by sequencing technology enables concurrent discovery and genotyping of genome-wide SNPs that are informative in the target populations at moderate scale and cost, and has been widely applied in aquaculture species for this purpose (Robledo et al. 2017).

The objective of this study was to assess the genetic diversity and constitution of the breeding lines currently used in the Mexican shrimp aquaculture industry and to develop a set of diagnostic SNP markers to identify those lines as a potential tool for stock management and traceability. To do so, we carried out Nextera-tagmented reductively-amplified DNA (nextRAD) sequencing in 95 individuals belonging to 21 breeding batches from 5 different hatcheries in the Northwest of Mexico. Our results provide a set of SNPs for shrimp Mexican population analyses and characterize current Mexican aquaculture shrimp stocks.

2. Materials and methods

2.1. Samples

A total of 95 *L. vannamei* individuals belonging to 21 breeding batches from five hatcheries located in Northwestern Mexico (3 – 7 individuals per breeding line) were sampled (Table 1). Those hatcheries, denoted 'A' – 'E' in the current study, represented the genetic makeup of most of the shrimp currently cultivated in Mexico based on the statistics generated by ANPLAC (National Larvae Producers Association, Mexico; pers. comm.). Information about the presumed origin of the founders was available for two hatcheries (Table 1). A pleopod was taken from each animal and kept in absolute ethanol until DNA isolation.

2.2. DNA sequencing

Approximately 20 mg of soft tissue were retrieved using forceps, and lysed in 500 µl of proteinase-K buffer (10 mM Tris, 50 mM NaCl, 5 mM $CaCl_2$, 50 µg mL^{-1} of Proteinase-K, pH 8.0) at 45°C overnight. The lysate was centrifuged, and the supernatant was incubated with 10 µl of silica glass milk suspension (Geneclean EZ, MP Biomedicals, Santa Ana, CA) for DNA binding. The suspension was centrifuged and washed twice in a washing solution (10 mM Tris, pH 7.5; 80 % ethanol), and the DNA was eluted in 20 µl of Milli-Q water. DNA was quantified by Qubit using the dsDNA broad range assay (Invitrogen, Palo Alto, CA), adjusted to 20 ng $µl^{-1}$ and shipped to SNPsaurus LLC (Eugene, OR) for nextRAD genotyping service. In brief, nextRAD involved the construction of individual Nextera libraries (Illumina), that were selectively amplified with modified multiplex Nextera primers that extend over insert regions with a restrictive 9 bp sequence in their 3' ends for a systematically reduced representation of the genome in the resulting libraries. Pooled libraries were sequenced in one lane of an Illumina HiSeq4000 platform, using the single-end 150 bp chemistry. Reads were demultiplexed and tags were removed by SNPsaurus using in-house scripts before reference construction and genotype calling.

111 *2.3. Data analysis*

112 The number of reads per individual sample ranged from 2.2 and 3.2 million, and were deposited
113 in NCBI-SRA (BioProject PRJNA492152, BioSample accessions SAMN10094370 -
114 SAMN10094464); each sample's identifier can be associated with the corresponding SRA
115 accession using the Supplemental Table 1. A reference Fasta file of single-end 27,196 fragments
116 of 150 bp length (Ref_shrimp.fasta; Supplemental File 1) was built by clustering the reads using
117 SNPSaurus' in-house scripts. Fastq files for each individual were analyzed for quality control
118 with the program FastQC version 0.11.2
119 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmomatic version 0.36
120 (Bolger et al., 2014) was used to trim low quality leading and trailing bases (Phred < 20; 4 bases
121 window) and the Nextera adapters. Only reads longer than 32 bp post-filtering were retained.

122 Trimmed sequences were aligned to the Ref_shrimp.fasta file using the Burrows-Wheeler
123 Aligner (bwa version 0.7.8; Li and Durbin, 2009). SAMtools version 1.6 (Li et al., 2009) was
124 used to generate the bam files and remove PCR duplicates. SNP identification (alignment and
125 base quality > 20, minimum depth 10×, maximum depth 150×) was performed using Samtools
126 (mpileup command) and the bcftools software (Li et al., 2009) version 1.4 (call command). The
127 obtained vcf file was filtered using vcftools version 0.1.15 (Danecek et al., 2011) with the
128 following criteria: minor allele frequency minimum = 0.016, minimum allele counts of the minor
129 allele = 3, minimum quality = 30, maximum missing sites = 0 %, indels removed, three SNPs or
130 fewer per fragment, and to only retain loci with two alleles. A BLAST search of the flanking
131 sequence of the obtained SNPs performed against the NCBI database and SNP regions matching
132 with mitochondrial DNA, repetitive regions of penaeid shrimps [e.g. ribosomal RNA (5S, 18S,

28S), non-LTR-like retrotransposons, and microsatellites] or of putatively exogenous DNA (bacteria) were removed.

Filtered vcf files were converted into Arlequin type files using PGDSpider version 2.1.1.3 (Lischer and Excoffier, 2012), treating each breeding stock as a separate population. Arlequin version 3.5 (Excoffier et al., 2005) was used to estimate allele frequencies and genetic diversity parameters (number of alleles per locus, observed and expected heterozygosity). Comparisons of these parameters among broodstocks were performed using t-Student tests. Significance for multiple pairwise tests was adjusted by the sequential Bonferroni correction (Rice, 1989). Cavalli-Sforza genetic distance among breeding lines with 1,000 bootstrap permutations were obtained to generate a Neighbor Joining tree with the software PHYLIP (Felsenstein, 1996), and visualized with the Interactive Tree Of Life tool v3 (Letunic and Bork, 2016).

Population genetic structure was inferred by a Bayesian model-based clustering method implemented by the software Structure version 2.3.4 (Pritchard et al., 2000). The model assumes the presence of K genetic clusters, each characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned probabilistically to clusters. First, the vcf files were converted to Structure type files with PGDSpider version 2.1.1.3. The running conditions for the Markov chain were 10,000 burnin steps and 100,000 repetitions for K = 1 to 12. The best K number of clusters was calculated by the Evanno method (Evanno et al., 2005) implemented in the Structure Harvester software version 0.6.94 (Earl and von Holdt, 2012). A Discriminant Analysis of Principal Component (DAPC) was also used to group broodstocks into putative genetic clusters using the Adegenet R package (Jombart, 2008), using the Structure file as input.

2.4. Candidate SNPs for broodstock identity and sex-linkage

Individual samples were classified into the K groups obtained from the Structure analysis. Pairwise F_{ST} between groups were calculated for each SNP using vcftools. Those SNPs showing high levels of between-group differentiation ($F_{ST} > 0.15$; Wright, 1978) were selected as candidates for group (and hence broodstock) membership identification. The candidate SNPs of the 10 pairwise analyses were merged into a single list.

Phenotypic sex was known for samples of broodstocks E-I – E-IV. These were used to perform an initial screen for putative candidate SNPs linked to sex. The 16 individuals of these stocks were pooled into a male group and a female group. F_{ST} analyses were performed between the two sexes for each SNP with vcftools. In this case the SNP set used for analysis was that obtained from the analysis with settings of 50 bp minimum separation between sites and a maximum of 30% of individuals without the site. Candidate SNPs were those being heterogametic in females and homogametic in males, consistent with the previously described ZW sex determining system (Zhang et al., 2007; Yu et al., 2017), and with $F_{ST} > 0.4$.

A BLAST search of the sequences containing the candidate broodstock-identity and sex-linked SNPs was performed using the current NCBI's nucleotide database, *Litopenaeus vannamei* sequences Ghaffari et al. (2014), Yu et al. (2015), and Jones et al. (2017), and *Penaeus monodon* sequences (Baranski et al. 2014) as reference databases. The databases of the two shrimp species were downloaded from the repositories of each of the journals.

3. Results

The initial variant calling analysis resulted in 398,125 putative SNPs. After applying the filtering criteria (remove indels, retain only biallelic loci, minimum MAF 0.016, SNP present in 100 % of

the individuals), and excluding the SNPs in repetitive regions (from BLAST) resulted in a filtered set of 2,619 SNPs distributed in 1,445 fragments (loci) that were used for downstream analyses.

The genetic diversity was significantly higher ($P < 0.025$) in the 6 lines of Hatchery B (B-I – B-VI) than in the rest of the lines (Fig. 1a-d). Hatcheries C and D showed the least values for both diversity indicators (Fig. 1a, c), and among these, the C-V line had the lowest values (Fig. 1b, d).

The Discriminant Analysis of Principal Components revealed three clearly differentiated genetic groups by using two linear discriminants that explained 92% of the variance (Fig. 2). At the hatchery level it can be seen that the stocks from hatcheries A and most of those from E belong to the same cluster. Likewise, all stocks from B are composed of individuals of a similar genetic makeup. Hatcheries C, D, and E maintain breeding lines of diverse origin.

The Bayesian clustering analysis showed a delta K bimodal shape (Supplemental Fig. 1). The first peak found at $K = 3$, was consistent with the DAPC analysis, but the best number of clusters was found at $K = 5$, indicating that beyond the three main groups (Fig. 3a), additional subdivision better explains the genetic structure (Fig. 3b). Four of the five genetic clusters would be composed by breeding lines from different hatcheries as shown in Table 2, highlighting diverse origin of the breeding lines within hatcheries. The other group is formed exclusively by all the breeding lines of Hatchery B, indicating a clear differentiation from the other hatcheries. Two breeding lines of Hatchery C seem to be a mixture of two different groups (Fig. 3b): C-I of C-III with C-V [groups 3 and 4, respectively (Table 2)]; and C-II of C-IV with C-V [groups 1 and 4, respectively (Table 2)].

The clustering pattern showed by the NJ tree based on Reynolds genetic distance was similar to that of the Bayesian clustering, with high bootstrap values for groups 1, 3, 4, and 5 (Fig. 4), but in this case C-I and C-II are clearly included in group 5, and therefore more related to C-V, rather than an admixture. A membership analysis with Structure at the individual level indicated that unexpected admixture within some breeding lines probably took place, such as individuals D-I-1, D-III-3, D-IV-2, and E-IV-2 (Supplemental Figure 2).

The pairwise F_{ST} analysis according to the groups of Table 2 and excluding individuals potentially coming from mixed stocks [D-I-1, D-III-3, D-IV-2, E-IV-2, C-I (1-4), C-II (1-4)], resulted in 1231 SNPs in 904 loci as candidates for breeding line identification (Supplemental Table 2).

The BLAST search of the SNP flanking sequences showed 26, 3, 151, 12, and 3 positive hits to the databases of NCBI, Ghaffari et al. (2014), Yu et al. (2015), Jones et al. (2017), and Baranski et al. (2014), respectively (Supplemental Table 3). Matches to genes from *Drosophila* or penaeids related to immune response, fatty acids synthesis, and growth, among others functions, were observed.

The analysis of potentially sex-linked SNPs resulted in a set of 50 loci showing $F_{ST} > 0.4$ between males and females, of which three loci are candidate sex-associated markers as they were exclusively heterozygous in females and homozygous in males (Supplemental Table 2); a linkage disequilibrium analysis indicated that the two loci are probably at the same chromosome ($r > 0.7$). None matched to LG18 of Yu et al. (2017).

4. Discussion

This is the first report, to our knowledge, on the characterization of the genetic makeup of shrimp hatchery-reared populations using SNPs as genetic markers. Previous reports have relied on microsatellites (e.g. Freitas and Galetti, 2005; Perez-Enriquez et al., 2009; Vela-Avitúa et al., 2013; Zhang et al., 2014; Ren et al., 2108) and mitochondrial DNA (Mendoza-Cano et al., 2013). Even though these markers have been adequate to assess the genetic diversity and structure of cultivated shrimp stocks, their potential to be used for other purposes, such as parentage analysis and genomic selection, is limited because a putative high mutation rate (particularly of microsatellites; Perez-Enriquez and Max-Aguilar, 2016), and a low genome coverage.

Two sets of SNP panels are currently available for *L. vannamei*. One, based on 164 SNPs derived from ESTs and offered as a genotyping service (J. Stannard, CTA, pers. comm.), has been used for parentage assignment in an study of heritability in growth and fatty acids content (Nolasco-Alzaga et al., 2018). The second one is a commercially-available Illumina SNP-chip (Jones et al., 2017), for which its usefulness to characterize breeding populations has not been reported. In both cases markers are located within expressed regions, and thus neutral variation might be underrepresented.

The SNP set developed in the present study comprises 2,619 SNPs distributed in 1,445 fragments (loci), and has clearly shown differential diversities of several *L. vannamei* breeding lines maintained in Mexico. Although sample sizes for each breeding stock were low, it was quite clear that the genetic diversity of the imported stocks of recent years was higher than that of the Mexican line (Group 4 in Table 2). Thus, it appears that the introduction of the new genetic variants can be beneficial to avoid a long term accumulation of inbreeding, which could hamper shrimp production traits (Moss et al., 2012).

241 Genetic differences among *L. vannamei* breeding lines were also clearly observed, and three to
242 five genetic clusters seem to be present in this sample of broodstocks used in the Mexican shrimp
243 industry. This provides a rather wide genetic background, either to inform baseline populations
244 or to test the performance of several types of between-line crosses. In China, Ren et al. (2018)
245 found enough genetic variation in 36 *L. vannamei* breeding stocks belonging to four genetic
246 clusters to inform the synthesis of a base population for future selective breeding.

247 The differentiation between the *L. vannamei* line of putatively Mexican origin with the other
248 lines is not surprising as it has founders from various origins and has been bred within-line
249 during more than 15 generations (Perez-Enriquez et al., 2009). However, the difference between
250 the *L. vannamei* Ecuadorian lines (B-I – B-VI) and those imported from Texas (lines of Group 1
251 in Table 2), which also originated from Ecuador (Gervais, 2014), is striking, but might be
252 explained by a previous mixing of WSSV-resistant animals from Ecuador and Panama (Gervais,
253 2014).

254 The acute differences among clusters found within the reduced set of 1,231 SNPs represent the
255 potential of having markers specific to *L. vannamei* breeding lines. This provides the opportunity
256 of devising a set of markers that can be used for stock traceability (in hatcheries, farms or even in
257 the market), for parentage analysis (e.g. Lapègue et al., 2014; Perez-Enriquez and Max-Aguilar,
258 2016; Bich Vo et al., 2018), and for the detection of hybrid stocks. For this purpose, individual-
259 locus assays (e.g. Taqman, KASP, integrated fluidic circuits) on a number of SNPs should be
260 tested for accuracy and repeatability in appropriate sample sizes. For the purpose of assessing
261 genetic variability, levels of inbreeding, and conformation to Hardy-Weinberg expectations, care
262 should be taken in the design of the SNP panel to avoid parameter biases.

Similarity searches against several published databases identified genes that may be related to traits of importance for shrimp aquaculture. In penaeids several cases of up- or down-regulation of genes related to the immune response during WSSV or *Vibrio parahaemolyticus* infections in comparison to negative controls have been reported [e.g. Core histone protein H4 (Feng et al., 2014); Argonaute-1, (Labreuche et al., 2010); Toll-4 (Ren et al., 2017); Vascular endothelial growth factor 1 (Wang et al., 2015); *Penaeus monodon* AV gene (Luo et al., 2007); Fatty acid synthase (Yang et al. 2011; Zuo et al., 2017)]. Large differences in the abundance of transcripts of the Hypoxia inducible factor 1 alpha were detected in *L. vannamei* under normoxia and hypoxia (Soñanez-Organis et al., 2009). In *P. monodon*, Magerd et al. (2013) reported the role of the Thrombospondin II gene in the induction of the sperm acrosome reaction. A sequence reported by Ghaffari et al. (2014) appears to be related to the ecdysis, and thus, probably with growth.

A sex determination region in *L. vannamei* has been proposed to be in a small section of a linkage group (LG18), in which a sex-associated marker has been found (Yu et al. 2017), supporting a ZW/ZZ sex determination mechanism previously reported (Zhang et al., 2007). The same sex determination system has been reported in *P. monodon* (Robinson et al. 2014). In our study, two loci were also heterogametic in females and homogametic in males, although they did not match to known sequences on LG18, nor to any public sequence in the Blast search. Further analyses of these two loci with a larger sample size are needed to confirm their association with the sex determination system.

In summary, we have developed a set of SNP markers which can effectively characterize the genetic diversity and composition of broodstocks introduced to Mexico during 2013-2015. A subset of these markers are potentially useful for stock identification, some of which were found

286 within genes of potential interest for the improvement of shrimp farming traits, such as immune
287 response, growth, or reproductive performance.

288

289 **Acknowledgements**

290 The work was supported by the Instituto Nacional de la Pesca, Mexico (INAPESCA, PIDETEC
291 Granting Program) to RLl-H. Thanks are due to Comisión Nacional de Acuacultura y Pesca,
292 Mexico (CONAPESCA), to the participant hatchery directors and managers that contributed with
293 the samples, and to the National Association of Shrimp Larvae Producers (ANPLAC) for their
294 support. RP-E received a CONACYT (Mexico) fellowship and facilities access from The Roslin
295 Institute during a sabbatical leave. RDH and DR are supported by the Biotechnology and
296 Biological Sciences Research Council (BBSRC), Institute Strategic Programme Grants
297 (BBS/E/D/20002172, BBS/E/D/30002275), and DR is supported by a Royal Society Newton
298 International Fellowship (NF160037).

299

300 **Literature Cited**

301 Baranski M., Gopikrishna G., Robinson N.A., Katneni V.K., Shekhar M.S., Shanmugakarthik J.,
302 Jothivel S., Gopal C., Ravichandran P., Kent M., Arnyasi M., Ponniah A.G. 2014. The
303 development of a high density linkage map for black tiger shrimp (*Penaeus monodon*) based on
304 cSNPs. PLoS ONE 9(1), e85413. doi:10.1371/journal.pone.0085413

305 Bich Vo T.T., Nguyen H.D., Nguyen T.T.H, Tran H.T.T. 2018. Identification and analysis of
306 snps in population of Vietnamese catfish (*Pangasianodon hypophthalmus*), using next generation

sequencing and snp validation. MOJ Current Research & Reviews 1, 12–19. DOI: 10.15406/mojcrr.2018.01.00003

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

CONAPESCA. 2013. Anuario Estadístico de Acuacultura y Pesca, Edición 2013. Comisión Nacional de Acuacultura y Pesca. Mazatlán, Sinaloa, México.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R., Lunter G., Marth G., Sherry S.T., McVean G., Durbin R. and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics, 27, 2156–2158.

Earl D.A., von Holdt B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4, 359–361. doi: https://doi.org/10.1007/s12686-011-9548-7

Esparza-Leal H.M., Magallón-Barajas F.J., Pérez-Enriquez R., Casillas-Hernández R., Cabanillas-Ramos J.A., Valenzuela-Quiñónez W. 2012. Región endémica y regímenes de infección con el Virus del Síndrome de la Mancha Blanca (WSSV) en las granjas camaronícolas del noroeste de México. Ra Ximhai 8, 117–129.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14, 2611–2620.

Excoffier L., Laval G., Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47–50.

FAO. 1993. FAO Report of the expert consultation on utilization and conservation of aquatic genetic resources. FAO Fisheries Report No. 491, Food and Agriculture Organization of the United Nations, Rome.

FAO. 2017. FAO yearbook. Fishery and Aquaculture Statistics. 2015/FAO annuaire. Statistiques des pêches et de l'aquaculture. 2015/FAO anuario. Estadísticas de pesca y acuicultura. 2015. Rome/Roma, Italy/Italie/Italia.

Felsenstein J., 1993. PHYLIP (Phylogeny Inference Package, Version 3. 5). Department of Genetics, University of Washington, Seattle, Washington.

Feng J., Tang X., Zhan W. 2014. Analysis and identification of tyrosine phosphorylated proteins in hemocytes from *Fenneropenaeus chinensis* (Decapoda: Penaeidae) infected with white spot syndrome virus. Journal of Crustacean Biology 34, 453–459. doi: 10.1163/1937240X-00002241

Freitas P.D., Galetti Jr. P.M. 2005. Assessment of the genetic diversity in five generations of a commercial broodstock line of *Litopenaeus vannamei* shrimp. African Journal of Biotechnology. 4, 1362–1367.

Gervais N. 2014. Neil Gervais comenta sobre los reproductores de PRIMO Broodstock. Industria Acuícola 11(1), 14.

Ghaffari N., Sanchez-Flores A., Doan R., Garcia-Orozco K.D., Chen P.L., Ochoa-Leyva A., Lopez-Zavala A.A., Carrasco J.S., Hong C., Brieba L.G., Rudiño-Piñera E., Blood P.D., Sawyer J.E., Johnson C.D., Dindot S.V., Sotelo-Mundo R.R., Criscitiello M.F. 2014. Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. Scientific Reports 4, 7081. doi: 10.1038/srep07081.

Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. doi:10.1093/bioinformatics/btn129

Jones D.B., Jerry D.R., Khatkar M.S., Raadsma H.W., Steen H.V., 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*. Scientific Reports 7, 10360. doi: 10.1038/s41598-017-10515–7.

Labreuche Y., Veloso A., de la Vega E., Gross P.S., Chapman R.W., Browdy C.L., Warr G.W. 2010. Non-specific activation of antiviral immunity and induction of RNA interference may engage the same pathway in the Pacific white leg shrimp *Litopenaeus vannamei*. Developmental and Comparative Immunology 34, 1209–1218. doi:10.1016/j.dci.2010.06.017

Lapègue S., Harrang E., Heurtebise S., Flahauw E., Donnadieu C., Gayral P., Ballenghien M., Genestout L., Barbotte L., Mahla R., Haffray P., Klopp C. 2014. Development of SNP genotyping arrays in two shellfish species. Molecular Ecology Resources 14, 820–830. doi: 10.1111/1755-0998.12230

Letunic I., Bork P. 2016. Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research44 W1, W242-245. doi: 10.1093/nar/gkw290

Li H., Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map

(SAM) format and SAMtools. Bioinformatics 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Lischer H.E.L., Excoffier L. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299. doi: 10.1093/bioinformatics/btr642

Lucien-Brun H. 2017. El cultivo camaronero de Ecuador, una historia de éxito. Industria Acuícola 14.1, 6–10.

Luo T., Li F., Lei K., Xua X. 2007. Genomic organization, promoter characterization and expression profiles of an antiviral gene *PmAV* from the shrimp *Penaeus monodon*. Molecular Immunology 44, 1516–1523. doi:10.1016/j.molimm.2006.09.015

Magerd S., Asuvapongpatana S., Vanichviriyakit R., Chotwiwatthanakun C., Weerachatyanukul W. 2013. Characterization of the thrombospondin (TSP)-II gene in *Penaeus monodon* and a novel role of TSP-like proteins in an induction of shrimp sperm acrosome reaction. Molecular Reproduction & Development 80, 393–402. doi: 10.1002/mrd.22173

Mendoza-Cano F., Grijalva-Chon J.M., Perez-Enriquez R., Ramos-Paredes J., Varela-Romero A. 2013. Genetic diversity of mitochondrial DNA from *Litopenaeus vannamei* broodstock used in northwestern Mexico. Ciencias Marinas 39, 401–412. http://dx.doi.org/10.7773/cm.v39i4.2269

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. Journal of Invertebrate Pathology 110, 247–250. doi: 10.1016/j.jip.2012.01.013

Nolasco-Alzaga H.R., Perez-Enriquez R., Palacios-Mechetnov E., Haffray P. 2018. Quantitative genetic parameters of growth and fatty acid content in the hemolymph of the Whiteleg shrimp *Litopenaeus vannamei*. Aquaculture 482, 17-23. https://doi.org/10.1016/j.aquaculture.2017.09.015

Perez-Enriquez R., Max-Aguilar A. 2016. Pedigree traceability in whiteleg shrimp (*Litopenaeus vannamei*) using genetic markers: a comparison between microsatellites and SNPs. Ciencias Marinas 42, 227–235. http://dx.doi.org/10.7773/cm.v42i4.2662

Perez-Enriquez R., Hernández-Martínez F., Cruz P. 2009. Genetic diversity status of White shrimp *Penaeus* (*Litopenaeus*) *vannamei* broodstock in Mexico. Aquaculture 297, 44–50. doi:10.1016/j.aquaculture.2009.08.038

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Ren Q., Huang X., Cui Y., Sun J., Wang W., Zhang X. 2017. Two White Spot Syndrome Virus microRNAs target the dorsal gene to promote virus infection in *Marsupenaeus japonicus* shrimp. Journal of Virology 91, e02261-16. doi: 10.1128/JVI.02261-16.

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. doi https://doi.org/10.1016/j.aquaculture.2018.03.036

Rice W.R. 1989). Analyzing tables of statistical tests. Evolution 41, 223–235.

Robinson N.A., Gopikrishna G., Baranski M., Katneni V.K., Shekhar M.S., Shanmugakarthik J., Jothivel S., Gopal C., Ravichandran P., Gitterle T., Ponniah A.G. 2014. QTL for white spot syndrome virus resistance and the sex-determining locus in the Indian black tiger shrimp (*Penaeus monodon*). BMC Genomics 15, 731.

Robledo D., Palaiokostas C., Bargelloni L., Martínez P., Houston R. 2017. Applications of genotyping by sequencing in aquaculture breeding and genetics. Reviews in Aquaculture https://doi.org/10.1111/raq.12193

Santana Navarro N. 2015. Factores de auge, crisis y recuperación del sector camaronero ecuatoriano en el periodo 1994 – 2014. Master in Sciences Thesis. University of Guayaquil, Ecuador. Available at: http://repositorio.ug.edu.ec/handle/redug/7158 (accessed on June 21, 2018).

Soñanez-Organis J.G.1, Peregrino-Uriarte A.B., Gómez-Jiménez S., López-Zavala A., Forman H.J., Yepiz-Plascencia G. 2009. Molecular characterization of hypoxia inducible factor-1 (HIF-1) from the white shrimp *Litopenaeus vannamei* and tissue-specific expression under hypoxia. Comparative Biochemistry and Physiology Part C 150, 395–405. doi: 10.1016/j.cbpc.2009.06.005

Soto-Rodriguez S.A., Gomez-Gil B., Lozano-Olvera R, Betancourt-Lozano M., Morales-Covarrubias M.S. 2015. Field and experimental evidence of *Vibrio parahaemolyticus* as the causative agent of acute hepatopancreatic necrosis disease of cultured shrimp (*Litopenaeus vannamei*) in northwestern Mexico. Applied Environmental Microbiology 81, 1689–1699.

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. Electronic Journal of Biotechnology 16(4).

Wang Z., Li S., Li F.3, Yang H., Yang F., Xiang J. 2015. Characterization of two types of vascular endothelial growth factor from *Litopenaeus vannamei* and their involvements during WSSV infection. Fish & Shellfish Immunology 47, 824–832. doi: 10.1016/j.fsi.2015.10.026.

Wright S. 1978. Evolution and the Genetics of Populations. Volume 4, Variability within and among natural populations. The University of Chicago Press. Chicago.

Yang L., Liu J., Liu M., Qian M., Zhang M., Hu H. 2011. Identification of fatty acid synthase from the Pacific white shrimp, *Litopenaeus vannamei* and its specific expression profiles during White Spot Syndrome Virus infection. Fish & Shellfish Immunology 30, 744–749. doi: 10.1016/j.fsi.2010.12.026

Yu Y., Zhang X., Yuan J., Li F., Chen X., Zhao Y., Huang L., Zheng H., Xiang J. 2015. Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp, *Litopenaeus vannamei*. Scientific Reports 5, 15612. doi: 10.1038/srep15612

Yu Y., Zhang X., Yuan J., Wang Q., Li S., Huang H., Li F., Xiang J., 2017. Identification of sex-determining loci in Pacific white shrimp *Litopeneaus vannamei* using linkage and association analysis. Marine Biotechnology 19, 277. https://doi.org/10.1007/s10126-017-9749-5

Zhang L., Yang C., Zhang Y., Li L., Zhang X., Zhang Q., Xiang J. 2007. A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates. Genetica 131, 37–49. doi 10.1007/s11802-014-2208-2

453 Zhang K., Wang W., Li W., Zhang Q., Kong J. 2014. Analysis of genetic diversity and
454 differentiation of seven stocks of *Litopenaeus vannamei* using microsatellite markers. Journal of
455 Ocean University of China 13, 647–656.

456 Zuo H., Gao J., Yuan J., Deng H., Yang L., Weng S., He J., Xu X. 2017. Fatty acid synthase
457 plays a positive role in shrimp immune responses against *Vibrio parahaemolyticus* infection.
458 Fish & Shellfish Immunology 60, 282–288. doi: 10.1016/j.fsi.2016.11.054.

459

Figures captions

Fig. 1. Genetic diversity. a) and b) Number of alleles per locus per hatchery and per breeding line, respectively; c) and d) Expected heterozygosity per hatchery and per breeding line, respectively.

Fig. 2. Discriminant analysis of principal components (DAPC) for 95 genotypes of shrimp breeding lines in Mexico. The axes represent the first two Linear Discriminants (LD). Each circle represents a cluster and each dot represents an individual. Legends represent the breeding lines of the individuals.

Fig. 3. Cluster memberships of breeding lines from the Bayesian genetic structure analysis. a) K = 3; b) K = 5.

Fig. 4. Neighbor-Joining tree based on the Cavalli-Sforza genetic distance among the breeding lines. Numbers along color stripes correspond to the groups of Table 3. Bootstrap percentages shown in the nodes.

Supplementary Figures

Supplementary Fig. 1. Change rate (delta K) for each value of K obtained from the Evanno method (Evanno et al. 2005) in the Structure analysis.

Supplementary Fig. 2. Individual membership to genetic clusters of the 95 samples from the Bayesian genetic structure analysis at K = 5.

Table 1. Number of individuals obtained from 21 shrimp breeding batches sampled in Mexico in 2016.

Broodstock	Batch	Sample size	Founder ID[a]
A	I	4	NA
	II	4	NA
B	I	6	EC-1
	II	6	EC-2
	III	6	EC-3
	IV	5	EC-4
	V	6	EC-5
	VI	7	EC-6
C	I	4	NA
	II	4	NA
	III	4	NA
	IV	4	NA
	V	4	NA
D	I	4	CA-1
	II	4	MX
	III	3	TX
	IV	4	CA-2
E	I	4[b]	NA
	II	4[b]	NA
	III	4[b]	NA
	IV	4[b]	NA

[a] Origin of founder stocks: NA: Not available; EC-1 - EC-6: Ecuador; CA-1 and CA-2: Central America; TX: Texas; MX: Mexico

[b] Sex proportion 1:1

479

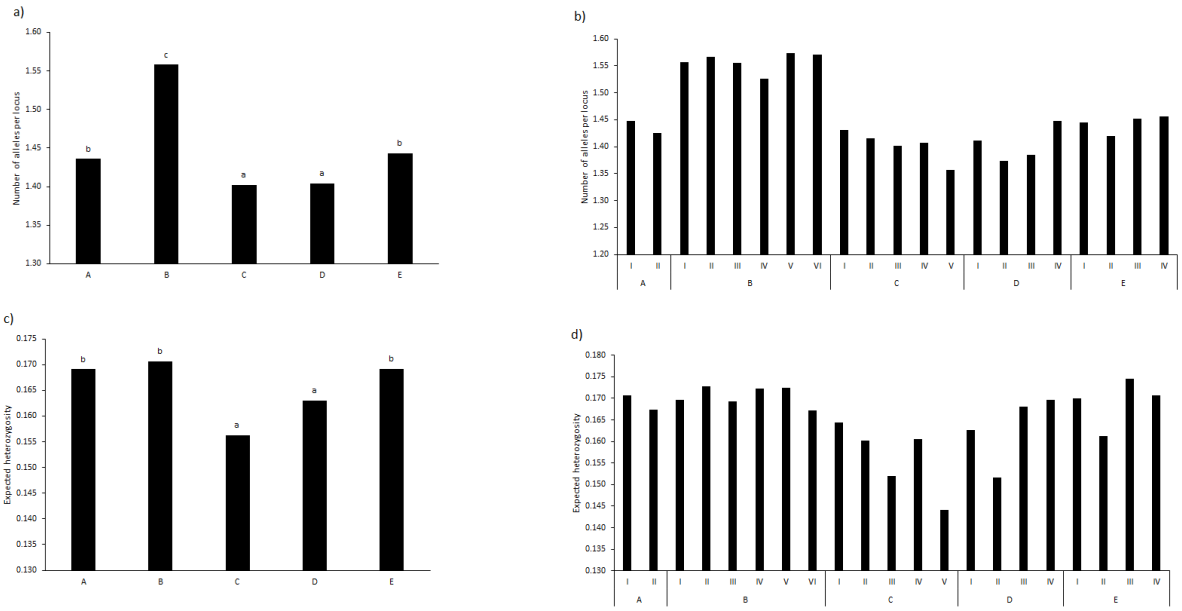
480

Table 2. Proposed genetic groups based on the Bayesian clustering analysis.

Group	Member breeding lines	Putative origin[a]
1	A-I, A-II, C-IV, DIII, E-I, E-II, E-III	TX
2	B-I, B-II, B-III, B-IV, B-V, B-VI	EC
3	C-III, D-I	CA-1
4	C-V, D-II	MX
5	D-IV, E-IV	CA-2

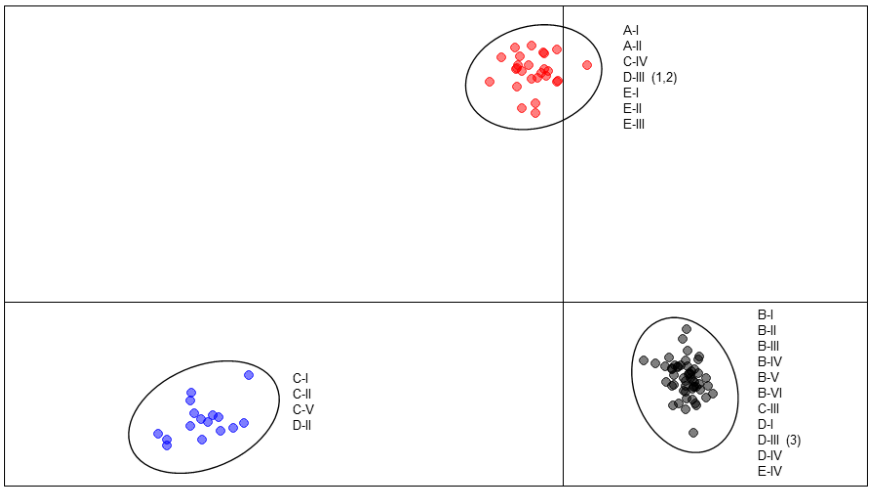
481 [a] Nomenclature according to Table 1

482


a)
Number of alleles per locus
1.60
1.55
1.50
1.45
1.40
1.35
1.30
b
c
a
a
b
A
B
C
D
E
b)
Number of alleles per locus
1.60
1.55
1.50
1.45
1.40
1.35
1.30
1.25
1.20
I II I II III IV V VI I II III IV V I II III IV I II III IV
A
B
C
D
E
c)
Expected heterozygosity
0.175
0.170
0.165
0.160
0.155
0.150
0.145
0.140
0.135
0.130
b
b
a
a
b
A
B
C
D
E
d)
Expected heterozygosity
0.180
0.175
0.170
0.165
0.160
0.155
0.150
0.145
0.140
0.135
0.130
I II I II III IV V VI I II III IV V I II III IV I II III IV
A
B
C
D
E

483

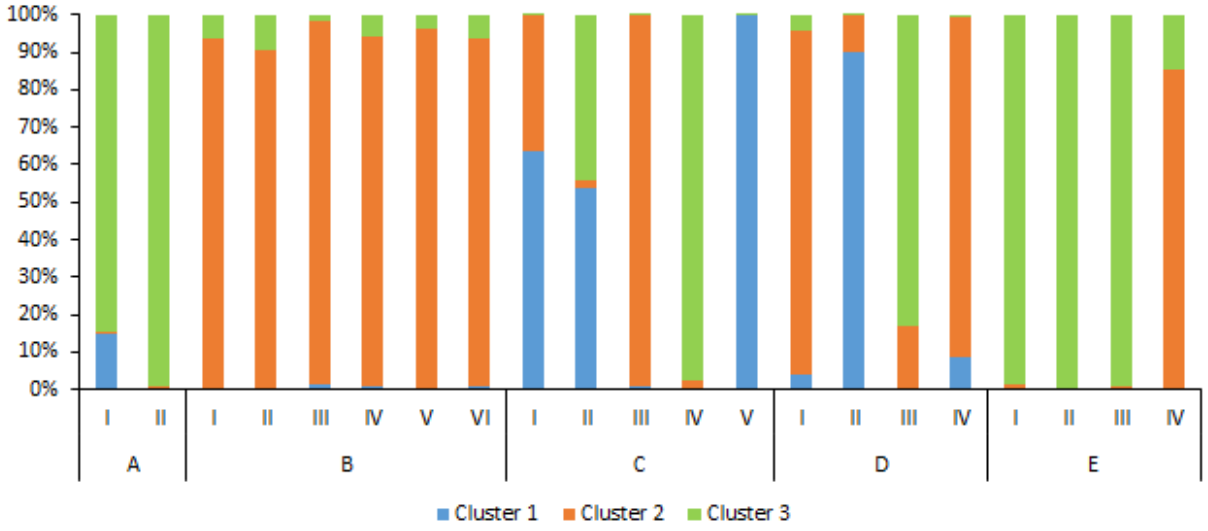
484


A-I
A-II
C-IV
D-III (1,2)
E-I
E-II
E-III
C-I
C-II
C-V
D-II
B-I
B-II
B-III
B-IV
B-V
B-VI
C-III
D-I
D-III (3)
D-IV
E-IV

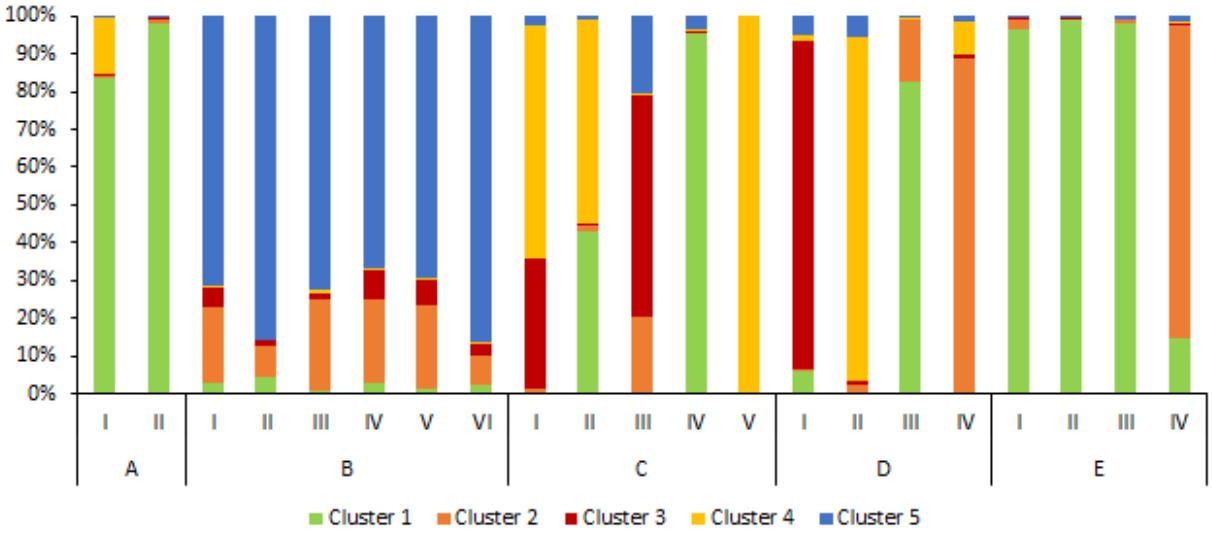
485

486

a)

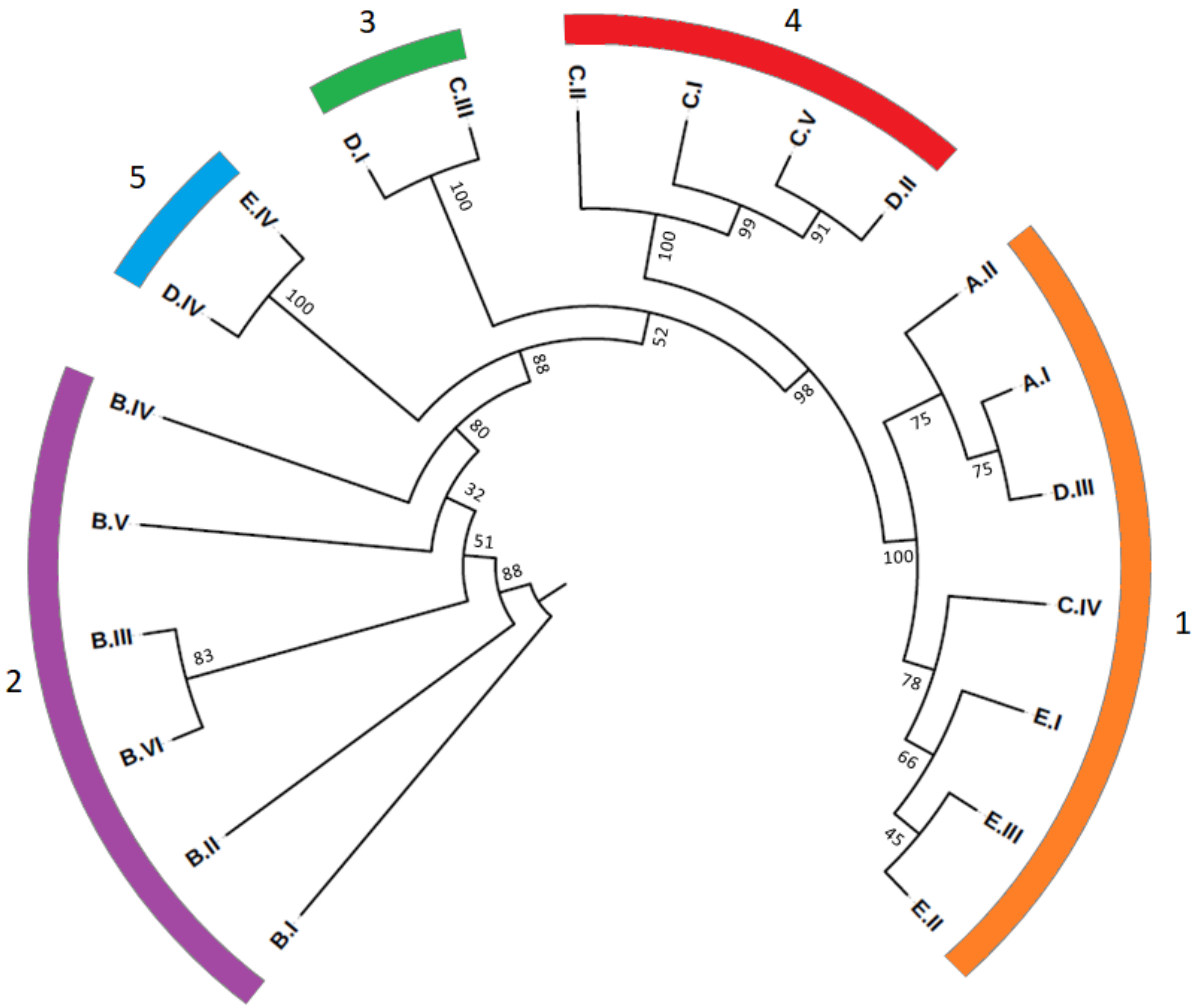

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
I II I II III IV V VI I II III IV V I II III IV I II III IV
A B C D E
Cluster 1 Cluster 2 Cluster 3

b)


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
I II I II III IV V VI I II III IV V I II III IV I II III IV
A B C D E
Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

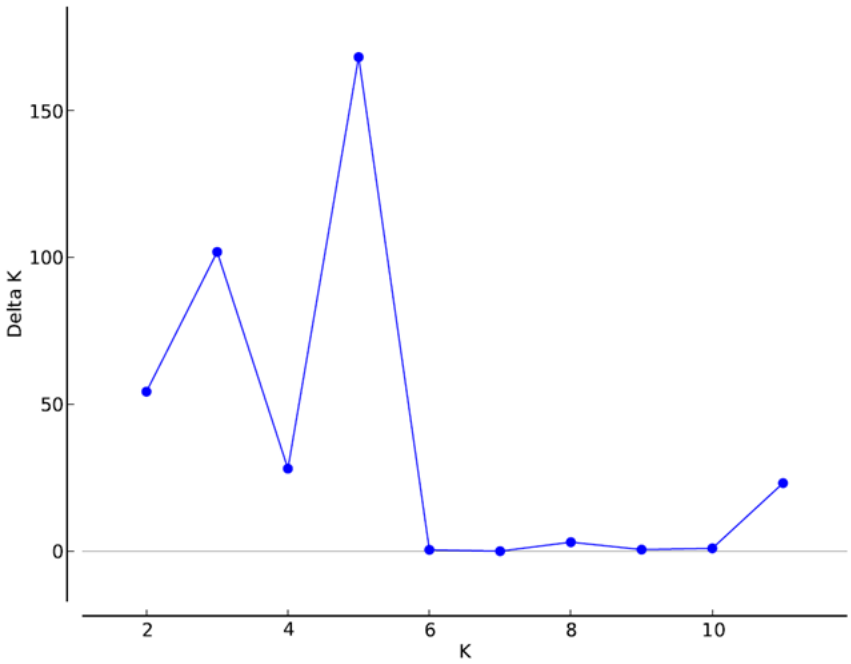
487

488


3
4
5
2
1
C.III
D.I
C.II
C.I
C.V
D.II
E.IV
D.IV
A.II
A.I
D.III
B.IV
B.V
B.III
B.VI
B.II
B.I
C.IV
E.I
E.III
E.II
100
52
98
88
80
32
51
88
100
100
99
91
100
75
75
83
78
66
45

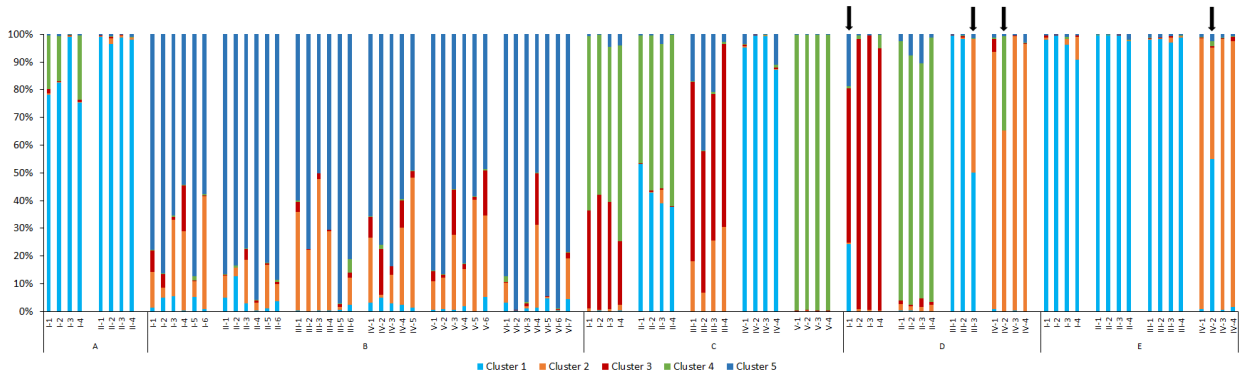
489

490



491

492


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A
B
C
D
E
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5

493

494